



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


Andréa Maria de Oliveira

**Estudo da ancestralidade paterna em amostras de populações
do Estado do Rio de Janeiro e do Oeste Africano:
“uma dinâmica populacional”**

Rio de Janeiro
2015

Andréa Maria de Oliveira

**Estudo da ancestralidade paterna em amostras de populações
do Estado do Rio de Janeiro e do Oeste Africano:
“uma dinâmica populacional”**



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

Orientadora: Dra. Leonor Gusmão

Coorientador: Prof. Dr. Elizeu Fagundes de Carvalho

Rio de Janeiro

2015

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

O48 Oliveira, Andréa Maria de.
Estudo da ancestralidade paterna em amostras de populações do Estado do Rio de Janeiro e do Oeste Africano: “uma dinâmica populacional” / Andréa Maria de Oliveira. – 2014.
175 f.

Orientadora: Leonor Gusmão.

Coorientador: Elizeu Fagundes de Carvalho.

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

1. África – Teses. 2. Rio de Janeiro (RJ) – População - Teses. 3. Grupos Populacionais - Genética. 4. Variações Genéticas. 5. Genética das Populações - Teses. I. Gusmão, Leonor. II. Carvalho, Elizeu Fagundes de. III. Universidade do Estado do Rio de Janeiro. Instituto de Biologia Roberto Alcântara Gomes. IV Título.

CDU 515.1

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

Assinatura

Data

Andréa Maria de Oliveira

**Análise Estudo da ancestralidade paterna em amostras de populações
do Estado do Rio de Janeiro e do Oeste Africano: “uma dinâmica
populacional”**

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

Coorientador: Prof. Dr. Elizeu Fagundes de Carvalho

Instituto de Biologia Roberto Alcântara Gomes – UERJ

Banca Examinadora:

Dra. Leonor Gusmão (Orientadora)

Instituto de Biologia Roberto Alcântara Gomes – UERJ

Prof.^a Dra Cintia Barros Santos Rebouças

Instituto de Biologia Roberto Alcântara Gomes – UERJ

Prof. Dr. Pedro Hernan Cabello Acero

Fundação Oswaldo Cruz

Prof.Dr. Milton Ozório Moraes

Fundação Oswaldo Cruz

Rio de Janeiro

2015

AGRADECIMENTOS

À Universidade do Estado do Rio de Janeiro, por incentivar a qualificação dos seus servidores;

Ao Laboratório de Diagnósticos por DNA, representado por seu Coordenador, o Prof. Dr. Elizeu Fagundes de Carvalho e sua equipe, por fornecer as condições necessárias para a execução deste trabalho;

Ao Instituto de Patologia e Imunologia da Universidade do Porto, representado pelo Prof. Dr. António Amorim e pela Prof^a. Dr^a. Leonor Gusmão, por me receberem e por permitirem a execução de parte desta Tese;

Ao Instituto de Medicina Legal da Universidade de Colônia, representado pelo Prof. Dr. Peter Schneider e pela Dr^a Iva Gomes, por me receberem para a execução de parte deste trabalho;

À FAPERJ, que no ano de 2013, designou verba referente à aprovação de Projeto APQ1, para a execução de parte deste trabalho;

À Universidade Federal de Ouro Preto (UFOP), representada pelo Prof. Dr George Luiz Lins Machado Coelho, Vice-Diretor da Escola de Medicina e sua equipe, pela recepção calorosa e pela colaboração inestimável neste trabalho;

Ao Prof. Dr. Antenor Rodrigues Barbosa Júnior, que pela amizade desinteressada e confiança no nosso trabalho, abriu as portas da Universidade Federal de Ouro Preto (UFOP).

RESUMO

OLIVEIRA, Andréa Maria de. **Estudo da ancestralidade paterna em amostras de populações do Estado do Rio de Janeiro e do Oeste Africano: “uma dinâmica populacional”**. 2014. 175 f. Tese (Doutorado em Biociências) - Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2014.

O uso de marcadores do tipo STR e SNP tem se revelado de grande importância na discriminação entre indivíduos de uma mesma população, assim como para estudos evolutivos. A utilização de um conjunto de 17 STRs e 46 SNPs específicos de cromossomo Y permitiu a caracterização de um conjunto de amostras representativas das populações do Rio de Janeiro e do oeste africano, com uma avaliação mais ampla sobre a ancestralidade de origem paterna. Na primeira parte deste estudo foram analisados 605 indivíduos do sexo masculino do estado do Rio de Janeiro. Como resultado, não foram observadas diferenças significativas entre as populações do sudeste e do Rio de Janeiro, que apresentou uma alta diversidade de haplótipos ($0,9999 \pm 0,0001$) e de haplogrupos ($0,7589 \pm 0,0171$). A comparação da população miscigenada do Rio de Janeiro com diferentes grupos étnicos ou populacionais mostrou que a frequência de indivíduos com marcadores tipicamente Europeus é de 77%, africanos é de 14,87% e em ameríndios é de 2,31%. A segunda parte do estudo revelou uma grande diversidade haplotípica ($1,0000 \pm 0,0018$) numa amostra do Oeste africano. Quanto ao valor da diversidade de haplogrupos ($0,6895 \pm 0,0200$), este foi similar aos observados em populações de origem Bantu do oeste e centro africanos, principalmente de Benin, Nigéria e Costa do Marfim. A terceira parte deste estudo mostrou que não existem diferenças significativas entre o componente africano da amostra do Rio de Janeiro e as populações africanas do sudeste, oeste e centro oeste. Por outro lado, observamos diferenças significativas quando comparamos o componente africano do Rio de Janeiro e o oeste africano com populações de Uganda, Quênia e África do Sul. A ampliação de estudos genéticos nas populações da África se fazem necessários para o entendimento da diversidade genética no mundo. Este trabalho contribuiu para fornecer mais alguns dados genéticos, que podem ser somados aos estudos mundiais que estão sendo realizados, ampliando os nossos conhecimentos sobre a formação das populações que também foram influenciadas pelo fenômeno da Diáspora Africana.

Palavras-chave: África. Y-STR. Y-SNP. Estudos populacionais. Ancestralidade.

ABSTRACT

OLIVEIRA, Andréa Maria de. **Study of paternal ancestry in samples from Rio de Janeiro and West Africa: "a populational dynamic"**. 2014. 175 f. Tese (Doutorado em Biociências) - Instituto de Biologia Roberto Alcantara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2014.

The use of STR and SNP markers has proved to be of great importance to discriminate individuals of the same population as well as for evolutionary studies. The use of a set of 17 STRs and 46 SNPs specific from the Y chromosome allowed the characterization of a group of samples representative of the Rio de Janeiro and the West African populations, with a deep assessment of the paternal ancestry. The first part of this study focused in the analysis of 605 males of the state of Rio de Janeiro. The results showed no significant differences between the Brazilian southeastern populations and Rio de Janeiro, which showed high values of haplotype (0.9999 ± 0.0001) and haplogroup (0.7589 ± 0.0171) diversities. The second part of the study revealed a high haplotype diversity (1.0000 ± 0.0018) in a sample from West Africa. The value of haplogroup diversity (0.6895 ± 0.0200) was similar to those previously seen in the West and Center African Bantu populations, mainly from Benin, Nigeria and Ivory Coast. The third part of this study showed no significant differences between the African component of our sample from Rio de Janeiro and the Southeastern, Western and Midwestern African populations. On the other hand, significant differences were observed when comparing the samples of the African component in Rio de Janeiro and of West Africa with population samples from Uganda, Kenya and South Africa. The increase of genetic studies in African populations is important for a better understanding of world genetic diversity. This work provided new genetic data, which can be added to those already available, expanding our knowledge about the formation of other populations also influenced by the African Diaspora.

Keywords: Africa. Y-STR. Y-SNP. Population studies. Ancestry.

LISTA DE FIGURAS

Figura 1 –	Exemplificação das possíveis rotas de chegada ao continente americano.....	19
Figura 2 –	Diáspora africana transatlântica.....	25
Figura 3 –	Mapa da África atual.....	27
Figura 4 –	Mapa esquemático da África - Separação em quadro humano e econômico.....	27
Figura 5 –	Comparações esquemáticas da diversidade genética na África e no mundo.....	28
Figura 6 –	Quadro com os quatro grupos etnolinguísticos da África, com alguns subgrupos representados.....	29
Figura 7 –	Famílias de línguas africanas e o domínio do grupo Bantu.....	30
Figura 8 –	Mapa esquemático indicando as direções da expansão/dispersão Bantu.....	32
Figura 9 –	Localização aproximada das 16 zonas Bantu descritas por Guthrie (1948).....	33
Figura 10 –	Estrutura do cromossomo Y - regiões pseudo-autossômicas (PAR1 e PAR2); região não recombinante (NRY) ou região específica masculina do cromossomo Y.....	43
Figura 11 –	Representação gráfica da distribuição dos haplogrupos do cromossomo Y, indicando as regiões e onde a frequência é mais elevada.....	50
Figura 12 –	Exemplo de alterações filogenéticas a partir de estudos do haplogrupo E1b1.....	54
Figura 13 –	Mapa indicando a origem da amostra incluída e analisada neste estudo: Oeste da África.....	64
Figura 14 –	Representação gráfica (eletroferograma) do resultado da eletroforese capilar dos produtos de amplificação de uma amostra controle (DNA 007), utilizando o kit comercial AmpFISTR® Y-Filer™.....	67
Figura 15 –	Representação esquemática dos multiplexes utilizados e a estratégia seguida para as genotipagens das amostras do Rio	

	de Janeiro e oeste da África.....	68
Figura 16 –	Visualização da amplificação da inserção do tipo Alu (YAP) em gel de agarose a 2%, corado com brometo de etídeo e visualizado em <i>scanner</i> de fluorescência.....	69
Figura 17 –	Gel de poliacrilamida a 6% corado com nitrato de prata, mostrando a análise do SNP M269, após digestão com enzima de restrição <i>MVAI</i>	71
Figura 18 –	Minissequenciamento com reação de SNaPshot, baseada em extensão de base única (SBE) com dideoxynucleotídeos marcados com corantes fluorescentes, que se anelam imediatamente após o final do oligonucleotídeo.....	72
Figura 19 –	Exemplo de genotipagem de Y-SNP, visualizada após eletroforese capilar, nas condições descritas em material e métodos, utilizando o multiplex Q.....	74
Figura 20 –	Distribuição de haplogrupos encontrados na amostra da população do Rio de Janeiro, separados por provável origem étnica.....	128
Figura 21 –	Distribuição de haplogrupos africanos encontrados na amostra da população do Rio de Janeiro e apresentados na forma de frequência relativa (%).....	125
Figura 22 –	Mapa com frequências de haplogrupos principais encontrados na população do Rio de Janeiro e em populações da África.....	130
Figura 23 –	Representação gráfica das distâncias genéticas, entre as amostras estudadas e países da África (MDS).....	135
Figura 24 –	Representação gráfica das distâncias genéticas, através da análise do escalonamento multidimensional - <i>Multidimensional Scalling Plot</i> (MDS).....	138

LISTA DE TABELAS

Tabela 1 - Africanos transportados pelo tráfico transatlântico, segundo nacionalidade do navio. Período do Século XVI ao Século XIX.....	22
Tabela 2 - Exemplos de marcadores bialélicos do cromossomo Y, representativos de alguns dos haplogrupos (HG) principais, e respectiva distribuição populacional, indicando as regiões e/ou populações onde a sua frequência é mais elevada.....	48
Tabela 3 - Comparação entre as frequências dos principais haplogrupos encontrados nas amostras estudadas do Rio de Janeiro e noutras populações brasileiras.....	127
Tabela 4 - Matriz das distâncias genéticas (F_{st}) para haplótipos entre pares de populações.....	136
Tabela 5 - Matriz das distâncias genéticas (F_{st}) para haplótipos entre pares de populações agrupadas por região.....	137

LISTA DE ABREVIATURAS E SIGLAS

AC -	Antes de Cristo
FUNAI -	Fundação Nacional do Índio
IBGE -	Instituto Brasileiro de Geografia e Estatística
ISFG -	Sociedade Internacional de Genética Forense (do inglês, <i>International Society of Forensic Genetics</i>)
Kb -	Kilobase
Kv -	Kilovolt
LDD -	Laboratório de Diagnósticos por DNA da UERJ
LINES -	<i>Long Interspersed Elements</i>
Ma -	Milhões de anos
MDS -	Escalonamento Multidimensional (do inglês, <i>MultiDimensional Scaling</i>)
Mb -	Megabase
µL -	Microlitro
mtDNA -	DNA mitocondrial (do inglês, <i>Mitochondrial DNA</i>)
MSY -	Região do cromossomo Y específica masculina (do inglês, <i>Male Specific Y Chromosome</i>)
nm -	Nanômetro
ng -	Nanograma
NR-Y-	Região não recombinante do cromossomo Y (do inglês, <i>Non-recombining Region of the Y chromosome</i>)
pb -	Par de bases
PCR -	Reação em cadeia da polimerase (do inglês, <i>Polymerase Chain Reaction</i>)
RFLP -	<i>Restriction Fragment Length Polymorphism</i>
TCLE -	Termo de Consentimento Livre e Esclarecido
RNA -	Ácido ribonucléico (do inglês, <i>Ribonucleic Acid</i>)
SNP -	Polimorfismo de base única (do inglês, <i>Single Nucleotide Polymorphism</i>)
STR -	Repetição curta em tandem (do inglês, <i>Short Tandem Repeat</i>)
VNTR -	Repetição em tandem de comprimento variável (do inglês, <i>Variable Number of Tandem Repeat</i>)

SUMÁRIO

	INTRODUÇÃO.....	12
1	REVISÃO DA LITERATURA.....	16
1.1	População brasileira- Elementos formadores.....	16
1.1.1	<u>Os portugueses - Ocupação e influência europeia.....</u>	16
1.1.2	<u>Os ameríndios - População nativa.....</u>	18
1.1.3	<u>Africanos no Brasil.....</u>	22
1.1.3.1	Diáspora africana e a entrada no Brasil.....	22
1.1.3.2	África – Geografia, migrações internas, diferenciações étnicas e linguísticas.....	26
1.2	Entrada de escravos no sudeste do Brasil - Rio de Janeiro.....	34
2	ESTUDOS FILOGEOGRÁFICOS E MARCADORES POLIMÓRFICOS DO DNA.....	36
2.1	Polimorfismos do DNA.....	37
2.2	Cromossomo Y.....	42
2.2.1	<u>Evolução dos cromossomos sexuais e estrutura do cromossomo Y....</u>	42
2.2.2	<u>Marcadores genéticos específicos do cromossomo Y.....</u>	44
2.2.3	<u>Filogenia e distribuição geográfica de haplogrupos do cromossomo Y</u>	47
2.2.4	<u>Principais haplogrupos africanos.....</u>	52
2.2.5	<u>Principais haplogrupos europeus.....</u>	55
2.2.6	<u>Principais haplogrupos nativos da América.....</u>	56
2.3	Estado da arte dos estudos de polimorfismos genéticos no Brasil.....	57
2.3.1	<u>Estudos com marcadores polimórficos.....</u>	57
2.3.2	<u>Estudos com polimorfismos do cromossomo Y.....</u>	58
3	OBJETIVOS GERAIS.....	62
3.1	Objetivos Específicos.....	62
4	MATERIAL E MÉTODOS.....	64
4.1	Amostragem e obtenção do DNA.....	66
4.2	Genotipagem.....	66
4.2.1.	<u>Marcadores do tipo STR do cromossomo Y.....</u>	67
4.2.2	<u>SNPs do cromossomo Y.....</u>	68
4.2.2.1	YAP - Análises da Inserção polimórfica do tipo <i>Alu</i>	69

4.2.2.2	M269 - Análise do Y-SNP por RFLP.....	71
4.2.2.3	Genotipagem por sequenciamento de base única (SBE).....	75
4.3	Análises estatísticas dos resultados	76
5	RESULTADOS E DISCUSSÃO	76
5.1	Caracterização da população do Rio de Janeiro	
5.2	Caracterização das linhagens masculinas de uma amostra do oeste africano	99
5.3	Análise da origem das linhagens africanas presentes na população do Rio de Janeiro	125
5.3.1	<u>Haplogrupos africanos na população do Rio de Janeiro</u>	
5.3.2	<u>Comparações das frequências dos haplogrupos africanos no Rio de Janeiro com populações da África</u>	128
5.3.3	<u>Comparações das frequência dos haplótipos dos grupos africanos no Rio de Janeiro com populações da África</u>	132
5.3.3.1	Comparações entre a amostra do Rio de Janeiro e países da África.....	132
		137
5.3.3.2	Comparações entre regiões da África.....	139
5.4	Discussão final dos resultados obtidos	144
	CONCLUSÕES	145
	REFERÊNCIAS	162
	ANEXO A - Supergrupos linguísticos e subdivisões.....	
	ANEXO B - Representação Esquemática da Divisão do Grupo Niger-Congo (grupo etnolinguístico dos <i>Bantu-Speakers</i>).....	170
		171
	ANEXO C - Árvore filogenética de haplogrupos do cromossomo Y.....	
	ANEXO D - Representação esquemática dos SNPs escolhidos e utilizados nos 5 multiplexes descritos de forma hierárquica.....	173
	ANEXO E - Oligonucleotídeos para a amplificação das regiões alvo e tamanho dos produtos de PCR dos multiplexes utilizados neste estudo.....	174
	ANEXO F - Oligonucleotídeos para a detecção dos SNPs por SBEs com a concentração utilizada e tamanho dos fragmentos.....	175

INTRODUÇÃO

Os navegantes, barbudos, hirsutos, fedentos de meses de navegação oceânica, escalavrados de feridas do escorbuto, olhavam, em espanto, o que parecia ser a inocência e a beleza encarnadas. Os índios vestidos da nudez emplumada, esplêndidos de vigor e de beleza, tapando as ventas contra a pestilência, viam, ainda mais pasmos, aqueles seres que saíam do mar [...]. Para os que chegavam, o mundo em que entravam era a arena dos seus ganhos, em ouro e glórias, ainda que estas fossem principalmente espirituais, ou parecessem ser, como ocorria com os missionários. (RIBEIRO, 1995, p.44).

Podemos assim perceber que o Brasil surgiu a partir da confluência de povos distintos, onde por muitas vezes foram observados entrecosques, matrizes raciais díspares, tradições culturais distintas, mas que sem dúvida, serviram para amalgamar o que hoje chamamos de povo brasileiro. Este povo ímpar, formado pela junção do colonizador europeu que aqui chegou, com a indulgência do silvícola que já aqui habitava e, posteriormente com um influxo maciço de negros africanos trazidos como objeto de trabalho e comércio. Desta miscelânea cultural nasceu este povo único, mestiço, sincrético e totalmente diferenciado de qualquer outro povo distribuído ao redor do globo.

A sociedade e a cultura brasileiras são conformadas como variantes da versão lusitana da tradição civilizatória europeia ocidental, diferenciadas por coloridos herdados dos índios americanos e dos negros africanos. O Brasil emerge, assim, como um renovo mutante, remarcado de características próprias, mas atado geneticamente à matriz portuguesa [...] (RIBEIRO, 1995).

Historicamente, muitas informações acerca das origens, traduzidas em número de indivíduos que aportaram ou através de dados sobre as suas regiões geográficas de origem, se mostram truncadas, pouco claras ou com pouca documentação. Muito foi perdido no período pós-escravidão com a queima ou destruição de arquivos sobre este período (BARBOSA, 1988). De um lado, por parte do clero, numa tentativa de esconder todos os tormentos e atrocidades cometidos durante a escravatura, ou de outro lado, como forma política de inibir os pedidos de indenização por parte dos antigos senhores de escravos (LACOMBE, 1988).

Portanto, verificar a verdadeira influência étnica, em um povo onde o processo de miscigenação mostrou-se completamente distinto de qualquer processo colonizador em outras partes do globo, pode mostrar-se extremamente difícil em termos puramente históricos. Cientes de que os registros de viagem nem sempre refletem com precisão o passado, nos últimos tempos os historiadores passaram a recorrer também à Genética na tentativa de compreender melhor o que de fato ocorreu.

Em vista disto, hoje podemos utilizar ferramentas biológicas muito informativas que permitem uma análise populacional extremamente confiável. O DNA humano possibilita a apresentação de toda a história de um indivíduo e sua ancestralidade, sendo também responsável pela expressão dos seus caracteres físicos como cor de pele, estatura, cor dos olhos ou sua propensão às doenças genéticas.

Apesar da maior parte do DNA humano ser invariável, é a porção variável do genoma que tem atraído a atenção, por ser a responsável pelas diferenças interindividuais e populacionais. Estas diferenças que se apresentam como variações ou polimorfismos surgiram gradativamente ao longo do tempo, sendo transmitidas para as gerações seguintes. À medida que a espécie humana foi se espalhando pelos continentes por meio de migrações, guerras ou diásporas, muitas alterações do genoma observadas sob a forma de mutações, acrescentaram novos polimorfismos aos já existentes. Por meio da análise destas variações presentes no DNA é possível recuperar informações sobre a história das populações humanas, pois estas permanecem como registros moleculares das rotas migratórias seguidas durante a expansão mundial.

Desde a época da descoberta da estrutura do DNA (WATSON & CRICK, 1953), muitos métodos de estudo foram desenvolvidos e muito evoluíram desde os primeiros usos em estudos forenses. A introdução da técnica da Reação em Cadeia da Polimerase ou PCR (do inglês *Polymerase Chain Reaction*) (MULLINS *et al.*, 1986; SAIKI *et al.*, 1985) permitiu a análise de grande número de fragmentos específicos de DNA, mesmo partindo de pequenas quantidades de amostras biológicas. Esta técnica permitiu a introdução de novos tipos de marcadores no campo forense, tais como as repetições curtas em *tandem*, também conhecidos como microssatélites ou STRs (do inglês *Short Tandem Repeats*) e os polimorfismos de base única ou SNPs (do inglês *Single Nucleotide Polymorphisms*),

O uso de marcadores polimórficos presentes em linhagens cromossômicas que não são sujeitas à recombinação e se apresentam em estado haplóide, tais como o DNA mitocondrial (mtDNA) e o cromossomo Y, permitem uma análise mais acurada e detalhada dos movimentos migratórios que originaram as populações atuais (SANTOS & TYLER-SMITH, 1996; UNDERHILL *et al.*, 2001). Por causa de seus padrões de herança, estes marcadores também chamados de uniparentais contam histórias complementares e independentes acerca dos ancestrais maternos (mtDNA) e paternos (cromossomo Y) dos povos (SCHURR & SHERRY, 2004). Além disto, estes marcadores permitem traçar linhagens independentes e ligadas ao sexo dos indivíduos ancestrais, e deste modo podemos detectar diferenças no padrão demográfico diferencial de homens e mulheres no passado, assim como nos padrões de acasalamento, poliginia ou poliandria, que afetam diferentemente estas linhagens ancestrais (MOORAD *et al.*, 2013).

Os polimorfismos localizados na região não recombinante do cromossomo Y (NRY) são transmitidos em bloco de uma geração para a outra, sendo modificados apenas por mutação. Tanto os microssatélites quanto os polimorfismos de uma única base (SNPs) têm sido amplamente utilizados na investigação da ancestralidade paterna de populações humanas. Os SNPs apresentam uma taxa de mutação muito menor, quando comparados aos STRs, mostrando-se muito elucidativos no estabelecimento de relações entre populações separadas por muitas gerações (KAYSER & DE KNIJFF, 2011)

Podemos considerar, então, que os dados obtidos a partir de estudos genéticos, podem contar a história do povoamento dentro de suas peculiaridades, como qualquer conjunto de dados, tais como os da antropologia, arqueologia ou linguística e fornecem valiosas informações sobre a nossa origem ou sobre a história evolucionária.

Estudos com haplótipos constituídos de microssatélites e de haplogrupos constituídos de SNPs, da região não recombinante do cromossomo Y, são muito utilizados na investigação populacional e se constituem ferramentas valiosas na investigação das proporções das populações ancestrais. Atualmente, muito se tem pesquisado sobre a formação das populações no Brasil e no mundo, através de análises moleculares do DNA humano. Informações históricas disponíveis (IBGEa, 2007) descrevem que a população brasileira foi formada pela miscigenação entre povos da Europa (principalmente da Península Ibérica), África e autóctones da

América (Ameríndios). Os dados históricos que temos disponíveis podem ser comparados através de estudos genéticos, com o objetivo de confirmar, corroborar ou até mesmo refutar as teorias sobre a colonização.

A base da formação da sociedade brasileira foi caracterizada por uma família numerosa, composta não só do núcleo conjugal e de seus filhos, mas incluindo um grande número de criados, parentes, agregados e escravos, submetidos todos ao poder absoluto do chefe do clã, que era, ao mesmo tempo, marido, pai e patriarca. Este foi o modelo que contribuiu para a predominância da linhagem masculina europeia no Brasil. Este tipo de formação familiar, com as mulheres e os filhos em estado de submissão, permitiu o encorajamento da mestiçagem, sendo um processo basicamente unidirecional e que se deveu à permissividade no cruzamento entre homens brancos, principalmente com mulheres negras, pois estas estavam estreitamente ligadas ao ambiente doméstico. Neste contexto, se define o perfil genético do Rio de Janeiro, com Portugal sendo a principal referência genética masculina. Nos momentos posteriores, com a abolição da escravatura e com a evolução social, este processo tornou-se mais diluído com a miscigenação ocorrendo nos dois sentidos (homem branco e mulher negra; mulher branca e homem negro). Hoje existe um grande interesse em determinar as proporções de cada componente das diferentes populações, e suas influências, na formação do perfil genético masculino brasileiro. Portanto, caracterizar a herança genética mostra-se de extrema importância na tentativa de contribuir com mais estudos, ampliando conhecimento sobre o nosso país e contribuindo com novos dados genéticos que servirão para ampliar a nossa base de dados e de conhecimentos, tanto na Genética Forense quanto na Clínica.

1 REVISÃO DA LITERATURA

1.1 População brasileira - Elementos formadores

Aos habitantes ameríndios originários do Brasil foram incorporados elementos culturais, étnicos e/ou genéticos, tanto de europeus quanto de africanos que, sem dúvida, moldaram a nossa população com uma característica única e ao mesmo tempo extremamente complexa.

A população brasileira é considerada uma das mais miscigenadas do mundo e deriva de um processo extenso e específico de mistura étnica, promovido a partir de três populações básicas compostas por índios nativo-americanos, europeus (principalmente de Portugal) e indivíduos de origem africana (ABE-SANDES *et al.*, 2004).

1.1.1 Os portugueses - Ocupação e influência europeia

A partir dos anos de 1501/1502, no início do período colonial, grande parte da costa brasileira passou a ser explorada, de forma extrativista, pelos portugueses. Mais tarde, entre os anos de 1515 e 1516, as expedições à chamada Terra de Santa Cruz se tornaram mais frequentes e de maior importância. A ocupação portuguesa que ocorreu no território brasileiro foi mínima durante os primeiros 50 anos de colonização, se constituindo como um considerável problema para Portugal. Isto se deveu ao fato de que a população da metrópole na época era pouco superior a 1 milhão de habitantes, o que não permitiu povoar de maneira efetiva a América, além de todas as colônias que já possuía na Ásia e na África. Além disto, havia ainda o desinteresse de Portugal com relação à nova terra, pois o comércio com as Índias apresentava-se, ainda, como mais vantajoso e menos dispendioso.

O perfil dos primeiros habitantes portugueses em solo brasileiro foi particularmente relacionado aos chamados degredados que, invariavelmente, eram

obrigados a vir como forma de castigo ou de perdão à pena de morte, deixando para trás as mulheres e os seus filhos (DEL PRIORE, 2001). Calcula-se que, durante os dois primeiros séculos de povoamento, nas regiões centrais da economia colonial, como Bahia e Pernambuco, os degredados corresponderam a 10 ou 20 % da população reinol (IBGE a, 2007). Neste momento, o número de habitantes portugueses em território brasileiro era mínimo, quando comparados aos índios nativos que aqui viviam. Na verdade, não é possível determinar com exatidão o número dos habitantes portugueses nos primeiros dois séculos de colonização brasileira, face aos dados discordantes da literatura, mas segundo o IBGE (2007), ocorreu a vinda de cerca de 500 a 5 mil portugueses por ano.

Muitos historiadores apontam que, nesta época, a união entre homens europeus e mulheres indígenas tornou-se corriqueira, sendo encorajada como estratégia para o processo de ocupação do território (TORERO & PIMENTA, 2000). Obviamente, este fato pode ser facilmente compreendido porque neste momento histórico, existia uma proporção de uma mulher para cada grupo de oito ou nove homens que aqui aportaram (TORERO & PIMENTA, 2000).

A política colonialista portuguesa em relação ao Brasil passou a tomar novos rumos sob o reinado de D. João III, no ano de 1531 e pôde ser observada a partir da expedição de Martim Afonso de Sousa, que manteve os mesmos objetivos das expedições anteriores, que era de combater os traficantes franceses, reconhecimento e exploração da costa, penetrar nas terras na direção do Rio da Prata e para procurar metais preciosos. Contudo, ainda adicionou-se a tais objetivos iniciais, a efetivação da ocupação portuguesa em solo brasileiro (MARANHÃO & KEATING, 2011).

O cultivo de cana-de-açúcar, a construção de engenhos e o uso da mão de obra escrava foi o modelo econômico colonial utilizado inicialmente. Assim, com estas intenções, começou a tentativa de escravizar o indígena, para em seguida, promover a implantação do tráfico maciço de negros africanos (PRADO JÚNIOR, 1997). A partir de então, na segunda metade do século XVI, as uniões com escravas africanas foram estimuladas e os indícios históricos mostraram uma prevalência de relações entre colonizadores europeus, que ainda vinham em grupos muito grandes de homens com poucas mulheres, ou nenhuma. Este aspecto somado ao relacionamento, na maioria das vezes forçado, destes europeus com mulheres indígenas e africanas, visto que, ocupavam posição social inferior e de

submissão, favoreceu a miscigenação por um processo de reprodução direcional na qual o homem europeu tinha filhos com as mulheres africanas e indígenas. O cruzamento entre índios e negros também ocorreu, porém em menor escala (PRADO JÚNIOR, 1997). A imigração europeia, basicamente portuguesa, foi estimada em aproximadamente 700.000 indivíduos entre os períodos de 1500 a 1808 (IBGE a, 2007). Após a abertura dos portos brasileiros às nações amigas, promulgada por meio de Carta Régia pelo Príncipe Regente D.João VI, em 28 de janeiro de 1808, o Brasil passou a receber um número considerável de imigrantes de outras partes do globo. Contudo, em relação à Europa, a contribuição portuguesa foi a principal, sendo posteriormente, seguida por italianos, espanhóis e alemães (DEL PRIORE, 2010).

1.1.2 Os ameríndios - População nativa

Existem algumas teorias sobre a chegada e instalação dos primeiros habitantes no continente americano. Especulações arqueológicas recentes sugerem a existência de uma cultura instalada no Brasil há milhares de anos antes da chegada de Cabral (SÓ HISTÓRIA – Origem do nome índio na América).

De acordo com a teoria mais aceita pelos historiadores, os primeiros habitantes das Américas, ou os antepassados dos atuais índios (conhecidos como Paleoíndios) eram povos originários do leste da Ásia que entraram no continente americano por intermédio de uma ponte terrestre, denominada de Beríngia, que se formou ligando a Ásia e o Alasca, conectando os dois continentes por volta de 24.000 a 9.000 anos atrás (BIANCHI *et al.*, 1998; DILLEHAY, 2009).

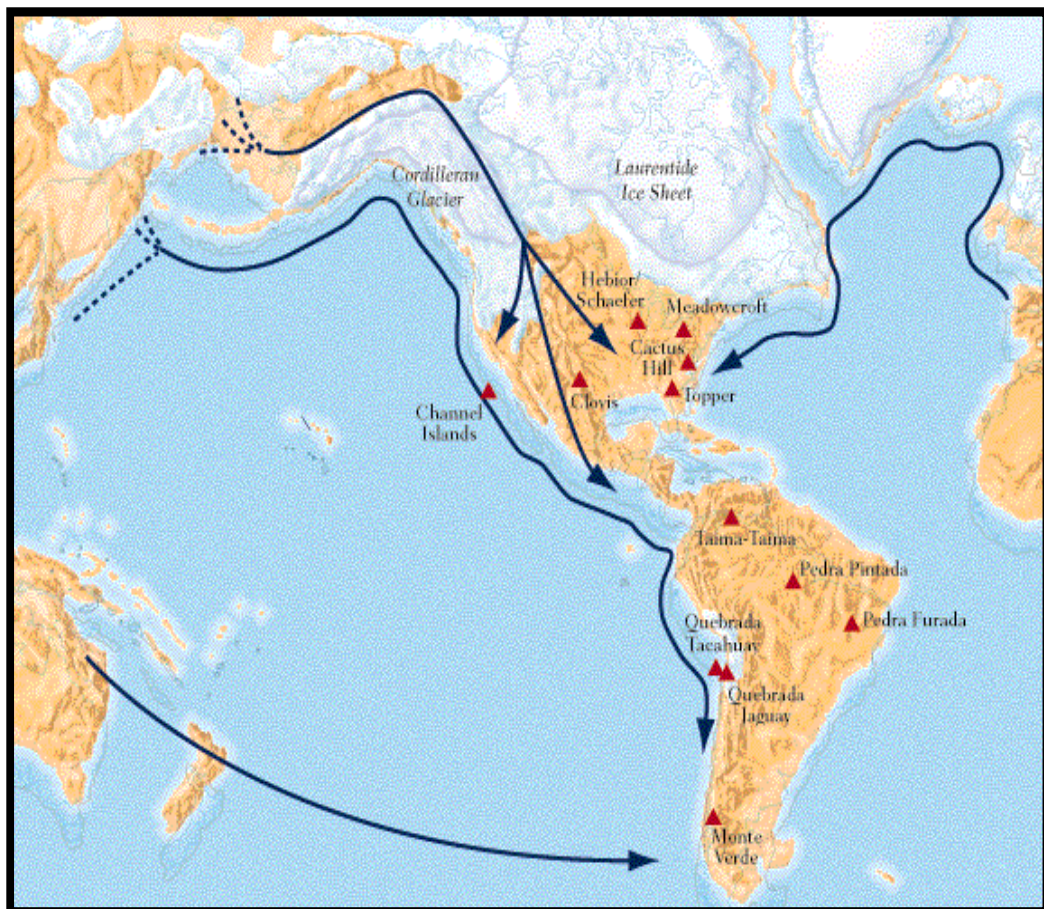
Segundo a teoria transoceânica ou Teoria Malaiopolinésia, há cerca de 10 mil anos atrás, povos que habitavam a atual região da Polinésia teriam navegado pelo Oceano Pacífico até alcançar o litoral sul-americano em pequenos barcos. Nessa travessia os indivíduos teriam se deslocado até aqui movidos por meio das correntes marítimas que os conduziram (RIVET, 2008).

Por outro lado, também existe a hipótese de Bradley & Stanford (2004), que postularam uma rota alternativa através do Atlântico Norte, possivelmente originária da Europa durante a Era Glacial, e que ocorreu através de migrações sobre a

camada de gelo que cobriria o atlântico conjuntoentrional em direção às Américas. Esta nova possibilidade foi apresentada com base na comparação de artefatos de pedra e osso da cultura do povo Clóvis (Novo México), com outros que foram encontrados entre os povos da cultura Solutreana na Europa (WATERS & STAFFORD, 2007). Com estes achados, sugere-se que a origem dos povos ameríndios poderia ser também a partir de regiões onde atualmente se encontram a França e a Espanha (Figura 1).

As análises de polimorfismos do cromossomo Y e do DNA mitocondrial sustentam as hipóteses de que provavelmente ocorreram duas principais ondas migratórias com origem na Ásia, sendo que a primeira se estendeu até à América do Sul e a segunda se restringiu à América do Norte (LELL *et al.*, 2002; BORTOLINI *et al.*, 2003)

Figura 1 - Exemplificação das possíveis rotas de chegada ao continente americano



Fonte: The 2011 Harriet–Elliott Lecture Series Blog.

A origem da palavra "índio" não é bem conhecida. Alguns historiadores acreditam que este vocábulo remonta à Idade Média, onde era normalmente empregado para designar de forma genérica todas as pessoas originárias do Extremo Oriente. Quando Cristóvão Colombo chegou às Américas acreditou que havia encontrado um novo caminho para as Índias e resolveu chamar os nativos que encontrou de "índios" (Etimologia; IBGE a, 2007). Existe ainda a idéia de que se referia também ao termo usado pelos colonizadores que durante muito tempo consideraram as Américas como as "Índias Ocidentais" (Só História – origem do nome índio na América). Portanto, podemos considerar que o conceito da palavra "índio" é uma denominação errônea ou simplesmente uma invenção européia.

O fato é que os habitantes originários encontrados nas Américas não constituíam um povo uno e se apresentavam como vários grupos distintos, que nutriam grande animosidade recíproca e que, constantemente guerreavam entre si (IBGE a, 2007).

No Brasil do século XVI, quando da chegada dos portugueses, os índios tupis predominavam no litoral e possuíam costumes semelhantes (IBGE a, 2007), sendo o "Tupinambá" a língua mais falada pelos indígenas, e conhecida, segundo denominação dos jesuítas, como a "língua geral" (IBGE a, 2007). Os índios tapuias eram os indivíduos que pertenciam a outros grupos que não falavam esta "língua geral" (O MOVIMENTO TAPUIA). Em realidade, "tapuia" é uma denominação genérica para indígenas que não se integraram à civilização luso-brasileira ou tupi-guarani e que, na sua grande maioria, eram povos de outros troncos ou famílias linguísticas, que falavam línguas dos grupos macro-jê, caribe, aruaque e arauá (IBGE a, 2007).

Enquanto em outros locais do globo, os grupos de povos europeus, africanos e asiáticos já interagiam em seus próprios continentes há muito tempo, fazendo com que as diferentes tecnologias se espalhassem e difundissem entre eles, os índios ficaram confinados por milênios sem interação com outras populações humanas e sem poder ensinar ou aprender novas técnicas (NARLOCH & TEIXEIRA, 2011). Portanto, a chegada dos europeus representou um rompimento de milhares de anos de isolamento. A vida junto aos "brancos" era muito atrativa e muitos indígenas abandonavam voluntariamente as suas aldeias e iam viver junto com os portugueses. Em muitos casos, os índios nem precisavam sair de suas tribos pois, com a expansão da colonização, essas aldeias foram assimiladas dentro da

sociedade ocidental. Temos como exemplos os centros urbanos, como Niterói ou Guarulhos, que eram aldeias indígenas e que se transformaram em cidades (NARLOCH & TEIXEIRA, 2011).

“A história do povoamento indígena no Brasil é antes de tudo, uma história de despovoamento” [...] (IBGE a, 2007), embora pouco se saiba ao certo, qual era o total da população que habitava o atual território brasileiro em 1500. A estimativa varia entre 1 milhão e 6,8 milhões de pessoas, numa população que se estendia da Amazônia, ao Brasil central e costa nordeste (IBGE b, 2007). No Brasil, é corriqueiro o senso comum afirmar que os índios foram "exterminados" e que, atualmente, restam alguns poucos representantes dessa população. É inegável que, com a chegada dos europeus, milhares de índios morreram em decorrência de guerras ou pela escravidão. Mas, a grande mortalidade indígena se deu pelo contágio involuntário de doenças trazidas pelos europeus, contra as quais os índios não tinham imunidade, por terem vivido durante milênios isolados de outras populações. Porém, o que muitas vezes se ignora no Brasil é que grande parte da população indígena não pereceu, mas foi assimilada pela sociedade brasileira. Muitos índios foram viver ao lado de portugueses e africanos, com eles se miscigenando e dando origem a grande parcela da população brasileira, sendo que seus descendentes não mais se identificam como "índios" (NARLOCH & TEIXEIRA, 2011).

É factível que a miscigenação entre portugueses e indígenas não tenha sido tão intensa quanto a que ocorreu entre portugueses e africanos. Quando comparada com outros países da América Latina a contribuição indígena no Brasil é bem menor, mas indubitavelmente, ela é existente em maior ou em menor grau (PRADO JÚNIOR, 1997). Segundo a Funai (2009), cerca de 25% da população indígena da Amazônia já mora em cidades e só metade desse contingente se considera indígena, mesmo falando uma segunda língua e praticando rituais. Esse processo de assimilação indígena, que já terminou na maior parte do Brasil, ainda está em curso na região Norte. Atualmente, existem cerca de 300.000 ameríndios no Brasil vivendo em reservas protegidas pelo Governo Federal (IBGE b, 2007; FUNAI - CENSO DEMOGRÁFICO, 2010).

1.1.3 Africanos no Brasil

1.1.3.1 Diáspora africana e a entrada no Brasil

O Brasil foi, no continente americano, a região que mais importou escravos africanos durante os 300 anos de duração do tráfico transatlântico, entre o século XVI e meados do século XIX (IBGE a, 2007) (Tabela 1).

Tabela 1 - Africanos transportados pelo tráfico transatlântico, segundo nacionalidade do navio. Período do Século XVI ao Século XIX

Nacionalidade do navio	Africanos Transportados (em milhares)	
	Partiram da África	Chegaram às Américas
Total	11 348,8	9 682,6
Inglaterra	3 536,2	3 009,4
América		
Inglesa/EUA	220,6	205,5
Caribe Inglês	59,4	51,3
Portugal/Brasil	4 942,2	4 335,8
França	1 456,6	1 127,8
Holanda	533,5	449,5
Espanha	513,3	429,6
Dinamarca	82	69,7
Outras	5	4

Fonte: “500 anos de povoamento” (IBGE, 2007)

A chegada dos africanos em território brasileiro teve início na metade do século XVI, mas não se sabe ao certo o ano da chegada dos primeiros negros ao Brasil. Goulart (1975) sugere que este evento tenha ocorrido entre 1516 e 1526, com o início da exploração da cana-de-açúcar, em substituição ao trabalho escravo dos indígenas. Segundo o IBGE a (2007), a substituição ocorreu devido ao declínio no número da população nativa, sua inexperiência e a pouca resistência ao trabalho sistemático da lavoura intensiva. Mas também ocorreu pelo interesse português de

que progredisse a atividade que, talvez fosse a mais lucrativa do sistema colonial: o tráfico de escravos africanos. Dados indicam que os primeiros africanos se estabeleceram nos engenhos do Recôncavo Baiano e foram trazidos das ilhas atlânticas portuguesas (Canárias, Madeira e Açores) que funcionavam como entreposto de escravos trazidos da África e onde já se cultivava o açúcar em regime de monocultura (IBGE a, 2007).

Nesta época, ocorria um grande comércio negreiro das regiões da Mina e Accra (Gana), Benin e Calabar (Nigéria), Gabão e Camarões para as plantações da ilha de São Tomé e alguns destes indivíduos também foram transportados para o Brasil (OGOT, 2000). Não é impossível que houvesse alguns cativos africanos, entre os tripulantes da esquadra de Cabral, uma vez que Portugal já traficava escravos nas costas africanas antes de 1500 (IBGE a, 2007). Contudo, o mais antigo registro da entrada de escravos é datado de 1532, com a chegada do navio capitaneado por Tomé de Souza (DEL PRIORE, 2001).

O tráfico transatlântico promoveu o povoamento do Brasil por gente oriunda de diversas regiões do continente africano. Regiões essas que contribuíram para o povoamento em graus variados de intensidade, dependendo do período considerado e das conexões comerciais mantidas pelos traficantes portugueses, brasileiros e africanos de um e outro lado do Atlântico. Assim, os portos do Brasil podiam por vezes, e em certos períodos, se especializar em determinadas direções do fluxo do comércio de pessoas (IBGE a, 2007).

Aproximadamente quatro milhões de pessoas, dos cerca de 10 milhões de indivíduos traficados, sobreviveram às condições da viagem e foram utilizados no trabalho escravo entre os séculos XVI e XIX (HEYWOOD, 2008). Eram povos que aqui foram denominados de congos, angolas, benguelas, cabindas, cassanges, monjolos, rebolos, moçambiques e outros. Os chamados angolas — em geral, indivíduos traficados através do porto de Luanda — e benguelas — traficados através de entrepostos situados mais ao sul da atual Angola — vieram a predominar nas levas do comércio oitocentista, em especial os que desembarcavam no Rio de Janeiro. Os traficantes envolvidos no comércio baiano, por outro lado, a partir de meados do Século XVII, e até o fim do tráfico, foram se especializando cada vez mais na região do Golfo do Benin (sudoeste da atual Nigéria), de onde importaram escravos que aqui foram denominados de dogomés, jejes, ussás, bornos, tapas e nagôs, entre outros (IBGE a, 2007).

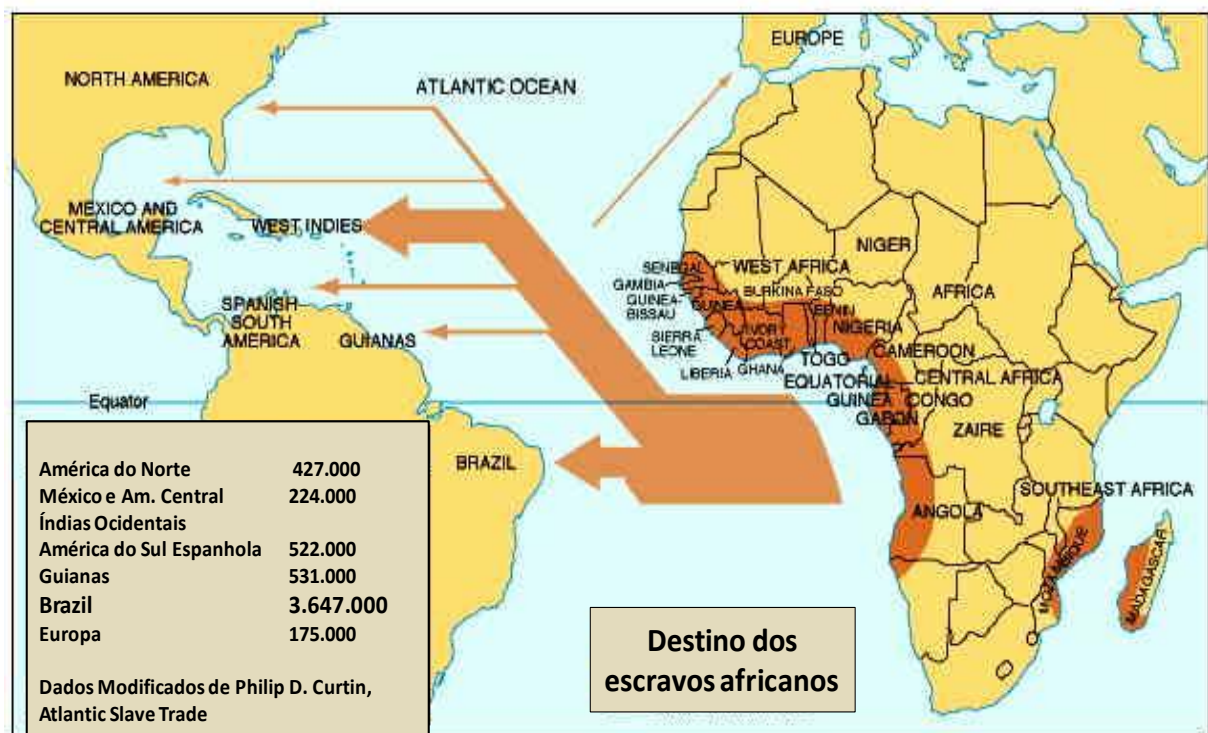
Registros históricos, ainda que incompletos, indicam que a maioria dos indivíduos era proveniente da região que compreende atualmente o Congo e Angola (Centro-Oeste e Sudoeste Africano), onde já havia o estabelecimento de colônias portuguesas, sendo as mais importantes fontes de escravos e as únicas mantidas durante todo o período do tráfico (KLEIN, 2002). Acredita-se que mais de 2/3 de todos os indivíduos foram oriundos desta região, sendo trazidos entre 1701 e 1810. Esta região manteve um fluxo constante de escravos durante todo o período escravista no Brasil, contudo, esta não foi a sua única fonte. Foram enviados também indivíduos oriundos das principais regiões da África Ocidental (Senegal até Nigéria), bem como da região da África Oriental (Moçambique). Como a presença portuguesa na colônia africana de Moçambique (1752-1975) foi fraca, com um número extremamente baixo de imigrantes (mas não menos do que o número de imigrantes que vieram para o Brasil), essa região se tornou uma das últimas fornecedoras de escravos no século XIX, devido ao alto custo do transporte pelo fato da rota ser mais distante do Brasil (KLEIN, 2002).

Estudos sugerem que mais da metade dos escravizados – em número de 15 mil por ano até 1790 – que alcançaram a região sudeste do Brasil, entre 1595 e início de 1800, procedia da parte ocidental da África Central, tendo recebido um influxo maior de povos oriundos dos portos de Luanda, Cabinda, Benguela, Congo, Malingo e Loango (HEYWOOD, 2008) (Figura 2). O porto do Rio de Janeiro foi o maior importador de escravos do Brasil do final do século XVII até quase metade do século XIX, recebendo predominantemente escravos enviados da África Ocidental. Na composição da população negra do Rio de Janeiro, no período de 1822 a 1833, 86% dos escravos eram recém-chegados do continente africano e 14% nascidos ou já residentes na América portuguesa (FLORENTINO, 1997). Desses, a maioria era do sexo masculino, com faixa etária predominante de 15 a 49 anos (HEYWOOD, 2008).

Um fato digno de nota refere-se ao fato de que muitas tribos dominadoras, principalmente da região do Congo, após as guerras tribais que ocorriam com frequência, podiam exterminar os capturados ou mantê-los como seus próprios escravos para executar os trabalhos considerados de menor importância (SILVA, 2008). Frequentemente, estes indivíduos foram utilizados como objeto de comércio com os europeus e esta prática forneceu grande quantidade de escravos para o tráfico transatlântico (HEYWOOD, 2008).

De um modo geral, as mulheres capturadas nas guerras tribais eram inseridas na nova sociedade, a fim de incrementar o crescimento populacional, com o objetivo principal de procriar e fornecer novos guerreiros. Deste modo, estas mulheres incorporaram a cultura dominante, perdendo a sua própria identidade ou, muitas vezes, introduziram elementos de seus costumes e da sua cultura original à cultura do dominador (HEYWOOD, 2008). Portanto, podemos ter em mente que o comércio negreiro promovido pelas tribos dominadoras muitas vezes facilitava o tráfico de um número maior de homens em relação ao de mulheres.

Figura 2 - Diáspora africana transatlântica



Fonte: ARREGLO - La Etnia Negra de Panama.

A diversidade linguística e cultural dos contingentes negros introduzidos no Brasil, somada às hostilidades recíprocas que eles traziam da África e a política de evitar a concentração de escravos oriundos de uma mesma etnia nas mesmas propriedades e até mesmo nos navios negreiros, impediu a formação de núcleos solidários que retivessem o patrimônio cultural africano (RIBEIRO, 1995). Os escravos encontravam-se dispersos na terra nova, ao lado de outros, seus iguais na

cor e na condição servil, mas diferentes na língua, na identificação tribal, e frequentemente, hostis, pelos referidos conflitos de origem. Desse modo, os africanos foram compelidos a incorporar-se no universo cultural da nova sociedade (RIBEIRO, 1995).

Uma grande quantidade de informações relacionadas às origens na África pode ter sido perdida, devido ao agrupamento de indivíduos de várias procedências regionais nos portos de saída. Portanto, não deve ser desprezado o fato de que os escravos que foram comercializados, muitas vezes recebiam como denominação étnica o local do porto de saída e não o da sua verdadeira origem geográfica ou tribal (HEYWOOD, 2008).

1.1.3.2 África – Geografia, migrações internas, diferenciações étnicas e linguísticas

O continente africano costuma ser regionalizado de duas formas: a primeira forma valoriza a localização dos países e os dividem em quatro grupos, que são: 1) África Conjuntoentrional ou do Norte, 2) África Ocidental, 3) África Oriental e 4) África Meridional. A segunda regionalização, utiliza-se de critérios étnicos e culturais (religião, linguagem e etnias predominantes em cada região), dividindo a África em dois grandes grupos: 1) África Branca ou Conjuntoentrional, formada pelos oito países da África do Norte, mais a Mauritânia e o Saara Ocidental, e 2) África Negra ou subsaariana, formada pelos outros 44 países do continente (PEREIRA, 2006) (Figura 3).

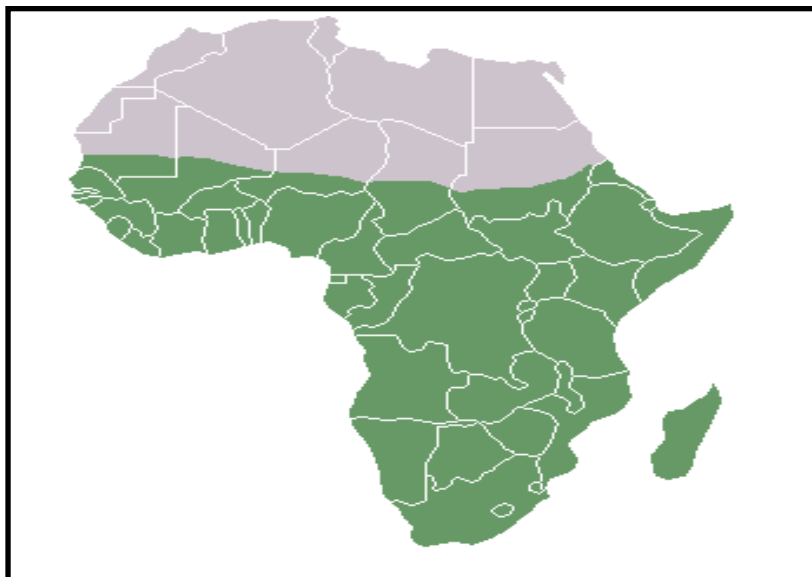
Efetivamente, o Deserto do Saara com os seus cerca de 9 milhões de quilômetros quadrados, surgido em período relativamente recente ocorrido na metade do Holoceno (9.000 - 6.000 atrás) (CLAUSSEN *et al.*, 1999), forma uma espécie de barreira natural que dividiu o continente africano em duas partes muito distintas quanto ao quadro humano e econômico (Figura 4). Enquanto a região norte manteve a possibilidade de se relacionar com o Oriente Médio, a porção subsaariana transformou-se numa área relativamente isolada, onde a Mesopotâmia e o Mediterrâneo não exerceram influência.

Figura 3 - Mapa da África atual



Fonte: Sua Pesquisa.com – Continente Africano.

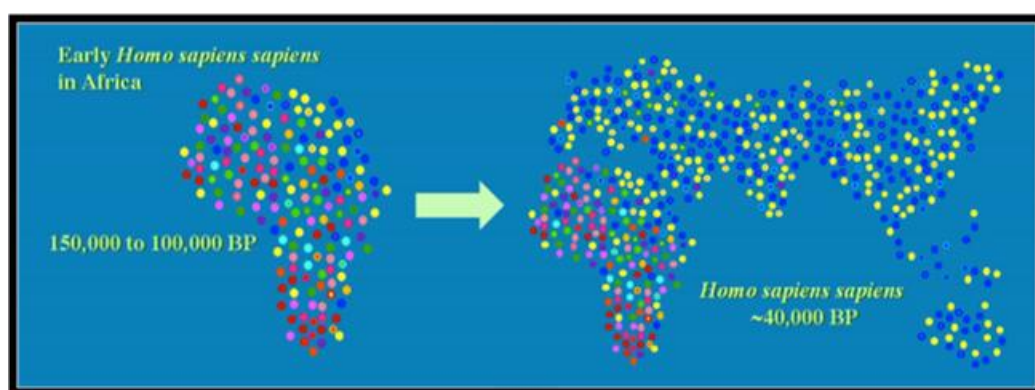
Figura 4 - Mapa esquemático da África mostrando a separação em quadro humano e econômico



Legenda: Em cinza = África do Norte; Em verde = África Negra ou subsaariana.
 Fonte: Wikipedia – África Subsaariana.

A população da África subsaariana tem uma história extremamente rica, sendo considerada uma região de extensa diversidade genética, fenotípica, cultural e linguística (TISHKOFF *et al.*, 2002), e onde existem aproximadamente 2.000 grupos étnicos e linguísticos que constituem cerca de um terço das línguas mundiais (CAMPBELL & TISHKOFF, 2010) (Figura 5).

Figura 5 - Comparações esquemáticas da diversidade genética na África e no mundo



Nota: Na figura encontra-se indicado o surgimento do *Homo sapiens sapiens* em África há cerca de 100.000 a 150.000 anos (BP, do inglês *Before Present*) e sua expansão a outros continentes há aproximadamente 40.000 anos, transportando somente parte da diversidade genética acumulada em África.

Fonte: Yale School of Medicin.

A grande maioria dos linguistas concorda que, no continente africano, as línguas faladas se dividem em quatro grandes famílias (GREENBERG, 1948) (Figura 6):

a) Afroasiática (que inclui as línguas Berberes do Norte da África, as Cushitas da Etiópia e da Somália e ainda as Semitas, abrangendo o hebreu, o árabe e o aramaico);

b) Nilo-Sahariana (ou Nilo-saariana), constituída pelo Nilota, Sudanês, o Saaariano e o Songhai;

c) Niger-Congo (ou Congo-Khordofanian, em nomenclatura obsoleta), que inclui numerosos grupos predominantes para o sul do Saara, do qual destacamos os povos Bantu, para o sul do Equador;

d) Outras, que são representadas principalmente pelos Khoisan (ou Khoi-San), língua falada pelos Pigmeus da floresta tropical do Congo Democrático e as

línguas faladas com “estalinhos” pelos povos !Kung, vulgarmente conhecidos como Hotentotes, Bosquímanos ou Mucancalas em Angola (GREENBERG, 1948; Wikipedia – Ethnic Groups; Comunidade Inzo Nkongombila)

Figura 6 - Quadro com os quatro grupos etnolinguísticos da África, com alguns subgrupos representados



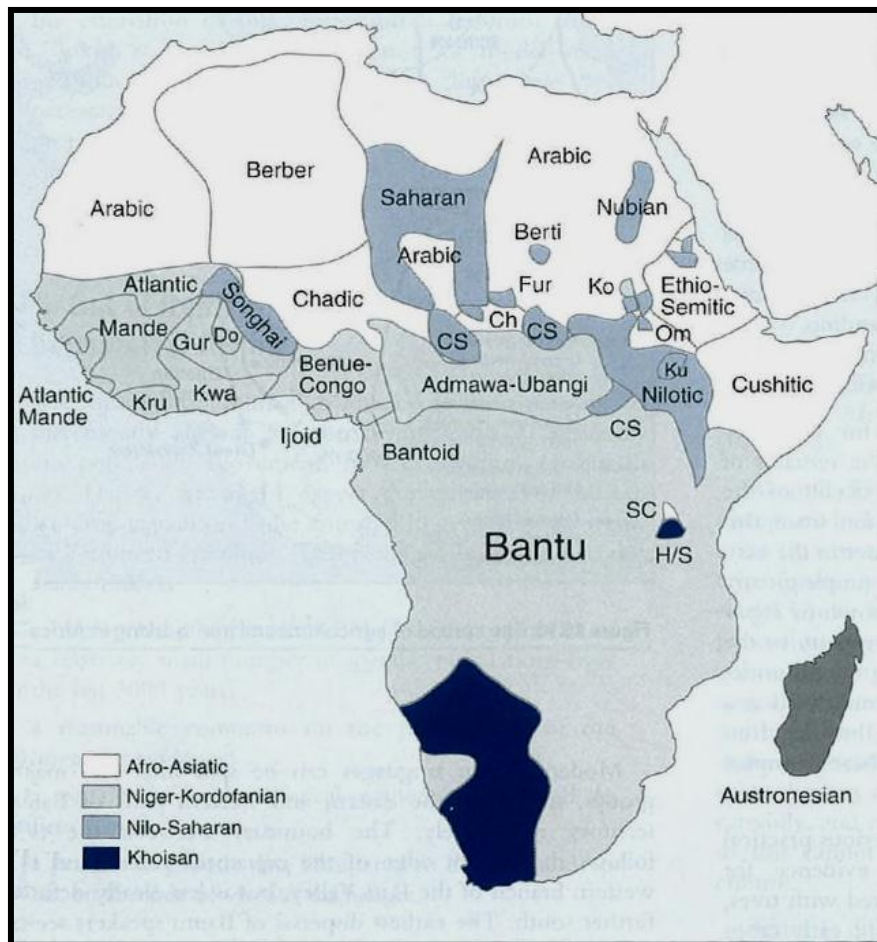
Fonte: Wikipedia - Africans.

Para entender com mais detalhes sobre os quatro principais grupos africanos e suas divisões, apresentamos informações adicionais no Anexo A.

Um dos grupos mais estudados em Genética de Populações é o dos originalmente classificados como *Bantu-speakers* (grupos de povos que falam a

língua Bantu). Este termo refere-se a uma classificação linguística relacionada a um grupo de famílias que se distribuem no sentido centro-sul do continente africano (GREENBERG, 1948). Os *Bantu-speakers* são derivados do grupo Niger-Congo (Anexo B), um dos mais variados e complexos grupos etnolinguísticos presentes no continente africano. Atualmente, mais do que uma identidade cultural ou racial, este grupo é retratado segundo a reunião de seus caracteres físicos e genéticos correlacionados com a distribuição linguística (Figura 7).

Figura 7 - Famílias de línguas africanas e o domínio dos grupos Bantu



Fonte: JOBLING *et al.*, 2004.

Segundo Blench (2006), estas características de homogeneidade do grupo Bantu, foram originárias da expansão em larga escala de povos agricultores, que empreenderam a maior corrente migratória ocorrida na África.

Esta expansão ocorreu em diversas ondas migratórias e em várias direções, sendo responsável pela dispersão da agricultura e da metalurgia para as regiões sul e central da África (PEREIRA, 2006). Através de evidências linguísticas, observou-se que a dispersão dos povos Bantu ocorreu a cerca de não mais de 5.000 anos atrás e seguiu em duas direções: do sudeste em direção à floresta equatorial (Nigéria e Camarões) e a leste penetrando na região lacustre do país que é conhecido atualmente como Uganda. Deste modo, se formou um núcleo Bantu na região oriental da África. A partir daí novas expansões ocorreram em direção ao Sul da África e ao leste do Zimbábue seguindo bem próximo ao Transvaal (região nordeste da África do Sul) (VANSINA, 1995; BLENCH, 2006) (Figura 8).

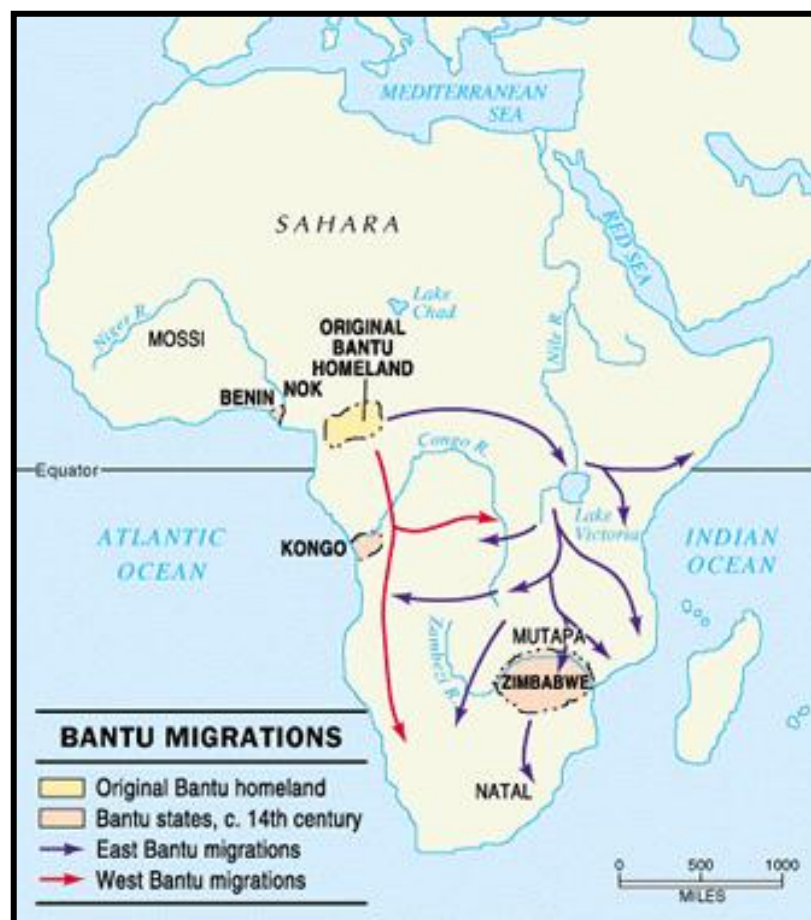
Além do fato da expansão populacional, esse êxodo promoveu a absorção das populações nativas de povos coletores-caçadores durante os deslocamentos (HEWLETT & CAVALLI-SFORZA, 1986). O resultado disso foi uma expansão simultânea da cultura, linguagem e genes Bantu através da África subequatorial (DIAMOND & BELWOOD, 2003), onde atualmente a maioria dos povos subsaarianos fala a língua Bantu.

Os povos do grupo dos Khoisan atualmente são encontrados nas florestas do Camarões, no Gabão e em algumas regiões da República Centro Africana (RCA), República Democrática do Congo (antigo Zaire), Namíbia, África do Sul e Botswana. Contudo, no passado, estes grupos nômades ocupavam uma área muito mais extensa do continente africano. Dentro deste grupo encontramos os Pigmeus, San e Khoi-Khoi, que apesar de serem considerados como idênticos, devido ao fato de se expressarem por “estalinhos” ou cliques, apresentam variações antropológicas e linguísticas que podem caracterizar suas diferentes origens (TISHKOFF *et al.*, 2002). Como resultado da expansão Bantu, os povos da região da floresta equatorial e tropical foram cedendo lugar a novas populações constituídas principalmente de indivíduos que falavam a língua Bantu.

Existe outra classificação dos povos Bantu, proposta por Guthrie em 1948 e posteriormente atualizada por Guthrie (1971) e Maho (2009), que se baseia na análise comparativa de fonemas e sons. Segundo os parâmetros deste estudo, os grupos de língua Bantu podem ser discriminados em 16 zonas geográficas através de estudos da gramática e envolve a comparação de diferentes padrões de linguagem, com a busca de palavras em comum e as variações que são encontradas em cada grupo linguístico e também com a fonética. Em suma, são

estudos que determinam as distâncias linguísticas comparando as similaridades ou diferenças existentes entre as línguas, determinando distâncias entre as populações. Segundo esta classificação, os povos Bantu podem ser separados em zonas geográficas assinaladas com as letras de A até S, que são subdivididas em grupos (A10, A20, etc), em linguagens individuais (A11, A21, etc) e dialetos (A11a, A11b, etc) (Figura 9).

Figura 8 - Mapa esquemático indicando as direções da expansão/ dispersão Bantu

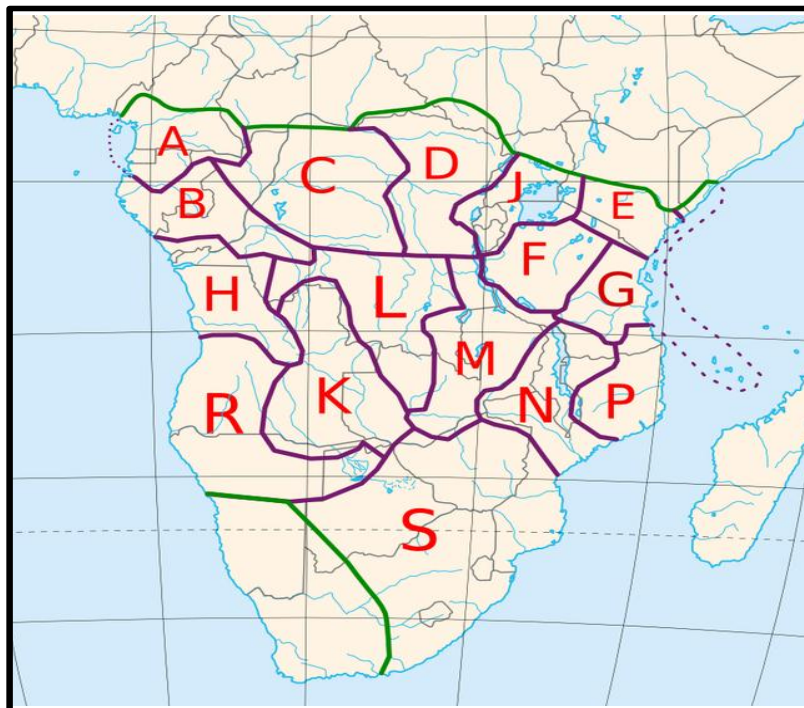


Fonte: The Bantu expansion 3000-1000 BC.

Em Angola, que foi a principal colônia portuguesa no século XV, foram fundadas as cidades de Luanda (1576) e Benguela (1617), que serviram de base para o comércio negreiro, entre os séculos XVI e XIX, e que enviaram um total de cerca de três milhões de escravos ao Brasil. Os Bantu angolanos são divididos em 9

grupos etnolinguísticos que por seu turno estão subdivididos em cerca de 100 subgrupos, tradicionalmente chamados de tribos. Além disto, dentro desta divisão podemos ainda ter várias nações, como por exemplo, os Bakongo, Mbundu, Lunda-Tchokwe, Ovimbundu e Ganguela entre outros (www.inzotumbansi.org).

Figura 9 - Localização aproximada das 16 Zonas Bantu descritas por Guthrie (1948)



Fonte: Guthrie classification of Bantu languages – Wikipedia.

Durante a colonização de Moçambique (a partir do século XVI), os primeiros descobridores portugueses observaram que os angolanos conseguiam se comunicar com facilidade com os povos originários da costa moçambicana, através de linguagem derivada do grupo Niger-Congo (HEYWOOD, 2008). Isto reflete e indica que a unidade linguística nem sempre está associada a uma região geográfica ou a um país com as fronteiras que temos atualmente, pois o fenômeno da expansão Bantu promoveu uma homogeneização linguística e populacional através do continente africano.

Atualmente, os mapas linguísticos da África não coincidem com a distribuição dos “tipos raciais”, embora tal concordância possa ter ocorrido num passado

remoto. Mas, durante um longo período, os antigos grupos étnicos se multiplicaram, migraram e cruzaram, favorecendo a divergência linguística ou de fenótipos. Isto é patente no caso dos povos do Sudão, que apresentam uma zona de confluência de dois tipos diferentes de famílias linguísticas, representadas pelos Afro-Asiáticos e Nilo-Saarianos (CAMPBELL & TISHKOFF, 2010).

1.2 Entrada de escravos no sudeste do Brasil - Rio de Janeiro

O sudeste brasileiro recebeu o maior número de cativos provenientes da África (FREIRE, 2012) e que teriam vindo essencialmente do centro-oeste africano (HEYWOOD, 2008).

O porto do Rio de Janeiro foi o maior centro importador de escravos do Brasil do final do século XVII até quase a metade do século XIX, recebendo predominantemente escravos enviados da África Centro-Occidental ou Centro-Atlântico e, durante o século XVIII, tornou-se um porto de ligação importante com Benguela (FLORENTINO, 1997). Em vista disto, houve no Rio de Janeiro, um incremento na importação, ocorrendo grande multiplicação do comércio de escravos, tornando-se assim, mais importante do que o comércio praticado inicialmente em Pernambuco, no século XVI (HEYWOOD, 2008).

A população predominante de escravos que permaneceu no Rio de Janeiro era originária dos portos de Luanda e Benguela, assim como a mão de obra específica para as minas de ouro e diamantes descobertas em Minas Gerais (HEYWOOD, 2008). Contudo, cada vez mais as pessoas traficadas eram também provenientes do interior das terras montanhosas ao sul da África Central (HEYWOOD, 2008). No início do século XIX (1800-1840), muitos dos escravos que foram comercializados vieram da região de Cabinda (costa atlântica) e de várias partes de Moçambique (costa leste), juntando-se às populações de cativos que já se encontravam estabelecidas e distribuídas, e sendo também enviados às novas regiões produtoras de café do Vale do Paraíba (São Paulo e sul do Rio de Janeiro) (HEYWOOD, 2008). O estado de Minas Gerais é considerado pelos pesquisadores como a região com maior número de escravos oriundos, principalmente, da Costa da Mina (golfo da Guiné) e de Angola, sendo o principal ponto de recebimento de

cativos a partir do porto do Rio de Janeiro, nos primeiros anos do século XIX (FREIRE, 2012).

Portanto, existem indicativos de que, após a chegada ao Rio de Janeiro, verificava-se uma distribuição seletiva dos escravos africanos para as várias regiões localizadas no Sudeste Brasileiro. Esta diferença estava essencialmente relacionada com as atividades desenvolvidas nas regiões, as quais eram marcadamente distintas (lavoura e extração de ouro/diamantes) e que, hoje, correspondem às regiões pertencentes aos estados do Rio de Janeiro e Minas Gerais, respectivamente.

Sendo assim, qual seria a real origem geográfica desta etnia ou etnias que, geneticamente, se apresentam de forma tão expressiva nestes estados? Podemos considerar como origem os portos de saída da África? Ou podemos considerar esta expressão associada somente com a etnia? Em vista de todas estas questões devemos ter em mente que, para responder estas perguntas elaboradas acima, é necessário que haja uma correta seleção de marcadores genéticos que permitam discriminar e estabelecer critérios sobre a origem étnica da população, de modo a permitir o estudo sistemático da contribuição africana no Brasil.

2 ESTUDOS FILOGEOGRÁFICOS E MARCADORES POLIMÓRFICOS DO DNA

A filogeografia é uma disciplina que usa dados genéticos para elucidar a história da dispersão das populações no mundo. Baseia-se na premissa de que a presente distribuição geográfica de indivíduos é determinada não somente pela biologia atual e fatores ecológicos, mas também por fatores históricos (AVISE *et al.*, 1987). Através da combinação de informações provenientes de dados genéticos, históricos, evolutivos e geográficos, podemos conhecer os padrões da diversidade genética atual e reconstruir os eventos passados.

Muitos avanços ocorreram em todas as áreas da Biologia, principalmente na Genética, a partir dos estudos feitos por Avery e colaboradores (1944), que determinaram que a hereditariedade relaciona-se diretamente a uma molécula de ácido desoxirribonucléico (DNA) encontrada nas células de todos os seres vivos. Em 1953, Watson e Crick propuseram um modelo para a estrutura da molécula do DNA, aceito até hoje, e sugeriram ainda um mecanismo de cópia do material genético, conhecido como replicação semiconservativa (WATSON & CRICK, 1953).

O material genético ou genoma divide-se em cromossomos, os quais são constituídos por uma molécula de DNA e proteínas histonas e não-histonas (LEWIN, 2014). O genoma das células diplóides humanas consiste em 22 pares de cromossomos autossômicos e um par de cromossomos sexuais. O sequenciamento do genoma mostrou que os 23 cromossomos do *Homo sapiens* são compostos por, aproximadamente, três bilhões de pares de bases, dos quais 1,1% correspondem a éxons, 24% a íntrons e 75% a seqüências intergênicas (VENTER *et al.*, 2001).

A partir da conceituação do modelo em dupla hélice da estrutura do DNA, foram desenvolvidas metodologias e ferramentas biotecnológicas que permitiram que o DNA fosse estudado de forma mais minuciosa. Na década de 70, Sanger desenvolveu sua técnica de sequenciamento do DNA (SANGER *et al.*, 1977) e em 1986 foi publicado o método de replicação de DNA *in vitro* denominado de PCR (MULLIS *et al.*, 1986). Em 1985, Jeffreys e colaboradores introduziram na ciência forense uma nova técnica de Biologia Molecular pela qual, através da análise direta do DNA, era obtido um padrão típico para cada indivíduo, denominado *DNA fingerprinting* ou “perfil de DNA” (JEFFREYS *et al.*, 1985). Na obtenção do perfil de DNA, foram utilizadas as técnicas de RFLP, ou do inglês, *Restriction Fragment*

Length Polymorphism, Southern Blotting e hibridização molecular (JEFFREYS *et al.*, 1985; KOBILINSKY *et al.*, 2005).

Nestes últimos anos, têm sido descritas novas técnicas de sequenciamento conhecidas por técnicas de sequenciamento de nova geração, do inglês *Next-Generation Sequencing Technologies* (NGS), com múltiplas aplicações em genômica e Biologia Molecular. Este termo refere-se a um conjunto de técnicas que permitem sequenciar o DNA e o RNA de forma mais rápida e barata, quando comparados com os princípios da técnica de sequenciamento de Sanger (BERGLUND *et al.*, 2011). Estas técnicas são consideradas como tipos de sequenciamento de alto rendimento, podendo ser utilizadas uma série de diferentes tecnologias que incluem algumas plataformas, tais como: *Sequencing by Synthesis* (Illumina sequencing), *Ion Semiconductor Sequencing* (Ion Torrent) and *Sequencing by Ligation* (SOLiD sequencing). O crescimento destas novas tecnologias abre um novo patamar em estudos de Genética de População baseados no sequenciamento de genomas completos de um grande número de indivíduos, assim como de amostras de DNA antigo, tanto de humanos como de outras espécies.

2.1 Polimorfismos do DNA

O DNA humano apresenta variações que são definidas como sequências de nucleotídeos que apresentam duas ou mais formas diferentes e se originam a partir de deleções, substituições e inserções de um ou mais nucleotídeos. Diz-se que um *locus* é polimórfico sempre que existam pelo menos duas formas alélicas com frequência igual ou superior a 1%. O termo polimorfismo, por vezes, também é aplicado aos alelos com frequência superior a 1%, sendo que aqueles com frequências inferiores a 1% são denominados de variantes alélicas. Vários destes tipos de polimorfismos (também denominados de marcadores genéticos) podem ser constituídos por sequências únicas, sequências repetitivas ou sequências altamente repetitivas, podendo estar dispersas ou arranjadas em *tandem* ao longo do genoma, apresentando variações tanto na sequência, quanto no número de elementos repetitivos (FOWLER *et al.*, 1988; ARCOT *et al.*, 1996). A maioria das enzimas e proteínas estruturais é codificada por sequências únicas, e algumas sequências

altamente repetitivas podem ser responsáveis pela codificação de alguns RNAs (KASS & BATZER, 2001). No entanto, existe um grande número de sequências repetitivas que estão localizadas em regiões intrônicas não codificantes e que, por isso, são seletivamente neutras, tendendo a apresentar uma elevada variabilidade intrapopulacional.

Por muito tempo predominou o uso de polimorfismos protéicos ou polimorfismos clássicos em estudos que utilizavam os grupos sanguíneos, proteínas do complexo de histocompatibilidade (HLA), imunoglobulinas e outras proteínas plasmáticas ou das hemácias (CAVALLI-SFORZA *et al.*, 1996; 1998).

Atualmente, com o conhecimento de outros marcadores polimórficos do DNA que se encontram distribuídos por todo genoma, abriu-se uma nova possibilidade de estudos moleculares, o que oferece uma nova perspectiva aos estudos de variabilidade genética. Devido ao fato destes novos marcadores serem mais numerosos, torna-se possível selecionar marcadores em regiões não codificantes, possibilitando assim, um número maior de variações do que aqueles permitidos através dos polimorfismos clássicos.

Como exemplo destes polimorfismos, temos os microssatélites ou STRs, que consistem de sequências em *tandem*, repetidas várias vezes ao longo do genoma, constituídas por unidades básicas com até seis pares de bases (pb) de tamanho. No genoma humano, ocorrem em média a cada 10.000 nucleotídeos e estima-se que compreendam cerca de 3% do genoma (ARCOT *et al.*, 1996; SCHLÖTTERER, 1998; HUMMEL, 2003; BUTLER, 2005). Com o advento da técnica da amplificação de regiões específicas do DNA através da PCR, houve um grande incremento no uso destes microssatélites em várias áreas, sobretudo na identificação individual e em análises forenses sendo preferencialmente utilizados a partir da década de 1990 (GOLDSTEIN & SCHLÖTTERER, 1999; SCHOSKE, 2003).

Os STRs são multialélicos, com alto grau de polimorfismo e apresentam uma alta taxa de mutação na ordem de aproximadamente 2×10^{-3} /geração (BUTLER, 2005). Estudos realizados nas áreas da Biologia Molecular, da Genética Clássica e da Genética de Populações, utilizando estes marcadores em particular, têm revelado a sua grande importância na discriminação entre indivíduos de uma mesma população (KAYSER *et al.*, 1997, BUTLER, 2005).

Estes marcadores também estão presentes no cromossomo Y e, devido ao fato de não apresentarem recombinação, possuem a particularidade de permitir a

discriminação de conjuntos de alelos (denominados haplótipos) com alta resolução (HAMMER, 2001), possibilitando assim a diferenciação de populações com uma história de separação recente (REDD *et al.*, 2006).

Durante os últimos anos, centenas de variações nas regiões repetitivas do DNA têm sido identificadas (REDON *et al.*, 2006). Recentes descobertas revelaram que o número de cópias de um determinado gene pode variar de um indivíduo para outro, e foram denominadas de variações no número de cópias – CNVs (do inglês “*Copy Number Variations*”). Esta variação é responsável por aproximadamente 13% do DNA genômico humano e cada variação pode ter cerca de uma quilobase (1.000 bases nucleotídicas) a até várias megabases de tamanho (REDON *et al.*, 2006). Esta forma de variação estrutural são modificações do genoma que resultam na alteração do número de cópias para certos genes, o que pode determinar uma anormalidade ou normalidade em uma seção do DNA. Os CNVs correspondem a regiões relativamente grandes do genoma que foram suprimidas (apresentando-se com número menor que o normal) ou duplicadas (mais do que o número normal) em certos cromossomos. Por exemplo, o cromossomo que normalmente tem seções representadas por ABCD pode em vez disto ter seções ABCCD (uma duplicação de "C") ou ABD (uma supressão de "C") (FREEMAN *et al.*, 2006).

Os polimorfismos de base única, ou SNPs (do inglês, *Single Nucleotide Polymorphisms*) são originados por mutações pontuais que produzem diferenças em um único par de bases entre as sequências de cromossomos (COLLINS *et al.*, 1998, BRUMFIELD *et al.*, 2003; DAAL, 2008). Os SNPs já foram considerados como a forma mais prevalente e importante da variação genética, contudo, os estudos atuais revelam que os CNVs compreendem, pelo menos, três vezes o teor total de nucleotídeos dos SNPs (FREEMAN *et al.*, 2006).

Até 2003, mais de quatro milhões de SNPs haviam sido mapeados e depositados em bases de dados públicas e privadas, sendo que apenas três por cento se localizam nos genes (MELTON, 2003). O projeto internacional denominado de *HapMap* (International HapMap Consortium, 2007) registrou aproximadamente 3,1 milhões de SNPs que foram mapeados e genotipados em 270 indivíduos de quatro regiões geográficas do planeta. O Centro Nacional para as Informações Biotecnológicas ou NCBI (do inglês, *The National Center for Biotechnology Information*) possui dados de mais de dezessete milhões de SNPs que já foram

mapeados, sendo as respectivas informações depositadas em uma base de dados específica, denominada de dbSNP (<http://www.ncbi.nlm.nih.gov/SNP>). Os SNPs têm sido muito utilizados como marcadores genéticos em estudos de associação com doenças, ou como marcadores para elucidação da história das populações (BRUMFIELD *et al.*, 2003; MELTON, 2003). Mais recentemente, estes polimorfismos também passaram a ser utilizados como ferramenta para prever a origem étnica e até mesmo características físicas dos indivíduos (PEREIRA *et al.*, 2012; KIDD *et al.*, 2014; PHILLIPS *et al.*, 2012).

Devido à sua baixa taxa de mutação, de aproximadamente 10^{-8} /geração (BUTLER, 2003), acredita-se que a maioria dos SNPs tenha surgido como resultado de eventos mutacionais únicos, razão pela qual também são chamados de polimorfismos de evento único – UEPs (do inglês, *Unique Event Polymorphisms*), geralmente apresentando apenas dois alelos (bialélicos) (BUTLER, 2003). O alelo derivado apresenta uma distribuição geográfica característica, ocorrendo com frequência mais elevada na região onde provavelmente surgiu e em frequências baixas ou ausentes em outras regiões (JOBLING *et al.*, 2004). Estes marcadores, quando localizados em regiões não sujeitas à recombinação, tanto no cromossomo Y quanto no DNA mitocondrial, possibilitam a definição de haplogrupos (grupos de haplótipos) que reúnem linhagens relacionadas por descendência (KNIJFF, 2000; RUIZ-PESINI *et al.*, 2007). As informações fornecidas por estes polimorfismos também são valiosas para o estudo das migrações humanas, permitindo assim a diferenciação entre populações provenientes de regiões geográficas distintas (UNDERHILL *et al.*, 1996; BOSCH *et al.*, 2001; BRIÓN *et al.*, 2003; BRUMFIELD *et al.*, 2003).

Os produtos de PCR dos SNPs podem ter menos de 100 pb de tamanho, o que lhes confere vantagem sobre os microssatélites para a genotipagem de amostras de DNA altamente degradado. Além disto, podem ser analisados a partir de reações de PCR multiplexes com maior número de *loci* alvos do que os STRs (BUTLER, 2005). Por este motivo, estes marcadores têm sido cada vez mais utilizados, tanto em Genética Forense como em estudos genealógicos com base em amostras de DNA antigo – do inglês *Ancient DNA* -, pois a amplificação de regiões maiores, tais como dos STRs, mostra-se mais difícil com amostras de DNA mais degradadas (COBLE *et al.*, 2009; ROMANINI *et al.*, 2012).

Os polimorfismos denominados Indels, do inglês, *Insertion-Deletion*, são variações de comprimento geradas por inserção ou deleção de nucleotídeo(s) em uma sequência de DNA, cujos tamanhos podem variar desde uma base a até milhões de bases ou mesmo grandes pedaços do cromossomo. Assim como os SNPs, também possuem uma baixa taxa de mutação, na ordem de $2,3 \times 10^{-9}$ (JOBILING *et al.*, 2004), e também são de particular interesse para estudos filogenéticos (WEBER *et al.*, 2002; MILLS *et al.*, 2006).

Outro grupo de polimorfismos relaciona-se às inserções de elementos retrotransponíveis, divididos em *Long Interspersed Elements* (LINEs) e *Short Interspersed Elements* (SINEs). Elementos de transposição constituem também uma importante classe de INDELS que tem nos SINEs da família *Alu* os seus mais numerosos representantes, com mais de um milhão de cópias no genoma humano (*International Human Genome Sequencing Consortium*, 2001).

Estas inserções de sequência do tipo *Alu* são uma classe de DNA repetitivo, originalmente caracterizada pela ação produzida por uma endonuclease de restrição de mesmo nome. Diferentes tipos destas inserções ocorrem em grande quantidade, dispersas no genoma de mamíferos e são compostas principalmente por sequências que se originaram como transcritos reversos de moléculas de RNA 7SL (componente de uma partícula de reconhecimento de sinal citoplasmático) encontrado de um modo geral em primatas (KRIEGS *et al.*, 2007). Estas inserções polimórficas do tipo *Alu* também são bastante utilizadas como marcadores para o estudo da variação genética em populações humanas (BATZER *et al.*, 1994; ARCOT *et al.*, 1996; TRIPATHI *et al.*, 2008).

Segundo Hammer *et al.* (2006), o estudo das frequências alélicas e seu padrão de distribuição ao longo do planeta, possibilitam a caracterização das populações em termos de afinidades genéticas. É bem caracterizado que a proporção dos alelos varia de um lugar para outro, mas geralmente a diferença é pequena entre populações vizinhas, de forma que as maiores diferenças são observadas entre populações distantes. Portanto, a filogeografia tem sido fortemente influenciada por avanços analíticos recentes, como a análise de STRs e SNPs que permitem a observação de sequências com um ancestral comum (haplótipos e haplogrupos, respectivamente) que frequentemente apresentam um padrão específico de localização geográfica ou étnica.

Em relação à espécie humana, muitos estudos têm oferecido informações importantes sobre nossa origem ou nossa história evolucionária. Em um contexto particular, marcadores com herança uniparental ou não recombinante como aqueles localizados no cromossomo Y e no DNA mitocondrial (mtDNA) tem reconhecida importância. Estes dois marcadores são capazes de determinar as linhagens de origem paterna e materna, respectivamente, e permitem ainda a discriminação das características genéticas de populações em áreas geográficas diferentes. Estes dois marcadores de ancestralidade podem auxiliar na elucidação das diferentes influências promovidas por ambos os sexos na história das populações humanas.

Em suma, esta tese foi centrada no estudo de ancestralidade, utilizando polimorfismos genéticos específicos do cromossomo Y, através do uso de marcadores do tipo STRs, Indels e a inserção polimórfica do tipo *Alu*.

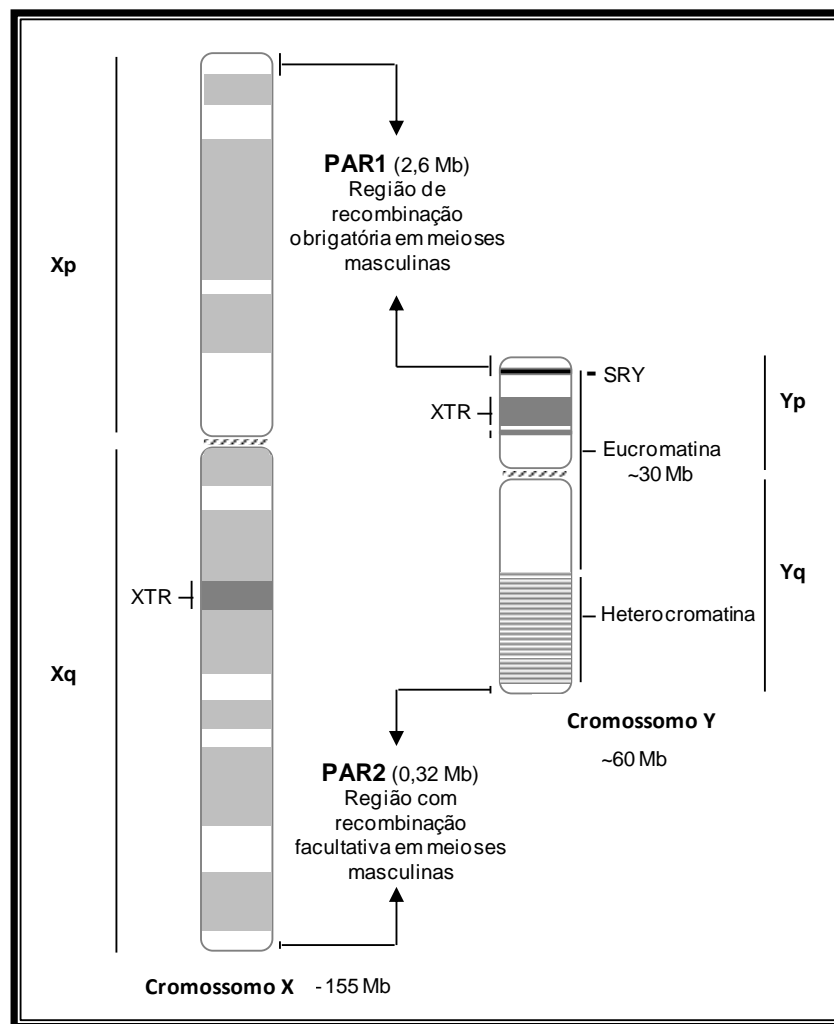
2.2 Cromossomo Y

2.2.1 Evolução dos cromossomos sexuais e estrutura do cromossomo Y

O cromossomo Y é produto da divergência do seu homólogo, o cromossomo X, num processo que se iniciou há aproximadamente 240-320 milhões de anos (Ma), com a aquisição da função da determinação do sexo masculino (JOBILING *et al.*, 2004). Após o primeiro evento no processo da divergência, o protocromossomo Y sofreu a repressão da recombinação que causou a expansão da região não-recombinante, representando hoje 95% do seu total. Esta perda de recombinação também levou a uma degeneração do cromossomo, tanto em tamanho (com aproximadamente 60 milhões de pares de bases, sendo o menor cromossomo em humanos), quanto em relação ao conteúdo gênico (JOBILING *et al.*, 2004). Em humanos, a parte eucromática do cromossomo Y apresenta aproximadamente 23 megabases (Mb) de tamanho e contém 78 genes que codificam proteínas, sendo este conteúdo muito menor do que aquele presente no cromossomo X, que apresenta 150Mb e cerca de 800 genes que codificam proteínas (BACHTROG, 2013). A maior parte do cromossomo Y não tem homologia com nenhum outro

cromossomo e, portanto, não sofre recombinação, sendo denominada região não recombinante do Y (NRY) ou região específica masculina (MSY) (LAHN *et al.*, 2001; JOBLING, 2003) (Figura 10).

Figura 10 - Estrutura do cromossomo Y - regiões pseudoautossômicas (PAR1 e PAR2); região não recombinante (NRY), ou específica do sexo masculino (MSY)



Fonte: Gusmão & Alves, 2010.

Embora grande parte seja não recombinante, durante a meiose masculina os cromossomos Y e X sofrem algum tipo de recombinação em regiões onde a homologia ainda se encontra preservada. Estas regiões são transmitidas de forma semelhante a outras regiões localizadas em autossomos. Por esta razão são referidas como regiões pseudoautossômicas, denominadas PAR1 no braço curto

com aproximadamente 2,6 Mb e PAR2 no braço longo com aproximadamente 0,32 Mb (ROZEN *et al.*, 2003; SKALETSKY *et al.*, 2003) (Figura 10).

Apesar de apresentar menos polimorfismos em relação aos demais cromossomos (JOBLING, 1994), o cromossomo Y ainda assim apresenta mais de 28.000 SNPs (JOBLING & TYLER-SMITH, 2003), o minissatélite MSY1 (JOBLING *et al.*, 1998), centenas de microssatélites (REDD *et al.*, 2002) e marcadores bialélicos entre os quais o elemento polimórfico *Alu* YAP (HAMMER, 1994).

2.2.2 Marcadores genéticos específicos do cromossomo Y

A caracterização de indivíduos e populações masculinas por meio dos polimorfismos do cromossomo Y é extremamente relevante em várias áreas da Genética, sendo amplamente utilizados em estudos populacionais ou forenses. Para isto, é usado um grande número de marcadores genéticos polimórficos, dentre os quais os STRs e os SNPs, por exemplo.

Em relação aos marcadores do tipo Y-SNP, referimo-nos de um modo geral a qualquer tipo de marcador, geralmente bialélico, que apresentam outras variações de sequência além daquelas que envolvam uma única base. Sendo assim, neste grupo também são incluídos os marcadores do tipo Indel e a inserção polimórfica do tipo *Alu*. Estes marcadores são mais adequados para construção de árvores filogenéticas pelo fato de serem relativamente estáveis e fáceis de genotipar [e.g. VAN OVEN *et al.*, 2011; 2014].

A combinação de alelos de Y-STRs encontrados nos distintos *loci* para um mesmo indivíduo é denominada de haplótipo (QUINTANA-MURCI *et al.*, 2001). Estes marcadores genéticos definem a variação dentro dos haplogrupos e tendem a apresentar uma maior diversidade quanto maior for o número de marcadores analisados. Por outro lado, definimos e consideramos o conjunto de marcadores polimórficos binários encontrados no cromossomo Y como haplogrupo. Por definição, um haplogrupo é um tipo de haplótipo formado por polimorfismos binários (SNPs) do cromossomo Y encontrados nas linhagens humanas e que são mais estáveis que os STRs (JOBLING & TYLER-SMITH, 2003).

Existe um número elevado de Y-STRs descritos na região não recombinante do cromossomo Y, sendo que, alguns deles foram incluídos em kits comerciais amplamente utilizados, tanto em Genética de Populações quanto em Forense. Atualmente, encontram-se disponibilizados no *Y-STR Haplotype Reference Database* (YHRD, 2014), um elevado número de dados de distribuição de frequência de haplótipos de Y-STRs que foram compilados e são utilizados, tanto na avaliação quantitativa dos resultados no tratamento de casos forenses e de parentesco, quanto na avaliação da estratificação da população masculina mundial. Este banco de dados é formado a partir de análises realizadas em amostras relativas a 126 países, representando 942 populações e 134.215 haplótipos depositados (YHRD, 2014).

Recentemente, Ballantyne e colaboradores (2010), efetuando estudos sobre taxa de mutação em STRs, definiram 13 Y-STR com uma taxa de mutação excepcionalmente alta (acima de 10^{-2} por *locus*/geração), denominados de *Rapidly Mutating* (RM) Y-STRs. Estes marcadores, em particular, permitem a individualização de amostras masculinas relacionadas por via paterna, facilitando a identificação nos casos de paternidade ou violações com a possibilidade da participação de indivíduos aparentados (BALLANTYNE *et al.*, 2012).

De um modo geral, em Genética de Populações tende-se a utilizar alguns grupos de marcadores segundo as diferenças que apresentam em relação às taxas de evolução, o que permite o seu uso com diferentes propósitos:

- Os Y-STRs, por apresentarem taxas de mutação mais elevadas, são utilizados para estudos históricos mais recentes. Como já foi dito, eles definem as variações (haplótipos) dentro dos haplogrupos, e permitem a separação de grupos de populações com terão divergido recentemente, assim como a definição do tempo de coalescência de haplogrupos bem definidos a partir de marcadores mais estáveis como os SNPs (QUINTANA-MURCI *et al.*, 2001).

- Os Y-SNPs, com taxas de mutação mais baixas, são considerados os marcadores mais adequados para estudos evolutivos. São usados na comparação de populações que tenham divergido há mais tempo ou em estudos de ancestralidade de linhagens paternas. Eles são considerados como “sinalizadores” de pequenos grupos ou indivíduos, definindo as raízes filogenéticas de haplogrupos com distribuição geográfica restrita, ou em alguns casos, agrupamentos regionais (QUINTANA-MURCI *et al.*, 2001).

Em resumo, os marcadores SNP, tornam-se desta forma, um grupo de marcadores de evolução lenta, ocorrendo em escalas de grandeza de 10^{-7} a 10^{-9} por geração, enquanto os microssatélites são marcadores de evolução mais rápida com frequência de mutação de cerca de 2×10^{-3} por geração (DUPUY *et al.*, 2004). Devido a esta característica, os Y-SNP permitem a identificação de haplogrupos, que são mais estáveis que os haplótipos de STRs, sendo de vital importância nos estudos populacionais, devido ao alto poder de diferenciar as principais populações continentais em nível mundial.

Diversos estudos sobre as variações do cromossomo Y conduzidos nas últimas duas décadas permitiram a construção de uma detalhada e robusta árvore filogenética. Muitos esforços têm sido despendidos no sentido de padronizar o sistema de nomenclatura dos haplogrupos (Y-SNPs) do cromossomo Y, através do *Y Chromosome Consortium* (YCC), de modo a permitir a inclusão das novas mutações e novos haplogrupos que vão sendo descritos. A primeira versão da árvore filogenética foi publicada em 2002 (*The Y Chromosome Consortium*, 2002), com atualização subsequente publicada em 2003 (JOBBLING & TYLER-SMITH, 2003). A mais recente modernização foi feita por Karafet e colaboradores (2008) que apresentaram uma árvore filogenética com 20 haplogrupos principais (A-T), representando a variação mundial do cromossomo Y. Esta última versão inclui 311 haplogrupos definidos por aproximadamente 600 Y-SNPs, podendo ser visualizado de forma hierárquica na árvore filogenética - YCC Tree (Anexo C).

Os haplogrupos são definidos como os ramos, grupos ou clados principais da árvore filogenética da linhagem paterna humana, que podem ser discriminados por SNPs característicos. Cada haplogrupo conjuga alguns sub-ramos também denominados subgrupos ou subclados. Cada haplogrupo e seus subclados (ramos) marcam as migrações humanas ao longo do globo (YCC, 2002). Ao longo destes últimos anos têm sido feitos estudos numerosos no sentido de se pesquisar novos SNPs, aprofundando a discriminação de linhagens dentro de certos haplogrupos específicos, com a conseqüente revisão da nomenclatura dos mesmos.

O termo Paragrupo é usado em Genética de Populações para descrever linhagens dentro de um haplogrupo que não são definidos por marcadores únicos. Nos haplogrupos do cromossomo Y, os paragrupos são tipicamente representados por um asterisco (*) colocado após o haplogrupo principal. Os paragrupos podem não possuir quaisquer marcadores únicos adicionais, além das mutações que

definem o haplogrupo principal, mas é claro que isto não significa que mutações adicionais não possam ser descobertas. Desta maneira, ao serem encontrados novos marcadores que os caracterizem, podem ser renomeados como um novo subclado independente.

2.2.3 Filogenia e distribuição geográfica de haplogrupos do cromossomo Y

Como já foi descrito, os marcadores localizados na NRY são passados de pai para filho, em bloco, na forma de haplótipo, possibilitando a aquisição de relevante informação filogenética (JOBILING *et al.*, 1997; SCHULTES *et al.*, 1998; JOBILING & TYLER-SMITH, 2003). Estes marcadores localizados na NRY são mais susceptíveis à deriva genética do que marcadores localizados nos demais cromossomos. Isto se deve ao fato de que, numa população com aproximadamente o mesmo número de homens e mulheres, os cromossomos Y se encontram na proporção de 1/4 em relação aos autossomos e de 1/3 em relação aos cromossomos X (JOBILING & TYLER-SMITH, 2003). Esta susceptibilidade produz uma maior divergência entre populações consideradas a partir de marcadores da NRY em relação àquelas comparadas através de marcadores autossômicos (JORDE *et al.*, 2000).

Na Tabela 2, são apresentados exemplos de marcadores bialélicos do cromossomo Y representativos de alguns dos haplogrupos (HG) principais, e na Figura 11 a respectiva distribuição populacional, indicando as regiões e/ou populações onde a sua frequência é mais elevada.

Os Y-SNPs têm sido utilizados há vários anos e novos polimorfismos são introduzidos a cada dia. Após os estudos realizados por Karafet e colaboradores (2008), várias outras publicações têm atualizado as ramificações da árvore filogenética [e.g. TROMBETTA *et al.*, 2011; YAN *et al.*, 2011; MENDEZ *et al.*, 2013]. Diversas mudanças na filogenia foram introduzidas a partir de estudos de resequenciamento do cromossomo Y e, nos últimos anos, novos Y-SNPs têm sido definidos melhorando o nível de resolução da árvore filogenética. Este fato favorece a frequente atualização dos haplogrupos, expandindo de forma significativa a resolução das ramificações.

Tabela 2 – Exemplos de marcadores bialélicos do cromossomo Y, que representam alguns dos haplogrupos (HG) principais, e a respectiva distribuição populacional, indicando as regiões e/ou populações onde a sua frequência é mais elevada

HG	SNP	número rs	Mutação	Distribuição populacional
A	M91	rs2032651	9T→8T	África (populações Khoisan, Etiópia, Sudão)
B	M60	rs2032623	Ins 1bp	África subsariana (elevada frequência em Pigmeus, África central)
C	M216	rs2032666	C→T	Ásia, Austrália, Oceânia e populações Na-Dene da região norte da América do norte.
D	M174	rs2032602	T→C	Ásia Central (Himalaias, Tibete) e Japão
E	SRY ₄₀₆₄	rs9786608	G→A	África (elevadas frequências em todo o continente), Médio Oriente e Mediterrâneo
G	M201	rs2032636	G→T	Médio Oriente, Mediterrâneo e Cáucaso
H	M69	rs2032673	T→C	Sub-continente Indiano e populações Roma (ciganas) europeias
I	M170	rs2032597	A→C	Europa (frequências elevadas no norte da Europa, Escandinávia)
J	M304	rs13447352	A→C	Médio Oriente, Norte de África, Europa, Ásia Central, Paquistão e Índia
L	M22	rs3913	A→G	Médio Oriente, Norte de África, Europa, Ásia Central, Mediterrâneo
M	P256	-	G→A	Oceânia e Indonésia Oriental (Papua Nova Guiné e Melanésia)
N	M231	rs9341278	G→A	Norte da Eurásia
O	M175	rs2032678	Del 5bp	Ásia Oriental, Ásia Central e Oceânia
Q	M242	rs8179021	C→T	Norte da Eurásia (Sibéria), elevada frequência em populações nativas da América.
R	M207	rs2032658	A→G	Europa (frequências elevadas no oeste Europeu), Ásia central e ocidental
S	P202	-	T→A	Oceânia e Indonésia
T	M70	rs2032672	A→C	Médio Oriente, África, Europa, Ásia Central, Mediterrâneo

Fonte: Adaptado de Gusmão & Alves, 2010.

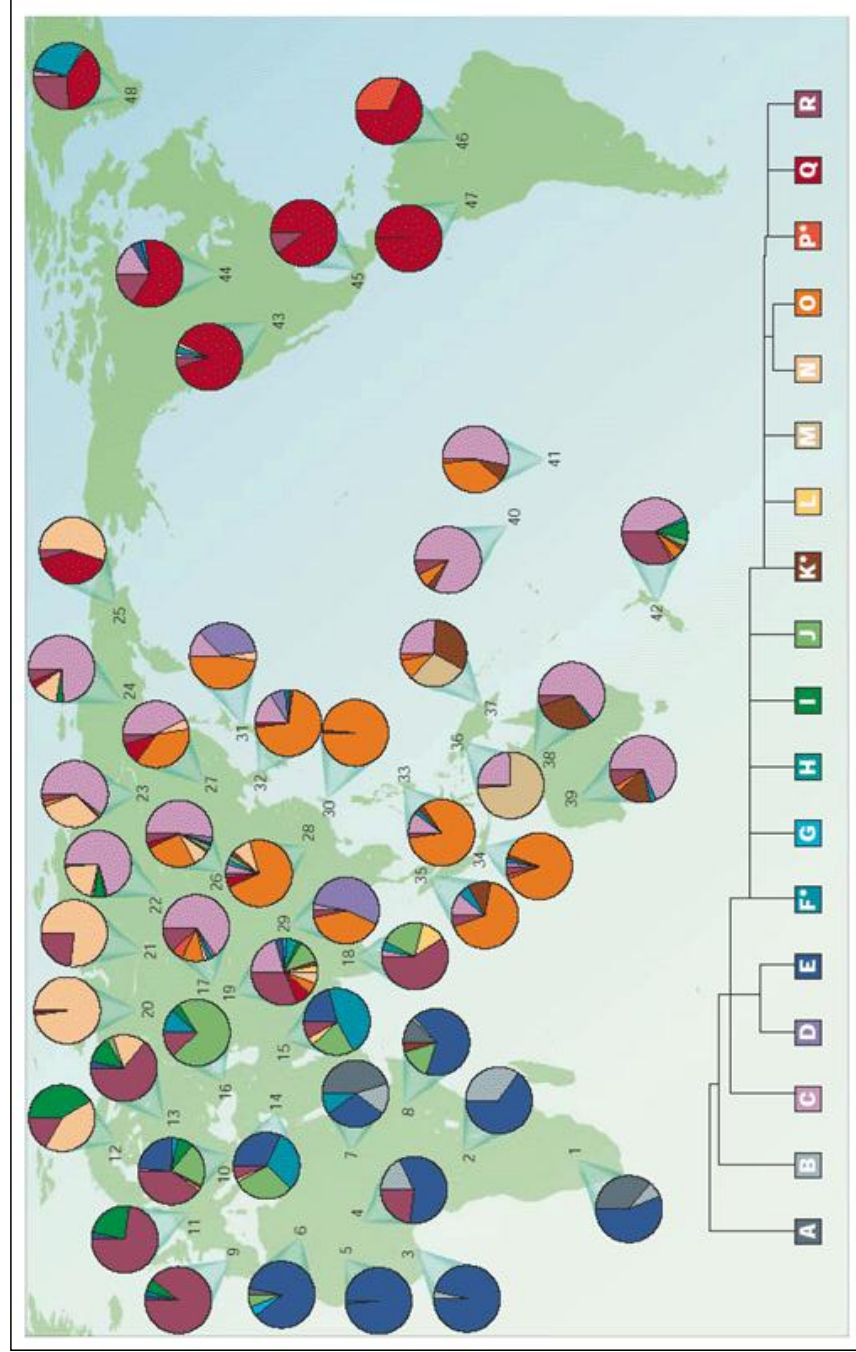
Todos os clados que não são exclusivamente africanos pertencem ao macro-haplogrupo C - T, que são definidos pelas mutações M168, M294 e P9 (SCOZZARI *et al.*, 2012). O haplogrupo C e suas linhagens são encontradas entre populações do continente asiático, na Oceania, entre Ameríndios (HAMMER *et al.*, 2001; ZEGURA *et al.*, 2004), Nova Guiné e na Austrália (HUDJASHOV *et al.*, 2007). A mutação M168 traduz a assinatura da mais recente e bem sucedida migração humana através da África e, além disto, representa a raiz dos haplogrupos de D até T (UNDERHILL *et al.*, 2001).

O haplogrupo F é a origem dos haplogrupos de G a T, presentes em mais de 90% da população mundial. Foi originário da migração para fora da África (“*Out of Africa*”), sendo também observado no subcontinente Indiano em frequências moderadas e baixas (KIVISILD *et al.*, 2003; SENGUPTA *et al.*, 2006). Estudos sugerem que a origem do haplogrupo G talvez seja a Índia ou Paquistão e que houve a sua dispersão para Ásia Central e Oriente Médio (KARAFET *et al.*, 2008), sendo frequente também em populações do Mediterrâneo e nas Montanhas do Cáucaso (JOBILING & TYLER-SMITH, 2003; BEHAR *et al.*, 2004; CINNIUGLU *et al.*, 2004; SENGUPTA *et al.*, 2006).

O haplogrupo H é definido pela mutação M69 (KARAFET *et al.*, 2008) e na população da Índia observa-se elevada frequência, com estimativas entre 15,8% (SENGUPTA *et al.*, 2006) e 21% (KIVISILD *et al.*, 2003), sendo ainda encontrado no Paquistão (6,3%) e no Leste da Ásia (0,57%) (SENGUPTA *et al.*, 2006). Em ciganos da Europa sua frequência também é elevada chegando a valores que variam de 17% em Portugal a até 80% na Bulgária (GUSMÃO *et al.*, 2008). A distribuição do haplogrupo J estende-se pelo Oriente Médio, Norte da África, Europa, Ásia Central, Paquistão e Índia (HAMMER *et al.*, 2001; UNDERHILL *et al.*, 2001) sendo que o haplogrupo mais comum presente na Europa é o J2 - M172 (SEMINO *et al.*, 2004).

A maioria das linhagens do haplogrupo L é encontrada no subcontinente Indiano, ocorrendo também no Oriente Médio, Ásia Central, Norte da África e ao longo da costa do Mediterrâneo na Europa (KIVISILD *et al.*, 2003; BEHAR *et al.*, 2004; CINNIUGLU *et al.*, 2004; SENGUPTA *et al.*, 2006; KARAFET *et al.*, 2008).

Figura 11- Representação gráfica da distribuição dos haplogrupos do cromossomo Y, indicando onde as frequências são mais elevadas



Fonte: Jobling & Tyler-Smith, 2003).

Com distribuição geográfica restrita à Oceania e ao Leste da Indonésia, o haplogrupo M também se encontra presente na maioria das linhagens masculinas em Papua, Nova Guiné, e na Melanésia (SU *et al.*, 2000; KAYSER *et al.*, 2006; SCHEINFELTD *et al.*, 2006).

Na população finlandesa e de nativos siberianos encontramos com mais frequência o haplogrupo N (CINNIOGLU *et al.*, 2004). Estas linhagens N são encontradas no Norte da Eurásia, estando ausentes ou pouco frequentes em outros lugares (KARAFET *et al.*, 2001; ROOTSI *et al.*, 2007).

Relativamente ao oeste asiático encontramos o haplogrupo O, com alta frequência (cerca de 80% da população) sendo um grupo particularmente restrito a estas populações. Predomina no Leste da Ásia ocorrendo também, mas com menor frequência, na Ásia Central (SU *et al.*, 1999; KARAFET *et al.*, 2001; UNDERHILL *et al.*, 2001).

Existe um haplogrupo que é considerado indiferenciado e que raramente é encontrado, denominado de haplogrupo P e de onde derivam os haplogrupos Q e R. Estes dois últimos haplogrupos representam linhagens com amplas distribuições geográficas, sendo caracterizados pelas mutações 92R7, M45 e M74, entre outras (KARAFET *et al.*, 2001). O haplogrupo Q é definido pela mutação M242 (KARAFET *et al.*, 2008), sendo amplamente distribuído no Norte da Eurásia e ocorrendo com frequências elevadas em algumas populações da Sibéria. É encontrado também na Europa, Leste da Ásia e no Oriente Médio, contudo nestas regiões a sua frequência é baixa (KARAFET *et al.*, 2002; KARAFET *et al.*, 2008). O haplogrupo derivado Q1a3a-M3 é a principal linhagem entre ameríndios e sua presença é quase totalmente restrita às Américas (JOBILING & TYLER-SMITH, 2003; ZEGURA *et al.*, 2004). Battaglia e colaboradores (2013) realizaram novos estudos ampliando as ramificações do haplogrupo Q, e definiram novas sublinhagens mais modernas e com distribuição geográfica restrita, tendo como exemplo o haplogrupo Q1a3a1-L54 (x M3), que não é encontrado na Ásia, mas somente no sudeste siberiano.

2.2.4 Principais haplogrupos africanos

O polimorfismo do tipo *Alu* localizado no cromossomo Y, ou também conhecido como *YAP element* (HAMMER, 1994), têm sido um dos mais usados para diferenciar as populações africanas das não africanas. Este polimorfismo foi resultado de uma simples e estável inserção de um membro da família *Alu* em um sítio específico (*locus* DYS287) no braço longo do cromossomo Y entre 29.000-334.000 anos atrás (HAMMER, 1995). A frequência deste elemento varia grandemente nas populações humanas de diferentes regiões geográficas (HAMMER 1994; SPURDLE *et al.*, 1994a, 1994b; HAMMER & HORAI 1995; HAMMER *et al.*, 1997). Pesquisas globais mostram que, as populações da África subsaariana têm maior frequência deste marcador, presente em 80% das populações de origem Bantu, sendo seguida pelas populações do nordeste africano, Ásia, Europa, Américas e Oceania. Postula-se que a inserção YAP tenha ocorrido inicialmente no cromossomo Y africano e, subseqüentemente, se espalhado por outros continentes (HAMMER *et al.*, 1994).

Na África, temos as raízes da filogenia das populações atuais e também as impressões dos fenômenos de movimentação das populações, em especial a Bantu, que se expandiram pelo continente há pelo menos 2.700 anos (JOBILING & TYLER-SMITH, 2003). O haplogrupo A, que representa o mais antigo e diverso dos haplogrupos do cromossomo Y, é considerado o haplogrupo fundador (Cromossomo Y de Adão), dos quais todos os outros derivaram. É restrito basicamente à África e apresenta-se mais frequentemente entre populações Etíopes, Sudanesas e grupos de Khoisan (HAMMER *et al.*, 2001; UNDERHILL *et al.*, 2001; KARAFET *et al.*, 2008). Há alguns anos, estudos feitos por Scozzari e colaboradores (2012) promoveram a definição de novos haplogrupos que permitiram a ampliação das ramificações do haplogrupo A.

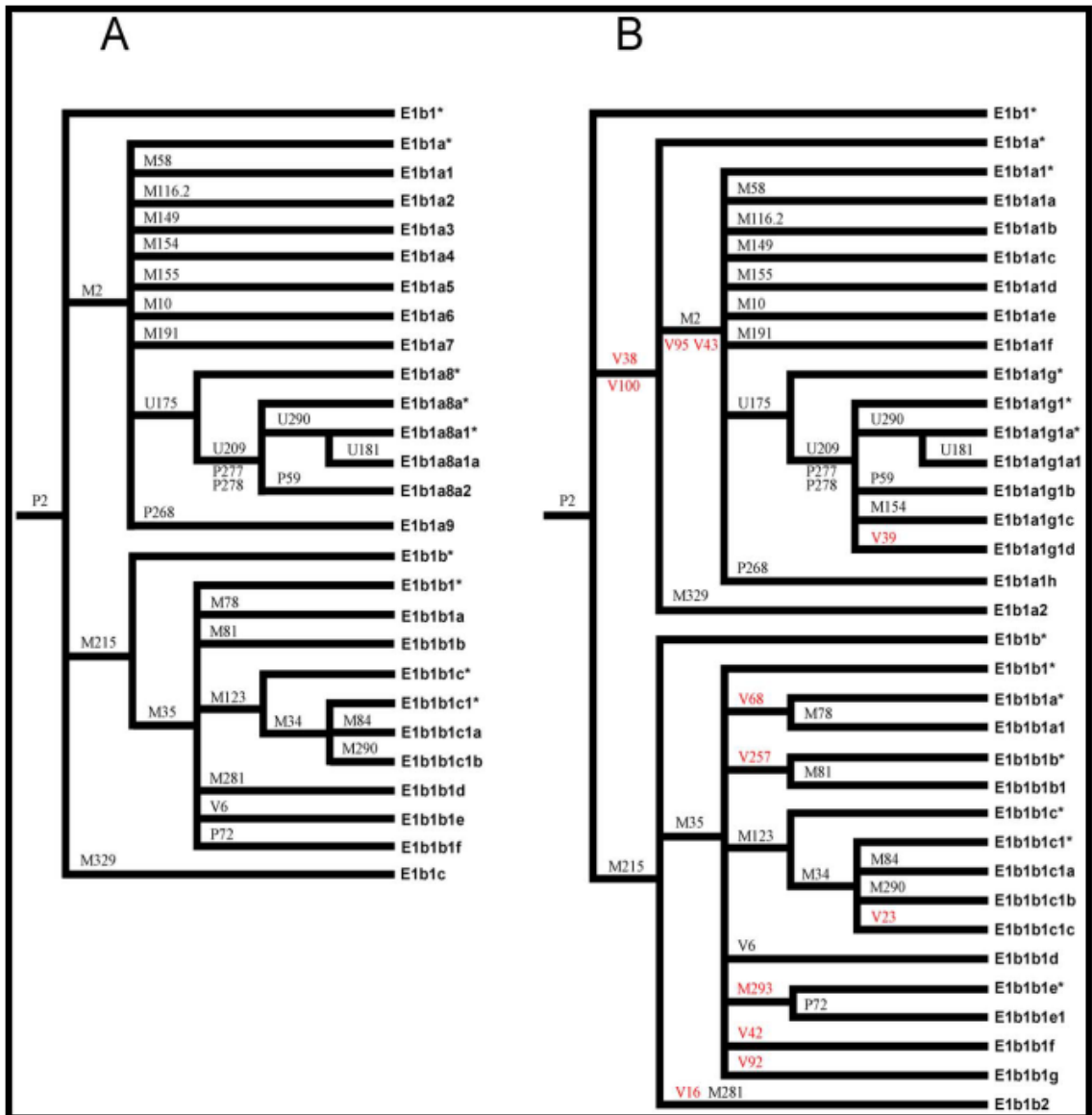
O haplogrupo B é um dos mais antigos das linhagens do cromossomo Y humano e a maioria destas é bem definida e restrita, tanto geograficamente quanto etnicamente. Encontra-se exclusivamente na África, sendo o primeiro haplogrupo a se dispersar pelo continente e aparecendo em baixas frequências. Contudo, é um haplogrupo muito prevalente entre a população de Pigmeus, ocorrendo também em

algumas populações do Centro-Oeste, Centro e Sul da África (UNDERHILL *et al.*, 2001; SEMINO *et al.*, 2002; CRUCIANI *et al.*, 2002).

A grande maioria dos cromossomos Y africanos pertence ao haplogrupo E, que é o grupo que possui a maior diversidade de todos, ramificando-se em 55 haplogrupos caracterizados por 83 marcadores, mais o parágrafo E* (KARAFET *et al.*, 2008). Muitas mutações descritas anteriormente, com elevadas frequências na África, moderadas no Oriente Médio e no Sul da Europa e ocorrendo raramente nas regiões Central e Sul da Ásia (HAMMER *et al.*, 1998; UNDERHILL *et al.*, 2001; CRUCIANI *et al.*, 2002; JOBLING & TYLER-SMITH, 2003). Muitas destas mutações foram resequenciadas, resultando em novas alterações nas ramificações da árvore filogenética (TROMBETTA *et al.*, 2011) (Figura 12).

O haplogrupo tipicamente africano que apresenta a maior distribuição geográfica é o haplogrupo E1b1, caracterizado por um alto grau de diversidade interna, sendo o mais representativo. Existem três ramificações deste haplogrupo que são bem elucidativas quanto às frequências no continente africano. O haplogrupo E1b1a*- M2 é o mais comum e amplamente distribuído na África subsaariana, o que tem sido atribuído à expansão Bantu (UNDERHILL *et al.*, 2001; CRUCIANI *et al.*, 2002). Sua frequência é de cerca de 80% na região oeste e 60% na região central, sendo encontrado também no norte da África, embora sendo bem mais baixas (usualmente abaixo de 10%). Outra ramificação representativa é do haplogrupo E1b1b - M215 que apresenta uma larga distribuição desde o sudeste da Europa até o nordeste e sudeste africano e o E1b1c - M329 que foi observado quase exclusivamente no sudeste africano.

Figura 12 - Exemplo de alterações filogenéticas a partir de estudos do haplogrupo E1b1 feitos por Trombetta *et al.* (2011)



Legenda: (A) é a representação segundo Karafet *et al.* (2008) e (B) é a representação com as alterações do autor (Trombetta *et al.*, 2011).

2.2.5 Principais haplogrupos europeus

Vários tipos de evidências sugerem que o *pool* gênico da população europeia atual surgiu de grupos caçadores-coletores do período paleolítico (46.000-35.000 anos atrás) e agricultores do Neolítico (há 12.000-4.000 anos), vindos do leste da região do Oriente Médio (atualmente, Israel, Cisjordânia e Líbano, partes da Jordânia, Iraque, Egito e sudeste da Turquia) (SEMINO *et al.*, 2000). Dois haplogrupos I* e R* (mutações M170 e M173, respectivamente) estariam presentes na Europa desde a época do Paleolítico (SEMINO *et al.*, 2000, CRUCIANI *et al.*, 2002). O haplogrupo I é um dos haplogrupos do cromossomo Y mais predominantes na Europa (JOBLING & TYLER-SMITH, 2003) sendo quase ausente em outros locais, o que leva a crer que seu surgimento ocorreu no último período de glaciação (SEMINO *et al.*, 2000; ROOTSI *et al.*, 2004).

O haplogrupo R é principalmente representado por duas linhagens: R1a provavelmente originada das estepes da Eurásia, no norte dos mares Negro e Cáspio, e o R1b. O haplogrupo R1b é o haplogrupo europeu mais comum, e caracteriza-se por apresentar o Haplótipo Modal Atlântico (HMA), que é um conjunto de alelos de Y-STRs mais representativos (DYS19=14; DYS385a=11; DYS385b=14; DYS389I=13; DYS389II=29; DYS390=24; DYS391=11; DYS392=13 e DYS393=13) que caracterizam esta população (WILSON *et al.*, 2001). O haplogrupo R1 corresponde a cerca de 50% dos cromossomos Y da Europa (SEMINO *et al.*, 2000), sendo o haplogrupo R1b1b2 - M269 mais frequente em ibéricos, incluindo os bascos (CINNIOGLU *et al.*, 2004). É encontrado com baixas frequências na África, Oriente Médio, Austrália, Ásia Central, Paquistão, Índia e América (UNDERHILL *et al.*, 2001). Em Portugal, o haplogrupo R1b1b2-M269 é o mais frequente, correspondendo a 57,7% da população (BELEZA *et al.*, 2006), sendo também o haplogrupo mais frequente no Brasil (51,2%) (SILVA *et al.*, 2006).

As linhagens pertencentes ao haplogrupo S são encontradas principalmente na Oceania e Polinésia (KAYSER *et al.*, 2006; SCHEINFELDT *et al.*, 2006). Da mesma maneira, o haplogrupo T também apresenta frequências em regiões específicas, sendo baixas no Oriente Médio, África e Oeste da Eurásia, sendo mais encontrado no Oriente Médio e Nordeste da África, onde ocorre com frequência média de 6,8% (KING *et al.*, 2007).

2.2.6 Principais haplogrupos nativos da América

Estudos realizados por Zegura *et al.* (2004) mostraram que a diversidade genética das populações da América é relativamente baixa, sendo a maioria dos indivíduos pertencentes ao haplogrupo Q (Q1a3*- M346 e Q1a3a-M3) e com uma baixa porcentagem de indivíduos sendo pertencentes ao haplogrupo C3*-M217 (ROEWER *et al.*, 2013), que é muito frequente em grupos mongóis e bastante disperso entre populações do extremo leste da Rússia e Ásia Oriental (GEPPERT *et al.*, 2011). Contudo, algumas sublinhagens (Q1a3a2-M194, Q1a3a3-M199 e Q1a3a4-SA01) podem ser vistas com distribuição restrita em algumas populações (BISSO-MACHADO *et al.*, 2012).

Em populações andinas foram encontrados novos haplogrupos, tais como o Q-M557 e Q-PV2 (ROEWER *et al.*, 2013). O haplogrupo Q-M19 foi originalmente encontrado na população de indígenas das tribos Ticuna (60%) e Wayuu (10%) na Colômbia e mais recentemente em indivíduos da população Toba na Argentina (4%) (GEPPERT *et al.*, 2011). No geral, a América do Sul apresenta uma homogeneidade de distribuição de haplogrupos dentro de Q, com uma grande prevalência do haplogrupo Q-M3. O haplogrupo Q1a3*-M346 é o segundo mais representado, e se apresenta com uma larga distribuição na América do Sul, sendo encontrado principalmente no nordeste e na região do Chaco (região dos pampas compreendida entre Argentina, Bolívia, Paraguai e Brasil) (ROEWER *et al.*, 2013).

Estudos realizados em populações brasileiras demonstraram que o haplogrupo Q1a3a-M3 é o mais frequente, sendo encontrado em populações com origem ameríndia dispersas em todo território brasileiro (BISSO-MACHADO *et al.*, 2012).

Devido a esta grande variedade de marcadores presentes no cromossomo Y, que podem contar a nossa história genética, deve-se ter em mente que, para o estudo sistemático da contribuição genética utilizando os Y-SNPs dentre todos os haplogrupos do YCC, alguns devem ser destacados em razão da composição da população a ser estudada. Em vista da extensa miscigenação ocorrida durante o processo de colonização brasileira e da impossibilidade de conhecer as origens geográficas de algumas populações, marcadores preferenciais devem ser

escolhidos e devem contemplar o maior número de haplogrupos possível, de modo a servir como referência para estabelecer critérios de origem étnica.

2.3 Estado da arte dos estudos de polimorfismos genéticos no Brasil

2.3.1 Estudos com marcadores polimórficos

No Brasil, estudos utilizando polimorfismos de proteínas, grupos sanguíneos, RFLP, sequências de mtDNA e alguns STRs, realizados entre as décadas de 1970 e 1990 em populações do nordeste e sudeste, já indicavam um indiscutível grau de miscigenação, mas o padrão apresentava-se de forma variável de acordo com a região estudada (SALZANO, 1997; GUERREIRO *et al.*, 1999).

Estudos com marcadores recombinantes localizados nos autossomos revelaram que as populações miscigenadas brasileiras, apresentaram uma maior prevalência de ancestralidade europeia, com frequências comparativamente mais elevadas no sul do país, do que na região norte (PENA *et al.*, 2011), sendo observado que a contribuição africana no nordeste do Brasil foi maior do que a contribuição do componente nativo americano, ao passo que, no norte o componente nativo americano é bastante alto em comparação com as demais regiões do país (SALZANO *et al.*, 2002; PENA *et al.*, 2011). Costa-Francez *et al.* (2011), em estudos realizados na população do Macapá (região amazônica), baseados na frequência alélica utilizando 12 STRs, estimaram a mistura étnica de três grupos parentais. A partir destas análises, concluíram que a população estudada apresentava 35% de componente nativo americano, com 46% e 19% de contribuição europeia e africana, respectivamente. Por outro lado, os estudos em populações prioritariamente brancas da região sul do Brasil (Rio Grande do Sul e Santa Catarina) revelaram, como seria esperado, maior contribuição de genes de origem europeia (DORNELLES *et al.*, 1999), com as influências africana e nativo americanas reduzidas, quando comparadas com outras regiões do Brasil (SALZANO *et al.*, 2002).

Algumas destas análises, também efetuadas em comunidades afrobrasileiras, revelaram uma maior contribuição de genes de origem africana em relação à população em geral (SCHEINDER *et al.*, 1987; GUERREIRO *et al.*, 1999). Algumas estimativas sobre a contribuição africana na população também foram feitas a partir de estudos em genes que codificam a β -globina, mostrando que os afrodescendentes dos remanescentes de quilombos do Vale do Ribeira (Sudeste do Brasil) com traços falcêmicos, apresentaram uma frequência de 81% do haplótipo Bantu (conhecido como tipo da República Central da África - CAR), seguidos dos haplótipos Benin, Senegal e Camarões (8,1%, 8,1%, 2,3%, respectivamente), que são regiões africanas tipicamente Bantu (ZAGO *et al.*, 1992). Esta proporção também corrobora os relatos históricos de que, grande parte dos escravos, seria de origem Bantu (AURICCHIO *et al.*, 2008).

Manta e colaboradores (2013) revelaram que, quando as populações brasileiras foram analisadas a partir de marcadores de linhagem do tipo Indels (AIMs), observou-se uma acentuada assimetria de cruzamento, com uma predominância de linhagens paternas de origem europeia e uma contribuição africana e nativa de origem materna. Este padrão tem sido encontrado na maioria das populações miscigenadas da América, onde predominaram os cruzamentos entre homem europeu e mulheres de origens africana e nativa.

2.3.2 Estudos com polimorfismos do cromossomo Y

A partir de estudos com STRs do cromossomo Y, tem sido verificada a ocorrência de similaridade genética entre as várias populações de diferentes regiões do Brasil e a de Portugal (GRATTAPAGLIA *et al.*, 2005; GÓES *et al.*, 2005; SCHWENGBER *et al.*, 2009, PALHA, *et al.*, 2012) devido à predominância de linhagem masculinas de origem europeia. Mesmo em indivíduos do Rio de Janeiro que se autodeclararam afrodescendentes, observa-se que existe também uma maior proximidade genética com a população masculina portuguesa do que com populações africanas de um modo geral (DOMINGUES *et al.*, 2007). Podemos assim perceber que, em relação aos Y-STR, estes resultados mostraram que existe uma concordância entre os fatos históricos e os resultados moleculares,

evidenciando uma assimetria de cruzamento que envolveu principalmente homens de origem europeia com as mulheres de origem tanto africana quanto nativo americana.

Utilizando marcadores bialélicos (SNPs) da NRY em amostras populacionais brasileiras foi observado que a maioria dos haplogrupos encontrados eram de origem europeia (entre 88,2% e 100%), ao passo que os haplogrupos de origem africana apresentaram-se menos frequentes (entre 0% e 7,9%) e os de origem ameríndia bastante raros (entre 0% e 1,6%) (CARVALHO-SILVA *et al.*, 2001; ABE-SANDES *et al.*, 2004; SILVA *et al.*, 2006). CARVALHO-SILVA *et al.* (2001), estudando populações brasileiras constatou que, em grupos de pessoas classificadas fenotipicamente como caucasianas, os marcadores bialélicos pertencentes aos haplogrupos de origem europeia se apresentavam numa frequência de 54%. Contudo, no Rio Grande do Sul esta frequência foi ainda mais alta, variando entre 58% e 93% (MARRERO *et al.*, 2005; MARRERO, 2006).

Outro estudo feito com 16 Y-SNPS em afrobrasileiros do estado de São Paulo (GONÇALVES *et al.*, 2008) demonstrou que 42% das linhagens de cromossomo Y eram originárias da África subsaariana (linhagens E1b1a7 –M191 e E1b1a1-M2), 3% de outra origem africana (haplogrupo B) e apenas 1,7% eram de origem ameríndia (haplogrupo Q). Por outro lado, em relação às linhagens de origem europeia, foi observada uma frequência de 42,5% de linhagens pertencentes a haplogrupos europeus incluídos no parágrafo F, excluindo o haplogrupo Q1a3a-M3 (tipicamente nativo americano). Apesar de elevada, a contribuição paterna de origem africana se mostrou significativamente mais baixa do que a observada para linhagens mitocondriais (85%), analisadas através do estudo das regiões hipervariáveis HV1 e HV2 do DNA mitocondrial. Esta diferença de ancestralidade acentuada entre linhagens de origem paterna e materna indicam o fenômeno de cruzamento direcionado entre homens europeus e mulheres africanas, o qual tem sido frequentemente descrito em populações miscigenadas da América do Sul (VULLO *et al.*, 2014).

Nos trabalhos apontados, foram investigados marcadores que permitiram uma classificação inequívoca destas linhagens como sendo europeias. No entanto, o baixo número de SNPs analisados não permitiu uma discriminação mais minuciosa, que fosse capaz de descortinar a origem precisa das linhagens encontradas, dentro daquelas encontradas tipicamente na Europa. Na realidade,

não foram investigados marcadores dentro do haplogrupo R*, o que incluiria o seu principal representante em populações da península ibérica (R1b1b2 - M269). SILVA e colaboradores (2006) incluíram em sua investigação o marcador M269 e observaram que o haplogrupo R1b1b2 era o mais frequentemente encontrado (58%) em uma amostra de indivíduos do estado do Rio de Janeiro. De fato, este estudo permitiu concluir que o componente europeu observado se devia, principalmente, a uma grande contribuição de origem portuguesa no processo de formação da população do Rio de Janeiro, o que mais uma vez, se apresenta em conformidade com os dados históricos.

Por outro lado, OLIVEIRA *et al.* (2009), em estudo preliminar, utilizando o marcador binário do tipo *Alu*-YAP, que tem uma frequência 10 vezes maior em populações da África subsaariana (~80% em populações Bantu), demonstraram que 44% dos indivíduos autodeclarados afrodescendentes do estado do Rio de Janeiro apresentavam esta mutação. De maneira semelhante, NASCIMENTO *et al.* (2009), também em estudo preliminar realizado em uma amostra da população do estado da Bahia, observaram que o haplogrupo E1b1a - M2 estava representado em 19% dos cromossomos Y analisados.

A análise de marcadores bialélicos do Y em amostras populacionais de indivíduos afrobrasileiros de Salvador e das comunidades remanescentes de quilombolas do Rio das Contas e Barra revelou uma variação considerável nas proporções de linhagens africanas, europeias e ameríndias. Nestas populações a frequência do haplogrupo africano E1b1a - M2 variou entre 41,1 e 77,3% e o haplogrupo ameríndio Q1a3a1a* - M3, encontrado somente na amostra de Salvador, apresentou uma frequência de 4% (ABE-SANDES *et al.*, 2004).

Quando pensamos na história da formação de uma população, frequentemente nos remetemos aos fatos históricos disponíveis. Mas, a partir do surgimento de novas tecnologias utilizando o DNA e o crescimento das análises através dos seus polimorfismos, muitos eventos históricos podem ser confirmados, revalidados ou até mesmo reavaliados.

Como já visto, a colonização brasileira foi o resultado de fenômenos muito particulares de miscigenação. Às populações mais ou menos numerosas de indígenas existentes em muitas regiões, vieram juntar-se números também variáveis de indivíduos de diferentes regiões do globo, sendo a grande maioria deles de origem portuguesa e africana. Isto pode ser constatado em qualquer região

do país, evidenciando uma elevada heterogeneidade genética ao longo do território. Hoje temos registros históricos até certo ponto confiáveis, mas ocorrem algumas lacunas acerca da origem de um dos principais elementos que contribuíram na nossa formação, os africanos. Muitos trabalhos foram publicados a partir de estudos sobre ancestralidade em diversas populações no Brasil, contudo, os focos de interesse geralmente baseiam-se na determinação da contribuição de cada grupo continental para a formação da nossa população. Quanto aos locais de origem a partir do continente africano, pouco se tem estudado em relação às populações afrodescendentes brasileiras atuais. Em vista disto, muitas perguntas devem ser respondidas de modo a entender como ocorreu esta dispersão dos povos/etnias africanos no Brasil e, quais as proporções destes que foram trazidos como escravos, assim como a sua distribuição nas diferentes localidades do país.

Com base nas lacunas existentes no tocante à contribuição para o patrimônio genético vigente no Brasil dos povos que formaram a base inicial da população brasileira e a falta de dados concretos relacionados aos marcadores de ancestralidade, principalmente de populações afrodescendentes, a proposição deste estudo é acrescentar informações, associando a ancestralidade africana e as possíveis regiões de origem no continente africano, em relação às populações afrobrasileiras do Rio de Janeiro, e avaliar em que medida o patrimônio genético de origem africana prevalente está de acordo com os dados históricos disponíveis – relativamente aos locais de origem na África.

Por meio de análises moleculares de polimorfismos presentes no cromossomo Y utilizando STRs e SNPs, novas informações sobre a origem da nossa ancestralidade poderão ser conhecidas.

À exceção de pequenas comunidades nativas ou grupos de populações afrodescendentes, todos os trabalhos feitos até o presente momento em populações brasileiras miscigenadas, apontam para uma maior prevalência da ancestralidade europeia. Em relação às linhagens paternas, esta prevalência é ainda mais acentuada, evidenciando o papel das mulheres como mediadoras da herança africana e nativo americana das populações brasileiras.

3 OBJETIVOS GERAIS

Muitas perguntas devem ser respondidas de modo a entender quais as contribuições dos diferentes povos/etnias que foram trazidos como escravos e como se deu a sua distribuição nas diferentes regiões do país. Em vista disto, ao tentar responder a estas perguntas, contribuiremos para aumentar o entendimento do mecanismo da dispersão das linhagens masculinas dos povos africanos no Brasil.

Pretendeu-se, por um lado, caracterizar de forma minuciosa uma amostra representativa da população masculina do estado do Rio de Janeiro, utilizando para isto um número elevado de marcadores genéticos específicos do cromossomo Y, tanto para STRs quanto para SNPs.

Tendo em conta a escassez de estudos e, conseqüentemente, a falta de dados genéticos relativos às populações do oeste africano, pretendeu-se ainda, estudar amostras de indivíduos naturais dos principais locais de origem dos escravos que foram trazidos para a América. Então, além das amostras do Rio de Janeiro, também foram incluídas neste trabalho, amostras biológicas provenientes de casos de paternidade de indivíduos originários de 11 países do oeste africano (Senegal, Gâmbia, Guiné, Serra Leoa, Costa do Marfim, Gana, Togo, Benin, Nigéria, Níger e Libéria), que também foram caracterizadas quanto aos mesmos marcadores específicos do cromossomo Y. Estas análises foram feitas em colaboração com a equipe de pesquisadores coordenada pelo Professor Dr. Peter Schneider, sendo todo o trabalho experimental efetuado no laboratório da Divisão de Genética Molecular Forense, situado no Instituto de Medicina Legal de Colônia na Alemanha.

3.1 Objetivos Específicos

De forma a cumprir o objetivo geral, foram assim determinados os três seguintes objetivos específicos:

a) Caracterizar o patrimônio genético de uma amostra de indivíduos do sexo masculino, não relacionados, da população do estado do Rio de Janeiro, quanto aos polimorfismos do tipo STR e SNP localizados na região não recombinante do cromossomo Y. Elaborar uma base de dados para a comparação com outras populações brasileiras. Classificar as diferentes linhagens masculinas existentes na população do Rio de Janeiro e identificar quais são as de origem africana, calculando as correspondentes frequências. Os dados de frequência dos marcadores das nossas populações serão comparados com aqueles já publicados entre várias populações do Brasil e do mundo. Com base no patrimônio genético comprovadamente africano, comparar as amostras brasileiras com aquelas já descritas na literatura, através da análise das distâncias genéticas entre pares de populações.

b) Caracterizar o patrimônio genético de amostras de indivíduos não relacionados do Oeste da África, quanto aos polimorfismos do tipo STR e SNP localizados na região não recombinante do cromossomo Y. Conjuntamente, com outros dados já publicados para os mesmos marcadores em outras populações africanas, elaborar uma base de dados representativa das três principais regiões da África que, historicamente, foram fonte de envio de escravos para o Brasil. Avaliar a estratificação existente entre populações da África subsaariana.

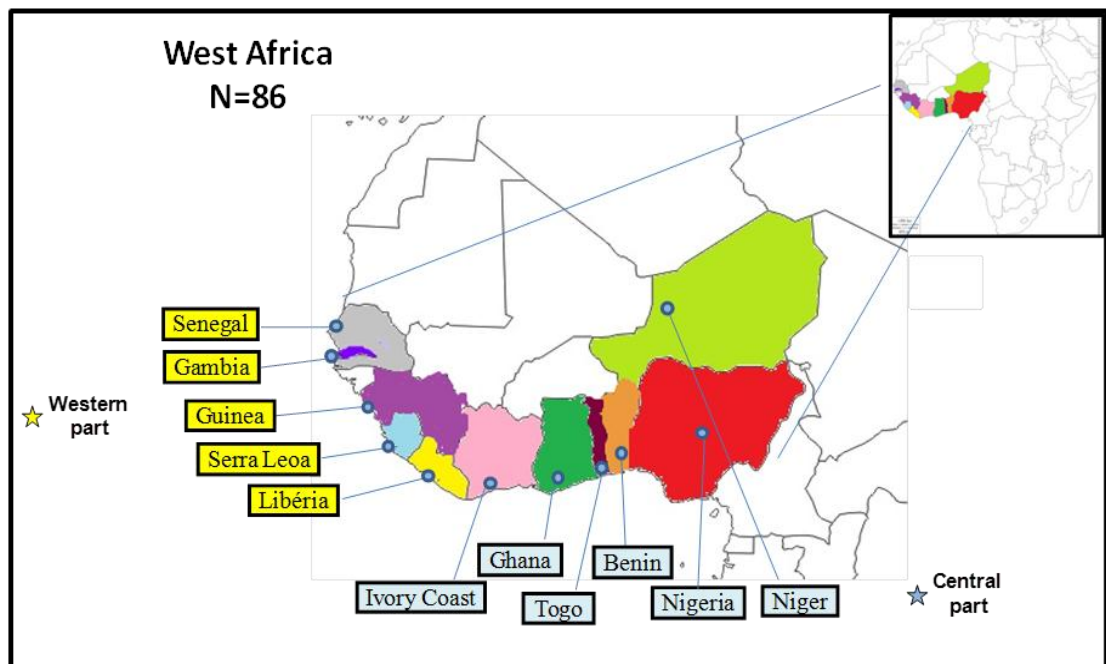
c) Efetuar, no patrimônio genético comprovadamente africano presente na amostra brasileira, a análise estatística dos dados que permitam inferir, através de análises comparativas, quais as principais populações da África subsaariana que terão contribuído para a composição das populações afrobrasileiras.

4 MATERIAL E MÉTODOS

4.1 Amostragem e obtenção do DNA

No âmbito desta tese, foram estudadas duas amostras populacionais, compostas por indivíduos do sexo masculino, não relacionados, representativos do estado do Rio de Janeiro e da região oeste da África subsaariana (Figura 13).

Figura 13 – Mapa indicando a origem da amostra incluída e analisada neste estudo:
Oeste da África



Legenda: África subsaariana, com os países que foram amostrados representados em várias cores.

As amostras biológicas dos indivíduos da população do Rio de Janeiro, incluídas neste estudo, são provenientes de exames de paternidade realizados no Laboratório de Diagnóstico por DNA da Universidade do Estado do Rio de Janeiro (LDD-UERJ). Foram coletadas amostras de 605 indivíduos do sexo masculino sem vínculo genético conhecido. A utilização destas amostras biológicas para estudo científico, de forma não identificada, foi feita com o consentimento por escrito do

doador (TCLE - Termo de Consentimento Livre e Esclarecido). O TCLE foi aprovado pelo Comitê de Ética da Universidade do Estado do Rio de Janeiro (CAAE:0067.0.228.000-09), para serem utilizadas no presente projeto. Amostras de sangue periférico foram coletadas através de punção venosa em tubos contendo EDTA e foram mantidas a 4° C até a data da extração do DNA.

Foram utilizados dois métodos de extração de DNA: *salting-out* descrito por Miller *et al.* (1988) e também o método de extração utilizando Chelex®, uma resina derivada do estireno (WALSH *et al.*, 1991), para as amostras da população do Rio de Janeiro. As amostras de DNA extraídas por *salting-out* foram quantificadas por espectrofotometria utilizando o NanoDrop modelo 2000 (Thermo Scientific), sendo posteriormente armazenadas e mantidas em concentração padrão de 2 ng/μL. As demais amostras do Rio de Janeiro (extraídas por Chelex) não foram quantificadas, mas todas as análises apresentaram resultados compatíveis com os obtidos com as amostras extraídas por *salting-out*, com concentrações finais de aproximadamente 2 ng/μL.

De forma semelhante, o DNA das amostras de 86 indivíduos de onze países do oeste africano (Senegal=2, Gâmbia=3, Guiné=1, Serra Leoa=2, Costa do Marfim=15, Gana=17, Togo=18, Benin=1, Nigéria=19, Níger=6 e Libéria=1 e um indivíduo sem país especificado) foi extraído de forma automatizada, utilizando o equipamento EZ1-Advanced XL da Qiagen, seguindo as especificações de extração fornecidas pelo fabricante. Todas as análises foram efetuadas na Divisão de Genética Molecular Forense da Universidade de Colônia (Alemanha) e também foram obtidas com consentimento do doador, com a sua utilização previamente aprovada pelo seu Comitê de Ética. O DNA das amostras dos indivíduos da região oeste da África foi extraído e foram quantificados por espectrofotometria, utilizando o equipamento *NanoDrop* modelo 2000 (Thermo Scientific), sendo também armazenados em concentrações de 2 ng/μL.

4.2 Genotipagem

4.2.1 Marcadores do tipo STR do cromossomo Y

Todas as amostras de DNA dos indivíduos foram genotipadas com o sistema comercial AmpF/STR® Y-Filer™ (Thermo Fisher Scientific Inc). Este sistema inclui 17 marcadores do tipo STR, DYS389I, DYS389II, DYS390, DYS456, DYS19, DYS385 (incluindo os *loci* DYS385a e DYS385b), DYS458, DYS391, DYS392, DYS393, DYS439, DYS635, DYS437, DYS438, DYS448 e GATA_H4.

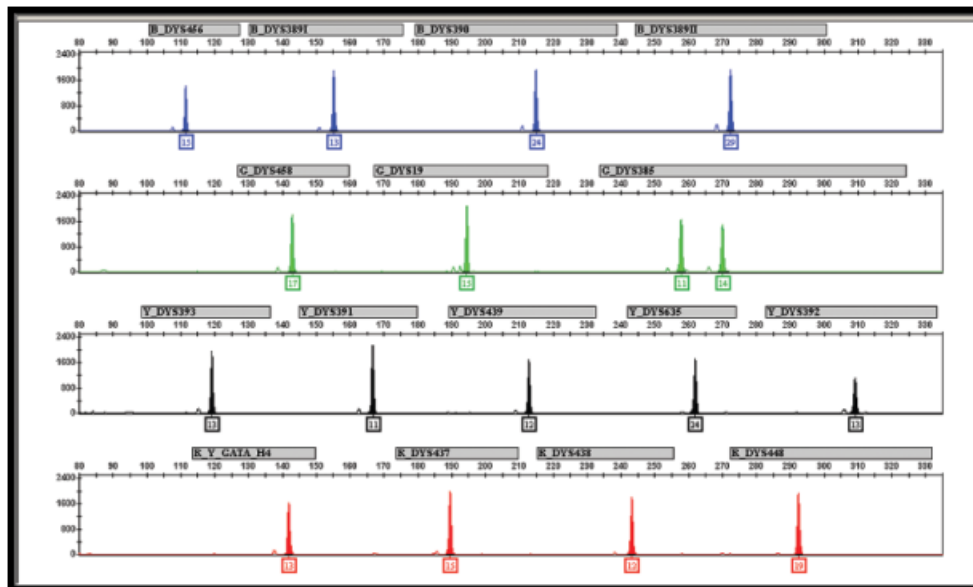
A reação de PCR foi realizada em volume final de 6,25 µL contendo 0,2 µL de AmpliTaq Gold DNA polymerase [5 U/µL] (Thermo Fisher Scientific Inc), 2,3 µL da mistura de reação de PCR e 1,25 µL da mistura de oligonucleotídeos, ambas fornecidas com o kit AmpF/STRs Y-Filer (Thermo Fisher Scientific Inc), e 1 µL de DNA (2 ng/µL). As amostras foram amplificadas em termocicladores PCR system 9700 (Thermo Fisher Scientific Inc).

As condições da PCR foram: 95°C por 11 minutos, seguido de 30 ciclos de desnaturação de 94°C por 1 minuto, pareamento a 61°C por 1 minuto e extensão a 72°C por 1 minuto. O tempo de extensão final a 60°C foi de 80 minutos.

Todos os produtos de amplificação ou *amplicons* foram separados por eletroforese capilar nos sequenciadores automáticos ABI3100, ABI3130 e ABI3500 (Thermo Fisher Scientific Inc) e os resultados foram analisados através do programa GeneMapper® Software v4.1 (Thermo Fisher Scientific Inc). As designações dos alelos foram baseadas na comparação com escalas alélicas padronizadas pelo fabricante, também utilizando como controle positivo o DNA controle do kit (DNA 007) (Figura 14).

Para seguir a recomendação da ISFG (*International Society of Forensic Genetics*) (GUSMÃO *et al.*, 2006), os alelos GATA H4.1 amplificados através do kit AmpF/STRs Y-Filer, tiveram seus valores corrigidos adicionando-se aos resultados obtidos um fator de correção com o valor de 9 (MULERO *et al.*, 2006).

Figura 14 - Representação gráfica (eletroferograma) do resultado da eletroforese capilar dos produtos de amplificação de uma amostra controle (DNA 007), utilizando o kit comercial AmpF/STR® Y-Filer™ (ThermoFisher Scientific Inc)



Fonte: Adaptado de Mulero *et al.*, 2006.

4.2.2 SNPs do cromossomo Y

Com base em estudos prévios (SANCHEZ *et al.*, 2003; BRIÓN *et al.*, 2004; BRIÓN *et al.*, 2005; BLANCO-VEREA *et al.*, 2006; SILVA *et al.*, 2006) e em dados provenientes da história da nossa colonização, foram escolhidos 46 Y-SNPs que caracterizam os principais haplogrupos europeus, africanos e ameríndios que poderiam ser encontrados na nossa população, permitindo definir vários haplogrupos de origens distintas (Anexo D).

Estes 46 Y-SNPs foram genotipados de forma hierárquica, de acordo com o esquema indicado na figura 15.

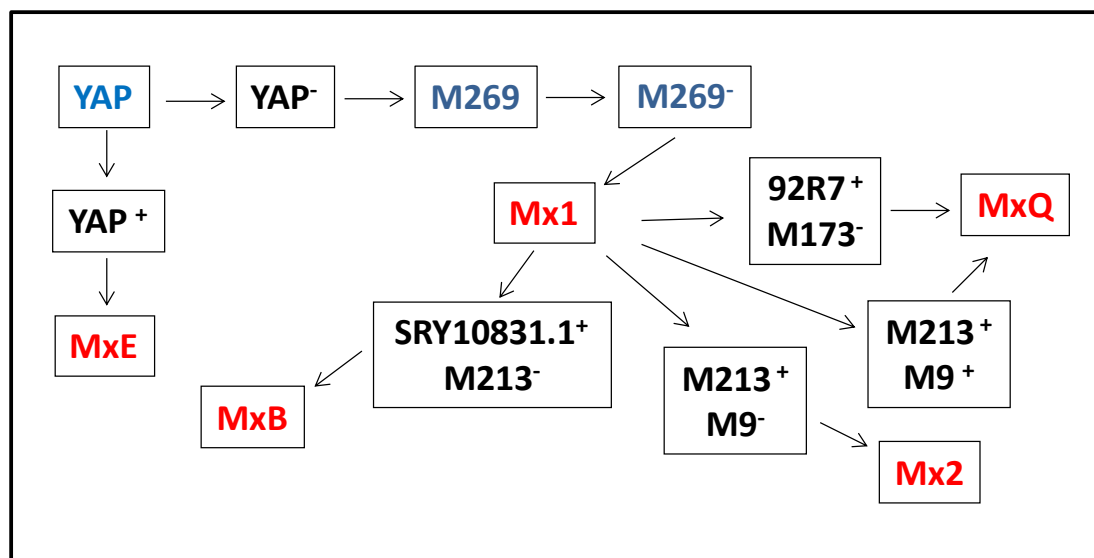
De acordo com a estratégia descrita, seguimos as etapas:

4.2.2.1 YAP - Análises da Inserção polimórfica do tipo *Alu*

Inicialmente, a inserção polimórfica do tipo *Alu* (YAP) foi o primeiro marcador analisado em todas as amostras, em PCR monoplex, ou do inglês *singleplex*, segundo o protocolo de Hammer & Horai (1995).

Como já descrito anteriormente, podemos detectar ou não a inserção polimórfica do tipo *Alu* (YAP) em uma região específica do cromossomo Y. Nas condições que serão descritas abaixo, o alelo selvagem, observado em gel de agarose, apresentou-se com um tamanho observado de 150 pb, enquanto o alelo mutado apresentou um tamanho de 450 pb (equivalente a uma inserção de 300 pb). Isto nos permitiu diferenciar os dois polimorfismos através do tamanho do seu produto de amplificação, caracterizando as amostras como não mutada (YAP⁻) e mutada (YAP⁺). Esta mutação (YAP⁺) mostra-se muito frequente em populações africanas de origem Bantu (com frequência acima de 80%).

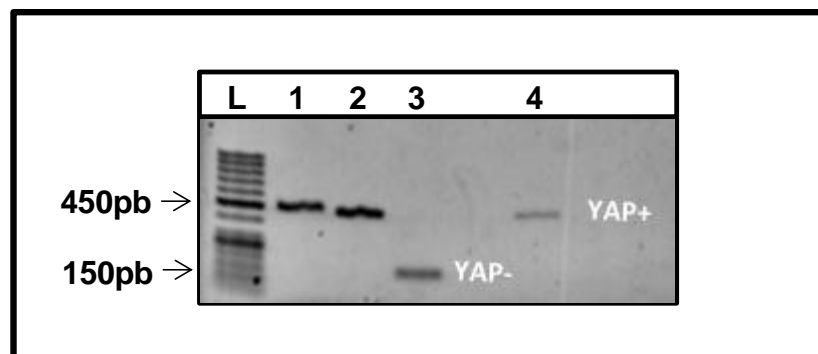
Figura 15 - Representação esquemática dos multiplexes utilizados e a estratégia seguida para as genotipagens das amostras do Rio de Janeiro e oeste da África



Nota: Em Azul = Monoplex YAP e M269; em Vermelho = Multiplexes usados após análise preliminar com YAP e M269.

Através de PCR monoplex, as amostras foram amplificadas e a análise revelou diferenças de tamanho (em pares de bases) nos produtos de PCR referentes aos dois alelos (YAP^+ , 455bp e YAP^- , 150bp), o que permitiu a sua discriminação através da eletroforese dos produtos amplificados em gel de agarose a 2%. O gel foi imerso por 10 minutos em solução de brometo de etídeo na concentração de 10 mg/mL (OLIVEIRA *et al.*, 2009), com a visualização das bandas amplificadas através da fluorescência emitida, observadas após a análise em um *scanner* de fluorescência (FMBIO II- Multi View - HITACHI) (Figura 16). A partir desta identificação, direcionamos as análises para os multiplexes com os SNPs que seriam os mais esperados. Deste modo, as amostras consideradas YAP^+ , foram direcionadas aos multiplexes de haplogrupos africanos, enquanto as amostras consideradas como YAP^- foram direcionadas para a análise do SNP M269.

Figura 16 - Visualização da amplificação da inserção do tipo *Alu* (YAP) em gel de Agarose a 2%, corado com brometo de etídeo e visualizado em *scanner* de fluorescência (FMBIO- Multi View- Hitachi)



Legenda: L= Ladder; 1, 2 e 4 = Amostras amplificadas (YAP^+ = 450 pb); 3 = Amostra amplificada (YAP^- = 150 pb).

4.2.2.2 M269 - Análise do Y-SNP por RFLP

Baseado nos resultados observados para o YAP, as amostras sem a inserção foram consideradas negativas (YAP^-), sendo posteriormente analisadas por RFLP para o marcador M269.

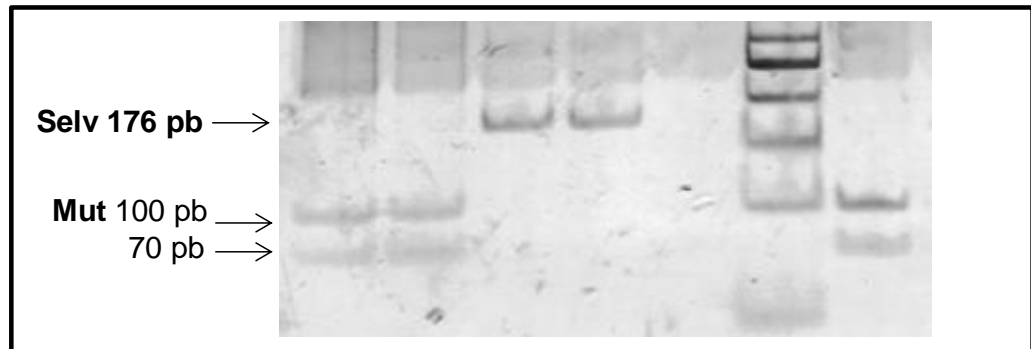
Esta análise foi feita por PCR monoplex, com a posterior utilização de enzima de restrição *Mval* (ou seu isoesquisômero *BstM*) (Thermo Scientific). Com os oligonucleotídeos para a amplificação do M269 utilizado para esta PCR, observa-se um produto de amplificação do alelo ancestral, selvagem ou T(0), com um tamanho de 184 pb. Por outro lado, a mutação no sítio de restrição (**T para C**), detecta o outro alelo (mutado), que é denominado de C(1). Em resumo, quando visualizados em gel de acrilamida a 6%, podem ser observados um fragmento de 184 pb referente ao alelo selvagem ou dois fragmentos, com aproximadamente 70 pb e 100 pb, nas amostras com a mutação.

As reações para amplificação por PCR da região alvo (M269) foram realizadas com o kit de amplificação da Qiagen® Multiplex kit. Para um volume final de 10µL, foram utilizados: 5 µL de Master mix [2x], 1 µL de oligonucleotídeo senso 5' (TCATGCCTAGCCTCATTCT)3' e reverso 5'(GCCCGGCCACTATACTTCTT)3' [10x] (0,2 µM final), 2,7 µL H₂O e 0,5 µL DNA (2 ng/µL). As condições de termociclagem foram de 95°C por 10 min, ciclos de 94°C por 30 segundos, 59°C por 30 segundos, 70°C por 30 segundos, com uma extensão final a 65°C por 15 minutos.

Os produtos de PCR foram analisados em gel de poliacrilamida a 6%, (SAMBROOK & MANIATIS, 1989) e a visualização dos fragmentos amplificados foi feito através do método de coloração com nitrato de prata, conforme descrito por BASSAM *et al.* (1991). Após a confirmação da amplificação do produto (184 pb), os mesmos foram digeridos com a enzima de restrição *Mval* a 37°C *overnight*, na proporção de 0,5 unidades de enzima (10 U/µL) adicionada a cada 2 µL de produto amplificado. A visualização desta digestão também foi feita através de gel de acrilamida a 6% (BASSAM *et al.*, 1991) (Figura 17).

No decorrer do trabalho laboratorial, a utilização do *software* “*Haplogroup Predictor*” (<http://www.hprg.com/hapest5/index.html>) disponível gratuitamente na Internet, se mostrou uma nova e eficaz ferramenta de pré-análise e foi incorporada ao estudo, mostrando-se muito eficaz para diminuir o gasto de tempo e de reagentes com análises moleculares preliminares (YAP e M269).

Figura 17 - Gel de poliacrilamida a 6% corado com nitrato de prata, mostrando a análise do SNP M269, após digestão com enzima de restrição *MVAI* (37°C *overnight*)



Legenda: **Selv**= Selvagem com um fragmento com aproximadamente 176 pb, **Mut**= Mutado com a observação de dois fragmentos de aproximadamente 100 e 70 pb.

Este programa computacional permitiu prever o haplogrupo ao qual cada amostra poderia pertencer, segundo a análise prévia dos 17 Y-STRs. Todos os possíveis haplogrupos indicados pelo programa foram então confirmados por PCR e suas reações de SNaPshot, usando 5 diferentes multiplexes (item 4.2.2.3). Contudo, esta ferramenta se mostrou ineficiente na previsão de alguns haplogrupos, como por exemplo, os haplogrupos B e J, e estas amostras seguiram todo o protocolo conforme a estratégia inicial proposta.

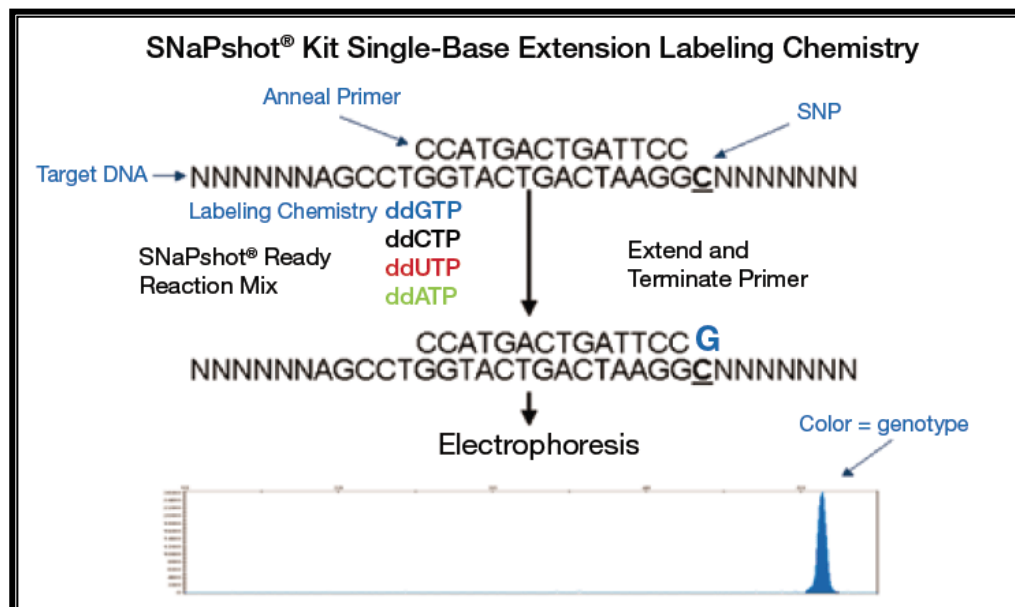
4.2.2.3 Genotipagem por sequenciamento de base única (SBE) - Multiplexes de Y-SNPs

Além dos marcadores YAP e M269, os demais Y-SNPs foram genotipados por SBE (do inglês *Single Base Extension*).

A escolha dos multiplexes a serem utilizados para a genotipagem dos Y-SNPs foi feita após as análises preliminares efetuadas com os marcadores YAP e M269, assim como pela previsão do haplogrupo através do programa computacional Haplogroup Predictor. As amostras foram então amplificadas através da utilização de cinco reações multiplex, tal como foi descrito no Anexo D.

As amostras foram então analisadas através da técnica de minissequenciamento utilizando o kit SNaPshot (Thermo Fisher Scientific Inc). A técnica consiste em uma primeira amplificação por PCR da região alvo do DNA genômico. A segunda etapa consiste de um minissequenciamento (com o mesmo princípio do sequenciamento de Sanger), a partir do anelamento do oligonucleotídeo SBE, seguido pela adição de uma única base. Por princípio, o oligonucleotídeo SBE tem como alvo a sequência imediatamente anterior ao sítio do SNP. A extensão de uma única base ocorre pela adição de um nucleotídeo complementar de terminação de cadeia (dideoxinucleotídeo – ddNTP). Os ddNTPs no kit SNaPshot são marcados com os seguintes fluoróforos: ddATP, dR6G (verde); ddGTP, dR110 (azul); ddCTP, dTAMRA (amarelo); ddTTP, dROX (vermelho), o que possibilita a detecção de qualquer alelo em uma única reação (Figura 18).

Figura 18 - Minissequenciamento com reação de SNaPshot, baseada em extensão de base única (SBE) com dideoxinucleotídeos marcados com corantes fluorescentes, que se anelam imediatamente após o final do oligonucleotídeo



Fonte: Thermo Fisher Scientific Inc.

Os cinco multiplexes de PCR-SBE foram executados como previamente descrito por Gomes *et al.* (2010) e segundo Vullo *et al.* (2014) para o multiplex Q. Todos os oligonucleotídeos dos PCRs e SBEs utilizados neste estudo foram descritos nos Anexos E e F.

Todas as reações multiplex utilizadas para a amplificação das regiões alvo foram realizadas com o kit Qiagen® Multiplex (Qiagen). Para um volume final de 10µL de reação de PCR, foram utilizados: 5 µL de Master mix [2x], 1 µL de oligonucleotídeo mix [10x] (2 µM de cada oligonucleotídeo), 2,7 µL de H₂O e 0,5 µL de DNA (2 ng/µL).

As condições de termociclagem para a PCR foram de 95°C por 15 min, seguindo-se de 35 ciclos de 94°C, por 30 s, 60°C, por 1 min 30 s, 72°C, por 1 min e uma extensão final a 72°C, por 10 min. Para a confirmação da amplificação, os produtos de PCR foram submetidos à eletroforese em gel de poliacrilamida com concentração de 6%, sendo que, para a detecção dos fragmentos amplificados foi utilizado o método de coloração com nitrato de prata, conforme descrito por BASSAM *et al.* (1991).

O preparo dos produtos de PCR para as reações de minissequenciamento requer a purificação para remoção de oligonucleotídeos e dNTPs não incorporados com a utilização de *Exonuclease I/Shrimp Alkaline Phosphatase* (ExoSAP-IT- USB® products) nas seguintes condições: para cada 1 µl de produto de PCR devem ser adicionados 0,5 µl de ExoSAP-IT, sob temperatura de 37°C durante 15 minutos, sendo posteriormente inativada em temperatura de 85°C por 15 minutos.

A partir disto as reações de minissequenciamento foram realizadas pela adição de 2,5 µl de SNaPshot Ready Reaction Mix (Thermo Fisher Scientific Inc) e 1,0 µl da mistura de oligonucleotídeos em concentrações de 2,0 µM ao volume de 1,5 µl de produto de PCR purificado. As condições de termociclagem foram as seguintes: 25 ciclos de 96°C, por 10s, 50°C, por 5s, e 60°C, por 30s. Para a remoção de ddNTPs não incorporados, foi adicionada 1 unidade da enzima *Shrimp Alkaline Phosphatase* (SAP) (GE Healthcare, USA) e, a seguir, a preparação foi incubada a 37°C, por 1h, e, em seguida, a 85°C, por 15min, para inativação da enzima.

O preparo da amostra para a eletroforese foi feito pela adição de 1 µl do produto de minissequenciamento purificado a uma solução contendo 8,9 µl de

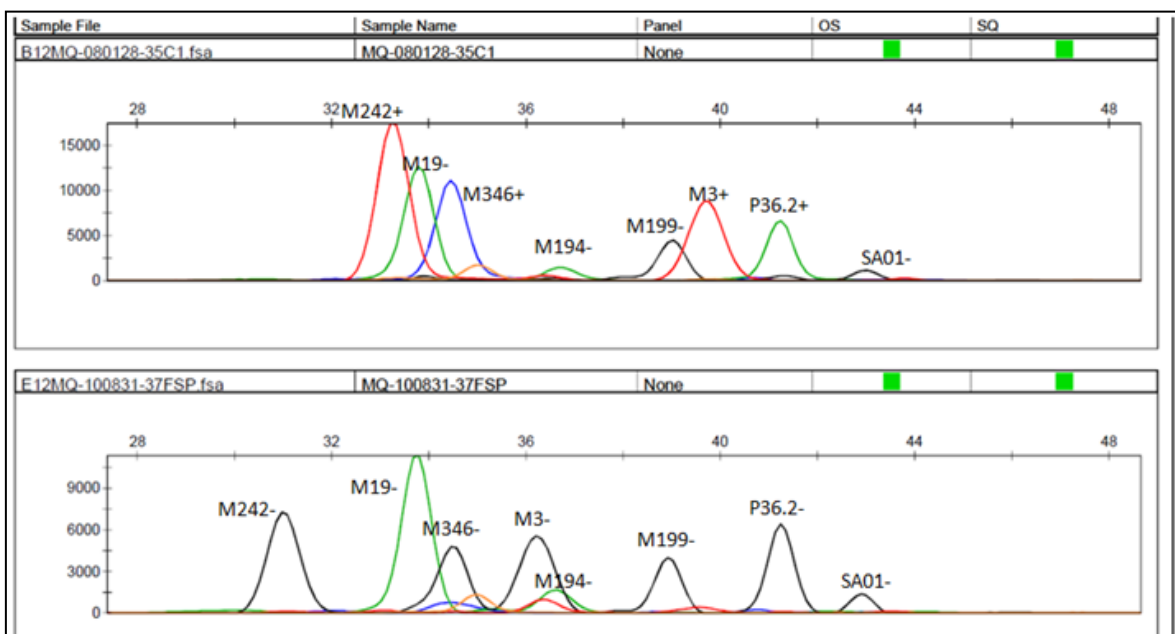
Formamida HI-DI (Thermo Fisher Scientific Inc) e 0,1 µl de padrão interno de tamanho Genescan 120 LIZ (Thermo Fisher Scientific Inc).

A eletroforese e a detecção dos produtos de minissequenciamento purificados foram realizadas em sequenciadores automáticos ABI3100, ABI3130 e ABI3500 (Thermo Fisher Scientific Inc), sendo utilizado o capilar de 50 cm e polímero POP-7 (Thermo Fisher Scientific Inc). O tempo de injeção foi de 8s e o tempo de eletroforese foi de 9 min, mantendo-se a corrente elétrica em 15 kv.

A análise dos eletroferogramas foi realizada com o *software* GeneMapper v. 4.1 (Thermo Fisher Scientific Inc). De forma ilustrativa, na Figura 19 apresenta-se um exemplo de imagem da eletroforese capilar que pode ser observada para o Multiplex Q.

A nomenclatura dos haplogrupos foi baseada em Dulik *et al.* (2012), para o haplogrupo Q, Trombetta *et al.* (2011) para o haplogrupo E, Myres *et al.* (2011) para o haplogrupo R e Karafet *et al.* (2008) para os demais haplogrupos.

Figura 19 - Exemplo de genotipagem de Y-SNP que pode ser visualizada após eletroforese capilar, nas condições descritas em material e métodos, utilizando o multiplex Q



Legenda: Os marcadores podem ser exemplificados como: (+) = Mutado, apresentando a base derivada ou (-) = Não mutado ou selvagem, apresentando a base ancestral.

4.3 Análises estatísticas dos resultados

As frequências haplotípicas e de diversidade foram estimadas de acordo com Nei (1987) usando o *Software* ARLEQUIN Ver 3.01 (EXCOFFIER, 2005) que também foi usado para computar as distâncias genéticas.

Os cálculos de AMOVA (Análise de Variância Molecular), frequências haplotípicas, diversidade e distâncias genéticas entre pares de populações (F_{ST} e R_{ST}) foram realizados através do programa Arlequin Ver 3.11 (EXCOFFIER *et al.*, 2005). As frequências dos haplogrupos foram estimadas por contagem direta utilizando-se para isto o programa Microsoft Excel (Microsoft Corporation, USA).

Para a análise do escalonamento multidimensional (MDS) das distâncias genéticas (F_{ST} e R_{ST}) entre as populações foi utilizado o programa SPSS Statistics v.17.0 (SPSS Incorporation, Chicago).

O nível de significância foi testado para 10.000 permutações e foi utilizado o valor de 0,05 para testes simples, tendo sido aplicada a correção de Bonferroni (1936) em casos de testes múltiplos.

5 RESULTADOS E DISCUSSÃO

Uma parte dos resultados apresentados neste trabalho já se encontra publicado, sendo que os demais resultados se encontram em preparação para a elaboração de artigo científico.

Optou-se então, por apresentar os resultados em três seções, que foram pontuadas de acordo com os objetivos propostos:

- a) Caracterização das linhagens masculinas na população do Rio de Janeiro;
- b) Caracterização das linhagens masculinas das amostras de referência do oeste africano;
- c) Estudo do aporte africano ao Rio de Janeiro com base na comparação com populações africanas de referência, que historicamente, foram associadas à diáspora africana em direção à América.

5.1 Caracterização da população do Rio de Janeiro

Ao longo deste trabalho, foram caracterizados, quanto aos polimorfismos do tipo STRs e SNPs localizados na região não recombinante do cromossomo Y, o patrimônio genético de indivíduos da população do estado do Rio de Janeiro, assim como foram efetuadas análises comparativas com dados relativos a outras populações do Brasil. Demonstrou-se que o uso destes dois tipos de marcadores (STRs e SNPs), de forma combinada pode ser uma boa estratégia para predizer subestrutura populacional e que deve ser sempre considerada quando se delineiam bases de dados forenses para haplótipos de cromossomo Y. A comparação dos resultados obtidos para a população do Rio de Janeiro com diferentes amostras populacionais permitiram, ainda, demonstrar a elevada variabilidade associada aos polimorfismos em estudos genéticos e a sua utilidade em estudos de Genética Populacional e Forense.

Os resultados do estudo realizado na população de 605 indivíduos do estado do Rio de Janeiro, bem como a sua análise exaustiva, foram publicados no artigo científico, que é apresentado em seguida, juntamente com o seus dados suplementares.

Artigo I:

Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs.

Forensic Science International: Genetics 13 (2014), 264–268.

RESUMO:

O Brasil é um grande território dividido em 5 regiões geográficas e com uma população altamente diversa originária de diferentes graus e modos de miscigenação entre os Nativos Americanos, Europeus e Africanos. Neste estudo, uma amostra de 605 indivíduos do sexo masculino, não aparentados, foi genotipada para 17 Y-STRs e 46 Y-SNPs visando uma profunda caracterização do pool genético da população do Rio de Janeiro e uma comparação com outras populações do Brasil.

Foi observado um alto valor de diversidade haplotípica (0.9999 ± 0.0001) com uma diversidade de haplogrupos de Y-SNPs de 0.7589 ± 0.0171 . As comparações entre populações tanto no nível de haplótipos quanto de haplogrupos mostraram diferenças significativas entre as populações do Sudeste e Norte que podem ser explicadas por diferenças na proporção de cromossomos Y de origem Africana e Nativo Americano. Diferenças estatisticamente significativas entre as amostras de populações miscigenadas urbanas de cinco regiões do Brasil não foram previamente detectadas em nível de haplótipos quando baseadas em amostras de pequeno número amostral do Sudeste, o que enfatiza a importância do tamanho amostral para detectar estratificação populacional para uma interpretação exata de perfis de similaridade ou concordância em casos forenses e de parentesco. Embora não apresentando um poder de discriminação intrapopulacional tão alto quanto o dos Y-STRs, os Y-SNPs são bastante poderosos para descortinar diferenças em populações miscigenadas.

Neste estudo, as análises combinadas destes dois tipos de marcadores, provaram ser uma boa estratégia para predizer a subestruturação populacional, que

deve ser levado em conta quando se delinea bancos de dados forenses para haplótipos de cromossomo Y.



ELSEVIER

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs



Andréa M. Oliveira^a, Patricia M. Domingues^a, Verónica Gomes^b, António Amorim^{b,c}, Juliana Jannuzzi^a, Elizeu F. de Carvalho^a, Leonor Gusmão^{a,b,*}

^a DNA Diagnostic Laboratory (LDD), Institute of Biology, State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

^b IPATIMUP – Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal

^c FCUP – Faculty of Sciences of the University of Porto, Porto, Portugal

ARTICLE INFO

Article history:

Received 22 July 2014

Received in revised form 28 August 2014

Accepted 29 August 2014

Keywords:

Y-STRs

Y-SNPs

Brazil

Population stratification

ABSTRACT

Brazil has a large territory divided in five geographical regions harboring highly diverse populations that resulted from different degrees and modes of admixture between Native Americans, Europeans and Africans. In this study, a sample of 605 unrelated males was genotyped for 17 Y-STRs and 46 Y-SNPs aiming a deep characterization of the male gene pool of Rio de Janeiro and its comparison with other Brazilian populations. High values of Y-STR haplotype diversity (0.9999 ± 0.0001) and Y-SNP haplogroup diversity (0.7589 ± 0.0171) were observed. Population comparisons at both haplotype and haplogroup levels showed significant differences between Brazilian South Eastern and Northern populations that can be explained by differences in the proportion of African and Native American Y chromosomes. Statistical significant differences between admixed urban samples from the five regions of Brazil were not previously detected at haplotype level based on smaller size samples from South East, which emphasizes the importance of sample size to detected population stratification for an accurate interpretation of profile matches in kinship and forensic casework. Although not having an intra-population discrimination power as high as the Y-STRs, the Y-SNPs are more powerful to disclose differences in admixed populations. In this study, the combined analysis of these two types of markers proved to be a good strategy to predict population sub-structure, which should be taken into account when delineating forensic database strategies for Y chromosome haplotypes.

© 2014 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Brazil is a genetically highly diverse country as a result of extensive admixture between Native American, European and African populations in a vast and very heterogeneous territory [1,2].

Different patterns of genetic ancestry were revealed among the Brazilian populations from different regions, when assessed through mtDNA and autosomal markers (e.g. [3–9]). Nevertheless, no important population stratification have been disclosed until now for the Y chromosome gene pool, except for few African and

Native American descendent groups that are spread all over the country (e.g. [10–12]). A recent study from Palha et al. [13] showed no significant differentiation for 17 populations, distributed among the 5 geopolitical regions of Brazil, which was attributed to the high frequency of European male lineages in all populations.

Owing to the non-recombining nature of the Y chromosome, in forensic genetics it is important to study a large number of markers at the same time to increase levels of discrimination within populations. Concerning Brazil, data exist for a total of 17 states including 23 Y-STRs [13,14]. For the 17 Y-STR loci in the Yfiler kit, there are data available for a total of 18 out of the 26 Brazilian states [13–18]; and haplotypic data for the remaining states have been published for smaller Y-STR sets [19].

Although less powerful to discriminate male lineages inside populations due to their low mutation rates and recurrence, Y-SNPs are useful to define haplogroups with high geographical specificity and are therefore more apt to detect population stratification in admixed populations from diverse continental origins.

* Corresponding author at: DNA Diagnostic Laboratory (LDD), Pavilhão Haroldo Lisboa da Cunha, UERJ – Universidade do Estado do Rio de Janeiro, São Francisco Xavier, 524, Maracanã, CEP: 20550-900 Rio de Janeiro, RJ, Brazil.

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

E-mail address: lgusmao@ipatimup.pt (L. Gusmão).

<http://dx.doi.org/10.1016/j.fsigen.2014.08.017>

1872–4973/© 2014 Elsevier Ireland Ltd. All rights reserved.

Studies using Y-SNPs to characterize admixed populations in Brazil are few, and covering a small number of populations (e.g. Rio de Janeiro [20], Alagoas [21], Bahia [22], Manaus and Ribeirão Preto [23]).

In this work, a large sample of 605 non-related males from Rio de Janeiro population was genotyped for an extended group of both Y-STRs and Y-SNPs aiming a better characterization of Brazilian population stratification in what concerns male lineages contributions, allowing a more accurate interpretation of profile matches in kinship and forensic casework, as well as a deeper insight on the microevolution and demographic history of South America male gene pool.

2. Methods and materials

2.1. Population samples

A total of 605 samples from unrelated males living in Rio de Janeiro district, Brazil, were selected for this study (Fig. S1). Samples were obtained from paternity casework from the DNA Diagnostic Laboratory (LDD) with written informed consent of the donors. This study, together with the informed consent, was approved by the Ethics Committee of the State University of Rio de Janeiro (CAAE: 0067.0.228.000-09).

A total of 149 samples from Rio de Janeiro from previous reports [20,24,25] were also included in the present study, after being typed for 5 new Y-STRs and the full set of Y-SNPs.

2.2. Y chromosome STR typing

The AmpFISTR Yfiler kit was used to analyze 17 Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and GATA H4), following manufacturer's recommendations. Typing was carried out using ABI 3130 and ABI 3500 Genetic Analysers (Applied Biosystems) and analysis was performed using GeneMapper[®] ID Software v4.0 (Applied Biosystems). The Y-STR alleles were designated according to the ISFG recommendations [26].

Y haplotype data are included in the Y-STR Haplotype Reference Database (<http://www.yhrd.org>) under the following accession number: Rio de Janeiro, Brazil [Admixed] YA003720.

2.3. Y-chromosome SNP typing

In this study, 46 Y-SNPs were analyzed (Fig. S2). Haplogroup nomenclature follows Dulik et al. [27] for haplogroup Q, Trombetta et al. [28] for haplogroup E, Myres et al. [29] for haplogroup R, and Karafet et al. [30] for the remaining haplogroups. The Y-SNPs were chosen to discriminate the most frequent Native American, European and African haplogroups.

The 46 Y-SNPs were genotyped by PCR and SNaPshot reactions using the different multiplexes indicated in Supplementary Fig. S2. Multiplexes were carried out as previously described by Gomes et al. [31], except for multiplex Q that was performed as described by Vullo et al. [32]. The first multiplex to be typed in each sample was chosen after inferring the most likely haplogroup employing the freely available haplogroup predictor software (<http://www.hprg.com/hapest5/index.html>). In cases where no derived alleles were found using the first multiplex, the remaining multiplexes were also investigated and haplogroup classification was based on the results obtained. Besides not discriminating some sub-lineages inside haplogroups, the prediction program misclassified approximately 1% of the samples belonging to European haplogroups, and all samples belonging to the African haplogroups B2a*-M150 and E2b1-M85.

2.4. Data analysis

Haplogroup frequencies were determined by direct counting. Population comparisons, haplotype frequencies and genetic diversities according to Nei [33] were estimated with the software Arlequin 3.5.1.2 [34]. In population comparisons, the number of repetitions in DYS389I was subtracted from DYS389II and DYS385 was excluded. To match Arlequin STR data format: (a) duplicated alleles were represented as missing data; (b) null and non-consensus alleles were also represented as missing data and an additional locus was inserted with non-null/consensus alleles coded as 1 and null/intermediate alleles as 10 repeats.

Population pairwise genetic distances were conducted based on F_{ST} and R_{ST} and the significance was tested with 10,000 permutations. Pairwise genetic distances were visualized in two dimensional space using the multi-dimensional scaling (MDS) analysis included in the StatSoft, Inc. (2007) program, STATISTICA (data analysis software system), ver.8.0 (www.statsoft.com).

3. Results and discussion

The genotyping results of a population sample of 605 unrelated individuals from Rio de Janeiro for 17 Y-STR and 46 Y-SNP markers are listed in Table S1.

3.1. Y-chromosomal diversity

A high Y-STR based haplotype diversity was observed (0.9999 ± 0.0001); with just 12 haplotypes being shared by two individuals and just one observed three times. The haplotype diversity value observed in Rio de Janeiro was similar to those previously obtained for other samples from Brazil that have been studied for the same set of Y chromosome STR markers [13–18].

The observed haplogroup diversity was 0.7589 (± 0.0171) that can be considered high when compared to the values estimated (for the same group of Y-SNPs) in Portuguese (0.652) [35], in Sub-Saharan African Bantu groups (0.681 and 0.753 in Angola [36] and Equatorial Guinea [37], respectively) and in Native Americans (0.265 and 0.709 for Toba and Colla populations, respectively) [38], which is explained by the presence of haplogroups from very diverse continental origins.

The analysis of the Y-SNPs did not allow discriminating lineages with identical Y-STR haplotypes, i.e. all chromosomes sharing the same haplotype also belong to the same haplogroup.

3.2. Comparison of the Y chromosome haplotypes in Brazil

Haplotype frequencies in Rio de Janeiro were compared with those from samples from other 17 Brazilian states, which have been studied for the same markers (see states location and Y-STR haplotype data references in Fig. S1). Population comparisons were performed through F_{ST} analysis based on number of different alleles (FST-like) and sum of squared differences (RST-like), after excluding the DYS385 from the AmpFISTR Yfiler markers' set (see results in Table S2 and Fig. S3). The results revealed no significant differences between Rio de Janeiro and two other populations from South East Brazil (São Paulo and Minas Gerais), for a significance level of 0.00031 obtained after applying the Bonferroni correction for 161 pairwise comparisons. When comparing Rio de Janeiro with samples from other regions, significant differences were only found with two populations from the North region (Pará and Roraima), when accessed by both F_{ST} and R_{ST} values.

Samples were pooled to represent the 5 geopolitical Brazilian regions (see the delimitations of the geopolitical regions in Fig. S1) and compared with each other, as well as with reference samples from Portuguese (pooled data from [13,14,39]), Africans [13,37,40]

Table 1

Matrix showing the pairwise F_{ST} values (below diagonal) between Brazilian regions and reference samples from Africa (AFR), Europe (EUR) and Native America (NAM), obtained for the haplotypes defined by the 16-markers set. The corresponding differentiation p values are indicated in the above diagonal.

	N	NE	CW	SE	S	AFR	EUR	NAM
FST-like								
N	*	0.00020	0.00525	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06
NE	0.00568	*	0.07841	0.05267	0.04366	<5E-06	0.00436	<5E-06
CW	0.00246	0.00152	*	0.03366	0.30334	<5E-06	0.39452	<5E-06
SE	0.00383	0.00131	0.00133	*	0.01525	<5E-06	0.00178	<5E-06
S	0.00418	0.00148	0.00018	0.00091	*	<5E-06	0.06772	<5E-06
AFR	0.14984	0.13059	0.15733	0.12772	0.14404	*	<5E-06	<5E-06
EUR	0.00551	0.00505	0.00005	0.00346	0.00112	0.17371	*	<5E-06
NAM	0.14155	0.14924	0.15930	0.13724	0.14544	0.20796	0.17148	*
RST-like								
N	*	<5E-06	0.00218	<5E-06	<5E-06	<5E-06	0.05039	<5E-06
NE	0.01917	*	0.00208	0.03485	0.00337	<5E-06	<5E-06	<5E-06
CW	0.00482	0.00804	*	0.01713	0.06950	<5E-06	0.06732	<5E-06
SE	0.00851	0.00249	0.00271	*	0.01416	<5E-06	0.00119	<5E-06
S	0.00763	0.00498	0.00160	0.00146	*	<5E-06	0.00693	<5E-06
AFR	0.27659	0.20360	0.26386	0.21337	0.24235	*	<5E-06	<5E-06
EUR	0.00205	0.01310	0.00238	0.00582	0.00399	0.29182	*	<5E-06
NAM	0.17778	0.22505	0.18906	0.16791	0.18235	0.48162	0.21607	*

Note: Significant value after applying Bonferroni's correction ($p < 0.00179$) are in bold.

and Brazilian Native Americans [13,41]. South East, the region to each Rio de Janeiro belongs, presented significant differences with the North, which is indeed the most differentiated region of Brazil (Table 1). When compared with the three reference samples, South East Brazil is more close to the Europeans than to Africans or Native Americans, although significant differences could be detected with European samples from Portugal, based on F_{ST} genetic distance (Table 1). Nonetheless, it is worth noting that South East region presents the smallest distance observed with Africans (Table 1).

3.3. Y chromosome haplogroups in Rio de Janeiro

A total of 23 different haplogroups were detected in the sample from Rio de Janeiro population, which ancestry can be traced to Europe, Africa or Native America (Table 2). As expected from the STR results, as well as from previous publications concerning the paternal ancestry of Brazilian admixed populations (e.g. [1,20,21,23,42,43]), the Y chromosomes in Rio de Janeiro belong

mainly to typically European haplogroups, followed by Africans, and just a small proportion of about 2% belong to Native American haplogroups. However, the frequency of African haplogroups increased from 8 to 15% when compared with the results of a previous study on a smaller sample from the same population [20].

Concerning Y chromosome SNP studies in Brazil, most have been focused on small isolated Afro-descendent or Native groups, or self-declared ethnic groups inside admixed populations (e.g. [3,6,44]). Moreover, few of them report the three main continental ancestral compositions of Brazilians in large admixed populations, which are the most relevant information in the forensic perspective. In Table 3, we compare the results from different studies that inferred ancestry proportions of Y chromosomes in Brazilian population groups. The results show a higher proportion of African haplogroups in Rio de Janeiro than in other populations; a value more than three times higher than the previous estimate for a general sample of 50 individuals from the same geopolitical region. The Native American component is consistently low in all studies, except for Manaus, in the North region, where it reaches a frequency of almost 10%.

Table 2

Absolute and relative frequencies of 23 Y chromosome haplogroups found in 605 unrelated male samples from Rio de Janeiro population; and inferred recent African, European and American continental origin of each haplogroup.

Haplogroup	Inferred origin	N	Freq. (%)
B2a*-M150	Africa	4	0.66
DE*-YAP	Africa	3	0.50
E1b1a1*-M2	Africa	47	7.77
E1b1a1f-M191	Africa	26	4.30
E1b1a1g1c-M154	Africa	2	0.33
E1b1b1*-M35	Africa	3	0.50
E1b1b1a1-M78	Africa/Europe	29	4.79
E1b1b1b1-M81	Europe	32	5.29
E1b1b1c-M123	Europe	6	0.99
E2b1-M85	Africa	5	0.83
F*-M213	Europe	1	0.17
G-M201	Europe	22	3.64
I*-M170	Europe	41	6.78
I2a2-M26	Europe	10	1.65
J*-I2f2a	Europe	10	1.65
J2-M172	Europe	39	6.45
K*-M9	Europe	4	0.66
L-M22	Europe	4	0.66
Q1a3a1a*-M3	America	14	2.31
R1a-SRY10831.2	Europe	8	1.32
R1b1b-M269	Europe	282	46.61
T-M70	Europe	12	1.98
Y*(xA,B,DE*,F*)-M168	Unknown	1	0.17

4. Conclusions

In forensic genetics, it is important to accurately estimate the frequency of relevant profiles in reference populations, taking into account population boundaries and possible effects of sub-structure.

The size and demographic complexity of Brazil requires large studies including samples from different representative regions to disclose the full genetic diversity and its stratification across the country.

Concerning Y chromosome haplotype frequencies in Rio de Janeiro, previous publications only included a restricted number of

Table 3

Comparison of the Y chromosome ancestry proportions inferred based on Y-SNP haplogroups found in different population groups from Brazil.

Population	n	NAM (%)	EUR (%)	AFR (%)
Rio de Janeiro [this study]	605	2.32	81.95	15.73
Alagoas [21]	247	0.81	94.74	4.45
Ribeirão Preto [23]	65	1.54	95.38	3.08
Manaus [23]	42	9.52	88.10	2.38
North region [43]	49	0	98.00	2.00
North East [43]	49	0	95.90	4.10
South [43]	52	0	100.00	0.00
South East [43]	50	0	96.00	4.00

markers and/or samples. In the present work, the study of a large sample of 605 unrelated males living in Rio de Janeiro, for a high number of Y-STRs and Y-SNPs, showed no significant differences between Rio de Janeiro, São Paulo and Minas Gerais, all from South East region. Nevertheless, significant differences have been detected when comparing South East and North region populations, which were not previously noticed when using smaller size samples. Because of the very high levels of intra-population diversity that are associated with haplotypes defined by a high number of STRs, large samples are necessary to statistically support small differences between populations, or to detect population stratification.

The study of Y-SNPs in Rio de Janeiro revealed a high percentage of European haplogroups (81.95%), together with a noticeable proportion of African haplogroups (15.73%), which is usually lower in most urban admixed populations.

The main European contribution to the Y chromosome gene pool of most Brazilian populations explains the difficulty to detect significant population differences for Y-haplotypes in most studies, based on small size samples for extremely diverse haplotypes associated to large numbers of Y-STR markers. However, differences in Native American, European and African derived pools can be easily detected using SNPs, foreseeing differences in haplotype frequencies in large population databases.

Acknowledgements

This work was supported by FAPERJ (Fundação de Amparo à Pesquisa do Rio de Janeiro), Programa APQ1 2013/01, processo E-26/111.395/2013. IPATIMUP is an associate laboratory of the Portuguese Ministry of Science, Technology and Higher Education, and is partially supported by FCT, the Portuguese Foundation for Science and Technology. V.G. is supported by FCT through grant SFRH/BPD/76207/2011.

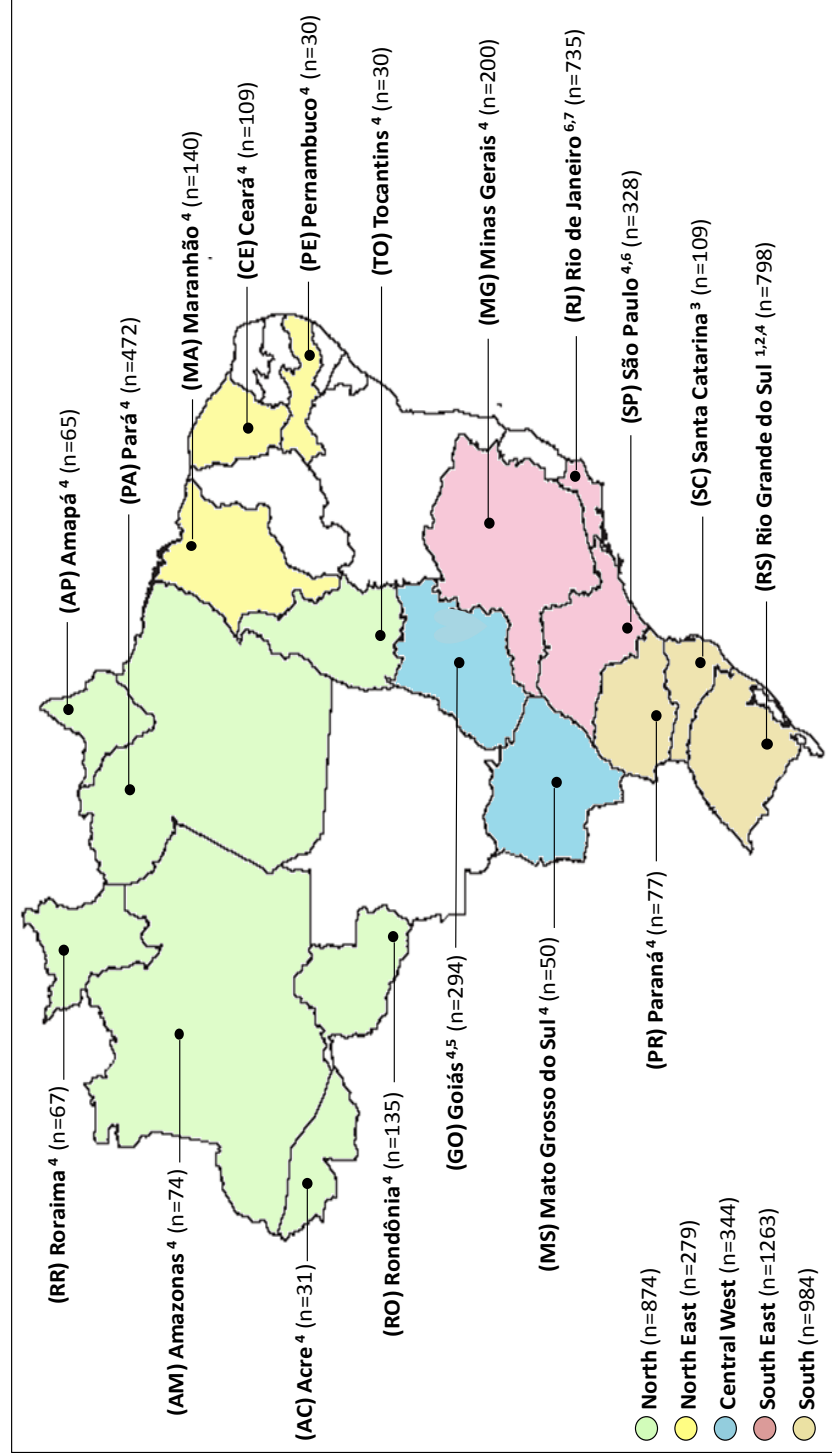
Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2014.08.017.

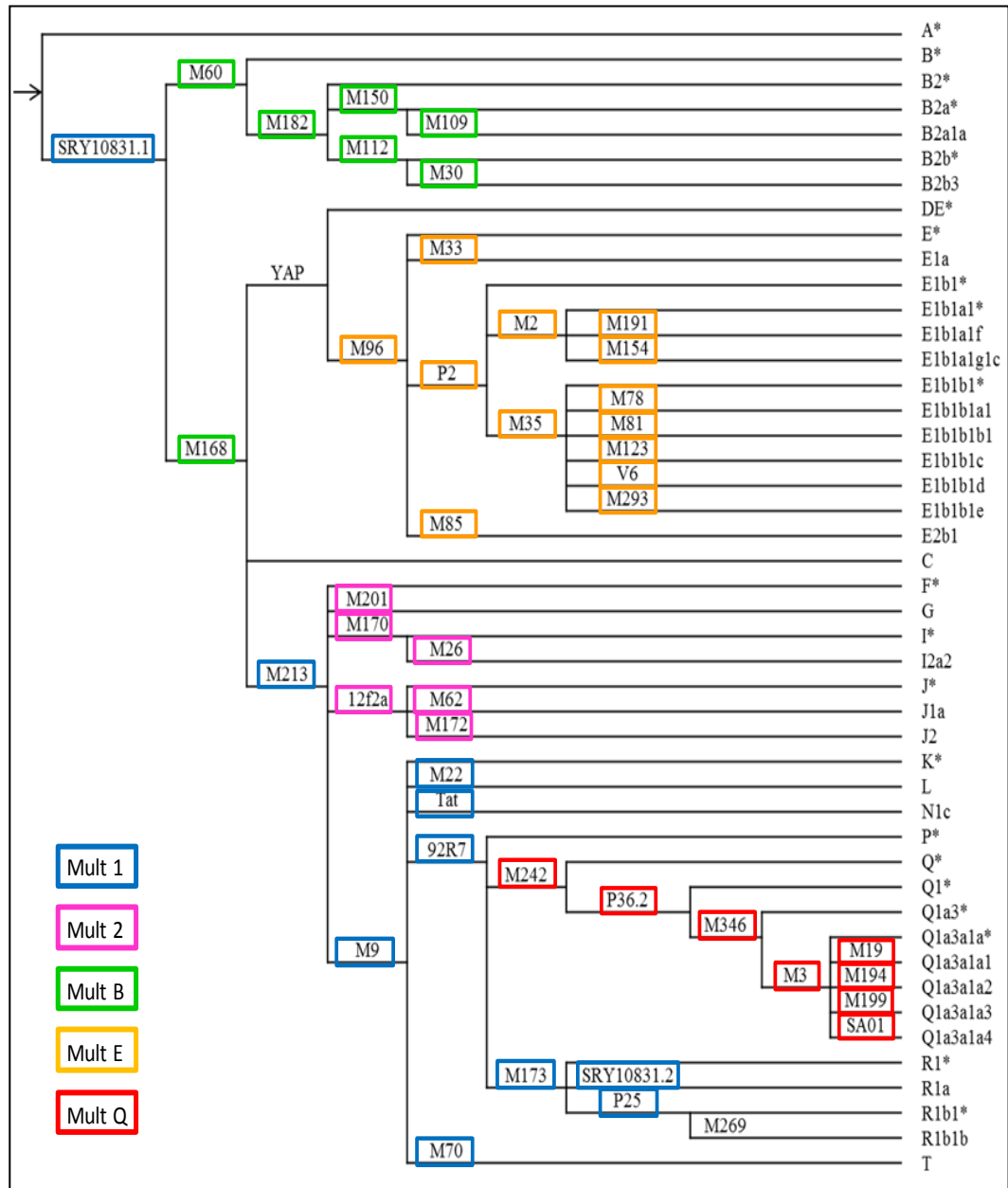
References

- [1] K. Abe-Sandes, W.A. Silva Jr., M.A. Zago, Heterogeneity of the Y chromosome in Afro-Brazilian populations, *Hum. Biol.* 76 (2004) 77–86.
- [2] M.D. Priore, R. Venâncio, Uma Breve História do Brasil, first ed., Planeta do Brasil, São Paulo, 2010.
- [3] M.C. Bortolini, W.A. Silva Jr., D.C. De Guerra, G. Remonato, R. Mirandola, African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers, *Am. J. Hum. Biol.* 11 (1999) 551–563.
- [4] J. Alves-Silva, M. da Silva Santos, P.E. Guimarães, A.C. Ferreira, H.J. Bandelt, The ancestry of Brazilian mtDNA lineages, *Am. J. Hum. Genet.* 67 (2000) 444–461.
- [5] A.K. Ribeiro-dos-Santos, J.M. Pereira, M.R. Lobato, B.M. Carvalho, J.F. Guerreiro, S.E. Batista Dos Santos, Dissimilarities in the process of formation of Curiaú, a semi-isolated Afro-Brazilian population of the Amazon region, *Am. J. Hum. Biol.* 14 (2002) 440–447.
- [6] T. Hunemeier, C. Carvalho, A.R. Marrero, F.M. Salzano, S.D. Pena, Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data, *Am. J. Phys. Anthropol.* 133 (2007) 854–867.
- [7] N.P. Santos, E.M. Ribeiro-Rodrigues, A.K. Ribeiro-Dos-Santos, R. Pereira, L. Gusmão, Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel, *Hum. Mutat.* 31 (2010) 184–190.
- [8] T.C. Lins, R.G. Vieira, B.S. Abreu, D. Grattapaglia, R.W. Pereira, Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs, *Am. J. Hum. Biol.* 22 (2010) 187–192.
- [9] F.S. Manta, R. Pereira, A. Caiafa, D.A. Silva, L. Gusmão, E.F. Carvalho, Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers, *Ann. Hum. Biol.* 40 (2013) 94–98.
- [10] F.P. Leite, S.M. Callegari-Jacques, B.A. Carvalho, T. Kommer, C.H. Matte, Y-STR analysis in Brazilian and South American populations, *Am. J. Hum. Biol.* 20 (2008) 359–363.
- [11] T.J. Palha, E.M. Ribeiro-Rodrigues, A. Ribeiro-dos-Santos, J.F. Guerreiro, L.S. Moura, Male ancestry structure and interethnic admixture in African-descent communities from the Amazon as revealed by Y-chromosome STRs, *Am. J. Phys. Anthropol.* 144 (2011) 471–478.
- [12] C.E. Wiesel, M.R. Luizon, S.M. Sousa, L.M. Santos, Y.C. Muniz, Y-linked microsatellites in Amazonian Amerindians applied to ancestry estimates in Brazilian Afro-derived populations, *Am. J. Hum. Biol.* 25 (2013) 313–317.
- [13] T.J. Palha, L. Gusmão, E. Ribeiro-Rodrigues, J.F. Guerreiro, A. Ribeiro dos Santos, Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes, *PLoS ONE* 7 (2012).
- [14] J. Purps, S. Siegert, S. Willuweit, M. Nagy, C. Alves, R. Salazar, S.M. Angustia, L.H. Santos, K. Anslinger, B. Bayer, Q. Ayub, W. Wei, Y. Xue, C. Tyler-Smith, M.B. Bafalluy, B. Martínez-Jarreta, B. Egyed, B. Balitzki, S. Tschumi, D. Ballard, D.S. Court, X. Barrantes, G. Bäßler, T. Wiest, B. Berger, H. Niederstätter, W. Parson, C. Davis, B. Budowle, H. Burri, U. Borer, C. Koller, E.F. Carvalho, P.M. Domingues, W.T. Chamoun, M.D. Coble, C.R. Hill, D. Corach, M. Caputo, M.E. D'Amato, S. Davison, R. Decorte, M.H. Larmuseau, C. Ottoni, O. Rickards, D. Lu, C. Jiang, T. Dobosz, A. Jonkisz, W.E. Frank, I. Furac, C. Gehrig, V. Castella, B. Grskovic, C. Haas, J. Wobst, G. Hadzic, K. Drobnic, K. Honda, Y. Hou, D. Zhou, Y. Li, S. Hu, S. Chen, U.D. Immel, R. Lessig, Z. Jakovski, T. Ilievski, A.E. Klann, C.C. Garcia, P. de Knijff, T. Kraaijenbrink, A. Kondili, P. Miniati, M. Vourpoulou, L. Kovacevic, D. Marjanovic, I. Lindner, I. Mansour, M. Al-Azem, A.E. Andari, M. Marino, S. Furfuro, L. Locarno, P. Martín, G.M. Luque, A. Alonso, L.S. Miranda, H. Moreira, N. Mizuno, Y. Iwashima, R.S. Neto, T.L. Nogueira, R. Silva, M. Nastainczyk-Wulf, J. Edelman, M. Kohl, S. Nie, X. Wang, B. Cheng, C. Núñez, M.M. Pancorbo, J.K. Olofsson, N. Morling, V. Onofri, A. Tagliabracci, H. Pamjav, A. Volgyi, G. Barany, R. Pawlowski, A. Maciejewska, S. Pelotti, W. Pepinski, M. Abreu-Glowacka, C. Phillips, J. Cárdenas, D. Rey-Gonzalez, A. Salas, F. Brisighelli, C. Capelli, U. Toscanini, A. Piccinini, M. Pigiionica, S.L. Baldassarra, R. Ploski, M. Konarzewska, E. Jastrzebska, C. Robino, A. Sajantila, J.U. Palo, E. Guevara, J. Salvador, M.C. Ungria, J.J. Rodriguez, U. Schmidt, N. Schlauderer, P. Saukko, P.M. Schneider, M. Sirker, K.J. Shin, Y.N. Oh, I. Skitsa, A. Ampati, T.G. Smith, L.S. Calvit, V. Stenzl, T. Capal, A. Tillmar, H. Nilsson, S. Turrina, D. De Leo, A. Verzeletti, V. Cortellini, J.H. Wetton, G.M. Gwynne, M.A. Jobling, M.R. Whittle, D.R. Sumita, P. Wolańska-Nowak, R.Y. Yong, M. Krawczak, M. Nothnagel, L. Roewer, A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, *Forensic Sci. Int. Genet.* 12 (2014) 12–23.
- [15] P. Sánchez-Diz, C. Alves, E.F. Carvalho, M. Carvalho, R. Espinheira, O. Garcia, L. Pontes, M.J. Porto, O. Santapa, Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study, *Int. J. Legal Med.* 122 (2008) 529–533.
- [16] S.P. Schwengber, T. Kommer, C.H. Matte, P.E. Raimann, B.A. Carvalho, F.P. Leite, M.A. Medeiros, L.F. Souza, C.S. Castro, F.G. Chassot, S.L. Bonatto, Population data of 17 Y-STR loci from Rio Grande do Sul state (South Brazil), *Forensic Sci. Int. Genet.* 4 (2009) e31–e33.
- [17] L.M. Cainé, M.M. de Pancorbo, F. Pinheiro, Y-chromosomal 17R haplotype diversity in males from Santa Catarina, *J. Forensic Leg. Med.* 17 (2010) 92–95.
- [18] G.B. Barra, C.F. Chianca, T.H. Santa Rita, L.F. Velasco, C.F. de Souza, Haplotype diversity of 23 Y-chromosomal STRs in a population sample from the Federal District (Brazil) – a territory that arose from nothing, *Int. J. Legal Med.* (2014), <http://dx.doi.org/10.1007/s00414-013-0960-2> (Epub ahead of print).
- [19] T.C. Vieira, M.A. Gigonza, D.M. Silva, R.G. Rodovalho, G.S. Santos, A.D. da Cruz, Y-STR haplotype diversity and population data for Central Brazil: implications for environmental forensics and paternity testing, *Genet. Mol. Res.* 13 (2014) 3404–3410.
- [20] D.A. Silva, E.F. Carvalho, G. Costa, L. Tavares, A. Amorim, L. Gusmão, Y-chromosome genetic variation in Rio de Janeiro population, *Am. J. Hum. Biol.* 18 (2006) 829–837.
- [21] D. Azevedo, L.A.F. Silva, L. Gusmão, E.F. Carvalho, Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil, *Forensic Sci. Int. Genet. Suppl. Ser.* 2 (2009) 421–422.
- [22] E. Nascimento, E. Cerqueira, E. Azevedo, W. Freitas, D. Azevedo, The Africa male lineages of Bahia's people – North East Brazil: a preliminary SNPs study, *Forensic Sci. Int. Genet. Suppl. Ser.* 2 (2009) 349–350.
- [23] M. Carvalho, P. Brito, V. Lopes, L. Andrade, M.J. Anjos, F.C. Real, L. Gusmão, Analysis of paternal lineages in Brazilian and African populations, *Genet. Mol. Biol.* 33 (2010) 422–427.
- [24] A.C.S. Góes, E.F. Carvalho, I. Gomes, D.S. Silva, E.H.F. Gil, A. Amorim, L. Gusmão, Population and mutation analysis of 17 Y-STR loci from Rio de Janeiro (Brazil), *Int. J. Legal Med.* 119 (2005) 70–76.
- [25] P.M. Domingues, L. Gusmão, D.A. da Silva, A. Amorim, R.W. Pereira, E.F. Carvalho, Sub-Saharan Africa descendants in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR loci, *Int. J. Legal Med.* 121 (2007) 238–241.
- [26] L. Gusmão, J.M. Butler, A. Carracedo, P. Gill, M. Kayser, W.R. Mayr, N. Morling, M. Prinz, L. Roewer, C. Tyler-Smith, P.M. Schneider, DNA commission of the International Society for Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, *Int. J. Legal Med.* 120 (2006) 191–200.
- [27] M.C. Dulik, A.C. Owings, J.B. Galeski, M.G. Vilar, A. Andre, C. Lennie, M.A. Mackenzie, I. Kritch, S. Snowshoe, R. Wright, J. Martin, N. Gibson, T.D. Andrews, T.G. Schurr, Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations, *Proc. Natl. Acad. Sci. U S A* 109 (2012) 8471–8476.
- [28] B. Trombetta, F. Cruciani, D. Sellitto, S. Scozzari, A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms, *PLoS ONE* 6 (2011) e16073.
- [29] N.M. Myres, S. Rootsi, A.A. Lin, M. Järve, R.J. King, I. Kutuev, V.M. Cabrera, E.K. Khusnutdinova, A. Pshenichnov, B. Yunusbayev, O. Balanovsky, E. Balanovska, P. Rudan, M. Baldovico, R.J. Herrera, J. Chianoni, J. Di Cristofaro, R. Villems, T. Kivisild, P.A. Underhill, A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe, *Eur. J. Hum. Genet.* 19 (2011) 95–101.

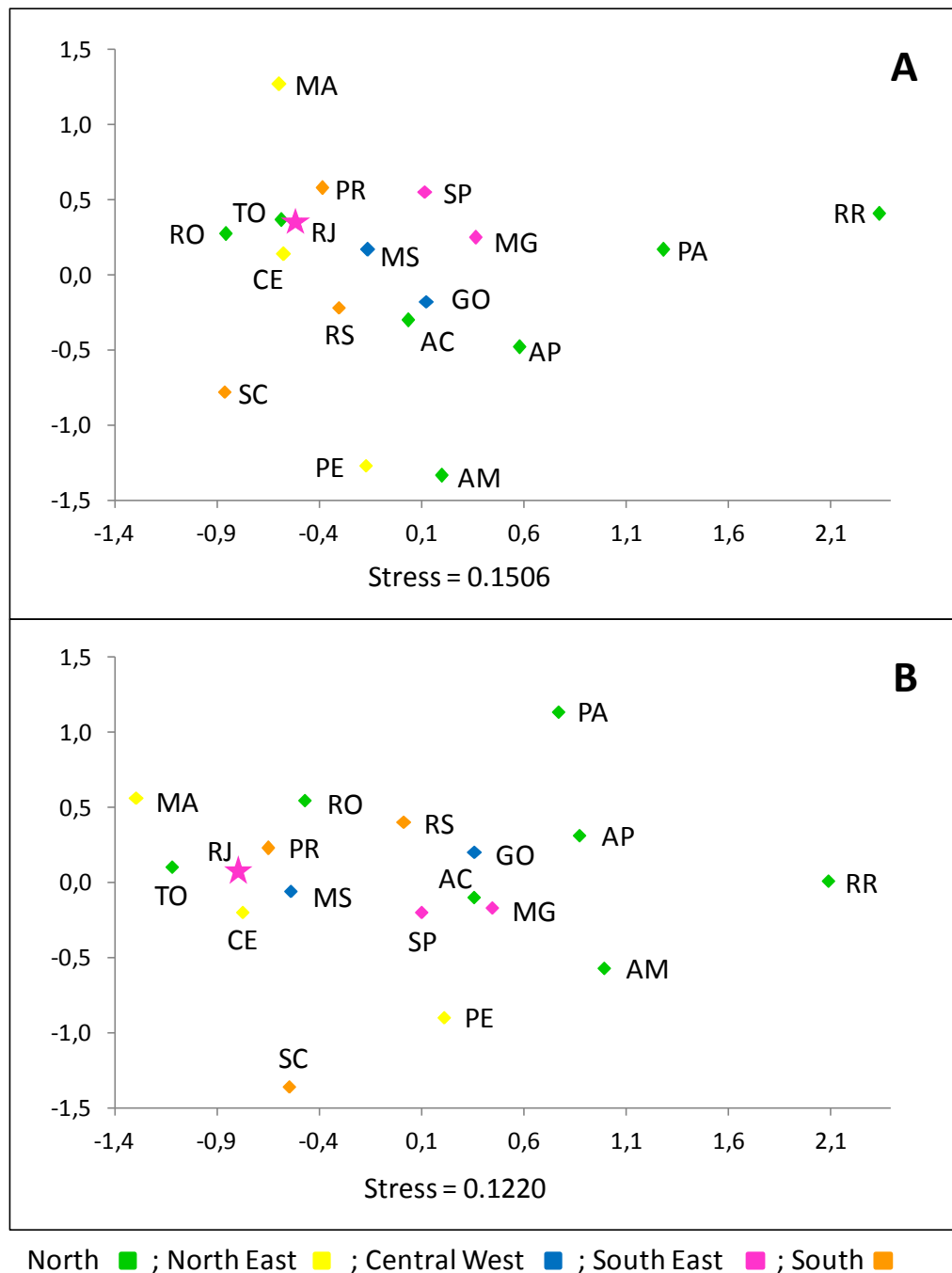
- [30] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, *Genome Res.* 18 (2008) 830–838.
- [31] V. Gomes, P. Sánchez-Diz, A. Amorim, A. Carracedo, L. Gusmão, Digging deeper into East African human Y chromosome lineages, *Hum. Genet.* 127 (2010) 603–613.
- [32] C. Vullo, V. Gomes, C. Romanini, A.M. Oliveira, O. Rocabado, J. Aquino, A. Amorim, L. Gusmão, Association between Y haplogroups and autosomal AIMs reveals intra-population substructure in Bolivian populations, *Int. J. Legal Med.* (2014). <http://dx.doi.org/10.1007/s00414-014-1025-x> (Epub ahead of print).
- [33] M. Nei, *Molecular Evolutionary Genetics*, Columbia University Press, New York, 1987.
- [34] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [35] S. Beleza, L. Gusmão, A. Lopes, C. Alves, I. Gomes, M. Giouzeli, F. Calafell, A. Carracedo, A. Amorim, Micro-phylogeographic and demographic history of Portuguese male lineages, *Ann. Hum. Genet.* 70 (2006) 181–194.
- [36] S. Beleza, L. Gusmão, A. Amorim, A. Carracedo, A. Salas, The genetic legacy of western Bantu migrations, *Hum. Genet.* 117 (2005) 366–375.
- [37] M. González, V. Gomes, A.M. López-Parra, A. Amorim, A. Carracedo, P. Sánchez-Diz, E. Arroyo-Pardo, L. Gusmão, The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y Chromosome Haplogroup R-V88, *Eur. J. Hum. Genet.* 21 (2013) 324–331.
- [38] U. Toscanini, L. Gusmão, G. Berardi, V. Gomes, A. Amorim, A. Salas, E. Raimondi, Male lineages in South American native groups: evidence of M19 traveling south, *Am. J. Phys. Anthropol.* 146 (2011) 188–196.
- [39] M.L. Pontes, L. Cainé, D. Abrantes, G. Lima, M.F. Pinheiro, Allele frequencies and population data for 17 Y-STR loci (AmpF/STR® Y-filer™) in a Northern Portuguese population sample, *Forensic Sci. Int.* 170 (2007) 62–67.
- [40] M. Carvalho, P. Brito, A.M. Bento, V. Gomes, H. Antunes, H.A. Costa, V. Lopes, A. Serra, F. Balsa, L. Andrade, M.J. Anjos, F. Corte-Real, L. Gusmão, Paternal and maternal lineages in Guinea-Bissau population, *Forensic Sci. Int. Genet.* 5 (2011) 114–116.
- [41] L. Roewer, M. Nothnagel, L. Gusmão, V. Gomes, M. Gonzalez, D. Corach, A. Sala, E. Alechine, T. Palha, N. Santos, A. Ribeiro-dos-Santos, M. Geppert, S. Willuweit, M. Nagy, S. Zweynert, M. Baeta, C. Núñez, B. Martínez-Jarreta, F. Gonzalez-Andrade, E.F. Carvalho, D.A. Silva, J.J. Builes, D. Turbon, A. Lopez-Parra, E. Arroyo-Pardo, U. Toscanini, L. Borjas, C. Barletta, E. Ewart, S. Santos, M. Krawczak, Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans, *PLoS Genet.* 9 (2013) e1003460.
- [42] S.E. Batista-dos-Santos, J.D. Rodrigues, A.K. Ribeiro-dos-Santos, M.A. Zago, Differential contribution of indigenous men and women to the formation of an urban population in the Amazon region as revealed by mtDNA and Y-DNA, *Am. J. Phys. Anthropol.* 109 (1999) 175–180.
- [43] D.R. Carvalho-Silva, F.R. Santos, J. Rocha, S.D. Pena, The phylogeography of Brazilian Y-chromosome lineages, *Am. J. Hum. Genet.* 68 (2001) 281–286.
- [44] A.R. Marrero, F.P.N. Leite, B.A. Carvalho, L.M. Peres, T.C. Kommers, I.M. Cruz, F.M. Salzano, A. Ruiz-Linares, W.A. Silva Jr., M.C. Bortolini, Heterogeneity of the genome ancestry of individuals classified as White in the State of Rio Grande do Sul, *Am. J. Hum. Biol.* 17 (2005) 496–506.



Supplementary Figure S1. Y chromosome studies (including the Y-filer markers) across Brazilian Geopolitical regions and States: distribution and population sample sizes (inside brackets). Superscript numbers refer to the following references: ¹Sanchez-Diz et al. [15]; ²Schwengber et al. [16]; ³Cainé et al. [17]; ⁴Palha et al. [18]; ⁵Barra et al. [14]; ⁶Purps et al. [14]; ⁷This study.



Supplementary Figure S2. Phylogenetic tree of Y-haplogroups analyzed in the present study. The haplogroups are named in accordance with Karafet et al. (2008) and updated according to Trombetta et al. (2011), Dulik et al. (2012) and Myres et al. (2011) for haplogroups E, Q and R, respectively.



Supplementary Figure S3. Multidimensional scaling plot of the pairwise F_{ST} (**A**) and R_{ST} (**B**) genetic distances calculated between samples from different Brazilian states: Acre (AC), Pará (PA), Rondônia (RO), Roraima (RR), Tocantins (TO), Amazonas (AM), Amapá (AP), Ceará (CE), Pernambuco (PE), Maranhão (MA), Goiás (GO), Mato Grosso do Sul (MS), São Paulo (SP), Minas Gerais (MG), Rio de Janeiro (RJ), Rio Grande do Sul (RS), Paraná (PR) and Santa Catarina (SC).

Table S1. List of the Y-STR haplotypes and Y-SNP haplogroups for 605 unrelated males from the admixed population of Rio de Janeiro, Brazil.

Sample name	DYS393	DYS390	DYS19	DYS391	DYS385	DYS439	DYS389I	DYS392	DYS389II	DYS458	DYS437	DYS448	GATAH4	DYS456	DYS438	DYS635	Haplogroup
021211-14C	13	22	14	11	11-14	12	13	13	29	17	14	18	11	16	12	23	R1b1b - M269
030109-5SP	13	22	14	10	13-14	13	12	11	28	15	16	20	11	15	10	22	I* (xI2a2) - M170
030109-11C	12	23	15	9	13-16	11	13	11	28	14	14	20	11	16	9	23	J2 - M172
030115-12SP	13	24	14	11	11-14	12	13	13	29	18	14	18	11	15	12	24	R1b1b - M269
030128-10SP	13	24	14	11	11-14	12	13	13	30	16	15	18	12	16	12	25	R1b1b - M269
030129-2SP	14	24	14	11	11-14	12	13	13	29	16	14	19	12	15	12	24	R1b1b - M269
030204-6SC	12	24	13	10	12-19	13	12	11	29	16.2	14	20	10	15	10	21	J* (xJ1a,2) - 12f2a
030205-2C	13	22	15	10	14-14	12	13	11	30	20	16	21	11	15	10	21	G - M201
030205-4C	13	22	15	10	12-14	12	12	11	29	19	16	18	11	15	10	21	G - M201
030205-11C	13	24	14	10	14-16	11	13	13	26	18	14	19	11	16	9	21	T - M70
030206-6C	13	22	16	11	11-12	13	13	11	28	17	15	20	12	14	10	21	I2a2 - M26
030210-10SP	13	24	14	10	11-14	12	13	13	30	18	15	19	12	15	12	23	R1b1b - M269
030210-11SP	12	23	14	10	15-17	12	14	14	31	16	16	20	11	16	11	21	L - M22
030211-5SP	13	24	12	11	12-14	11	13	13	29	16	15	20	12	15	12	23	R1b1b - M269
030213-1SP	12	23	14	11	12-14	13	13	13	30	16	14	19	12	16	12	23	R1b1b - M269
030217-13SP	14	21	17	11	16-17	13	13	11	30	15	13	21	11	16	11	21	E1b1a1f - M191
030220-8SP	13	24	14	11	11-11	13	14	13	30	19	15	19	12	16	13	23	R1b1b - M269
030220-11SP	13	25	14	11	11-14	12	13	13	29	18	14	18	11	16	12	24	R1b1b - M269
030224-4SP	14	21	16	10	15-18	11	14	11	31	18	14	21	12	15	11	21	E1b1a1* (xE1b1a1f,g1c) - M2
030324-2C	14	22	15	10	13-17	11	14	12	31	16	14	19	11	13	10	21	I* (xI2a2) - M170
030326-3SP	13	23	13	9	13-14	10	14	11	30	18	14	20	12	16	10	21	E1b1b1b1 - M81
030327-2C	13	23	14	10	13-14	11	12	11	29	16	16	20	11	14	10	23	I* (xI2a2) - M170
030401-10SP	13	25	14	10	11-14	12	13	13	29	17	15	18	12	16	12	23	R1b1b - M269
030409-6C	13	23	14	11	11-14	12	13	13	29	16	14	19	11	15	12	23	R1b1b - M269
030409-10SP	13	24	14	11	11-14	14	14	13	29	17	15	19	12	17	12	24	R1b1b - M269
030409-12C	13	24	14	11	11-14	12	13	13	29	18	15	19	12	16	11	24	R1b1b - M269
030414-1SP	13	25	14	10	11-14	12	13	13	29	18	15	19	12	16	12	23	R1b1b - M269
030417-4C	13	21	15	10	16-17	11	13	11	30	17	14	21	12	15	11	21	E1b1a1g1c - M154
030424-6SP	13	23	14	11	11-14	12	13	13	29	17	15	19	12	15	12	23	R1b1b - M269
030424-8SP	15	23	15	10	15-15	11	14	12	31	15	14	20	11	15	10	22	I* (xI2a2) - M170
030425-14SP	13	24	15	11	11-14	11	13	14	29	17	15	19	11	16	12	23	R1b1b - M269
030428-10SP	13	24	15	11	14-17	12	12	11	28	16	16	19	11	13	9	21	J2 - M172
030505-1C	13	24	14	11	11-14	12	13	13	28	17	15	19	12	15	12	23	R1b1b - M269
030505-12C	13	23	14	10	11-14	11	13	15	30	17	14	19	12	15	12	23	R1b1b - M269
030507-26C	13	24	13	10	16-16	11	13	11	30	16	14	20	12	17	10	22	E1b1b1a1 - M78
030514-8C	12	24	14	10	13-14	13	13	14	28	18	15	19	12	16	12	25	R1b1b - M269
030514-9SP	13	24	14	11	12-14	12	14	13	30	17	15	19	11	16	12	23	R1b1b - M269
030514-10C	13	24	15	10	11-15	12	14	13	31	18	15	19	12	16	12	23	R1b1b - M269
030514-13SP	13	24	15	11	14-15	12	13	11	30	17	15	19	11	15	10	23	I* (xI2a2) - M170
030514-17SP	13	23	15	10	14-19	11	13	14	31	17	14	19	12	16	9	21	T - M70
030515-17C	13	23	14	11	11-14	12	13	13	30	17	15	19	12	15	12	23	R1b1b - M269
030515-19SP	12	23	14	10	13-18	11	13	11	31	17.2	14	20	11	15	10	19	J* (xJ1a,2) - 12f2a
030520-5SP	13	24	13	10	16-16	11	13	11	30	17	14	20	12	17	10	23	E1b1b1a1 - M78
030521-9SP	12	24	15	9	13-16	12	13	11	29	15	14	20	12	16	9	23	J2 - M172
030526-6SP	13	24	14	11	12-14	12	13	13	30	18	15	19	12	16	12	23	R1b1b - M269
030602-9C	13	24	14	11	12-14	12	13	13	29	18	16	19	12	16	12	23	R1b1b - M269
030602-12C	13	24	14	11	11-11	11	13	13	29	18	15	19	12	15	12	23	R1b1b - M269

120315-44C	13	24	14	10	11-12	14	12	13	28	19	15	19	12	15	12	23	R1b1b - M269
120330-31C	13	23	14	10	14-17	11	14	11	31	17	14	22	11	18	9	23	J2 - M172
120411-33C	12	24	15	10	13-18	12	13	11	30	15	14	20	10	15	10	22	E1b1b1c - M123
120411-42C	14	25	14	10	11-15	12	13	13	29	17	15	20	12	15	12	23	R1b1b - M269
120416-03C	13	22	14	10	11-14	13	13	13	29	18	14	18	11	16	12	23	R1b1b - M269
120426-04C	13	24	14	10	11-14	13	14	13	30	18	14	18	11	15	12	23	R1b1b - M269
120426-10C	13	24	14	11	12-14	11	13	13	29	17	14	19	12	15	11	23	R1b1b - M269

5.2 Caracterização das linhagens masculinas de uma amostra do oeste africano

Na segunda parte deste trabalho de tese, foi feita a caracterização do patrimônio genético de amostras do oeste africano, de indivíduos não relacionados de diferentes países, quanto aos polimorfismos do tipo STR e SNP localizados na região não recombinante do cromossomo Y. Como já citado, este foi um trabalho realizado em colaboração com a Universidade de Colônia na Alemanha. Aos dados encontrados foram adicionados outros dados já publicados para os mesmos marcadores em outras populações africanas, para elaborar uma base de dados representativa das 3 principais regiões da África (Sudeste = Moçambique; Sudoeste = Angola; Centro-Oeste = Benin, Costa do Marfim e Nigéria) que, historicamente, foram relacionadas aos locais de saída/envio de escravos para o Brasil (KLEIN, 2002).

Foi demonstrado, pelos resultados obtidos, que existe uma homogeneidade genética entre a amostra do oeste africano estudada e os diferentes países da região oeste (compreendida entre a Guiné Equatorial e a Guiné-Bissau). Segundo Guthrie (1948), os países que estão nesta região, pertencentes à África subsaariana, compreendem majoritariamente etnias de origem Bantu, incluídas no grupo linguístico denominado como Niger-Congo B, o que explica a homogeneidade genética encontrada.

Por outro lado, observamos diferenças significativas quando comparamos a amostra do oeste africano com populações de Uganda, Quênia e África do Sul, o que seria esperado visto que são países que compreendem outras etnias além das de língua Bantu, pertencendo majoritariamente a outros grupos linguísticos (Nilo-saariano e Khoisan). Contudo, também foram observadas diferenças entre a amostra estudada e a população da Guiné Bissau. Apesar de também apresentarem uma grande influência Bantu, estes dois países possuem populações pertencentes a outra divisão linguística, também segundo classificação de Guthrie (1948), denominada de grupo linguístico Niger-Congo A.

Os dados obtidos deste estudo compreendendo 86 indivíduos do oeste africano (Senegal, Gâmbia, Guiné, Serra Leoa, Costa do Marfim, Gana, Togo, Benin, Nigéria, Niger e Libéria), bem como a sua análise cuidadosa, foram

organizados no formato de artigo, para a submissão em revista científica. Este trabalho encontra-se em processo de finalização para avaliação por todos os autores, e é apresentado em seguida, juntamente com o seu resumo em Português e material suplementar associado.

As revistas para a submissão propostas são: *Electrophoresis* ou *International Journal of Legal Medicine*.

Artigo Científico II:

Y chromosome genetic profile of Western Africa using SNP and STR loci

RESUMO:

O padrão de variação humana na África é o resultado de uma série de eventos demográficos complexos. Um dos mais importantes se refere à dispersão maciça dos povos Bantu, que representou um dos maiores fenômenos de migração humana na África, e que esteve ligado à expansão da agricultura e difusão da metalurgia (manufatura do ferro).

O cromossomo Y tem sido amplamente estudado exhaustivamente em populações humanas e o conhecimento produzido a partir dos seus marcadores genéticos têm sido uma ferramenta importante no estudo das populações, assim como, na história demográfica masculina e seus padrões de migração.

Construir bases de dados referentes aos marcadores do cromossomo Y é um esforço produzido para conhecer as diferenças entre populações. As diferenças observadas, tanto entre como dentro das populações, dependem dos tipos e número de marcadores utilizados. Alguns estudos mostram uma significativa correlação entre a distribuição da linguagem e a variação genética associada aos cromossomos Y na África, o que pode ser visto claramente através da semelhança das linhagens paternas entre membros de grupos situados na rota da expansão Bantu.

Em vista dos poucos dados populacionais existentes sobre os povos da África, principalmente da região oeste e central, as análises baseadas em diferentes marcadores podem contribuir para o entendimento da estratificação destas

populações e também para a ampliação das bases de dados, assim como promover novos conhecimentos na história demográfica deste continente.

A partir de uma amostra de 86 indivíduos não relacionados da região oeste da África foram feitas análises de 17 Y-STR que foram comparados com outras populações representativas de outras regiões africanas. A diversidade haplotípica observada para esta região foi de 1,0000 ($\pm 0,0018$). Este valor elevado de diversidade é similar aos obtidos para outras amostras africanas já estudadas, utilizando o mesmo conjunto de marcadores genéticos. De fato, este resultado confirma o que seria esperado, devido ao fato de que, as populações africanas são relativamente mais antigas do que as dos demais continentes.

Esta amostra também foi genotipada para 31 Y-SNPs. Como o esperado, a partir dos resultados publicados previamente, referentes à ancestralidade paterna africana em populações subsaarianas de origem Bantu, a maioria das linhagens de Y-SNP encontradas pertence ao haplogrupo E, onde o alelo derivado M2 relaciona-se de forma intrínseca às populações tipicamente de origem Bantu. Outras linhagens encontradas incluíram 3 representantes dos paragrupos A e B, típicos de populações Khoisan e Nilota, bem como dois cromossomos pertencentes aos haplogrupos R1b1b e T, cuja presença na nossa amostra é explicada por influxo europeu recente.

A diversidade de haplogrupos encontrada foi de 0,6895 ($\pm 0,0200$), sendo da mesma magnitude que o valor encontrado para o mesmo conjunto de marcadores estudados em outras regiões da África, especificamente os países onde encontramos os grupos Bantu (0,681 and 0,753 em Angola e Guiné Equatorial, respectivamente). No entanto, o valor de diversidade encontrado nesta região é mais baixo do que o valor anteriormente observado em uma população do Leste africano (e.g. 0.831 \pm 0.022 em Uganda), região que é considerada como a de origem provável do homem moderno e possível região da expansão “*Out of Africa*”.

Em relação às populações da África subsaariana, para as quais se dispõe de dados genéticos para os mesmos marcadores de cromossomo Y analisados neste trabalho, foi feita uma análise comparativa entre amostras de diferentes países pertencentes às seguintes regiões geográficas: Oeste, Centro, Sudeste, Sudoeste e Leste. Esta análise comparativa foi feita com base em distâncias genéticas calculadas a partir das frequências encontradas, tanto de haplótipos como de haplogrupos.

Em relação às populações do Oeste, foram observadas diferenças significativas entre a Guiné-Bissau e as populações restantes desta região. Em vista disto, esta amostra foi excluída do agrupamento representativo do Oeste. Da mesma forma, as amostras de Uganda e Quênia não foram agrupadas como representantes da região Leste, por apresentarem diferenças estatisticamente significativas entre si. Desta forma, foram comparadas com as outras regiões e países de forma separada.

Quanto à amostra relativa ao nosso estudo, verificou-se que esta apresenta uma composição genética semelhante à da amostra representativa do Oeste Africano. Além disso, observou-se não existirem diferenças significativas entre a nossa amostra e as amostras do Sudeste Africano e da África Central. Mais uma vez, a nossa população mostrou-se significativamente diferente da população da Guiné Bissau, apesar de ambas pertencerem à região Oeste da África.

Existe uma grande afinidade genética entre as regiões Oeste e Centro africanas, em parte devido a uma predominância dos povos de origem Bantu. Contudo, em relação à Guiné-Bissau, quando comparada a todas as populações das regiões Oeste e Central, verifica-se que existem diferenças significativas, evidenciando uma diferenciação genética relativamente às restantes populações, que poderá estar relacionada com uma maior distância relativamente às rotas da expansão Bantu.

Os resultados encontrados neste trabalho demonstram uma homogeneidade genética entre a maioria dos países pertencentes ao Oeste africano. Isto permite então, a reunião dos dados disponíveis numa só base de dados genéticos representativa desta região.

Verificou-se uma alta diversidade das linhagens de cromossomo Y no Oeste africano, contudo esta diversidade apresentou-se mais baixa do que aquela observada em populações do Leste africano.

Title: “Genetic profile of Western Africa using Y-chromosomal SNP and STR loci”**Abstract**

The pattern of human variation in Africa is the result of a complex series of demographic events. There are few databases related to the peoples of Africa, especially in western and central regions. In this study, 17 Y-STR and 31 Y- SNP loci were genotyped in a sample from Western Africa, which was compared with other representative populations of other African regions. The haplotype diversity observed in this region was 1.0000 (± 0.0018), a value that is similar to those obtained in other African samples already studied. The most frequent haplogroup found in our sample belongs to E, showing the M2 derived allele. The haplogroup diversity was 0.6895 (± 0.0200), a value that is very similar to that observed in other parts of Africa, namely in countries where the Bantu groups are predominant (0.681 and 0.753 in Angola and Equatorial Guinea, respectively). Concerning the Western Africa, significant differences were observed between Guinea-Bissau and the remaining populations of this region. In addition, there was no significant difference between our sample and the samples from Eastern and Central Africa. The genetic composition of the population studied was significantly different from the population of Guinea-Bissau, although a genetic homogeneity was found among most countries belonging to West Africa.

1. Introduction

The African continent contains a tremendous cultural, linguistic and genetic diversity, with more than 2,000 distinct ethnic groups and languages. The demographic history has consisted of fluctuations in population size, short- and long-range migrations and admixture, which have resulted in complex patterns of variation in modern populations [Campbell & Tishkoff, 2010; Campbell *et al.*, 2014]. The pattern of human variation in Africa was the result of a *complex series of demographic events, and the massive dispersal of Bantu-speakers characterizes one of the principal examples of human migration* [Coelho *et al.*, 2009]. Africa has

therefore been characterized as the most genetically diverse region of the world [Campbell & Tishkoff, 2008]. All of these features make African populations extremely interesting for genetic studies allowing for a better understanding of the complex diversity of this continent.

The identification of thousands of genetic markers, particularly short tandem repeat (STR) loci and single nucleotide polymorphisms (SNPs), distributed throughout the human genome have become efficient genetic tools for individual identification and determination of genetic relationships among individuals in population and forensic genetics [Góes *et al.*, 2005; Rosa *et al.*, 2007; Filippo *et al.*, 2010]. Studies using autosomal STRs are also important to perceive population-related evolutionary forces, such as population bottlenecks and genetic drift (among others), which impact on the dynamics of these loci. The dynamics of STRs within specific populations is also relevant in forensic genetics as these are the most used markers for individual identification [Sun *et al.*, 2003]. In addition, investigations of uniparental markers, such as Y chromosome or mitochondrial DNA (mtDNA) markers, offer valuable insights into human history that cannot be revealed by autosomal markers alone [Filippo *et al.*, 2011]. These systems allow the reconstruction of phylogenies based on mutations and the tracing of the origins of polymorphisms as well as the identification of the geographical spread, which is not possible with autosomal data due to recombination.

The male-specific region of the Y chromosome (MSY) has been extensively studied in human populations [e.g., Scheinfeldt *et al.*, 2010], and the study of markers in this region have proved to be a very useful tool in human evolutionary studies [Hammer & Zegura, 1996], disclosing male demographic histories and migration patterns [Carvalho-Silva *et al.*, 2001, Abe-Sandes *et al.*, 2004; Góes *et al.*, 2005]. Furthermore, demic migrations often involve higher numbers of males than females, with a sex biased mating between male migrants and local females, and, therefore, Y chromosome markers are more sensitive than mtDNA or autosomal markers to trace that movements [Cavalli-Sforza & Minch, 1997].

The Y chromosome has a non-recombining nature and it is necessary to study a large number of STR markers at the same time to increase levels of discrimination within populations. Very high levels of diversity is not always desirable in population genetics, because it is necessary to study a large number of samples

to statistically support small differences between populations, or to detect population stratification [Oliveira *et al.*, 2014].

Despite the lower power to discriminate different male lineages, the Y-SNPs are very useful to define haplogroups with high geographical specificity and are therefore more apt to detect population stratification in admixed populations from diverse continental origins. Some studies showed a significant correlation between language and the variation in Y chromosome in Africa, which can be clearly seen within the Bantu language family [Rosa *et al.*, 2007; Filippo *et al.*, 2010], mainly by events connected with haplogroups defined by the single nucleotide polymorphisms M2 (E1b1a1*) and M150 (B2a*) (Berniell-Lee *et al.*, 2009).

Despite the extensive efforts on worldwide allele/haplotype frequencies' data basing, differences observed within and between populations can change depending on the regions and the different type and numbers of markers under analysis. In this way, analyses based on different markers can provide new hints on populations' stratification, as well as offer deeper insights on their demographic history, namely in what concerns the poorly studied populations of Africa.

This study aimed to characterize samples from West African region, using Y-chromosomal specific markers with different mutation rates, namely 17 Y-STR and 31 Y-SNP loci, in order to better discern genetic stratification in this region. This is important for a better understanding of the genetic profile of the current African populations, and a more accurate interpretation of profile matches in kinship and forensic casework.

2. Methods and Materials

2.1. Population samples and DNA extraction

Samples from unrelated individuals from different countries in West Africa (Figure 1) were selected from routine paternity cases from the Institute of Legal Medicine in Cologne (Germany). This study was approved by the Ethics Committee of this Institute and samples were collected with the informed consent from the donors. DNA extraction was performed using the EZ1 Investigator kit and EZ1 Robot (Qiagen) according to the recommended protocol.

2.2. Y chromosome STR typing

The AmpF/STR Yfiler kit (Thermo Fisher Scientific Inc) was used to analyze 17 Y-STR loci in 86 male samples (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and GATA H4), following manufacturer's recommendations. Detection was carried out using an ABI 3130 capillary electrophoresis (Thermo Fisher Scientific Inc) and genotyping was performed using GeneMapper ID Software v3.2 (Thermo Fisher Scientific Inc). Y haplotype data were submitted to the Y-STR Haplotype Reference Database (<http://www.yhrd.org>).

2.3. Y chromosome SNP typing

In this study, 31 Y-SNPs were chosen to discriminate the most frequent African haplogroups (Figure 2). These 31 Y-SNPs were typed in the same male samples studied for Y-STRs using PCR and SNaPshot methods with different multiplex reactions previously described by Gomes *et al.* [2010]. The haplogroup nomenclature followed Trombetta *et al.* [2011] for haplogroup E, Myres *et al.* [2011] for haplogroups R and Karafet *et al.* [2008] for the remaining haplogroups.

2.4. Data analysis

The number of haplotypes/haplogroups was determined by direct counting. Population comparisons, haplotype frequencies and genetic diversities according to Nei [1987] were calculated using the Arlequin v3.5 software (Excoffier & Lischer, 2010). For population comparisons using the Y chromosome STR data, the number of repetitions in DYS389I was subtracted from DYS389II and DYS385 was excluded. To match Arlequin STR data format duplicated, null and non-consensus alleles were represented as missing data. Population pairwise genetic distances were conducted based on F_{ST} and the significance was tested with 10,000 permutations. Pairwise genetic distances were visualized in two dimensional space using the multi-dimensional scaling (MDS) analysis included in the StatSoft, Inc. (2007) program, STATISTICA (data analysis software system), ver.8.0 (www.statsoft.com). When

applicable p values were corrected using the Bonferroni's correction for multiple testing calculations.

3. Results and Discussion

3.1. Y-chromosomal diversity

The genotyping results of a population sample of 86 unrelated males from Western Africa for 17 Y-STR and 31 Y-SNP markers are listed in Supplementary Table S1. A high Y-STR based haplotype diversity was observed (1.0000 ± 0.0018). The observed haplotype diversity was similar to that obtained for other samples from African countries that have been studied for the same conjunto of Y chromosome STR markers (Figure 3).

In the West African population sample that we have analysed, a total of 8 different haplogroups were detected, which ancestry could be traced to sub-Saharan Africa, except for two chromosomes that belong to European haplogroups R1b1 and T (Figure 2). The haplogroup diversity was 0.6895 ± 0.0200 , a value of the same magnitude of those usually found in other Western populations (e.g. 0.681 and 0.753 in Angola and Equatorial Guinea, respectively), when studying the same set of markers. This value is, nevertheless, lower than the one obtained in a sample from Uganda (0.831 ± 0.022), which support previous studies pointing to East Africa as the most genetically diverse region of the world (Salas *et al.*, 2002; Gonder *et al.*, 2007; Hassan *et al.*, 2008; Poloni *et al.*, 2009).

3.2. Comparison of the Y-STR haplotypes in African populations

Before comparing our data with the remaining data available from Africa, for the same group of Y-STRs, populations from different countries have been compared and grouped in 5 regions (Table 1), whenever non-significant differences were observed among population from the same geographic region. Population comparisons were performed through F_{st} analysis based on the number of different alleles (F_{st} -like) after excluding the DYS385 from the AmpF/STR Yfiler markers' conjunto. MDS plot was performed to allow an easier interpretation of the pairwise distances (Figure 4).

No significant differences were observed between Mozambique and Zimbabwe – grouped in the Southeast African region – and between Equatorial Guinea, Cameroon and Democratic Republic of Congo – representing the Central region of Sub-Saharan Africa. Concerning West African populations, including Guinea-Bissau, Ivory Coast, Nigeria and Benin, a significant difference was detected between Nigeria and Guinea-Bissau. For this reason, Ivory Coast, Nigeria and Benin were pooled together - representing West Africa – separated from Guinea-Bissau.

Finally, the two countries from East Africa showed significant differences and, therefore, they were not joined together. The sample from Uganda belongs to an ethnic group of Eastern Nilotes (the Karimojong) and the sample from Kenia includes individual belonging to two different ethnic groups: the Maasai (Eastern Nilotes) and the Luhya (Bantu). In order to elucidate if the differences found between Uganda and Kenia were due to different ethnic compositions, the two groups from Kenia were separated and compared with the Karimojong from Uganda (data not shown). As expected, the highest genetic distances were observed between the Bantu and both Nilotic groups ($F_{ST}=0.1419$, for Uganda; $F_{ST}=0.1398$, for Kenia). Although lower, the genetic distance among the Nilotes from Uganda and Kenia remained high ($F_{ST}=0.0604$).

After grouping samples by regions following the previous results, the haplotype frequencies in the West African sample included in the present work were compared with other African data, representing: Guinea Bissau; West Africa (Nigeria, Benin and Ivory Coast); Central Africa (Equatorial Guinea, Cameroon and Democratic Republic of Congo); Southwest Africa (Angola); South Africa; Southeast Africa (Mozambique e Zimbabwe); Kenia and Uganda (Tabela 2 and Figure 5).

The results revealed no significant differences between all populations studied (for a significance level obtained by applying Bonferroni's correction), except to Uganda, Kenia and South Africa. These three populations revealed significant differences between each other as well as when compared with the remaining African regions.

3.3. Comparison of the Y-SNP haplogroups in Africa

As expected from the STR results, as well as from previous publications concerning the paternal ancestry of sub-Saharan populations (e.g. [Beleza *et al.*,

2005; Coelho *et al.*, 2009; Filippo *et al.*, 2010, 2011; Gonzalez *et al.*, 2010; Carvalho *et al.*, 2011; Tishkoff *et al.*, 2011]), the majority of the Y-SNP lineages found in this study belong to the haplogroup E. The SNP frequencies found in our sample and in other African samples are in Table 3. Inside haplogroup E, most lineages hold the M2-derived allele that is typical of Sub-Saharan Bantu populations (55.8%); followed by samples with the M191-derived allele (32.56%).

A single individual was found to belong to haplogroup A, a haplogroup that reach its highest frequencies in populations with Khoisan or Nilote ancestry. Two other individuals have been found to also carry haplogroups that are well represented in Khoisan and Nilote groups, belonging to the haplogroup B2a*-M150 ((Underhill *et al.*, 2000; Cruciani *et al.*, 2002; Gomes *et al.*, 2010). Both haplogroup A and B lineages that we found in our sample are most probably of Khoisan origin, since the migration routes involved in the dispersion of Nilote people from Ethiopia to Kenia, Tanzania, Uganda and did not went further West of Sudan (Pazzaglia, 1982; Maxon, 1994)

A small proportion of about 2% of the lineages in the studied sample from Western Africa belong to haplogroup R1b1-P25, which is rare in Africa and mainly found in Europe and Asia (Gonzalez *et al.*, 2013). The presence of these lineages in our sample can be explained by admixture with European men during the colonial period.

Rosa *et al.* (2007) described a high frequency of M2 in Central and Eastern Africa, reaching its highest values in nearby Senegal as well as in a region between Nigeria and Gana. This data correspond to the same region that we have sampled in this study, showing similarity between populations that are in the same region, except when comparing the Guinea-Bissau population. In fact, all samples of the western region that were studied can be grouped into a single database, excluding Guinea-Bissau.

In summary, the results of this study demonstrate a genetic homogeneity among most countries belonging to West Africa. This allows the compilation of available data of different studies from Western African populations to produce a representative genetic database of this region, increasing the available information with relevance for genetic population and forensic studies.

Conclusions:

The results of this study demonstrate a genetic homogeneity among most countries belonging to West Africa. The high haplotypic diversity in West Africa was observed but it is lowest when compared with populations from East Africa. We showed a high male mediated genetic affinity between the West and Center of Africa, influenced by the characteristic phenomenon of the Bantu expansion in these directions. However, significant differences were found between Guinea-Bissau and populations of the West and Central regions, contrasting with the high genetic homogeneity created by the Bantu expansion that happened further south, not reaching this particular region.

References

Berniell-Lee G, Calafell F, Bosch E, Heyer E, Sica L, Mouguiama-Daouda P, van der Veen L, Hombert JM, Quintana-Murci L, and Comas D. (2009). Genetic and Demographic Implications of the Bantu Expansion: Insights from Human Paternal Lineages. *Mol. Biol. Evol.* 26 (7): 1581–1589.

Bonferroni, CE (1936) Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8:3–62.

Beleza S, Gusmão L, Amorim A, Carracedo A, Salas A. (2005). The genetic legacy of western Bantu migrations. *Hum. Genet.* 117: 366-375.

Carvalho M, Brito P, Bento AM, Gomes V, Antunes H, Costa HA, Lopes V, Serra A, Balsa F, Andrade L, Anjos MJ, Corte-Real F, Gusmão L. (2011). Paternal and maternal lineages in Guinea-Bissau population, *Forensic Sc. Int: Genet.* 5: 114-116.

Cavalli-Sforza LL, Minch E (1997). Paleolithic and Neolithic Lineages in the European Mitochondrial Gene Pool. *Am. J. Hum. Genet.* 61: 247-251,

Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J. (2009). On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol. Biol.* 9:80, doi: 10.1186/1471-2148-9-80.

Cruciani F, Santolamazza O, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace D.C, Oefner P.J., Torroni A, Cavalli-Sforza L. L., Scozzari R, Underhill P.A. A back migration from Asia to Sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *A Hum Genet* 70:1197–1214, 2002

Desmarais D, Zhong Y, Chakraborty R, Perreault C, Busque L (1998) Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). *J Forensic Sci* 43(5):1046-9.

Excoffier L, Lischer HEL (2010) Arlequin ver. 3.5, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567.

Filippo C, Heyn P, Barham L, Stoneking M, Pakendorf B. (2010). Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia. *Am. J. Phys. Anthropol.* 141:382-394.

Filippo C, Barbieri C, Whitten M, Mpoloka SW, Gunnarsdóttir ED, Bostoen K, Nyambe T, Beyer K, Henning S, Knijff P, Luiselli D, Stoneking M, Pakendorf B. (2011). Y-Chromosomal Variation in Sub-Saharan Africa: Insights Into the History of Niger-Congo Groups. *Mol. Biol. Evol.* 28(3):1255-1269.

Gomes I, Alves C, Maxzud K, Pereira R, Prata MJ, Sánchez-Diz P, Carracedo A, Amorim A, Gusmão L (2007). Analysis of 10 X-STRs in three African populations. *Forensic Sci Int Genet.* 1 (2): 208-11.

Gomes I, Pereira V, Gomes V, Prata MJ, Pinto N, Carracedo A, Amorim A, Gusmão L (2009). The Karimojong from Uganda: genetic characterization using an X-STR decaplex system. *Forensic Sci Int Genet.*3(4):e127-8.

Gomes V, Sánchez-Diz P, Amorim A., Carracedo A. Gusmão L. (2010). Digging deeper into East African human Y-chromosome lineages. *Hum. Genet. Mar;* 127(5):603-13.

Gonder M K, Mortensen H M, Reed F A, de Sousa A, Tishkoff S A. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol.* 24: 757-768.

González M, Gomes V, López-Parra AM, Amorim A, Carracedo A, Sánchez-Diz P, Arroyo-Pardo E, and Gusmão L (2013). The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y chromosome haplogroup R-V88. *Eur. J. Int. Hum. Genet.* (2013) 21, 324–331.

Hassan H Y, Underhill P A, Cavalli-Sforza L L, Ibrahim M E. 2008. Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography, and history. *Am. J. Phys. Anthr.* 137: 316-323.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, *Genome Res.* 18 830-838.

Laan M, Wiebe V, Khusnutdinova E, Remm M, Pääbo S (2005) X-chromosome as a marker for population history: linkage disequilibrium and haplotype study in Eurasian populations. *Eur. J. Hum. Genet.* 13(4):452-62.

Lahn, BT & Page DC (1999). Four evolutionary strata on the human X chromosome. *Science* 286 (5441): 964-967.

Maxon, RM. East Africa: an introductory history, West Virginia, University Press. 1994.

Palha, TJ, Gusmão, L, Ribeiro-Rodrigues, E, Guerreiro, JF, Ribeiro-dos-Santos, A (2012). Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes. *PLoS ONE* 7.

Pazzaglia A. The Karimojong some aspects. EMI, Bologna. 1982.

Poloni E S, Naciri Y, Bucho R, Niba R, Kervaire B, Excoffier L, Langaney A, Sanchez-Mazas A. 2009. Genetic evidence for complexity in ethnic differentiation and history in East Africa. *Ann Hum Genet* 73: 582-600.

Purps, J *et al.*, 2014. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci.Int.Genet.* 12: 12–23.

Salas A, Richards M, De la Fe T, Lareu M V, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. *Am. J. Int. Hum. Genet.* 71: 1082-1111.

Schaffner, SF (2004) The X chromosome in population genetics. *Nat. Rev. Genet.* 5(1): 43-51.

Szibor R, Plate I, Edelman J, Hering S, Kuhlisch E, Michael M, Krause D. (2003) Chromosome X haplotyping in deficiency paternity testing principles and case report. International Congress Series 1239, 815–820.

Szibor, R (2007) X-chromosomal markers: past, present and future. *Forensic Sci Int Genet* 1(2): 93-99.

Trombetta B, Cruciani F, Sellitto D, Scozzari R. (2011). A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms, *PLoS One* 6 e16073.

Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Oligonucleotide3 - new capabilities and interfaces. *Nucleic Acids Research* 40(15):e115.

Saitou, N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonn -Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ. Y chromosome sequence variation and the history of human populations (2000). *Nat Genet*. Nov; 26 (3):358-61.

Vullo C, Gomes V, Romanini C, Oliveira A.M, Rocabado O, Aquino J, Amorim A, Gusm o L (2014). Association between Y haplogroups and autosomal AIMs reveals intra-population substructure in Bolivian populations, *Int. J. Legal Med.* [Epub ahead of print].

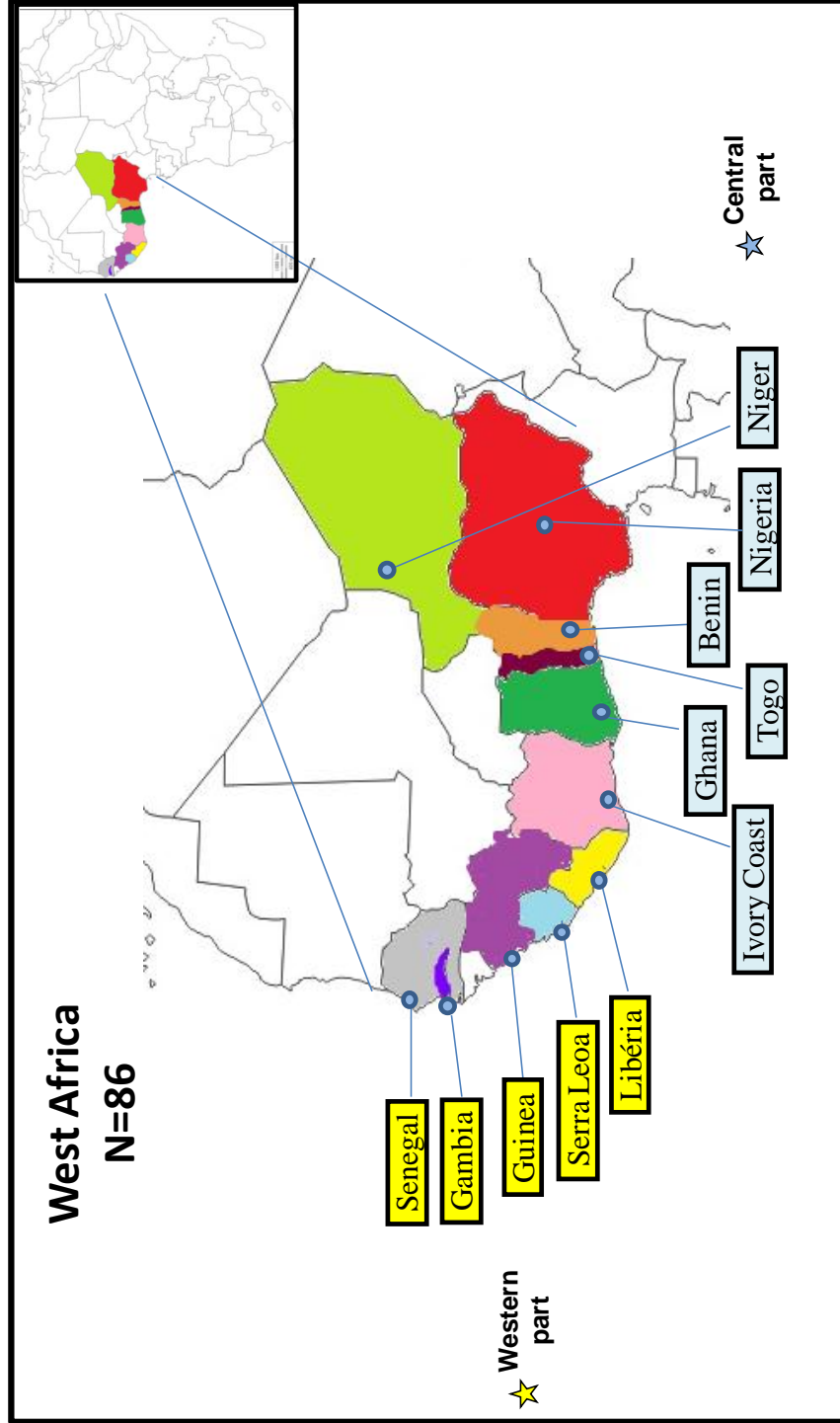


Figure 1 – Map of West Africa showing the countries of origin of the samples studied. The samples were organized in two groups following affinities observed in STR.

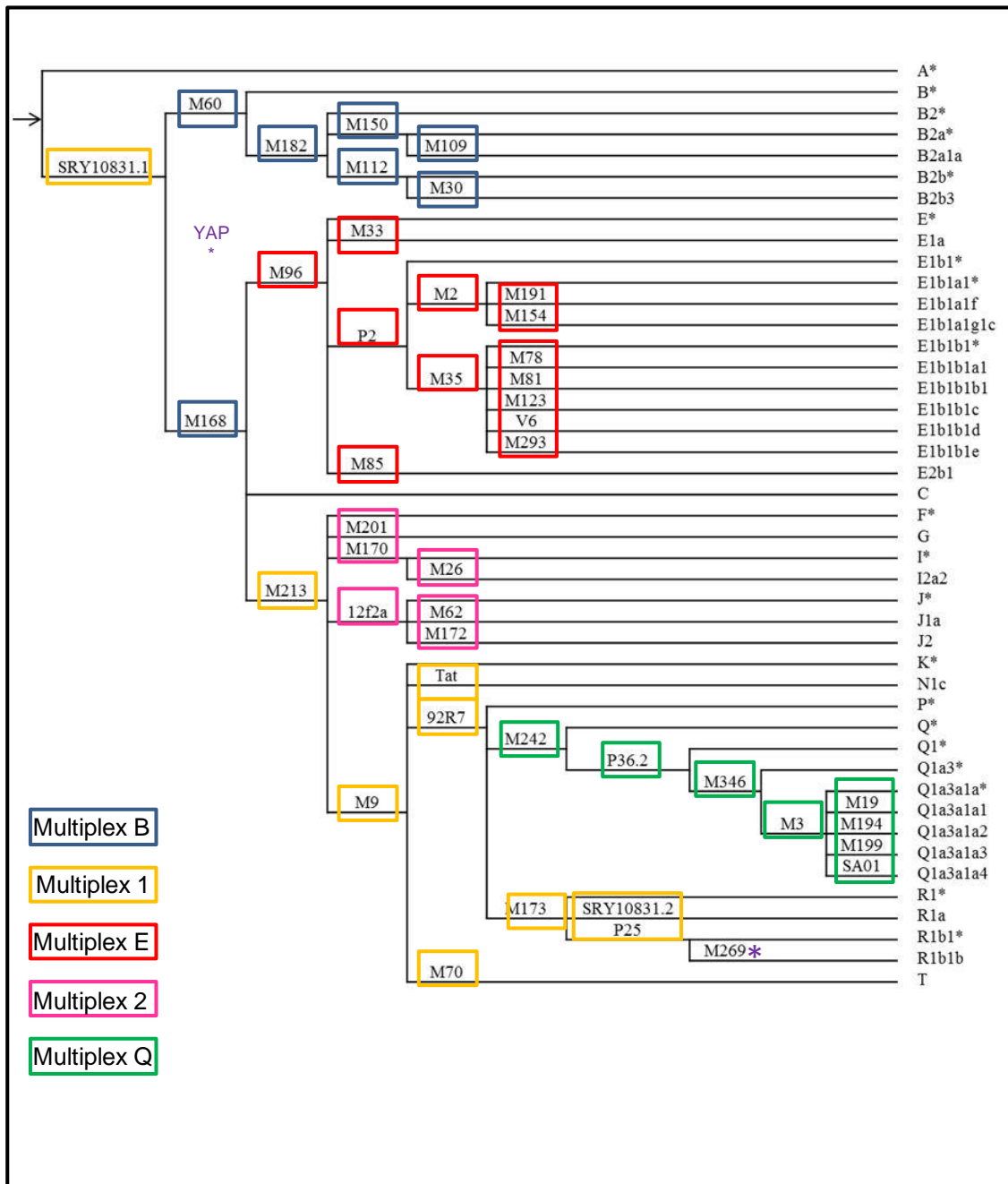


Figure 2 - Phylogenetic tree of Y-haplogroups analyzed in the present study. The haplogroups are named in accordance with Karafet *et al.* (2008) and updated according to Trombetta *et al.* (2011) for haplogroup E and Myres *et al.* (2011) for haplogroup R.

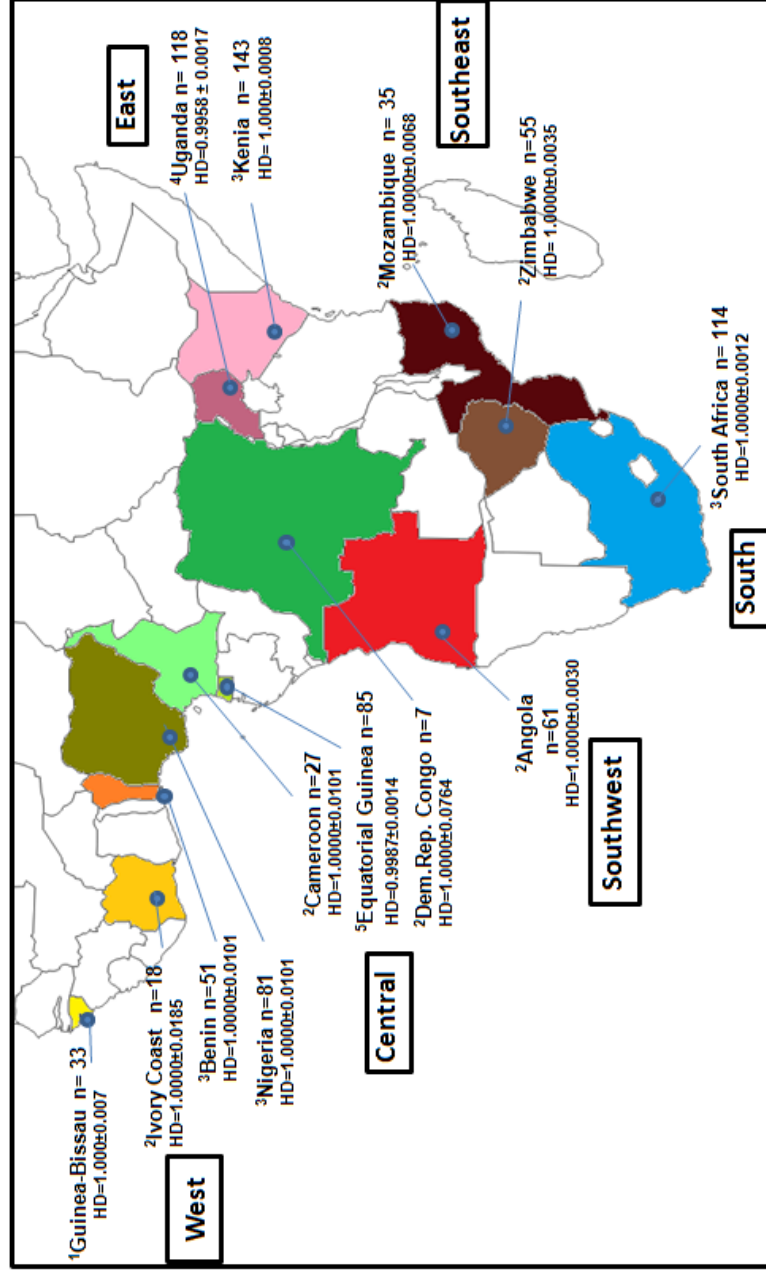


Figure 3 - Map of the distribution of the Y chromosome data (Y-Filer) that have been used for comparisons across African geographic regions and countries: distribution, population sample sizes (n) and haplotype diversities (HD). ¹ Carvalho *et al.*, 2011; ² Palha *et al.*, 2012; ³ Purps *et al.*, 2014; ⁴ Gomes *et al.*, 2010, ⁵ Gonzalez *et al.*, 2010.

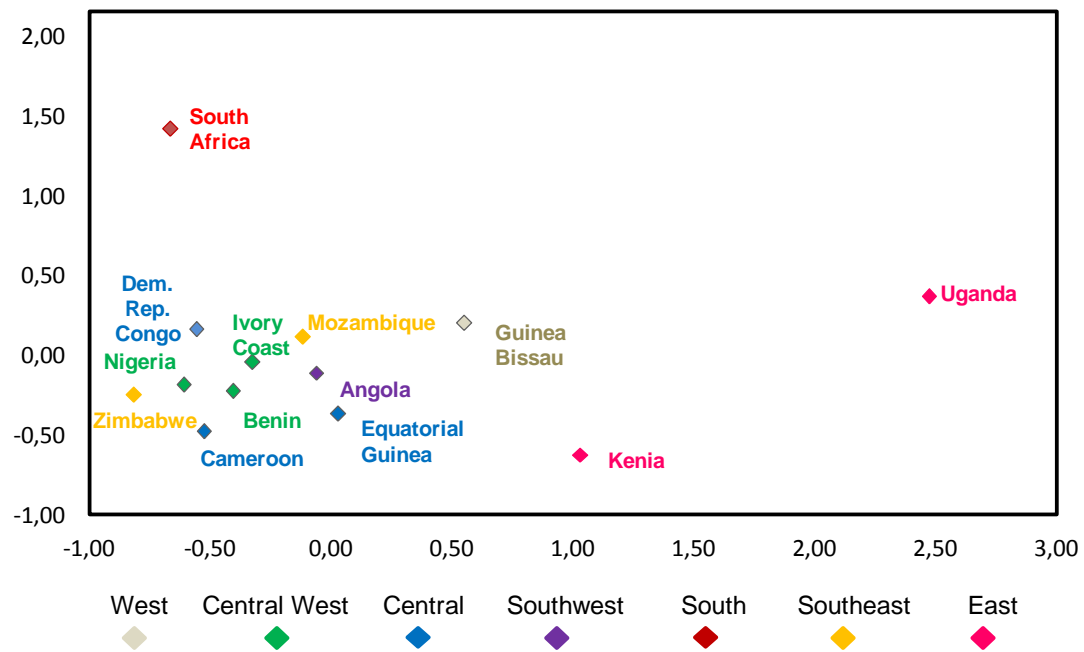


Figure 4 - Multidimensional scaling plot of the pairwise F_{st} genetic distances calculated among samples from different African countries. (*Stress* = 0,0755).

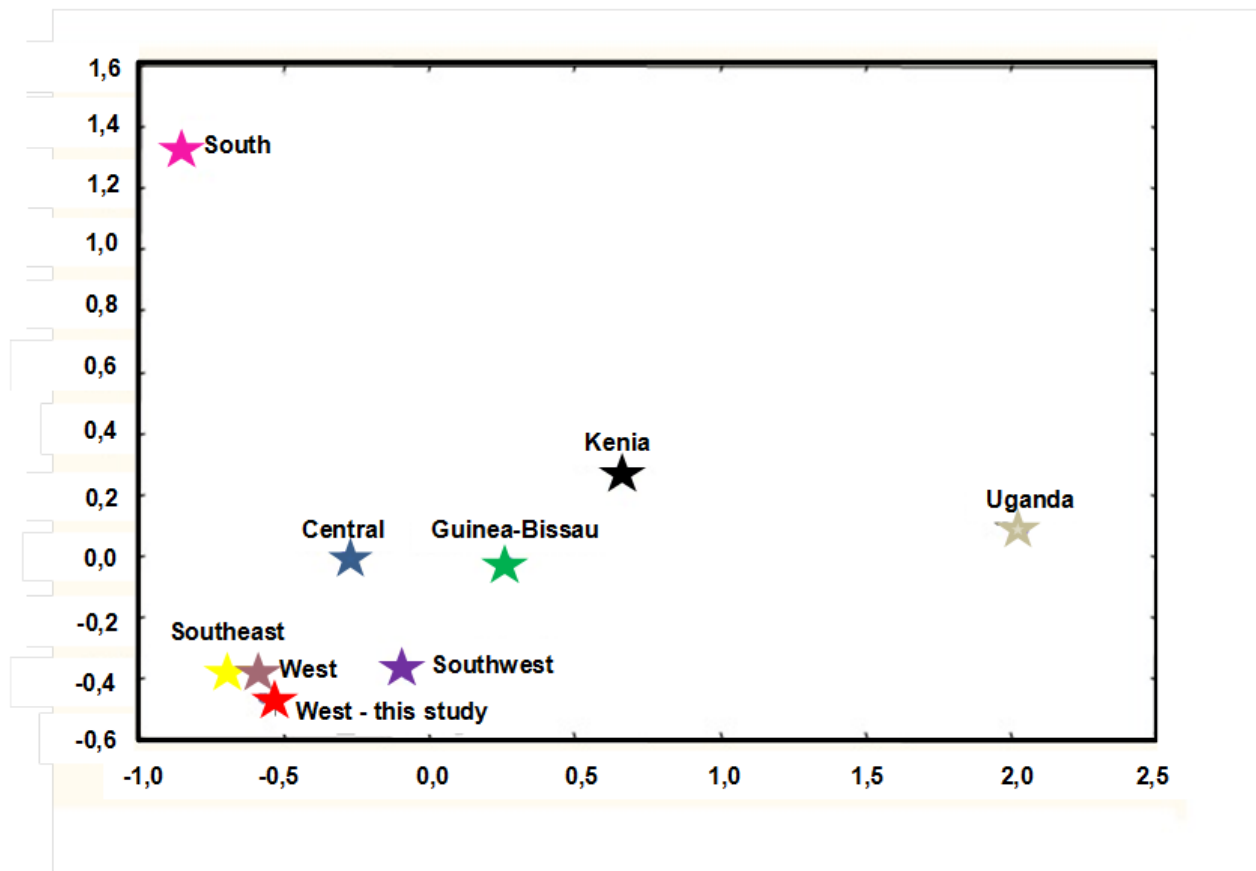


Figura 5 - Multidimensional scaling plot of the pairwise F_{st} genetic distances calculated between samples from different African countries and regions ($Stress = 0,0110921$). **Southeast**= Mozambique and Zimbabwe; **Central**= Equatorial Guinea, Cameroon, Democratic republic of Congo; **Southwest**= Angola; **South** = South Africa.

F _{st}	SOUTHERN		SOUTH	SOUTHWEST	CENTRAL		CENTRAL WEST		WEST	This study		
	Uganda	Kenia	South Africa	Angola	Equat. Guinea	Cameroon	D.R.Congo	Ivory Coast	Nigeria	Benin	Guinea Bissau	West Africa
Uganda	0.05909	< 5E-06	< 5E-06	< 5E-06	< 5E-06	< 5E-06	0.00040	< 5E-06	< 5E-06	< 5E-06	< 5E-06	< 5E-06
Kenia	0.08463	0.03692	< 5E-06	< 5E-06	< 5E-06	0.00010	0.01574	0.00366	< 5E-06	0.00040	< 5E-06	< 5E-06
Mozambique	0.14142	0.07207	0.00149	0.04089	0.00347	0.01416	0.72448	0.88061	0.08554	0.27324	0.06712	0.08108
Zimbabwe	0.14630	0.08314	< 5E-06	0.01683	0.00010	0.13880	0.49262	0.14771	0.06287	0.40907	< 5E-06	0.00901
South Africa	0.09140	0.04065	0.08219	< 5E-06	< 5E-06	0.00040	0.16543	0.01218	< 5E-06	< 5E-06	< 5E-06	< 5E-06
Angola	0.09140	0.04065	0.08219	*	0.02831	0.00921	0.46134	0.15197	0.04811	0.04855	< 5E-06	< 5E-06
Equat. Guinea	0.09552	0.04024	0.03486	0.01187	*	0.04237	0.21087	0.05851	0.04297	0.04485	< 5E-06	< 5E-06
Cameroon	0.11606	0.04647	0.00958	0.02775	0.01667	*	0.20226	0.17969	0.11345	0.10682	0.04455	0.19820
D.R.Congo	0.12480	0.06693	0.03239	0.00275	0.01525	0.02214	*	0.38600	0.53411	0.52698	0.13702	0.45946
Ivory Coast	0.09491	0.04100	0.04729	0.01123	0.01926	0.01207	0.00352	*	0.18553	0.49767	0.07643	0.26126
Nigeria	0.13007	0.06389	0.07494	0.02556	0.02080	0.00885	0.00535	0.00831	*	0.49767	0.00040	0.36495
Benin	0.11264	0.04999	0.00338	0.01053	0.00970	0.01111	0.00637	0.00181	0.00125	*	0.00168	0.61261
Guinea Bissau	0.06438	0.03351	0.06599	0.02775	0.01375	0.03848	0.02856	0.01989	0.03754	0.03094	*	0.61261
West Africa	0.13152	0.06078	0.07027	0.02783	0.01998	0.00557	0.00166	0.00204	0.00489	0.00123	0.03675	< 5E-06

Note: Significant value after applying Bonferroni's correction ($p < 0.00041$) are in bold.

References: ¹Carvalho et al., 2011; ²Palha et al., 2012; ³Purps et al., 2014; ⁴Gomes et al., 2010; ⁵Gonzalez et al., 2010.

Table 1 – Pairwise F_{st} values between African populations for the haplotypes defined by the Y-chromosome 16-set markers (below diagonal) and corresponding differentiation p values (above diagonal).

	Uganda	Kenia	Southeast	South Africa	Angola	Central	West	Guinea-Bissau	This study
Uganda	*	< 5-E6	< 5-E6	< 5-E6	< 5-E6	< 5-E6	< 5-E6	< 5-E6	< 5-E6
Kenia	0.05909	*	< 5-E6	< 5-E6	< 5-E6	< 5-E6	< 5-E6	< 5-E6	< 5-E6
Southeast	0.12118	0.05674	*	< 5-E6	0.01660	0.00098	0.44141	< 5-E6	0.03320
South Africa	0.14630	0.08314	0.06287	*	< 5-E6	< 5-E6	< 5-E6	< 5-E6	< 5-E6
Angola	0.09140	0.04065	0.01114	0.06863	*	0.02734	0.00195	0.00488	< 5-E6
Central	0.10342	0.04145	0.01706	0.05321	0.01071	*	0.00195	0.01074	0.00293
West	0.12708	0.05924	0.00010	0.06553	0.01915	0.01018	*	< 5-E6	0.89160
Guinea-Bissau	0.06438	0.03351	0.04674	0.06599	0.02775	0.01845	0.03584	*	< 5-E6
This study	0.13152	0.06078	0.00767	0.07027	0.02783	0.01103	0.00308	0.03675	*

Table 2 - Pairwise *Fst* values between populations from regions and countries for the haplotypes defined by the Y-chromosome set markers (below diagonal) and corresponding differentiation *p* values (above diagonal).

Note: Significant value after applying Bonferroni's correction ($p < 0.0011$) are in bold .

Legend: **Central** = Equatorial Guinea, Cameroon, Democratic Republic of Congo; **West** = Nigeria, Benin, Ivory Coast.

COUNTRY	HAPLOGROUP (%)																N																
	A1	A32-M13	B-M60	B2a-M150	B2a1a-M109	B2b-M112	C-RPS4y	DE-YAP	E-M96	E1a-M33	E1b1a1-M2	E1b1a1f-M191	E1b1b1-M35	E1b1b1a1-M78	E1b1b1b1-M81	E2-M75		E3-P2	F-M213	G-M201	J--122a	J2-M172	K-M9	M1c-Tat	P-92R7	R	R1a-SRY10831,2	R1b1-P25	R1b1a'-V88	R1b1b-M269	T-M70	Y*(X,A,B,DE',F')	
West Africa ^{This study}	1.16			2.32	4.65	55.8	32.56											2.50									1.16				1.16	1.16	86
Angola ^a			4.66		4.23		86.00																										236
Angola ^g			1.33	8.00			45.3	32.00													1.33												75
Botswana ^b	2.50		20.00				2.50	60.00	5.00																							40	
Burkina Faso ^b	0.60		5.70				0.60	8.10	35.5	46.00	2.10																					335	
Cameroon ^c				10.00					75.0																							88	
CAR ^b			22.6								52.00																					31	
DRC ^c			12.0							1.50	77.00																						69
Ethiopia ^b			1.00							31.0	1.00																						98
Equatorial Guinea ^d			1.20							1.20	3.52	29.41	41.20																				85
Gabon ^c	0.60		6.53							1.40	0.41	79.31																					735
Gabon+ Cameroon ^e	5.00		6.70							10.0	28.30																						60
Guinea Bissau ^f			2.70	0.39	0.39					0.39	6.60	1.93	77.22	3.00	0.39	4.63																	259
Kenia ^b	20.22		5.62								1.12	31.50	25.84																				89
Senegal ^b			4.37								13.33	73.33	6.70	6.70																			15
Zambia ^{b1}	0.18		2.70	3.38	0.67						33.78	0.67	3.38																				549
Khoinan			2.43	0.81	4.07						4.07	15.45																					148
Niger-Congo			2.32								2.32	9.30	18.60																				123
Tanzania ^d			3.33								20.00	13.33	26.67																				43
Nilo-saharan																																	30
Afro-Asiatic																																	

^a Coelho et al., 2009; ^b Filippo et al., 2011; ^c Berniell-Lee et al., 2009; ^d Gonzalez et al., 2010; ^e Carvalho et al., 2011; ^f Filippo et al., 2010; ^g Belezza et al., 2005; ^h Tishkoff et al., 2010

Table 3 - Relative frequencies of Y-chromosome haplogroups found in 86 unrelated male samples from Western African population, showing comparisons with 16 different populations from Africa.
** References used were cited in the table.*

Table S1 - List of the Y-STR haplotypes and Y-SNP haplogroups in a male population sample from Western Africa.

Sample name	Y-STR markers																Last mutation	Haplogroup	
	DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS439	DYS389I	DYS392	DYS389II	DYS458	DYS437	DYS448	GATAH4	DYS456	DYS438			DYS635
WA1	13	21	15	10	15	16	11	13	11	30	17	14	20	11	15	11	23	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA2	13	21	15	10	14	16	12	14	13	32	16	14	21	11	14	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA3	13	21	15	10	18	20	12	13	11	32	17	14	20	12	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA4	13	22	16	9	14	17	11	14	11	31	18	14	21	11	14	10	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA5	14	21	16	10	15	16	11	13	11	30	15	14	20	11	15	11	23	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA6	13	21	15	10	13	18	12	14	12	32	17	14	21	12	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA7	14	21	17	10	17	18	13	13	11	30	16	13	21	11	16	11	22	M191	E1b1a1f - M191
WA8	14	21	17	10	17	19	12	13	11	30	18	14	21	11	17	11	22	M191	E1b1a1f - M191
WA9	15	21	17	10	18	19	12	14	11	31	16	14	21	11	15	11	20	M191	E1b1a1f - M191
WA10	14	21	16	10	14	17	12	13	12	30	17	14	21	11	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA11	14	21	16	10	17	19	12	13	11	30	17	14	22	11	15	11	20	M191	E1b1a1f - M191
WA12	13	25	15	10	10	11	12	11	12	11	30	19	14	20	10	16	10	M150	B2a1*
WA13	14	21	17	10	15	16	10	14	11	31	17	14	20	11	15	11	23	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA14	13	21	16	10	15	16	11	13	11	29	16	14	21	11	14	11	22	M2	E1b1a1f - M191
WA15	15	21	15	11	17	17	12	13	11	29	17	13	21	11	16	11	21	M191	E1b1a1f - M191
WA16	14	21	17	10	17	18	13	13	11	30	16	13	21	11	16	11	22	M191	E1b1a1* (xE1b1a1f,g1c) - M2
WA17	13	21	15	10	16	18	13	14	11	32	15	14	21	11	14	11	22	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA18	13	21	16	10	17	17	12	14	11	31	19	14	21	11	19	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA19	13	21	16	10	16	17	12	13	11	30	16	14	21	12	15	11	21	M2	E1b1a1f - M191
WA20	14	21	15	10	17	19	11	13	11	30	14	14	21	11	15	11	21	M191	E1b1a1f - M191
WA21	13	21	15	10	15	17	12	13	11	31	17	14	21	11	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA22	11	22	17	10	13	15	11	12	12	30	15	17	19-20	11	16	10	20	M33	E1a-M33
WA23	14	21	15	11	15	20	12	13	11	31	17	14	21	11	14	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA24	13	21	15	10	15	20	11	13	11	31	17	14	21	12	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA25	14	21	16	10	16	16	12	13	11	30	17	14	21	11	15	11	21	M191	E1b1a1f - M191
WA26	13	21	10	10	17	17	12	13	11	31	16	14	21	12	16	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA27	14	21	17	10	17	19	13	13	11	30	16	14	20.2	11	15	11	21	M191	E1b1a1f - M191
WA28	13	21	15	10	17	17	12	13	11	31	16	14	21	12	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA29	13	21	15	10	16	18	13	13	10	31	17	14	21	11	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA30	14	21	17	10	17	19	12	13	11	32	16	14	21	11	15	11	21	M191	E1b1a1f - M191
WA31	14	21	17	10	17	18	12	13	11	30	16	14	20	11	16	11	21	M191	E1b1a1f - M191
WA32	15	21	16	10	16	18	13	13	12	30	17	14	21	11	16	11	21	M191	E1b1a1f - M191
WA33	13	21	15	10	14	16	12	14	13	32	16	14	21	11	14	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA34	13	21	15	10	17	18	11	13	12	30	17	14	21	12	14	11	22	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA35	15	21	17	10	17	18	12	13	11	30	16	14	21	11	14	11	21	M191	E1b1a1f - M191
WA36	14	21	16	10	16	18	11	13	11	29	19	14	20	13	15	11	22	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA37	14	21	15	10	15	17	13	13	11	30	18	14	20	11	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA38	15	20	17	10	16	19	13	13	11	30	16	14	21	11	16	11	21	M191	E1b1a1f - M191
WA39	13	23	17	10	14	15	10	12	11	29	16	16	20	11	15	8	21	M168	Y(XA,B,E,F) - M168
WA40	14	21	16	10	15	17	13	13	11	31	16	14	21	11	16	10	22	M191	E1b1a1f - M191
WA41	13	21	15	10	14	16	12	14	13	32	16	14	21	11	14	11	22	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA42	13	21	17	10	17	18	13	13	11	32	16	14	21	11	15	12	21	M191	E1b1a1f - M191
WA43	13	22	16	9	14	17	11	14	11	31	18	14	21	11	14	10	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA44	15	21	17	10	18	19	12	14	11	31	16	14	21	11	15	11	20	M191	E1b1a1f - M191
WA45	14	22	16	10	14	19	12	13	11	30	17	14	21	10	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA46	13	22	16	11	14	17	12	12	11	30	19	17	19-20	11	15	8	23	M33	E1a-M33
WA47	13	22	15	10	15	17	12	13	11	30	17	14	21	12	15	11	20	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA48	13	24	15	9	10	10	12	12	12	31	17	14	19	10	15	10	22	M150	B2a1*
WA49	12	21	15	10	17	17	13	13	11	31	18	14	21	10	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA50	14	20	14	12	18	18	12	14	11	31	16	14	15	10	12	11	17	Sem mutações A (x A3b2)	
WA51	14	21	15	10	14	17	11	13	11	30	17	14	20	11	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA52	13	21	15	10	15	16	11	13	11	30	17	14	20	11	15	11	23	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA53	13	21	16	10	17	17	12	13	11	31	16	14	21	12	16	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA54	13	23	15	10	12	14	11	12	14	28	18	14	20	11	15	9	21	M70	M70-T
WA55	12	21	15	10	16	17	12	13	11	31	16	14	21	11	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA56	14	21	16	10	16	17	11	12	11	30	19	14	22	11	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA57	13	21	15	10	17	17	13	14	11	31	17	14	21	12	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA58	14	21	17	10	17	19	12	13	11	30	18	14	21	11	17	11	22	M191	E1b1a1f - M191
WA59	14	21	17	10	13	15	13	13	11	31	18	14	21	11	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA60	14	21	15	10	15	18	12	14	11	31	17	14	21	12	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA61	15	21	17	10	16	18	11	13	11	29	15	14	21	11	16	11	21	M191	E1b1a1f - M191
WA62	13	24	14	11	11	15	11	13	13	29	17	14	18	11	18	12	23	P25	R1b1*-P25
WA63	15	21	16	10	16	17	12	13	11	30	18	14	20	11	16	11	22	M191	E1b1a1f - M191
WA64	15	22	16	10	18	18	13	13	11	31	14	14	21	12	15	11	22	M191	E1b1a1f - M191
WA65	13	21	14	11	15	17	12	13	11	30	17	14	20	11	16	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA66	14	21	15	10	16	17	12	13	11	31	15	14	21	11	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA67	14	21	16	10	16	16	13	13	11	30	17	14	21	11	15	11	23	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA68	14	21	16	10	15	18	12	13	11	30	16	14	20	11	15	11	21	M2	E1b1a1* (xE1b1a1f,g1c) - M2
WA69	13	22	15	9	16	16	13	12	12	30	15	16	19-20	12	16	10	21	M33	E1a-M33
WA70	14	21	16	10	16	16	12	13	11	30	17	14	21	11	15	11	21	M191	E1b1a1f - M191

(To be continued.)

Table S1 - List of the Y-STR haplotypes and Y-SNP haplogroups in a male population sample from Western Africa

Sample name	DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS439	DYS389I	DYS392	DYS389II	DYS458	DYS437	DYS448	GATAH4	DYS456	DYS438	DYS635	Last mutation	Haplogroup
WA71	13	20	15	10	16	17	12	13	11	30	16	14	20	11	16	11	21	M2	E1b1a1* (xE1b1a1f.g1c) - M2
WA72	13	21	15	10	14	16	12	11	11	28	18	14	21	12	15	11	22	M2	E1b1a1* (xE1b1a1f.g1c) - M2
WA73	15	21	17	11	17	19	11	13	11	29	15	14	20	11	16	11	21	M191	E1b1a1f - M191
WA74	13	21	15	10	16	16	12	14	11	32	15	14	21	11	14	11	22	M2	E1b1a1* (xE1b1a1f.g1c) - M2
WA75	13	21	15	10	16	17	12	14	11	32	19	14	22	12	15	11	22	M2	E1b1a1* (xE1b1a1f.g1c) - M2
WA76	14	21	15	10	16	16	11	13	11	31	19	14	20	11	15	11	21	M2	E1b1a1* (xE1b1a1f.g1c) - M2
WA77	14	21	15	10	17	19	11	13	11	30	14	14	21	11	15	11	21	M191	E1b1a1f - M191
WA78	12	22	17	10	13	15	11	12	12	30	15	17	19-20	11	16	10	20	M33	E1a-M33
WA79	14	21	15	10	16	17	12	13	11	30	16	14	20	12	15	11	22	M2	E1b1a1* (xE1b1a1f.g1c) - M2
WA80	14	21	16	10	15	17	13	13	11	31	16	14	21	11	15	10	22	M191	E1b1a1f - M191
WA81	13	21	16	10	16	16	13	11	11	29	16	14	21	11	15	11	21	M2	E1b1a1* (xE1b1a1f.g1c) - M2
WA82	15	21	16	11	16	17	12	13	11	31	18	14	21	11	16	11	21	M191	E1b1a1f - M191
WA83	15	21	17	10	16	17	12	14	11	31	17	13	21	11	16	11	21	M191	E1b1a1f - M191
WA84	14	21	15	10	17	18	10	13	11	31	19	14	22	11	15	11	21	M191	E1b1a1f - M191
WA85	12	21	15	10	16	17	12	13	11	31	16	14	21	12	15	11	21	M2	E1b1a1* (xE1b1a1f.g1c) - M2
WA86	13	21	16	11	16	18	13	13	11	31	18	14	21	11	15	11	21	M2	E1b1a1* (xE1b1a1f.g1c) - M2

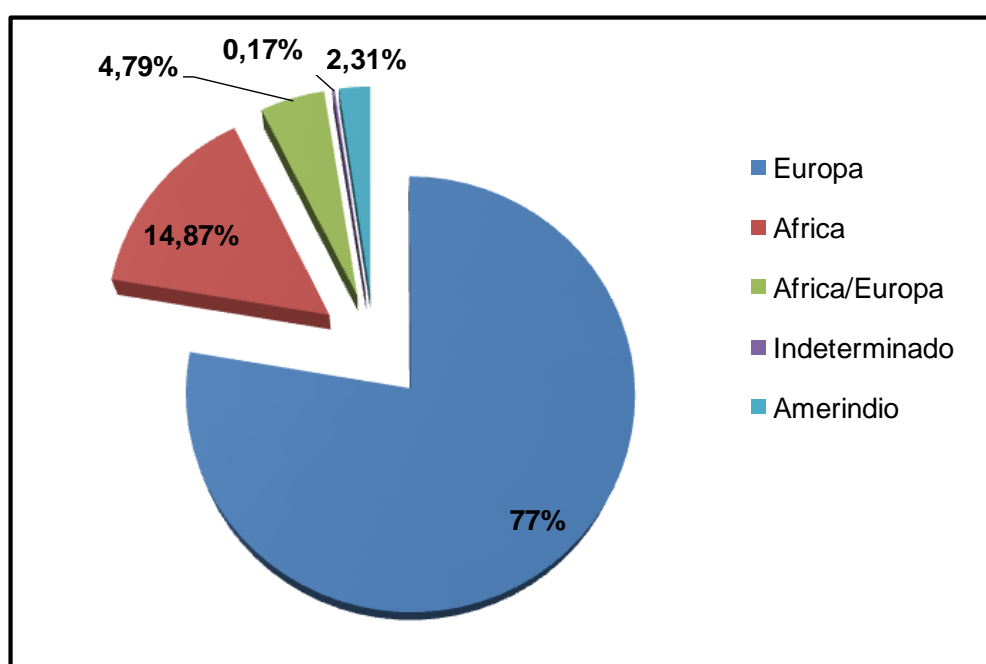
5.3 Análise da origem das linhagens africanas presentes na população do Rio de Janeiro

A terceira e última parte desta tese refere-se às análises comparativas entre o patrimônio genético comprovadamente africano da população do Rio de Janeiro, e do oeste africano, com outras populações da África subsaariana, que podem ter contribuído para a composição das populações afrobrasileiras atuais.

5.3.1 Haplogrupos africanos na população do Rio de Janeiro

A distribuição dos haplogrupos de cromossomo Y encontrados numa amostra de 605 indivíduos da população do Rio de Janeiro encontra-se descrita no Artigo I, e foi apresentada de forma esquemática na Figura 20.

Figura 20 - Distribuição de haplogrupos encontrados na amostra da população do Rio de Janeiro, separados por provável origem étnica



* Dados apresentados em frequência relativa

Observou-se claramente a maior influência europeia (77%) na população estudada, sendo seguida pela influência africana (14,87%) e de nativos americanos (2,31%), com uma pequena porcentagem de haplogrupos cuja origem não foi possível definir, representada pelos haplogrupos $Y^*(xA,B,DE^*,F^*)$ - M168 e E1b1b1a1 – M78 (0.17% e 4.79%, respectivamente). Estes resultados corroboraram os dados históricos e genéticos que conferem ao homem europeu a maior influência masculina na formação da população brasileira (PENA, 2000; NASCIMENTO *et al.*, 2009; PALHA *et al.*, 2012).

Em relação ao haplogrupo não determinado [$Y^*(xA,B,DE^*,F^*)$ - M168] e também a título de esclarecimento, esta amostra poderia ser considerada como pertencente ao haplogrupo C por simples exclusão após a utilização de todos os multiplexes descritos. Contudo, apesar da metodologia de tipagem seguida, as determinações inequívocas devem sempre ser feitas com a utilização dos oligonucleotídeos característicos para o haplogrupo a ser considerado. Deste modo, esta amostra não foi incluída no haplogrupo C pelo fato de não terem sido utilizados marcadores específicos para a classificação neste haplogrupo.

Para melhor avaliação dos dados encontrados, os mesmos foram comparados a outros dados originários de estudos efetuados em outras populações brasileiras miscigenadas. Nas amostras do Rio de Janeiro, um dos objetos deste estudo, foi encontrada uma elevada contribuição africana, quando comparada com outras populações miscigenadas já estudadas (Tabela 3) das regiões sudeste, norte e nordeste.

Indiscutivelmente, as contribuições africanas em populações miscigenadas na região sudeste são mais significativas, quando comparadas com as populações do sul e norte do país. Por outro lado, também pode ser visto na Tabela 3 que estudos realizados em populações que, historicamente, receberam uma grande influência de povos africanos na sua formação, tais como a população da Bahia, populações quilombolas e populações, cujos indivíduos se autodeclararam como negros, apresentaram uma alta porcentagem de haplogrupos tipicamente africanos, com valores de frequência muito superiores aos valores que encontramos no Rio de Janeiro.

Tabela 3 – Comparações entre as frequências dos principais haplogrupos encontrados entre as amostras do Rio de Janeiro e outras populações brasileiras

População	Região	Estado	Grupo Étnico	Autodelação	Marcadores	N	Europeu	Africano	Nativo	Asia	others	Referência
Urbana*	Sudeste	RJ	Admixed	Não declarado	STR+46 SNPs	605	77,8	14,8	2,3	-	5	Oliveira et al., 2014
Urbana	Sudeste	RJ	Admixed	Não declarado	STRs+ 28 SNPs	127	88,1	8,7	1,6	-	1,6	Silva et al., 2006
Urbana	Sudeste	MG	Admixed	Branco	STRs+SNPs	50	84	8	0	-	8	Carvalho-Silva et al., 2001/2006
Urbana	Norte	AM,AC,RO,PA	Admixed	Branco	STRs+SNPs	49	83,7	10,2	0	2	4,1	Carvalho-Silva et al., 2001/2006
Urbana	Nordeste	PE	Admixed	Branco	STRs+SNPs	49	89,8	4,1	0	-	6,1	Carvalho-Silva et al., 2001/2006
Urbana	Nordeste	BA	Admixed	Não declarado	6 SNPs	41	17,4	82,6	0	-	-	Wiezel et. al, 2013
Urbana	Nordeste	PI	Admixed	Não declarado	6 SNPs	76	79,5	20,5	0	-	-	Wiezel et. al, 2013
Urbana	Sul	RS_Porto Alegre	Admixed	Branco	11 SNPs	203	99	1	0	-	-	Guerreiro-Junior et al., 2009
Urbana	Sul	RS_Santa Catarina	Admixed	Branco	STRs+SNPs	52	78,8	3,8	0	-	17,4	Carvalho-Silva et al., 2001/2006
Urbana	Sudeste	RJ	Admixed	Negros	30 SNPs	94	44	56	0	-	-	Hunemeier, 2007
Urbana	Sul	RS_Porto Alegre	Admixed	Negros	11 SNPs	54	56	38	6	-	-	Guerreiro-Junior et al., 2009
Urbana	Sul	RS	Admixed	Negros	30 SNPs	107	59	36	5	-	-	Hunemeier, 2007
Nativa	Central e Sul	RS,MS,PR	Nativo	guarani	7 SNPs	200	9	14	77	-	-	Marrero,2007
Nativa	Central e Sul	RS,MS,PR	Nativo	kaingang	7SNPs	78	15	16	69	-	-	Marrero,2007
Afrodscendente	Nordeste	PI	quilombola		6 SNPs	31	47,1	52,9	0	-	-	Wiezel et. al, 2013
Afrodscendente	Nordeste	PI	quilombola		6 SNPs	23	39,8	60,2	0	-	-	Wiezel et. al, 2013
Afrodscendente	Nordeste	PI	quilombola		6 SNPs	7	31	59,8	9,2	-	-	Wiezel et. al, 2013
Afrodscendente	Nordeste	PI	quilombola		6 SNPs	17	40,3	59,7	0	-	-	Wiezel et. al, 2013

Nota: * Dados deste estudo.

É digno de nota que, estudos realizados por Wiezel *et al.* (2013), em populações do nordeste, apresentaram altas frequências de marcadores que caracterizam a influência africana em populações do Piauí (20,5%) e Bahia (82,6%). Contudo, este evento pode ser explicado pelo fato de que um baixo número de Y-SNPs foi analisado. Em vista disto, sugere-se a possibilidade de que tenha havido uma baixa resolução na detecção das frequências dos haplogrupos, ou também podemos especular que, pelos fatos históricos da colonização, no caso específico da população baiana, realmente tenha ocorrido um grande influxo africano, o que justificaria esta alta frequência. Portanto, estes dados devem ser questionados pelo baixo número de marcadores utilizados, o que pode favorecer a produção de dados com grandes desvios.

5.3.2 Comparações das frequências dos haplogrupos africanos no Rio de Janeiro com populações da África

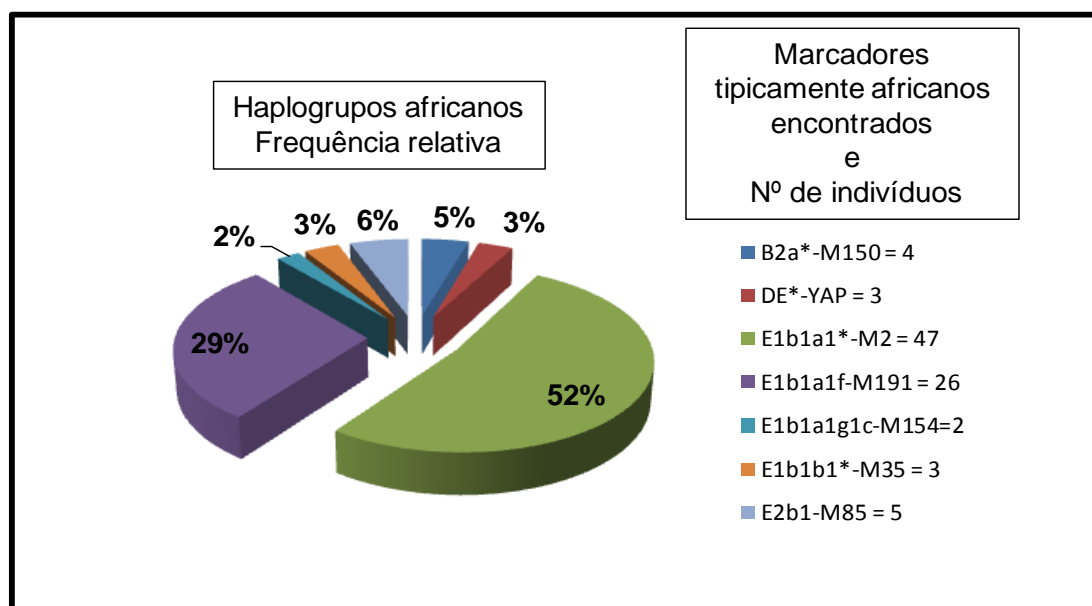
Como estratégia para relacionar as linhagens do cromossomo Y encontradas no Rio de Janeiro com os possíveis locais de origem na África, a amostra dos cromossomos referentes somente ao aporte africano encontrado na nossa amostra foi isolada, e as frequências relativas foram calculadas e comparadas com as encontradas em populações africanas que poderão ter contribuído para o *gene pool* brasileiro.

No total, dentre os 605 indivíduos estudados, foram encontrados 90 (14,88%) com haplogrupos de origem tipicamente africana (Figura 21).

Para os marcadores africanos, observamos que 47 indivíduos (52%) foram identificados como haplogrupo E1b1a1* - M2 (portadores da mutação M2). Este haplogrupo que foi encontrado com maior frequência na amostra do Rio de Janeiro, também foi observado de forma majoritária em populações da África subsaariana, exceto no norte, com maior frequência nas regiões do oeste e centro africanas. É associado com a expansão da agricultura pelos povos de língua Bantu subsaarianos e das regiões da África Equatorial, onde é prevalente (GONZALEZ *et al.*, 2013). Nos indivíduos afro-americanos (EUA) este é o haplogrupo mais comum,

como resultado do tráfico negreiro da África subsaariana em direção à América do Norte (HAMMER *et al.*, 2006).

Figura 21 - Distribuição de haplogrupos africanos encontrados na amostra da população do Rio de Janeiro (n=90 indivíduos) e apresentados na forma de frequência relativa (%) e absoluta (nº de indivíduos)



O haplogrupo E1b1a1f –M191, também com elevada frequência na nossa amostra, foi observado em 26 indivíduos (28,8%). Este haplogrupo também é encontrado de forma muito frequente em populações da África Central (WOOD *et al.*, 2005). Historicamente, esta região também foi muito importante para a formação do povo brasileiro, visto que, possuía portos importantes de saída de escravos, principalmente das regiões de Angola e do Congo que seriam destinados à região sudeste do Brasil, fato este bastante documentado (OGOT, 2000; HEYWOOD, 2008).

Os outros haplogrupos representam grupos com origens diversas, tais como, o haplogrupo B2a*- M150, característico de populações subsaarianas com alta frequência entre os Pigmeus e os Khoisan, sendo bem restrita a estes grupos étnicos (JOBILING & TYLER-SMITH, 2003). Outro exemplo a ser citado é o haplogrupo E1b1a1g1c – M154 que tem sido encontrado com frequências de

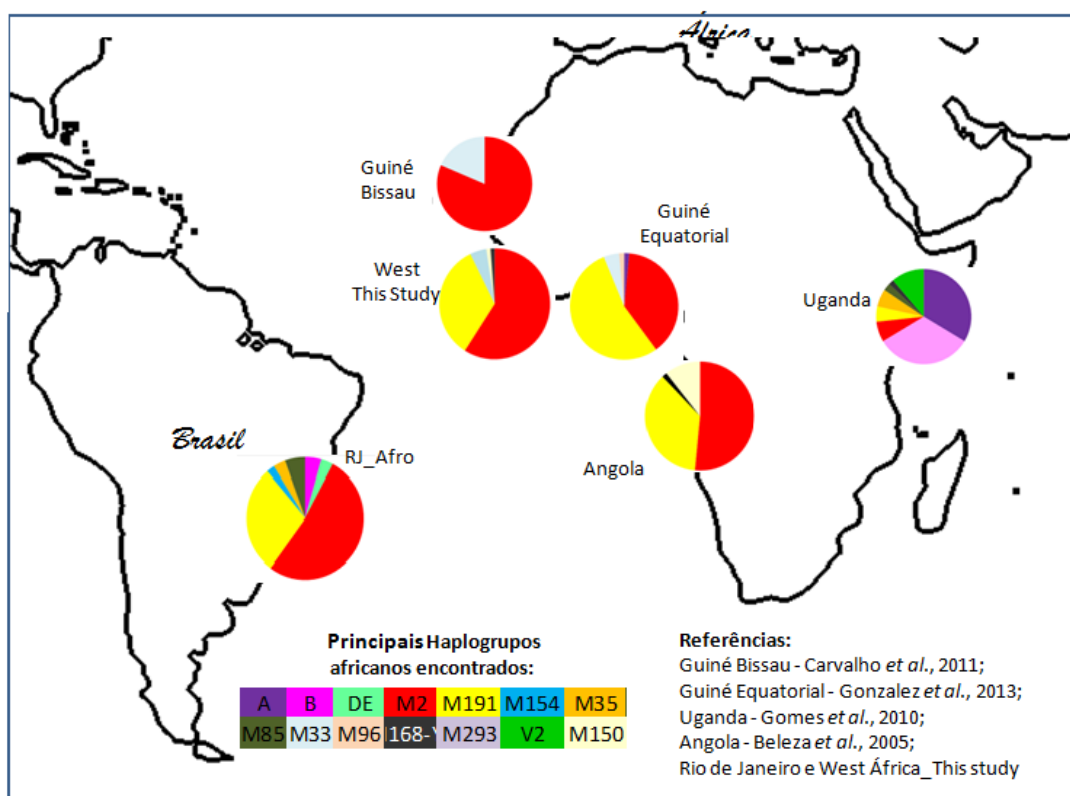
aproximadamente 9% em Camarões e de 4% em algumas regiões da África do Sul. Sugere-se que esta distribuição provavelmente reflete a dispersão por fazendeiros Bantu na rota de África Central e Ocidental da África do Sul.

E finalmente, o subclado E1b1b1* - M35 observado com uma frequência de 3% na nossa amostra e que pode ser encontrado tanto na região oriental quanto ocidental da África, mas que apresenta uma maior frequência no leste africano (KARAFET *et al.*, 2008; TROMBETTA *et al.*, 2011).

Existem poucos estudos realizados com Y-SNPs que sejam representativos de regiões específicas da África, mas as comparações com os dados provenientes de estudos realizados em diferentes países, que estão disponíveis na literatura, permitiram avaliar as particularidades de cada região.

Na Figura 22, é interessante observar que, as frequências de haplogrupos detectadas na amostra do Rio de Janeiro foram muito semelhantes às aquelas observadas em amostras das regiões oeste e central da África.

Figura 22 - Mapa com as frequências dos principais haplogrupos encontrados nas populações do Rio de Janeiro e da África, incluindo a amostra do oeste africano (West_This study), também genotipada neste trabalho



Os países da África que notoriamente apresentaram diferenças nas frequências, quando comparadas com nossas amostras, estão situados nas regiões do extremo oeste (Guiné-Bissau) e do leste (Uganda). A Guiné-Bissau e o Senegal, que no passado formavam a região que foi denominada de Alta-Guiné, durante o tráfico negreiro, foram países importantes que serviram como rota de envio de escravos entre os anos de 1441-1521, período anterior ao influxo negreiro destinado ao Brasil (OGOT, 2000). De fato, neste período que foi anterior ao tráfico transatlântico, os indivíduos provenientes desta região foram inicialmente enviados para Portugal e para as ilhas Atlânticas. Nos anos posteriores, iniciou-se o envio de pessoas desta região para a América Central e Antilhas, sendo que poucos foram direcionados à América do Sul e ao Brasil (OGOT, 2000; SILVA, 2008).

Quanto ao leste africano (e.g. Uganda e Quênia), não existem dados referentes ao envio de indivíduos escravizados desta região para o Brasil, ou para qualquer região das Américas. A ausência do tráfico português nestas regiões justificaria as grandes diferenças, em relação às nossas amostras, nas frequências dos haplogrupos A e B.

Por outro lado, temos somente alguns relatos históricos de algum aporte proveniente do sudeste africano, que estaria relacionado ao tráfico negreiro de indivíduos de Moçambique. Dados apontam que este país foi a última colônia africana conquistada pelos portugueses, de onde foram enviados menores contingentes de escravos ao Brasil, já no final do período da escravatura, mas a falta de dados genéticos disponíveis na literatura não nos permitiu fazer comparações em nível de Y-SNPs com esta região especificamente. Algumas inferências podem ser feitas a partir de dados haplotípicos, conforme já descrito no artigo número 2 e nos dados a seguir, onde encontramos similaridades entre as regiões do oeste e do sudoeste da África, países que sofreram grande influência da expansão dos grupos Bantu.

5.3.3 Comparações das frequências dos haplótipos dos grupos africanos no Rio de Janeiro com populações da África

Estudos prévios mostraram que para populações africanas, que são sujeitas a fortes efeitos de deriva genética, as medidas de distâncias genéticas são mais corretamente expressas pela análise do F_{ST} do que por R_{ST} (PÉREZ-LEZAUN *et al.*, 1997). A estimativa do parâmetro F_{ST} baseia-se na contagem do número de alelos diferentes entre dois haplótipos, ou seja, consideram-se somente as diferenças sem levar em conta os passos mutacionais necessários para cada possível alelo, enquanto que a estimativa do R_{ST} é calculada pela soma dos quadrados da diferença no número de repetições entre dois haplótipos (SLATKIN, 1995; MICHALAKIS & EXCOFFIER, 1996).

No presente estudo, as distâncias genéticas entre pares de populações foram calculadas utilizando-se a análise F_{ST} que se baseou no número de diferentes alelos para os haplótipos do cromossomo Y. Esta estimativa é uma variância padronizada da frequência de haplótipos e assume a deriva genética como sendo o agente que diferencia as populações (CAGLIÀ *et al.*, 2003). As análises de distância genética por F_{ST} são mais substanciais para a comparação de afinidades entre populações, pois a diversidade populacional apresentada por análises de *Rst* não se ajustam ao conhecimento antropológico e etnohistórico (CAGLIÀ *et al.*, 2003).

Devido a estas características de análise, onde se faz necessário o estudo de diferenças entre os haplótipos de Y e não entre os passos mutacionais de cada alelo, optou-se por fazer as comparações de distância genética tendo em conta os países e posteriormente os agrupamentos por regiões geográficas.

5.3.3.1 Comparações entre a amostra do Rio de Janeiro e países da África

As amostras utilizadas para a comparação com a amostra do Rio de Janeiro e nossas amostras do oeste africano são originárias dos seguintes países: Uganda (GOMES *et al.*, 2010), Guiné Equatorial (GONZALEZ *et al.*, 2013), Guiné Bissau (CARVALHO *et al.*, 2011), Moçambique, Costa do Marfim, República Democrática do Congo, Camarões e Angola (PALHA *et al.*, 2012), Nigéria, Benin, África do Sul e Zimbabwe (PURPS *et al.*, 2014), que foram todas genotipadas para 17 marcadores

Y-STRs (Y-Filer). As comparações entre populações foram feitas após a exclusão do DYS385 do conjunto de marcadores.

Com o objetivo de estudar o comportamento de cada grupo étnico, foram efetuadas análises de distância genética com o componente total da população do Rio de Janeiro (RJ_Total) e também com os componentes considerados africano (RJ_Afro) e europeu (RJ_Europa). Para a determinação e a consequente separação destes componentes, foram utilizados os resultados obtidos a partir da análise dos Y-SNPs, comparando-os com os dados relativos às distribuições em diferentes regiões do globo e em populações específicas. A separação destes componentes pode ser ilustrada a partir da observação das frequências já apontadas no Gráfico 19. Deste modo, os componentes foram determinados e as amostras foram separadas em grupos representativos associados aos principais grupos étnicos que apresentaram a maior frequência. Para isto também retiramos aqueles Y-SNPs que representam os nativo-americanos, os europeus e o que pode ser encontrado tanto na África quanto na Europa (e.g. M78 - E1b1b1a1).

A análise das distâncias genéticas foi efetuada para identificar o comportamento dos diferentes componentes frente à população total e aos dados referentes às amostras africanas também genotipadas neste estudo e de dados também disponíveis na literatura.

Os resultados obtidos a partir da análise de distâncias genéticas (F_{st}) entre os pares de populações podem ser observados na Tabela 4. Foi observado que as populações RJ_Total e RJ_Europa apresentaram diferenças significativas quando comparadas com todas as populações africanas utilizadas nas comparações.

Conforme já descrito em estudo anterior por Silva *et al.* (2006), não foi observada nenhuma diferença significativa entre uma amostra da população miscigenada do Rio de Janeiro, quando comparada com uma amostra da população portuguesa. Em vista disto, como seria esperado, o resultado encontrado na Tabela 4, indica a existência de grande semelhança entre a população total do Rio de Janeiro e a fração europeia extraída da amostra total do Rio de Janeiro. Devido a este fato, a amostra do Rio de Janeiro representativa dos haplogrupos europeus e a população total do Rio de Janeiro, encontraram-se geneticamente distantes de todas as populações africanas. Este mesmo resultado também foi observado nas populações de Uganda, Quênia e África do Sul em relação às demais, ou seja, são geneticamente distantes de todas as populações usadas nas comparações.

Quando o componente africano da nossa população foi destacado do total da amostra, não foram observadas diferenças significativas com outras populações, exceto para as populações da África do Sul, Uganda e Quênia.

Na nossa amostra, referente ao oeste africano (West_this study) também não apresentou diferenças significativas em relação às amostras do Benin, Moçambique, Camarões, Costa do Marfim e Nigéria sendo muito próximas entre si. Para melhor observação da posição relativa das populações no espaço genético em duas dimensões, foi feito o escalonamento multidimensional (MDS), que pode ser visto na Figura 23.

A análise do escalonamento multidimensional (MDS) permite visualizar graficamente, em duas dimensões, os resultados obtidos através da análise das distâncias genéticas (F_{st}). O nível de Stress deve ser menor do que 1 e o mais próximo de zero possível. Este nível de Stress indica o quanto os dados de distância genética foram adaptados/moldados de forma a permitir a melhor visualização espacial fornecida pelos valores de distância, com a menor “deformação” entre as análises.

Os resultados permitiram observar um agrupamento nítido formado pelas populações do oeste e centro-oeste: Costa do Marfim, Benin, Nigéria, Rio de Janeiro com aporte africano (RJ-Afro) e as amostras do oeste africano (West_This study).

Figura 23 - Representação gráfica das distâncias genéticas, entre as amostras estudadas e países da África, através da análise de escalonamento multidimensional-*Multidimensional Scalling Plot (MDS)*

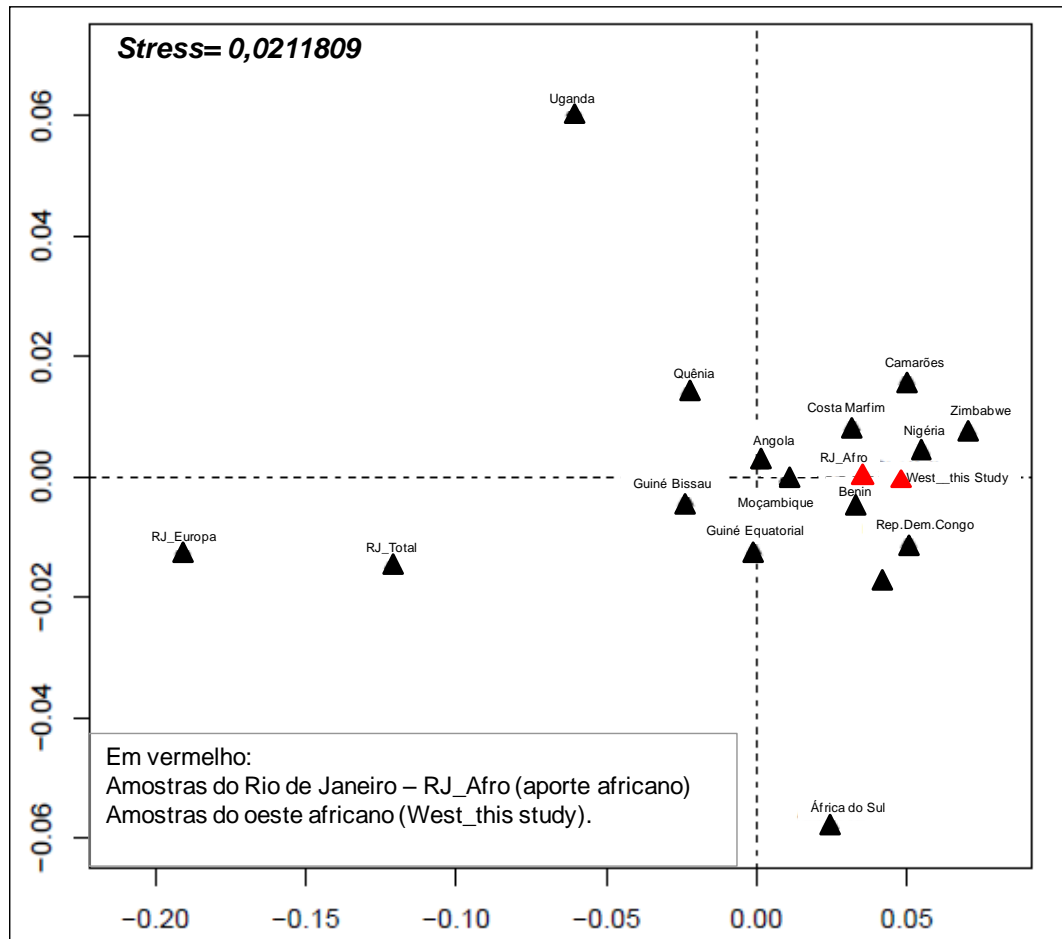


Tabela 4 - Matriz das distâncias genéticas (*Fst*) para haplótipos entre pares de populações. Abaixo da diagonal os valores de *Fst* e acima da diagonal os valores de ρ .

<i>Fst</i>	Países utilizados, para as comparações e número de indivíduos (N)																		
	Uganda N=118	Guiné Equatorial N=85	Guiné Bissau N=33	Moçambique N=35	Camargões N=27	Costa do Marfim N=18	Rep. Dem. Congo N=7	Angola N=61	Nigéria N=81	Quênia N=143	Benin N=51	África do Sul N=114	Zimbabwe N=55	West-this study N=86	RJ_total N=605	RJ_Afro N=90	RJ_Europa N=471		
Uganda	*																		
Guiné Equatorial	0.09552	* 0.04435																	
Guiné Bissau	0.06438	0.01375	* 0.06079																
Moçambique	0.08463	0.02462	0.01399	*															
Camargões	0.11606	0.01667	0.03848	0.02625	*														
Costa do Marfim	0.09491	0.01926	0.01999	-0.01324	0.01207	*													
Rep. Dem. Congo	0.12480	0.01525	0.02856	-0.01577	0.02214	0.00352	*												
Angola	0.09140	0.01187	0.02775	0.01364	0.02775	0.01123	0.00275	*											
Nigéria	0.13007	0.02080	0.03754	0.00833	0.00885	0.00831	0.00535	0.02556	*										
Quênia	0.05909	0.04024	0.03351	0.03692	0.04647	0.04100	0.06693	0.04065	0.06389	*									
Benin	0.11264	0.00970	0.03094	0.00338	0.01111	0.00181	0.00637	0.01053	0.00125	0.04999	*								
África do Sul	0.14630	0.05582	0.06599	0.04533	0.07368	0.04729	0.03239	0.06863	0.07494	0.08314	0.05954	*							
Zimbabwe	0.14142	0.03486	0.07174	0.02045	0.00958	0.01150	0.00463	0.01609	0.00791	0.07207	0.00045	0.08219	*						
West-this study	0.13152	0.01998	0.03675	0.00930	0.00557	0.00204	0.00166	0.02783	0.00489	0.06078	0.00129	0.07027	0.01432	*					
RJ_total	0.09480	0.11186	0.08461	0.12518	0.17038	0.14880	0.16625	0.12347	0.17378	0.09471	0.14848	0.15116	0.18742	0.17167	*				
RJ_Afro	0.11091	0.02448	0.03432	0.00721	0.02366	0.00348	0.01388	0.00059	0.01403	0.05048	0.00721	0.05453	0.00838	0.01924	0.15304	*			
RJ_Europa	0.13936	0.17289	0.14104	0.19117	0.24040	0.22055	0.23984	0.18909	0.24370	0.15330	0.21565	0.21458	0.25907	0.24117	0.00812	0.22317	*		

Correção de Bonferroni: $p \leq 0,00032$.

Referências utilizadas: Uganda – Gomes *et al.*, 2010; Guiné Equatorial - Gonzalez *et al.*, 2013; Guiné Bissau - Carvalho *et al.*, 2011; Moçambique, Costa do Marfim, República Democrática do Congo, Camargões e Angola - Palha *et al.*, 2012, Nigéria, Benin, África do Sul e Zimbabwe – Purps *et al.*, 2014

5.3.3.2 Comparações entre regiões da África

Após a análise comparativa entre as populações de diferentes países, foram efetuados agrupamentos, determinados segundo seus valores de distância genética, suas localizações geográficas no continente africano e também através de informações históricas sobre a sua composição étnica. Em vista destes elementos considerados, os diferentes países foram agrupados em 7 regiões distintas: Leste (East) = Uganda, Sudeste (Southeast) = Moçambique, Sudoeste (South West) = Angola, Sul (South) = África do Sul, Central (Central) = Camarões, Guiné Equatorial e República Democrática do Congo, Centro-Oeste (West Central) = Costa do Marfim e Oeste (West) = Guiné Bissau para a comparação com as amostras do Rio de Janeiro, com o objetivo de observar as diferenças regionais (Tabela 5).

Como já descrito, da população do Rio de Janeiro foram isolados dois componentes ancestrais, em função dos haplogrupos encontrados: RJ_Afro e RJ Europa. Estes dois componentes foram comparados com as 7 regiões distintas determinadas para esta análise, assim como com a população do oeste africano (West_This study), também parte desta tese. O resultado da análise de distâncias genéticas (F_{st}) entre os pares de populações pode ser observado na Tabela 5.

Tabela 5 - Matriz das distâncias genéticas (F_{st}) para haplótipos entre pares de populações agrupadas por região. Abaixo da diagonal os valores de F_{st} e acima da diagonal os valores de p

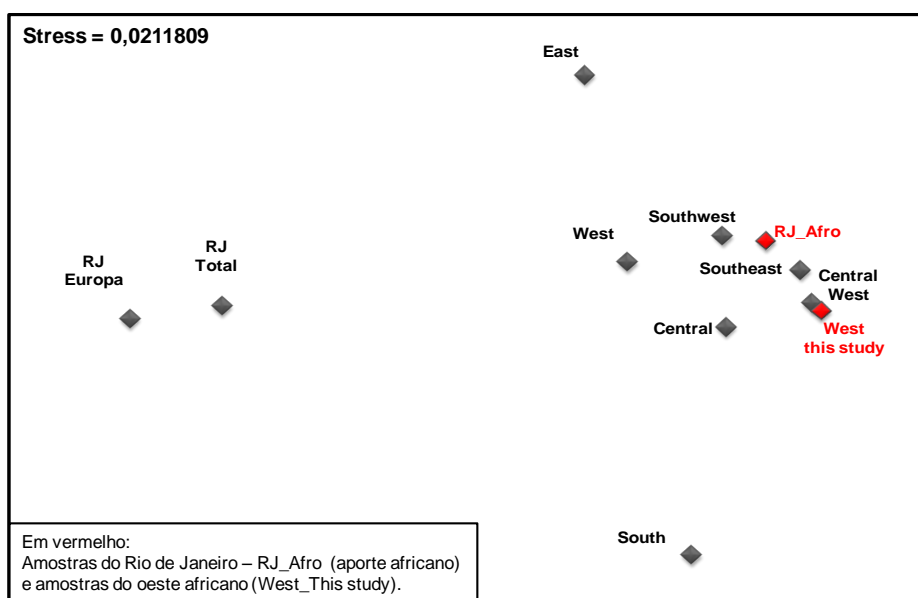
	East	Southeast	South	Southwest	Central	Central West	West	West_This Study	RJ_Total	RJ_Afro	RJ_Europa
East	*	5E-6	5E-6	5E-6	5E-6	5E-6	0.00010	5E-6	5E-6	5E-6	5E-6
Southeast	0.06856	*	5E-6	0.01673	0.00020	0.45689	5E-6	0.02871	5E-6	0.15226	5E-6
South	0.09347	0.06287	*	5E-6	5E-6	5E-6	0.00030	5E-6	5E-6	5E-6	5E-6
Southwest	0.04787	0.01114	0.06863	*	0.02366	0.00010	0.00535	5E-6	5E-6	0.48451	5E-6
Central	0.05412	0.01706	0.05321	0.01071	*	0.00079	0.01297	0.00436	5E-6	0.00040	5E-6
Central West	0.07268	0.00010	0.06553	0.01915	0.01018	*	5E-6	0.89833	5E-6	0.00406	5E-6
West	0.03267	0.04674	0.06599	0.02775	0.01845	0.03584	*	5E-6	5E-6	0.00030	5E-6
West_This Study	0.07621	0.00767	0.07027	0.02783	0.01103	0.00308	0.03675	*	5E-6	0.00030	5E-6
RJ_Total	0.08239	0.16449	0.15116	0.12347	0.12945	0.16693	0.08461	0.17167	*	0.00000	5E-6
RJ_Afro	0.06104	0.00333	0.05453	0.00059	0.01762	0.01022	0.03432	0.01924	0.15304	*	5E-6
RJ_Europa	0.13448	0.23484	0.21458	0.18909	0.19394	0.23696	0.14104	0.24117	0.00812	0.22317	*

Nota: Correção de Bonferroni: $p \leq 0.00076$.

Nas comparações, RJ_Total apresentou-se geneticamente diferente de todas as regiões usadas para a comparação. Contudo, a parcela desta população que corresponde ao componente africano mostrou maior proximidade com as populações do sudoeste africano (SouthWest), representado mais especificamente por Angola, assim como com a região sudeste (Southeast), representada por Moçambique.

Para melhor observação da posição relativa das populações no espaço genético em duas dimensões, foi feito o escalonamento multidimensional (MDS), que pode ser visto na Figura 24.

Figura 24 - Representação gráfica das distâncias genéticas, através da análise do escalonamento multidimensional - *Multidimensional Scalling Plot* (MDS)



Os resultados permitem observar a grande proximidade espacial entre a população do Rio de Janeiro com aporte africano e as regiões do Sudoeste (Angola) e Sudeste africano (Moçambique). Deste modo, podemos considerar que a fração da população do Rio de Janeiro com aporte comprovadamente africano (RJ_Afro) não apresenta diferenças genéticas em relação às regiões sudoeste e sudeste da África. Por outro lado, as amostras da população do oeste africano, também estudadas, são geneticamente mais próximas da região central da África, podendo

de maneira análoga, também serem consideradas como geneticamente semelhantes.

As outras populações se apresentaram mais dispersas, com grandes diferenças genéticas entre o Leste (East), o Sul (South) e outro agrupamento representado por RJ_Total e RJ_Europa.

5.4 Discussão final dos resultados obtidos

Em vista destes resultados, pode ser postulado que a colonização brasileira em muitas regiões foi o resultado de intensa miscigenação entre as populações de nativo-americanos que aqui existiam, mas que não deixaram uma grande herança por via paterna e, outros indivíduos representados em sua grande maioria pelos portugueses e africanos. Além disto, podemos considerar também uma pequena contribuição de outros povos de diferentes regiões do globo. Isto pode ser bem evidenciado pelo fato de uma elevada heterogeneidade genética ser encontrada na população brasileira em diversas regiões.

Em resumo, no Brasil existem registros históricos que nos permitem entender sobre como ocorreu a nossa colonização e a formação do povo brasileiro, mas existem algumas lacunas acerca da origem dos africanos que aqui aportaram, durante o período da diáspora africana em direção às Américas. Muito já foi publicado, contudo, os focos de interesse geralmente baseiam-se na determinação da contribuição de cada grupo continental (europeus, africanos, nativo-americanos) para a formação da nossa população.

O conjunto de oligonucleotídeos para Y-STRs e Y-SNPs utilizados permitiu a determinação de um painel significativo da população do Rio de Janeiro, com uma avaliação mais acurada dos principais marcadores de ancestralidade que poderiam nos permitir uma determinação mais aproximada da história da nossa origem paterna. Portanto, a partir destes estudos poderemos continuar o trabalho, na tentativa de refinar a nossa análise construindo uma base de dados que congreguem os STRs e SNPs mais significativos e informativos para a nossa população em particular. Além disto, novas amostras populacionais brasileiras poderão ser incorporadas, incrementando o nosso conhecimento sobre a formação

do povo e como ocorreram as migrações, além de detectar as variações intrapopulacionais e as suas diferenças regionais.

A primeira parte deste estudo relacionou-se à análise cuidadosa de 605 indivíduos do sexo masculino do estado do Rio de Janeiro, utilizando 17 Y-STRs e 46 Y-SNPs. Como resultado, não foram observadas diferenças significativas entre as populações do sudeste e os indivíduos do Rio de Janeiro, que apresentou uma alta diversidade haplotípica ($0,9999 \pm 0,0001$). Contudo, grandes diferenças foram observadas entre as regiões sudeste e norte do país (Roraima e Pará), com a utilização de um alto número de Y-STRs e Y-SNPs. O aumento do número amostral também foi determinante para encontrar estas diferenças. Deste modo, verificou-se que os altos níveis de diversidade intrapopulacional estão associados com haplótipos que são definidos por muitos marcadores do tipo STR. Isto significa que o aumento do número de marcadores nas análises populacionais permite detectar a ocorrência de subestruturação populacional.

A grande contribuição europeia no conteúdo gênico do cromossomo Y em muitas populações brasileiras explica a grande dificuldade de detectar diferenças populacionais significativas, quando o número amostral é reduzido para haplótipos extremamente reduzidos e associados com um grande número de Y-STRs utilizados. Contudo, as diferenças entre os três componentes formadores da nossa população (nativo-americanos, europeus e africanos) podem ser facilmente detectadas quando Y-SNPs são associados nestas análises populacionais, apontando diferenças nas frequências haplotípicas em grandes bases de dados.

A população total do Rio de Janeiro quando comparada com sua fração europeia, de um modo geral, não apresentou diferenças significativas, quando analisamos as suas distâncias genéticas (F_{st}). Estes resultados sugerem que, no que diz respeito às linhagens masculinas, o processo de formação da população do Rio de Janeiro, assim como de outros estados brasileiros, foi muito semelhante apresentando uma forte herança portuguesa. Contudo, em diferentes regiões, outras influências se apresentaram a partir do primeiro influxo de escravos no período colonial ou até mesmo a partir de correntes migratórias internas posteriores. Sem dúvida, a entrada de novos imigrantes, após a abertura dos portos, favoreceu esta grande variabilidade em nossa população. A maior proximidade genética da população do Rio de Janeiro com populações da região sudeste pode ser explicada pelo fato de que, após a entrada inicial pelos portos da Bahia e do Rio de Janeiro,

houve uma distribuição preferencial da população no sentido de todo o litoral e posteriormente em direção ao interior do continente.

Os portugueses foram os principais colonizadores europeus a introduzirem suas heranças genéticas à nossa população, inicialmente com as mulheres indígenas, mas a posterior entrada maciça de escravos africanos permitiu a miscigenação a partir do cruzamento destes primeiros colonizadores com as mulheres africanas. Na análise dos haplogrupos detectados na nossa amostra, foi percebido que existem três agrupamentos principais separados por origem étnica. Estes três grupos foram caracterizados, a partir da observação dos haplogrupos principais, relacionando as frequências observadas com as frequências descritas em diversas populações espalhadas pelo globo. Este agrupamento, segundo os dados observados em diferentes grupos étnicos ou populacionais, mostrou que a frequência de indivíduos da nossa população miscigenada com marcadores frequentemente observados na Europa é de 77%, de africanos é de 14,87% e de ameríndios é de 2,31% (não tendo sido possível determinar a origem de aproximadamente 5% dos cromossomos Y).

Quando avaliamos os Y-SNPs da fração da população com influência paterna potencialmente africana, observamos que 47 indivíduos (7,7%) foram identificados como haplogrupo E1b1a1* - M2 e 26 indivíduos (4,30%) identificados como E1b1a1f - M191, característicos de populações da África subsaariana. O haplogrupo E1b1b1b1 - M78, encontrado em 29 indivíduos (4,79%), é muito característico no norte e sul da África, oeste da Ásia e toda a Europa. O haplogrupo E1b1b1b1 - M81, com 32 indivíduos encontrados (5,29%), tem origem provável na região do chifre da África com possíveis migrações através do noroeste da África (Magreb) e Líbia, onde apresenta uma frequência com valores médios de 45% entre os berberes. Apesar desta característica africana, também é comumente encontrado em Portugal (8%) e Espanha (4%), com a mais alta frequência em Pasiegos (30%), numa comunidade isolada das regiões montanhosas da Cantábria, no extremo norte da Península Ibérica, indicando traços da presença muçulmana. Para o haplogrupo Q1a3a1a*- mutação M3, típico de populações ameríndias, somente quatorze indivíduos (2,31%) foram encontrados, demonstrando a pequena influência deste grupo étnico na formação da população brasileira, via herança paterna.

O gráfico do MDS apresentado na Figura 24 mostra um agrupamento nítido formado entre as populações africanas do sudeste, oeste e centro-oeste com o

componente africano de nossa amostra (RJ_Afro). Os resultados obtidos, que mostram uma maior proximidade genética entre a população do Rio de Janeiro (aporte africano) com as populações destas regiões, estão de acordo com o que foi documentado historicamente em relação à origem do influxo escravo no Brasil (OGOT, 2000; IBGE, 2007).

Em relação ao estudo com amostras do oeste africano (Artigo 2, em preparação), foi observado que existem correspondências destas amostras com a região do centro-oeste africano, principalmente com Benin, Nigéria e Costa do Marfim. Em vista dos poucos estudos relativos a esta região da África, a finalização deste artigo e sua subsequente publicação, acrescentará mais dados populacionais aos bancos de dados já existentes, ampliando nosso conhecimento nesta área.

As comparações entre as populações do Rio de Janeiro e amostras africanas mostraram a ocorrência de uma grande influência de grupos étnicos provenientes do sudoeste e centro africano, confirmando nossos dados históricos referentes à entrada de escravos no Brasil. Sem dúvida, apresentamos uma herança africana de origem Bantu, caracterizada em nossa população a partir da detecção de elevadas frequências de haplogrupos específicos deste grupo étnico-linguístico (E1b1a* - M2 e E1b1a1f - M191).

Em vista de todos estes dados, temos como perspectiva a continuação deste estudo, com populações urbanas e afrodescendentes da região de Minas Gerais (capital e interior) e também com amostras da região Norte do Rio de Janeiro. A continuação destas análises também deve ser realizada com outras populações do Brasil, permitindo ampliar os conhecimentos sobre a formação do povo brasileiro. As diferenças que foram observadas em estudos preliminares poderão ser confirmadas ou não, a partir da construção de bancos de dados populacionais utilizando Y-STR associados com Y-SNPs, mostrando a existência ou não de uma homogeneidade nacional ou se ocorreram possíveis subestruturas relacionadas ao influxo colonial.

O fenômeno da Diáspora Transatlântica promoveu a dispersão de genes africanos para várias partes do globo, principalmente para as Américas e Antilhas. Segundo Hammer *et al.* (2006), o estudo das frequências alélicas e seu padrão de distribuição possibilita a caracterização de populações derivadas destas miscigenações, promovendo o entendimento destes fluxos migratórios e a caracterização das afinidades genéticas.

Muitos outros estudiosos realizaram diversas análises, tanto nas áreas da Paleogeografia e Etnolinguística, quanto na área da Genética Populacional. Os resultados destes esforços mostraram que a população africana é uma das mais diversas do mundo, levando em conta as variações genéticas, étnicas e linguísticas. Em vista disto, a ampliação de estudos populacionais nas populações da África se fazem necessários para o entendimento desta diversidade genética, assim como o modo como ocorreram as dispersões pelo mundo e as suas influências sobre todas as outras populações mundiais. Sem nenhuma dúvida, este trabalho também contribuiu para fornecer mais alguns dados genéticos, que podem ser somados aos estudos mundiais que estão sendo realizados, ampliando não só os nossos conhecimentos sobre a formação do Brasil, enquanto Nação, mas também colaborando com uma visão ampla de todas as populações que também foram influenciadas pelo fenômeno da Diáspora Africana.

CONCLUSÕES

a) As linhagens masculinas de amostras do Rio de Janeiro foram caracterizadas quanto aos polimorfismos de Y-STRs e Y-SNPs e apresentaram grande influência europeia. A segunda maior influência é africana seguida pelos nativo-americanos. A população apresentou alta diversidade haplotípica, tendo sido encontrados 23 diferentes haplogrupos. Do ponto de vista das suas linhagens paternas, a população do Rio de Janeiro mostrou-se significativamente diferente de populações do norte. A utilização de Y-SNPs mostrou ser uma boa estratégia para detectar subestruturação populacional descortinando diferenças nas populações miscigenadas.

b) A população do oeste africano, analisada neste trabalho, apresentou uma alta diversidade haplotípica ($1,0000 \pm 0,0018$), não apresentando diferenças significativas quando comparada com outras populações das regiões do oeste e do centro da África. Um total de oito diferentes haplogrupos foi encontrado, demonstrando sua baixa diversidade ($0,6895 \pm 0,0200$), e mostrando uma homogeneidade populacional típica dos países da região central e oeste da África.

c) Estudos comparativos entre a amostra do Rio de Janeiro e populações da África mostraram grande similaridade entre o aporte africano impresso na população estudada e as amostras das regiões oeste e central do continente africano, mostrando que não apresentam diferenças genéticas, quando comparadas com populações que falam a língua Bantu.

d) Portanto, os dados genéticos confirmam a principal contribuição de povos originários do oeste e centro da África, onde estão localizados os países, que historicamente, foram os principais responsáveis pelo envio de escravos ao Brasil.

REFERÊNCIAS

- ABE-SANDES, K. *et al.* Heterogeneity of the Y chromosome in Afro-brazilian populations. *Hum. Biol.*, v. 76, n. 1, p. 77-86, 2004.
- ALONSO, A (Editor). *DNA Electrophoresis Protocols for Forensic Genetics Methods in Molecular Biology*, v. 830, p. 109-126, 2012.
- ALVES, C. *et al.* Evaluating the informative Power of Y-STRs: a comparative study using European and new African haplotype data. *Forensic Sci. Int.*, v. 134, n. 2, p. 126-133, 2003.
- ALVES-SILVA, J. *et al.* The ancestry of brasilian mtDNA lineages. *Am. J. Hum. Gen.*, v. 67, n. 2, p. 444-461, 2000.
- ARCOT, S.S. *et al.* Alu fóssil relics-distribution and insertion polymorphism. *Gen. Res.*, v. 6, p. 1084-1092, 1996.
- ARCOT, S.S. *et al.* Alu repeats: a source for the genesis of primate microsatellites. *Genomics*, v. 29, n. 1, p. 136-144, 1995.
- ARREGLO. La Etnia Negra de Panama. Disponível em:
<<http://diadelaetnia.homestead.com/ARREGLO.html>>. Acesso em 10 nov. 2014
- ARROYO-PARDO, E. *et al.* Genetic variability of 16 Y-chromosome STRs in a sample from Equatorial Guinea (Central Africa). *Forensic Sci. Int.*, v. 149, n. 1, p. 109-113, 2005.
- AURICCHIO, M.T.B.M. *et al.* Frequency and origins of HBB*S mutation in African-derived Brazilian populations. *Hum. Biol.*, v. 25, n. 2, p. 468-74, 2008.
- AVERY, O.T., MACLEOD, C.M., MCCARTY, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.*, v. 79, n. 2, p. 137-158, 1944.
- AVISE, J.C. *et al.* Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, v. 18, p. 489-522, 1987.
- BACHTROG, D. Y chromosome evolution: emerging insights into processes of Y chromosome degeneration. *Nat Rev Genet*, v. 14, n. 2, p. 113–124, 2013.
- BALLANTYNE, K. N. *et al.* A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci. Int. Genet.*, v. 6, p. 208–218, 2012.

BALLANTYNE, K. N. *et al.* Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications. *AJHG*, v. 87, p. 341–353, Sep 2010.

BASSAM, B.J. *et al.* Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Bioch.*, v. 196, p. 80, 1991.

BASTOS-RODRIGUES, L., PIMENTA, J.R., PENA, S.D.J. The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Ann. Hum. Genet.*, v. 70, p.1-20, 2006.

BATZER, M.A. *et al.* African origin of human-specific polymorphic *Alu* insertions. *Proc. Natl. Acad. Sci.*, v. 91, p. 12288-12292, 1994.

BEHAR, D.M. *et al.* Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. *Hum. Genet.*, n. 114, p. 354–365, 2004.

BELEZA, S. *et al.* Micro-phylogeographic and demographic history of Portuguese male lineages. *Ann. Hum. Genet.*, v.70, n. 2, p. 181-194, 2006.

BELEZA, S. *et al.* A: The genetic legacy of western Bantu migrations. *Hum. Genet.*, n. 117, p. 366–375, 2005.

BERNIELL-LEE, G. *et al.* Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol. Biol. Evol.*, v.26, p. 1581–1589, 2009.

BERGLUND, E.C. *et al.* Next-generation sequencing technologies and applications for human genetic history and forensics. *Investig. Genet.*, v. 2, n. 23, 2011.

BIANCHI, N.O. *et al.* Characterization of ancestral and derived Y-chromosome haplotypes of new world native populations. *Am. J. Hum. Genet.*, v. 63, n. 6, p. 1862-1871, 1998.

BLANCO-VEREA, A. *et al.* Analysis of Y chromosome lineages in native South American population. *Prog. Forensic Genet.*, v. 1288, n. 11, p. 222-224, 2006.

BLENCH, R. *Archaeology, language, and the African past.* Lahham (MD): Alta Mira Press, 2006.

BORTOLINI, M.C. *et al.* Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am. J. Hum. Genet.*, v. 73, n. 3, p. 524-539, 2003.

BOSCH E. *et al.* High-Resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between Northwestern Africa and the Iberian Peninsula. *Am. J. Hum. Genet.*, v. 68, n. 4, p. 1019-1029, 2001.

BRADLEY, B. & STANFORD, D. The North Atlantic ice-edge corridor: a possible Paleolithic route to the New World. *World Archaeol.*, v. 36, n. 4, p. 459-478, 2004.

BRYC, K. *et al.* Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc. Natl. Acad. Sci.*, n. 107, p. 786-791, 2010.

BRIÓN, M. *et al.* Introduction of a single nucleotide polymorphism-based “Major Y-chromosome haplogroup typing kit” suitable for predicting the geographical origin of male lineages. *Electrophoresis*, v. 26, n. 23, p. 4411-4420, 2005.

BRIÓN, M. *et al.* Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *Int. J. Legal Med.*, Berlin, v. 119, n. 1, p. 10-15, 2004.

BRIÓN, M. *et al.* Insights into iberian population origins through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite. *Am. J. Phys. Anthropol.*, v. 122, n. 2, p. 147-161, 2003.

BRUMFIELD, R.T., BEERLI, P., NICKERSON, D.A., EDWARDS S. The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol.*, v. 18, n. 5, p. 249-256, 2003.

BUDOWLE, B. & VAN DAAL, A. Forensically relevant SNP classes. *Bio Techniques*, v. 44, n. 5, p. 603-610, 2008.

BUTLER, J.M. *Forensic DNA typing*. 2nd ed. Burlington: Elsevier Academic Press, 2005. 660 p.

CAGLIÀ, A. *et al.* A study of Y chromosome microsatellite variation in sub-Saharan Africa: a comparison between *Fst* and *Rst* genetic distances. *Hum. Biol.*, n. 75, p. 313-330, 2003.

CAINÉ, L. *et al.* Allele frequencies and haplotypes of 8 Y-chromosomal STRs in the Santa Catarina population of southern Brazil. *Forensic Sci. Int.*, v. 148, n. 1, p. 75-79, 2005.

CAMPBELL, M.C. *et al.* African genetic diversity: implications for human demographic history, modern human origins and complex disease mapping. *Ann. Rev. Genet. Hum.*, v. 9, p. 403-433, 2008.

CAMPBELL, M.C. & TISHKOFF, S.A. The evolution of human genetic and phenotypic variation in Africa. *Curr. Biol.*, v. 20, p. 166-73, 2010.

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

CARVALHO, M. *et al.* Paternal and maternal lineages in Guinea-Bissau population. *Forensic Sci. Int. Genet.*, n. 5, p. 114–116, 2011.

CARVALHO-SILVA, D.R. *et al.* Y chromosome diversity in Brazilian, switching perspectives from slow to fast evolving markers. *Genetica*, n. 126, p. 251-260, 2006.

CARVALHO-SILVA, D.R. *et al.* The phylogeography of brazilian Y-chromosome lineages. *Am. J. Hum. Genet.*, v. 68, n.1, p. 281-286, 2001.

CAVALLI-SFORZA, L.L. *et al.* The History and geography of human genes. USA. Princeton University Press. 1996

CAVALLI-SFORZA, L.L. *et al.* The DNA revolution in population Genetics. *Trends Genet.*, v. 14, n. 2, p. 60-65, 1998.

CINNIOGLU, C. *et al.* Excavating Y-chromosome haplotype strata in Anatolia. *Hum. Genet.*, v. 114, n. 2, p. 127-148, 2004.

CENSO DEMOGRÁFICO – 2010- BRASIL INDÍGENA - FUNAI.GOV.BR. Disponível em <<http://www.funai.gov.br/arquivos/conteudo/ascom/2013/img/12-Dez/pdf-brasil-ind.pdf>>. Acesso em: 2 jan 2015.

CLAUSSEN, M. *et al.* Simulation of an abrupt change in Saharan vegetation in the mid-Holocene. *Geo. Res. Lett.*, v. 26, n. 14, p. 2037-2040, Jul 1999.

COBLE, M.D. *et al.* Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PLoS One*, n. 4, p.e4838, 2009.

COELHO, M. *et al.* On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genética variation in southwestern Angola. *BMC Evol. Biol.*, v. 9, n. 80, p. 1-18, 2009.

COLLINS, F.S., BROOKS, L.D., CHAKRAVARTI, A. A DNA polymorphism discovery resource for research on human genetic variation. *Gen. Res.*, v. 8, n. 12, p. 1229-1231, 1998.

COMUNIDADE INZO NKONGBILA.

Disponível em <<http://www.inzotumbasi.org>>. Acesso em: 5 fev 2012.

CRUCIANI, F. *et al.* A Back Migration from Asia to Sub-Saharan Africa Is Supported by High-Resolution Analysis of Human Y-Chromosome Haplotypes. *Am. J. Hum. Genet.*, v. 70, n. 5, p. 1197–1214, 2002.

DA COSTA FRANCEZ, P.A. *et al.* Allelic frequencies and statistical data obtained from 12 codis STR loci in an admixed population of the Brazilian Amazon. *Genet. Mol. Biol.*, n. 34, p. 35-39, 2011.

DE FILIPPO, C. *et al.* Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Mol. Biol. Evol.*, n. 28, p. 1255–1269, 2011.

DE FILIPPO, C. *et al.* Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia. *Am. J. Phys. Anthropol.*, n. 141, p. 382-394, 2010.

DEL PRIORE M.D. & VENÂNCIO R. *Uma Breve História do Brasil*. Ed. São Paulo, Planeta do Brasil, 2010.

dbSNP – The National Center for Biotechnology Information.

Disponível em <<http://www.ncbi.nlm.nih.gov/SNP>>. Acesso em: 6 ago 2014.

DEL PRIORE, M.D. & VENÂNCIO, R.P. *O Livro de Ouro da História do Brasil*: Rio de Janeiro: Ediouro Publicações S.A., 2001. 408 p.

DIAMOND, J. & BELWOOD, P. Farmers and their languages: the first expansions. *Science*, v. 300, p. 597-603, 2003.

DILLEHAY, T.D. Probing deeper into first American studies. *Proc. Natl. Acad. Sci.*, v.106, n. 4, p. 971-978, 2009.

DOMINGUES, P.M. *et al.* Sub-saharan Africa descendentes in Rio de Janeiro (Brazil): population and mutational data for 12 STR loci. *Int. J. Legal Med.*, v. 121, p. 1437-1596, 2007.

DORNELLES, C.L. *et al.* Genetics, surnames, grandparents nationalities, and ethnic admixture in Southern Brazil: do the parents of variation coincide? *Genet. Mol. Biol.*, v. 22, p. 151-161, 1999.

DULIK, M.C. *et al.* Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations, *Proc. Natl. Acad. Sci.*, v. 109, p. 8471-8476, 2012.

DUPUY, B.M. *et al.* Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. *Hum. Mutat.*, n. 23, p.117-124, 2004.

ETIMOLOGIA. Disponível em <<http://www.sk.com.br/sk-hist.html>>. Acesso em 10 fev 2012.

EXCOFFIER, L., LAVAL, G., SCHNEIDER, S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform Online*, v. 1, p. 47-50, 2005.

EXPLORING THE NIGER-CONGO LANGUAGES. Disponível em <<http://linguistics.byu.edu/classes/ling450ch/reports/niger-congo.html>>. Acesso em 7 dez 2014.

FLORENTINO, M. *Em costas negras: uma história do tráfico atlântico de escravos entre África e Rio de Janeiro (séculos XVIII e XIX)*, Rio de Janeiro: Companhia das Letras, 1997.

FOWLER, J.C.S. *et al.* Repetitive deoxyribonucleic acid (DNA) and human genome variation – A concise review relevant to forensic biology. *J. Forensic Sci.*, v. 33, n. 5, p. 1111-1126, 1988.

FREEMAN, J. L., *et al.* Copy number variation: New insights in genome diversity. *Gen. Res.*, v. 16, p. 949–961, 2006.

FREIRE, J. Família, parentesco espiritual e estabilidade familiar entre cativos pertencentes a grandes posses de Minas Gerais século XIX. *Afro-Ásia*, n.46, p. 9-59, 2012.

GEPPERT, M. *et al.* Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America. *Forensic Sci. Int. Genet.*, n. 5, p. 100–104, 2011.

GÓES, A.C.S. *et al.* Population and mutation analysis of 17 Y-STR loci from Rio de Janeiro (Brazil). *Int. J. Legal Med.*, v. 119, n. 2, p. 70-76, 2005.

GOIS, C.C. *et al.* Genetic population data of 12 STR loci of the PowerPlex1 Y system in the state of São Paulo population (Southeast of Brazil). *Forensic Sci. Int.*, v.174, p. 80–85, 2008.

GOLDSTEIN, D.B., SCHLÖTTERER, C. *Microsatellites evolution and applications*. New York: Oxford University Press, 1999. 352 p.

GOMES, V. *et al.* Digging deeper into East African human Y chromosome lineages. *Hum. Genet.*, v.127, n. 5, p. 603- 613, 2010.

GONZÁLEZ, M. *et al.* The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y chromosome haplogroup R-V88. *Eur. J. Hum. Genet.*, v. 21, p. 324–331, 2013.

GONÇALVES, V. F. *et al.* The phylogeography of African Brazilians. *Hum. Herd.*, v. 65, n.1, p.23-32, 2008.

GOULART, M. *A escravidão Africana no Brasil: das origens à extinção do tráfico*. Editora Alfa-Omega. São Paulo-SP, 1975. 316 p.

GRATTAPAGLIA, D. *et al.* Y-chromosome STR haplotype diversity in Brazilian populations. *Forensic Sci. Int.*, v. 149, p. 99-107, 2005.

GRESHAM, D. *et al.* Origins and divergence of the Roma (Gypsies). *Am. J. Hum. Genet.*, v. 69, n. 6, p. 1314-1331, 2001.

GREENBERG, J.H. *The Languages of Africa*. 2^a ed. Bloomington: Indiana University, 1966.

GREENBERG, J.H. The classification of African languages. *Am Anthropol*, v. 50, p. 24-30, 1948.

GUERREIRO, J.F. *et al.* Genetical demographic data from two Amazonian populations composed of descendants of African slaves: Pacoval and Curiaú. *Genet. Mol. Biol.*, v. 22, p. 485-492, 1999.

GUSMÃO, L., ALVES, C. Marcadores dos cromossomos sexuais. Relevância na identificação genética. In *Genética Forense. Perspectivas da Identificação Genética*: Porto. Edições Universidade Fernando Pessoa. 2010.

GUSMÃO, L. *et al.* A perspective on the history of the Iberian Gypsies provided by phylogeographic analysis of Y chromosome lineages. *Ann. Hum. Genet.*, v. 72, n. 2, p. 215-227, 2008.

GUSMÃO, L. *et al.* DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Int. J. Leg. Med.*, n. 120, p. 191-200, 2006.

GUTHRIE, M. *Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages*. 4 vols. Farnborough: Gregg Press, 1971.

GUTHRIE, M. *The classification of the Bantu languages*. London: Oxford University Press for the International African Institute, 1948.

HAMMER, M.F. *et al.* Populations structure of Y-Chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. *Forensic Sci. Int.*, v. 164, p. 45-55, 2006.

HAMMER, M.F. *et al.* Hierarchical patterns of global human Y-Chromosome diversity. *Mol. Biol. Evol.*, v. 18, n. 7, p. 1189–1203, 2001.

HAMMER, M.F. *et al.* Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.*, v. 15, n. 4, p. 427-441, 1998.

HAMMER MF, *et al.* The geographic distribution of human Y chromosome variation. *Genetics*, n.145, p.:787-806, 1997.

HAMMER, M.F. *et al.* The role of the Y chromosome in human evolutionary studies. *Evol. Anthropol.*, n. 5, p. 116–134, 1996.

HAMMER, M.F. & HORAI. Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.*, n. 56, p. 951-962, 1995.

HAMMER, M.F. A recent insertion of a Alu element on the Y chromosome is a useful marker for human population studies. *Mol. Biol. Evol.*, v. 11, n. 5, p. 749-761, 1994.

HEYWOOD, Linda M. (Org). *Diáspora negra no Brasil*. São Paulo: Ed. Contexto, 2008. 222p.

HEWLETT, B.S. & CAVALLI-SFORZA, L.L. Cultural transmission among Aka pygmies. *Am. Anthropol.*, v. 88, p. 922-934, 1986.

HUDJASHOV, G. *et al.* Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Natl. Acad. Sci.*, v. 104, n. 21, p. 8726-8730, 2007.

HUNEMEIER, T. *et al.* Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. *Am. J. Phys. Anthropol.*, v. 133, p. 854-867, 2007.

HUMMEL, S. *Ancient DNA typing: methods, strategies and applications*. Germany: Springer-Verlag, 2003. 298 p.

HURLES, M.E. & JOBLING, M.A. A singular chromosome. *Nature Genet.*, v. 34, p. 246-247, 2003.

IBGE a – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Brasil 500 anos de povoamento*. Centro de Documentação e Disseminação de Informações. - Rio de Janeiro: IBGE, 2007. 232 p.

IBGE b – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Tendências demográficas: uma análise da população com base nos resultados dos censos demográficos 1940 e 2000*. Rio de Janeiro: IBGE, 2007. 116 p.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Initial sequencing and analysis of the human genome. *Nature*, v. 409, p. 860-921, Feb 2001.

JEFFREYS, A.J., WILSON, V., THEIN, S.L. Individual-specific fingerprints of human DNA. *Nature*, v. 316, p. 76-79, 1985.

JOBLING, M.A. *et al.* *Human Evolutionary Genetics: Origins, Peoples & Disease*. Garland Publishing, 2004.

JOBLING, M.A. & TYLER-SMITH, C. The human Y chromosome: an evolutionary marker comes of age. *Nature Rev. Genet.*, n. 4, p. 598-612, 2003.

JOBLING, M.A. *et al.* Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y-specific minisatellite, MSY1 (DYF155S1). *Hum. Mol. Genet.*, v. 7, n. 4, p. 643-653, 1998.

JOBLING, M.A., PANDYA, A., TYLER-SMITH, C. The Y chromosome in forensic analysis and paternity testing. *Int. J. Legal. Med.*, v. 110, n. 3, p. 118-124, 1997.

JOBLING, M.A. A survey of long-range DNA polymorphisms on the human Y chromosome. *Hum. Mol. Genet.*, v. 3, n. 1, p. 107-114, 1994.

JORDE, L.B. *et al.* The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.*, v. 66, p. 979-988, 2000.

KARAFET, T.M. *et al.* Improved phylogenetic resolution and rapid diversification of Y-chromosome haplogroup K-M526 in Southeast Asia. *Eur. J. Hum. Genet.*, v. 4, 2014.

KARAFET, T.M. *et al.* New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Gen. Res.*, v. 18, p. 830-838, 2008.

KARAFET, T.M. *et al.* High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum. Biol.*, v. 74, n. 6, p. 761-789, 2002.

KARAFET, T.M. *et al.* Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.*, v. 69, p. 615-628, 2001.

KARAFET T, *et al.* Y chromosome markers and trans-Bering Strait dispersals. *Am. J. Phys. Anthropol.*, n. 102, p. 301-314, 1997.

KASS, D.H. & BATZER, M.A. Genome Organization/Human. Encyclopedia of Life Science. Nature Publishing Group / www.els.net. 2001.

KAYSER M & DE KNIJFF P. Improving human forensics through advances in genetics, genomics and molecular biology. *Nat. Rev. Genet.*, v.12, p.179-192, 2011.

KAYSER, M. *et al.* Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol. Biol. Evol.*, v. 23, n.11, p. 2234-2244, 2006.

KAYSER, M. *et al.* Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis*, v.18, p. 1602-1607, 1997.

KENNETH, K.K. *et al.* Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Sci. Int. Genet.*, v. 12, p. 215–224, 2014.

KING, T.E. *et al.* Thomas Jefferson's Y chromosome belongs to a rare European lineage. *Am. J. Phys. Anthropol.*, v. 132, n. 4, p. 584-589, 2007.

KIVISILD, T. *et al.* The genetic heritage of the earliest conjuntotlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.*, v. 72, n.2, p. 313-322, 2003.

KLEIN, H.S. As origens africanas dos escravos brasileiros. In: Pena S.D.J. (org.). *Homo brasílis: aspectos genéticos, linguísticos, históricos e socioantropológicos da formação do povo brasileiro*. São Paulo: Funpec, 2002.

KNIJFF, P. Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic Markers on the human Y chromosome. *Am. J. Hum. Genet.*, v. 67, n. 5, p. 1055-1061, 2000.

KOBILINSKY, L., LIOTTI, T.F., OESER-SWEAT, J. *DNA forensic and legal applications*. New Jersey: Willey-Interscience, 2005.

KRIEGS, J.O. *et al.* Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet.*, n. 23, p. 158-161, 2007.

LACOMBE, A. J., SILVA, E., BARBOSA, F.S. Rui Barbosa e a Queima dos Arquivos. Brasília: Ministério da Justiça, Rio de Janeiro: Fundação Casa de Rui Barbosa, 1988, p. 11-26.

LAHN, B.T., PEARSON, N.M., JEGALIAN, K. The human Y chromosome in the light of evolution. *Nature Genet.*, v. 2, n. 3, p. 207-216, 2001.

LELL, J.T. *et al.* The dual origin and Siberian affinities of native American Y chromosomes. *Am. J. Hum. Genet.*, v. 70, n. 1, p. 192-206, 2002.

LEWIN, B., *et al.* GENES XI. Jones & Bartlett Publishers, 2014. 940 p.

MAHO, J. A Comparative Study of Bantu Noun Classes. Göteborg: Acta Universitatis Gothoburgensis, 1999.

MANTA, F.S. *et al.* Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. *Ann. Hum. Biol.*, n. 40, p. 94-98, 2013.

MARANHÃO, R. & KEATING, V. Diário de Navegação - Pero Lopes e a Expedição de Martim Afonso de Sousa (1530 - 1532), 1ª ed. São Paulo: Ed. Terceiro Nome, 2011. 266 p.

MARRERO, A.R. *et al.* Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. *Am. J. Phys. Anthropol.*, v. 132, n. 2, p. 310-310, 2006.

MARRERO, A.R. *et al.* Heterogeneity of the genome ancestry of individuals classified as White in the State of Rio Grande do Sul. *Am. J. Hum. Biol.*, n. 17, p. 496-506, 2005.

MELTON, L. On the trail of SNPs. *Nature*, v. 422, n. 917 - 923, 2003.

MENDEZ, F.L. *et al.* An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am. J. Hum. Genet.*, n. 92, p. 454-459, 2013.

MICHALAKIS, Y & EXCOFFIER, L. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, v. 142, n. 3, p. 1061-1064, 1996.

MILLS, R.E. *et al.* An initial map of insertion and deletion (INDEL) variation in the human genome. *Gen. Res.*, v. 16, p 1182-1190, 2006.

MOORAD, JA. Multi-level sexual selection: individual and family-level selection for mating success in a historical human population. *Evolution*, v. 67, n. 6, p. 1635-48, 2013.

MULERO, J.J. *et al.* Development and Validation of the AmpF'STRs Yfiler™ PCR Amplification Kit: A Male Specific, Single Amplification 17 Y-STR Multiplex System. *J. Forensic Sci.*, v. 51, n. 1, 2006.

MULLIS, K. *et al.* Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*. New York, v. 51, p. 263-273, 1986.

MYRES, N.M. *et al.* A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur. J. Hum. Genet.*, v. 19, p. 95-101, 2011.

NASCIMENTO, E. *et al.* The Africa male lineages of Bahia's people-Northeast Brazil: a preliminary SNPs study. *Forensic Sci. Int. Genet.*, v. 2, n. 1, p. 349-350, 2009.

NARLOCH, L & TEIXEIRA, D. Guia politicamente incorreto da América Latina. São Paulo: Editora Leya, 2011.336 p.

OGOT, B.A (Editor). General History of Africa I – Africa from the Sixteenth to the Eighteenth Century – France: UNESCO 2000. 735 p.

O MOVIMENTO TAPUIA.

Disponível em <http://rpg_ficcao.sites.uol.com.br/Outro/Nomestapuia.htm>. Acesso em: 3 abr 2012.

OLIVEIRA, A.M. *et al.* Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs. *Forensic Sci. Int. Genet.*, n. 13, p. 264-268, 2014.

OLIVEIRA, A.M. *et al.* Analysis of Y chromosome lineages in a sample from Sub-Saharan Africa descendents in Rio de Janeiro. *Forensic Sci. Int. Genet.*, n. 442-443, 2009.

OTHA, T., KIMURA, M. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Gen. Research*, v. 22, n. 2, p. 201-204, 1973.

OVEN, M.N & KAYSER, M. Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. *Hum. Mutat.*, v. 30, n. 2, p. 386- 394, 2009.

PALHA, T. *et al.* Disclosing the Genetic Structure of Brazil through Analysis of Male Lineages with Highly Discriminating Haplotypes. *PLoSOne*, Jul 10, 2012.

PENA, S.D. *et al.* The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected. *Plos One*, v. 6, p. e17063, 2011.

PENA, S.D. *et al.* Retrato molecular do Brasil. *Ciência Hoje*, São Paulo, v. 159, p. 16-25, 2000.

PEREIRA J.M.N. *África - um novo olhar*. RJ: Cadernos CEAP, 2006. 88 p.

PEREIRA, R. *et al.* Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One*, v. 7, 2012.

PÉREZ-LEZAUN, A. *et al.* Population genetics of Ychromosome short tandem repeats in humans. *J. Mol. Evol.*, n. 45, p. 265-270, 1997.

PHILLIPS, C. *et al.* A 34-plex Autosomal SNP Single Base Extension Assay for Ancestry Investigations. *Methods Mol. Biol.*, n. 830, p. 109-26, 2012.

PONTES, M.L. *et al.* Allele frequencies and population data from 17 Y-STR loci (AmpF[®]STR Y-filer) in Northern Portuguese population sample. *Forensic Sci. Int.*, v. 170, p. 62–67, 2007.

PRADO JÚNIOR, C. *Formação do Brasil contemporâneo*. São Paulo: Braziliense, 1997. 660 p.

PURPS, J. *et al.* A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci. Int. Genet.*, n. 12, p. 12-23, 2014.

QUADRO GERAL DOS POVOS - POVOS INDÍGENAS NO BRASIL. Disponível em <<http://pib.socioambiental.org/pt/c/quadro-geral>>. Acesso em: 20 out 2014.

QUINTANA-MURCI, L. *et al.* The human Y chromosome function, evolution and disease. *Forensic Sci. Int.*, n. 118, p. 169-181, 2001.

REDD, A.J. *et al.* Genetic structure among 38 populations from United States based on 11 U.S core Y chromosome STRs. *J. Forensic Sci.*, v. 51, n. 3, p. 580-585, 2006.

REDD, A.J. *et al.* Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci. Int.*, v. 130, n. 2-3, p. 97-111, 2002.

REDON, R. *et al.* Global variation in copy number in the human genome. *Nature*, v. 444, p. 444–454, 2006.

RIBEIRO, D. *O Povo Brasileiro: A formação e o sentido do Brasil*. Rio de Janeiro: Companhia das Letras, 1995. 435p.

RIVET, P. *As Origens do Homem Americano*. Editora Anhambi. 2008. 152p.

ROEWER, L. *et al.* Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. *PLoS Genet.*, n. 9, 2013.

ROMANINI, C. *et al.* Ancestry informative markers: Inference of ancestry in aged bone samples using an autosomal AIM-Indel multiplex. *Forensic Sci. Int. Genet.*, n. 5, p. 58-63, 2014.

ROOTSI, S. *et al.* A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur. J. Hum. Genet.*, v. 15, n. 2, p. 204-211, 2007.

ROOTSI, S. *et al.* Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am. J. Hum. Genet.*, v. 75, n. 1, p. 128-137, 2004.

ROSA, A. *et al.* Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. *BMC Evol. Biol.*, v. 7, n. 124, 2007.

ROZEN, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*, v. 423, p. 873-876, 2003.

RUIZ-PESINI, E. *et al.* An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res.*, v. 35, p. 823-828, 2007.

SAIKI, R. K. *et al.* Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle-Cell Anemia. *Science*, v. 230, p. 1350-1354, 1985.

SALZANO, F.M. Molecular variability in Amerindians widespread but uneven information. *An. Acad. Bras. Cienc.*, v. 74, n. 2, p.223-263, 2002.

SALZANO, F.M. Human races: myth, invention or reality? Rio de Janeiro: Interciência, v. 22, p. 221-227, 1997.

SAMBROOK, J. *et al.* Molecular cloning: A laboratory manual. New York, USA. Cold Spring Harbor Press, 1989.

SANCHEZ, J.J. *et al.* A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis*, v. 27, n. 9, p. 1713-1724, 2006.

SANCHEZ, J.J. *et al.* Multiplex PCR and minisequencing of SNPs – a model with 35 Y chromosome SNPs. *Forensic Sci. Int.*, n.137, p.74–84, 2003.

SANGER, F. *et al.* DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, v. 74, n. 12, p. 5463-5467, 1977.

SANTOS, F.R. & TYLER-SMITH, C. Reading the human Y chromosome: the emerging DNA markers and human history. *Braz. J. Genet.*, v. 19, n. 4, p. 665-670, 1996.

SANTOS, M.V. *et al.* Y chromosome DNA haplotypes in human samples from Bahia, Brazil. International Congress Series, v. 1239, n. 9, p. 443– 448, 2003.

SCHNEIDER, H. *et al.* Isolate breakdown in Amazonia. The blacks of the Trombetas river. *Braz. J. Genet.*, v. 10, p. 565-574, 1987.

SCHEINFELDT, L. *et al.* Unexpected NRY chromosome variation in Northern Island Melanesia. *Mol. Biol. Evol.*, v. 23, n. 8, p. 1628-1641, 2006.

SCHLÖTTERER, C. Genome evolution: are microsatellites really simple sequences? *Curr. Biol.*, v. 8, n. 4, p. 132-134, 1998.

SCHOSKE, R. *The design, optimization and testing of Y chromosome short tandem repeat megaplexes.* 2003. 255 f. PhD Thesis (Doctorate in Chemistry) – Faculty of the College of Arts and Sciences, American University, Washington D.C, 2003.

SCHULTES, T. *et al.* Amplification of Y-chromosomal STRs from ancient skeletal material. *Hum. Genet.*, v. 104, n. 2, p. 164-166, 1998.

SCHWENGBER, S.P. *et al.* Population data of 17 Y-STR loci from Rio Grande do Sul state (South Brazil). *Forensic Sci. Int. Genet.*, v. 4, n. 1, p. 31-33, 2009.

SCOZZARI, R. *et al.* Molecular Dissection of the basal clades in the human Y chromosome Phylogenetic Tree. *PLoS One*, v. 7, n. 11, 2012.

SEMINO, O. *et al.* Origin, diffusion and differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.*, v. 74, p. 1023-1034, 2004.

SEMINO, O. *et al.* Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.*, v. 70, n. 1, p. 265-268, 2002.

SEMINO, O. *et al.* The genetic legacy of paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science*, v. 290, n. 5494, p. 1155-1159, 2000.

SENGUPTA, S. *et al.* Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am. J. Hum. Genet.*, v. 78, n. 2, p. 202-221, 2006.

SHURR, T.G., SHERRY, S.T. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am. J. Hum. Biol.*, v. 16, p. 420- 439, 2004.

SILVA, D.A. *et al.* Y-chromosome genetic variation in Rio de Janeiro Population. *Am. J. Hum. Biol.*, v. 18, n. 6, p. 829-837, 2006.

SILVA, A.C. *A África explicada aos meus filhos*: Rio de Janeiro: Agir, 2008. 129 p.

SKALESTSKY, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, v. 423, p. 825-822, 2003.

SLATKIN, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, v. 139, n. 1, p. 457-462, 1995.

SÓ HISTÓRIA – ORIGEM DO NOME ÍNDIO NA AMÉRICA. Disponível em <<http://www.sohistoria.com.br/curiosidades/indio/>>. Acesso em: 5 out 2014.

SPURDLE, A.B. *et al.* The Y Alu polymorphism in southern African populations and its relationship to other Y-specific polymorphisms. *Am. J. Hum. Genet.*, n. 54, p. 319-330, 1994a

SPURDLE, A.B. *et al.* The genetic affinity of Polynesians: evidence from Y chromosome polymorphisms. *Ann. Hum. Genet.*, n. 258, p. 251-263, 1994b.

STURROCK, K., ROCHA, J. A multidimensional scaling stress evaluation table. *Field Methods*, v. 12, n. 1, p. 49-60, 2000.

SU, B. *et al.* Polynesian origins: insights from the Y chromosome. *Proc. Natl. Acad. Sci. U.S.A.*, v. 97, n. 15, p. 8225-8228, Jul 2000.

SU, B. *et al.* Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am. J. Hum. Genet.*, v. 65, n. 6, p.1718-1724, 1999.

SUA PESQUISA.COM – CONTINENTE AFRICANO. Disponível em <www.sua.pesquisa.com/geografia/continente_africano.htm>. Acesso em: 5 set 2014.

THE BANTU EXPANSION.

Disponível em <<http://www.zonu.com/detail-en/2010-01-06-11637/The-Bantu-expansion-3000-1000>>. Acesso em: 9 set 2014.

THE 2011 HARRIET – ELLIOTT LECTURE SERIES BLOG. Disponível em <<https://harrietelliottuncg.wordpress.com/2011/02/10/genetics-archaeology-and-the-peopling-of-the-americas/>>. Acesso em: 10 nov 2014.

THE INTERNATIONAL HAPMAP CONSORTIUM: A second generation human haplotype map of over 3.1 million SNPs. *Nature*, v. 449, 851–861, 2007.

TISHKOFF, S.A. *et al.* History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol. Biol. Evol.*, n. 24, p. 2180–2195, 2007.

TISHKOFF, S.A. *et al.* Genetic analysis of African populations: Human Evolution and Complex Disease. *Nature reviews*, v. 3, p. 611-621, 2002.

TORERO, J. R. & PIMENTA, R. A. *Terra papagali*. Rio de Janeiro: Objetiva, 2000. 191p.

TRIPATHI, M. *et al.* Role of Alu element in detecting population diversity. *Int. J. Hum. Genet.*, v. 8, n. 1-2, p. 61-74, 2008.

TROMBETTA, B. A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms. *PLoS One*, v. 6, 2011.

TYLER-SMITH, C., BROWN, W.R. Structure of the major block of alphoid satellite DNA on the human Y chromosome. *J. Mol. Biol.*, v. 195, n. 3, p. 457-470, 1987.

UNDERHILL, P.A. *et al.* Maori origins, Y-chromosome haplotypes and implications for human history in the pacific. *Hum. Mutat.*, v. 17, n. 4, p. 271-280, 2001.

UNDERHILL, P.A. *et al.* The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.*, v. 65, p. 43-62, 2001.

UNDERHILL, P.A. *et al.* A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci.*, v. 93, n.1, p. 196-200, 1996.

VAN OVEN, M. *et al.* Multiplex genotyping assays for fine-resolution subtyping of the major human Y-chromosome haplogroups E, G, I, J, and R in anthropological, genealogical, and forensic investigations. *Electrophoresis*, n. 34, p. 20-21, 2013.

VAN OVEN, M. *et al.* An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. *Int. J. Legal Med.*, n. 125, p.879-885, 2011.

VAN OVEN, M *et al.* Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Hum Mutat*, n. 35, p.187-191, 2013. VANSINA, J. New linguistic evidence and the Bantu expansion. *J. African Hist.*, n. 36, p. 173-195, 1995.

VENTER, J.C. *et al.* The sequence of the human genome. *Science*, v. 16, p. 1304-1351, 2001.

VULLO, C. *et al.* Association between Y haplogroups and autosomal AIMs reveals intra-population substructure in Bolivian populations. *Int. J. Legal Med.* [Epub ahead of print], 2014.

WALSH, P.S., METZGER, D.A., HIGUCHI, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*, v. 10, n. 4, p. 506-13, 1991.

WATERS, M.R. & STAFFORD, T.W. JR. Redefining the age of Clovis: implications for the peopling of the Americas. *Science*, v. 23, n. 315, p. 1122-1126, 2007.

WATSON, J.D., CRICK, F.H.C. Molecular structure of nucleic acids. *Nature*, v. 171, p. 737-738, 1953.

WEBER, J.L. *et al.* Human diallelic insertion/deletion polymorphisms. *Am. J. Hum. Genet.*, v. 71, p. 854-862, 2002.

WIESEL, C.E. *et al.* Y-linked microsatellites in Amazonian Amerindians applied to ancestry estimates in Brazilian Afro-derived populations. *Am. J. Hum. Biol.*, n. 25, p. 313-317, 2013.

WIKIPEDIA – ÁFRICA SUBSAARIANA. Disponível em <http://pt.wikipedia.org/wiki/%C3%81frica_subsaariana>. Acesso em: 5 out 2014.

WIKIPEDIA-ETHNIC GROUPS. Disponível em <http://en.wikipedia.org/wiki/Ethnic_groups_in_Africa>. Acesso em: 2 ago 2014.

WOLFGRAMM, E. V. *et al.* Genetic analysis of 15 autosomal and 12 Y-STR loci in the Espírito Santo State population. *Brazil Forensic Sci. Int. Genet.*, v. 5, p. 41– 43, 2011.

WOOD, E.T., *et al.* Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur. J. Hum. Genet.*, v. 13, p. 867–876, 2005.

Y-CHROMOSOME BIALLELIC HAPLOGROUPS. Disponível em <<http://www.roperld.com/ybiallelichaplogroups.htm>>. Acesso em: 5 fev 2012.

YALE SCHOOL OF MEDICIN. Disponível em <<http://info.med.yale.edu/genetics/kkidd/point.html>>. Acesso em: 14 out 2014.

YAN, S. *et al.* An updated tree of Y-chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. Genographic Consortium. *Eur. J. Hum. Genet.*, n. 19, p.1013-1015, 2011.

Y-CHROMOSOME CONSORTIUM. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Gen. Res.*, v. 12, n. 2, p. 339-348, 2002.

YHRD. Disponível em <http://yhrd.org/pages/resources/release_history>. Acesso em: 29 out 2014.

ZAGO, M.A. *et al.* Bantu beta s cluster haplotypes predominates among Brazilian blacks. *Am. J. Phys. Anthropol.*, v. 88, n. 3, p.295-298, 1992.

ZEGURA, S.L. *et al.* High-resolution SNPs and microsatellite haplotypes point to a single recent entry of Native American Y chromosomes into the Americas. *Mol. Biol. Evol.*, v. 21, n. 1, p. 164-175, 2004.

ANEXO A - Supergrupos linguísticos e subdivisões. Adaptado de Greenberg, Joseph H. (1966) *The Languages of Africa* (2nd ed), com adições e correções.

Apresentados de forma hierárquica:

I. Congo–Kordofanian

I.A Niger–Congo

I.A.1 West Atlantic

I.A.1.a Northern: Wolof, Serer-Sin, Fulani, Serer- Non,
Konyagi, Basari, Biafada, Badyara (Pajade), Dyola, Mandyak, Balante, Banyu
n, Nalu, Cobia, Cassanga, Bidyogo

I.A.1.b Southern: Temne, Baga, Landoma, Kissi, Bulom,
Limba, Gola

I.A.2 Mande

I.A.2.a Western

I.A.2.a.1 Malinke, Bambara, Dyula, Mandinka, Numu,
Ligbi, Huela, Vai, Kono, Koranko, Khasonke

I.A.2.a.2 Bobo

I.A.2.a.3 Mende, Loko, Gbandi, Loma, Kpelle (Guerze)

I.A.2.a.4 Susu, Dyalonke

I.A.2.a.5 Soninke, Bozo

I.A.2.a.6 Duun, Dzuun, Jo, Seenku (Sembla), Kpan, Banka

I.A.2.b Eastern

I.A.2.b.1 Mano, Dan (Gio), Guro (Kweni), Mwa, Nwa,
Beng, Gban, Tura (Wen), Yaure

I.A.2.b.2 Samo, Bisa, Busa, Kyenga, Shanga

I.A.3 Voltaic

I.A.3.a Senoufo: Minianka, Tagba, Foro, Tagwana
(Takponin), Dyimini, Nafana

I.A.3.b Lobi, Dyan, Puguli, Gan, Gouin, Turuka, Doghosie,
Doghosie- Fing, Kyan, Tara, Bwamu, Wara, Natioro,
Kulango

I.A.3.c Grusi: Awuna, Kasena, Nunuma, Lyele, Tamprusi,

- Degha, Siti, Kurumba (Fulse), Sisala
 I.A.3.d Mossi, Dagomba, Kusasi, Nankanse, Talensi,
 Mamprusi, Wala, Dagari, Birifo, Namnam, Kanjaga (Bulea) (moved from
 group c)
- I.A.3.e Tem, Kabre, Delo, Chala
 I.A.3.f Bargu (Bariba)
 I.A.3.g Gurma, Tobote (Basari), Kasele (Chamba), Moba
 I.A.3.x Dogon
- I.A.4 Kwa
- I.A.4.a Kru: Bete, Bakwe, Grebo, Bassa, De, Kru (Krawi)
 I.A.4.b Avatime, Nyangbo, Tafi, Logba, Likpe, Ahlo, Akposo, Lefana, Bowili, A
 kpafu, Santrokofi, Adele, Kebu, Anyimere, Ewe, Aladian, Avikam, Gwa, Kyama, Aky
 e, Ari, Abe,
 Adyukru, Akan (Twi, Anyi, Baule, Guang, Metyibo, Abure), Ga, Adangme
- I.A.4.c Yoruba, Igala
 I.A.4.d Nupe, Gbari, Igbira, Gade
 I.A.4.e Bini, Ishan, Kukuruku, Sobo
 I.A.4.f Idoma, Agatu, Iyala
 I.A.4.g Ibo
 I.A.4.h Ijo
- I.A.5 Benue–Congo
- I.A.5.A Plateau
- I.A.5.A.1
- I.A.5.A.1.a Kambari, Dukawa, Dakakari, Basa,
 Kamuku, Reshe
- I.A.5.A.1.b Piti, Janji, Kurama, Chawai, Anaguta,
 Buji Amap, Gure, Kahugu, Ribina, Butawa, Kudawa
- I.A.5.A.2 Afusare, Irigwe, Katab, Kagoro, Kaje, Kachicheri,
 Morwa, Jaba, Kamantan, Kadara, Koro, Afo
- I.A.5.A.3 Birom, Ganawuri (Aten)
- I.A.5.A.4 Rukuba, Ninzam, Ayu, Mada, Kaninkwom
- I.A.5.A.5 Eggon, Nungu, Yeskwa

- I.A.5.A.6 Kaleri, Pyem, Pai
- I.A.5.A.7 Yergam, Basherawa
- I.A.5.B Jukunoid: Jukun, Kentu, Nyidu, Tigong, Eregba, Mbembe, Zumper (Kutev, Mbarike), Boritsu
- I.A.5.C Cross-River
 - I.A.5.C.1 Boki, Gayi (Uge), Yakoro
 - I.A.5.C.2 Ibibio, Efik, Ogoni (Kana), Andoni, Akoiyang, Ododop, Korop
 - I.A.5.C.3 Akunakuna, Abine, Yako, Asiga, Ekuri, Ukelle, Okpoto-Mteze, Olulomo
- I.A.5.D Bantoid: Tiv, Bitare, Batu, Ndoro, Mambila, Bute, Bantu
- I.A.6 Adamawa–Eastern
 - I.A.6.A Adamawa
 - I.A.6.A.1 Tula, Dadiya, Waja, Cham, Kamu
 - I.A.6.A.2 Chamba, Donga, Lekon, Wom, Mumbake
 - I.A.6.A.3 Daka, Taram
 - I.A.6.A.4 Vere, Namshi, Kolbila, Pape, Sari, Sewe, Woko, Kotopo, Kutin, Durru
 - I.A.6.A.5 Mumuye, Kumba, Gengle, Teme, Waka, Yendang, Zinna
 - I.A.6.A.6 Dama, Mono, Mbere, Mundang, Yasing, Mangbei, Mbum, Kpere, La kka, Dek
 - I.A.6.A.7 Yungur, Mboi, Libo, Roba
 - I.A.6.A.8 Kam
 - I.A.6.A.9 Jen, Munga
 - I.A.6.A.10 Longuda
 - I.A.6.A.11 Fali
 - I.A.6.A.12 Nimbari
 - I.A.6.A.13 Bua, Nielim, Koke
 - I.A.6.A.14 Masa
 - I.A.6.B Eastern
 - I.A.6.B.1 Gbaya, Manja, Mbaka
 - I.A.6.B.2 Banda

I.A.6.B.3 Ngbandi, Sango, Yakoma

I.A.6.B.4 Zande, Nzakara, Barambo, Pambia

I.A.6.B.5 Bwaka, Monjombo, Gbanziri, Mundu, Mayogo,
Bangba

I.A.6.B.6 Ndogo, Bai, Bviri, Golo, Sere, Tagbo, Feroqe,
Indri, Mangaya, Togoyo

I.A.6.B.7 Amadi (Madyo, Ma)

I.A.6.B.8 Mondunga, Mba (Bamanga)

I.B Kordofanian

I.B.1 Koalib: Koalib, Kanderma, Heiban, Laro, Otoro, Kawama,
Shwai, Tira, Moro, Fungor

I.B.2 Tegali: Tegali, Rashad, Tagoi, Tumale

I.B.3 Talodi: Talodi, Lafofa, Eliri, Masakin, Tacho, Lumun, El Amira

I.B.4 Tumtum: Tumtum, Tuleshi, Keiga, Karondi, Krongo,
Miri, Kadugli, Katcha

I.B.5 Katla: Katla, Tima

II. Nilo-Saharan

II.A Songhai

II.B Saharan

II.B.a Kanuri, Kanembu

II.B.b Teda, Daza

II.B.c Zaghawa, Berti

II.C Maban: Maba, Runga, Mimi of Nachtigal,
Mimi of Gaudefroy-Demombynes

II.D. Fur

II.E. Chari–Nile

II.E.1 Eastern Sudanic

II.E.1.1 Nubian

II.E.1.1.a Nile Nubian (Mahas-Fadidja and
Kenuzi-Dongola)

II.E.1.1.b Kordofanian Nubian: Dair, Dilling, Gulfan,

Garko, Kadero, Kundugr

II.E.1.1.c Midob

II.E.1.1.d Birked

II.E.1.2 Murle (Beir), Longarim, Didinga, Suri, Mekan,
Murzu, Surma (including Tirma and Zulmanu), Masongo

II.E.1.3 Barea

II.E.1.4 Ingassana (Tabi)

II.E.1.5 Nyima, Afitti

II.E.1.6 Temein, Teis-um-Danab

II.E.1.7 Merarit, Tama, Sungor

II.E.1.8 Dagu of Darfur, Baygo, Sila,

Dagu of Dar Dagu (Wadai),

Dagu of Western Kordofan, Njalgulgule, Shatt, Liguri

II.E.1.9 Nilotic

II.E.1.9.a Western

II.E.1.9.a.1 Burun

II.E.1.9.a.2 Shilluk, Anuak, Acholi, Lango, Alur, Luo, Jur, Bor

II.E.1.9.a.3 Dinka, Nuer

II.E.1.9.b Eastern

II.E.1.9.b.1 Bari, Fajulu, Kakwa, Mondari

II.E.1.9.b.2a Jie, Dodoth, Karamojong, Teso,

Topotha, Turkana

II.E.1.9.b.2b Masai

II.E.1.9.b.3 Southern: Nandi, Suk, Tatoga

II.E.1.10 Nyangiya, Teuso

II.E.2 Central Sudanic

II.E.2.1 Bongo, Baka, Morokodo, Beli, Gberi,

Sara dialects (Madjinngay, Gulai, Mbai, Gamba, Kaba,
Dendje, Laka), Vale, Nduka, Tana, Horo, Bagirmi, Kuka,

Kenga, Disa, Bubalia

II.E.2.2 Kreish

II.E.2.3 Binga, Yulu, Kara [= Tar Gula]

II.E.2.4 Moru, Avukaya, Logo, Keliko, Lugbara, Madi

II.E.2.5 Mangbetu, Lombi, Popoi, Makere, Meje, Asua

II.E.2.6 Mangbutu, Mamvu, Lese, Mvuba, Efe

II.E.2.7 Lendu

II.E.3 Berta

II.E.4 Kunama

II.F Koman/Coman: Komo, Ganza, Uduk, Gule, Gumuz, Mao

III. Afroasiatic[edit]

III. Afro-Asiatic

III.A Semitic

III.B Egyptian

III.C Berber

III.D Cushitic

III.D.1 Northern Cushitic: Beja (Bedauye)

III.D.2 Central Cushitic: Bogo (Bilin), Kamir, Khamta, Awiya, Damot, Kemant, Kayla, Quara

III.D.3 Eastern Cushitic: Saho-Afar,

Somali, Galla, Konso, Geleba, Marille, (Reshiat, Arbore), Gardula,

Gidole, Gowaze, Burji, Sidamo, Darasa, Kambata, Alaba, Hadya, Tambaro, M

ogogodo (added 1966)

III.D.4 Western

Cushitic: Janjero, Wolamo, Zala, Gofa, Basketo, Baditu, Haruro,

Zaysse, Chara, Gimira, Benesho, Nao, Kaba, Shako, She, Maji, Kafa, Garo,

Mocha, Anfillo (Mao), Shinasha, Bako, Amar, Bana,

Dime, Gayi, Kerre, Tsamai, Doko, Dollo

III.D.5 Southern

Cushitic: Burungi (Mbulungu), Goroa (Fiome), Alawa (Uwassi),

Iraqw, Mbugu, Sanye [= Dahalo], Ngomvia (added 1966)

III.E Chad

III.E.1

III.E.1.a Hausa, Gwandara

III.E.1.b Ngizim, Mober [= Kanuri, not Chadic], Auyokawa,

Shirawa, Bede

III.E.1.c

III.E.1.c.i Warjawa, Afawa, Diryawa, Miyawa, Sirawa

III.E.1.c.ii Gezawa, Seiyawa, Barawa of Dass

III.E.1.d

III.E.1.d.i Bolewa, Karekare, Ngamo, Gerawa,

Gerumawa, Kirifawa, Dera (Kanakuru), Tangale, Pia, Pero, Chongee, Maha (added 1966)

III.E.1.d.ii Angas, Ankwe, Bwol, Chip, Dimuk, Goram, Jorto, Kwolla, Miriam, Montol, Sura, Tal, Gerka

III.E.1.d.iii Ron

III.E.2 Kotoko group:

Logone, Ngala [= Mpade?], Buduma, Kuri, Gulfei, Affade, Shoe, Kuseri

III.E.3 Bata–Margi group

III.E.3.a Bachama, Demsa, Gudo, Malabu, Njei (Kobochi, Nzangi, Zany), Zumu (Jimo), Holma, Kapsiki, Baza, Hiji, Gude (Cheke), Fali of Mubi, Fali of Kiria, Fali of Jilbu, Margi, Chibak, Kilba, Sukur, Vizik, Vemgo, Woga, Tur, Bura, Pabir, Podokwo

III.E.3.b Gabin, Hona, Tera, Jera, Hinna (Hina)

III.E.4

III.E.4.a Hina, Daba, Musgoi, Gauar

III.E.4.b Gisiga, Balda, Muturua, Mofu, Matakam

III.E.5 Gidder

III.E.6 Mandara, Gamergu

III.E.7 Musgu

III.E.8 Bana, Banana (Masa), Lame, Kulung

III.E.9

III.E.9.a Somrai, Tumak, Ndam, Miltu, Sarwa, Gulei [= Tumak?]

III.E.9.b Gabere, Chiri, Dormo, Nangire

III.E.9.c Sokoro (Bedanga), Barein

III.E.9.d Modgel

II.E.9.e Tuburi

III.E.9.f Mubi, Karbo, (added 1966: Jegu, Jonkor,
Wadai-Birgid)

IV Khoisan[

IV. Khoisan

IV.A South African Khoisan

IV.A.1 Northern South African Khoisan

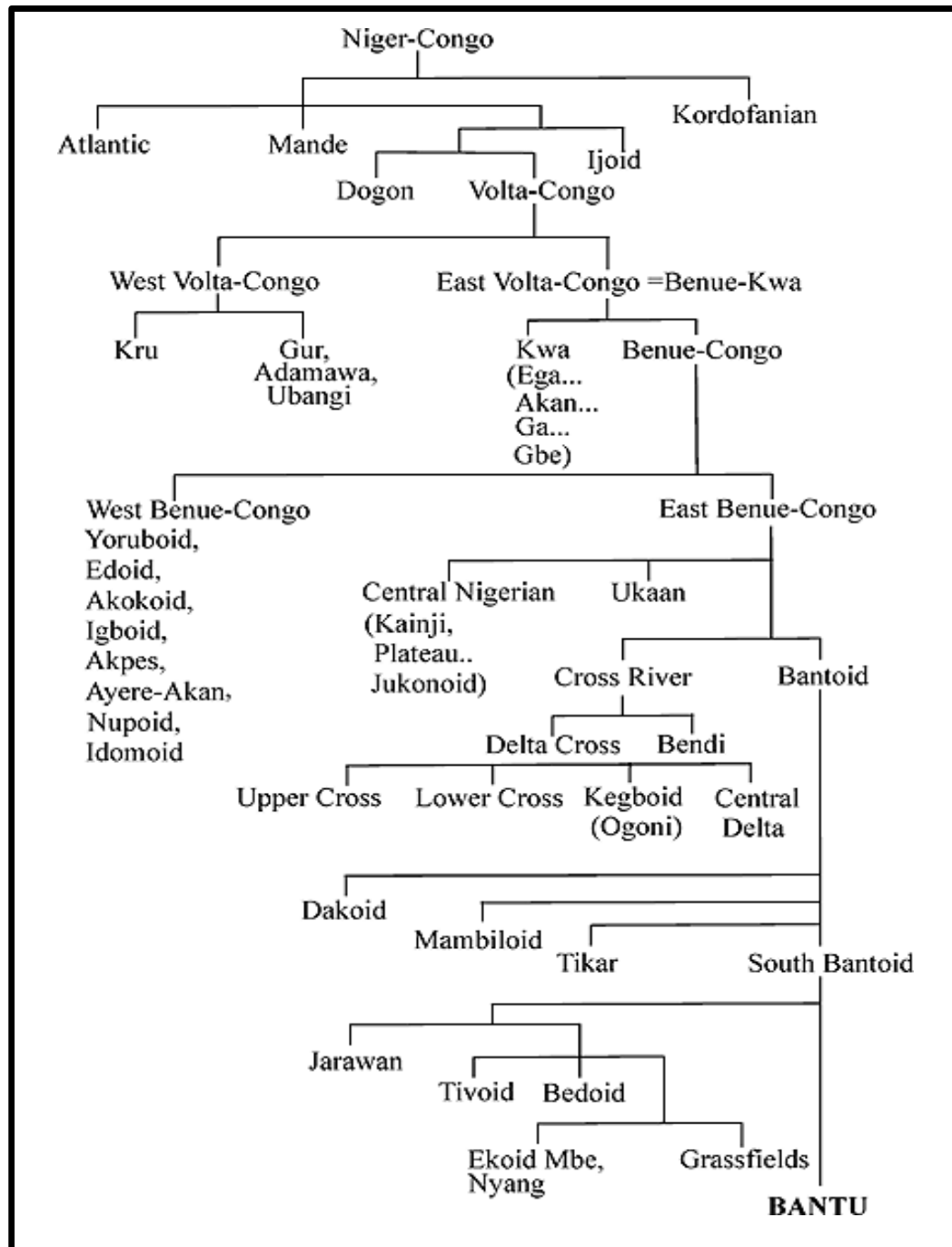
IV.A.2 Central South African Khoisan

IV.A.3 Southern South African Khoisan

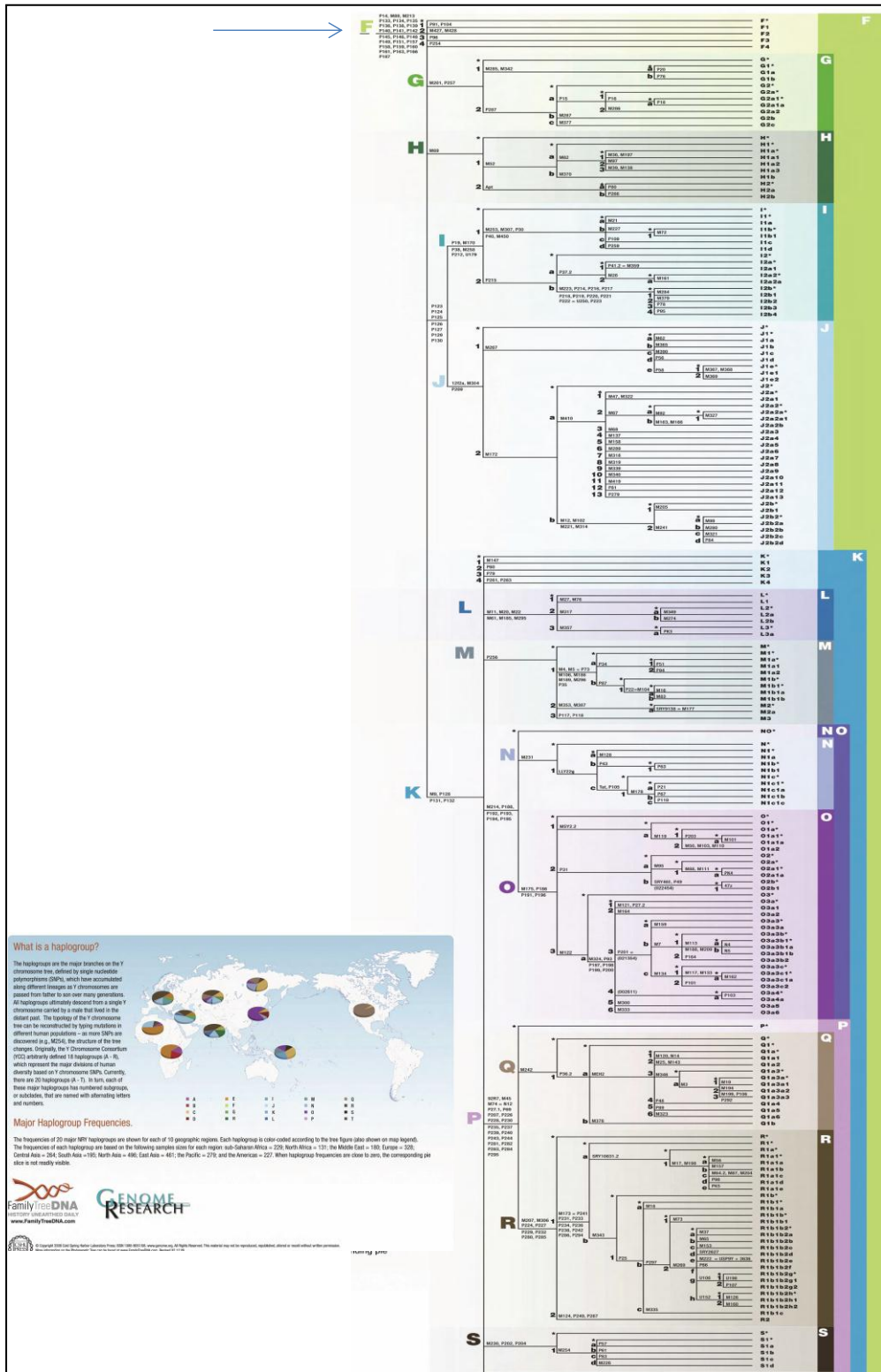
IV.B Sandawe

IV.C Hatsa

ANEXO B - Representação esquemática da divisão do grupo Niger-Congo (grupo etnolinguístico dos *Bantu-Speakers*)

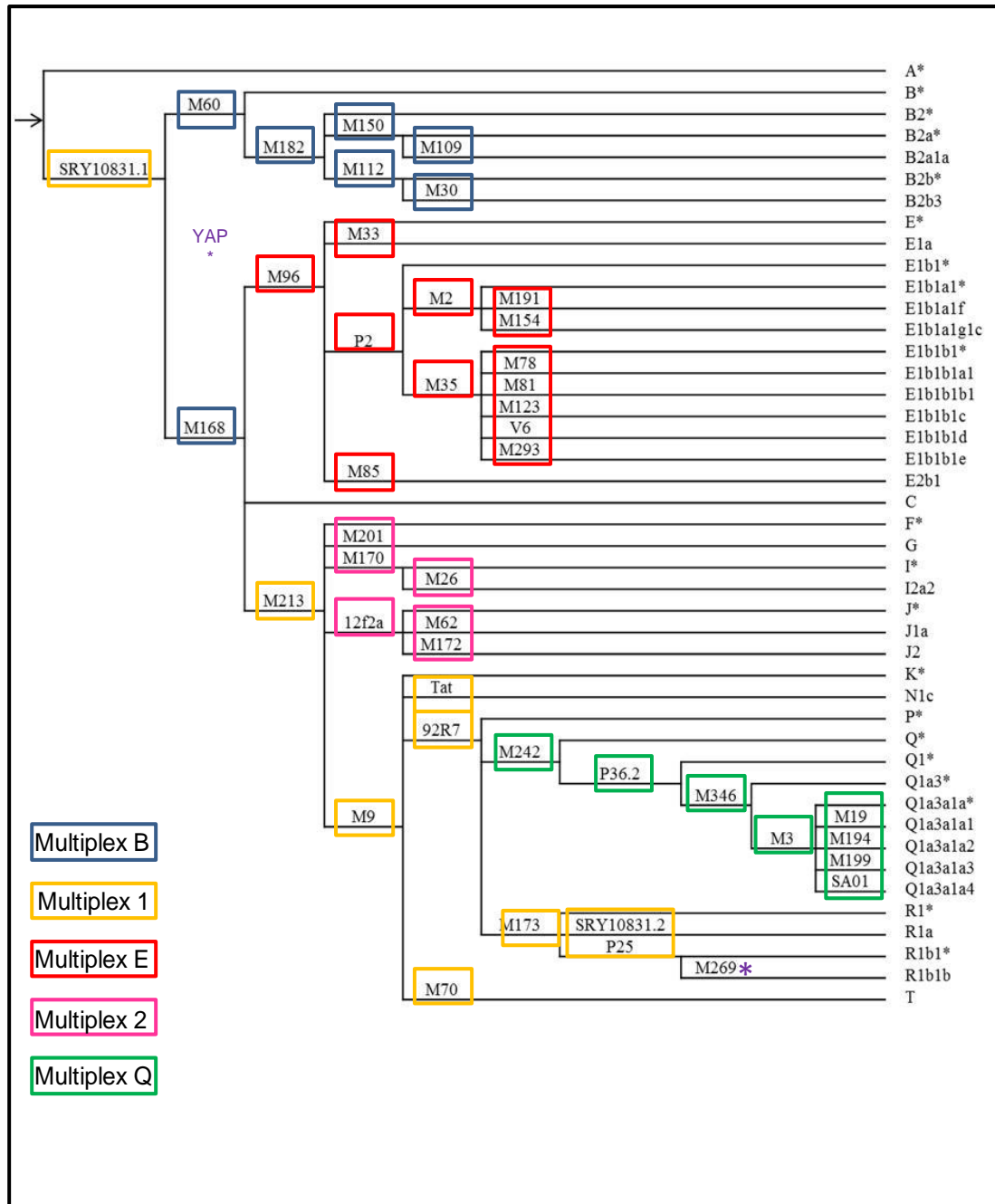


Fonte: Exploring the Niger-Congo Languages



Fonte: KARAFET, 2008

ANEXO D - Representação esquemática dos 46 Y-SNPs escolhidos e utilizados em 5 diferentes multiplexes para a determinação dos haplogrupos das populações estudadas



Legenda: YAP e M269 foram tipados em reações “singleplex”, como descrito em material e métodos.

ANEXO E - Oligonucleotídeos para a amplificação das regiões alvo e tamanho dos produtos de PCR dos multiplexes utilizados neste estudo

Y-SNP	Primer forward (5' – 3')	Primer reverse (5' – 3')	PCR Fragment Size	PCR-Amplif
SRY10831.1/2, M213, M9, Tat, 92R7, M70, M173, P25	Brion et al. (2004)	Brion et al. (2004)	-	Multiplex 1, sem 22 (Brion et al., 2004), mais M13
M13	TCCTAACCTGGTGGTCTTTCA	GCCATGATTTTATCCAACCAC	230	
M168	GGAGTATGTGTGGAGGTGAGT	CATCTCTTACCCAAACTGCTAAA	72	
M60	CCAACACTGAGCCCTGATG	GAGAAGGTGGGTGGTCAAGA	215	
M182	TTCAAAGACTTAAAGCAGTGGTTA	TGTGCCATCACACCATCTTT	176	
M150	AAGTGAGACTGGGCTTTGGA	AAGGAAGGGGAACAGAAAGGA	235	MULTIPLEX B
M109	GTCCCAACTCTGCAACAAT	GGGTATCAAATGTCTTCAACCT	96	
M112	AAAAGCAAAGAGAATGCCTCT	TTCAATTCTGTCTGTGCAGAA	160	
M30	GTGTGGTAGACAGAACAGCAG	GCACAGCCAGATAACCCCTAC	155	
M96	Brion et al. (2004)	Brion et al. (2004)	88	
M33	CACAACCTCATGGCTACGG	GTTGAAGCCCCAAGAGAGAC	190	
P2	GCTCCAGCCATCTTTTCCTTA	CTTCTCTCATGAGGGTTTTGGA	180	
M2	AAGTCCAGACCCAGGAAGGT	ACAGCTCCCCCTTATCCTC	162	
M154	TACTCACACAAACCAAGAAGAAACA	AACCATTGTGTTACATGGCCTA	130	
M191	AAAAATGGAGTGTATCAGAGCTT	CCCAGACACACAAAATATCTC	122	MULTIPLEX E
M35	Brion et al. (2004)	Brion et al. (2004)	198	
M78	GGATGGCTGTATGGGTTTCT	Brion et al. (2004)	235	
M81	GCACTATCATACTCAGCTACACATCTC	AACCATTGTGTTACATGGCCTA	203	
M123	TGCTCTCAGGGGAAAATCTG	AGCAAAGTTGAGGTTGCACA	213	
V6	GATGGCACAGTGTTCGACAG	CTTCTCTCCAAATGCCTGCT	102	
M293	AAAGAGATTGATCGGTGCATA	GCTGGCTAATACTCCACAGAG	230	
M85	TGGCATCCAATACTAGCTGATAAAC	AATGCTCACGCTTGTGTCT	283	
M242	TTGTGCAAAAAGGTGACCAA	TTTCGCTTTAAGGGCTTTCA	155	
P36.2	GAGGAGGGGAGAGAGAAAA	TTCAAACAGCCCACCAGATA	299	
M346	GGCCTGAAAATGTGAAAGA	AGCCTGGGAGACAGACACAG	247	
M3	CATTAAAGCCGGTCACAGGT	CTGCCAGGGCTTTCAAATAG	304	MULTIPLEX Q
M19	TCACCAGAGTTGTGGTTGC	ACAGACACAAAGGGCCAAC	277	
M194	TTACCTGAACCTGGGAGGTG	CATGTAATGACCGCCACTG	338	
M199	CACTACAGGCCTGGTTGGAT	TGCAAGCTGTGTGATTTCAAG	216	
12f2	CACTGACTGATCAAAATGCTTACAGAT	GGATCCCTTCTTACACCTTATAC	90	
M201	TCAAATTGTGACTGCAATAGTT	CATCCAACACTAAGTACCTATTAC	144	
M26	AGCAGAAGAGACCAAGACAGC	GACGAAATCTGCAGCAAAAA	147	MULTIPLEX 2
M170	TGCAGCTTTATTAAGTTATGTTTTCA	CCAATTACTTTCAACATTTAAGAC	158	
M172	TCCTCATTACCTGCCTCTC	TCCATGTTGGTTTGAACAG	187	
M62	ACTAAAACACCATTAGAAACAAAGG	CTGAGCAACATAGTGACCCC	309	
YAP	Hammer et al. (1995)	Hammer et al. (1995)	150/455	SINGLEPLEX

ANEXO F - Oligonucleotídeos para a detecção dos SNPs por SBEs com a concentração utilizada e tamanho dos fragmentos

Y-SNP	oligonucleotídeo SBE (5' a 3')	Concentração na reação [µM]	SBE tamanho do fragmento	Multiplex
SRY10831.1/2, M213, M9, Tat, 92R7, M70, M173, P25	Brion et al. (2004)	-	-	Multiplex 1, sem M22 (Brion et al. 2004) mais M13
M13	AAGGCAAGACGGTTA	0.36	16	
M60	ACCACTGTGTCCTGAT	0.35	18	
M112	GAGGTGAGATAAAAACAAAGCAGT	0.07	24	
M109	AATCAAGGAAGTCAAGTCATTCCAAA	0.22	26	
M182	TCTGACAAGCAGTGGTTAATGTAACAAA	0.35	29	MULTIPLEX B
M168	AAGTCTGACAAGGAGTATGTGTTGGAGGTGAGT	0.28	33	
M30	GTCGTGAAAGTCTGACAACACCTTTTCCCAATATGATAATT	0.44	41	
M150	GTGCCACGTCGTGAAAGTCTGACAAGCCACACACACAGATAGAAGT	0.29	47	
M96	Brion et al. (2004)	0.03	40	
M33	GTGCCACGTCGTGAAAGTCTGACAACAGTTACAAAAGTATAATATGTCTGAGAT	0.10	54	
P2	GCCCCTAGGAGGAGAA	0.04	16	
M2	GTGCCACGTCGTGAAAGTCTGACAATTTATCTCCACAGATCTCA	0.08	45	
M154	AAACATGGCCTATAATATTCAGTACA	0.04	26	
M191	GGTGCCACGTCGTGAAAGTCTGACAACATTTTTTTCTTTACAACCTTGACTA	0.11	51	MULTIPLEX E
M35	CCCCCCCCCCCCCCCCAGTCTCTGCCTGTGTC	0.44	36	
M78	Brion et al. (2004)	0.44	34	
M81	Brion et al. (2004)	0.44	27	
M123	TAGGTGCCACGTCGTGAAAGTCTGACAATTCTAGGTATTCAGGCGATG	0.11	48	
V6	GCCACGTCGTGAAAGTCTGACAATGCTGTGATTCTGATGTG	0.07	42	
M293	AAAGAGATTGATCGGTGCATA	0.04	21	
M85	CTTGTGTCTATTAAGTGTAGTTTGTAG	0.04	30	
M242	AAAAAAGGTGACCAAGGTGCT	0.1	23	
P36.2	GTCTGACAACATCTATCTATCCATTATTCTCTCT	0.2	35	
M346	CTGACAACAGCCAAGAGGACAGTAAGA	0.2	28	
M3	TCACCTCTGGGACTGA	0.4	17	MULTIPLEX Q
M19	TGACAAGTAGAGACATCTGAAACCCAC	0.05	28	
M194	ACAATGCCTTCTCGGGGGAA	0.4	22	
M199	GTCGTGAAAGTCTGACAATTTGTCAGTGTAGCAGTGG			
SA01	GTCTGACAAATGTTAAATGGCTTACACTTG	0.2	32	
12f2	Visualização em gel	-	-	
M201	AAGTGACAATAATAATCCAGTATCAACTGAGG	0.05	42	
M26	GCCACGTCGTGAAAGTCTGACAATAGGCCATTACAGTGTCTCTG	0.25	37	
M170	ACACAACCCCACTGAAAAAAA	0.45	22	MULTIPLEX 2
M172	AAACCCATTTTGTGCTT	0.10	32	
M62	CTGACAACAATGTTTGTGGCCATGGA	0.50	27	