



UNIVERSIDADE DE BRASÍLIA - UnB
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA ANIMAL

**ANÁLISE DE DIFERENCIAÇÃO INTRA-POPULACIONAL
EM AMOSTRAS DE DIFERENTES ESTRATOS
SÓCIO-ECONÔMICOS DO DISTRITO FEDERAL**

Gustavo de Carvalho Dalton

Brasília
2010



UNIVERSIDADE DE BRASÍLIA - UnB
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA ANIMAL

**ANÁLISE DE DIFERENCIAÇÃO INTRA-POPULACIONAL
EM AMOSTRAS DE DIFERENTES ESTRATOS
SÓCIO-ECONÔMICOS DO DISTRITO FEDERAL**

Tese apresentada ao Programa de Pós-Graduação em Biologia Animal do Instituto de Ciências Biológicas da Universidade de Brasília, como requisito parcial para a obtenção do título de Doutor em Ciências Biológicas, área de concentração: Genética.

Gustavo de Carvalho Dalton

Orientadora: Prof^a. Dr^a. Silviene Fabiana de Oliveira

Brasília

2010



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA ANIMAL

Tese de Doutorado

Gustavo de Carvalho Dalton

Título:

**“ANÁLISE DE DIFERENCIAÇÃO INTRA-POPULACIONAL EM AMOSTRAS DE
DIFERENTES ESTRATOS SÓCIO-ECONÔMICOS DO DISTRITO FEDERAL”**

Comissão Examinadora:

Prof^ª. Dr^ª. Silvíene Fabiana de Oliveira
Presidente / Orientadora
Universidade de Brasília

Prof. Dr. Celso Teixeira Mendes Júnior
Membro Titular Externo não Vinculado ao Programa
USP – Ribeirão Preto

Prof^ª. Dr^ª. Maria de Nazaré Klautau Guimarães
Membro Titular Interno Vinculado ao Programa
Universidade de Brasília

Prof. Dr. Luzitano Brandão Ferreira
Membro Titular Externo não Vinculado ao Programa
UniCEUB / DF

Prof^ª. Dr^ª. Juliana Forte Mazzeu Araújo
Membro Titular Interno não Vinculado ao Programa
Universidade de Brasília

Prof. Dr. Antônio Francisco Pereira de Araújo
Membro Suplente Interno não Vinculado ao Programa
Universidade de Brasília

Brasília, setembro de 2010.

Dedico esse trabalho aos meus pais, exemplo que sempre foram para minha formação. Ao meu pai, eterno aprendiz com quem descobri que aprender nunca cansa e nos mantém jovens de mente e espírito. À minha mãe, pela inquebrantável fé em Deus e em sua família, zelando sempre por todos nós.

Agradecimentos

Agradeço a todos que, de uma forma ou outra, contribuíram para mais este passo de minha formação, e em especial:

À professora Silviene, pela orientação, ensinamentos e companheirismo, com quem dividi dificuldades e conquistas.

Aos componentes da banca, que tanto contribuíram para a conclusão deste trabalho, com seus ensinamentos e considerações tão importantes: Nazaré, Luzitano, Celso e Juliana.

Ao estimado professor Antonio Francisco, o Chicão, meu amigo da graduação e mestrado, que sempre acompanhou minhas conquistas acadêmicas, e tão desprendidamente possibilitou que a apresentação final fosse possível – a você e toda sua família, muito obrigado.

Aos professores e pesquisadores Dario Grattapaglia e Sídia Callegari-Jacques, por terem disponibilizado as informações necessárias às comparações genotípicas, sem os quais muito deste trabalho não teria sido possível.

Aos professores do curso de Ciências Biológicas que opinaram, sugeriram e acompanharam as etapas necessárias à conclusão desse trabalho: Nazaré, César, Zulmira, Rosana, Nilda, César e Sônia.

Aos companheiros de curso, que sempre dividiram mais que o tempo com nossa orientadora, ajudando na organização de aulas, cursos, exposições e prateleiras: Carol, Diana, Arcanjo, Guilherme, Rafaela, Gabriel, Gegê, Ricardo,... e todos os demais que fizeram parte do grupo nestes anos.

À turma tão especial do Laboratório de Genética, que sempre ajudou para um alto astral e ótimo ambiente de convívio: Arthur, Ana Luísa, Bélin, Cássia, Graciana, Lane, ... e tantos outros que dividem seu tempo na UnB com a família e o trabalho.

Aos incansáveis funcionários do Laboratório de Genética, responsáveis por aquela rotina que possibilita o funcionamento de tudo: Ornil, Eliza, Adriana, Danúbia, ...

Aos inúmeros colegas de curso, com quem convivi nestes anos em salas de aula, laboratórios, e nos produtivos debates após o horário de aulas.

A todos os meus amigos que, mesmo distantes, sei que torceram muito pela conclusão e sucesso deste trabalho: Georgios, Spartaco, Renata, Márcio, Ricardo, Roberto, ...

Aos peritos criminais companheiros de pós-graduação, que conhecem mais do que ninguém as dificuldades de conciliar trabalho e estudo: Mila, Ramalho, Rejane, Neide, Guilherme e Büch.

Aos colegas de profissão que de diversas maneiras auxiliaram nesta empreitada, principalmente pela compreensão no tempo necessário à sua conclusão.

Aos sempre presentes amigos da Associação Brasiliense de Peritos em Criminalística, pelo apoio e paciência na divisão de horários que este curso exigiu: Iremar, Dércio, Patrícia, July e Rosi.

À Polícia Civil do DF, pelas oportunidades de desenvolvimento deste trabalho e por considerar a importância de profissionais academicamente qualificados para a excelência do mister pericial.

Aos meus sogros, cunhados, tios, primos e sobrinhos, que participaram com seus pensamentos, orações e corações desta empreitada.

À minha esposa e meu filho, pela imensa compreensão e o imprescindível estímulo para terminar este trabalho. Sem seu apoio não teria sido possível sua conclusão.

Aos meus pais e irmãos.

A Deus.

“Many inquiries have already been undertaken into interracial conflicts and the factors that produce them. The time has now come for us to consider the societies which have in large measure succeeded in resolving antagonisms by overriding racial differences. Thus, the General Conference of Unesco in Florence recommended for the 1951 programme of the Organization a study of racial relations in Brazil. This great republic has a civilization which has been developed by the direct contributions of different races. And it suffers less than other nations from the effects of those prejudices which are at the root of so many vexatious and cruel measures in countries of similar ethnic composition. We are as yet illinformed about the factors which brought about such a favourable and, in many ways, exemplary situation. But in the present state of the social sciences, general speculations no longer suffice. We must have specialists make searching inquiries in the field. We must learn from them exactly why and how social, psychological and economic factors have contributed in varying degrees to make possible the harmony which exists in Brazil. Then the results of their inquiries can be set forth in publications in order to stimulate those who are still struggling elsewhere to introduce more peaceable and happier inter-racial relations.”

“The Race Question”, UNESCO, 1950.

RESUMO

A história do povoamento do Brasil teve relação direta com a composição atual de sua população. Inicialmente com a presença indígena (ameríndia), e posteriormente com a colonização de europeus (notadamente da península ibérica) e contribuição africana decorrente da política escravagista, a população resultante traz a participação básica destes três grupos parentais. O histórico de ocupação mais recente do Centro-Oeste, com o deslocamento de migrantes de diversos outros Estados brasileiros, tem sugerido que esta região apresente um retrato genético da população do país. Um questionamento acerca dos estudos populacionais é se a amostra utilizada em um dado trabalho representa a população pretendida. Amostras visando uma descrição populacional buscam adotar condições de aleatoriedade, porém devido às singularidades de coleta questiona-se se diferentes amostras da mesma população podem refletir alguma estratificação nela existente. Outra questão é se perfis definidos por marcadores considerados neutros, como os STR autossômicos do sistema CODIS, fornecem informações sobre a ancestralidade de um dado indivíduo. Dessa forma, o objetivo geral desse trabalho foi investigar se amostras coletadas em uma mesma localidade geográfica, porém de estratos sócio-econômicos distintos, apresentam diferenças quanto à caracterização genética para marcadores genéticos do tipo STR autossômicos e se perfis genéticos gerados para esses marcadores fornecem informações sobre ancestralidade dos indivíduos em análise. Para tanto, foram utilizadas duas amostras do Distrito Federal, uma oriunda de pessoas amostradas em laboratório público e que apresentam menor poder aquisitivo e outra de um laboratório privado, onde os indivíduos apresentam um maior poder aquisitivo. Análises de diferenciação genética mostraram que essas duas amostras, assim como a comparação dessas com outras localidades brasileiras, não apresentam diferença quanto à constituição genética para 12 marcadores STR autossômicos. Entretanto, análises de contribuição genética dos grupos parentais na constituição das duas amostras, assim como a análise de estrutura populacional (STRUCTURE) mostraram diferenças. A amostra de origem de laboratório privado apresentou uma maior contribuição europeia (77,7%) do que a do laboratório público (61,2%). Por outro lado, a contribuição africana e ameríndia foram maiores no público (26,6% e 12,2%, respectivamente) em contraste com o

privado (13,2% e 9,0%, respectivamente). Análise obtida com o auxílio do programa STRUCTURE corrobora essa visão. Quanto à análise de perfis genéticos obtidos por STR autossômicos descritos como neutros serem ou não informativos de ancestralidade em população miscigenada, foram utilizadas as planilhas OMNIPOP e MPGO/EHSTRAFD. Foi possível identificar um potencial de indicação de ancestralidade a partir de alelos sugeridos no presente trabalho como indicativo de ancestralidade europeia, africana ou ameríndia. Deste modo, foram estabelecidos alelos e genótipos característicos de determinadas origens geográficas ou parentais, estabelecendo-se que os marcadores autossômicos do sistema CODIS, notadamente cinco deles (TPOX, D5S818, TH01, D13S317 e FGA), podem ser utilizados como indicadores de ancestralidade.

Palavras-chave: genética de populações, STR autossômicos, mistura genética, subestruturação populacional, estruturação sócio-econômica, população do Distrito Federal, marcadores indicativos de ancestralidade.

ABSTRACT

The history of the populating of Brazil had a direct relation with the current composition of its population. Originally with indigenous (Amerindian), and subsequently with the colonization of Europeans (mainly from the Iberian peninsula) and the African contribution arising from slavery politics, the population shares the involvement of these three basic groups of parents. The history of more recent occupation of the Midwest, with the displacement of many migrants from other Brazilian states, has suggested that this region presents a portrait of the genetic population. A questioning of population studies is whether the sample used in a given work represents the intended population. Sampling aiming at a description of population seeks to adopt conditions of randomness, but due to the singularity of collecting methods there is a questioning whether different samples of the same population may reflect some stratification that might exist within it. Another question is whether profiles defined by markers allegedly neutral, such as autosomal STR CODIS system, provide information about the ancestry of a given individual. Thus, the general objective of this study was to investigate whether the samples collected in the same geographic location, but in different socio-economic strata, show differences in genetic characterization for genetic markers of autosomal STR type and if genetic profiles generated by these markers provide information on ancestry of these individuals in the analysis. To this end, we used two samples of the Federal District, one coming from people sampled in a public laboratory with a lower purchasing power and another from a private laboratory, where individuals have a higher purchasing power. Analysis of genetic differentiation showed that these two samples, as well as the comparison among other locations in Brazil, have no difference in the genetic constitution for 12 autosomal STR markers. However, analysis of the genetic contribution of parental groups in the constitution of the two samples, as well as analysis of population structure (STRUCTURE) showed differences. The sample source from the private laboratory showed a greater European contribution (77.7%) than the one from the public laboratory (61.2%). On the other hand, African and Amerindian contributions were higher in public (26.6% and 12.2% respectively) in contrast with the private laboratory (13.2% and 9.0% respectively). Analysis obtained with the help of the STRUCTURE program corroborates this view. As for the analysis of genetic profiles obtained by autosomal

STR described as neutral being or not ancestry informative in a mixed population, we used OMNIPOP and MPGO/EHSTRAFD spreadsheets. It was possible to identify a potential indication of ancestry from allele suggested in this study as indicative of European, African or American ancestry. Thus, alleles and genotypes characteristic of certain geographic origins or parents were established, indicating that the CODIS system autosomal markers, especially five of them (TPOX, D5S818, TH01, D13S317 and FGA), can be used as indicators of ancestry.

Keywords: population genetics, autosomal STR markers, genetic admixture, population substructure, socio-economic structure, Federal District population, ancestry indicative markers.

SUMÁRIO

RESUMO	7
ABSTRACT	9
1. INTRODUÇÃO	12
1.1 MARCADORES MOLECULARES E SUA UTILIZAÇÃO NA GENÉTICA HUMANA	14
1.2 POVOAMENTO DO BRASIL	19
1.2.1 SITUAÇÃO DEMOGRÁFICA BRASILEIRA	22
1.2.2 ASSOCIAÇÃO ENTRE ASPECTOS SOCIAIS E COR OU RAÇA	25
1.2.2.1 Reflexos da influência social em análises genéticas de composição étnica.....	28
2. JUSTIFICATIVAS E OBJETIVOS	32
3. MATERIAL E MÉTODOS.....	34
3.1 AMOSTRAS POPULACIONAIS DO DISTRITO FEDERAL	34
3.1.1 AMOSTRA ORIUNDA DE LABORATÓRIO PÚBLICO DO DISTRITO FEDERAL	34
3.1.2 AMOSTRA DE LABORATÓRIO PRIVADO DO DISTRITO FEDERAL	35
3.1.3 DIFERENÇAS SÓCIO-ECONÔMICAS ENTRE OS INDIVÍDUOS QUE COMPÕEM AS DUAS AMOSTRAS REPRESENTATIVAS DO DISTRITO FEDERAL.....	36
3.1.4 AMOSTRAS POPULACIONAIS DE OUTROS ESTADOS BRASILEIROS.....	37
3.2 DADOS DEMOGRÁFICOS.....	38
3.3 ANÁLISE ESTATÍSTICA	38
3.3.1 ANÁLISES UTILIZANDO FREQUÊNCIAS ALÉLICAS	39
3.3.2 ANÁLISES UTILIZANDO DADOS GENOTÍPICOS	40
3.4 APLICAÇÕES À INFERÊNCIA POPULACIONAL	41
4. RESULTADOS E DISCUSSÃO	44
4.1 LABORATÓRIOS PÚBLICO E PRIVADO: CARACTERIZAÇÃO GENÉTICA E COMPARAÇÕES	44
4.2 MISTURA GENÉTICA.....	53
4.2.1 ANÁLISE DE ESTRUTURA POPULACIONAL COM O PROGRAMA STRUCTURE	54
4.2.1.1 Populações parentais	54
4.2.1.2 Análise de estrutura populacional	55
4.2.2 ANÁLISE COMPARATIVA PELO PROGRAMA DISTRUCT	60
4.3 ALELOS E GENÓTIPOS CARACTERÍSTICOS DE POPULAÇÕES	63
4.3.2 CARACTERÍSTICAS POPULACIONAIS ASSOCIADAS À ORIGEM GEOGRÁFICA.....	68
4.3.3 PODER DE RESOLUÇÃO DE LOCOS E ALELOS NAS AMOSTRAS ANALISADAS	71
5. CONSIDERAÇÕES FINAIS	78
6. REFERÊNCIAS BIBLIOGRÁFICAS	80

1. INTRODUÇÃO

O desenvolvimento da genética de populações foi fundamental para a formulação da Teoria Sintética da Evolução (ou NeoDarwinismo), resultado da união das ideias de diversos cientistas, cuja base é a síntese entre a teoria inicialmente propostas por Charles Darwin e os conhecimentos advindos da genética. Foi a partir do melhor conhecimento da genética populacional que se iniciou a associação entre as duas áreas, permitindo a compreensão de como a variabilidade individual se associava às pressões evolutivas incidentes sobre uma determinada população.

A diversidade populacional (entre indivíduos de um mesmo grupo ou entre diferentes grupos), assim como sua composição genética, é avaliada a partir da distribuição das frequências alélicas de marcadores genéticos. As variações destas frequências permitem estimar o grau de divergência entre populações, variabilidade que pode ser decorrente de diversos fatores.

As análises em populações humanas vem demonstrando que a maior parte da variabilidade genética ocorre dentro das populações, tendo menor influência as diferenças inter-populacionais (Rosenberg *et al.*, 2002). No caso de populações humanas não isoladas do continente americano, a própria história do povoamento esclarece como essa alta variabilidade genética surgiu. A principal causa para isso é o fato de que ao menos três grupos populacionais geograficamente distintos foram reunidos: ameríndios nativos, imigrantes europeus (ibéricos, em sua maioria) e africanos subsaarianos (trazidos como escravos, em sua maioria). As frequências alélicas refletem o grau de miscigenação das populações parentais na composição do grupo formado (Hartl & Clark, 2006).

Dentre os fatores evolutivos que afetam a diferenciação genética entre populações, está exatamente a migração, um dos processos fundamentais no povoamento do continente americano. A introdução de alelos da população migrante para a receptora (fluxo gênico) em proporção diferente da original tende a estabelecer uma frequência alélica intermediária em relação às parentais, num processo de homogeneização (Wang & Caballero, 1999).

As primeiras populações humanas foram derivadas da expansão do homem moderno a partir da África subsaariana oriental, num processo que levou milhares de

anos. Com o surgimento dos meios de transporte, o homem tornou-se capaz de, em curtos intervalos de tempo, percorrer grandes distâncias geográficas. Apesar dos rápidos deslocamentos hoje existentes, as novas gerações continuam sendo influenciadas por laços sociais, culturais e geográficos. Por isto, ainda é possível visualizar-se que as variações genéticas entre populações apresentam uma distribuição aproximadamente contínua (Cavalli-Sforza *et al.*, 1994), com relação inversamente proporcional à distância geográfica entre elas (Handley *et al.*, 2007; Serre & Pääbo, 2004).

Outros fatores evolutivos importantes na evolução das populações são as mutações e a deriva genética. O surgimento de novos alelos, e a posterior modificação em suas frequências nas populações, decorrentes de fatores evolutivos como a deriva genética, promove mudanças que podem, posteriormente, ser disseminadas pelo fluxo gênico às populações vizinhas. Já o efeito fundador, um exemplo de deriva genética, é outro dos fatores evolutivos presentes no processo de dispersão humana, explicando a ausência de alelos em certas populações, assim como a alta ocorrência de alelos raros em populações derivadas. Isso porque a migração de um grupo pequeno originário de uma população qualquer dificilmente representará uma amostra perfeita dos alelos e suas respectivas frequências. Esta amostra dos alelos originais certamente produzirá uma população diferente, trazendo consigo apenas parte de sua constituição inicial. Um exemplo característico na formação das populações americanas é representado pelo sistema sanguíneo ABO, demonstrando que provavelmente os migrantes bem-sucedidos deste continente apresentavam uma maior frequência do alelo O, a ponto de não serem detectados outros grupos (A, B ou AB) entre os nativos da região (Olson, 2002).

A constatação de que existem variações genéticas detectáveis entre diferentes populações humanas já foi verificada em diversos trabalhos, confirmando que divergências populacionais entre regiões geográficas são maiores que diferenças dentro de uma mesma área analisada (Bowcock *et al.*, 1991; Tishkoff & Kidd, 2004, Witherspoon, 2007). Entretanto, ainda é pequeno o número de análises de estruturação dentro de uma mesma população. São escassos, também, os estudos que abordam a relação entre a composição genética de uma população e estratos sócio-econômicos.

A estrutura de castas existente na Índia, onde o casamento preferencial é comum, é fonte de estudos que buscam associar padrões genéticos a classes sócio-culturais (Eaaswarkhanth *et al.*, 2008; Khan *et al.*, 2008; Watkins *et al.*, 2008). O estudo de marcadores genéticos uniparentais (DNA mitocondrial e cromossomo Y) demonstrou que as castas tem origem materna comum (>98%), mas paterna bastante diversificada, mas distinta dentro de cada casta. As heranças genéticas paternas presentes na Índia refletem, inclusive, a formação de linhagens antes da colonização europeia, com perfis genéticos locais mais frequentes nas castas inferiores e padrões europeus nas castas dominantes (Thanseem *et al.*, 2006, Bamshad *et al.*, 2001).

Um estudo realizado na Venezuela abordou a diferença entre estratos sociais numa mesma população. Helios Martínez e colaboradores (2007) analisaram duas amostras da cidade de Caracas, uma oriunda de uma maternidade pública e outra de uma maternidade particular, com relação aos sistemas ABO e Rh, dezessete marcadores mitocondriais (mtDNA), três do cromossomo Y e três STRs (vWA, F13A01 e FESFPS). O conjunto dos três locos de minissatélites autossômicos demonstrou uma evidente diferença de contribuição genética em relação às populações parentais: participação europeia (espanhóis) de 75,0% (privada) e 32,8% (maternidade pública), associada a contribuição ameríndia de 16,8% (clínica particular) e 39,7% (pública), complementada pela participação africana (subsaarianos) de 8,2% na maternidade privada e 14,0% na pública. As diferenças indicam uma participação maior que o dobro de europeus da maternidade privada em relação à pública, associada a menos da metade de ameríndios e africanos na maternidade pública (em relação à particular). Este trabalho desenvolvido pelo grupo de Martínez evidencia que os dois grupos amostrais estudados apresentam variações genéticas entre si, reflexo da diferença quanto ao nível sócio-econômico, e alerta para os critérios de seleção na escolha dos indivíduos que comporão uma amostra.

1.1 Marcadores moleculares e sua utilização na genética humana

A diversidade de indivíduos numa população já havia sido relatada por Darwin, em seus estudos sobre evolução das espécies. A genética de populações adotou a observação e quantificação destas características, determinando os polimorfismos existentes. Até a década de 1980, estes estudos baseavam-se em

caracteres fenotípicos, o que incluía os polimorfismos protéicos, enzimáticos e imunológicos, como os grupos sanguíneos, proteínas séricas, imunoglobulinas e sistemas de histocompatibilidade, entre outros, denominados marcadores genéticos clássicos.

O uso das técnicas de biologia molecular possibilitou a análise da variação genética analisando diretamente o genoma, tornando sua aplicação em estudos populacionais uma ferramenta capaz de proporcionar informações mais objetivas (Salzano e Callegari-Jacques, 1988; Pérez-Lezaun *et al.*, 1997; Jorde *et al.*, 1997; Hutz *et al.*, 1999). Tais marcadores passaram a ser denominados marcadores genéticos moleculares.

A identificação humana na área forense foi uma das primeiras aplicações dos marcadores genéticos moleculares, elucidando casos criminais na Inglaterra (Jeffreys *et al.*, 1985), por meio da análise das variações de comprimento dos fragmentos de restrição, o RFLP (*Restriction Fragment Length Polymorphism*), em regiões multilocais do DNA genômico. A mesma metodologia encontrou aplicação em diversas outras áreas, como por exemplo o mapeamento genético (Donis-Keller *et al.*, 1987; Telenius *et al.*, 1990; Silva & Moura-Neto, 1998) e posterior diagnóstico de patologias. Com o advento da técnica da PCR (*Polymorphism Chain Reaction*), desenvolvida em 1985 por Kary Mullis (1990), a identificação pelos microssatélites (ou STRs – *Short Tandem Repeats*) tomou o lugar do RFLP, por sua praticidade, rapidez e ótima relação de custo-benefício (Pena & Jeffreys, 1993; Pena & Chakraborty, 1994). À análise dos STRs somaram-se os estudos dos polimorfismos do tipo Indels (inserção/deleção – Weber *et al.*, 2002) e SNPs (*Single Nucleotide Polymorphisms* – Rapley & Harbron, 2004), muito utilizados hoje na identificação humana e genética de populações. O desenvolvimento dos marcadores moleculares permitiu uma rápida evolução na pesquisa genética, como *kits* capazes de cobrir mais de 900.000 SNPs e outros 946.000 marcadores em um único conjunto (Affymetrix, 2009), ou mesmo alcançando os cidadãos comuns, capazes de conseguir seu perfil genético para 600.000 SNPs a preços acessíveis (23andMe, 2007).

O uso de marcadores moleculares alcançou o estudo da formação das populações atuais, permitindo a estimativa matemática da contribuição de seus diferentes grupos parentais, de acordo com as frequências dos alelos presentes em

cada uma. A análise genética por meio de marcadores moleculares possibilita estudar a composição de uma população (em termos de herança ancestral), e também inferir processos de colonização e migração histórica (Salzano & Callegari-Jacques, 1988; Cavalli-Sforza *et al.*, 1994; Da Silva Jr. *et al.*, 1999; Cavalli-Sforza & Feldman, 2003, Bamshad *et al.*, 2003). Para tanto, busca-se nas populações parentais em estudo (geográfica ou culturalmente distintas), frequências alélicas bastante distintas, típicas de cada grupo (Shriver *et al.*, 1997). Os marcadores assim definidos são denominados AIMs (*Ancestry Informative Markers*) ou PSAs (*Population-Specific Alleles*), e suas frequências alélicas podem variar de acordo com as populações parentais analisadas, sendo mais informativos para estudos de mistura genética que marcadores aleatórios (Shriver *et al.*, 1997, Bonilla *et al.*, 2004).

Os principais marcadores moleculares utilizados na genética de populações são os STRs, as Indels (Mills *et al.*, 2006; Whitterspoon *et al.*, 2006) e os SNPs (Underhill *et al.*, 2001; Moskvina *et al.*, 2010). Estes últimos caracterizam-se por sua menor taxa de evolução (Nielsen, 2000; Weber *et al.*, 2002; Bastos-Rodrigues *et al.*, 2006), sendo muito úteis em análises populacionais de eventos de dispersão e migração (Romualdi *et al.*, 2002). As indels, que ocorrem em quase qualquer parte do genoma, podem sujeitar o organismo a mutações capazes de afetar o poder adaptativo de um indivíduo, sendo um dos marcadores mais ligados às pressões evolutivas (Britten *et al.*, 2003), por consistirem em sequências longas, afetando em maior grau o conteúdo do genoma. Os STRs, por sua praticidade de aplicação e maior potencial de individualização que os demais, tornaram-se a principal metodologia nas aplicações forenses, seja nos casos de paternidade ou correlação entre suspeito e vestígio. Também são usados os RFLPs (Jeffreys *et al.*, 1985) e os elementos *Alu* (Bydlowski *et al.*, 1996; Parra *et al.*, 1998; Oliveira, 1999), inserções que ocorrem numa região do genoma como evento único, incorporando-se ao patrimônio genético daquele indivíduo. Por ser unidirecional (Tishkoff *et al.*, 2000), é analisado de maneira diferente na genética de populações, tendo importância pela quase impossibilidade estatística de existir como inserção recorrente.

Dos marcadores citados, os STRs são os mais utilizados não só na área de genética de populações, mas também na esfera forense, por diversas características: seu elevado polimorfismo (Ferreira & Grattapaglia, 1995;), com

grande variação entre indivíduos (Lins *et al.*, 1996) e, em menor grau, entre populações (Charlesworth *et al.*, 1994). Regiões de STRs apresentam taxa de mutação mais alta por causa da sua principal característica, as repetições em tandem (Santos & Tyler-Smith, 1996; Kayser *et al.*, 2000). Nessas regiões a replicação tem maior probabilidade de sofrer um erro em decorrência de um processo denominado derrapagem da polimerase, o que leva ao aumento ou diminuição de, geralmente, uma unidade de repetição. Os STRs mais utilizados em genética de populações estão presentes em regiões que não codificam proteínas, estando menos sujeitos às pressões evolutivas. Dessa forma, espera-se que mesmo populações que sofreram diferenciação mais recentemente na história evolutiva humana apresentem alta diversidade.

Esta vantagem dos microssatélites é também motivo para análises genéticas conservadoras, visto que com frequência os alelos observados podem existir por mutações recorrentes. Os STRs são, portanto, ideais para o trabalho de identificação humana e análises de subestruturação populacional (Rosenberg *et al.*, 2003; Rosenberg *et al.*, 2002), mais do que inferências filogenéticas. Seu uso, entretanto, exige uma quantidade grande de regiões, de modo a reduzir os efeitos das mutações recorrentes (Budowle & Brown, 2001), aumentando seu poder de discriminação.

Os estudos genéticos aplicados à genética de populações exigem o uso de marcadores com diferentes características de mutação e recombinação, de modo a proporcionar uma visão mais completa de sua história genética, incluindo subestruturação, migrações e variabilidade decorrente de outros processos evolutivos, como a deriva genética (Salzano, 1998).

As análises populacionais, por si só, não são suficientes para considerações conclusivas acerca de um processo de miscigenação, por exemplo. Inferências sobre contribuições de distintas populações na formação de uma população complexa devem considerar o processo histórico de ocupação e migração dos componentes das respectivas populações parentais.

Os marcadores do tipo STR vem sendo utilizados para a reconstrução da história demográfica do Brasil e da América Latina, como demonstraram Hutz e colaboradores (2002) nas diversas regiões brasileiras, Godinho e colaboradores (2008a) com países da América do Sul (Brasil, Argentina e Colômbia), bem como

estudos acerca das populações da América Latina (Wang *et al.*, 2008) e outros alcançando todo o continente americano (Yang *et al.*, 2010). Análises de composição genética intracontinental na constituição das populações americanas demonstram grandes variações entre áreas costeiras, interiores, urbanas e rurais (Sans, 2000).

Estudos como o de Corach e colaboradores (2010) mostram que a história de colonização do continente sul-americano é semelhante à do Brasil, obtendo resultados para a composição étnica argentina com poucas diferenças em relação à população brasileira. A diferença se verifica numa menor participação africana, consistente com a distância em relação aos pontos de início da colonização (Godinho e colaboradores, 2008a), e maior percentual de ancestralidade genética ameríndia.

A possibilidade de individualização por meio de marcadores moleculares popularizou esta ferramenta no trabalho de investigação forense, tanto na área cível (como os processos de paternidade) quanto no ambiente criminal (no trabalho policial de identificação de autoria de crimes).

Marcadores genéticos utilizados em bancos de dados criminais com finalidade de identificação humana forense tornaram-se uma tendência em diversos países do mundo. O intercâmbio de informações entre diferentes nações, com vistas a investigações, é uma prática antiga, possibilitando a identificação e eventual prisão de criminosos que atuem em mais de um país. Para que as informações de perfis genéticos fossem úteis entre diferentes países, identificou-se a necessidade de uso de marcadores em comum. Atualmente, alguns marcadores do tipo STR são quase um padrão entre nações, a maioria deles presente no sistema norte-americano CODIS (*Combined DNA Index System*), implementado pelo FBI (*Federal Bureau of Investigation*) a partir de marcadores conhecidos e internacionalmente utilizados (Budowle *et al.*, 1998). Treze marcadores STR foram selecionados para compor o banco de dados genético: CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 e D21S11. O estabelecimento destes marcadores como utilizados para a identificação humana baseou-se em sua grande capacidade de diferenciação, ao mesmo tempo em que não trazem nenhuma informação fenotípica, por não estarem em regiões codificantes do genoma

(requisito para que seu uso fosse aprovado sem consequências negativas na área de direitos humanos).

A definição destes STRs como marcadores de uso comum em diversos países produziu uma série de pesquisas e estudos para verificar a frequência de seus alelos em diferentes populações. Tais análises estatísticas são necessárias para demonstrar a viabilidade de uso destes marcadores em cada população, bem como estabelecer seus respectivos índices de identificação e paternidade, instrumentos matemáticos utilizados na resolução de investigações criminais e cíveis.

Como as pesquisas com esse conjunto de STRs foram basicamente realizadas no hemisfério norte, aos poucos foi-se verificando que, apesar de suas qualidades como locos individualizadores, quando em grupo eram passíveis de denotar outras características populacionais. Estudos como os citados anteriormente (na Índia, na América Latina e mesmo no Brasil) demonstraram sua capacidade de atuar também na genética de populações, nas análises de composição genética (Godinho, 2008b), processos de povoamento (Godinho *et al.*, 2008a), e até mesmo na identificação de estratos sociais (Khan *et al.*, 2008).

1.2 Povoamento do Brasil

O processo de ocupação pré-colombiana do continente americano só agora tornou-se mais claro em função das análises genéticas moleculares. Vestígios arqueológicos e dados baseados em polimorfismos não-moleculares indicavam uma quantidade não inferior a três ondas migratórias de ocupação, a partir do continente asiático (Greenberg *et al.*, 1986) através do estreito de Bering. Análises mais recentes levando em consideração marcadores moleculares autossômicos e sexuais sugerem uma hipótese compatível com apenas duas levas migratórias (O'Rourke & Raff, 2010; Yang *et al.*, 2010).

Por outro lado, análise das regiões hipervariáveis do DNA mitocondrial (Bonatto & Salzano, 1997) e de marcadores genéticos Y-específicos, demonstram que tal processo migratório deve ter ocorrido em um único momento (entre 30.000 e 40.000 anos atrás), levando à colonização do continente americano nos milênios seguintes (Silva Jr *et al.*, 2002; Ruiz-Linares *et al.*, 1999). Mesmo nos dias de hoje, a variedade de haplótipos no cromossomo Y ainda é pequena entre os ameríndios, evidenciando (apesar do processo de miscigenação) uma provável origem única

(Bortolini *et al.*, 2002, Santos *et al.*, 1999), consistindo de uma grande leva migratória com posteriores fluxos de menor intensidade.

Análises moleculares com marcadores do cromossomo Y permitiram inferências sobre a história da ocupação do continente americano, mostrando evidências de uma grande migração masculina com fluxos menores posteriores, e um processo de diferenciação associado à diversidade lingüística e cultural (Santos *et al.*, 1999; Tarazona-Santos *et al.*, 2001). Esta teoria de um único fluxo migratório recentemente ganhou força com estudos que complementam as análises do cromossomo Y com o DNA mitocondrial (Fagundes *et al.*, 2008).

Os ameríndios colonizaram o futuro território brasileiro há pelo menos 12 mil anos, constituindo diversas tribos distintas, com diversidade cultural. A partir do século XIV, tendo uma população estimada de dois milhões de indivíduos (Cunha, 1998), somou-se um enorme contingente de europeus e africanos, levando a um intrincado processo de mistura genética, motivo de diversos estudos com a utilização de marcadores genéticos (Salzano, 1998, Pena, 2000; Oliveira *et al.*, 2006). Somase a este processo a redução acentuada da população indígena, seja pelo óbito decorrente de doenças, ou pelo extermínio no processo de ocupação do continente pelos colonizadores e ainda pela incorporação via mistura às novas populações brasileiras.

É importante ressaltar que o contingente de migrantes europeus e africanos do sexo masculino foi bem maior do que do sexo feminino o que, em teoria, deveria ter como reflexo uma notória distinção entre a contribuição observada na população atual dos componentes materno ou paterno-específicos. Estima-se que o contingente de escravos trazidos para o Brasil foi da ordem de quatro milhões de indivíduos (aproximadamente metade do total de africanos levados para o continente americano como um todo), dos quais cerca de dois terços eram do sexo masculino (Reis e Gomes, 1996; Queirós Mattoso, 1982). Por outro lado, como análises de populações brasileiras vem demonstrando, os homens africanos tiveram pouco sucesso reprodutivo quando comparados aos portugueses. A consequência disso é o encontro atual de uma frequência muito maior de cromossomos Y de origem europeia do que africana e ameríndia no Brasil (Barcelos *et al.*, 2006; Callegari-Jacques *et al.*, 2003; Carvalho-Silva *et al.*, 2001).

A participação de outras etnias além destas três principais (ameríndios, europeus e africanos) na composição da população brasileira foi bastante reduzida, sendo que processos de migração asiática, por exemplo, só ocorreram de maneira significativa nas últimas décadas. Com isso, estima-se que dos imigrantes chegados ao Brasil entre 1500 e 1978, aproximadamente 58% eram europeus, 40% africanos e 2% asiáticos (Callegari-Jacques & Salzano, 1999). De forma natural, ocorreu um processo lento e gradual de miscigenação, incorporando-se à população brasileira diversas características genotípicas e fenotípicas (Salzano, 1986).

Verifica-se que, passados 500 anos do início de processo de colonização por populações não-ameríndias no Brasil, a composição étnica matrilinear apresenta uma mistura relativamente equilibrada no país. Face à história recente de colonização do continente americano, suas populações tem sido objeto de crescente número de estudos (Bonatto & Salzano, 1997; Pena *et al.*, 2000; Barcelos, 2006; Godinho *et al.*, 2008a), visando esclarecer o processo de constituição do povo brasileiro.

Estudos acerca da atual composição da população brasileira em suas diversas regiões concluem que a participação materna é representada pelas três populações parentais, enquanto que a linhagem patrilinear é primordialmente de origem europeia (Bortolini *et al.*, 1997a e 1997b; Carvalho-Silva *et al.*, 2001; Abe-Sandes *et al.*, 2004; Barcelos *et al.*, 2006).

O que se percebe, em termos regionais, é que a população parental africana possui menor representação na região Sul, enquanto que no Norte a proporção indígena é responsável por mais de 50% da composição populacional (Alves-Silva *et al.*, 2000; Parra *et al.*, 2003). A participação do genoma africano na composição étnica da região Norte é bastante reduzida quando comparada às demais regiões brasileiras, como se verifica de estudos realizados a partir do mtDNA (Alves-Silva *et al.*, 2000; Callegari-Jacques *et al.*, 2003).

A análise de amostras populacionais da região Centro-Oeste do Brasil por meio de marcadores moleculares do tipo AIM também vem contribuindo para o entendimento desses processos de colonização e miscigenação mais lentos e recentes que os das demais regiões brasileiras. Nordeste, Sudeste e Sul, por compreenderem regiões costeiras e de fácil acesso, foram colonizadas mais rapidamente após o descobrimento, apresentando processos de ocupação

semelhantes entre si, e cujo exaustivo estudo muito elucidou sobre sua composição e história (Godinho, 2008b).

Os estudos geralmente demonstram que amostras populacionais da região Norte apresentam uma maior quantidade de alelos caracteristicamente ameríndios que as amostras do Centro-Oeste (Callegari-Jacques *et al.*, 2003; Parra *et al.*, 2003; Arcanjo *et al.*, 2009). Análises por meio de marcadores com poder de discriminação de ancestralidade africana ou ameríndia demonstram de forma clara este padrão (Godinho, 2008b; Carvalho-Silva *et al.*, 2001). Estudos com STRs (Muniz *et al.*, 2008) e SNPs (Lins *et al.*, 2010) demonstram a existência de gradientes de participação genética ameríndia, africana e europeia na composição da população brasileira.

Porém, apesar da existência de um bom volume de estudos populacionais das diversas regiões brasileiras, estes ainda acontecem de modo por demais autônomo. Dessa forma, considerando-se que o processo de formação entre as diversas regiões brasileiras é distinto uma das outras, é necessário um melhor conhecimento das diferenças e similaridades entre a composição étnica das dessas regiões, para uma compreensão maior da dinâmica de colonização do Brasil.

1.2.1 Situação demográfica brasileira

Dados sobre a constituição da população brasileira tem sido obtidos pelo IBGE (Instituto Brasileiro de Geografia e Estatística) desde 1872, quando ainda se denominava Directoria Geral de Estatística. A partir de 1960 foi incluída a informação acerca da cor ou raça, sendo que desde o Censo de 2000 na forma de auto-declaração dos entrevistados. De acordo com o banco de dados do IBGE, a proporção das diferentes cores de pele é similar quando compara-se o Brasil como um todo e a região Centro-Oeste. Esta similaridade deve ser decorrente, dentre outros fatores, do fato da região ter sido ocupada apenas recentemente, sendo povoada em período posterior ao início do processo de miscigenação. Ainda de acordo com a PNAD (Pesquisa Nacional por Amostra de Domicílios) do IBGE, a região Centro-Oeste foi a única a apresentar um maior percentual de imigrantes do que de pessoas naturais da região, em 2007 e 2008. As informações do IBGE referentes à distribuição de raça ou cor do Censo 2000 e do PNAD 2007-2008 são mostradas na Tabela 1.1.

TABELA 1.1 – Percentuais de participação das raças (auto-declaradas) branco, preto, pardo e indígena, relativas ao Censo 2000 e ao PNAD 2007-2008. Não estão apresentadas as raças Amarela e Outros.

	Censo 2000				PNAD 2007-2008			
	Branco	Preto	Pardo	Indígena	Branco	Preto	Pardo	Indígena
Brasil	53,7	6,2	38,5	0,4	48,4	6,8	43,8	0,3
Centro-Oeste	49,4	5,0	43,7	0,8	41,6	6,5	50,6	-
DF	49,2	4,8	44,8	0,3	-	-	-	-

Fonte: IBGE – Censo 2000 e PNAD 2007-2008

As auto-declarações mais frequentes foram a branca, reflexo da alta migração europeia para o Brasil desde o início da ocupação em 1500 e do sucesso reprodutivo dos europeus, e a parda, produto da intensa miscigenação encontrada no país. Na Figura 1.1 estão apresentadas as distribuições dessas cores no Brasil, bem como dos negros e indígenas (também por auto-declaração). Observa-se que a cor branca apresenta maior frequência na região sul e menor ocorrência na região norte, evidenciando uma gradação entre as duas. Padrão inverso é observado para a cor denominada Parda. As cores pretas e a etnia indígena mostram baixa frequência em todo o território brasileiro, com pequenas áreas de maior concentração, como os indígenas no extremo noroeste do país (IBGE, 2000b). Os grupos denominados Amarelos (descendentes de orientais) e Outros não são considerados nas análises de miscigenação por representarem, em conjunto, cerca de 1,0% da população brasileira. Revela-se na Figura 1.1, com relação à região Centro-Oeste, uma participação populacional semelhante para brancos e pardos, com índices entre 25 e 50%.

Há que se ressaltar que a parcela da população auto-declarada parda no Brasil apresenta composição genética de brancos e negros (descendentes respectivamente de populações europeias e africanas, de um modo geral), sendo difícil determinar o grau de participação de cada parental, em termos genéticos ou étnicos.

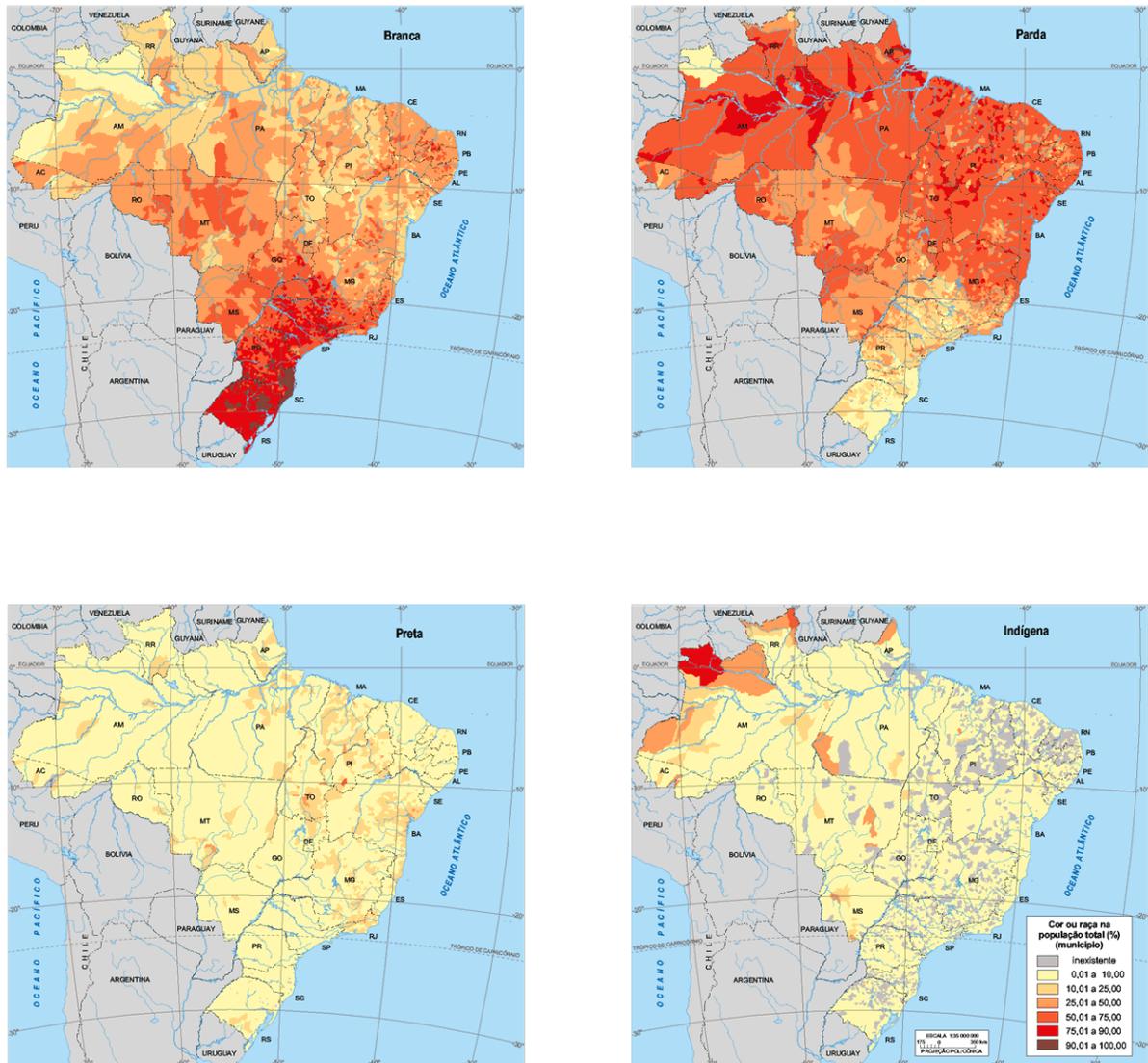


FIGURA 1.1 – Mostra a distribuição percentual da população por cor ou raça auto-declarada: Branca, Parda, Preta e Indígena. FONTE: Atlas do Censo Demográfico 2000 - www.ibge.gov.br/home/estatistica/populacao/censo2000/atlas

Cabe destacar que os índices da presença indígena (ou ameríndia) nos censos brasileiros são, via de regra, discrepantes em relação aos cálculos de mistura baseados em frequências alélicas para marcadores moleculares. Tal fato decorre tanto da incorporação via miscigenação, que leva a auto-definição como pardo, quanto também ao que tem sido denominado pouca dominância fenotípica das características tipicamente ameríndias. Estudo realizado por Klimentidis & Shriver (2009) demonstrou que a aparência ameríndia é uma das que menos é reconhecida na sociedade, tanto por auto-declaração quanto por visualização de outras pessoas.

1.2.2 Associação entre aspectos sociais e cor ou raça

Entre os primeiros registros demográficos da população brasileira, destacava-se a presença de cerca de 20% de população preta no final do século XIX, em sua maioria escravos. A carência de direitos civis dessa parcela da população brasileira, que incluía a educação e possibilidade de empregos com maior renda, incutiu na sociedade brasileira uma associação de cor-de-pele e classe social, apesar da atual constituição do Brasil prever iguais direitos a qualquer tipo de cidadão, independente de sua raça ou cor.

No Brasil, a educação é um fator reconhecidamente influenciado por fatores econômicos e sociais. Segundo o IBGE (Censo 2000), o reflexo da relação entre educação e cor ou raça é também histórico, sendo evidente nos censos. Os maiores índices de escolaridade entre os brasileiros estão entre os auto-declarados amarelos, com uma média de 9 anos de estudo *per capita*. Entre os brancos a média foi de 7 anos, caindo para 5 anos entre os indivíduos de cor preta ou parda e 4,3 anos entre os indígenas.

Os dados se repetem quanto ao analfabetismo. Entre os brancos, segundo o Censo 2000, atingia 8,3% das pessoas, subindo para 21,5% entre os auto-declarados pretos e 26,1% na etnia indígena. Os auto-declarados amarelos (de origem predominantemente asiática) no Brasil tem índice de analfabetismo inferior a 5%. Com relação aos ameríndios, a alta frequência de analfabetismo é consequência direta da manutenção de culturas próprias independentes em graus variáveis por parte das populações circundantes, isso é, a barreira cultural de forma diversificada, decorrente também do isolamento geográfico. Quanto aos afro-descendentes, trata-se principalmente de consequências da desigualdade social ainda existente no país.

Na apresentação do IV Relatório Nacional de Acompanhamento dos Objetivos de Desenvolvimento do Milênio (projeto do governo federal brasileiro), a coordenadora do Sistema das Nações Unidas no Brasil, Marie Pierre Poirier, afirmou que “O gênero, a raça, a etnia e o local de nascimento de uma criança brasileira ainda determinam, em grande parte, suas oportunidades futuras.” O relatório mostra, por exemplo, que a proporção de pessoas definidas como pretas que vivem abaixo da linha da pobreza estabelecida pela ONU é o dobro do observado entre os definidos como brancos (IPEA, 2010).

Mas a característica mais marcante da relação cor-de-pele e fatores sociais é a renda mensal. Observa-se valores desproporcionais na distribuição de renda da população brasileira, de acordo com o IBGE (Censo 2000): enquanto entre as pessoas com renda superior a cinco salários-mínimos (IBGE, Censo 2000) aparece apenas a metade de auto-declarados pretos, pardos e indígenas esperados, os brancos apresentam valores acima da média esperada justamente neste segmento de renda (Figura 1.2). Entre os brasileiros com curso superior, a renda dos não-afro-descendentes supera a dos afro-descendentes em cerca de 40%, um dado alarmante de desigualdade social.

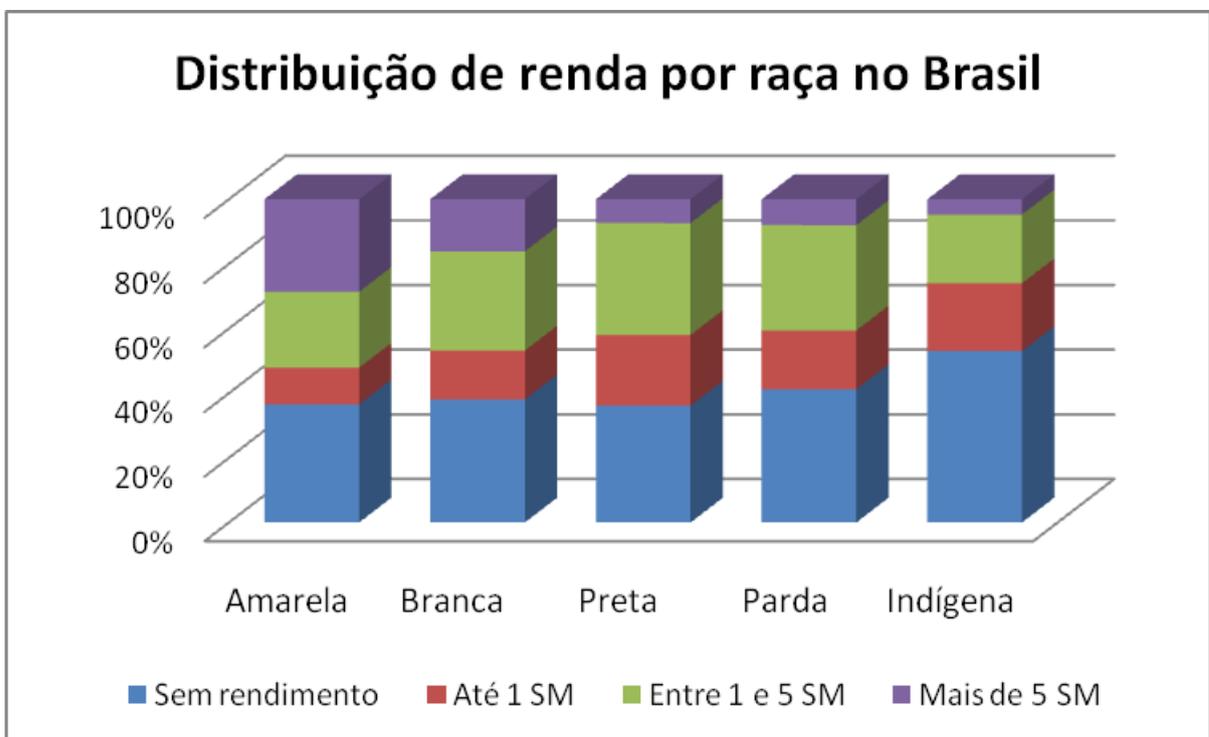


FIGURA 1.2 - Proporção observada de distribuição de renda no Brasil, nas cinco raças auto-declaradas (conforme IBGE).

Trabalhos brasileiros buscam um retrato da situação brasileira sob a ótica dos levantamentos de raça ou cor, escolaridade e renda. Andrade & Dachs (2007) dividem os dados da PNAD (2003) em “brancos” (englobando os auto-declarados brancos e amarelos) e os “não-brancos” (que incluem os pardos, negros e indígenas), representando respectivamente 48% e 52% da população brasileira. Neste estudo, identificam que o acesso ao ensino médio e superior está associado à

renda e à raça (mais ao primeiro fator que ao segundo), demonstrando uma inter-relação entre estes.

A união dos auto-declarados pardos e pretos nas análises estatísticas dos dados demográficos brasileiros (pelo IBGE) é comum, apesar das diferenças genéticas existentes. Yara Muniz e colaboradores (2008) mostraram que, em termos de composição genética, os pretos apresentam similaridade maior com as populações africanas (63% de participação africana e 37% de contribuição europeia) do que os pardos, cuja maior semelhança genética é com os europeus (26% de constituição parental africana e 62% de participação europeia). Esses resultados levantam questões sobre a adequação de agrupar os pardos com os pretos, ao invés dos brancos.

Percebe-se que, de fato, a história da escravidão de africanos no Brasil apresenta repercussões até os dias atuais. Mesmo no Império, os ameríndios possuíam mais direitos civis que os afro-descendentes (apesar de serem mão-de-obra mais barata, já presente no território), parcialmente protegidos que eram pela iniciativa católica, associada a uma postura colonialista de compensação e omissão em relação à exclusão social gerada pela escravidão (Alencastro, 2000).

Nos séculos XVII e XVIII, com o ciclo do ouro e o crescimento de cidades decorrentes do aquecimento econômico, surgiu na sociedade brasileira o equivalente a uma classe média, de comerciantes que não dependiam das culturas agrícolas e dos escravos. Este período de oportunidades econômicas e sociais ainda excluía os escravos negros, com raras exceções históricas, como a história de Galanga (transformada no filme “Chico Rei”), escravo trazido do Congo a quem se atribuía um reino africano cuja corte ele reconstruiu em Ouro Preto (antiga Vila Rica), após pagar sua alforria e comprar uma mina aurífera na década de 1740 (Silva, 2007).

Nesse mesmo período, o comércio de escravos dentro do país cresceu enormemente, talvez superado apenas pelos negócios com ouro e açúcar. A fuga de escravos e/ou alforria destes iniciou um processo de negação da condição de ser afro-descendente, surgindo no país a autodenominação de “pardo”. Tratou-se de um processo social de tentativa de inclusão na sociedade dominante, buscando a fuga da condição de submissão social. A abolição da escravatura não resolveu a exclusão social dos afro-descendentes, que de escravos passaram a ser

desempregados ou contratados com salários ínfimos em condições semelhantes à escravidão. Permaneceram, pois, à margem da sociedade, sem nenhum tipo de processo de inclusão cultural ou econômica (Alencastro, 2000).

A história do Brasil reflete claramente a exclusão dos ex-escravos com a discriminação à cor da pele. Dados do IBGE sempre evidenciaram que entre pessoas com a mesma formação acadêmica, os brancos recebem salários maiores que os não-brancos.

As evidências da relação entre cor-da-pele e aspectos sociais/econômicos são claras. Visto que a cor-da-pele apresenta um componente genético e que, dependendo da população em análise, pode indicar ancestralidade (Parra *et al.*, 2003; Machado, 2008; Muniz *et al.*, 2008; Santos *et al.*, 2009), é possível que ocorra diferença na composição genética entre as diferentes classes sociais. Sua detecção, entretanto, depende de amostras escolhidas em uma população de maneira não-aleatória, com um viés econômico. Neste sentido, espera-se que duas amostras oriundas de uma mesma população, caso apresentem características de seleção com algum fator sócio-econômico, possam apresentar padrões genéticos distintos.

1.2.2.1 Reflexos da influência social em análises genéticas de composição étnica

Estudos com marcadores moleculares visando inferências de composição étnica estão ficando mais comuns, à medida que se tornam disponíveis genótipos de populações miscigenadas. E a análise de amostras representativas de uma mesma população pode começar a ser feita de maneira mais crítica, discutindo a forma de amostragem dos indivíduos estudados.

No estudo de Callegari-Jacques e colaboradores (2003), uma amostra de 1037 indivíduos não-relacionados foi analisada quanto à sua composição étnica. É preciso ressaltar que esses indivíduos amostrados buscavam exames de paternidade em laboratório particular, o que já produz um viés sócio-econômico na amostras, decorrente do custo de tal exame. Neste estudo, portanto, é de se esperar resultados com perfil social e econômico superiores aos da média populacional regional, cujo poder aquisitivo é inferior. Como exposto na Figura 1.1 desta Introdução, há na população brasileira uma participação média entre 25 e 50% de brancos e pardos (auto-declarados) e inferior a 10% de ameríndios. Ao comparar-se esses dados aos do Censo 2000 e PNAD 2007-2008 (Tabela 1.1), observa-se uma

expectativa de 40-50% de brancos e entre 45 e 55% de pardos e pretos, na região Centro-Oeste. Presume-se aqui uma realidade (ainda que por meio de informações prestadas por meio de auto-declaração) de equilíbrio destas duas categorias de cores de pele (pardos e brancos), com os demais grupos compondo um percentual ao redor de 5%.

As estimativas de contribuição das populações parentais na composição genética da amostra populacional obtida neste laboratório privado, utilizando os marcadores do tipo STR autossômicos (Tabela 1.2), mostram valores bastante diversos, coerentes entre si pela utilização de dois métodos estatísticos distintos (Chakraborty, 1985; Dupanloup & Bertorelle, 2001). Ambos fornecem estimativa de ancestralidade a partir das frequências alélicas dos locos analisados, sendo que o segundo leva em conta também o modelo mutacional de alelo simples (característico de STRs). Pelos genótipos obtidos, foi estimada uma participação de europeus da ordem de 71%, de africanos em torno de 18% e ameríndia ao redor de 11%. O fato de duas análises diferentes produzirem valores tão semelhantes e com desvios proporcionalmente mínimos indica que a análise estatística dos genótipos obtidos está robusta.

TABELA 1.2 – Proporção de ancestralidade estimada para a região Centro-Oeste a partir de STRs, por duas metodologias (extraído de Callegari-Jacques *et al.*, 2003).

Ancestralidade	por Chakraborty (1985)	por Dupanloup & Bertorelle (2001)
Europeia	0,71 ± 0,002	0,72 ± 0,032
Africana	0,18 ± 0,001	0,18 ± 0,029
Ameríndia	0,11 ± 0,001	0,10 ± 0,018

Ao se comparar estes resultados (Tabela 1.2) com as informações de auto-declaração do IBGE (Figura 1.1 e Tabela 1.1), observa-se uma discrepância na proporção de brancos (expectativa de ancestralidade predominantemente europeia) e pardos (expectativa de ancestralidade intermediária entre brancos e negros – Muniz *et al.*, 2008) e negros (maior ancestralidade africana), respectivamente da ordem de 45% e 50%. Esta diferença provavelmente decorre de um das seguintes possibilidades:

- viés na auto-avaliação dos recenseados (voluntário ou involuntário), produzindo esta discrepância em relação aos dados observados nas análises genéticas; ou

- erro amostral decorrente de um ou mais fatores, dos quais destaco dois: a coleta não-aleatória e a provável não-representatividade destas grandes cidades como retrato da população regional.

A primeira possibilidade produzirá erros sistemáticos em todas as avaliações, quando comparados seus padrões genéticos aos de auto-avaliação dos recenseados. Neste caso, a análise de qualquer outra amostra similar produzirá resultados semelhantes aos de Callegari-Jacques e colaboradores (2003), com índices ao redor de 71% de europeus e 18% de africanos. A reprodução destes valores percentuais indicará a presença de um erro (na forma de um viés) no procedimento de auto-declaração utilizado pelo IBGE. Neste caso, considera-se erro até mesmo o procedimento hoje utilizado de auto-avaliação, que é subjetivo e sujeito a discrepâncias voluntárias (cultural) ou involuntárias (pela grande variação entre fenótipo e genótipo no quesito cor-da-pele).

No segundo caso, admitindo-se que a amostra populacional obtida apresente algum tipo de erro, valores gerados a partir de amostras teoricamente “semelhantes” podem produzir índices de composição étnica bastante díspares (desde que não cometido o mesmo tipo de erro na escolha dos indivíduos que comporão a amostra populacional).

Admite-se, aqui, que se a comparação do resultado de ancestralidade (a partir de três populações parentais) entre as amostras obtidas em laboratórios público e privado produzir resultados diversos, supõe-se uma não-aleatoriedade na escolha de seus indivíduos, seja em apenas uma delas ou em ambas. É fato que as amostras em questão contemplam (de um modo geral) pessoas com situação econômica distinta. Havendo discrepância nos resultados de composição genética, fica detectada uma diferença que poderá ser associada ao poder aquisitivo destas classes sócio-econômicas (principal fator que distingue as amostras dos laboratórios público e privado).

Callegari-Jacques e colaboradores (2003) já alertavam para essa possibilidade na discussão de seu trabalho, lembrando que o poder aquisitivo da

população de pardos e pretos é de apenas 50 a 60% em comparação aos auto-declarados brancos. Sugeriam, no bojo do trabalho, a possibilidade de que a população com melhores condições financeiras poderia apresentar um maior percentual de composição étnica europeia.

2. JUSTIFICATIVAS E OBJETIVOS

Em decorrência das particularidades do processo de colonização, as populações americanas constituem uma grande fonte de informações para a genética de populações. Um questionamento acerca dos estudos populacionais é se as amostras são, de fato, representativas das populações. Os processos de coleta buscam adotar condições de aleatoriedade, para tornarem-se realmente representativas da população da qual fazem parte. No entanto, devido às singularidades de coleta que distintos grupos de pesquisa adotam, questiona-se também se diferentes amostras da mesma população podem refletir a estratificação que uma dada população apresenta.

Propõe-se, aqui, a hipótese de que amostras, mesmo sendo estatisticamente consideradas como oriundas de uma mesma população, podem apresentar diferenças decorrentes de subestruturação, advindas de fatores sócio-econômicos, detectáveis a partir de estudos genéticos. Dessa forma, a hipótese inicial é que as pessoas amostradas em laboratório público e que apresentam menor poder aquisitivo apresentarão constituição genética quanto aos grupos parentais distinta das amostras coletadas em laboratório privado.

O diagnóstico de que a distribuição sócio-econômica da sociedade brasileira se reflete na análise de sua composição genética abre espaço para uma discussão acerca das divergências sociais. A constatação genética do fato permite discutir as origens de tais diferenças, podendo abrir espaço nos debates acerca da história da formação do povo brasileiro.

Deste modo, e considerando-se que:

- os estudos genéticos de populações urbanas no Brasil tem buscado especialmente a formação de bancos de dados e reconstituição histórica do processo de miscigenação;
- os procedimentos de coleta de amostras em populações buscam ser aleatórios, mas geralmente desconsideram as classes sociais;
- há pouco ou nenhum estudo de comparação entre amostras obtidas a partir de uma mesma população.

O presente projeto apresenta como objetivos específicos, dentro de um contexto de um grupo de estudos sobre populações brasileiras:

- Comparar duas amostras teoricamente representativas da população do Distrito Federal, coletadas em laboratórios que traçam perfis genéticos individuais para finalidade, em maior número, de investigação de paternidade, sendo que um dos laboratórios é público (Instituto de Pesquisa de DNA Forense da Polícia Civil do Distrito Federal) e o outro privado (Laboratório Heréditas) quanto à:
 - diversidade genética relativa a marcadores do tipo STRs autossômicos;
 - porcentagem de contribuição das populações parentais africanas, ameríndias e europeias na formação destas amostras populacionais;
- Comparar essas duas amostras representativas do Distrito Federal com outras oito amostras populacionais referentes a seis estados brasileiros e à população brasileira como um todo;
- Avaliar a possibilidade de utilização de bancos de dados de amostras populacionais para inferir a origem geográfica (população parental mais provável) de um determinado perfil genético.

3. MATERIAL E MÉTODOS

3.1 Amostras populacionais do Distrito Federal

Foram utilizadas neste trabalho duas amostras populacionais obtidas por laboratórios do Distrito Federal (Grattapaglia *et al.*, 2001 e Dalton *et al.*, 2009) já coletadas e genotipadas para marcadores STRs (incluindo os locos utilizados no sistema CODIS). Todas as coletas realizadas foram feitas após autorização individual escrita dos doadores.

Ambas as amostras (Grattapaglia *et al.*, 2001 e Dalton *et al.*, 2009) serviram de padrão comparativo com outras frequências alélicas de trabalhos com amostras populacionais brasileiras.

3.1.1 Amostra oriunda de laboratório público do Distrito Federal

A amostra de laboratório público foi coletada no Laboratório de Pesquisa de DNA Forense da Polícia Civil do Distrito Federal (PCDF). Esta amostra (n=915) tem como característica principal o fato de originar-se em sua maioria de pessoas sem condições de pagar por um exame de paternidade em laboratório particular. Em menor quantidade (aproximadamente 20%) encontram-se amostras relacionadas à investigação criminal, constituídas basicamente por perfis de referência (cujo perfil é, majoritariamente, semelhante ao dos indivíduos analisados nos casos de paternidade). A amostra consiste de indivíduos não aparentados, tendo sido excluídos das análises para fins estatísticos os indivíduos que apresentavam parentesco entre si, como é o caso dos filhos questionados em casos de paternidade (com vínculo genético em relação à mãe e/ou ao pai).

Os indivíduos cujo perfil constitui o banco de dados de laboratório público são, portanto, pessoas encaminhadas à Polícia Civil do DF pelas Varas de Família do Tribunal de Justiça do DF e Territórios ou pelas Delegacias de Polícia, envolvidas (como vítimas, suspeitos ou autores) em casos criminais em apuração por parte da PCDF. Para tanto, ao menos uma das partes interessadas deve ser beneficiária da justiça gratuita, segundo o Juiz.

Com relação ao sexo dos indivíduos amostrados, nos casos de paternidade a maioria apresenta quantidade igual de homens e mulheres (mães e supostos pais dos casos de investigação de paternidade). Nos casos de origem criminal, cerca de

2/3 é de homens e 1/3 de mulheres, resultando num total aproximado de 53% de perfis masculinos e 47% de perfis femininos.

Quanto às idades, dos 915 indivíduos amostrados, há registros de data de nascimento de 522 deles, sendo 33 menores de idade (6,3% menores de 18 anos), 231 entre 19 e 30 anos (44,3%), 161 entre 31 e 40 anos de idade (30,8%), 66 entre 41 e 50 anos (12,6%) e 31 com mais de 51 anos de idade (5,9%); a média de idade dos 522 com registro é de 32 anos.

Dentre os 915 indivíduos que constituem a amostra do laboratório público, há registro de local de nascimento de 720 deles, assim divididos: 44,7% nascidos no Centro-Oeste (desses, 69,3% de Brasília - Distrito Federal), 38,1% no Nordeste (quase a metade nativos do Piauí e da Bahia) e 17,2% oriundos da região Sudeste (desses, 78,2% nascidos em Minas Gerais).

Foram analisados 18 marcadores microssatélites - D16S539, D7S820, D13S317, D5S818, CSF1PO, TPOX, TH01, vWA, F13A01, FESFPS, F13B, LPL, Penta E, D18S51, D3S1358, D8S1179, D21S11 e FGA. Nem todos os indivíduos amostrados foram genotipados para todos os marcadores, visto que a metodologia de trabalho incluiu análises complementares em casos de índice de paternidade abaixo do utilizado na emissão de laudos, bem como diferentes *kits* de amplificação usados no decorrer dos exames. Desses dezoito locos STR, doze foram utilizados no desenvolvimento deste trabalho: D16S539, D7S820, D13S317, D5S818, CSF1PO, TPOX, TH01, vWA, D3S1358, D8S1179, D21S11 e FGA.

Os bancos de dados assim como os parâmetros forenses associados foram publicados pelo autor desse trabalho, encontrando-se em anexo (Dalton *et al.*, 2009 – Anexo D).

3.1.2 Amostra de laboratório privado do Distrito Federal

A amostra de laboratório de investigação de paternidade privado foi coletada pelo Laboratório Heréditas, situado no Distrito Federal. O referido laboratório particular é conhecido no Distrito Federal por ser o primeiro a realizar os exames de paternidade na própria Unidade da Federação, tendo por isso sido o responsável por atender uma demanda reprimida com relação aos exames de investigação de paternidade. A partir do ano 2000, com outros laboratórios concorrentes na região, o

custo dos exames (em todos eles) foi reduzido, devido à competição (inclusive junto ao poder judiciário).

Trata-se de uma amostra de 800 indivíduos genotipados para treze marcadores do tipo STR autossômicos, dos quais doze (D16S539, D7S820, D13S317, D5S818, CSF1PO, TPOX, TH01, vWA, D3S1358, D8S1179, D21S11 e FGA) foram utilizados no trabalho comparativo. O artigo contendo esse banco foi publicado por Grattapaglia e colaboradores, em 2001.

3.1.3 Diferenças sócio-econômicas entre os indivíduos que compõem as duas amostras representativas do Distrito Federal

Considerando que o acesso aos laboratórios público e privado dependem de aprovação por juiz, no caso do público, ou condições financeiras, no caso do privado, tem-se como premissa que a origem sócio-econômica dos usuários desses laboratórios seja distinta. Esta mesma relação de acesso e condição sócio-econômica já foi observada em outros países, como os Estados Unidos (House *et al.*, 1994), em que ficou evidente a ligação que educação e renda tem com a saúde dos indivíduos, medida pela exposição a fatores de risco durante toda a vida. Em linhas gerais, laboratórios públicos são procurados, em sua maioria, por pessoas com menor poder aquisitivo, que tem dificuldades para pagar por um exame particular. Já os laboratórios privados são mais utilizados por pessoas de uma classe econômica capaz de pagar pelo serviço. À época das coletas de ambas as amostras (final de década de 1990), o custo de um exame de paternidade em laboratório particular era de aproximadamente R\$500,00 (quinhentos reais), que equivalia a cerca de três salários-mínimos. No caso de processos cíveis de reconhecimento de paternidade, famílias que não tinham condição de pagar por um exame de paternidade biológica eram encaminhadas pelas Varas de Família do Tribunal de Justiça do Distrito Federal majoritariamente ao laboratório público do DF, no caso o laboratório de DNA da Polícia Civil. Os beneficiários da justiça gratuita eram atendidos por este laboratório, a despeito de um prazo maior para o atendimento e, como consequência, uma maior demora no andamento do processo.

Em síntese, a diferença primordial entre a composição das duas amostras em análise no presente estudo é o fato de que advém de pessoas cujo poder aquisitivo é distinto.

3.1.4 Amostras populacionais de outros estados brasileiros

Face à padronização proposta pelos laboratórios forenses do uso de marcadores STRs (Budowle, 1998; Butler, 2006), e principalmente pelo fato da maioria dos bancos de dados para fins forenses publicados enfocarem marcadores genéticos em comum, tornou-se possível analisar as amostras em questão face às frequências alélicas dos treze locos utilizados pelo sistema CODIS: CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 e D21S11.

Para avaliar as semelhanças e diferenças entre essas duas amostras e outras representativas do Brasil como um todo, bem como de diferentes estados brasileiros, foram utilizados dados concernentes às seguintes publicações:

- Brasil (Whittle *et al.*, 2004): amostra composta por cerca de 13.000 indivíduos obtida a partir de casos de paternidade oriundos de todo o Brasil, avaliada para os marcadores F13B, TPOX, D3S1358, FGA, CSF1PO, D5S818, F13A01, D7S820, D8S1179, D10S1237, TH01, vWA, D13S317, FESFPS, Penta E, D16S539, D18S51, D19S253 e D21S11;
- Estado de Alagoas (Silva *et al.*, 2002): amostra composta por 598 indivíduos analisada para os marcadores D16S539, D7S820, D13S317, CSF1PO, TPOX, TH01, F13A01, FESFPS e vWA;
- Estado do Mato Grosso do Sul (Silva *et al.*, 2004): amostra composta por 208 indivíduos analisada para os marcadores FGA, D8S1179, D3S1358, D18S51, D21S11, Penta E, TH01, TPOX, vWA, CSF1PO, D16S539, D7S820, D13S317 e D5S818;
- Estado de Pernambuco (Dellalibera *et al.*, 2004): amostra de 546 indivíduos (casos de paternidade de todo o Estado) analisada para os marcadores CSF1PO, TH01, TPOX, F13A01, FESFPS, vWA, D16S539, D7S820, D13S317, D18S51, D21S11, D8S1179 e FGA;
- Estado do Piauí (Da Silva *et al.*, 2003): amostra composta por 203 indivíduos (vivendo no Estado por ao menos três gerações) analisada para os marcadores CSF1PO, TPOX, TH01, D16S539, D7S820, D13S317, F13B e LPL;

- Estado do Rio de Janeiro (Góes *et al.*, 2004): amostra composta por 300 indivíduos analisada para os marcadores D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX e D3S1358;
- Estado do Rio de Janeiro (Silva *et al.*, 2004): amostra de 234 indivíduos analisada para os marcadores FGA, D8S1179, D3S1358, D18S51, D21S11, Penta E, TH01, TPOX, vWA, CSF1PO, D16S539, D7S820, D13S317 e D5S818;
- Estado do Rio Grande do Sul (Leite *et al.*, 2003): amostra composta por 104 indivíduos analisada para os marcadores D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11, FGA e vWA.

Dados de frequências alélicas obtidos nos trabalhos citados foram utilizados para análise de F_{ST} em relação à amostra do laboratório público do Distrito Federal. Para tanto, foram utilizados os programas Arlequin (Schneider *et al.*, 2000; Excoffier *et al.*, 2005) e GENEPOP (Raymond & Rousset, 1995). Para essa análise, foram usados os dezoito marcadores da amostra do laboratório público, comparadas aos locos em comum com cada um dos outros nove trabalhos citados.

3.2 Dados demográficos

Com vistas à comparação dos dados estatísticos relativos aos marcadores STRs e os valores demográficos populacionais (sociais, econômicos e culturais), foram obtidas as informações disponibilizadas pelo IBGE (Instituto Brasileiro de Geografia e Estatística) relativas às pesquisas populacionais de 1991, de 2000 e 2008 (respectivamente os Censos de 1991 e de 2000, e a PNAD 2007-2008).

Face às semelhanças observadas e discutidas na Introdução acerca dos dados econômicos e da raça, e em função das amostras analisadas apresentarem origem econômica distinta, as informações disponíveis no IBGE sobre raça foram usadas comparativamente com as inferências de ancestralidade geográfica realizadas no desenvolvimento do trabalho.

3.3 Análise estatística

Para a análise dos dados gerados quanto à distribuição alélica dos marcadores STR, foram utilizados os programas estatísticos Arlequin 2000 (Excoffier

et al., 2005; Schneider *et al.*, 2000), GENEPOP (Raymond & Rousset, 1995) e STRUCTURE (Pritchard *et al.*, 2000). As frequências genóticas, diferenciação populacional e desequilíbrio de ligação foram obtidos com auxílio do programa Arlequin 2000, enquanto a aderência ao equilíbrio de Hardy-Weinberg foi avaliada pelo GENEPOP 3.4.

3.3.1 Análises utilizando frequências alélicas

A partir das frequências alélicas, foram estimados para cada loco das duas amostras representativas do Distrito Federal, os parâmetros da estatística F: o F_{IS} , o F_{ST} e o F_{IT} utilizando o programa GENEPOP. Estes valores permitem avaliar o índice de endocruzamento nas amostras testadas, em relação aos indivíduos (I), subpopulações (S) e a população total (T).

A análise comparativa da amostra coletada em laboratório público do Distrito Federal (Dalton *et al.*, 2009) e outras amostras brasileiras foi efetuada com os programas GENEPOP 3.4 (Raymond & Rousset, 1995) e Arlequin 2000 (Schneider *et al.*, 2000). As amostras foram submetidas à comparação par a par, de acordo com suas frequências alélicas para cada um dos locos disponíveis, verificando se havia diferenciação entre elas, pelo F_{ST} . Este tratamento estatístico visou verificar se as duas amostras estudadas apresentavam maior ou menor semelhança entre si, quando comparadas a outras amostras populacionais brasileiras, avaliando se poderiam ou não ser consideradas semelhantes em termos do F_{ST} aplicado às frequências alélicas dos locos.

As análises de composição genética e das proporções de contribuição das populações parentais das duas amostras genotipadas em estudo foram efetuadas a partir de amostras das populações parentais estabelecidas em trabalhos anteriores (Godinho, 2008b; Gontijo, 2008). Foram consideradas as frequências ponderadas, levando em consideração o tamanho das amostras (número de indivíduos) genotipadas para as populações utilizadas. As frequências parentais e das amostras estudadas foram utilizadas para estimar a contribuição dos grupos parentais na composição das populações utilizando o programa ADMIX 3 (Chakraborty, 1986). Este programa estima as contribuições a partir da média ponderada de frequências alélicas das populações parentais (Chakraborty, 1986). Para esta análise, foram considerados os três grupos parentais de maior contribuição à constituição genética

da população brasileira: africanos (subsaarianos), europeus (basicamente portugueses e espanhóis) e ameríndios (da América do Sul).

Frequências alélicas de populações destas regiões (oeste da África subsaariana, península ibérica e ameríndios sul-americanos) foram obtidos para os marcadores analisados a partir de dezenas de trabalhos publicados, compilados por Godinho (2008b) e atualizados com auxílio do sítio “*Short Tandem Repeat Internet DataBase*” (<http://www.cstl.nist.gov/div831/strbase/>). Ressalte-se que nem todas as amostras utilizadas apresentavam frequências para os doze marcadores aqui estudados, havendo uma composição de diferentes referências populacionais para cada loco. As frequências alélicas para os 12 locos obtidas para cada conjunto de parentais estão apresentadas no Anexo A.

3.3.2 Análises utilizando dados genotípicos

Análises das proporções de contribuição das populações parentais na constituição dos indivíduos pertencentes às duas amostras populacionais foram estimadas com o programa STRUCTURE (desenvolvido por Pritchard *et al.*, 2000), associados ao DISTRUCT (Rosenberg *et al.*, 2002)

Para cálculo de ancestralidade no STRUCTURE, foi necessária a utilização de genótipos das populações parentais utilizadas. No presente trabalho, foram considerados os genótipos de 209 ibéricos (compostos por espanhóis da Andaluzia, da Catalunha e Bascos, bem como portugueses - Budowle & Moretti, 1999) de 210 indivíduos afro-americanos (Budowle & Moretti, 1999) e de 150 ameríndios (Hutz *et al.*, 2002), estes últimos gentilmente cedidos por Sídia Callegari-Jacques, constituindo-se de cinco tribos brasileiras (Gavião, Suruí, Zoró, Wai Wai e Xavante) estudadas para 15 marcadores (CD4, CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, F13A01, FGA, TH01, TPOX e vWA). Uma vez que a genotipagem desses marcadores em populações ameríndias só foi identificada para estas amostras, admitiu-se seu uso como parental nestas análises. Alternativamente, foi testado como parental africano um conjunto de 103 genótipos correspondendo a amostras populacionais africanas de sete países daquele continente (Quênia, Senegal, Nigéria, Namíbia, Congo, República Centro-Africana e África do Sul), obtidos do Centro de Estudos do Polimorfismo Humano (CEPH, 2010).

Para os cálculos do programa, os locos foram considerados independentes, face às análises anteriores que demonstram não haver ligação entre esses marcadores (Grattapaglia *et al.*, 2001; Butler, 2006; Dalton *et al.*, 2009; dentre outros), e após serem testados quanto ao Equilíbrio de Hardy-Weinberg (pelo programa *Genepop*). Os genótipos dos 2460 indivíduos para os 12 locos (D16S539, D7S820, D13S317, D5S818, CSF1PO, TPOX, TH01, vWA, D3S1358, D8S1179, D21S11 e FGA) foram analisados visando à inferência de sua composição genética com 1000 repetições, considerando três agrupamentos populacionais, em adequação às amostras consideradas parentais. Dentre os parâmetros do programa, utilizou-se uma indicação pré-cálculo de origem geográfica das amostras; deste modo, o programa assume que as parentais são de origens territoriais distintas, e tenta agrupar as duas outras amostras (de laboratórios do Distrito Federal) de acordo com os três grupos parentais.

O programa DISTRUCT (Rosenberg, 2004) possibilitou a visualização gráfica dos resultados produzidos pelo STRUCTURE, buscando mostrar por meio de diferentes cores a composição das amostras populacionais, pela visualização de cada um dos indivíduos lado a lado.

3.4 Aplicações à inferência populacional

Analisou-se neste trabalho a possibilidade de utilização dos doze marcadores em tela para inferências de origem populacional de um dado indivíduo com determinado perfil alélico. Foi adotada a suposição de que eventuais diferenças observadas entre as duas populações e as parentais poderiam ser avaliadas quanto à sua associação com bancos populacionais, possibilitando inferir acerca da origem de uma determinada amostra. Para isso, foram selecionadas as frequências alélicas e genóticas mais comuns entre cada amostra, e comparadas com bancos de dados estabelecidos como referência para identificação humana. Tais bancos são capazes de, a partir de um genótipo dado (mesmo incompleto), indicar a qual população ou região geográfica há maior possibilidade dele pertencer. Neste trabalho, para evitar erros de avaliação a partir de um indivíduo isolado (que pode não ser representativo da amostra), propôs-se analisar os genótipos de maior frequência de cada uma das amostras para estabelecer a maior afinidade deste grupo a uma região geográfica.

O uso de marcadores moleculares baseados em dinucleotídeos geralmente apresenta resultados melhores para inferências deste tipo (Rosenberg *et al.*, 2003), mas buscou-se aqui demonstrar a possibilidade do uso de tetranucleotídeos com esta finalidade. Dos doze marcadores STR usados neste trabalho, apenas quatro foram estudados por Rosenberg e colaboradores (2003) na confecção de planilhas que identificassem o potencial de resolução na determinação de origem populacional: D13S317, D16S539, D7S820 e D8S1179. O trabalho concluiu que o marcador com o maior poder de identificação para os nativos americanos era o D13S317 e o menor o D8S1179. Na discriminação dos europeus em relação aos africanos e americanos, o mais eficiente foi o D8S1179. Para os africanos, o marcador com menor potencial de identificação para isolar das demais parentais foi o D16S539, e o maior o D13S317. Para uso com as três parentais simultaneamente, o marcador mais indicado dentre os cinco tetranucleotídeos foi o D13S317, e o menos eficiente, o D16S539. Do exposto nos resultados do trabalho de Rosenberg *et al.* (2003), espera-se melhores resultados para determinação de inferência populacional para o marcador D13S317 e o menor poder de resolução para o loco D16S539.

No intuito de verificar a viabilidade de uso destes alelos característicos para determinados grupos populacionais, buscou-se selecionar os de frequências extremas (maiores e menores) dentre os resultados obtidos, e estudá-los. Dentre os três grupos populacionais distintos (ameríndios, europeus e africanos), as frequências alélicas e genotípicas apresentam diferenças sutis, em sua maioria, não podendo ser usadas para fins de caracterização de origem étnica. Entretanto, algumas frequências destacam-se em relação às demais, por valores percentuais expressivamente acima ou abaixo das médias parentais. A escolha destes para as análises seguintes buscou selecionar uma quantidade viável para os estudos propostos, ao mesmo tempo que apresentassem diferenças significativas. Para tanto, admitiu-se que diferenças de frequência de 10% (dez por cento) entre ao menos duas das populações parentais seriam testados.

Procedimento semelhante foi feito para os genótipos de maior e menor frequência entre os grupos parentais, estabelecendo-se que seriam escolhidos apenas genótipos com frequência de ocorrência superior a 2% no conjunto total de 2460 indivíduos. Foram selecionados para esta análise genótipos ausentes em dado

grupo parental ou alternativamente com uma frequência bem maior em um grupo parental em relação aos dois outros. Dessa forma, foi estabelecido o limite de ao menos 80% do genótipo em questão ocorrer em um dado grupo parental ou, alternativamente, que um dado genótipo não ocorresse em um dado grupo (frequência 0%), mas estivesse presente nos demais. Os genótipos assim selecionados foram buscados no banco de dados dos 2460 indivíduos. Indivíduos que apresentavam os genótipos selecionados foram avaliados acerca da amostra populacional de origem, de modo a verificar a eficiência dos alelos ou genótipos para a associação à população parental de origem do conjunto alélico/genotípico. O objetivo foi verificar se esses alelos/genótipos apresentavam bom poder de identificação sobre a origem populacional dos indivíduos selecionados.

A planilha OmniPop, compilada por Brian Burritt (2006) para o programa Microsoft Excel, calcula a frequência de ocorrência de um dado perfil genético para mais de 200 populações publicadas (incluindo oito amostras populacionais brasileiras). A planilha indica as respectivas frequências de ocorrência para cada uma das amostras populacionais, em ordem decrescente de probabilidade, o que possibilita inferir maior ou menor afinidade com populações de certa região geográfica. Com base nos alelos e/ou genótipos mais comuns para cada uma das duas amostras, laboratório público e privado, foram traçados perfis genéticos, e estes foram avaliados com o auxílio dessa planilha.

De forma similar, o site “*Earth Human Short Tandem Repeat Allele Frequencies Database*” (EHSTRAFD, 2009) apresenta uma ferramenta que visa indicar a “Origem Geográfica Mais Provável” (*MPGO - Most Probable Geographical Origin*) de um perfil genético, a partir de quinze marcadores STRs (os doze estudados no bojo deste trabalho e outros três, D2S1338, D18S51 e D19S433), para 451 populações previamente analisadas (incluindo 14 amostras populacionais brasileiras). Os mesmos perfis genéticos utilizados para OmniPop foram utilizados na planilha o EHSTRAFD.

4. RESULTADOS E DISCUSSÃO

4.1 Laboratórios Público e Privado: Caracterização genética e comparações

A Tabela 4.1 apresenta dados de frequências alélicas para os doze marcadores autossômicos e o teste exato de aderência ao equilíbrio de Hardy-Weinberg na amostra populacional do laboratório público em comparação ao previamente descrito para o laboratório privado (Grattapaglia *et al.*, 2001). No Anexo D está apresentada a reprodução integral da publicação do banco de dados da amostra populacional do laboratório público (Dalton *et al.*, 2009), onde também podem ser encontrados os parâmetros forenses. Os marcadores encontram-se em equilíbrio de Hardy-Weinberg em ambas as amostras, com exceção dos locos TH01 ($p = 0,005 \pm 0,002$) e D5S818 ($p = 0,006 \pm 0,001$) no laboratório privado e TPOX ($p = 0,028 \pm 0,009$) no laboratório público.

Com relação à diferenciação gênica e genotípica, observou-se divergência entre as amostras para quatro dos 12 marcadores avaliados: D13S317, CSF1PO, FGA e TH01 (Tabela 4.2).

A análise de desequilíbrio de ligação entre os doze locos analisados para as amostras dos laboratórios público e privado produziu resultados dentro do estatisticamente esperado (6,8% das combinações). Mesmo considerando-se que os marcadores CSF1PO e D5S818 estão localizados no cromossomo 5, os resultados não demonstram ligação entre os locos, fato já relatado anteriormente por outros estudos (por exemplo, Butler, 2006). Dentre as 132 possíveis combinações de pares de locos para as amostras populacionais representativas do Distrito Federal, nove apresentaram valores de p inferiores a 5%, nenhum destes associados aos marcadores do cromossomo 5 (Tabela 4.3).

TABELA 4.1 – Frequências alélicas e parâmetros estatísticos para os 12 marcadores genéticos do tipo STR avaliados para as amostras populacionais representativas do Distrito Federal, obtida de um laboratório público e um privado.

STR	D16S539		D7S820		D13S317		D5S818		CSF1PO		Alelos	FGA	
	Público	Privado	Público	Privado	Público	Privado	Público	Privado	Público	Privado		Público	Privado
Amostra													
Alelos													
5	-	0,0006	-	-	-	-	-	-	-	-	16	0,0009	-
6	-	-	0,0019	-	0,0019	-	-	-	0,0017	-	17	0,0037	-
7	-	-	0,0097	0,0120	0,0010	-	0,0165	0,0159	0,0132	0,0104	18	0,0129	0,0059
8	0,0232	0,0227	0,1754	0,1625	0,0902	0,1135	0,0272	0,0121	0,0281	0,0117	18.2	-	0,0007
9	0,1770	0,1591	0,1143	0,1239	0,0806	0,0950	0,0291	0,0292	0,0290	0,0175	19	0,0790	0,0635
10	0,1025	0,0844	0,2810	0,2585	0,0480	0,0555	0,0631	0,0628	0,2513	0,2633	20	0,1085	0,1119
11	0,2930	0,2955	0,2355	0,2560	0,2831	0,2978	0,3214	0,3280	0,2900	0,3067	21	0,1379	0,1682
12	0,2621	0,2500	0,1521	0,1625	0,3109	0,2736	0,3757	0,3687	0,3304	0,3294	22	0,1710	0,1774
13	0,1170	0,1558	0,0291	0,0215	0,1392	0,1135	0,1524	0,1688	0,0501	0,0538	22.2	-	0,0020
14	0,0242	0,0292	0,0010	0,0032	0,0441	0,0478	0,0117	0,0127	0,0053	0,0071	23	0,1654	0,1427
15	0,0010	0,0026	-	-	0,0010	0,0032	0,0029	0,0019	0,0009	-	24	0,1351	0,1499
											25	0,1103	0,1099
											26	0,0404	0,0445
											27	0,0257	0,0151
											28	0,0046	0,0039
											29	0,0018	0,0007
											30	0,0009	0,0026
											30.2	-	0,0007
											31.2	0,0009	-
											32	-	0,0007
											43.2	0,0009	-
<i>Ho</i>	0,749	0,783	0,800	0,802	0,772	0,786	0,736	0,709	0,731	0,722	<i>Ho</i>	0,853	0,859
<i>He</i>	0,790	0,793	0,798	0,799	0,786	0,797	0,727	0,723	0,740	0,725	<i>He</i>	0,875	0,867
<i>p</i>	0,352 ±	0,057 ±	0,570 ±	0,767 ±	0,788 ±	0,994 ±	0,813 ±	0,006 ±	0,279 ±	0,849 ±	<i>p</i>	0,375 ±	0,118 ±
	0,023	0,015	0,040	0,023	0,036	0,001	0,020	0,001	0,023	0,015		0,041	0,032
<i>n</i>	517	770	516	791	521	784	515	788	569	771	<i>n</i>	544	764

TABELA 4.1 (continuação)

STR	TPOX		TH01		vWA		D3S1358		D8S1179		Alelos	D21S11	
	Público	Privado	Público	Privado	Público	Privado	Público	Privado	Público	Privado		Público	Privado
Amostra													
Alelos											Alelos		
5	-	0,0007	-	0,0014	-	-	-	-	-	-	14	-	0,0006
6	0,0191	0,0111	0,1952	0,2062	-	-	-	-	-	-	15	-	0,0006
7	0,0076	0,0026	0,2697	0,2220	-	-	-	-	-	-	24.2	0,0037	0,0026
8	0,4454	0,4452	0,1721	0,1236	-	-	-	-	0,0065	0,0079	25	0,0019	0,0013
9	0,1246	0,1258	0,1409	0,1753	-	-	-	-	0,0102	0,0112	26.2	-	0,0006
9.3	0,0623	0,0737	0,2155	0,2579	-	-	-	-	-	-	27	0,0317	0,0281
10	0,2902	0,2920	0,0055	0,0136	-	-	-	-	0,0605	0,0745	28	0,1825	0,1411
11	0,0497	0,0482	0,0011	-	0,0038	0,0019	-	0,0006	0,0670	0,0884	29	0,1955	0,2171
12	0,0011	0,0007	-	-	0,0011	-	0,0028	0,0025	0,1313	0,1214	29.2	0,0037	0,0006
13	-	-	-	-	0,0044	0,0058	0,0056	0,0031	0,2551	0,2850	30	0,2440	0,2522
14	-	-	-	-	0,0760	0,0845	0,0950	0,0926	0,2766	0,2447	30.2	0,0298	0,0377
15	-	-	-	-	0,1344	0,1229	0,3017	0,2922	0,1462	0,1266	31	0,0689	0,0626
15.2	-	-	-	-	-	-	-	0,0006	-	-	31.2	0,0931	0,0990
16	-	-	-	-	0,2760	0,2458	0,2709	0,2519	0,0419	0,0350	32	0,0102	0,0121
17	-	-	-	-	0,2661	0,2740	0,1946	0,2185	0,0047	0,0053	32.2	0,0801	0,1003
18	-	-	-	-	0,1677	0,1844	0,1136	0,1266	-	-	33	0,0037	0,0019
19	-	-	-	-	0,0503	0,0608	0,0149	0,0107	-	-	33.2	0,0326	0,0326
20	-	-	-	-	0,0164	0,0186	0,0009	0,0006	-	-	34	0,0028	0,0013
21	-	-	-	-	0,0038	0,0013	-	-	-	-	34.2	0,0037	0,0026
											35	0,0056	0,0038
											35.2	0,0009	-
											36	0,0037	0,0013
											37	0,0009	-
											38	0,0009	-
<i>Ho</i>	0,691	0,671	0,797	0,797	0,811	0,795	0,782	0,762	0,793	0,831	<i>Ho</i>	0,827	0,836
<i>He</i>	0,696	0,693	0,794	0,796	0,799	0,805	0,776	0,779	0,810	0,814	<i>He</i>	0,847	0,842
<i>p</i>	0,028 ±	0,116 ±	0,198 ±	0,005 ±	0,572 ±	0,432 ±	0,176 ±	0,769 ±	0,208 ±	0,568 ±	<i>p</i>	0,780 ±	0,801 ±
	0,009	0,011	0,028	0,002	0,034	0,034	0,036	0,039	0,022	0,024		0,064	0,048
<i>n</i>	915	767	912	695	915	781	537	794	537	758	<i>n</i>	537	782

TABELA 4.2 – Diferenciação gênica e genotípica (teste G exato) por loco entre as amostras populacionais dos laboratórios público e privado do Distrito Federal. Em negrito são mostrados os valores de $p < 5\%$.

Loco	Diferenciação gênica (valor de p)	Diferenciação genotípica (valor de p)
D16S539	0,107 ± 0,014	0,111 ± 0,014
D7S820	0,274 ± 0,020	0,257 ± 0,017
D13S317	0,032 ± 0,009	0,026 ± 0,006
D5S818	0,388 ± 0,023	0,379 ± 0,021
CSF1PO	0,018 ± 0,004	0,022 ± 0,004
TPOX	0,225 ± 0,018	0,264 ± 0,021
TH01	0,000 ± 0,000	0,000 ± 0,000
vWA	0,185 ± 0,016	0,203 ± 0,018
D21S11	0,291 ± 0,025	0,300 ± 0,021
D3S1358	0,794 ± 0,017	0,809 ± 0,014
FGA	0,012 ± 0,003	0,018 ± 0,004
D8S1179	0,185 ± 0,021	0,131 ± 0,015

TABELA 4.3 – Valores de p menores que 5% para as 132 análises de desequilíbrio de ligação entre os locos das amostras populacionais dos laboratórios público e privado.

Amostra populacional	Locos	Valor de p
Laboratório público	D13S317 x TPOX	0,0233 ± 0,0118
	TH01 x vWA	0,0337 ± 0,0167
	D13S317 x D21S11	0,0220 ± 0,0154
	D16S539 x D3S1358	0,0356 ± 0,0126
	D21S11 x D3S1358	0,0000 ± 0,0000
	D16S539 x FGA	0,0435 ± 0,0236
Laboratório privado	D13S317 x FGA	0,0401 ± 0,0216
	TH01 x D8S1179	0,0000 ± 0,0000
	D3S1358 x D8S1179	0,0165 ± 0,0125

A análise comparativa das frequências alélicas de maior discrepância entre as amostras dos laboratórios público e privado buscou também avaliar os alelos que apresentaram um diferencial de frequência maior que 2%. O resultado dessa análise está na Tabela 4.4. Observa-se que a população parental africana apresentou uma maior participação na constituição genética da amostra do laboratório público enquanto que a europeia no laboratório privado. Exemplificando: no caso do Alelo 10 do loco D7S820 (D7S820*10), a frequência alélica do grupo africano (34,4%) é superior à dos laboratórios público (28,10%) e privado (25,85%), de modo que para este loco esta parental deve ter contribuído para elevar o percentual de africanos no laboratório público. Em princípio, quanto maior o percentual de frequência do alelo 10 em D7S820, maior a participação africana na composição genética.

Com base nas frequências alélicas dos 12 marcadores, foram estimados, F_{IS} , o F_{ST} e o F_{IT} para cada loco analisado em ambas as amostras (Tabela 4.5). Com exceção do TH01, os demais locos apresentaram valores de F_{IS} e F_{IT} bastante semelhantes, o que reflete a ocorrência de endogamia dentro das subpopulações, e relativamente superiores aos valores de F_{ST} . Isto parece indicar que as maiores diferenças entre as amostras são decorrentes de diferenças entre os indivíduos que compõem as duas, e não entre as amostras como um todo. O maior valor de F_{ST} , estimado para o loco TH01, foi de 0,0043, que apresentou um p significativo (Tabela 4.5), enquanto que o valor global, considerando todos os locos, foi de 0,0007. Desta forma, apesar das pequenas diferenças observadas para alguns locos, as duas amostras parecem representativas de uma mesma população, já que os valores obtidos são relativamente baixos, indicando haver fluxo gênico e, em princípio, frequências alélicas semelhantes.

TABELA 4.4 - Alelos cujas frequências observadas (em %) apresentaram um diferencial de frequência acima de 2% na comparação da amostra do laboratório público com o privado e valores dos mesmos nas parentais.

Marcador	Alelos	Público	Privado	Europeus	Africanos	Ameríndios
D16S539	13	11,70	15,58	14,2	16,6	0,8
D7S820	10	28,10	25,85	27,8	34,4	22,5
	11	23,55	25,60	24,7	16,8	41,0
D13S317	8	9,02	11,35	11,4	1,0	0,8
	12	31,09	27,36	26,4	66,4	12,9
	13	13,92	11,35	11,6	14,8	21,4
TH01	7	26,97	22,20	20,8	62,9	47,9
	8	17,21	12,36	13,8	24,9	0,4
	9	14,09	17,53	16,9	7,6	0,4
	9.3	21,55	25,79	26,0	1,0	18,0
vWA	16	27,60	24,58	25,9	26,3	50,7
D21S11	28	18,25	14,11	13,7	32,0	0,5
	29	19,55	21,71	22,0	9,9	16,9
	32.2	8,01	10,03	9,4	3,5	24,4
D3S1358	17	19,46	21,85	20,2	17,7	14,8
FGA	21	13,79	16,82	16,8	5,7	0,8
	23	16,54	14,27	15,2	11,5	9,9
D8S1179	11	6,70	8,84	8,6	1,2	1,4
	13	25,51	28,50	29,1	9,1	34,0
	14	27,65	24,47	25,3	37,7	20,5

TABELA 4.5. – Comparação entre as duas amostras populacionais do Distrito Federal por loco com base na estatística de Wright (F_{IS} , F_{ST} e F_{IT}).

Loco	F_{IS}	F_{ST}	F_{IT}
D16S539	0,0280	0,0006	0,0286
D7S820	-0,0026	0,0001	-0,0025
D13S317	0,0155	0,0011	0,0166
D5S818	0,0066	-0,0004	0,0062
CSF1PO	0,0071	-0,0002	0,0070
TPOX	0,0181	-0,0004	0,0176
TH01	-0,0024	0,0043	0,0018
vWA	-0,0028	0,0004	-0,0024
D21S11	0,0139	0,0009	0,0148
D3S1358	0,0102	-0,0000	0,0102
FGA	0,0159	0,0004	0,0163
D8S1179	-0,0037	0,0011	-0,0025
Total	0,0083	0,0007	0,0090

Não foram observadas diferenças significativas no perfil genético das amostras populacionais do Distrito Federal quando comparadas às outras populações urbanas brasileiras. A Tabela 4.6 apresenta os valores de p , quanto às estimativas de diferenciação populacional (F_{ST}) entre a amostra do laboratório público em relação à amostra do privado e nove outras amostras populacionais, as quais: representam a população brasileira como um todo, Alagoas, Mato Grosso do Sul, Pernambuco, Piauí, Rio de Janeiro e Rio Grande do Sul (Whittle *et al.*, 2004; Silva *et al.*, 2002; Silva *et al.*, 2004; Dellalibera *et al.*, 2004; Da Silva *et al.*, 2003; Silva *et al.*, 2004 e Góes *et al.*, 2004; Leite *et al.*, 2003). Observa-se que, comparadas loco a loco, as amostras são muito similares, destacando-se a população do Piauí com um terço das estimativas de F_{ST} significativas. Porém, esse é o banco de dados com o menor número de marcadores analisados, o que pode justificar o alto valor observado como sendo o acaso de terem sido comparados

alguns dos locos com maiores diferenças; provavelmente com maior número de marcadores os resultados seriam mais semelhantes. O TH01 novamente foi o marcador genético que se destacou, pois apresentou diferenças entre a amostra do laboratório público e três outras amostras: a do laboratório privado, a representativa do Brasil e a do Rio de Janeiro.

As amostras populacionais do Distrito Federal analisadas neste trabalho apresentaram diferenciação baixa (indicação de serem amostras similares), resultando em diferença significativa ($p < 2\%$) apenas para o loco TH01. Esta comparação estatística inicial demonstrou, a princípio, que ambas poderiam ser consideradas oriundas de uma mesma população. As diferenças entre os marcadores STR autossômicos foram mínimas, dentro das expectativas de 95% nos testes de chi-quadrado. Estes resultados repetiram-se com as outras amostras populacionais comparadas (Tabela 4.6), demonstrando que a população do DF apresenta um padrão semelhante ao das regiões Centro-Oeste, Sudeste, Sul e Nordeste.

Depreende-se do até aqui exposto que as duas amostras populacionais dos laboratórios do Distrito Federal apresentam semelhanças, a ponto de poderem ser consideradas oriundas de uma única população.

TABELA 4.6 – Valores de p para as análises de F_{ST} entre pares de populações para um máximo de 18 marcadores do tipo STRs autossômicos entre a amostra populacional do laboratório público do Distrito Federal e a amostra do laboratório privado, amostras representativas do Brasil e populações urbanas de seis estados brasileiros. Em negrito os valores de p inferiores a 5%, evidenciando diferenças entre as amostras (% = n° de p significativos / n° de marcadores avaliados).

	CSF1PO	D3S1358	D5S818	D7S820	D8S1179	D13S317	D16S539	D18S51	D21S11	TH01	TPOX	vWA	%
Privado	0,749	0,760	0,893	0,733	0,275	0,273	0,410	0,997	0,338	0,004	0,926	0,493	8,3
BR	0,561	0,796	0,329	0,424	0,032	0,258	0,074	0,924	0,253	0,000	0,166	0,354	16,7
AL	0,601	-	-	0,475	-	0,978	0,535	-	-	0,165	0,529	0,674	0,0
MS	0,449	0,299	0,540	0,918	0,847	0,742	0,496	0,812	0,249	0,245	0,139	0,855	0,0
PE	0,243	-	-	0,856	0,081	0,209	0,283	0,603	0,000	0,670	0,998	0,872	10,0
PI	0,146	-	-	0,000	-	0,000	0,915	-	-	0,925	0,994	-	33,3
RJ1	-	0,980	0,925	0,689	0,809	0,892	0,308	0,898	-	0,974	0,591	0,676	0,0
RJ2	0,136	0,240	0,542	0,895	0,776	0,647	0,181	0,002	0,797	0,000	0,200	0,238	16,7
RS	-	0,896	0,399	0,898	0,125	0,862	-	0,995	-	-	-	0,994	0,0

Privado (Grattapaglia *et al.*, 2001); BR – Brasil (Whittle *et al.*, 2004); AL – Alagoas (Silva *et al.*, 2002); MS –Mato Grosso do Sul (Silva *et al.*, 2004); PE –Pernambuco (Dellalibera *et al.*, 2004); PI –Piauí (da Silva *et al.*, 2003); RJ1 –Rio de Janeiro (Góes *et al.*, 2004); RJ2 –Rio de Janeiro (Silva *et al.*, 2004); RS –Grande do Sul (Leite *et al.*, 2003).

4.2 Mistura genética

Apesar dos resultados estatísticos indicarem que as amostras dos laboratórios público e privado podem ser consideradas semelhantes, as estimativas de mistura genética indicaram dissimilaridades. Na Tabela 4.7 estão apresentadas as estimativas (e os respectivos erros) de contribuição das populações parentais na composição das duas amostras populacionais.

TABELA 4.7 – Estimativas (percentuais e respectivos erros) de contribuição das populações parentais na composição das amostras dos laboratórios público e privado do Distrito Federal e a variação entre as contribuições.

	Europeu	Africano	Ameríndio	R ²
Público (n=915)	61,2 ± 1,6	26,6 ± 1,3	12,2 ± 0,8	0,99915
Privado (n=800)	77,7 ± 0,7	13,2 ± 0,6	9,0 ± 0,3	0,99985
Variação	+ 16,5 %	- 13,4 %	- 3,2 %	

A constituição genética das duas amostras evidenciou diferença entre elas. A maior variação observada foi com relação à contribuição europeia, mais acentuada na amostra do laboratório privado em 16,5%. Por outro lado, as contribuições das parentais africanas e ameríndias apresentaram-se mais elevadas no laboratório público. Na Tabela 1.2, observa-se que a estimativa de contribuição das populações parentais para a região Centro-Oeste foi de aproximadamente 71% de europeus, (Callegari-Jacques *et al.*, 2003), isto é, 10% acima do estimado para a amostra do laboratório público. A variação para ameríndios foi bem menor, tendo sido estimada em 11% (Callegari-Jacques *et al.*, 2003) enquanto que no laboratório público foi de aproximadamente 12%.

O trabalho de Lins e colaboradores (2010), feito com SNPs informativos de ancestralidade mostrou valores intermediários para a região Centro-Oeste em relação aos dois laboratórios (69,5% de europeus, 18,7% de africanos e 11,8% de ameríndios). Esse resultado (apesar de obtido com SNPs adequados para análise de ancestralidade) indica que ambos os laboratórios (público e privado) podem apresentar pequenos desvios em relação à média da região Centro-Oeste.

Há, aqui, que se considerar o aspecto (mesmo que não em sua totalidade) do nível econômico de procura e utilização deste laboratório privado para realização de exames de paternidade. À época das coletas e exames, o custo desses testes era usualmente inacessível às classes de menor poder aquisitivo.

O estudo de Martínez e colaboradores (2007) realizado na cidade de Caracas (Venezuela) evidenciou diferenciação semelhante de amostras coletadas em uma mesma população. Apesar de utilizar apenas três locos STR entre os 23 marcadores moleculares analisados, o resultado daqueles já é suficiente para mostrar que a coleta com viés econômico-social evidencia diferenças na composição genética.

Como citado anteriormente, e uma vez que a diferença fundamental entre as amostras dos laboratórios do Distrito Federal é a composição majoritária de estratos sócio-econômicos distintos, pode-se fazer uma associação de participação ancestral e classe social. Relaciona-se a maior composição europeia à amostra obtida pelo laboratório privado, e a maior participação de frequências alélicas caracteristicamente africanas na amostra de laboratório público. O resultado obtido evoca a relação discutida na Introdução acerca das classes de menor poder aquisitivo apresentarem maior ancestralidade africana, em decorrência da história brasileira. Esse mesmo histórico relaciona a ascendência europeia às classes dominantes do período de colonização, mantendo até hoje dominância financeira e étnica entre as classes de maior poder aquisitivo.

4.2.1 Análise de estrutura populacional com o programa STRUCTURE

4.2.1.1 Populações parentais

Na Tabela 4.8 estão apresentadas as análises quanto à aderência ao Equilíbrio de Hardy-Weinberg do agrupamento de genótipos das três populações. Essas análises revelaram desequilíbrio em 50% dos marcadores na população de ameríndios. No caso das populações ameríndias é importante frisar que os genótipos utilizados foram obtidos da análise de cinco diferentes tribos pertencentes por vezes a línguas diferentes e situadas em regiões geográficas tão dissimilares quanto Amazônia e Mato Grosso. As demais amostras parentais apresentaram-se em equilíbrio de Hardy-Weinberg. No caso dos europeus (conjunto de portugueses e espanhóis catalães, bascos e da Andaluzia), mesmo com as diferenças culturais

históricas de alguns grupos, todos os locos apresentaram equilíbrio de Hardy-Weinberg.

TABELA 4.8 – Valores de p e respectivos desvios-padrão para o teste exato quanto ao Equilíbrio de Hardy-Weinberg nas populações parentais.

Loco	Afro-americanos	Ibéricos	Ameríndios
D16S539	0,097 ± 0,009	0,339 ± 0,019	0,384 ± 0,018
D7S820	0,456 ± 0,020	0,789 ± 0,018	0,023 ± 0,003
D13S317	0,292 ± 0,020	0,767 ± 0,013	0,030 ± 0,004
D5S818	0,498 ± 0,028	0,409 ± 0,023	0,456 ± 0,020
CSF1PO	0,982 ± 0,003	0,162 ± 0,020	0,356 ± 0,018
TPOX	0,456 ± 0,016	0,977 ± 0,004	0,006 ± 0,002
TH01	0,144 ± 0,011	0,212 ± 0,024	0,844 ± 0,009
vWA	0,322 ± 0,020	0,853 ± 0,016	0,600 ± 0,015
D21S11	0,612 ± 0,035	0,461 ± 0,032	0,001 ± 0,001
D3S1358	0,744 ± 0,021	0,640 ± 0,025	0,002 ± 0,001
FGA	0,998 ± 0,01	0,085 ± 0,020	0,100 ± 0,014
D8S1179	0,719 ± 0,021	0,274 ± 0,022	0,300 ± 0,019

4.2.1.2 Análise de estrutura populacional

A Tabela 4.9 apresenta os dados obtidos a partir da análise realizada com o programa STRUCTURE. A parental ameríndia foi a que mais se concentrou em um único agrupamento (Grupo 3), revelando ser a população mais homogênea (96,7% no mesmo grupo), seguramente reflexo de menor mistura genética com outras populações consideradas parentais, apesar da amostra não refletir um agrupamento real, como já discutido. Os genótipos que compuseram a amostra de ibéricos também apresentaram bom agrupamento (no Grupo 1, que caracterizou a origem europeia), com índice de 87,0%. Porém, a parental africana (representada por um grupo de afro-americanos) não apresentou agrupamento bem definido, mas distribuiu-se entre os Grupos 1 e 2 (que estão representando, respectivamente, os padrões genéticos europeus e africanos). Este resultado sugeriu a existência de mistura genética destes dois continentes na formação dos genótipos destes

indivíduos, o que pode estar refletindo um viés nas análises de composição genética utilizando estas populações como parentais.

Com a utilização de outro banco de dados genotípicos para a parental africana (mantendo-se os demais como na análise anterior), obtido a partir de 103 genótipos de amostras do CEPH (*Centre d'Etude du Polymorphisme Humain* - <http://www.cephb.fr>), mostrou resultado semelhante com relação ao agrupamento das populações parentais. O grupo 2 concentrou bem a amostra parental ibérica (87,9%) e as dos laboratórios público (87,7%) e privado (95,8%). O agrupamento 3 repetiu o resultado com as parentais afro-americanas, concentrando a amostra ameríndia (96,8%). O conjunto de genótipos africanos do CEPH dividiu-se entre os grupos 1 (60,6%) e 2 (39,1%), também demonstrando que não está adequado para análises de mistura tri-híbrida para amostras da população brasileira, neste caso possivelmente por não refletir de maneira exata as populações que compuseram a fração africana no processo de miscigenação do Brasil.

Procedeu-se também à análise pelo programa STRUCTURE utilizando o conjunto dos genótipos das duas referências (CEPH e Budowle & Moretti, 1999). Neste caso, os resultados foram inconsistentes, mesmo com o uso de diferentes valores de k , com agrupamentos variando de 4 a 9. Ficou claro que, a princípio, o uso em conjunto desses dois bancos genotípicos é inadequado como parental africano para a análise da composição genética da população brasileira (Anexo C).

Os melhores resultados obtidos foram com os genótipos do CEPH, que apesar de não serem ideais (possivelmente por agruparem também indivíduos de populações de pigmeus entre os subsaarianos), foram utilizados para as demais análises dos programas STRUCTURE e DISTRICT, como parental africana. A Tabela 4.9a mostra os resultados produzidos pelo programa STRUCTURE para $k=3$ (considerando as amostras populacionais como resultado de mistura tri-híbrida).

Foram feitas, ainda, análises com o programa STRUCTURE utilizando-se outros valores de k , que são os agrupamentos estabelecidos para verificação de existência de sub-estruturas evidentes. No caso dessas amostras da população brasileira, e considerando sua origem tri-híbrida (com apenas três parentais), os agrupamentos adicionais não produziram resultados diferentes da análise com $k=3$. As Tabelas 4.9b e 4.9c, mostram os resultados do programa para 4 e 6 agrupamentos, respectivamente. A mesma análise (para $k=4$ e $k=6$) foi feita para os

parentais afro-americanos (ao invés dos africanos), produzindo resultados quase idênticos.

TABELA 4.9 – Estrutura dos cinco grupos populacionais com relação à contribuição genética advinda dos três grupos parentais – africanos, europeus e ameríndios - obtidas pela análise de estruturação individual (STRUCTURE), para diferentes números de agrupamentos (k) estabelecidos para o programa: $k=3$ (a), $k=4$ (b); e $k=6$ (c).

4.9a ($k=3$)

Amostra populacional	Grupo 1	Grupo 2	Grupo 3
Laboratório público (n=915)	0,877	0,107	0,016
Laboratório privado (n=800)	0,958	0,033	0,009
Africanos (n=103)	0,391	0,606	0,003
Ibéricos (n=384)	0,879	0,002	0,119
Ameríndios (n=150)	0,030	0,002	0,968

4.9b ($k=4$)

Amostra populacional	Grupo 1	Grupo 2	Grupo 3	Grupo 4
Laboratório público (n=915)	0,876	0,101	0,023	0,001
Laboratório privado (n=800)	0,960	0,028	0,010	0,002
Africanos (n=103)	0,404	0,591	0,003	0,003
Ibéricos (n=384)	0,874	0,001	0,123	0,001
Ameríndios (n=150)	0,026	0,002	0,971	0,000

4.9c ($k=6$)

Amostra populacional	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6
Lab. público (n=915)	0,863	0,086	0,048	0,001	0,001	0,001
Lab. privado (n=800)	0,942	0,021	0,035	0,001	0,001	0,001
Africanos (n=103)	0,458	0,537	0,002	0,002	0,000	0,001
Ibéricos (n=384)	0,838	0,002	0,156	0,002	0,001	0,002
Ameríndios (n=150)	0,029	0,001	0,968	0,001	0,000	0,001

As Figuras 4.1a e 4.1b foram obtidas a partir dos dados de análise individual gerados pelo STRUCTURE. As figuras evidenciam a questão da parental africana, que foi a que menos se agrupou num dos extremos. Com a utilização dos 103 genótipos africanos do CEPH (Figura 4.1a e Tabela 4.9a), os percentuais foram de 40% no grupo 1 e 60% no grupo 2, e com o uso dos 211 genótipos afro-americanos (Figura 4.1b), os percentuais inverteram-se. Como consequência, os genótipos dos indivíduos africanos (e ainda mais nos afro-americanos, cujos percentuais eram respectivamente de 57% e 43%), representados pela cor azul nas Figuras 4.1a e 4.1b, posicionaram-se entre dois dos cantos do triângulo. Tal fato provavelmente decorreu da presença de mistura de material genético europeu nos indivíduos da amostra africana, ou de sua inadequação como parental para a população tri-híbrida brasileira, distanciando seu perfil de uma maior participação no Grupo 2 da Tabela 4.9a e no canto superior do triângulo das Figuras 4.1 (a e b).

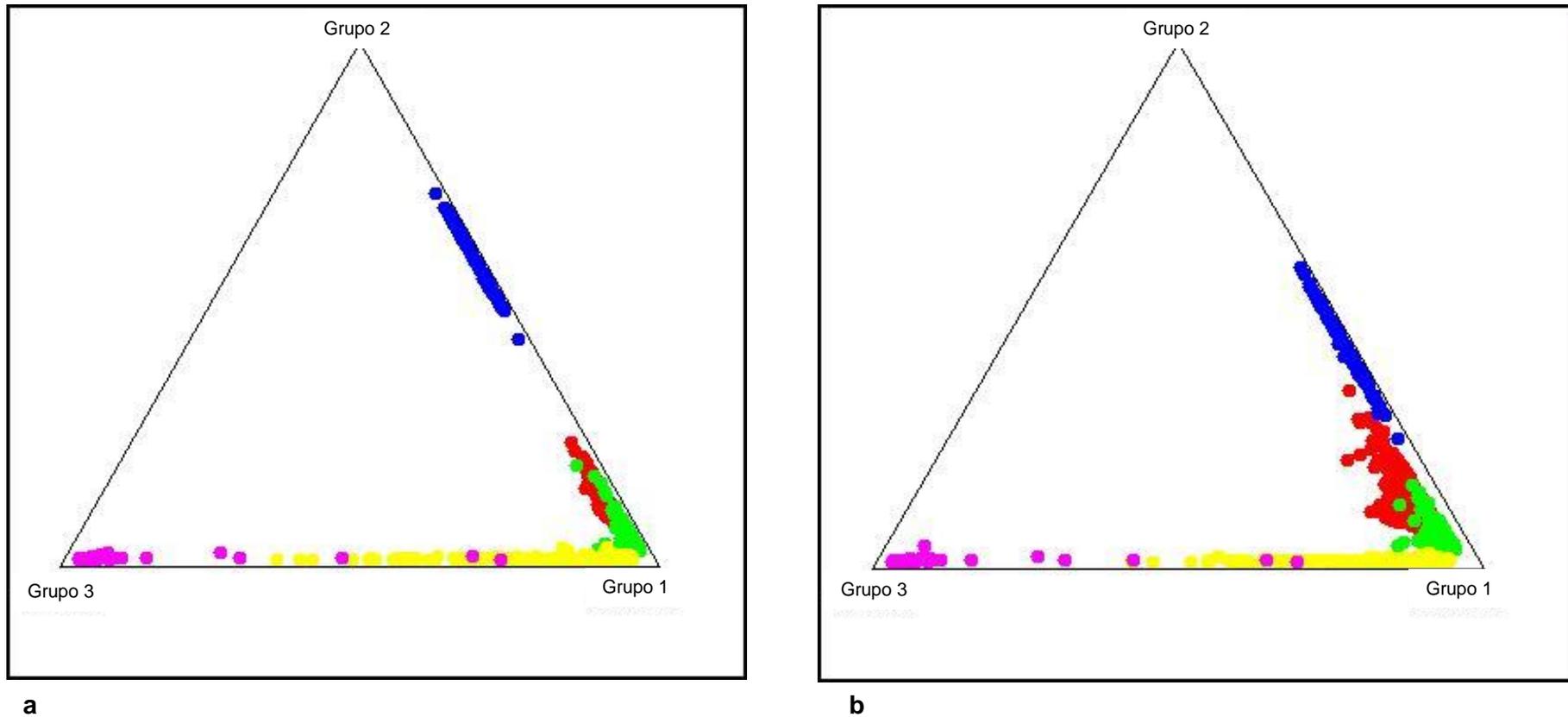


FIGURA 4.1 – Gráficos representativos da composição individual (STRUCTURE) das amostras populacionais do laboratório público (vermelho), do privado (verde), de ameríndios (lilás), ibéricos (amarelo) e africanos (azul) para genótipos relativos a 12 marcadores STRs autossômicos. A Figura 4.1a foi obtida com o parental africano de 103 genótipos obtidos do Centro de Estudos do Polimorfismo Humano (CEPH, 2010), enquanto a Figura 4.1b foi produzida pelo mesmo programa com parental africano a partir de 211 indivíduos afro-americanos (Budowle & Moretti, 1999).

Isto também se reflete no resultado das amostras aqui analisadas (laboratórios público e privado), uma vez que considera-se como parental uma amostra que aparenta possuir mistura oriunda também do Grupo 1, ou de não ser exatamente representativa das populações africanas que compõem a genética brasileira. Esta distribuição dos indivíduos africanos entre os grupos 1 e 2 (representativos de europeus e africanos) se apresenta nos dois grupos parentais analisados (CEPH, 2010 e Budowle & Moretti, 1999). Outro fator a ser considerado é que os marcadores utilizados para estudos de identificação humana foram, na sua grande maioria, selecionados na América do Norte e na Europa. Dessa forma, esses marcadores estão otimizados para essas populações em especial. Um outro fator a ser considerado é que o tamanho amostral das populações parentais é inferior ao tamanho das populações em teste, o que pode contribuir para desvios com relação a composição efetiva dessas populações.

A amostra do laboratório privado mostrou um percentual de ancestralidade europeia superior ao apresentado pela amostra parental de europeus (Tabela 4.9 e Figura 4.1), nas duas situações com parentais africanos diferentes. No caso dos africanos do CEPH, o laboratório público apresentou percentual de ancestralidade semelhante ao dos hispânicos. Por outro lado, a esperada maior contribuição africana na amostra pública, que corroboraria as estimativas obtidas pelo Admix, foi observada. A amostra do laboratório público apresenta um índice de participação africana maior que o do laboratório privado, situando seus respectivos indivíduos em pontos intermediários aos parentais europeus e africanos. Há uma sobreposição entre indivíduos das duas amostras do Distrito Federal (canto inferior direito da Figura 4.1, representando o agrupamento 1), por outro lado observa-se uma ampla separação dos indivíduos que compõem as duas amostras.

Dessa forma, fica evidente que a amostra de laboratório privado está mais próxima dos parentais europeus que a pública. Não foi observada contribuição ameríndia evidente na composição genética do Distrito Federal.

4.2.2 Análise comparativa pelo programa DISTRUCT

Com o objetivo de visualizar melhor o resultado obtido pelo programa STRUCTURE na imagem do triângulo, foi utilizado o *software* DISTRUCT (Rosenberg, 2004), que produziu uma imagem composta pelos 2460 genótipos

analisados (ou 2342 genótipos, no caso da parental africana obtida a partir do CEPH), atribuindo a cada um dos indivíduos amostrados (por meio de linhas verticais lado a lado) as proporções de cada agrupamento, indicados por diferentes cores.

O resultado obtido está mostrado na Figura 4.2, com as amostras do laboratório público aparecendo na parte esquerda, seguidas dos genótipos do laboratório privado, e a seguir os padrões parentais de africanos (em verde), europeus (em rosa) e ameríndios (em azul). O que se observa, mais uma vez, é a dificuldade do programa STRUCTURE em isolar os genótipos africanos ou afro-americanos como parentais, sugerindo a presença de contribuição europeia na amostra ou a inadequação das populações africanas consideradas para a análise tri-híbrida do povo brasileiro (constatada pela presença das cores verde e rosa em proporções quase iguais na média dos indivíduos).

O padrão europeu apresenta uma fração de coloração azul, atribuição dada pelo programa a uma característica comum entre estes e os ameríndios. Este padrão indica, antes de tudo, uma característica genotípica ausente no perfil africano, mais do que uma semelhança entre europeus e ameríndios.

Já a amostra denominada parental ameríndia foi a que melhor se agrupou num vértice, apresentando índice de 96,8% (Tabela 4.9), sendo provavelmente a que apresenta menos grau de mistura genética nos locos avaliados, para os fins desta análise.

O que se vê em relação às amostras analisadas neste trabalho é que em ambas a participação genética europeia é majoritária, mas no laboratório público há uma contribuição de ancestralidade africana proporcionalmente maior que no laboratório privado (Figura 4.2). Os perfis genotípicos de ambas as amostras, quando comparados, indicam presença de subestruturas distintas, com contribuições europeias e africanas distintas. Tal fato indica, a princípio, que duas amostras de uma mesma população, diferenciadas basicamente pelo seu poder aquisitivo, apresentam um certo grau de endocruzamentos (por meio de relacionamentos preferenciais dentro da mesma classe social), caracterizado por diferenças genéticas sutis.

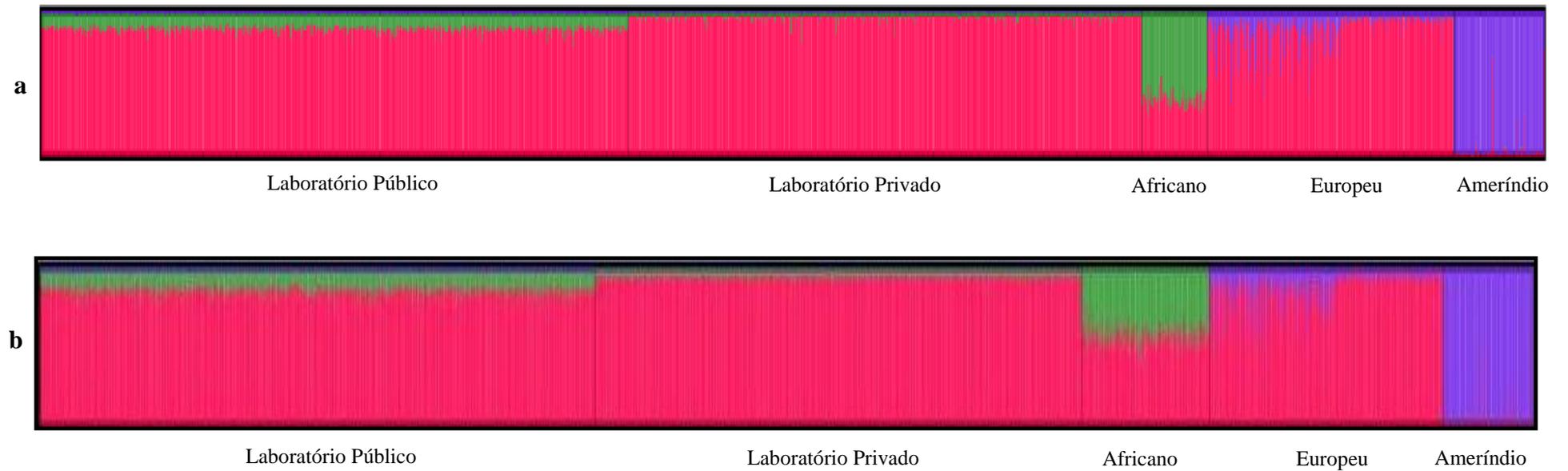


FIGURA 4.2 - Constituição genética dos indivíduos que compuseram as duas amostras do Distrito Federal e cada uma das populações parentais, obtidas do programa DISTRUCT, a partir dos dados gerados pelo STRUCTURE. Cada indivíduo genotipado é representado por uma linha vertical no gráfico, com seu percentual de composição genética, baseado nos agrupamentos realizados pelo programa STRUCTURE, assim distribuídos: Grupo 1 – rosa; Grupo 2 – verde; e Grupo3 – azul. Na Figura 4.2a a parental africana utilizada foi a do CEPH (2010), com 103 genótipos africanos; na Figura 4.2b a parental africana utilizada foi a de Budowle & Moretti (1999), a partir de 211 genótipos afro-americanos.

4.3 Alelos e genótipos característicos de populações

Ao se efetuar o conjunto de análises pelo programa STRUCTURE, foi possível obter as frequências alélicas estimadas para cada um dos agrupamentos (Anexo B). O volume de dados ausentes, isso é, a porcentagem de indivíduos que não apresenta genotipagem para um dado marcador, variou de 2% para o TPOX a 23,6% para o D16S539.

A partir das frequências alélicas obtidas para as populações parentais (Anexo B) foram estimados valores de deltas (δ - diferença de frequências alélicas) para alelos que apresentem um diferencial de frequência de no mínimo 10% entre ao menos dois agrupamentos. Os resultados obtidos constam da Tabela 4.10, onde estão sugeridos alelos informativos de ancestralidade populacional para estes doze marcadores STR autossômicos.

Foram buscados, dentre os 2460 genótipos em estudo, aqueles que apresentavam simultaneamente os alelos aqui levantados como os mais frequentes para o grupo dos europeus - D13S317*8, D13S317*11, TPOX*8, D3S1358*18 e FGA*21. Nove indivíduos apresentaram esse conjunto, sendo que cinco são do grupo parental europeu (quatro espanhóis e um português) e os demais são da amostra de laboratório privado (com alto percentual europeu). Portanto, a presença em conjunto desses cinco alelos num determinado indivíduo pode ser considerada indicativa de um genótipo de origem europeia.

De forma similar, foi possível selecionar na Tabela 4.10 os alelos mais frequentes no grupo dos ameríndios - D16S539*10, D7S820*11, D13S317*9, D5S818*7, D5S818*9, D5S818*11, CSF1PO*12, TPOX*12, TH01*6, vWA*16, D21S11*32.2 e D3S1358*15. Dentre os 2460 indivíduos, apenas dois deles apresentaram todos estes alelos em conjunto, com exceção ao marcador D5S818 (em que os indivíduos apresentam um dos alelos), sendo que ambos eram ameríndios.

De maneira complementar, a exclusão dos indivíduos que portavam ao menos um dos alelos menos frequentes no grupo de africanos - D16S539*12, D7S820*12, D5S818*11, CSF1PO*11, TPOX*11, TH01*6, TH01*9.3, vWA*17, D21S11*30 e D8S1179*13 resultou em sete indivíduos que atenderam os critérios.

Destes, três foram da parental africana (afro-americanos), três oriundos das amostras miscigenadas (laboratórios público e privado) e um dentre os europeus (portugueses). Há que lembrar que a amostra afro-americana apresentou mistura com genótipos europeus, como demonstrado pelos programas STRUCTURE e DISTRICT (Figuras 4.1 e 4.2).

TABELA 4.10 - Diferencial de frequência alélica (δ) entre os agrupamentos (presumidas populações parentais) para os alelos cujo δ foi superior a 10% entre ao menos dois grupos e proposta de população-característica para cada alelo.

Marcador	Alelo	Afr - Eur	Afr - Amer	Eur - Amer	Pop-característica
D16S539	10	0,045	0,179	0,224	Ameríndios
	12	0,201	0,163	0,038	Africanos
	13	0,024	0,158	0,134	Ameríndios
D7S820	8	0,040	0,195	0,155	Ameríndios
	11	0,079	0,242	0,163	Ameríndios
	12	0,129	0,159	0,030	Africanos
D13S317	8	0,104	0,002	0,106	Europeus
	9	0,078	0,346	0,268	Ameríndios
	11	0,214	0,010	0,224	Europeus
	12	0,400	0,535	0,135	Afr + Amer
D5S818	7	0,000	0,248	0,248	Ameríndios
	8	0,136	0,134	0,002	Africanos
	9	0,013	0,133	0,120	Ameríndios
	11	0,208	0,377	0,169	Afr + Amer
	12	0,033	0,292	0,325	Ameríndios
	13	0,140	0,280	0,140	Afr + Amer
CSF1PO	8	0,200	0,200	0,000	Africanos
	11	0,259	0,202	0,057	Africanos
	12	0,063	0,211	0,148	Ameríndios
TPOX	6	0,162	0,161	0,001	Africanos
	8	0,333	0,184	0,149	Eur + Afr
	9	0,213	0,313	0,100	Afr + Amer
	11	0,111	0,163	0,052	Africanos
	12	0,031	0,270	0,239	Ameríndios
TH01	6	0,196	0,306	0,110	Afr + Amer
	7	0,421	0,150	0,271	Eur + Afr
	8	0,111	0,245	0,134	Afr + Amer
	9.3	0,250	0,170	0,080	Africanos
vWA	15	0,292	0,339	0,047	Africanos
	16	0,004	0,244	0,248	Ameríndios
	17	0,186	0,203	0,017	Africanos
D21S11	28	0,183	0,315	0,132	Afr + Amer
	30	0,233	0,208	0,025	Africanos
	32.2	0,059	0,209	0,150	Ameríndios
D3S1358	15	0,050	0,239	0,189	Ameríndios
	18	0,113	0,013	0,126	Europeus
FGA	21	0,111	0,049	0,160	Europeus
	22	0,044	0,210	0,166	Ameríndios
	24	0,115	0,026	0,141	Europeus
	25	0,036	0,150	0,114	Ameríndios
D8S1179	13	0,200	0,249	0,049	Africanos
	14	0,124	0,172	0,048	Africanos
	15	0,194	0,111	0,083	Africanos

(Afr – africano; Eur – europeu; Amer – ameríndio)

Com relação aos genótipos, foram selecionados aqueles que apresentaram uma frequência de ocorrência superior a 80% ou igual a 0%, mas apenas para os genótipos que apresentaram uma ocorrência de mais de 2% do total de indivíduos analisados, em um dado grupo parental. A Tabela 4.11 mostra os genótipos selecionados e sua frequência relativa. Esses genótipos também podem ser considerados indicativos de ancestralidade. Observa-se que a parental ameríndia é a que apresenta maior riqueza de genótipos com frequências dentro dos parâmetros estabelecidos (84,5%) e a menor é a africana. Em ameríndios, 24 dos 27 genótipos com frequências dentro dos parâmetros estabelecidos referem-se a ausência do genótipo em questão.

Foi realizada uma busca de indivíduos no banco de 2460 genótipos que apresentassem os genótipos com frequência superior a 80% no grupo dos ameríndios (genótipos 9/12 para D13S317 e 7/11 ou 9/11 para D5S818) e os genótipos ausentes nesta parental (listados na Tabela 4.11). Sete indivíduos se ajustaram aos critérios, sendo quatro deles do grupo de ameríndios, um de ibéricos e dois da amostra do laboratório público. Considerando apenas as amostras de parentais, 80% (oitenta por cento) dos indivíduos selecionados se incluíram no perfil genotípico ameríndio.

É fato que na população brasileira a composição étnica de um indivíduo não necessariamente se reflete em seu fenótipo (Parra *et al.*, 2003), de modo que sua aplicação em identificação física é indeterminada. Entretanto, seu uso em programas de bancos de dados possibilita associar um indivíduo a uma determinada população regional ou local (Burritt, 2006; EHSTRAFD, 2009).

TABELA 4.11 - Genótipos que apresentaram frequência de ocorrência superior a 80%, considerando o número total de ocorrências do genótipo em questão, ou igual a 0%* em um dado grupo parental. Os valores entre parênteses referem-se às porcentagens em que estes genótipos foram observados considerando o número total de ocorrência do genótipo em questão. N = número de genótipos; % = percentual de genótipos no total.

Marcador	Europeus	Africanos	Ameríndios
D16S539	10/13 (89)		
D7S820			8/8 (0), 8/9 (0)
D13S317			9/12 (87)
D5S818			7/11 (81), 9/11 (86), 10/12 (0), 12/12 (0),12/13 (0)
TPOX		6/8 (90)	8/9 (0), 8/10 (0), 9/11 (0)
TH01		7/8 (80), 8/8 (82)	6/8 (0), 9/9.3 (0)
vWA			14/15 (0), 16/19 (0)
D21S11		28/31 (89)	28/29 (0), 28/31.2 (0)
D3S1358			14/15 (0), 14/16 (0), 14/17 (0),15/18 (0), 16/18 (0)
FGA			20/21 (0), 21/23 (0), 22/22 (0), 22/23 (0)
D8S1179			11/14 (0)
N	1	4	27
%	3,1	12,5	84,5

* somente para os casos em que o genótipo ocorreu em mais de 2% do total de indivíduos analisados.

4.3.2 Características populacionais associadas à origem geográfica

A partir do levantamento de alelos e genótipos (Tabelas 4.10 e 4.11), foi verificada a probabilidade de inferência de origem populacional para cada amostra (parentais e laboratórios). Cabe lembrar que um dos parâmetros utilizados para a seleção desses marcadores STR para o sistema CODIS foi a similaridade da distribuição de frequências entre as populações, isso é, baixos valores de F_{ST} , definição oposta à dos AIMS (Marcadores Informativos de Ancestralidade). Neste aspecto, as análises para associação de origem geográfica são tentativas de seu uso no estabelecimento de informações extras sobre os indivíduos, em especial em populações miscigenadas, não previstas inicialmente.

O conjunto de alelos e genótipos mais frequentes para o grupo parental europeu (Tabelas 4.10 e 4.11), foi submetido à planilha OmniPop (Burritt, 2006). O perfil mais frequente apresenta os seguintes alelos e genótipos: D13S317 - genótipo 8/11, TPOX - alelo 8, D3S1358 - genótipo 17/18, FGA - genótipo 21/23, D16S539 - genótipo 10/13, D7S820 - genótipo 8/12, TH01 - genótipo 9/9.3, vWA - genótipo 16/19 e D8S1179 - genótipo 11/13. Dentre as mais de 200 populações comparadas, a maior probabilidade de se encontrar esse perfil foi entre: indianos (Rajput), brasileiros não-negros, norte-americanos (Michigan), iraquianos (curdos), bósnios, albanianos e árabes (Marrocos), com índices variando entre $7,38 \times 10^{-10}$ e $1,74 \times 10^{-11}$. As com menor probabilidade de ocorrência foram aborígenes canadenses, nativos do Alaska, navajos, apaches e peruanos, com frequências de ocorrência entre $1,17 \times 10^{-14}$ e $1,29 \times 10^{-16}$. Com exceção da amostra de indianos, fica evidente que o perfil é comum às populações europeias (e delas derivadas), e distante das amostras americanas nativas (tanto da América do Norte quanto do Sul).

O mesmo perfil genético quando analisado pela ferramenta *MPGO* (*Most Probable Geographical Origin*, ou “Origem Geográfica Mais Provável”) resultou em um índice de 100% para a Bósnia (frequência de ocorrência de $2,28 \times 10^{-13}$) e 90,52% para italianos, com outras cinco populações apresentando valores de semelhança entre 61 e 71%: espanhóis, malaios, albanianos e norte-americanos de Michigan. Ressalte-se que os valores percentuais deste programa identificam o índice de semelhança entre as amostras populacionais disponíveis em seu banco de dados, em que a mais próxima sempre é identificada pelo programa como tendo um

índice de 100%. Constatou-se que os resultados gerados por ambas ferramentas (OmniPop e MPGO/EHSTRAFD) foram similares.

O mesmo procedimento foi reproduzido para o grupo parental africano, a partir dos dados obtidos das Tabelas 4.10 e 4.11. Os genótipos e alelos do perfil considerado africano foram D13S317 - 10/12, D5S818 - 8/13, CSF1PO*8, TPOX - 6/8, TH01 - 7/8, vWA*15, D21S11 - 28/31, FGA - 22/24 e D8S1179 - 12/15. Populações com maior probabilidade de ocorrência foram, na maioria, afro-americanas as quais perfizeram 18 das 24 com maior índice, entremeadas por populações de Moçambique, Guiné-Bissau, Somália, marroquinos e espanhóis (Andaluzia), com valores entre $2,36 \times 10^{-09}$ e $1,71 \times 10^{-11}$. As mais distantes em termos de índice de frequência foram: apaches, navajos, nativos norte-americanos, vietnamitas, asiáticos, caucasianos norte-americanos, japoneses, hispânicos, escoceses e chineses, com valores de frequência de ocorrência entre $2,74 \times 10^{-14}$ e $1,88 \times 10^{-17}$. Mostrou-se evidente a relação do perfil comum ao grupo parental africano e as amostras da mesma região, com valores bastante distantes notadamente das populações asiáticas e norte-americanas. Com a ferramenta *MPGO* obteve-se um resultado de 100% de índice de proximidade para uma amostra da Guiné Equatorial (com frequência de ocorrência de $1,36 \times 10^{-13}$), seguido por Angola (75,6% e frequência de ocorrência de $1,03 \times 10^{-13}$) e outras amostras populacionais africanas e afro-americanas (com índices entre 26 e 49%), corroborando a planilha OmniPop.

O mesmo processo foi realizado para o grupo ameríndio, a partir dos alelos e genótipos selecionados (Tabelas 4.10 e 4.11), sendo o perfil: D16S539 - 9/9, D7S820*11, D13S317 - 9/12, D5S818 - 7/11, CSF1PO - 12/12, TPOX - 11/12, TH01 - 6/6, vWA - 16/17, D21S11 - 29/31.2, D3S1358 - 15/15 e FGA - 24/25. As amostras populacionais mais correlatas ao perfil foram: peruanos, apaches, equatorianos, navajos, nativos do Alaska, indianos (Karan), paraguaios e mexicanos, com índices variando entre $6,85 \times 10^{-09}$ e $2,42 \times 10^{-11}$. Dentre os com menor afinidade de perfil, destacam-se poloneses, bielorrussos, tunisianos, afro-americanos (Estados Unidos e Canadá), espanhóis, italianos e noruegueses, com valores de frequência de ocorrência entre $1,80 \times 10^{-14}$ e $1,15 \times 10^{-15}$. Evidente a distância entre este grupo parental e as populações europeias e africanas, com proximidade às amostras americanas (exceto novamente uma referência da Índia). Com a ferramenta *MPGO*

do site EHSTRAFD, os resultados foram semelhantes, obtendo-se um índice de 100% para uma amostra de Apache (frequência de ocorrência de $1,29 \times 10^{-12}$), seguida dos peruanos (67,6% e frequência de $8,70 \times 10^{-13}$), navajos e guatemaltecos – todas populações americanas nativas.

A planilha OmniPop foi então usada para as amostras dos laboratórios público e privado, com base nos alelos que apresentaram maior diferença entre elas. Para o laboratório público foi utilizado o perfil: D7S820*10, D13S317*12 e D13S317*13, TH01*7 e TH01*8, vWA*16, D21S11*28, FGA*23 e D8S1179*14. As 25 referências com maior índice de ocorrência foram amostras de Moçambique, Guiné-Bissau, afro-americanos (da Jamaica, Bahamas, Canadá e diversos estados norte-americanos) e marroquinos, com valores entre $5,67 \times 10^{-02}$ e $1,15 \times 10^{-04}$. Dentre as 27 populações mais distantes considerando o perfil, estavam amostras de nativos do Alasca, populações asiáticas (chineses, vietnamitas, indianos, coreanos, japoneses e indianos) e nativos da América do Norte, com índices variando entre $1,91 \times 10^{-05}$ e $2,16 \times 10^{-07}$. Utilizando a ferramenta *MPGO* do site EHSTRAFD, essa mesma combinação de alelos produziu um resultado também caracteristicamente africano: as quatro populações mais próximas (frequências de ocorrência entre $1,43 \times 10^{-05}$ e $1,32 \times 10^{-05}$) foram duas de Moçambique e duas da Guiné Equatorial, seguidas por outras 16 amostras de africanos ou afro-americanos, entremeadas por uma população da Colômbia. Apesar dos índices com valores de frequência de ocorrência menores (decorrentes de um perfil menos completo), fica clara a correlação desta amostra com populações de origem africana, visto que as 25 mais extremas (dentre as mais de 200) separam claramente os africanos dos asiáticos e americanos.

Para a análise da amostra do laboratório privado, foi submetido à planilha OmniPop o seguinte perfil: D16S539*13, D7S820*11, D13S317*8, TH01*9 e TH01*9.3, D21S11*29 e D21S11*32.2, D3S1358*17, FGA*21 e D8S1179*11 e D8S1179*13. Dentre as 39 referências com maior similaridade (com índices variando de $1,26 \times 10^{-06}$ a $3,53 \times 10^{-07}$), 33 foram europeus (hispânicos, italianos e belgas) e asiáticos (indianos, chineses, vietnamitas e japoneses), com apenas seis amostras americanas (argentinos, brasileiros e norte-americanos miscigenados). Das 28 referências com índices mais distantes ao do perfil do laboratório privado, 25 foram de populações americanas (apaches, navajos, nativos norte-americanos e

brasileiros) ou afro-americanas (Jamaica, Bahamas e diversos estados norte-americanos), sendo apenas três europeias (duas hispânicas e uma norueguesa). Pelo banco de dados da EHSTRAFD, o uso do MPGO identificou como mais próximas do perfil analisado populações da Lituânia ($2,07 \times 10^{-08}$) e da Espanha ($1,94 \times 10^{-08}$), seguidas de amostras populacionais de europeus e asiáticos.

Apesar das duas amostras em questão (laboratório público e particular) apresentarem, de modo geral, semelhanças a ponto de poderem ser consideradas oriundas de uma mesma população, as diferenças existentes (destacadas pelos alelos apresentados na Tabela 4.4) deixam claro o motivo da divergência. Os índices de maior semelhança entre a amostra de instituição pública são com as amostras de populações de origem africana, enquanto no caso do laboratório privado, a maior similaridade é com as populações europeias e asiáticas. De modo similar, as maiores diferenças são respectivamente entre os asiáticos (com relação ao laboratório público) e os africanos (comparado ao laboratório particular). Ambas apresentam diferenças em relação às amostras de populações americanas.

4.3.3 Poder de resolução de locos e alelos nas amostras analisadas

Duas análises foram estabelecidas com relação ao poder de informação de origem populacional dos genótipos parentais e amostras dos laboratórios analisadas neste trabalho. Um se referiu ao potencial de identificação de origem a partir dos genótipos utilizados como populações parentais, de acordo com os alelos e genótipos mais ou menos frequentes em cada uma. O outro tratou dos alelos e genótipos mais díspares entre as amostras laboratoriais estudadas, estabelecendo características identificadoras entre as duas.

No caso dos alelos e genótipos típicos das populações parentais utilizadas, os resultados podem ser confrontados com os estudos de Rosenberg e colaboradores (2003), visando corroborar ou acrescentar potenciais Alelos Indicativos de Ancestralidade. Em comum com o trabalho de Rosenberg (2003), quatro locos foram analisados neste trabalho: D7S820, D8S1179, D13S317 e D16S539. Todos apresentaram frequências que os habilitaram a diferenciar as populações parentais e as amostras laboratoriais.

O marcador D7S820 apresentou os alelos 8, 10, 11 e 12 como fortes indicadores de origem populacional. O alelo 8 esteve praticamente ausente no grupo

ameríndio (0,8%), e ainda apresentou diferença entre os grupos europeu (16,3%) e africano (20,3%). O alelo 10 apresentou frequências distintas nos três grupos, variando de 34,4% entre as amostras africanas subsaarianas a 22,5% nos ameríndios. O alelo 11, ao contrário, foi indicativo de origem ameríndia, ocorrendo com frequência de 41,0% neste grupo, e 16,8% entre os africanos (com valor intermediário de 24,7% entre os europeus ibéricos). O alelo 12 apresentou frequência de ocorrência bastante baixa entre os africanos (3,3%), e entre 16,2 e 19,2% nas outras. No caso dos quatro alelos com maior diferença de ocorrência entre as parentais, a frequência nos europeus foi sempre intermediária à de ameríndios e africanos, sendo que os alelos 8 e 10 foram mais comuns entre os nativos da África, enquanto 11 e 12 apareceram com mais frequência entre os americanos. Apesar do pequeno número de populações ameríndias utilizadas como parentais neste trabalho, verificou-se que os alelos inferiores ou iguais a 10 foram sempre mais frequentes entre os africanos, e do 11 em diante mais comuns entre os ameríndios. Por serem extremos da dispersão humana, talvez sejam também os extremos de uma distribuição gradual de frequências alélicas nos continentes. O trabalho de Rosenberg e colaboradores (2003) sugere o marcador D7S820 como mais eficiente na resolução de origem ameríndia do que na correlação com os outros dois continentes parentais (África e Europa).

O loco D8S1179 apresentou resultados semelhantes ao marcador D7S820 para os alelos 13 e 14. O alelo 13 foi mais frequente entre ameríndios (34,0%) e menos na África (9,1%), com valor intermediário entre os europeus (29,1%). O alelo 14 inverteu o observado, sendo mais comum entre africanos (37,7%) e ocorrendo com menor frequência nos nativos americanos (20,5%), mais uma vez ficando os europeus entre estes (25,3%). Semelhante também no fato de que os alelos iguais ou superiores a 14 serem mais frequentes entre africanos, e iguais ou inferiores a 13 ocorrerem mais entre os ameríndios (Anexo B). Neste ponto, os alelos 11 e 15 também foram eficientes na identificação de origem continental, no que tange a africanos e ameríndios. Entretanto, a ocorrência dos alelos do marcador D8S1179 entre os europeus não obedeceu o mesmo padrão observado para o marcador citado anteriormente, uma vez que não apresentavam frequência intermediária entre as populações africana e americana. Os alelos 11 e 15 foram flagrantes discrepâncias neste aspecto, apresentando frequências de ocorrência superior

(8,6% do alelo 11) e inferior (11,8% do alelo 15) às das parentais da África e América. Coerente com o indicado por Rosenberg (2003), este marcador parece ser eficiente na separação entre africanos e europeus, não sendo tão eficiente na indicação de origem ameríndia. De fato, em todos os alelos com maior frequência (do 10 ao 16), os percentuais de ocorrência em europeus foram sempre mais próximos dos ameríndios que dos africanos.

O marcador D13S317 apresenta padrão de distribuição de frequência alélica bastante distinto dos anteriores, com alelos indicativos de ancestralidade nas três populações parentais. Esta característica corrobora as conclusões de Rosenberg e colaboradores (2003) sobre o potencial de resolução deste marcador em termos de identificação de origem populacional. Pela análise de seu trabalho, o loco D13S317 só não seria muito eficiente na diferenciação entre europeus e africanos. No estudo aqui realizado, verificou-se alelos típicos para cada parental. O alelo 8 ocorreu com 11,4% de frequência em europeus, e apenas 1,0% em africanos e 0,8% em ameríndios, diferenciando aquela população das demais. O alelo 9 apresentou frequência de ocorrência de 35,6% em ameríndios, muito superior aos 8,8% em europeus e 1,0% em africanos – trata-se de um alelo bem típico de nativos americanos. O alelo 10 apresentou distribuição mais uniforme entre as três parentais, sendo um pouco mais presente em ameríndios (13,1%) do que em africanos (3,4%) e europeus (5,6%). O alelo 11 também se mostrou bem mais frequente em uma das três parentais, no caso os europeus, com 31,0% (e 9,6% em africanos e 8,6% em ameríndios). O alelo 12 foi talvez o com maior poder de resolução para as três populações parentais, com 66,4% de ocorrência em africanos e 12,9% em ameríndios, ficando os europeus com 26,4%. O alelo 13 destacou os americanos dos demais, com 21,4% de ocorrência, maior que os 11,6% dos europeus e 14,8% de africanos. Dentre os alelos citados (os demais apresentaram frequências baixas e semelhantes – Anexo B), os europeus apresentaram menor diferença de frequência em relação aos africanos que aos ameríndios, com exceção do alelo 12. Tal resultado corrobora mais uma vez o trabalho de Rosenberg e colaboradores (2003), que conclui que o marcador D13S317 é menos eficiente na diferenciação entre africanos e europeus.

O loco D16S539 foi descrito como sendo possivelmente o de menor poder de resolução para estas três parentais analisadas (Rosenberg *et al.*, 2003). Os

resultados aqui obtidos mostraram que apenas os alelos 10, 12 e 13 apresentam frequências capazes de auxiliar na indicação da origem populacional. O alelo 10 apresentou frequência de ocorrência de 31,2% entre os ameríndios, em contraste com os 8,8% de europeus e 13,3% nos africanos. O alelo 12 foi pouco frequente entre os africanos (7,2%) e relativamente comum entre europeus (27,3%) e ameríndios (23,5%). Já o alelo 13 pouco ocorreu entre os nativos americanos (0,8%), sendo mais frequente em europeus (14,2%) e africanos (16,6%). Trata-se de um marcador com pouca resolução para estas três populações parentais, quando comparado aos outros três locos (D7S820, D8S1179 e D13S317).

Dentre os marcadores utilizados neste trabalho e que não foram analisados por Rosenberg e colaboradores (2003), oito apresentam frequências que podem auxiliar na indicação de origem populacional.

O loco CSF1PO apresentou três alelos com frequências possíveis de serem usadas na indicação de origem continental. O alelo 8, de baixa frequência entre europeus e ameríndios (0,3%), ocorre com frequência de 20,3% nos africanos. De modo oposto, o alelo 11 é menos frequente exatamente entre os africanos (5,9%), distante das frequências de ocorrência em europeus (31,8%) e ameríndios (26,1%). Foram dois alelos com possibilidade de diferenciar indivíduos de origem africana, dentre as três populações parentais. O alelo 12, por sua vez, foi o menos informativo dos três, sendo comum nos três grupos parentais (47,4% em ameríndios, 32,6% em europeus e 26,3% em africanos).

No caso do marcador TPOX, houve grande variação de frequência de alelos entre os três grupos parentais. Os alelos 6 e 9 apareceram com frequência alta em africanos (respectivamente 16,6 e 31,6%), se comparados a europeus (0,4% e 10,3%) e ameríndios (0,5% e 0,3%). O grupo parental africano também pode ser separado dos demais pela baixa frequência de ocorrência dos alelos 8 (15,8%) e 11 (17,8%), quando comparado aos europeus (respectivamente 49,1% e 28,9%) e aos americanos nativos (34,2% e 34,1%). Tratou-se de um loco com bom poder de identificação de origem africana. Ainda, conseguiu destacar os europeus com a frequência mais alta no alelo 8 (49,1% contra 15,8% de africanos e 34,2% de ameríndios) e os ameríndios com sua típica frequência alta no alelo 12 (28,4%, bastante superior aos 4,5% de europeus e 1,4% nos africanos).

O marcador STR TH01 mostrou também bom poder de discriminação entre os três grupos parentais, principalmente na indicação de origem africana. Os alelos 6 e 9.3 foram relativamente raros entre os africanos (respectivamente 1,7% e 1,0%), quando comparados aos europeus (21,3% e 26,0%) e aos ameríndios (32,2% e 18,0%). Os alelos 7 e 8 também possibilitam distinguir a origem africana dos demais, mas pela sua frequência superior (62,9% e 24,9%) em relação aos demais grupos parentais (20,8% e 13,8% nos europeus e 47,9% e 0,4% nos ameríndios). Pode-se destacar, fora do grupo africano, a baixa frequência relativa do alelo 7 em europeus (20,8%) e do alelo 8 em ameríndios (raros 0,8%).

O loco D5S818 não possibilitou separar os europeus dos outros dois parentais, mas apresentou alelos característicos para ameríndios e africanos. O alelo 7, raro entre europeus e africanos (0,9% para ambos), apareceu com frequência de 25,7% entre os ameríndios. A situação é semelhante quanto ao alelo 8, mas com frequência superior observada entre os africanos (14,1%), comparativamente aos europeus (0,5%) e ameríndios (0,7%). Em menor grau, o alelo 9 repetiu o de número 7, com frequência de 14,7% em americanos nativos, bastante superior aos europeus (2,7%) e africanos (1,4%). O alelo 11 ocorreu em mais da metade dos ameríndios (51,2%), sendo bem mais baixo nos africanos (13,5%) e um intermediário entre os europeus (34,3%). O alelo 12 também foi eficiente para diferenciar os ameríndios, por ser menos frequente (4,6%) que nos europeus (37,1%) e africanos (33,8%). O alelo 13 permitiu distinguir bem os ameríndios (1,7%) dos africanos (29,7%), com os europeus apresentando frequência intermediária (15,7%). Trata-se de um loco eficiente na diferenciação entre americanos nativos e africanos, não tendo o mesmo poder para distinguir indivíduos de origem europeia.

O marcador vWA, apesar de seus 11 alelos, não apresentou bom poder de resolução para estes três grupos parentais, tendo apenas três alelos com indicação de capacidade de diferenciação. O alelo 15 apresentou frequência mais alta entre os africanos (39,7%), destacando-se dos europeus (10,5%) e ameríndios (5,8%). O alelo 16 apareceu em metade dos perfis ameríndios (50,7%) e em aproximadamente $\frac{1}{4}$ dos europeus (25,9%) e africanos (26,3%). Já o alelo 17 se destacou pela baixa frequência entre africanos (9,1%), quando comparado aos europeus (27,7%) e nativos americanos (29,4%). Não apresentou nenhuma frequência típica de populações ibéricas.

O loco D21S11 apresentou três alelos (dentre os vinte e cinco observados) capazes de auxiliar na indicação de origem parental entre os três grupos estudados. O alelo 28 foi raro em ameríndios (0,5%), enquanto apresentou frequência alta em africanos (32,0%) e intermediária em europeus (13,7%). O alelo 30 foi comum entre europeus (27,4%) e ameríndios (24,9%), mas raro entre os africanos (4,1%). O alelo 32.2 apresentou frequência significativamente maior em ameríndios (24,4%) do que nos europeus (9,4%) e africanos (3,5%). Apesar de seu limitado poder de resolução para os três grupos parentais, este marcador demonstrou possibilidade de diferenciar entre ameríndios e africanos.

O marcador FGA, a despeito de seus vinte e seis alelos, apresentou apenas quatro com possibilidade de indicar a origem continental para os três parentais estudados, sendo mais eficiente para diferenciar entre europeus e ameríndios. O alelo 21 foi bem mais frequente em europeus (16,8%) do que em africanos (5,7%) e ameríndios (0,8%). O alelo 22 foi raro em ameríndios (1,7%), em contraste com sua frequência em europeus (18,3%) e africanos (22,7%). O alelo 24 apresentou cerca da metade da frequência de ocorrência em europeus (12,4%) do que entre africanos (23,9%) e ameríndios (26,5%). E o alelo 25 apresentou frequência mais alta em nativos americanos (22,2%) do que nos europeus (10,8%) e africanos (7,2%).

Por fim, o loco D3S1358 apresentou apenas dois alelos com capacidade de indicação de origem populacional entre os três parentais. O alelo 15 foi mais frequente em ameríndios (49,2%) do que nos outros (30,3% em europeus e 25,3% em africanos), e o alelo 18 se destacou por sua frequência superior em europeus (12,9%), quando comparado a africanos (1,6%) e ameríndios (0,3%).

Foi possível, portanto, destacar alguns marcadores STR quanto à sua capacidade de indicação de origem parental, quando se trata dos três grupos que basicamente compõem a população brasileira. De acordo com o objetivo a que se propõe no estabelecimento de contribuição genética e/ou origem populacional, os marcadores puderam assim ser divididos.

Resolução entre europeus e africanos: D16S539*12, CSF1PO*8 e *11, D7S820*12, TPOX*6, *8, *9 e *11, D13S317*8, *11 e *12, TH01*6, *7, *8 e *9.3, D5S818*8, *11 e *13, vWA*15 e *17, D21S11*28 e *30, FGA*21 e *24, D3S1358*18 e D8S1179*13, *14 e *15.

Resolução entre europeus e ameríndios: D16S539*10 e *13, CSF1PO*12, D7S820*8 e *11, TPOX*8, *9 e *12, D13S317*8, *9, *11 e *12, TH01*6, *7 e *8, D5S818*7, *9, *11, *12 e *13, vWA*16, D21S11*28 e *32.2, FGA*21, *22, *24 e *25 e D3S1358*15 e *18. O loco D8S1179 não apresentou alelos adequados para separar origem europeia de ameríndia.

Resolução entre africanos e ameríndios: D16S539*10, *12 e *13, CSF1PO*8, *11 e *12, D7S820*8, *11 e *12, TPOX*6, *8, *9, *11 e *12, D13S317*9 e *12, TH01*6, *7, *8 e *9.3, D5S818*7, *8, *9, *11, *12 e *13, vWA*15, *16 e *17, D21S11*28, *30 e *32.2, FGA*22 e *25, D3S1358*15 e D8S1179*13, *14 e *15.

Foram observados 28 alelos, considerando os 12 locos aqui analisados, capazes de diferenciar entre europeus e africanos, 29 alelos que distinguiram entre europeus e ameríndios (em 11 locos), e 38 alelos capazes de auxiliar na indicação de origem entre africanos e ameríndios. O maior número de alelos com poder de diferenciação entre ameríndios e africanos pode ser explicado pela história da dispersão humana no planeta, estando os europeus com frequências (de um modo geral) intermediárias às populações dos continentes africano e sul-americano. A isto se soma o fato de que os STRs do sistema CODIS foram selecionados por pesquisadores e estudiosos europeus e norte-americanos, a partir de análises com bancos de dados genéticos disponíveis, cuja grande maioria era de populações do hemisfério norte. Face aos objetivos propostos quando da seleção destes locos, é de se esperar que sua eficiência para separação de populações daquela região seja menor que de africanos e ameríndios.

Dentre os doze locos usados nas avaliações de composição étnica das amostras laboratoriais, cabe destacar a eficiência dos cinco seguintes: TPOX (com cinco alelos identificadores), D13S317 (com quatro alelos), TH01 (com quatro alelos indicadores de origem), D5S818 (com seis alelos) e FGA (com quatro alelos com poder de resolução). Destes, apenas o D13S317 foi utilizado no trabalho de Rosenberg e colaboradores (2003), mas os outros quatro demonstram ter tanto poder de resolução quanto aquele.

Estes cinco marcadores podem ser recomendados para análises de mistura populacional, quando se trata de diferenciação entre europeus (ibéricos), africanos (subsaarianos) e ameríndios (sul-americanos).

5. CONSIDERAÇÕES FINAIS

A detecção de diferenças entre duas amostras populacionais do Distrito Federal coletadas por laboratórios diferentes (representantes de clientela distintas), mostrou que o padrão de coleta pode produzir discrepâncias no resultado final das análises. Tal confirmação sugere que a metodologia de coleta de amostras populacionais seja repensada, buscando adotar parâmetros ainda mais isentos que os atuais, considerados inicialmente como adequados.

As duas amostras aqui consideradas foram analisadas quanto à presença de diferenciação e sua composição étnica com relação aos três principais grupos parentais da população brasileira: ameríndios, africanos e europeus. Tal análise baseou-se nos genótipos de doze marcadores autossômicos teoricamente neutros (não ligados à expressão gênica e pressão evolutiva).

Observou-se que, a despeito das duas amostragens poderem ser consideradas estatisticamente como oriundas de uma mesma população, ficou patente a diferença entre elas no que diz respeito à composição genética, especificamente em relação ao percentual de europeus e africanos. Foi estimado que na amostra obtida pela coleta em laboratório público houve maior participação de contribuição africana que a amostra obtida em laboratório privado. Em princípio, a diferença básica entre as duas amostragens foi o estrato sócio-econômico, uma vez que o laboratório público efetua exames de forma gratuita, e o laboratório particular cobra pela realização da genotipagem.

O trabalho evidenciou a existência de uma ancestralidade distinta em amostras de uma mesma população, decorrente de uma segmentação social histórica. Esta diferenciação já era sugerida pelos dados do IBGE, mas traduzida em um contexto sócio-econômico vinculado à raça auto-declarada. A análise genotípica de indivíduos a partir das amostras estudadas retrata a mesma subestruturação de forma mais objetiva, por meio de análises estatísticas que confirmam a maior participação de contribuição genética europeia nas classes com maior poder aquisitivo, contrastando com um maior percentual africano entre os indivíduos de uma classe economicamente menos favorecida.

Demonstrou-se, ainda, que é possível a inferência de agrupamento parental a partir do genótipo de um indivíduo. Estes resultados mostram, a despeito dos objetivos iniciais quando da seleção dos locos que compõem o sistema CODIS, alelos pertencentes a estes locos podem também ser úteis como indicativos de ancestralidade individual. Aplicados a brasileiros cuja ascendência genética esteja incluída na composição tri-híbrida (africanos, europeus e/ou ameríndios), pode-se realizar uma estimativa consistente sobre sua composição genética (associada à origem geográfica parental), corroborada pelos resultados comparativos com outras amostras populacionais.

Ficou sugerida a potencialidade dos STRs estudados nesse trabalho como indicadores de ancestralidade para indivíduos e amostras populacionais, com destaque para os locos TPOX, D5S818, TH01, D13S317 e FGA.

Os resultados obtidos neste trabalho permitem ainda vislumbrar possibilidades de análises complementares, buscando informações genéticas associadas à presença ou ausência de outros fatores sócio-econômicos. Estudos com bancos de sangue ou doadores de medula, por exemplo, poderão evidenciar novos resultados ou corroborar os até aqui obtidos.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- 23andMe, em <https://www.23andme.com/store/>, acessado em fevereiro de 2010.
- Abe-Sandes K, Silva-Jr W & Zago MA (2004) Heterogeneity of the Y chromosome in Afro-Brazilian populations. *Hum. Biol.*, 76(1): 77-86.
- Affymetrix (2009) Affymetrix Genome-Wide Human SNP Array 6.0, em http://www.affymetrix.com/support/technical/datasheets/genomewide_snp6_datash eet.pdf (acessado em fevereiro de 2010).
- Alencastro, LF (2000) O Trato dos Viventes – Formação do Brasil no Atlântico Sul – Séculos XVI e XVII. *Companhia das Letras*, São Paulo.
- Alves-Silva J, Santos MS, Guimarães PEM, Ferreira ACS, Bandelt H-J, Pena SDJ & Prado VF (2000) The Ancestry of Brazilian mtDNA Lineages. *Am. J. Hum. Genet.* 67: 444-461.
- American Association of Blood Banks – AABB (2004) Annual Report Summary for Testing in 2003.
- Andrade CY & Dachs JNW (2007) Acesso à Educação por Faixas Etárias Segundo Renda e Raça. *Cadernos de Pesquisa*, 37(131): 399-422.
- Arcanjo AC, Dalton GC, Godinho NMO, Gontijo CC, Barcelos R, Astolfi-Filho S, Pontes IM, Klautau-Guimarães MN & Oliveira SF (2009) DRD2 TaqI haplotypes in three urban Brazilian populations. *Forensic Science International: Genetics Supplement Series*, 2: 481-482.
- Bamshad MJ, Wooding S, Scott-Watkins W, Ostler CT, Batzer MA & Jorde LB (2003) Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.*, 72: 578–589.
- Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA & Jorde LB (2001) Genetic evidence on the origins of Indian caste populations. *Genome Res.*, 11(6): 994-1004.
- Barcelos RSS (2006) *Diversidade genética das populações do Centro-Oeste: DNA mitocondrial e cromossomo Y*. 2006. Tese de Doutorado em Ciências Biológicas (área de concentração: Genética) – Universidade de Brasília (UnB).

- Barcelos RSS, Ribeiro GGBL, Silva-Jr WA, Abe-Sandes K, Godinho NM, Marinho-Neto F, Gigonzac MAD, Klautau-Guimarães MN & Oliveira SF (2006) Male contribution in the constitution of the Brazilian Centro-Oeste population estimated by Y-chromosome binary markers. In: International Society of Forensic Genetics. (Org.). *Progress in Forensic Genetics*, 11. 11 ed. Amsterdam: Elsevier Science, 2006, v. 11, p. 228-230
- Bastos-Rodrigues L, Pimenta JR & Pena SDJ (2006) The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Annals of Human Genetics*, London, 70: 658-665.
- Bonato SL & Salzano FM (1997) A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci U.S.A.* 94(5): 1866-1871.
- Bonilla C, Parra EJ, Pfaff CL, Dios S, Marshall JA, Hamman RF, Ferrell RE, Hoggart CL, McKeigue PM & Shriver MD (2004) Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann. Hum. Genet.*, 68(2): 139-153.
- Bortolini MC, Salzano FM, Bau CHD, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Bedoya G, & Ruiz-Linares A (2002) *Ann. Hum. Genet.* 66: 255-259.
- Bortolini, MC; Salzano, FM; Zago, MA; Silva-Jr, WA; Weiner, TA (1997a) Genetic variability in two Brazilian ethnic groups: a comparison of mitochondrial and protein data. *Am. J. Phys. Anthropol.*, 103: 147-156.
- Bortolini MC, Zago MA, Salzano FM, Silva-Jr WA, Bonatto SL, da Silva MC & Weimer TA (1997b) Evolutionary and anthropological implications of mitochondrial DNA variation in African-Brazilian populations. *Hum Biol*, 69(2): 141-159.
- Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, and Cavalli-Sforza LL (1991) Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. *Proc. Natl. Acad. Sci.*, 88: 839-843.
- Britten RJ, Rowen L, Williams J & Cameron RA (2003) Majority of divergence between closely related DNA samples is due to Indels. *Proc. Natl. Acad. Sci. U. S. A.*, 100 (8): 4661-4665.
- Budowle B & Brown BL (2001) El uso del análisis de ADN en la identificación forense. *Forénsica*, 1(1): 9-22.

- Budowle B & Moretti T (1999) Genotype Profiles for Six Population Groups at the 13 CODIS Short Tandem Repeat Core Loci and Other PCR-Based Loci. *Forensic Science Communications*, 1(2) – em www.fbi.gov/hq/lab/fsc/backissu/july1999/index.htm - acessado em fevereiro de 2010.
- Budowle B, Moretti TR, Niezgoda SJ, Brown BL (1998) CODIS and PCR-based short tandem repeat loci: law enforcement tools. *Proceedings of the Second European Symposium on Human Identification*, Innsbruck, Austria. Junho 1998. Madison, WI: Promega Corporation. 1998: 73-88; <http://www.promega.com/geneticidproc/eusymp2proc/17.pdf>
- Burritt B (2006) Planilha OmniPop 200.1 (em formato Excel); <http://www.cstl.nist.gov/biotech/strbase/population/OmiPop200.1.xls>
- Butler JM (2006) Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing. *J. Forensic Sci.*, 51(2): 253-265.
- Bydlowski SP, Novak EM, Issa JS, Forti N, Giannini SD & Diament J (1996) DNA polymorphisms of apolipoprotein B and AI-CIII-AIV genes in a Brazilian population: a preliminary report. *Braz. J. Med. Biol. Res.*, 29(10): 1269-1274.
- Callegari-Jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Crossetti SG, Ferreira ME & Hutz MH (2003) Historical Genetics: Spatiotemporal Analysis of the Formation of the Brazilian Population. *Am. J. Hum. Biol.*, 15: 824-834.
- Callegari-Jacques, SM & Salzano FM (1999) Brazilian Indian/non-Indian interactions and their effects. *Ciênc Cult* 51: 166-174.
- Carvalho-Silva DR, Santos FR, Jorge-Rocha J & Pena SDJ (2001) The Phylogeography of Brazilian Y-Chromosome Lineages. *Am. J. Hum. Genet.*, 68: 281-286.
- Cavalli-Sforza LL & Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.*, (Suppl 33): 266-275.
- Cavalli-Sforza LL, Menozzi P & Piazza A. *History and Geography of Human Genes*. Princeton. Princeton University Press. 1994.
- CEPH (Centre d'Etude du Polymorphisme Humain) - <http://www.cephb.fr> , acessado em setembro de 2010.

- Chakraborty R. (1985) Gene identity in racial hybrids and estimation of admixture rates. In: *Genetic Differentiation in Human and Other Animals* (eds Neel JV, Ahuja YR), pp. 171–180. Indian Anthropological Association, Delhi.
- Chakraborty R (1986) Gene admixture in human populations: Models and predictions. *Am J Phys Anthropol.* 29: 1-5.
- Charlesworth B, Sniegowski P & Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 371: 215-220.
- Corach D, Lao O, Bobillo C, van Der Gaag K, Zuniga S, Vermeulen M, van Duijn K, Goedbloed M, Vallone PM, Parson W, de Knijff P & Kayser M (2010) Inferring Continental Ancestry of Argentineans from Autosomal, Y-Chromosomal and Mitochondrial DNA. *Ann. Hum. Genet.*, 74: 65-76.
- Cunha, MC (1998) Introdução a uma história indígena. In: Cunha, MC (ed) História dos índios no Brasil. *Companhia das Letras*, São Paulo.
- Da Silva EB, Dellalibera E, Spouza M, Silva RDS & Mauricio-da-Silva L (2003) Population genetics of eight STR loci – CSF1PO, TPOX, YH01, D16S539, D7S820, D13S317, F13B and LPL in a Brazilian population from the State of Piauí, Northeast Brazil. *Forensic Sci. Int.*, 126: 90-92.
- Da Silva Jr WA, Bortolini MC, Meyer D, Salzano FM, Elion J, Krishnamoorthy R, Schneider MP, De Guerra DC, Layrisse Z, Castellano HM, Weimer TD & Zago MA (1999) Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable *loci*. *Am J Phys Anthropol*, 109(4): 425-437.
- Dalton GC, Maia FAS, Mendes CRBO, Pak HI, Paula KAA, Trindade-Filho A & Oliveira SF (2009) Genetic profile of Federal District of Brazil based on 18 STR autosomal loci. *Forensic Science International: Genetics Supplement Series*, 2: 373-375.
- Dellalibera E, Havro ML, Souza M, Kajihara K, Mauricio-da-Silva L & dos Santos Silva R (2004) Genetic analysis of 13 STR loci in then population from the State of Pernambuco, northeast Brazil. *Forensic Sci. Int.*, 146: 57-59.
- Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES, Botstein D, Akots G, Rediker KS, Gravius T, Brown VA, Rising MB, Parker C, Powers JA, Watt DE, Kauffman ER, Bricker A, Phipps P, Muller-Kahle H, Fulton TR, Ng S, Schumm JW, Braman JC, Knowlton

- RG, Barker DF, Crooks SM, Lincoln SE, Daly MJ & Abrahamson J (1987) A genetic linkage map of the human genome. *Cell*, 51(2): 319-337.
- Dupanloup I & Bertorelle G (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol. Biol. Evol.*, 18: 672-675.
- Eaaswarkhanth M, Vasulu TS & Haque I (2008) Genetic affinity between diverse ethnoreligious communities of Tamil Hadu, India: a microsatellite study. *Hum. Biol.*, 80(6): 601-609.
- EHSTRAFD - *Earth Human STR Allele Frequencies Database*, Release 2 (2009). www.ehstrafd.org – Acessado em agosto de 2010.
- Excoffier L, Laval G & Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis, *Evol. Bioinform.* Online 1: 47–50.
- Fagundes NJR, Kanitz R, Eckert R, Valls ACS, Bogo MR, Salzano FM, Smith DG, Silva WA, Zago MA, Ribeiro-dos-Santos AK, Santos SEB, Petzl-Erler ML & Bonatto SL (2008) Mitochondrial Population Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas. *Am. J. Hum. Genet.*, 82(3): 583-592.
- Ferreira ME & Grattapaglia D (1995) *Introdução ao uso de marcadores moleculares RAPD e RFLP em análise genética*. Brasília, EMBRAPA/CENARGEN, 220p.
- Godinho NMO, Gontijo CC, Diniz MCEG, Falcão-Alencar G, DaltonGC, Amorim CEG, Barcelos RSS, Klautau-Guimarães MN & Oliveira SF (2008a) Regional patterns of genetic admixture in South America. *Forensic Science International: Genetics Supplement Series*, 1: 329-330.
- Godinho, NMO. *O impacto das migrações na constituição genética de populações latino-americanas*. 2008b. Tese de Doutorado em Ciências Biológicas (área de concentração: Genética) - Universidade de Brasília.
- Góes ACS, Silva DA, Gil ÉHF, Silva MTD, Pereira RW & Carvalho EF (2004) Allele frequencies data and statistic parameters for 16 STR loci – D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179,, TPOX, D3S1358 – in the Rio de Janeiro population, Brazil. *Forensic Sci. Int.*, 140: 131-132.
- Gontijo, CC. *Composição Genética de Duas Populações Afro-derivadas Brasileiras Inferida a partir de Marcadores Informativos de Ancestralidade*. 2008. Trabalho de dissertação de Mestrado em Ciências Biológicas - Universidade de Brasília (UnB).

- Grandy DK, Litt M, Allen L, Bunzow JR, Marchionni M, Makam H, Reed L, Magenis RE & Civelli O (1989) The human dopamine D2 receptor gene is located in chromosome 11 at q22-q23 and identifies a *Taq I* polymorphism. *Am. J. Hum. Genet.*, 45: 778-785.
- Grattapaglia D, Schmidt AB, Costa e Silva C, Stringher C, Fernandes AP & Ferreira ME (2001) Brazilian population database for the 13 STR loci of AmpF/STR® Profiler Plus™ and Cofiler™ multiplex kits. *Forensic Sci. Int.*, 118: 91-94.
- Greenberg JH, Turner-II GC & Zegura SL (1986) The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence. *Current Anthropology*, 27: 477-497.
- Handley LJJ, Manica A, Goudet J & Balloux F (2007) Going the distance: human population genetics in a clinal world. *Trends in Genetics*, 23 (9): 432-439.
- Hartl DL, Clark AG. Principles of Population Genetics, 4^a ed., Sunderland, Massachusetts: Sinauer Associates, Inc. 2006.
- Hauge XY, Grandy DK, Eubanks JH, Evans GA, Civelli O & Litt M (1991) Detection and characterization of additional DNA polymorphisms in the dopamine D2 receptor gene. *Genomics*, 10: 527-530.
- House JS, Lepkowski JM, Kinney AM, Mero RP, Kessler RC & Herzog AR (1994) The social stratification of aging and health. *J. Health Soc. Behav.*, 35(3): 213-234.
- Hutz MH, Callegari-Jacques SM, Almeida SEM, Armborst T & Salzano FM (2002) Low levels of STRP variability are not universal in American Indians. *Hum. Biol.*, 74: 781-806.
- Hutz MH, Callegari-Jacques SM, Bortolini MC & Salzano FM (1999) Variability in nDNA, mtDNA, and proteins: a test case. *In: Genomic Diversity: Applications in Human Population Genetics*. Kluwer Academic/Plenum Publishers, New York, p. 23-32.
- Instituto Brasileiro de Geografia e Estatística (IBGE) – PNAD (Pesquisa Nacional por Amostra de Domicílios) 2008 - www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2008/default.shtm - Site acessado em fevereiro de 2010.
- Instituto Brasileiro de Geografia e Estatística (IBGE) – Atlas do censo Demográfico 2000. (2000b)

- Instituto Brasileiro de Geografia e Estatística (IBGE) - Censo 1991 e 2000. – www.ibge.gov.br/home/estatística/população - Site acessado em fevereiro de 2010.
- IPEA - Instituto de Pesquisa Econômica Aplicada (2010). *Objetivos de Desenvolvimento do Milênio - Relatório Nacional de Acompanhamento*. Presidência da República, Brasília.
- Jeffreys AJ, Wilson V & Thein SL (1985) Individual-specific “fingerprints” of human DNA. *Nature*, 316: 76-79.
- Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J & Harpending HC (1997) Microsatellite diversity and the demographic history of modern humans. *Proc Natl Acad Sci USA*, 94: 3100-3103.
- Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M & Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.*, 66(5): 1580-1588.
- Khan F, Pandey AK, Borkar M, Manorma T, Talwar S, Bisen PS & Agrawal S (2008) Effect of sociocultural cleavage on genetic differentiation: a study from North India. *Hum. Biol.*, 80(3): 271-286.
- Klimentidis YA & Shriver MD (2009) Estimating Genetic Ancestry Proportions from Faces. *PLoS ONE*, 4(2):e4460: 1-8.
- Leite FPN, Menegassi FJ, Schwenberger SP, Raimann PE & Albuquerque TK (2003) STR data for 09 autosomal STR markers from Rio Grande do Sul (southern Brazil). *Forensic Sci. Int.*, 132: 223-224.
- Lins AM, Sprecher CJ, Puers C & Schumm W (1996) Multiplex sets for the amplification of polymorphic short tandem repeat loci - Silver stain and fluorescence detection. *BioTechniques*, 20: 882-889.
- Lins TC, Vieira RG, Abreu BS, Grattapaglia D & Pereira RW (2010) Genetic composition of Brazilian population samples based on a set of twenty eight ancestry informative SNPs. *Am. J. Hum. Biol.*, 22: 187-192.
- Maca-Meyer N, González AM, Larruga JM, Flores C & Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.*, 2(1): 13.

- Machado TMB (2008) *Ancestralidade em Salvador – BA*. Trabalho de dissertação de Mestrado em Biotecnologia e Medicina Investigativa. Centro de Pesquisa Gonçalo Moniz – FIOCRUZ – BA.
- Martínez H, Rodríguez-Larralde A, Izaguirre MH & Guerra DC (2007) Admixture Estimates for Caracas, Venezuela, Based on Autosomal, Y-Chromosome, and mtDNA Markers. *Human Biol.*, 79(2): 201-213.
- Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS & Devine SE (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.*, Sep;16(9): 1182-1190.
- Morton NE (1991) Parameters of the human genome. *Proc. Natl. Acad. Sci. USA*, 88: 7474–7476.
- Moskvina V, Smith M, Ivanov D, Blackwood D, StClair D, Hultman C, Toncheva D, Gill M, Corvin A, O’Dushlaine C, Morris DW, Wray NR, Sullivan P, Pato C, Pato MT, Sklar P, Purcell S, Holmans P, O’Donovan MC, Owen MJ, Kirov G & International Schizophrenia Consortium (2010) Genetic Differences between Five European Populations. *Hum. Hered.*, 70: 141-149.
- Mullis KB. (1990) The unusual origin of the polymerase chain reaction. *Sci. Am.*, 262(4): 56-61, 64-65.
- Muniz YCN, Ferreira LB, Mendes-Junior CT, Wiezel CEV & Simões AL (2008) Genomic Ancestry in urban Afro-Brazilians. *Ann. Hum. Biol.*, 35(1): 104-111.
- Nei M (1987) Molecular evolutionary genetics. *Columbia University Press*, New York.
- Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics*. 154: 931-942
- Oliveira SF, Godinho NMO, Gontijo CC, Diniz MECG, Ribeiro GGBL, Motta DS, Pedrosa MAF, Barcelos RSS, Klautau MN & Simões AL (2006) Ancestralidade Genética da População Brasileira: Marcadores Autossômicos Indicam Baixa Contribuição Africana. Anais do IX Congresso da ALAB e I Congresso Brasileiro de Antropologia Biológica. Ouro Preto – Brasil (em <http://www.icb.ufmg.br/big/alab/resumos/resumo60831125422.htm> - visitado em jun/2010).
- Oliveira SF. (1999) *Inserções Alu em populações indígenas da Amazônia brasileira*. Tese de Doutorado em Ciências Biológicas - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo (USP).

- Olson S (2002) *Mapping human history: genes, race, and our common origins*. Mariner Books, New York/NY, 292p.
- O'Rourke DH & Raff JÁ (2010) The Human Genetic History of the Americas: The Final Frontier – Review. *Current Biol.*, 20: R202-R207.
- Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE & Shriver MD (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet*, 63: 1839-1851.
- Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM & Pena SDJ (2003) Color and genomic ancestry in Brazilians. *Proc. Natl. Acad. Sci. USA* 100(1): 177-182.
- Parsian A, Fisher L, O'Malley KL & Todd RD (1991) A new *TaqI* RFLP within intron 2 of human dopamine D2 receptor gene (DRD2). *Nucleic Acids Res.*, 19: 6977.
- Pena SDJ, Carvalho-Silva DR, Alves-Silva J, Prado VF & Santos FR (2000) Retrato Molecular do Brasil. *Ciência Hoje*, 159: 16-25.
- Pena SDJ & Chabraborty R (1994) Paternity testing in the DNA era. *Trends in Genet.*, 10: 204-209.
- Pena SDJ & Jeffreys AJ (1993) Breve introdução às impressões digitais de DNA. *Rev. Bras. Genet.*, 16: 857-879.
- Pérez-Lezaun A, Calafell F, Mateu E, Comas D, Ruiz-Pacheco R, Betranpetit J (1997) Microsatellite variation and the differentiation of modern humans. *Hum Gen*, 99: 1-7.
- PHYLIP - Phylogeny Inference Package (Version 3.65). Felsenstein, J. 1989. *Cladistics* 5: 164-166 (<http://evolution.genetics.washington.edu/phylip.html>)
- Pritchard JK, Stephens M & Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155: 945-959.
- Queirós Mattoso, K (1982) *Ser escravo no Brasil*. São Paulo, Editora Brasiliense.
- Rapley, R & Harbron, S (Eds.) (2004) *Molecular Analysis and Genome Discovery*. Chichester. John Wiley & Sons Ltd.
- Raymond, M & Rousset, F (1995) An exact test for population differentiation. *Evolution*, 49: 1280-1283.
- Reis JJ & Gomes F (1996) *Liberdade por um fio. História dos quilombos no Brasil*. São Paulo; Companhia das Letras, Editora Schwarz.

- Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzer MA & Barbujani G (2002) Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res.*, 12: 602–612.
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA & Feldman MW (2002) Genetic structure of human populations. *Science*, 298 (5602), 2381-2385.
- Rosenberg NA, Li LM, Ward R & Pritchard JK (2003) Informativeness of Genetic Markers for Inference of Ancestry. *Am J Hum Genet*, 73: 1402-1422.
- Rosenberg NA (2004) *Distruct*: a program for the graphical display of population structure. *Mol. Ecol. Notes*, 4: 137-138.
- Ruiz-Linares, A; Ortíz-Barrientos, D; Figueroa, M; Mesa, N; Múnera, JG; Bedoya, G; Vélez, ID; García, LF; Pérez-Lezaun, A; Bertranpetit, J; Feldman, MW & Goldstein, DB (1999) Microsatellites provide evidence for Y chromosome diversity among the founders of the New World. *Proc. Natl. Acad. Sci. USA* 96: 6312-6317.
- Salzano FM (1998) Genetic diversity of South-American human populations at the DNA and protein levels. *J. Experim. Zool.*, 282: 157-163.
- Salzano, FM (1986) Em busca das raízes. *Ciência Hoje* 5: 48-53.
- Salzano FM & Callegari-Jacques SM (1988) South American Indians: a case study in evolution. *Research Monographs on Human Population Biology*. Number 6. Clarendon Press. Oxford.
- Sans M (2000) Admixture studies in Latin America from the 20th to the 21st century. *Hum. Biol.*, 72(1): 155-157.
- Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, Crawford MH & Mitchell RJ (1999) The Central Siberian Origin for Native American Y-Chromosomes. *Am. J. Hum. Genet.*, 64: 619-628.
- Santos FR & Tyler-Smith C (1996) Reading the human Y chromosome: the emerging DNA markers and human genetic history. *Braz. J. Genet.*, 19 (4): 665-670.
- Santos RV, Fry PH, Monteiro S, Maio MC, Rodrigues JC, Bastos-Rodrigues L & Pena SD (2009) Color, race, and genomic ancestry in Brazil: dialogues between anthropology and genetics. *Curr. Anthropol.*, 50(6): 787-819.
- Schneider, S; Roessli, D; Excoffier, L (2000) ARLEQUIN, version 2.000: a software for population genetics data analysis. Geneva, Genetics and Biometry Laboratory, *University of Geneva*.

- Serre D & Pääbo S (2004) Evidence for gradients of human genetic diversity within and among continents. *Genome Research*, 14: 1679-1685.
- Short Tandem Repeat Internet DataBase - www.cstl.nist.gov/div831/strbase - Site acessado em agosto de 2010.
- Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R & Ferrell RE (1997) Ethnic-affiliation estimation by use of population-specific DNA markers. *Am. J. Hum. Genet.*, 60: 957-964.
- Silva DA, Crouse CA, Chakraborty R, Góes ACS & Carvalho EF (2004) Statistical analyses of 14 short tandem repeat loci in Brazilian populations from Rio de Janeiro and Mato Grosso do Sul states for forensic and identity testing purposes. *Forensic Sci. Int.*, 139: 173-176.
- Silva LAF, Pimentel BJ, Azevedo DA, Silva ENP & Santos SS (2002) Allele frequencies of nine STR loci – D16S539, D7S820, D13S317, CSF1PO, TPOX, TH01, F13A01, FESFPS and vWA – in the population from Alagoas, northeastern Brazil. *Forensic Sci. Int.*, 130: 187-188.
- Silva RA (2007) Chico Rei Congo do Brasil. In: Silva, VG (Ed) Memória Afro-Brasileira. Imaginário, cotidiano e poder. *Selo Negro Edições*. São Paulo.
- Silva R & Moura-Neto RS (1998) Allelic frequency distribution for three VNTR markers - D6S132, D7S467, D17S26 - in Rio de Janeiro population, Brazil. *Forensic Sci. Int.*, Jun 8; 94(1-2): 33-38.
- Silva Jr, WA; Bonatto, SL; Holanda, AJ; Ribeiro-Dos-Santos, AK; Paixao, BM; Goldman, H; Abe-Sandes, K; Rodriguez-Delfin, L; Barbosa, M; Paco-Larson, ML; Petzl-Erlar, ML; Valente, V; Santos, SE & Zago, MA (2002) Mitochondrial genome diversity of native Americans supports a single early entry of founder populations into America. *Am J Hum Genet* 71(1): 187-92.
- Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De-Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SDJ & Santos FR (2001) Genetic Differentiation in South Amerindians Is Related to Environmental and Cultural Diversity: Evidence from the Y Chromosome. *Am. J. Hum. Genet.*, 68: 1485-1496.
- Telenius H, Clark J, Marcus E, Royle N, Jeffreys AJ, Ponder BAJ & Mathew CGP (1990) Minisatellite DNA profiles: rapid sample identification in linkage analysis. *Hum. Hered.*, 40(2): 77-80.

- Thanseem I, Thangaraj K, Chaubey G, Singh VK, Bhaskar LV, Reddy BM, Reddy AG & Singh L (2006) Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genet.*, 7:42.
- Tishkoff S & Kidd K (2004) Implications of biogeography of human populations for 'race' and medicine. *Nature Genetics Supplement*, 11: S21-S27.
- Tishkoff AS, Pakstis AJ, Stoneking M, Kidd JR, Destro-Bisol G, Sanjantila A, Lu R-b, Deinard AS, Sirugo G, Jenkins T, Kidd KK & Clark AG (2000) Short Tandem-Repeat Polymorphism/Alu Haplotype Variation at the PLAT locus: Implications for Modern Human Origins. *Am. J. Hum. Genet.*, 67: 901-925.
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazón-Lahr M, Foley RA, Oefner PJ & Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.*, 65: 43-62.
- UNESCO (1950). "The Racial Question".
- Wang J & Caballer A (1999) Developments in predicting the effective size of subdivided populations. *Heredity*, 82: 212-226.
- Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camrena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Llop E, Rothhammer F, Excoffier L & Ruiz-Linares A (2008) Geographic Patterns of Genome Admixture in Latin American Mestizos. *PLoS Genet.*, 4(3): e1000037.
- Watkins WS, Thara R, Mowry BJ, Zhang Y, Witherspoon DJ, Tolpinrud W, Bamshad MJ, Tirupati S, Padmavati R, Smith H, Nancarrow D, Filippich C & Jorde LB (2008) Genetic variation in South Indian castes: evidence from Y-chromosome, mitochondrial, and autosomal polymorphisms. *BMC Genet.*, 9: 86.
- Weber JL, David D, Heil J, Fan Y, Zhao C & Marth G (2002) Human diallelic insertion/deletion polymorphisms. *Am. J. Hum. Genet.*, 71: 854-862.
- Whittle MR, Romano NL & Negreiros VAC (2004) Updated Brazilian genetic data, together with mutation rates, on 19 STR loci, including D10S1237. *Forensic Sci. Int.*, 139: 207-210.
- Whiterspoon DJ, Marchani EE, Watkins WS, Ostler CT, Wooding SP, Anders BA, Fowlkes JD, Boissinot S, Furano AV, Ray DA, Rogers AR, Batzer MA & Jorde LB

(2006) Human Population Genetic Structure and Diversity Inferred from Polymorphic *L1* (*LINE-1*) and *Alu* Insertions. *Hum. Hered.*, 62: 30-46.

Witherspoon DJ, Wooding S, Rogers AR, Marchani EE, Watkins WS, Batzer MA & Jorde LB (2007) Genetic Similarities Within and Between Human Populations. *Genetics*, 176: 351-359.

Yang NN, Mazières S, Bravi C, Ray N, Wahg S, Burley M-W, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Poletti G, Hill K, Hurtado AM, Petzl-Erler MLM, Tsuneto LT, Klitz W, Barrantes R, Llop E, Rothhammer F, Labuda D, Salzano FM, Bortolini M-C, Excoffier L, Dugoujon JM & Ruiz-Linares A (2010) Contrasting Patterns of Nuclear and mtDNA Diversity in Native American Population. *Ann. Hum. Genet.*, 00:1-14.

ANEXOS

ANEXO A – Frequências alélicas dos grupos parentais utilizadas na análise do programa ADMIX

Mostra as frequências alélicas para cada grupo parental. Em destaque, as frequências cujo δ (delta - diferencial de frequência alélica) é maior do que 10% em relação às outras duas parentais.

Loco **D16S539**

Alelo	Europeus	Africanos	Ameríndios
5	0,000	0,002	0,000
8	0,024	0,028	0,000
9	0,113	0,235	0,219
10	0,051	0,146	0,257
11	0,301	0,289	0,236
12	0,299	0,175	0,248
13	0,185	0,119	0,034
14	0,028	0,005	0,004
15	0,001	0,002	0,002

Loco **D5S818**

Alelos	Europeus	Africanos	Ameríndios
7	0,000	0,002	0,174
8	0,007	0,084	0,002
9	0,030	0,024	0,155
10	0,067	0,063	0,030
11	0,348	0,206	0,524
12	0,377	0,378	0,095
13	0,156	0,231	0,017
14	0,013	0,010	0,000
15	0,002	0,002	0,000
16	0,000	0,000	0,002

Loco **D7S820**

Alelo	Europeus	Africanos	Ameríndios
6	0,000	0,000	0,000
7	0,024	0,010	0,002
8	0,158	0,181	0,016
9	0,134	0,120	0,052
10	0,283	0,386	0,229
11	0,204	0,200	0,413
12	0,162	0,080	0,216
13	0,031	0,021	0,053
14	0,003	0,001	0,020

Loco **CSF1PO**

Alelos	Europeus	Africanos	Ameríndios
6	0,000	0,002	0,000
7	0,000	0,069	0,000
8	0,005	0,052	0,007
9	0,021	0,051	0,011
10	0,284	0,285	0,211
11	0,313	0,201	0,251
12	0,308	0,289	0,473
13	0,060	0,042	0,043
14	0,007	0,009	0,004
15	0,002	0,001	0,000

Loco **D13S317**

Alelos	Europeus	Africanos	Ameríndios
6	0,000	0,000	0,000
7	0,000	0,001	0,007
8	0,145	0,018	0,027
9	0,057	0,012	0,322
10	0,049	0,026	0,109
11	0,322	0,303	0,089
12	0,268	0,414	0,155
13	0,112	0,161	0,186
14	0,046	0,065	0,105
15	0,000	0,001	0,000

Loco **TPOX**

Alelos	Europeus	Africanos	Ameríndios
5	0,000	0,000	0,000
6	0,002	0,095	0,000
7	0,002	0,022	0,000
8	0,515	0,331	0,392
9	0,101	0,207	0,009
10	0,064	0,078	0,010
11	0,280	0,254	0,346
12	0,035	0,014	0,243
13	0,001	0,000	0,000

Loco TH01

Alelos	Europeus	Africanos	Ameríndios
5	0,004	0,001	0,000
6	0,210	0,112	0,376
7	0,161	0,398	0,444
8	0,138	0,283	0,004
9	0,199	0,154	0,015
9,3	0,275	0,040	0,159
10	0,011	0,011	0,001
11	0,001	0,000	0,001
12	0,000	0,002	0,000

Loco D3S1358

Alelos	Europeus	Africanos	Ameríndios
11	0,000	0,001	0,000
12	0,002	0,006	0,002
13	0,004	0,010	0,000
14	0,102	0,086	0,021
15	0,260	0,290	0,459
16	0,251	0,335	0,347
17	0,200	0,209	0,155
18	0,168	0,054	0,013
19	0,012	0,008	0,000
20	0,001	0,002	0,002

Loco vWA

Alelos	Europeus	Africanos	Ameríndios
11	0,000	0,011	0,002
12	0,001	0,002	0,001
13	0,002	0,012	0,002
14	0,107	0,077	0,014
15	0,121	0,214	0,052
16	0,231	0,252	0,443
17	0,271	0,189	0,349
18	0,181	0,151	0,108
19	0,073	0,076	0,028
20	0,012	0,014	0,003
21	0,001	0,003	0,000

Loco FGA

Alelos	Europeus	Africanos	Ameríndios
≤17,2	0,001	0,005	0,000
18	0,016	0,005	0,009
18,2	0,000	0,006	0,000
19	0,069	0,061	0,084
19,2	0,001	0,014	0,000
20	0,145	0,054	0,037
20,2	0,001	0,000	0,000
21	0,175	0,081	0,032
21,2	0,001	0,001	0,002
22	0,177	0,173	0,049
22,2	0,006	0,001	0,000
23	0,151	0,173	0,097
23,2	0,002	0,001	0,004
24	0,132	0,177	0,252
24,2	0,001	0,000	0,000
25	0,080	0,112	0,201
26	0,030	0,052	0,149
26,2	0,000	0,001	0,000
27	0,007	0,038	0,043
27,2	0,000	0,001	0,004
28	0,003	0,015	0,026
29	0,000	0,008	0,011
≥30	0,001	0,020	0,000

Loco D21S11

Alelos	Europeus	Africanos	Ameríndios
24,2	0,002	0,004	0,000
25	0,001	0,003	0,000
26	0,001	0,002	0,000
27	0,024	0,070	0,000
28	0,154	0,238	0,065
29	0,216	0,156	0,149
29,2	0,001	0,003	0,002
30	0,251	0,174	0,200
30,2	0,034	0,014	0,019
31	0,063	0,098	0,037
31,2	0,110	0,051	0,103
32	0,009	0,013	0,019
32,2	0,089	0,072	0,214
33	0,001	0,008	0,026
33,2	0,033	0,027	0,056
34	0,001	0,010	0,011
34,2	0,005	0,007	0,013
35	0,002	0,040	0,002
35,2	0,000	0,001	0,002
36	0,001	0,006	0,000
37	0,000	0,001	0,000
38	0,000	0,001	0,000

Loco D8S1179

Alelos	Europeus	Africanos	Ameríndios
8	0,010	0,000	0,000
9	0,008	0,000	0,002
10	0,090	0,006	0,065
11	0,089	0,042	0,032
12	0,129	0,143	0,185
13	0,299	0,186	0,339
14	0,218	0,346	0,205
15	0,128	0,212	0,149
16	0,023	0,058	0,024
17	0,004	0,007	0,000
18	0,001	0,000	0,000

ANEXO B – Frequências alélicas dos grupos parentais para os doze marcadores autossômicos analisados

Mostra as frequências alélicas para cada grupo parental. Os dados ausentes referem-se ao número de indivíduos que não possuíam genótipo para este marcador. Os valores em destaque correspondem aos valores de δ (delta) cuja diferença seja maior que 10% em relação aos demais grupos parentais. Em negrito, os maiores que 10%; em itálico, os menores que 10%.

Loco **D16S539** (23,6% dados ausentes)

Alelo	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
5	0,001	0,004	0,003
8	0,021	0,053	0,004
9	0,142	0,296	0,231
10	0,088	0,133	0,312
11	0,304	0,263	0,196
12	0,273	<i>0,072</i>	0,235
13	0,142	0,166	<i>0,008</i>
14	0,028	0,008	0,004
15	0,002	0,004	0,007

Loco **D5S818** (18,8% dados ausentes)

Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
7	0,009	0,009	0,257
8	0,005	0,141	0,007
9	0,027	0,014	0,147
10	0,073	0,043	0,005
11	0,343	<i>0,135</i>	0,512
12	0,371	0,338	<i>0,046</i>
13	0,157	0,297	<i>0,017</i>
14	0,013	0,011	0,003
15	0,002	0,007	0,003
16	0,000	0,005	0,002

Loco **D7S820** (17,2% dados ausentes)

Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
6	0,001	0,005	0,003
7	0,016	0,006	0,004
8	0,163	0,203	<i>0,008</i>
9	0,110	0,201	0,056
10	0,278	0,344	0,225
11	0,247	0,168	0,410
12	0,162	<i>0,033</i>	0,192
13	0,022	0,029	0,070
14	0,002	0,009	0,031

Loco **CSF1PO** (15,9% dados ausentes)

Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
6	0,000	0,006	0,003
7	0,004	0,098	0,003
8	0,003	0,203	0,003
9	0,015	0,080	0,009
10	0,266	0,241	0,206
11	0,318	<i>0,059</i>	0,261
12	0,326	0,263	0,474
13	0,060	0,036	0,029
14	0,007	0,011	0,006
15	0,001	0,003	0,006

Loco **D13S317** (18,7% dados ausentes)

Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
6	0,001	0,005	0,003
7	0,000	0,004	0,012
8	0,114	0,010	0,008
9	0,088	0,010	0,356
10	0,056	0,034	0,131
11	0,310	0,096	0,086
12	0,264	0,664	<i>0,129</i>
13	0,116	0,148	0,214
14	0,049	0,022	0,058
15	0,002	0,007	0,003

Loco **TPOX** (2,0% dados ausentes)

Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
5	0,000	0,003	0,002
6	0,004	0,166	0,005
7	0,003	0,045	0,005
8	0,491	<i>0,158</i>	0,342
9	0,103	0,316	<i>0,003</i>
10	0,063	0,116	0,014
11	0,289	<i>0,178</i>	0,341
12	0,045	0,014	0,284
13	0,001	0,004	0,004

Loco **TH01** (5,0% dados ausentes)

Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
5	0,001	0,002	0,004
6	0,213	0,017	0,323
7	0,208	0,629	0,479
8	0,138	0,249	0,004
9	0,169	0,076	0,004
9,3	0,260	0,010	0,180
10	0,010	0,014	0,003
11	0,001	0,003	0,003

Loco **D3S1358** (16,1% dados ausentes)

Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
11	0,001	0,008	0,002
12	0,002	0,010	0,003
13	0,004	0,020	0,005
14	0,098	0,130	0,009
15	0,303	0,253	0,492
15,2	0,000	0,003	0,002
16	0,247	0,366	0,327
17	0,202	0,177	0,148
18	0,129	0,016	0,003
19	0,014	0,008	0,004
20	0,001	0,007	0,004

Loco **vWA** (2,8% dados ausentes)

Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
11	0,002	0,013	0,003
12	0,000	0,005	0,003
13	0,004	0,013	0,003
14	0,083	0,052	0,008
15	0,105	0,397	0,058
16	0,259	0,263	0,507
17	0,277	0,091	0,294
18	0,190	0,060	0,114
19	0,062	0,056	0,004
20	0,016	0,040	0,003
21	0,002	0,010	0,003

Loco **FGA** (18,5% dados ausentes)

Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
16	0,000	0,004	0,002
17	0,001	0,008	0,003
17,2	0,000	0,005	0,002
18	0,009	0,016	0,003
18,2	0,001	0,012	0,002
19	0,072	0,040	0,072
19,2	0,000	0,005	0,002
20	0,118	0,027	0,024
20,2	0,000	0,003	0,004
21	0,168	0,057	0,008
21,2	0,000	0,003	0,004
22	0,183	0,227	0,017
22,2	0,002	0,008	0,003
23	0,152	0,115	0,099
23,2	0,000	0,003	0,008
24	0,124	0,239	0,265
25	0,108	0,072	0,222
26	0,041	0,037	0,134
27	0,014	0,046	0,067
28	0,001	0,037	0,030
29	0,001	0,011	0,015
30	0,001	0,009	0,003
30,2	0,001	0,006	0,002
31,2	0,000	0,004	0,002
32	0,000	0,004	0,002
43,2	0,000	0,004	0,002

Loco **D21S11** (18,1% dados ausentes)

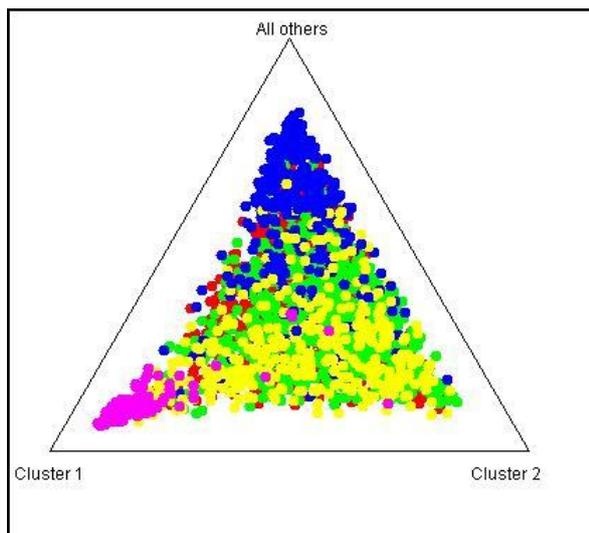
Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
14	0,000	0,003	0,002
15	0,000	0,003	0,002
24,2	0,003	0,008	0,003
25	0,001	0,007	0,003
26	0,001	0,005	0,003
26,2	0,000	0,004	0,002
27	0,022	0,107	0,003
28	0,137	0,320	0,005
29	0,220	0,099	0,169
29,2	0,001	0,011	0,003
30	0,274	0,041	0,249
30,2	0,035	0,009	0,004
31	0,062	0,114	0,032
31,2	0,093	0,055	0,138
32	0,013	0,009	0,004
32,2	0,094	0,035	0,244
33	0,001	0,022	0,003
33,2	0,034	0,031	0,093
34	0,001	0,020	0,003
34,2	0,003	0,007	0,020
35	0,001	0,053	0,003
35,2	0,000	0,004	0,005
36	0,001	0,019	0,003
37	0,000	0,009	0,002
38	0,000	0,004	0,003

Loco **D8S1179** (19,1% dados ausentes)

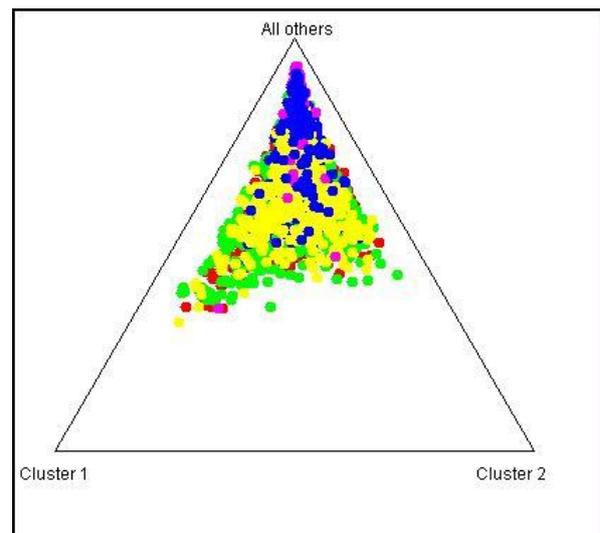
Alelos	Grupo 1 (europeu)	Grupo 2 (africanos)	Grupo 3 (ameríndios)
8	0,007	0,008	0,003
9	0,009	0,013	0,005
10	0,072	0,009	0,081
11	0,086	0,012	0,014
12	0,125	0,089	0,122
13	0,291	0,091	0,340
14	0,253	0,377	0,205
15	0,118	0,313	0,202
16	0,032	0,075	0,025
17	0,005	0,014	0,003

ANEXO C – Gráficos representativos da composição individual (pelo programa STRUCTURE) não utilizados nas conclusões e resultados do trabalho

Apresenta gráficos obtidos a partir do programa STRUCTURE, para $k=3$ e $k=6$ (k sendo o número de agrupamentos populacionais) com o uso da parental africana a partir do conjunto de genótipos do CEPH (*Centre d'Etude Du Polymorphisme Humain* – <http://www.cephb.fr>) e de afro-americanos (Budowle & Moretti, 1999). A combinação dos dois grupos demonstra não ser adequado seu uso para uso como parental africana nos estudos da composição genética da população brasileira.



$k=3$



$k=6$

O resultado para $k=3$ busca agrupar as cinco amostras populacionais (laboratório público, laboratório privado, parental africana, parental europeia e parental ameríndia) em três grupos. Para $k=6$, o programa STRUCTURE procura identificar sub-estruturas dentro dos grupos, de forma a mostrar a existência de mais subpopulações dentro das amostras.

ANEXO D -

Artigo publicado no periódico *Forensic Science International: Genetics*.

Dalton GC, Maia FAS, Mendes CRBO, Pak HI, Paula KAA, Trindade-Filho A & Oliveira SF (2009) Genetic profile of Federal District of Brazil based on 18 STR autosomal loci. *Forensic Science International: Genetics Supplement Series*, 2: 373-375.