



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

**O IMPACTO DAS MIGRAÇÕES NA CONSTITUIÇÃO
GENÉTICA DE POPULAÇÕES LATINO-AMERICANAS**

Neide Maria de Oliveira Godinho

Brasília

2008



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

**O IMPACTO DAS MIGRAÇÕES NA CONSTITUIÇÃO
GENÉTICA DE POPULAÇÕES LATINO-AMERICANAS**

Tese apresentada ao 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

Neide Maria de Oliveira Godinho

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

Brasília

2008

*Dedico a todas as pessoas
de minha família.
De modo especial:
Meu pai e minha mãe
Meu esposo e
minha querida filha SOPHIA,
que renovou e transformou
minha vida para melhor.*

Agradecimentos

Agradeço a Deus pela oportunidade de ter convivido com pessoas maravilhosas e muito melhores que eu, sem as quais eu não teria a oportunidade de crescer como pessoa, de aprender. A cada um que fez e faz parte da minha história, que unidas contribuíram para minha formação pessoal e profissional.

Agradeço a minha orientadora, Professora Silviene Fabiana de Oliveira, que indiscutivelmente me ajudou a ser uma pessoa melhor. Obrigada pela confiança e pelo vínculo de amizade que se formou nesses últimos anos.

À Polícia Técnico-Científica da Secretaria de Segurança Pública do Estado de Goiás pelo investimento em formação de servidor de seu quadro.

Agradeço aos professores do Instituto de Ciências Biológicas pela contribuição no meu aprendizado.

Agradeço ao Prof. Aguinaldo Luiz Simões e Marcelo R. Luizon pelo suporte financeiro, intelectual e estímulo.

Às instituições de fomento que permitiram a realização deste trabalho (FINATEC, FUNPE, CAPES, CNPQ).

Aos meus amigos do laboratório de Genética da UnB, que pacientemente me auxiliaram nos trabalhos de bancada e na compreensão da parte teórica, em especial Carol, Mila, Gabriel, Dudu, Diogo, Guilherme, Rejane.

Aos meus amigos do laboratório químico e de biologia do Instituto de Criminalística de Goiás, pelo incentivo e companheirismo.

À Prodímol pela doação de material para a pesquisa de bancada.

À Polícia Federal pelo apoio técnico na análise de parte dos dados obtidos.

Agradeço à SENASP pelo apoio acadêmico, científico e financeiro.

Ao laboratório Biologia Molecular da Universidade de Brasília pela contribuição na realização de algumas análises laboratoriais.

Agradeço à Associação dos Peritos em Criminalística de Goiás, pelo apoio financeiro e incentivo.

Aos doadores das amostras de sangue, sem o qual não seria possível realizar essa pesquisa.

Agradeço toda minha família, parentes, amigos e colegas.

De modo especial, agradeço a todas as pessoas que cuidaram da minha filha durante minha ausência: minhas queridas irmãs Norma e Meirilucy, que se doaram incondicionalmente para cuidar do meu bem mais precioso; minha amada mãe Maria Salomé pela vigilância constante; meu sogro e pai Altamiro; minha amiga Ana, a segunda mãe de minha filha; minhas queridas cunhadas Sabrina e Eliana. Meu irmão Hernando pelo apoio. Agradeço ainda minha sogra e tia por me apoiar em todas as minhas decisões e fornecer subsídios para o término do meu doutorado

Aos meus queridos sobrinhos e sobrinha, às minhas cunhadas.

Ao meu esposo Aguinaldo e minha querida filha Sophia, por me amarem, serem compreensivos e me apoiarem.

RESUMO

A América Latina tem sido palco de grandes transformações populacionais, sendo hoje caracterizada por ter uma constituição multiétnica, porém com maior influência de três populações parentais - européia, africana e ameríndia. Apesar das similaridades, cada país latino-americano apresentou uma história de povoamento singular, já que ocorreram variações na distribuição das populações parentais e na quantidade de fluxo gênico entre elas. Esse trabalho teve por objetivo avaliar a constituição genética de populações miscigenadas da América Latina e Caribe e a estruturação genética entre elas. Para tanto, foram considerados quatro conjuntos populacionais e dois conjuntos de marcadores. Primeiro, foram analisados treze países da América Latina e Caribe (Argentina, Bahamas, Brasil, Chile, Colômbia, Costa Rica, El Salvador, Equador, Jamaica, México, Peru, Porto Rico e Venezuela), utilizando marcadores do tipo STRs (CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX e vWA). Em um segundo momento, foram avaliadas as diferenças entre as regiões geográficas de três países da América do Sul (Argentina, Brasil e Colômbia) utilizando os mesmos marcadores. Para uma análise mais regional, populações urbanas da região centro-oeste brasileira foram analisadas utilizando os mesmo marcadores anteriores adicionados do Penta E. O povoamento da região centro-oeste foi também avaliado pela comparação de duas populações da região – Goiás e Distrito Federal – utilizando marcadores AIMs autossômicos - APO, AT3, D1, DRD2-A, ECA, FXIIIIB, FyNull, GC, LPL, OCA2, PV92, Rb2300, Sb19.3 e TPA25. As análises de distância genética entre as populações latino-americanas e entre as regiões dos três países mostraram um quadro bastante heterogêneo quanto à distribuição das frequências alélicas dos marcadores utilizados. A análise de mistura corroborou os dados de distância, sendo que a contribuição parental ameríndia variou de 5% a 73%, a africana de 4% a 90% e a européia de 4% a 66%. Com relação à análise das populações da região centro-oeste do Brasil, não foi observado diferenciação entre a distribuição gênica e genotípica, provavelmente devido ao intenso fluxo gênico entre elas. A mistura gênica estimada, considerando os marcadores microssatélites, foi de 11% para a parental ameríndia, 21% para a parental africana e de 68% para a parental européia. Os resultados obtidos com os marcadores AIMs foi de 15%, 21% e 63%, respectivamente. A maioria das populações analisadas apresentou modelo tri-híbrido de mistura. Esses resultados são concordantes com dados históricos e demográficos das populações latino-americanas.

ABSTRACT

The Latin American population has been under great transformations and today it can be characterized as having a multiethnic constitution, however with three main parental populations influence - European, African and Amerindian. Despite their similarities, each country has a unique history of settlement, since there have been variations in the distribution of the parental populations and in the amount of gene flow between them. This study aimed to evaluate the genetic constitution in Latin American and Caribbean populations and the genetic structure between them. For that matter, four groups of populations and two sets of markers were employed. First, thirteen countries of Latin America and Caribbean (Argentina, Bahamas, Brazil, Chile, Colombia, Costa Rica, El Salvador, Ecuador, Jamaica, Mexico, Peru, Puerto Rico and Venezuela), were analyzed with STR markers (CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX and vWA). In a second moment, the differences between the geographic regions of three South-American countries were evaluated (Argentina, Brazil and Colombia) using the same markers. Next, for a more regional analysis, urban populations from Brazilian Middle-Western region were studied using the same markers listed above and Penta E. The settlement of the Brazilian Middle-Western region was also evaluated regarding the comparison of two populations of this region - *Goiás* and Federal District - using autosomic AIMs - APO, AT3, D1, DRD2-A, ECA, FXIIB, FyNull, GC, LPL, OCA2, PV92, Rb2300, SB19.3 and TPA25. The analyses of the genetic distance between the Latin American populations and between the regions of the three countries had shown a heterogeneous picture of the distribution of the allelic frequencies. The admixture analysis corroborated the genetic distance data, with the Amerindian parental contribution presenting variations ranging from 5 to 73%, the African from 4 to 90% and the European from 4 to 66%. Regarding the analysis of the populations of the Brazilian Middle-Western region, differentiation between the genetic and genotypic distribution was not observed, probably due the intense gene flow between them. The genetic admixture estimate, considering the STRs markers, was 11% for the Amerindian parental, 21% for the African parental and 68% for the European parental. The results generated with the AIMs were 15%, 21% and 63%, respectively. The majority of the analyzed populations are tri-hybrid, having been formed by three parental groups. These results agree with historical and demographic data of the Latin American populations.

ABREVIATURAS

APS	Persulfato de amônia
CEP	Conselho de Ética em Pesquisa
CODEPLAN	Companhia de Planejamento do Distrito Federal
CONEP	Comissão Nacional de Ética em Pesquisa
DNA	Ácido desoxirribonucléico
dNTP	Dideoxi-nucleotídeo
EDTA	Ácido etilenodiamino tetra-acético
FUNAI	Fundação Nacional do Índio
IBGE	Instituto Brasileiro de Geografia e Estatística
NaCl	Cloreto de sódio
PCR	Reação em cadeia da polymerase
RFLP	Polimorfismo de comprimento de fragmento de restrição
TBE	Tampão tris-borato
TE	Tampão tris-EDTA
STR	Repetições curtas em <i>tandem</i>
UnB	Universidade de Brasília

LISTA DE TABELAS

Tabela 1.1	Análise de variância molecular (AMOVA) para 13 populações latino-americanas - Argentina, Bahamas, Brasil, Chile, Costa Rica, Colômbia, El Salvador, Equador, Jamaica, México, Peru, Porto Rico e Venezuela – utilizando 13 marcadores genéticos do tipo STR do CODIS	15
Tabela 1.2	Porcentagem de loci STR cujas frequências gênicas apresentaram valores de p para F_{st} na análise de variância molecular ($p < 0,05$) entre os pares de populações	16
Tabela 1.3	Estimativas de contribuição dos três principais grupos parentais na composição genética atual de treze países da América latina utilizando 13 marcadores microssatélites do sistema CODIS	17
Tabela 2.1	Estimativas de contribuição das populações parentais sul-americanas – europeu, africano e ameríndio – na constituição atual das regiões geográficas de Brasil, Colômbia e Argentina, utilizando 13 marcadores microssatélites do sistema CODIS	36
Tabela 3.1	Tribos indígenas localizadas nos Estados da região centro-oeste (Goiás, Mato Grosso e Mato Grosso do Sul) e número total de indivíduos por estado	51
Tabela 3.2.	Discriminação das mesorregiões, microrregiões, do número de habitantes de cada microrregião à época da coleta das amostras de sangue e do número de indivíduos coletados em cada região do Estado de Goiás	53
Tabela 3.3	Frequências alélicas e parâmetros estatísticos estimados para a população do estado de Goiás, correspondente aos 15 marcadores STRs do sistema PowerPlex 16	58
Tabela 3.4	Pares de populações da população brasileira que apresentaram diferenças estatisticamente significativas por marcador STR considerando 14 marcadores e 13 bancos de dados	61
Tabela 3.5	Teste de significância das análises de F_{st} considerando bancos de dados da população brasileira ($\alpha = 5\%$) referente à frequência para 14 marcadores STRs.....	63

Tabela 3.6	Estimativas de contribuição dos três principais grupos parentais na composição genética atual de dois estados brasileiros da região Centro-oeste – Goiás e Mato Grosso do Sul – e do Distrito Federal, utilizando 14 marcadores microsatélites	64
Tabela 4.1.	Distribuição dos indivíduos componentes da amostra do Distrito Federal com relação à região administrativa da residência dos pais à época do nascimento	75
Tabela 4.2	Caracterização das mesorregiões, microrregiões quanto ao número de habitantes de cada microrregião a época da coleta das amostras de sangue e do número de indivíduos coletados em cada microrregião do Estado de Goiás	76
Tabela 4.3	Diferenças nas frequências alélicas (δ) entre os grupos parentais (africano, europeu e ameríndio), dos 14 marcadores autossômicos analisados	78
Tabela 4.4	Características dos marcadores genéticos do tipo AIMs analisados: tipo de marcador, modo de identificação, localização cromossômica, seqüência dos iniciadores, temperatura de pareamento na PCR e referência	80
Tabela 4.5	Frequências alélicas, valores de heterozigose observada e esperada, valores de p do teste de aderência ao Equilíbrio de Hardy-Weinberg e respectivos desvios padrões com relação aos 14 marcadores AIMs autossômicos analisados para as populações de Goiás e Distrito Federal e nas três populações parentais	88
Tabela 4.6	Pares de loci que apresentaram desequilíbrio de ligação ($p < 0,05$) nas populações do Distrito Federal e do estado de Goiás	90
Tabela 4.7	Estatística F (f, F e Theta-P) e componentes da variância (Sigma-P, Sigma-I e Sigma-G) para os 14 loci autossômicos analisados nas populações do Distrito Federal e do estado de Goiás	90
Tabela 4.8	Diferenciação gênica e genotípica entre as populações do Distrito Federal e Goiás considerando 14 marcadores genéticos do tipo AIMs	91
Tabela 4.9	Estimativas de contribuições das parentais africana, ameríndia e europeia na constituição genética atual das populações do Distrito Federal e de Goiás, utilizando 14 marcadores de DNA autossômicos ...	91

Tabela 4.10	Dados da análise de estatística descritiva para os valores IAAs determinados nos indivíduos amostrados na população do Distrito Federal e Goiás	94
-------------	---	----

LISTA DE FIGURAS

Figura 1.1	Mapa da América Latina e Caribe mostrando as variações étnicas entre os países conforme as estimativas de contribuição parental africana, ameríndia e europeia	18
Figura 1.2	Matriz de distância genética (Reynolds <i>et al.</i> , 1983), entre treze populações miscigenadas latino-americanas e as três parentais com referência aos 13 marcadores STRs do CODIS	20
Figura 1.3	Diagrama das duas primeiras coordenadas principais, relativo à distância genética entre os 13 países latino-americanos e caribenhos e suas populações parentais, baseado na análise de 13 marcadores STRs autossômicos	21
Figura 2.1	Mapa da Argentina, Brasil e Colômbia e suas respectivas regiões geográficas adotadas para as análises desenvolvidas neste trabalho	33
Figura 2.2	Matriz de distância genética de Reynolds <i>et al.</i> (1983) mostrando as distâncias genéticas entre as regiões geográficas da Argentina, Brasil, Colômbia e as populações parentais.....	39
Figura 2.3	Primeira e segunda coordenadas principais obtidas a partir da análise de 13 marcadores STRs mostrando o relacionamento entre a distribuição das frequências alélicas nas regiões de Argentina, Brasil e Colômbia e as populações parentais	40
Figura 3.1	Mapa do estado de Goiás destacando as dezoito microrregiões, de acordo com classificação do sistema de cartografia do IBGE (2000)	52
Figura 3.2	Matriz de distância genética de Reynolds <i>et al.</i> (1983) para as populações do Distrito Federal, Goiás, Mato Grosso do Sul, da região centro-oeste como um todo, bem como das populações parentais utilizando 14 marcadores genéticos do tipo STRs.....	64

Figura 4.1	Regiões Administrativas que compunham o Distrito Federal em 2002 (Fonte: Codeplan, 1997), época da coleta de material biológico	75
Figura 4.2	Histograma da distribuição dos valores de índice de ancestralidade africana nas populações do Distrito Federal e do Estado de Goiás	93

SUMÁRIO

INTRODUÇÃO.....	1
Migração e evolução.....	1
Migração na América Latina.....	4
OBJETIVOS.....	6
CAPÍTULO 1.....	8
<i>Panorama étnico da América Latina pela análise de marcadores microssatélites autossômicos</i>	
1.1. INTRODUÇÃO.....	9
1.2 MATERIAL E MÉTODOS.....	11
Dados de frequências gênicas.....	11
Análise estatística.....	12
1.3. RESULTADOS.....	13
1.4. DISCUSSÃO.....	22
Diferenciação populacional na América latina.....	22
Mistura genética.....	23
CAPÍTULO 2.....	27
<i>Padrões regionais de mistura genética na américa do sul</i>	
2.1. INTRODUÇÃO.....	28
2.2. MATERIAL E MÉTODOS.....	32
Fonte de Dados	32
Análises populacionais.....	34
2.3. RESULTADOS.....	35
2.4. DISCUSSÃO.....	41
A presença européia.....	41
O africano e o ameríndio.....	43
CAPÍTULO 3.....	47
<i>Mistura gênica na região centro-oeste brasileira estimada utilizando dados de marcadores microssatélites autossômicos</i>	
3.1. INTRODUÇÃO.....	48
3.2. MATERIAL E MÉTODOS.....	52
Amostras populacionais.....	52
Levantamento de dados para realização das análises estatísticas.....	54
Análise estatística.....	55
Dos dados gerados para a população de Goiás.....	55
Análises interpopulacionais.....	56
3.3. RESULTADOS.....	57
Análise intrapopulacional.....	57

Análise interpopulacional.....	57
3.4. DISCUSSÃO.....	65
CAPÍTULO 4.....	69
<i>Constituição genética de Goiás e Distrito Federal, duas populações da região Centro-Oeste Brasileira, com relação a marcadores do tipo AIMS autossômicos</i>	
4. 1. INTRODUÇÃO.....	70
4. 2 MATERIAL E MÉTODOS.....	73
Caracterização da população e coleta das amostras.....	73
Distrito Federal.....	73
Goiás.....	76
Procedimentos metodológicos.....	77
Seleção dos marcadores genéticos	77
Extração de DNA e verificação em gel de agarose	77
Reação em Cadeia da Polimerase (PCR) e visualização dos produtos de amplificação	78
Reação de restrição	79
Eletroforese e avaliação fenotípica	82
Análise estatística.....	82
4. 3. RESULTADOS.....	86
4. 4. DISCUSSÃO.....	94
Mistura.....	98
Contextualização dos dados apresentados.....	102
CONSIDERAÇÕES FINAIS.....	104
REFERÊNCIA BIBLIOGRÁFICA.....	106
Sítios da internet.....	121
ANEXO 1: TABELAS SUPLEMENTARES.....	122
ANEXO 2. PROTOCOLOS DOS PROCEDIMENTOS LABORATORIAIS.....	133
ANEXO 3. TERMOS DE CONSENTIMENTO LIVRE E ESCLARECIDOS.....	141

INTRODUÇÃO

INTRODUÇÃO

Uma questão que intriga os estudiosos em populações humanas é porque as pessoas migram, visto que a maioria dos migrantes não gostaria, de fato, abandonar suas casas, terras, seus locais de origem. Fatores de expulsão e de atração são os mais importantes fatores que levam à migração. Exemplos de fatores de expulsão são escassez de alimentos e perseguições político-religiosas. Os principais fatores de atração são os relacionados à carência de mão-de-obra, facilidades na obtenção de terra, migração subsidiada, maiores salários na área para onde os migrantes se encaminham, dentre outros (Klein, 2000).

As formas de migração estão relacionadas com o destino dos indivíduos migrantes, isso é, podem ocorrer dentro de um território, sendo assim chamadas migrações internas, ou para outro território ou país, sendo chamadas de migrações externas. A mobilidade dentro do território geralmente não sofre restrições, mas as migrações externas, atualmente, dependem de normas estabelecidas pelos chefes do Estado receptor.

Restrições à migração são práticas mais recentes. O povoamento das Américas é um exemplo marcante de migração em massa com enorme modificação na constituição populacional humana da população receptora. Logicamente que restrições às migrações não se aplicaram nesse caso. Muito pelo contrário, houve incentivos à migração européia (migração subsidiada), assim como a migração forçada de povos africanos à época das grandes navegações em virtude do tráfico de indivíduos africanos para as Américas. O contexto da migração européia para a América pode ser melhor compreendido à luz do conhecimento de que os europeus viviam um momento de recessão, onde a dificuldade de acesso à terra e ao alimento era um fator de repulsão. Por outro lado, o descobrimento de um novo território, o americano, atraía e apresentava-se naquele momento histórico como solução. Assim, a possibilidade de acesso à terra, associado ao incentivo dado pelos governantes para a contratação de mão-de-obra de modo a suprir a economia do novo mercado de trabalho, além dos subsídios fornecidos para estimular a migração, foram os principais fatores de atração que contribuíram para a migração européia.

Migração e evolução

A migração é um fator evolutivo que pode levar populações à homogeneização (Wang & Caballero, 1999). Pode ser definida como o deslocamento físico de um indivíduo ou grupos de indivíduos de uma sociedade ou área geográfica para outra. Porém, o deslocamento de indivíduos de uma população para outra só faz sentido em termos de

evolução se tais indivíduos estiverem na idade reprodutiva. Isso torna possível a introdução de genes da população migrante para a população receptora.

Caso exista diferença na constituição genética entre a população receptora e a doadora, haverá a possibilidade de introdução de novos alelos na receptora, o que a tornará mais semelhante à população de origem dos migrantes. Portanto, necessário se torna também que os indivíduos migrantes tenham sucesso reprodutivo, pois somente assim haverá a modificação efetiva na constituição genética da população. Esse evento decorrente da migração é denominado fluxo gênico (Barbujani, 2005).

Geralmente a troca de material genético entre populações se dá entre aquelas que se encontram à uma menor distância geográfica. Portanto, o fluxo gênico entre grupos populacionais é inversamente proporcional à distância geográfica entre eles (Handley *et al.*, 2007; Serre & Pääbo, 2004). Um fator decorrente do fluxo gênico é que as modificações genéticas não se darão apenas com relação a um determinado gene, mas as alterações ocorrerão no genoma como um todo, já que cada indivíduo de um par de acasalamento doará 23 cromossomos.

Sabe-se que pessoas de diferentes lugares, em especial continentes, podem apresentar diferenças fenotípicas, e que parte dessas diferenças é decorrente das diferenças entre as frequências alélicas em determinados genes. Desse modo, frequências alélicas podem ser comparadas com o intuito de definir o quanto uma população difere da outra. Observa-se que há uma maior similaridade genética entre populações próximas geograficamente, isso é, as diferenças entre os agrupamentos populacionais tendem a seguir a distância geográfica que uma tem em relação às demais (Manica *et al.*, 2005; Jorde & Wooding, 2004). Dessa forma, populações mais distantes geograficamente têm menor probabilidade de trocarem migrantes. Conseqüentemente, as variações genéticas entre as populações tendem a ser geograficamente estruturadas, pois os indivíduos de uma região geográfica são mais similares entre si do que de indivíduos de uma região mais distante. Como a migração geralmente é constante entre as populações humanas, e tende a ocorrer para regiões vizinhas, as variações genéticas observadas se distribuem em um padrão contínuo, e raramente tem descontinuidades geográficas marcantes (Behar *et al.*, 2008).

Embora a migração ocorra na maioria das populações humanas, determinados grupos populacionais se diferenciaram, devido a mutações em alguns genes, deriva genética e seleção. Entre os fatores evolutivos que levam à diferenciação populacional, a deriva genética é a que promove as alterações entre as populações em um tempo menor e

de modo mais intenso, pois ela produz alteração em todo o genoma do indivíduo e não apenas em um dado gene (Jorde & Wooding, 2004; Destro-Bisol *et al.*, 2000). Por exemplo, a migração de um grupo de indivíduos para uma outra região inicialmente desabitada pode originar uma população distinta da população original, pois é provável que nesse grupo não estejam representados todos os alelos de um determinado gene da população original. Esse efeito é conhecido como efeito fundador, o que promove, conseqüentemente, a diferenciação entre as populações em virtude da perda de alelos e da fixação de outros. Exemplo disso são os ameríndios que ao migrarem para a América trouxeram consigo apenas parte dos alelos que existiam na população original. O grupo sanguíneo ABO mostra de modo muito claro este efeito. Os ameríndios atuais praticamente não apresentam os alelos A e B, sendo que a maior parte das populações apresenta monomorfismo do alelo O (Ollson, 2002). A teoria mais aceita para esse fenômeno é que o efeito do fundador de um pequeno número de migrantes que deu origem ao povoamento da América seja responsável por isso, embora exista a hipótese de que tenha ocorrido seleção natural em favor dos indivíduos do tipo O (Cavalli-Sforza *et al.*, 1996).

Pode ocorrer, entretanto, que as diferenças genéticas não sejam proporcionais às distâncias geográficas, mas decorrente de um fator que impede a reprodução entre esses grupos, como, por exemplo, um acidente geográfico ou uma barreira cultural (Comas *et al.*, 2000). O cruzamento preferencial é outro fator que pode promover ao longo do tempo pequenas variações entre as populações. É o que ocorreu, por exemplo, com populações latino-americanas. De acordo com dados genéticos e históricos, no início da colonização americana houve cruzamento preferencial entre homens europeus e mulheres ameríndias e africanas (Lind *et al.*, 2007; Martinez *et al.*, 2007; Bedoya *et al.*, 2006; Abe-Sandes *et al.*, 2004; Martinez-Marignac *et al.*, 2004). Tais estudos mostram que, nessas populações, as linhagens ameríndias e africanas foram passadas para as próximas gerações pela linhagem matrilinear, enquanto que o componente genético europeu predominou nas linhagens patrilineares. O cruzamento preferencial ocorreu em decorrência do predomínio de indivíduos do sexo masculino entre os migrantes europeus nos primeiros séculos de colonização.

De fato, o movimento migratório em massa mais marcante nos últimos séculos foi o de colonização da América, que resultou na formação de populações híbridas, como mostram vários estudos de estimativas de mistura em populações da América (Martinez *et al.*, 2007; Tang *et al.*, 2007; Bedoya *et al.*, 2006; Choudhry *et al.*, 2006; Gadelha *et al.*, 2005; Bonilla *et al.*, 2004). Para tais análises, freqüentemente tem-se usado marcadores

que apresentam diferenças nas frequências alélicas entre os grupos parentais que originaram a população miscigenada. Determinar um conjunto de marcadores que possam discriminar grupos étnicos é difícil, porque as maiores diferenças acontecem entre indivíduos dentro de uma população do que entre populações. A variabilidade entre grupos populacionais ou étnicos é pequena, justamente porque as diferenças entre as frequências alélicas dos genes/marcadores genéticos são pequenas na maioria dos casos (Barbujani, 2005). Entretanto, essas diferenças são suficientes para separar grupos geográficos nas análises filogenéticas (Luizon *et al.*, 2008; Rowold e Herrera, 2003; Comas *et al.*, 2000; Batzer *et al.*, 1996).

Um dos efeitos provocados pelo fluxo gênico é que poderão ocorrer associações alélicas entre dois *loci* que não estejam ligados. Chakraborty e Weiss (1988) denominaram tais associações como desequilíbrio de mistura para distinguir das associações que ocorrem entre *loci* ligados. Essas associações decaem com o tempo e depende da distância genética entre os grupos parentais, mas o desequilíbrio de ligação pode permanecer por longos períodos, caso a história de mistura não seja muito antiga. Portanto, o desequilíbrio de ligação entre dois *loci* em uma população híbrida pode ser decorrente apenas do processo de mistura, não significando necessariamente relação de ligação física entre eles. O desequilíbrio de ligação tende a diminuir com o tempo devido à recombinação recorrente em várias gerações de cruzamento. Compreender esse processo e discernir em uma população a origem do desequilíbrio entre *loci* tem ajudado no mapeamento de genes relacionados a doenças (Nalls *et al.*, 2008; Martinez-Marignac *et al.*, 2007; Bonilla *et al.*, 2004; Shriver *et al.*, 2003; Kidd *et al.*, 1998).

Migração na América Latina

O descobrimento da América pelos europeus, no século XV, promoveu uma onda migratória intensa envolvendo pelo menos dois continentes doadores, isto é, a Europa e a África, e a América como receptora de tais etnias. Estima-se que cerca de 15 milhões de europeus tenham cruzado o Atlântico até 1880 e que entre 1881 a 1915 o número de migrantes alcançou seu apogeu, chegando à pelo menos 31 milhões de indivíduos (Klein, 2000). A maioria dos migrantes era de jovens adultos do sexo masculino em busca de emprego, temporário ou permanente, em território americano, sendo que metade desses imigrantes permaneceu na América. Outro grupo étnico que veio para a América em massa foi o africano que foi trazido pela força, para trabalhar como escravo na agricultura e na

extração de metais preciosos. Estima-se que até 1880 tenham sido trazidos para o continente americano pelo menos 10 milhões de africanos. A migração em massa de europeus e de africanos, promoveu alterações genéticas tanto nas populações locais, quanto nas populações migrantes em decorrência da miscigenação, havendo o estabelecimento de populações com vários tipos e graus de mistura em toda América. Além disso, a migração em massa promoveu também mudanças demográficas, havendo redução da população nativa em aproximadamente 90% do que existia à época do início da colonização (Bedoya *et al.*, 2006).

Quanto à origem do grupo parental ameríndio existem controvérsias quanto a sua chegada à América, mas é consenso que todos ou a maioria vieram da Ásia via Estreito de Bering (Cavalli-Sforza *et al.*, 1996). Estudos utilizando marcadores do DNA mitocondrial, do cromossomo Y e de autossômicos buscam compreender a origem do homem ameríndio e como ele chegou à América (Fagundes *et al.*, 2008; Wang *et al.*, 2007; Bandelt *et al.*, 2003; Underhill *et al.*, 2003; Silva *et al.*, 2002).

Devido a essas diferenças históricas e ao padrão de colonização e migração na América, pode-se observar atualmente uma variação quanto às contribuições entre esses três grupos geográficos parentais na formação das populações urbanas e rurais da América Latina (Sans, 2000). O encontro desses três grupos parentais possibilitou o aparecimento de novos grupos étnicos ou novas populações, como a de indivíduos que se autodenominam pardos, que são produto de cruzamento por várias gerações de colonizadores (europeus e africanos) e seus descendentes, com populações locais. Outro fator que contribui para esse quadro complexo de mistura na América é a ocorrência de migrações internas motivadas por distintos fatores, dentre eles os políticos ou devido à emergência de novas oportunidades econômicas, como por exemplo, o que ocorreu na Venezuela na década de 70 do século passado. Associado à instabilidade política na América do Sul e à prosperidade experimentada pela Venezuela por causa da existência de petróleo, muitos indivíduos de vários países da América do Sul e do Caribe migraram para esse país, colaborando ainda mais para o processo de amalgamento experimentado na região (Martinez *et al.*, 2007). De um modo geral, não apenas nas populações latino-americanas, como da Venezuela, exemplo citado acima, a mistura étnica é um processo contínuo que ocorre ao longo de várias gerações.

OBJETIVOS

O objetivo do presente trabalho foi avaliar o impacto das migrações na constituição atual de populações latino-americanas utilizando marcadores genéticos moleculares do tipo microssatélites, inserções *Alu*, inserções/deleções e SNPs. A hipótese inicial é saber se as migrações contribuíram para homogeneizar as populações. Para tanto, foram analisados distintos níveis populacionais: 1. da América Latina; 2. de regiões geográficas da Argentina, Brasil e Colômbia; 3. região centro-oeste brasileira.

Visando atingir esse objetivo geral, foram buscados:

Capítulo 1 – Avaliar o impacto das migrações considerando países da América Latina e Caribe, utilizando dados de marcadores do tipo microssatélites (CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX e vWA). Para tanto buscou-se:

- a) Estimar a diversidade gênica intrapopulacional h_{sk} para cada *locus*;
- b) Avaliar a diferenciação populacional e estruturação populacional pela análise de variância molecular (AMOVA);
- c) Estimar mistura genética na constituição atual de 13 países;
- d) Estimar distâncias genéticas de Reynolds *et al.* (1983) entre as populações utilizando matriz de distância e análise do componente principal.

Capítulo 2 – Buscar compreender o impacto das migrações decorrentes da colonização em três países da América do Sul - Argentina, Brasil e Colômbia, utilizando os mesmos marcadores microssatélites do capítulo 1 baseado em:

- a) Estimativa de mistura étnica para cada país e suas respectivas regiões geográficas;
- b) Avaliação das distâncias genéticas de Reynolds *et al.* (1983) entre os pares de populações com base na matriz gerada e na análise do componente principal.

Capítulo 3 - Avaliar a região centro-oeste brasileira quanto ao seu processo de formação genética e estruturação populacional, com base em dois estados, Goiás e Mato Grosso do Sul, e o Distrito Federal utilizando marcadores do tipo microssatélites. Para alcançar esse objetivo, as estratégias utilizadas foram:

- a) Estimar as freqüências alélicas de 15 marcadores STRs (CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, Penta D, Penta E, TH01, TPOX e vWA) para a população do estado de Goiás;

-
- b) Estimar heterozigose observada e esperada e o conteúdo informativo de polimorfismo (PIC) para cada marcador;
 - c) Comparar as frequências genótípicas observadas com as esperadas de acordo com a Lei de Hardy-Weinberg;
 - d) Avaliar a ocorrência de desequilíbrio de ligação entre os 15 *loci* STRs analisados;
 - e) Utilizar os dados levantados para as populações da região centro-oeste brasileira para realizar uma série de comparações e estimativas:
 - I. AMOVA, índice de estruturação populacional (F_{st}) e o teste de significância do mesmo para comparar os estados de Goiás e Mato Grosso do Sul e o Distrito Federal;
 - II. Diversidade gênica intrapopulacional h_{sk} (Nei, 1987), para cada *locus* para as populações da região centro-oeste (Distrito Federal, Goiás e Mato Grosso do Sul);
 - III. AMOVA, índice de estruturação populacional (F_{st}) e o teste de significância do mesmo para comparar as populações da região centro-oeste com outras populações brasileiras;
 - IV. Índices de mistura genética para as três populações do centro-oeste.

Capítulo 4 - Avaliar a constituição e estruturação populacional do estado de Goiás e do Distrito Federal pela análise de 14 marcadores autossômicos informativos de ancestralidade - AIMS (APO, AT3, D1, DRD2-A, ECA, FXIIB, FyNull, GC, LPL, OCA2, PV92, Rb2300, Sb19.3, TPA25). Para atingir este objetivo definiu-se:

- a) Determinar a frequência alélica dos marcadores autossômicos AIMS, heterozigose observada e esperada para cada *locus* e avaliar se esses marcadores encontram-se em equilíbrio segundo o proposto por Hardy-Weinberg;
- b) Estimar o desequilíbrio de ligação entre todos os *loci* e entre os pares de *loci*;
- c) Comparar as populações do estado de Goiás e do Distrito Federal pelo teste exato de diferenciação populacional e da estatística F;
- d) Estimar mistura genética nas duas populações (Distrito Federal e Goiás), utilizando os 14 marcadores AIMS;
- e) Estimar o índice de ancestralidade individual nas duas populações e a distribuição em uma curva normal.

CAPÍTULO 1

Panorama étnico da América Latina e Caribe pela análise de marcadores microssatélites autossômicos

1.1. INTRODUÇÃO

A expansão de vários impérios europeus em decorrência das grandes navegações, a partir do século XV, promoveu transformações econômicas, culturais e genéticas nas populações dos vários continentes e em especial na América. Antes da chegada de Cristóvão Colombo à América, em 1492, a população local era constituída por ameríndios, uma população bastante heterogênea, tendo em vista não só a cultura praticada, mas também devido à existência de uma infinidade de grupos lingüísticos.

A migração européia para a América Latina foi marcada por um fluxo intenso de indivíduos de vários países, principalmente povos da península Ibérica (Portugal e Espanha), e, em um segundo momento, italianos e alemães. Além da migração européia acentuada até o início do século XX, outro fator contribuiu para alterar o perfil genético populacional da América: a migração forçada de povos africanos (Klein, 2000). Esses povos foram advindos da África subsaariana, especialmente dos atuais países de Congo e Angola e da antiga Costa do Ouro (atual Gana), e pertenciam a diferentes grupos étnicos e muitas vezes lingüísticos, favorecendo ainda mais uma formação complexa da nova população americana (Wang *et al.*, 2008).

A distribuição do contingente populacional africano não foi homogênea ao longo do território latino-americano, o que reflete a composição genética atual das populações. Observa-se uma forte influência africana, em especial cultural, em países do Caribe e no Brasil enquanto que em outros países latino-americanos, como aqueles localizados nas regiões andinas, essa influência é pequena (Wang *et al.*, 2008).

Uma estratégia para melhor compreender a história da constituição genética da população da América Latina é estimar a contribuição das populações parentais baseada na distribuição das frequências alélicas de marcadores genéticos. Dentre os marcadores, estão os STRs (*Short Tandem Repeats* – repetições curtas em tandem) com um grande volume de dados na literatura, em especial com relação aos utilizados pelo sistema CODIS (*Combined DNA Index Systems*). Cada *locus* STR constitui um *locus* genético altamente variável, multialélico, com expressão codominante, havendo assim a possibilidade de seus alelos poderem ser detectados e discriminados na população (Schröer *et al.*, 2000). O polimorfismo nos *loci* STRs resulta da variação do número de unidades de repetições (Drobnic & Budowle, 2000).

Essas variações, possivelmente decorrentes de mutações que favorecem pequenas mudanças, geralmente de uma unidade de repetição (Martinovic *et al.*, 1999), envolvem diferenças no comprimento do microssatélite em decorrência de erros da enzima

polimerase. Durante a replicação de uma região repetitiva, as cadeias de DNA podem dissociar-se e então se associarem incorretamente, levando a inserção ou deleção de unidades repetidas, alterando, assim, o comprimento do alelo (Ellegren, 2000; Hancock, 1996; Kruglyak *et al.*, 1998).

Sua abundância no genoma e sua alta taxa de mutação, comparado com outros tipos de polimorfismos, tornam-os úteis em estudos microevolutivos. Dentro de um curto período evolutivo, um número suficiente de eventos mutacionais acumulados pode ser observado em populações divergentes (Wang *et al.*, 2008; Agrawal & Khan, 2005; Rowold & Herrera, 2003). Por isso, recentemente está sendo dada atenção para a sua utilidade em estudos mais detalhados de grupos populacionais humanos, com objetivo de inferir alguns aspectos históricos demográficos e avaliar mistura étnica em populações miscigenadas.

A avaliação do processo das expansões humanas utilizando marcadores microssatélites é um dos trabalhos que demonstrou essa questão (Bowcock *et al.*, 1994). Quando árvores filogenéticas de distâncias foram construídas, foi observado que os indivíduos agrupavam-se de acordo com sua origem geográfica. A grande maioria dos indivíduos examinados foi agrupada em grupos discretos que coincidem com o continente de origem da amostra. Além disso, apresentou uma tendência para formar subgrupos dentro dos continentes que correspondiam à população de origem da amostra.

A dinâmica de mistura entre populações pode produzir desequilíbrio de ligação entre os *loci* STRs analisados mesmo quando são sintênicos mas não ligados e ainda quando os *loci* não se encontram em um mesmo cromossomo. O desequilíbrio produzido pela miscigenação será numericamente proporcional ao produto das diferenças das freqüências alélicas entre as populações que sofreram miscigenação, significando que o desequilíbrio poderá ser maior do que aquele produzido puramente por mutação e deriva em *loci* ligados, a menos que os *loci* sejam muito ligados (Pfaff *et al.*, 2001).

A possibilidade de avaliar as relações genéticas de indivíduos amostrados de uma população com base em genótipos multilocos abre novas perspectivas em análises genéticas populacionais. Este tipo de dados elimina a necessidade de definir população *a priori*, o que pode ajudar em estudos de movimentos populacionais recentes, bem como facilita a estimativa de mistura individual (Martinez *et al.*, 2007; Ferreira *et al.*, 2006; Callegari-Jaques *et al.*, 2003; Cerda-Flores *et al.*, 2002; Cavalli-Sforza, 1997). Dessa forma, os microssatélites têm sido considerados como marcadores genéticos ideais na identificação de indivíduos e também na descrição de processos históricos de migração e outros eventos evolutivos mais recentes, como o processo de formação étnica e

miscigenação, incluindo a que ocorreu na América Latina (Wang *et al.*, 2008; González-Andrade *et al.*, 2007; Martínez *et al.*, 2007; Toscanini *et al.*, 2007; Ferreira *et al.*, 2006; Ferreira *et al.*, 2005; Callegari-Jacques *et al.*, 2003; Cerda-Flores *et al.*, 2002).

Este capítulo teve por objetivo traçar um panorama da composição genética da América Latina com relação à porcentagem de contribuição das três principais populações parentais – europeus, africanos subsaarianos e ameríndios – e as similaridades e diferenças genéticas entre as populações de 13 países da América Latina. Para a realização das análises, que incluíram também estimativas de distância genética, foram utilizados dados de frequências alélicas de 13 marcadores genéticos do tipo STR do sistema CODIS.

1.2 MATERIAL E MÉTODOS

Dados de frequências gênicas

Os países considerados neste trabalho foram: da América do Sul - Argentina, Brasil, Chile, Colômbia, Equador, Peru e Venezuela; da América Central - Costa Rica, El Salvador, Bahamas, Jamaica e Porto Rico, estes três últimos localizados no Mar do Caribe, e da América do Norte - México. Para a realização das análises foi estruturado um banco de dados das frequências alélicas para 13 marcadores STRs adotado pelo CODIS – CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX e vWA – a partir de dados obtidos da literatura. O critério utilizado para a seleção dessas populações se baseou na disponibilidade dos dados publicados na literatura. O número de indivíduos analisados variou entre os países e entre os *locus*, podendo ser conferido na Tabela 1 do Anexo 1. As frequências gênicas para cada *locus* para cada país foram obtidas utilizando a média ponderada, quando dados de mais do que uma população do mesmo país estava disponível. Nos casos de Porto Rico, Bahamas, Jamaica, Equador e Peru havia disponibilidade de apenas um banco de dados.

A escolha das populações parentais baseou-se no histórico de povoamento da América Latina e Caribe, considerando como maiores contribuintes povos europeus, ameríndios e africanos subsaarianos. Com relação às populações parentais européias, foi considerado que o maior fluxo gênico para a América Latina veio da Península Ibérica. Para o cálculo das frequências gênicas da população parental européia, foi efetuada a média ponderada de frequências gênicas disponíveis para as populações de Portugal e da Espanha obtidos no banco de dados da Universidade de Düsseldorf (<http://www.uni->

duesselforf.de/WWW/MedFak/Serology/Database.html, acessado dia 03/05/2008), em Paredes *et al.*, 2003a e em Fernandes e Brehm (2002).

Com relação ao grupo parental ameríndio, foram consideradas diferentes populações a depender do país em questão. Para os países da América do Sul foram utilizadas populações deste subcontinente: *Kichwas*, *Cubeo*, *Desano*, *Puinave*, *Tehuelche*, *Tucano*, *Wichi* e *Curripaco* (González-Andrade *et al.*, 2006; ALFRED – *Allelic Frequency Database* <http://alfred.med.yale.edu> – acessado em 03/05/2008). Excepcionalmente para a população do Equador foram utilizados apenas dados de *Kichwas* por ser essa a população nativa daquele país e apresentar o conjunto completo de dados de frequência quanto aos marcadores analisados. Para as populações da América Central foi utilizada a média ponderada de frequências de populações ameríndias de El Salvador - Conchagua, San Alejo, Panchimalco e Izalco (Lovo-Gómez *et al.*, 2007). Para o México foram realizadas duas análises: uma considerando os dados de Navajo e Apache (sítio ALFRED) e outra com as populações de El Salvador supracitadas utilizadas para a América Central.

Para o grupo parental africano, a maioria utilizada foi de populações do oeste da África subsaariana. Os dados foram obtidos no banco de dados da Universidade de Düsseldorf - populações de Nigéria, Zimbábue, Benin, *Bamileke* e *Ewondo* de Camarões, *Ovambo* da Namíbia e uma amostra de Angola e Guiné Bissau - e em artigos científicos - Angola (Alves *et al.*, 2004), Guiné Bissau (Gonçalves *et al.*, 2002), Gabão (Steinlechner *et al.*, 2002), Guiné Equatorial (Alves *et al.*, 2005; Gene *et al.*, 2001) e Ruanda (Tofanelli *et al.*, 2003).

Análise estatística

As proporções de contribuição das populações parentais na constituição da população atual dos países da América Latina (estimativa de mistura genética) foram estimadas usando o método de Chakraborty (1985), que considera as probabilidades de identidade gênica. Essa análise foi realizada utilizando o programa ADMIX 3. Pelo fato de Argentina e Brasil possuírem dados disponíveis para a maioria de suas regiões geográficas, após a análise de mistura para cada região (será discutido no capítulo seguinte) foi feita análise de mistura total para cada país levando-se em consideração o número de habitantes de cada região geográfica. Esta última análise foi realizada devido às diferenças de densidade demográfica de cada região, bem como das diferenças na composição étnica entre as regiões.

As distâncias genéticas entre as populações foram estimadas utilizando a metodologia proposta por Reynolds (1983), a qual assume que as mudanças nas frequências são decorrentes apenas dos efeitos de deriva genética. Esta análise foi realizada com o uso do software PHYLIP versão 3.65 (Felsenstein, 1989). Para Análise de Coordenada Principal (PCA), gerada pelo programa NTSYS-pc versão 1.70 (Rohlf, 1992), foi utilizada essa matriz de distância genética.

A diversidade gênica intrapopulacional h_{sk} (Nei, 1987) foi estimada para cada *locus* para todas as populações. A análise de variância molecular (AMOVA) foi utilizada para avaliar a diferenciação populacional estimando a partição da variância que é decorrente da diferenciação intrapopulacional e a interpopulacional (Excoffier *et al.*, 1992). Foi estimado também o índice de estruturação populacional F_{st} (Weir e Cockerham, 1984) assim como realizado o teste de significância do mesmo, de acordo com o descrito por Excoffier *et al.* (1992). Para todas essas análises foi utilizado o programa Arlequin versão 2000 (Schneider, 2000). Dos resultados obtidos foram verificados quais valores de F_{st} apresentaram diferenças estatisticamente significativas e calculado o percentual dessas diferenças para cada par de populações. Para a estimativa da diversidade gênica média para os diferentes marcadores e de cada marcador para as diferentes populações foi realizada uma média ponderada.

1.3. RESULTADOS

A análise de variância molecular (AMOVA) indicou subestruturação populacional entre os 13 países analisados. O valor de F_{st} variou de 0,002, para o marcador CSF1PO, a 0,013 para o marcador TH01 (Tabela 1.1). O marcador com maior poder de discriminação foi o TH01, que mostrou diferenças para 82% dos 78 pares de populações possíveis, e o com menor poder de discriminação foi o CSF1PO, 35%. Apesar dos valores de F_{st} estimados serem baixos, o teste de significância mostrou um valor de $p < 0,00001$ para todos os marcadores STRs analisados individualmente. A maior causa das variações foram decorrentes das variações dentro das populações (Vb).

Análise ao nível intrapopulacional foi realizada com objetivo de observar as variações genéticas em cada população. Para isso, a média dos índices obtidos para diversidade em cada país foi estimada considerando todos os marcadores analisados. Os valores de diversidade gênica obtidos para os 13 marcadores (Tabela 1.1) variaram de 0,69 a 0,88. Com relação à diversidade gênica média estimada para cada país, os valores foram

altos e muito similares, variando de 0,76 para o Peru a 0,80 para Bahamas, Brasil, Porto Rico e Venezuela (Tabela 1 - Anexo 1).

A tabela 1.2 mostra a porcentagem de marcadores que apresentaram diferenças nas frequências alélicas considerando os pares de populações (F_{st} com $p < 0,05$). Jamaica e Bahamas não apresentaram diferenças. Peru apresentou apenas uma diferença (marcador TH01) com relação à El Salvador e Equador. Chile e Argentina apresentaram diferença quanto ao marcador D18S51. Diferença para 100% dos marcadores foi observada entre Jamaica e três países (El Salvador, Costa Rica e Colômbia), Brasil e dois países (Costa Rica e Colômbia) e Bahamas com El Salvador. Considerando os 78 pares de combinações possíveis entre as populações, verificou-se que em 64% deles houve diferença em pelo menos seis dos marcadores utilizados.

A tabela 1.3 mostra as estimativas de contribuição dos três grupos parentais principais na composição genética atual de 11 países latino-americanos e dois caribenhos com base nos 13 STRs e os erros padrões associados às mesmas. A figura 1.1 mostra o mapa da América Latina e Caribe com a distribuição das variações de mistura entre os países analisados segundo as estimativas de contribuição africana, ameríndia e europeia. As cores mais escuras indicam maior contribuição e as mais claras, menor contribuição da parental em questão.

Para a população do México foi realizada ainda a estimativa de mistura considerando como parentais o europeu, o africano e duas populações nativas da América do Norte (Apache e Navajo – dois grupos Na-Dené), devido à localização geográfica do mesmo e à proximidade com essas duas populações. O resultado obtido foi de $50\% \pm 0,02$ para a parental europeia, $12,4\% \pm 0,01$ para a parental africana e $37,6\% \pm 0,02$ para a parental Na-Dené.

A tabela de mistura e o mapa da América Latina e Caribe com as variações quanto à contribuição das parentais mostram que a maior contribuição africana foi observada nos países situados na região do Caribe: Jamaica e Bahamas. El Salvador, Peru, Equador e México foram os países com maior contribuição ameríndia. Com relação à contribuição europeia, os maiores valores foram observados no Brasil, Porto Rico, Venezuela, Argentina e Chile.

Tabela 1.1. Análise de variância molecular (AMOVA) para 13 populações latino-americanas - Argentina, Bahamas, Brasil, Chile, Costa Rica, Colômbia, El Salvador, Equador, Jamaica, México, Peru, Porto Rico e Venezuela – utilizando 13 marcadores genéticos do tipo STR do CODIS.

Marcadores	Componentes da variância		Porcentagem de variação		<i>F_{st}</i>	Diversidade Gênica (média ponderada)
	Va	Vb	Interpopulacional	Intrapopulacional		
CSF1PO	0,0006	0,3693	0,17	99,83	0,002	0,74
D3S1358	0,0019	0,3780	0,51	99,49	0,005	0,76
D5S818	0,0038	0,3649	1,03	98,97	0,010	0,73
D7S820	0,0017	0,3972	0,44	99,56	0,004	0,79
D8S1179	0,0008	0,4019	0,21	99,79	0,002	0,80
D13S317	0,0028	0,4023	0,69	99,31	0,007	0,81
D16S539	0,0010	0,3980	0,26	99,74	0,003	0,80
D18S51	0,0016	0,4400	0,35	99,65	0,004	0,88
D21S11	0,0018	0,4232	0,42	99,58	0,004	0,85
FGA	0,0012	0,4392	0,27	99,73	0,003	0,88
TH01	0,0052	0,3918	1,32	98,68	0,013	0,78
TPOX	0,0018	0,3470	0,50	99,50	0,005	0,69
vWA	0,0015	0,3950	0,37	99,63	0,004	0,79

Tabela 1.2. Porcentagem de *loci* STR cujas frequências gênicas apresentaram valores de p significativos para F_{st} na análise de variância molecular ($p < 0,05$) entre os pares de populações.

	Argentina	Bahamas	Brasil	Chile	Colômbia	Costa Rica	El Salvador	Equador	Jamaica	México	Peru	Porto Rico
Argentina	-											
Bahamas	0,923	-										
Brasil	0,923	0,615	-									
Chile	0,077	0,538	0,538	-								
Colômbia	0,846	0,846	1,000	0,231	-							
Costa Rica	0,538	0,846	1,000	0,231	0,538	-						
El Salvador	0,615	1,000	0,846	0,308	0,462	0,538	-					
Equador	0,615	0,923	0,923	0,231	0,692	0,692	0,231	-				
Jamaica	0,923	0,000	0,846	0,692	1,000	1,000	1,000	0,923	-			
México	0,308	0,846	0,846	0,231	0,308	0,385	0,231	0,385	0,846	-		
Peru	0,385	0,846	0,769	0,231	0,385	0,462	0,077	0,077	0,846	0,231	-	
Porto Rico	0,538	0,538	0,154	0,846	0,538	0,692	0,769	0,769	0,769	0,462	0,692	-
Venezuela	0,385	0,692	0,385	0,385	0,462	0,538	0,692	0,769	0,769	0,308	0,615	0,154

Tabela 1.3. Estimativas de contribuição dos três principais grupos parentais na composição genética atual de treze países da América Latina e Caribe utilizando 13 marcadores microssatélites do sistema CODIS.

Localização geográfica	Populações parentais		
	Africana	Ameríndia	Européia
América do Sul			
Argentina	0,090 ± 0,028	0,308 ± 0,051	0,601 ± 0,179
Brasil	0,248 ± 0,050	0,093 ± 0,010	0,659 ± 0,089
Chile	0,063 ± 0,017	0,421 ± 0,017	0,516 ± 0,027
Colômbia	0,203 ± 0,063	0,338 ± 0,057	0,459 ± 0,081
Equador	0,044 ± 0,005	0,646 ± 0,006	0,310 ± 0,007
Peru	0,119 ± 0,012	0,730 ± 0,006	0,151 ± 0,015
Venezuela	0,163 ± 0,031	0,230 ± 0,015	0,606 ± 0,040
América Central			
Bahamas	0,852 ± 0,023	0,080 ± 0,021	0,068 ± 0,039
Costa Rica	0,122 ± 0,012	0,497 ± 0,038	0,381 ± 0,039
El Salvador	0,097 ± 0,019	0,752 ± 0,018	0,151 ± 0,032
Jamaica	0,903 ± 0,010	0,057 ± 0,011	0,041 ± 0,001
Porto Rico	0,264 ± 0,010	0,132 ± 0,028	0,603 ± 0,026
América do Norte			
México	0,101 ± 0,027	0,601 ± 0,021	0,296 ± 0,023

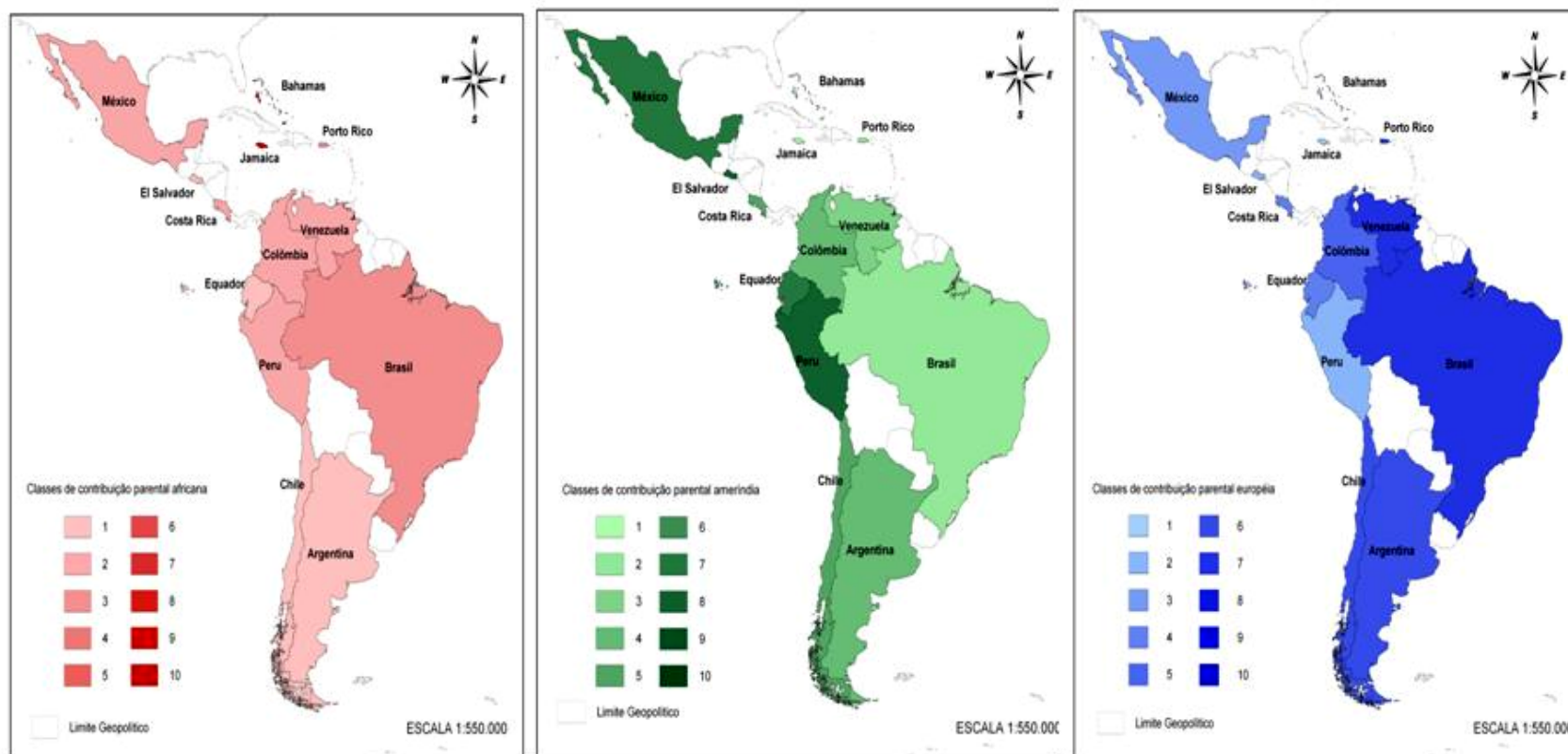


Figura 1.1. Mapa da América Latina e Caribe mostrando as variações étnicas entre os países conforme as estimativas de contribuição parental africana (vermelho), ameríndia (verde) e europeia (azul). As legendas indicam a proporção de mistura variando de 1 a 10, sendo que até 1 indica que a população possui até dez por cento de contribuição da respectiva parental e assim sucessivamente.

A matriz de distância genética de Reynolds *et al.* (1983), entre as 13 populações latino-americanas e caribenhas miscigenadas e as populações três parentais, está apresentada na Figura 1.2. Como esperado, as distâncias genéticas entre as populações parentais foi maior do que os valores de distância entre as populações miscigenadas. Exceções foram observadas nos casos em que as populações miscigenadas apresentavam grande diferencial no percentual de contribuição para as populações parentais, como é o caso de Jamaica, com alta contribuição africana, com relação às populações de Peru (0,044) e Equador (0,039), que apresentaram alta contribuição ameríndia.

Quando se comparam as populações miscigenadas com as parentais, observou-se uma maior distância genética entre estas e as parentais que menos contribuíram para a formação das mesmas e menor distância com as que mais contribuiriam. Por exemplo, verificou-se uma maior distância entre as populações caribenhas (Bahamas e Jamaica) e a parental ameríndia. Entretanto, essas duas populações miscigenadas apresentaram distâncias menores quando se compara com a parental africana. As populações de Chile, Peru, Equador, Colômbia, Argentina, México e El Salvador apresentaram as menores distâncias em relação à parental ameríndia, enquanto que o Brasil, Venezuela e Porto Rico apresentaram as menores distâncias com relação à parental européia. De um modo geral, na comparação entre os grupos parentais e as populações miscigenadas, as menores distâncias observadas ocorreram com a parental européia. Quando se comparam apenas as populações miscigenadas entre si, as menores distâncias observadas foram com o Brasil e a Venezuela e com a Colômbia e a Argentina e Costa Rica.

A Figura 1.3 mostra o diagrama da primeira e segunda coordenadas principais, que responderam por cerca de 83% da variação, sendo que a primeira coordenada representou 64% e a segunda 18% do total da variabilidade entre as populações. O eixo X mostrou a separação entre ameríndio e as outras parentais. A segunda coordenada principal separou o agrupamento europeu da parental africana, ficando assim localizado no quadrante superior à direita. Os dados mostrados na matriz de distância genética e no diagrama acima corroboraram os resultados gerados na análise de mistura. Todas as populações miscigenadas apresentaram valores de distância genética intermediários às três parentais. As populações com alto percentual africano (Jamaica e Bahamas) ficaram próximos da parental africana, localizados no quadrante inferior à direita. As com maior contribuição ameríndia apresentaram menor distância em relação a essa parental, ou seja, todas se localizaram nos dois quadrantes da esquerda. As populações mais próximas da parental européia foram Porto Rico, Brasil e Venezuela.

	EUR	AFR	AMC	AMS	VEZ	PER	CHI	EQU	BAH	JAM	COL	BRA	ARG	MEX	COS	SAL	PUE
EUR	-																
AFR	0,0255	-															
AMC	0,0503	0,0674	-														
AMS	0,0276	0,0503	0,0144	-													
VEZ	0,0047	0,0211	0,0299	0,0144	-												
PER	0,0327	0,0460	0,0097	0,0122	0,0171	-											
CHI	0,0188	0,0388	0,0276	0,0151	0,0111	0,0184	-										
EQU	0,0204	0,0415	0,0125	0,0078	0,0097	0,0068	0,0132	-									
BAH	0,0196	0,0023	0,0563	0,0398	0,0148	0,0367	0,0305	0,0324	-								
JAM	0,0239	0,0025	0,0671	0,0473	0,0199	0,0444	0,0372	0,0395	0,0032	-							
COL	0,0097	0,0250	0,0220	0,0087	0,0021	0,0120	0,0098	0,0057	0,0179	0,0234	-						
BRA	0,0028	0,0157	0,0392	0,0215	0,0017	0,0229	0,0149	0,0147	0,0106	0,0150	0,0049	-					
ARG	0,0086	0,0299	0,0202	0,0096	0,0023	0,0110	0,0086	0,0053	0,0216	0,0289	0,0020	0,0052	-				
MEX	0,0133	0,0313	0,0195	0,0100	0,0050	0,0097	0,0113	0,0064	0,0238	0,0304	0,0034	0,0085	0,0035	-			
COS	0,0092	0,0286	0,0240	0,0102	0,0029	0,0156	0,0091	0,0078	0,0213	0,0271	0,0019	0,0056	0,0028	0,0051	-		
SAL	0,0183	0,0376	0,0150	0,0036	0,0078	0,0094	0,0111	0,0041	0,0283	0,0350	0,0032	0,0126	0,0046	0,0047	0,0048	-	
PUE	0,0058	0,0182	0,0419	0,0261	0,0049	0,0270	0,0206	0,0193	0,0134	0,0180	0,0085	0,0034	0,0082	0,0121	0,0084	0,0161	-

Figura 1.2. Matriz de distância genética (Reynolds *et al*, 1983), entre treze populações miscigenadas latino-americanas e as três parentais com referência aos 13 marcadores STRs do CODIS. Legenda: EUR (europeu), AFR (africano), AMC (ameríndio da América Central), AMS (ameríndio da América do Sul), VEZ (Venezuela), PER (Peru), CHI (Chile), EQU (Equador), BAH (Bahamas), JAM (Jamaica), COL (Colômbia), BRA (Brasil), ARG (Argentina), MEX (México), COS (Costa Rica), SAL (El Salvador), PUE (Porto Rico).

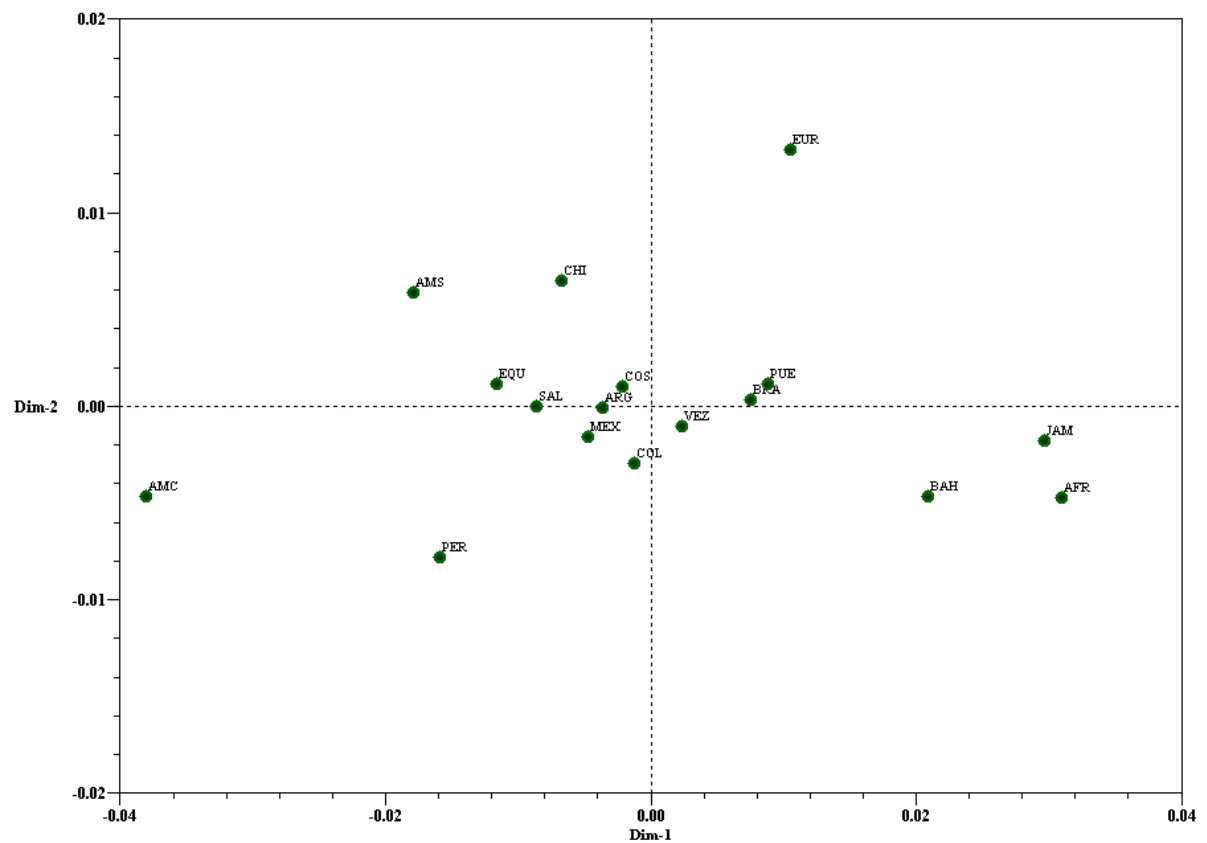


Figura 1.3. Diagrama das duas primeiras coordenadas principais, relativo à distância genética entre os 13 países latino-americanos e caribenhos e suas populações parentais (europeu, africano e ameríndio), baseado na análise de 13 marcadores STRs autossômicos. AFR (africano), AMEC (ameríndio da América Central), AMES (ameríndio da América do Sul), ARG (Argentina), BAH (Bahamas), BRA (Brasil), CHI (Chile), COL (Colômbia), COS (Costa Rica), EQU (Equador), EUR (europeu), JAM (Jamaica), MEX (México), PER (Peru), PUE (Porto Rico), SAL (El Salvador), VEN (Venezuela).

1.4. DISCUSSÃO

Diferenciação populacional na América Latina

Há uma grande semelhança na história geral dos países latino-americanos, em especial com relação à colonização, a escravidão e a origem geográfica dos colonizadores e escravos (Sans, 2000). Por outro lado, em decorrência de diversos fatores, dentre os quais o posicionamento geográfico e econômico dentro do continente americano, há uma heterogeneidade entre as regiões do continente (Wang *et al.*, 2008; Choudhry *et al.*, 2006; Salzano, 2004). No presente trabalho foi observado que há diferenças entre os países da América latina quando são considerados 13 marcadores STR do CODIS.

As diferenças entre os países não são reflexo da diversidade gênica dos marcadores analisados, pois esta foi alta em todos os países para todos os marcadores. A diversidade gênica alta foi decorrente do alto número de alelos que é uma característica do tipo de marcador empregado. Essa característica é um dos parâmetros que levou o CODIS a utilizar tais marcadores para análises forenses, ou seja, o alto polimorfismo principalmente em populações européias (Oliveira *et al.*, 2006; Dixon *et al.*, 2005; Destro-Bisol *et al.*, 2000). A diversidade gênica avalia a heterozigose que é dependente do número de alelos e sua frequência nas populações (Nei, 1987). Como o número de alelos variou pouco entre as populações, a única fonte de variação ficou restrita à frequência dos alelos.

A subestruturação populacional entre os países da América Latina e Caribe foi confirmada pela análise de variância molecular – AMOVA – que indicou que os valores de *Fst* estimados para os pares de populações foram estatisticamente significativos em várias combinações de populações. Como em outras análises envolvendo a espécie humana (Manica *et al.*, 2005; Jorde & Wooding, 2004; Comas *et al.*, 2000; Destro-Bisol *et al.*, 2000), o resultado mostrou que a diferença entre as populações pode ser melhor explicado pelas diferenças intrapopulacionais do que interpopulacionais. No presente trabalho a variação intrapopulacional estimada variou de 99,83% a 98,68. A maior variação intrapopulacional é um dos parâmetros utilizados na sustentação da inexistência de raças humanas (Barbujani, 2005; Bamshad *et al.*, 2004; Oliveira & Ferreira, 2004).

A subestruturação populacional foi também confirmada pela análise de distância genética. Observou-se que as três populações parentais ficaram perfeitamente bem diferenciadas, enquanto que as populações latino-americanas e do Caribe apresentaram uma posição no diagrama, diretamente relacionada à população parental que participou com a maior contribuição na sua constituição genética atual. De um modo geral, a análise

de coordenada principal foi concordante com dados antropológicos e históricos, assim como com as estimativas de mistura, isto é, o posicionamento das populações no diagrama apresentou-se relacionado à porcentagem de contribuição das populações parentais.

Mistura genética

A América Latina tem sido palco de grandes transformações populacionais, sendo hoje caracterizada por ter uma constituição multiétnica em decorrência da influência dos mais distintos povos ao longo da história. Por outro lado, a maior contribuição genética adveio de três populações parentais principais - europeu, africano e ameríndio.

A população americana pré-colombiana sofreu mudanças significativas na sua composição inicial. Paralelamente à diminuição numérica expressiva, relacionada principalmente ao processo da conquista, houve também à diminuição numérica importante relativa à perda da identidade étnica em decorrência da absorção destes por parte dos colonizadores. Entretanto, em algumas regiões a população nativa subsistiu e tem forte influência na formação da população. Isto pode ser observado em especial nas regiões onde o acesso foi mais difícil e/ou eram regiões inóspitas a época da colonização. Países como El Salvador, Chile, Costa Rica, Colômbia, Equador e Peru, que apresentavam florestas fechadas e/ou regiões montanhosas, apresentaram estimativas expressivas de contribuição ameríndia. Por outro lado, no Brasil, em algumas regiões da Argentina e da Colômbia, na Venezuela e em Porto Rico, a contribuição parental ameríndia é menor, em decorrência da extensiva migração europeia que houve para esses países.

Pode-se dizer que, em termos de contribuição populacional, a colonização europeia na América Latina obteve êxito no sentido de que as estimativas aqui obtidas revelam que sua presença é marcante em quase todos os territórios. Em determinados países a contribuição se sobrepôs às demais, com foi o caso do Brasil, Venezuela, Porto Rico e Argentina (65,9%, 61,5%, 61,3% e 60,1%, respectivamente). Essa sobreposição numérica tem relação com o número efetivo de migrantes e com o fato de que os homens europeus se reproduziram com mulheres de todas as etnias, como já demonstrado em outros estudos (Martínez *et al.*, 2007; González-Andrade *et al.* 2007; Bedoya *et al.*, 2006; Campos-Sánchez *et al.*, 2006; Alves-Silva *et al.*, 2000), o que teve como consequência a geração de um maior número de descendentes.

A contribuição africana é maior em regiões onde houve maior exploração econômica por parte dos europeus, sendo, portanto lugares de mais fácil acesso, isto é, pela

costa do Atlântico já que a rota de navegação naquela época entre a Europa e África com a América se ocorria no Oceano Atlântico. As maiores contribuições dessa parental foram encontradas nos países do Caribe, Jamaica (com mais de 90%) e Bahamas (quase 85%). Na América do Sul, países com acesso mais fácil à costa do Oceano Atlântico apresentaram estimativas maiores de contribuição africana na composição genética atual, como Brasil e Venezuela. Exceção foi a Argentina que, apesar de ser um país da costa Atlântica, apresenta apenas 9% de contribuição africana. Por outro lado, países em que o acesso se dá pela costa do pacífico, como Chile, Equador e Peru, apresentaram baixa contribuição africana. Estimativas de contribuição dessa parental para a população do Equador foram também obtidas por González-Andrade *et al.* (2007), os quais obtiveram resultado similar ao aqui apresentado. Essa observação reflete o fato histórico de que durante o tráfico de escravos, em especial no século XVI, os escravos africanos com destino à costa Pacífica vinham da costa Atlântica, passando pelo atual território colombiano e depois eram levados para o Peru de onde eram distribuídos para as regiões vizinhas, (Klein, 1986).

Estimativas de mistura genética calculadas anteriormente para a Argentina mostraram resultados similares aos encontrados neste trabalho. A contribuição africana foi estimada em 7% e ameríndia em torno de 30% na constituição da população de La Plata, região próxima a capital Buenos Aires, utilizando dados de marcadores genéticos autossômicos clássicos (Lopes-Camelo *et al.*, 1996). A avaliação dessa mesma população utilizando cinco marcadores do tipo AIMS confirmaram a baixa contribuição africana (6,5%), alta europeia (67,5%) e ameríndia intermediária (30%) (Martínez-Marignac *et al.*, 2004).

No Brasil, o processo de mistura foi complexo devido à sua extensão territorial e, ao contrário da Argentina, a contribuição parental africana foi significativa. A contribuição europeia aqui estimada foi concordante com a descrita em outros trabalhos (Ferreira *et al.*, 2006; Callegari-Jacques *et al.*, 2003) e é marcante em todas as regiões do Brasil, especialmente na região sul, sendo a contribuição ameríndia e africana variável.

Na Colômbia, há uma diversidade na constituição genética entre as regiões. Na Costa Colombiana, por exemplo, predomina a contribuição europeia e africana (capítulo 2), mas nos Andes a presença ameríndia é marcante. Na Antioquia, a contribuição europeia é elevada, acima de 69% (Bedoya *et al.*, 2006) quando se considera marcadores do cromossomo X, ao passo que na região de Chocó a contribuição europeia é baixa. Se considerarmos todas as regiões simultaneamente o que se verifica são valores intermediários para as três parentais.

Já na Venezuela, estudos mostram que a heterogeneidade genética é verificada com relação às diferentes regiões geográficas (Simmons *et al.*, 2007; Loyo *et al.*, 2004), e também entre diferentes grupos sociais (Martínez *et al.*, 2007), sendo as contribuições ameríndias e africanas maiores nas classes mais baixas do que nas classes mais ricas. Os dados do presente estudo foram similares aos estudos já realizados, mostrando contribuição europeia mais alta do que as outras parentais. Com relação ao Chile, os resultados encontrados no presente trabalho foram similares ao observado em trabalho utilizando marcadores genéticos clássicos (grupos sanguíneos, enzimas eritrocitárias, proteínas do plasma e marcadores do sistema HLA), onde foi estimada uma contribuição ameríndia que variou de 47 a 59% e uma africana baixa (Acuña *et al.*, 2000).

A escolha das parentais, como dito anteriormente, tem um reflexo direto nos resultados. Isso pode ser exemplificado com os resultados obtidos por Zúñiga *et al.* (2006) para Porto Rico, que mostraram diferenças para os resultados aqui relatados, apesar de tratar-se fundamentalmente do mesmo banco de dados, isto é, eles utilizaram 15 marcadores STRs, incluindo os 13 marcadores analisados neste estudo. A maior diferença, de fato, concentrou-se na escolha das parentais. Zúñiga *et al.* (2006) selecionaram como parental europeia um banco de dados de Andaluzia (Sul da Espanha), como africana um banco do oeste da África subsaariana (Guiné-Bissau) e como ameríndia dados de ameríndios do México (Estado de Hidalgo). Os resultados indicaram contribuição europeia alta (superior a 76%), africana intermediária (em torno de 17%) e ameríndia baixa (cerca de 6%). No presente estudo foi também observada contribuição europeia alta, porém com valores bem inferiores (61%), enquanto que a contribuição estimada da parental ameríndia foi o dobro do obtido por Zúñiga *et al.* (2006). Essa discrepância sugere que a população ameríndia que foi utilizada provavelmente apresenta miscigenação com europeu, o que pode ter superestimado a contribuição europeia obtida por aqueles autores.

Com relação à população mexicana, as estimativas para a região noroeste do México mostraram contribuição significativa da parental europeia e da parental ameríndia, com baixa contribuição africana (Cerdeira-Flores *et al.*, 2002). No presente estudo observou-se uma contribuição parental ameríndia maior que a europeia. Entretanto, nas análises de mistura do presente estudo foram incluídos dados de frequências alélicas de outras regiões mexicanas, isto é, da região central (Hernández-Gutiérrez *et al.*, 2005) e cidade do México (Luna-Vasquez *et al.*, 2003). Pode-se sugerir que exista heterogeneidade genética nas diferentes regiões do México, assim como que tais diferenças possam ser decorrentes da escolha das populações parentais.

Similarmente, trabalho anterior com a população da Costa Rica também apresentou resultados diferentes aos encontrados nesse estudo. Os resultados de análises de mistura utilizando 11 marcadores genéticos clássicos (ABO, RH, MNS, KEL, FY, jK, P, LE, SE, DI e HP) mostraram contribuição européia de mais de 60% na constituição étnica da população (Morera *et al.*, 2003), enquanto que os dados aqui obtidos indicaram que a parental ameríndia apresenta a maior contribuição.

Os resultados de estimativa de mistura obtidos para a região caribenha com os 13 marcadores STRs foi similar ao encontrado por Benn-Torres *et al.* (2007) com a utilização de 28 marcadores AIMs. Há uma concordância de que a contribuição africana é muito alta e as ameríndia e européia muito baixas.

É possível sugerir que os aspectos geográficos e climáticos influenciaram na europeização da América. Extensas barreiras naturais, como as regiões semi-áridas e os desertos do México, do noroeste da Argentina e o Atacama no Chile, a Cordilheira dos Andes e a floresta Amazônica aparentemente formaram barreiras que podem ter desviado o curso das migrações. Essas regiões são bastante inóspitas o que tornou as regiões interioranas de difícil acesso. Ainda assim, a América Latina sofreu forte transformação étnica, com grau de amalgamação intenso e ininterrupto que reflete na formação da população, possibilitando a existência de diversas combinações. Assim, os afro-americanos estão mais presentes no Caribe e ao longo da costa atlântica até o Brasil. Os ameríndios são mais numerosos na região da Cordilheira, e os europeus nas regiões de clima mais ameno ou menos inóspito como Argentina, sul do Brasil, Chile, Venezuela e Porto Rico.

Quando se analisa grandes regiões geográficas pode-se entender melhor todo o processo de transformação que as populações sofreram devido ao intenso fluxo gênico contínuo, como ocorreu na América nestes últimos 500 anos. Entretanto, as mudanças microrregionais são importantes para se compreender o todo, já que existem variações étnicas ao longo do continente. Estudando o processo de construção histórica regional pode-se fazer inferências de como ocorreu o processo de assimilação entre os diferentes grupos étnicos. Nos próximos capítulos serão discutidos dois modelos distintos de colonização: primeiro, o processo de mistura étnica em três países considerando-se as diferenças regionais dos mesmos e, segundo, será discutido um modelo regional de colonização, em que populações da região centro-oeste brasileira serão o objeto de estudo. O objetivo é mostrar que, assim como ocorreu com toda a América Latina, a região centro-oeste foi palco de profundas transformações étnicas.

CAPÍTULO 2

PADRÕES REGIONAIS DE MISTURA GENÉTICA NA AMÉRICA DO SUL

2.1. INTRODUÇÃO

As sucessivas migrações ocorridas para a América do Sul levaram à formação de grupos miscigenados e a um quadro populacional diversificado. É notável a heterogeneidade populacional decorrente dessas migrações e do processo de amalgamento que se deu em virtude dos contatos entre os diferentes grupos étnicos a partir do início da colonização da América pelos europeus. Embora o início da colonização tenha sido similar, posteriormente o povoamento se desenvolveu por diferentes caminhos nas diferentes regiões da América, incluindo diferenças regionais dentro de um mesmo país (Salzano, 2004). Exemplo disso foi o que ocorreu na Argentina, Brasil e Colômbia, países localizados na América do Sul.

Antes da ocupação pelos europeus, a partir do século XV, a América do Sul era ocupada exclusivamente pelos ameríndios, porém de diferentes grupos lingüísticos. Quando os espanhóis chegaram ao atual território colombiano eles se depararam com populações ameríndias divididas em diversos grupos lingüísticos, sendo que a maioria pertencia aos grupos Chibchan e Caribe (Melo, 1996). Os povos indígenas dessa região já conheciam a agricultura, sendo que os que habitavam as regiões litorâneas praticavam uma agricultura mais rudimentar, enquanto que aqueles que viviam nas regiões mais altas (Chibchan) praticavam a agricultura utilizando técnicas mais avançadas e intensivas, cultivando milho, batata e algodão.

Na Argentina, os espanhóis encontraram populações ameríndias agricultoras, na região noroeste, nas serras centrais e nas margens dos rios, e nômades, que viviam da caça e da coleta, localizadas principalmente nas regiões de pampa, nos chacos e na patagônia. A maioria pertencia ao grupo lingüístico Tupi-guarani e outros pequenos grupos como os Tehuelche (Salzano & Callegari-Jacques, 1988).

No Brasil, diferentemente dos demais países latino-americanos, o principal colonizador foi o português. Quando os colonizadores chegaram ao território, encontraram na costa atlântica grupos ameríndios que pertenciam a diversos grupos lingüísticos especialmente Tupi-guarani. Ao adentrarem o país encontraram diversos grupos indígenas, dentre os quais os do grupo Jê, habitantes da região mais central do Brasil. A maior parte dos indígenas brasileiros vivia da caça, coleta e da pesca, sendo que vários conheciam a agricultura, cultivando hortaliças e tubérculos como a mandioca (Vainfas, 2000; Salzano & Callegari-Jacques, 1988).

Paralelamente a essa diferença inicial entre os três países em decorrência dos diferentes grupos lingüísticos que habitavam a região antes do início da colonização, dois outros fatores contribuíram de maneira fundamental na definição da constituição genética populacional desses países. Primeiro foi a forma com que se deu a conquista desses territórios pelos europeus e segundo foi o tráfico de escravos e a distribuição dos africanos nos mesmos.

Quanto à colonização pelos europeus, a elaboração de um tratado em 1494 estabelecendo a divisão das terras até então descobertas e aquelas por descobrir entre Portugal e Espanha, o Tratado de Tordesilhas, definiu que Portugal teria direito de ocupar as terras localizadas a leste da linha de 370 léguas traçada a partir de Açores e Cabo Verde, enquanto que a Espanha ocuparia todas as terras que ficassem do lado ocidental dessa linha. Por isso, Brasil foi colonizado pelos portugueses, enquanto que o restante da América seria de direito da Espanha (Herman, 2000).

A Colômbia foi o primeiro país da América do Sul a ser dominado pela Espanha. Quando Cristóvão Colombo chegou à América em 1492, os espanhóis se fixaram primeiro nas ilhas do mar do Caribe. Somente em 1498, em sua terceira viagem à América, Colombo alcançou a atual costa venezuelana. Entre 1501 e 1535 vários contatos já haviam sido estabelecidos com os ameríndios da costa atlântica da Colômbia e das savanas adjacentes. A ocupação espanhola se estendeu para o interior do território entre 1535 a 1560, especialmente nos Vales del Cauca e Madalena e dos altiplanos ao sul e oriente. O estabelecimento dos espanhóis no território foi rápido, sendo que nas primeiras décadas de colonização houve a fundação de várias cidades como Santa Marta (1526), Cartagena (1533) e Santa Fé de Bogotá (1539), atual capital Bogotá. A Colômbia ocupava uma posição estratégica, pois era a porta de acesso para as demais regiões a oeste e ao sul da América do Sul, onde se localizavam os impérios ameríndios mais desenvolvidos, Astecas e Incas (Melo, 1996).

A chegada dos portugueses à costa atlântica do Brasil se deu em 1500. Como o território e a costa marítima eram vastos, a coroa portuguesa criou o regime de capitânicas hereditárias em 1532 com o objetivo de explorar e ocupar o território, propiciando, assim, a criação dos primeiros núcleos de colonização portuguesa no Brasil. Diferente do que ocorreu com as colônias espanholas, não houve planejamento de cidades e a ocupação ocorreu apenas na costa atlântica nas primeiras décadas. A união entre Espanha e Portugal entre 1580 e 1640 permitiu a expansão do território brasileiro além dos limites fixados pelo Tratado de Tordesilhas, época em que ocorreram expedições, conhecidas como Entradas e

Bandeiras, que promoveram a ocupação das terras a oeste, além das terras localizadas ao norte do país e ao sul, na região platina. No século XVIII muitas cidades já haviam sido estabelecidas, tanto na costa marítima (Fortaleza, São Luiz, Natal, Recife, Olinda, Ilhéus, Vitória, Rio de Janeiro entre outras) quanto no interior do território (Mariana, Vila Rica, Goiás, Cuiabá, entre outras) (Bertran, 2000; Herman, 2000; Palacín & Moraes, 1994).

Os portugueses foram os europeus que mais contribuíram para a colonização do Brasil. Estima-se que em um primeiro momento aproximadamente 700 mil portugueses tenham migrado para o Brasil. A partir de 1700 o fluxo migratório se tornou mais intenso em virtude da descoberta do ouro, sendo que entre os períodos de 1880 a 1960 houve migração em massa, com um número de imigrantes estimado em mais de 1,5 milhão de indivíduos. Contribuíram ainda para povoar o Brasil, principalmente a partir de 1890, os italianos, que formaram o segundo maior contingente de migrantes europeus, seguido pelos espanhóis, que iniciaram a migração para o Brasil em épocas mais remotas desde o século XVI (Venâncio, 2000) e os japoneses que migraram principalmente para a região sudeste.

O processo de colonização da Argentina ocorreu a partir do século XVI. Pelo fato de haver poucos grupos camponeses indígenas estabelecidos e, inicialmente, poucos recursos minerais disponíveis ao sul do rio da Prata, até 1520 a coroa espanhola estava mais voltada para a conquista do rico império inca localizado na região da Cordilheira dos Andes, na América do Sul. Somente em 1527 se deu a construção do primeiro povoado da Argentina, chamado forte de *Sancti Spiritus*. O povoamento do território se deu graças às concessões para explorar a região fornecida pela coroa espanhola a particulares (Rock, 1985).

O atual território argentino fazia parte da vasta região de colonização da Espanha que abrangia também, por direito concedido pelo tratado de Tordesilhas, a Bolívia, Paraguai, Uruguai, Chile e parte do Brasil. Assim, até o início do século XVII os espanhóis fundaram alguns assentamentos que acabaram por originar as principais cidades argentinas, como Buenos Aires (1536), Santa Fé e Córdoba (1573), Catamarca (1603), entre outras. A imigração européia para a Argentina foi contínua, o que contribuiu com a política de “embranquecimento” da população, sendo que os migrantes eram originários principalmente da Espanha e da Itália. A colonização pelos europeus foi intensa até o início do século XX, época da segunda guerra mundial (Rock, 1985).

Quanto à contribuição africana, a vinda para a Argentina, Brasil e Colômbia se deu em diferentes épocas e proporções. Todos esses três países apresentam costa marítima atlântica, o que facilitou a entrada de africanos, decorrente do tráfico de escravos pelos

européus, a partir do século XVI. A origem dos africanos foi semelhante, isto é, eram de povos que viviam na África subsaariana, principalmente do Congo e Angola e, em menor número, da Guiné, Gana, Nigéria, Serra Leoa e Moçambique. Pertenciam a diferentes grupos lingüísticos, como bantus e sudaneses. O Brasil foi a colônia que mais utilizou a mão-de-obra escrava. Dados estatísticos mostram que cerca de 40% de todos os africanos trazidos para a América vieram para o Brasil (Klein, 1986). Enquanto que mais de quatro milhões de escravos entraram no Brasil entre 1451 e 1870, nesse mesmo período a América espanhola recebeu menos da metade deste contingente. Isso foi teoricamente decorrente do menor êxito português na utilização de mão-de-obra indígena quando comparado com os espanhóis (Klein, 2000).

A Colômbia foi uma rota importante para a entrada de escravos africanos na América do Sul. Antes da chegada dos escravos até o Peru e sua distribuição a outras regiões da colônia espanhola, o porto de Cartagena na costa caribenha era o primeiro local de desembarque dos escravos. Os escravos que ficavam na Colômbia eram utilizados principalmente na costa caribenha e, posteriormente, na região oeste ao norte do pacífico, que compreende o atual estado de Chocó. A descoberta de ouro no século XVIII nesta região incrementou o fluxo de escravos. Embora o local fosse de difícil acesso, Chocó se tornou o principal centro de extração de ouro nessa época e a população afro-descendente nesse local cresceu rapidamente: de dois mil escravos mineiros em 1720 a mais de sete mil em 1782. Esse número respondia por mais de 13% da população escrava total de todo o vice-reinado de Nova Granada, onde hoje se localiza as terras baixas da costa nordeste da Colômbia (Klein, 1986). Portanto, nessa época a população escrava total na Colômbia era em torno de 54 mil indivíduos.

Na Argentina houve inicialmente pouco interesse na mão-de-obra escrava, pois esta era uma região pouco explorada devido à falta de metais preciosos exportáveis disponíveis de imediato. A demanda se deu quando da implantação de estâncias de gado no Rio da Prata e da descoberta de prata no alto Peru em 1545. O porto de entrada dos escravos foi a cidade de Buenos Aires, sendo que a maioria deles era proveniente de Angola. Porém os números são imprecisos em decorrência de uma ampla perda de documentos. Estima-se que entre 1600 até 1806 tenham entrado na Argentina pouco mais de 52 mil escravos (Rock, 1985).

O avanço de povos europeus na América do Sul provocou mudanças significativas nas populações locais. À medida que os europeus avançaram no processo de conquista, trazendo junto consigo escravos africanos, os povos ameríndios que se encontravam na

costa atlântica foram sendo forçados a migrar para o interior do continente, isto quando conseguiam escapar da morte provocada pelas guerras ou pelas doenças trazidas pelos europeus (Alencastro, 2000). Isso provocou a diminuição do número de ameríndios onde a presença européia ocorria. Uma questão então pode ser levantada: quais foram as conseqüências em termos de constituição genética populacional do predomínio europeu em determinadas regiões? Assim, buscou-se responder às seguintes indagações: Argentina, Brasil e Colômbia, devido à similaridade nas suas histórias, apresentam hoje o mesmo padrão populacional? A participação dos grupos parentais presentes durante a colonização da América Latina ocorreu de forma homogênea nos três países analisados?

Este capítulo teve por objetivo mostrar que o processo de colonização e mistura em três países da América do Sul – Argentina, Brasil e Colômbia - ocorreu de forma heterogênea. Os instrumentos para tal estudo foram marcadores genéticos do tipo microsatélites, pois seu alto grau de polimorfismo permite observar as variações entre indivíduos numa dada população e entre as populações.

2.2. MATERIAL E MÉTODOS

Fonte de Dados

Argentina, Brasil e Colômbia são três países da América do Sul que apresentaram semelhanças quanto ao seu povoamento. O Brasil é dividido em cinco regiões geográficas: norte, nordeste, centro-oeste, sudeste e sul. A Argentina pode ser dividida em seis regiões geográficas: Pampa (região centro-oriental), Mesopotâmia (na bacia hidrográfica formada pelos rios Uruguai e Paraná), Gran Chaco (terras do Chaco argentino), Noroeste (região montanhosa localizada a noroeste, próximo às divisas com Chile e Bolívia), Cuyo (região andina) e Patagônia (sul do país). A Colômbia é dividida em 32 estados, mas não apresenta divisão administrativa ou geográfica em macrorregiões como é o caso da Argentina e do Brasil. No presente trabalho foi adotado a divisão populacional em quatro principais regiões colombianas sugerida por Paredes *et al.* (2003) que são: Costa do Caribe, Costa do Pacífico Norte (afro-descendentes do Estado de Chocó) e Ilha Caribenha de San André, região a sudoeste dos Andes e região dos Andes, Amazônica e do Orinoco, as quais foram denominadas A, B, C e D, respectivamente, no presente trabalho. As divisões acima citadas foram utilizadas para as análises descritas a seguir e estão apresentadas na Figura 2.1.



Figura 2.1. Mapa da Argentina, Brasil e Colômbia e suas respectivas regiões geográficas adotadas para as análises desenvolvidas neste trabalho. A divisão adotada para a Argentina é referente às regiões geográficas do mesmo, para o Brasil é a divisão político-administrativa adotada pelo governo federal do país, enquanto que a Colômbia foi dividida de acordo com o proposto por Paredes *et al.* (2003).

Para as análises populacionais relativas às regiões dos três países, foram realizados levantamentos da literatura referentes às frequências gênicas de treze marcadores STRs do sistema CODIS (*Combined DNA Index System*) - CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX e vWA (Tabela 1, Anexo 1). Os dados foram compilados de acordo com as regiões geográficas de cada país e as frequências gênicas foram estimadas pelas médias ponderadas dos dados obtidos para cada região. A região de Cuyo (Argentina) não foi considerada devido à ausência de dados na literatura com relação aos marcadores analisados.

Análises populacionais

A mistura gênica foi estudada com base nas diferenças e similaridades das frequências gênicas dos marcadores genéticos entre as populações de interesse e suas parentais (grupos étnicos parentais), que no caso aqui apresentado são europeus, africanos e ameríndios. Para estimar as frequências alélicas das parentais, foram calculadas as médias ponderadas a partir de bancos de dados de Portugal e da Espanha, para a parental europeia, da África subsaariana, para a africana, e de povos ameríndios da América do Sul para a ameríndia. Os dados de frequência foram obtidos no banco de dados da Universidade de Düsseldorf (<http://www.uni-duesseldorf.de/WWW/MedFak/Serology/Database.html>, acessado dia 03/05/2008), em Paredes *et al.* (2003a) e em Fernandes e Brehm (2002) para as populações de Portugal e da Espanha; no banco de dados ALFRED (*Allelic Frequency Database* - <http://alfred.med.yale.edu> – acessado em 03/05/2008) e em González-Andrade *et al.* (2006) para as populações ameríndias da América do Sul (*Kichwas, Cubeo, Desano, Puinave, Tehuelche, Tucano, Wichi e Curripaco*). Para a população parental africana, os dados foram obtidos no banco de dados da Universidade de Düsseldorf para as populações de Nigéria, Zimbábue, Benin, *Bamileke* e *Ewondo* de Camarões, *Ovambo* da Namíbia e uma amostra de Angola e Guiné Bissau - e em artigos científicos - Angola (Alves *et al.*, 2004), Guiné Bissau (Gonçalves *et al.*, 2002), Gabão (Steinlechner *et al.*, 2002), Guiné Equatorial (Alves *et al.*, 2005; Gene *et al.*, 2001) e Ruanda (Tofanelli *et al.*, 2003).

A análise de mistura foi realizada utilizando o programa ADMIX 3, baseado no método de Identidade Gênica (Chakraborty, 1985) entre as populações ancestrais e a miscigenada. Esse programa gera duas estimativas, a viesada e a não viesada. É considerada como melhor estimativa aquela que apresentar o maior *R-square* e nos casos em que o *R-square* é igual escolheu-se a estimativa não viesada. As estimativas de mistura

foram realizadas para cada região geográfica dos países e dos países como um todo. Para essa última análise, foram considerados os resultados de mistura obtidos para cada região, ponderado pela proporção de habitantes em cada região geográfica. Esta última análise foi realizada devido às diferenças do número de habitantes de cada região, bem como das diferenças na composição étnica entre as regiões. O desvio padrão foi calculado considerando a contribuição parental relativa em cada região utilizando para isso uma planilha Excel, sendo:

$$\sqrt{\frac{\sum (x - \bar{x})^2}{(n-1)}}$$

onde \bar{x} é média de amostra média (valores informados) e n é o tamanho da amostra.

As distâncias genéticas entre os pares de populações foram estimadas como proposto por Reynolds *et al.* (1983), utilizando o software PHYLIP versão 3.65 (Felsenstein, 1989). A matriz de distância obtida foi utilizada para realizar a Análise de Coordenada Principal (PCA) gerada no programa NTSYS-pc versão 1.70 (Rohlf, 1992).

2.3. RESULTADOS

As estimativas de mistura gênica obtidas para as regiões dos três países e nos países como um todo estão apresentadas na tabela 2.1. De um modo global, as maiores estimativas obtidas foram as de contribuição européia para os três países, sendo que o Brasil apresentou o maior valor. A maior contribuição africana foi estimada também para o Brasil, enquanto que a ameríndia foi alta e similar na Argentina e Colômbia (38%) e baixa no Brasil.

A maior homogeneidade entre as estimativas foi observada no Brasil, onde apenas a região sul se diferencia das demais, porém mantendo a alta contribuição européia e baixa africana e ameríndia. A região sul do país apresentou elevada contribuição européia (81,5), enquanto que as regiões norte e sudeste apresentaram contribuições menores (60,6% e 60,7% respectivamente). A contribuição africana, intermediária em quase todas as regiões, foi mais alta na região sudeste do país (32%), sendo que na região sul do país essa contribuição foi menor que 10%. Quanto à contribuição ameríndia, a região norte do país foi a que apresentou maior contribuição (18%), enquanto que a região sudeste apresentou menos da metade do observado na região norte deste país.

A Argentina possui uma população triíbrida, como pode ser observado na tabela 2.1. A contribuição africana estimada foi baixa, menor que 10%, quando considerada a análise de mistura ponderada pelo tamanho populacional de cada região. A contribuição parental ameríndia por sua vez foi três vezes superior à contribuição africana, ultrapassando 30%. A contribuição parental europeia foi superior às demais em todas as regiões geográficas, exceto para a região noroeste que apresentou menos de 25% de

Tabela 2.1. Estimativas de contribuição das populações parentais sul-americanas – europeu, africano e ameríndio – na constituição atual das regiões geográficas de Brasil, Colômbia e Argentina, utilizando 13 marcadores microsatélites do sistema CODIS.

Populações analisadas	Representatividade populacional (%)*	Grupos Parentais			
		Europeu	Africano	Ameríndio	
Argentina	Pampa	63,3	0,678±0,050	0,103±0,039	0,219±0,019
	Gran Chaco	6,3	0,531±0,018	0,089±0,014	0,379±0,007
	Mesopotamia	8,4	0,630±0,009	0,064±0,007	0,305±0,003
	Noroeste	10,1	0,228±0,014	0,085±0,013	0,686±0,007
	Patagônia	4,8	0,554±0,004	-	0,446±0,004
	Total	92,9***	0,601±0,179	0,090±0,028	0,308±0,051
Brasil	Centro-Oeste	7,0	0,663±0,019	0,217±0,015	0,120±0,008
	Norte	5,0	0,606±0,055	0,213±0,044	0,181±0,021
	Nordeste	29,0	0,667±0,045	0,234±0,035	0,100±0,017
	Sudeste	44,0	0,607±0,025	0,320±0,017	0,074±0,021
	Sul	15,0	0,815±0,009	0,093±0,007	0,092±0,003
	Total	100,0	0,659±0,089	0,248±0,050	0,093±0,010
Colômbia**	Região A	33,9	0,405±0,047	0,350±0,037	0,245±0,018
	Região B	1,2	0,116±0,016	0,730±0,013	0,154±0,006
	Região C	23,2	0,457±0,042	0,137±0,033	0,406±0,016
	Região D	41,7	0,549±0,023	0,036±0,018	0,415±0,009
	Total	100,0	0,459±0,081	0,203±0,063	0,338±0,057

*Fonte: National Geographic, 2008, para Argentina e Colômbia; contagem demográfica 2007 IBGE para Brasil. **Região A: Costa do Caribe; região B: Costa do Pacífico Norte (afro-descendentes do Estado de Chocó) e Ilha Caribenha de San André; Região C: região a sudoeste dos Andes; Região D: região dos Andes, Amazônica e do Orinoco (de acordo com Paredes *et al.*, 2003). ***A região de Cuyo não foi incluída na análise.

contribuição para essa parental. Entretanto, essa região foi a que apresentou maior contribuição para a parental ameríndia (68,6%), inclusive quando considerada as demais regiões de Brasil e Colômbia.

Com relação à Colômbia, quando se considerou a proporção de mistura para o país como um todo se observou uma maior contribuição européia na constituição atual seguida pela ameríndia e finalmente a africana. Mesmo representando em termos de percentual populacional 1,2% da população total, ao considerar essa região na análise de mistura o valor da parental africana foi de 20,3%. Esse resultado é reflexo da elevada contribuição africana na população de Chocó e Ilha de Santo André (73%). A região caribenha apresentou valores de mistura entre as três parentais mais homogêneos quando se compara todas as regiões geográficas dos três países, enquanto que as duas outras regiões (C e D), tiveram percentuais de contribuição européia e ameríndia superiores a 40%.

Na figura 2.2 estão apresentadas as estimativas de distâncias genéticas entre todas as populações das regiões geográficas dos três países analisados, bem como entre essas populações miscigenadas e as populações parentais. Os dados de distância genética corroboraram os resultados de mistura obtidos e ambos foram coerentes com a história de colonização das regiões nos três países analisados. Comparando-se as populações miscigenadas com as três parentais, foi observado que as menores distâncias genéticas ocorreram entre todas as populações e a parental européia, especialmente as regiões brasileiras e duas regiões argentinas, isto é, pampa e mesopotâmia. As maiores distâncias foram observadas, de um modo geral, entre as populações miscigenadas e a parental ameríndia, novamente destacando-se o Brasil além de nesse caso também duas regiões da Colômbia (Costa Caribenha e Chocó e Ilha de Santo André). Na comparação entre as regiões dos três países e a parental africana, duas regiões da Argentina (Patagônia e noroeste), bem como uma região da Colômbia (região D), apresentaram as maiores distâncias. De modo contrário, as menores distâncias foram observadas em duas regiões da Colômbia (regiões A e B) e valores intermediários nas regiões brasileiras.

Com relação à distância genética entre as regiões dos três países, os menores valores foram observados entre as regiões brasileiras. Destacaram-se as regiões noroeste da Argentina e Chocó e Ilha de Santo André da Colômbia, pois estas apresentaram as maiores distâncias genéticas quando comparadas às diferentes regiões entre os três países. Por exemplo, na comparação entre a região noroeste da Argentina e as regiões brasileiras os valores de distância genética foram mais altos do que quando as mesmas regiões foram

comparadas com outras duas regiões da Argentina (pampa e mesopotâmia). O mesmo se observou com relação a região B da Colômbia.

Na Figura 2.3 está apresentado o gráfico relativo a análise da primeira e segunda coordenadas principais obtidas a partir de dados de 13 marcadores STRs mostrando o relacionamento entre a distribuição das frequências alélicas nas microrregiões de Argentina, Brasil e Colômbia e as populações parentais. A primeira e a segunda coordenadas principais responderam por cerca de 88% da variação, sendo que a primeira coordenada representou 63% e a segunda 25% do total da variabilidade entre as populações. A separação das populações parentais foi bem definida, sendo que o eixo X mostrou a separação entre ameríndio e africano e europeu, localizados nos dois quadrantes à direita. O segundo componente principal separou o agrupamento europeu da parental africana, estando a primeira localizada no quadrante superior à direita. Os dados mostrados na análise de coordenada principal estão de acordo com os resultados de mistura e distância genética apresentados. Observou-se que as populações que apresentaram os maiores valores de contribuição para a parental ameríndia ficaram situadas nos dois quadrantes à esquerda, como é o caso de todas as regiões da Argentina e de duas regiões da Colômbia (sudoeste dos Andes e Andes, bacias dos rios Amazônia e Orinoco). As regiões da Argentina ficaram posicionadas entre as parentais ameríndia e européia, colocando-se de maneira oposta à parental africana, principalmente a região da Patagônia. Esse resultado era esperado tendo em vista que nas análises de mistura genética as regiões da Argentina apresentaram baixa contribuição africana, sendo que na Patagônia não foi observada contribuição desta parental.

As regiões que apresentaram maior contribuição africana ou européia ficaram posicionadas no lado direito do diagrama. Observou-se que duas regiões da Colômbia (Chocó e Ilha de Santo André e costa caribenha) que apresentaram maior contribuição da parental africana, em relação às demais regiões, ficaram localizadas no quadrante inferior direito, junto da parental africana, enquanto que a maioria das regiões brasileiras se localizou no mesmo quadrante da parental européia. As regiões colombianas ficaram espalhadas por três dos quatro quadrantes.

	EUR	AFR	AME	Col A	Col B	Col C	Col D	PAM	NOR	MES	GCH	PAT	BRSE	BRNE	BRN	BRS	BRCO
EUR																	
AFR	0,0256																
AME	0,0503	0,0674															
Col A	0,0086	0,0155	0,0305														
Col B	0,0192	0,0060	0,0492	0,0084													
Col C	0,0110	0,0254	0,0214	0,0034	0,0149												
Col D	0,0107	0,0306	0,0204	0,0046	0,0189	0,0008											
PAM	0,0046	0,0245	0,0302	0,0042	0,0158	0,0048	0,0046										
NOR	0,0295	0,0453	0,0119	0,0161	0,0309	0,0122	0,0117	0,0144									
MES	0,0059	0,0259	0,0252	0,0042	0,0162	0,0040	0,0036	0,0020	0,0144								
GCH	0,0101	0,0285	0,0215	0,0052	0,0187	0,0044	0,0041	0,0031	0,0112	0,0029							
PAT	0,0126	0,0359	0,0177	0,0070	0,0226	0,0035	0,0027	0,0048	0,0108	0,0042	0,0035						
BRSE	0,0039	0,0143	0,0415	0,0033	0,0089	0,0067	0,0078	0,0037	0,0229	0,0048	0,0071	0,0107					
BRNE	0,0040	0,0179	0,0412	0,0044	0,0129	0,0076	0,0083	0,0044	0,0228	0,0050	0,0074	0,0103	0,0027				
BRN	0,0042	0,0186	0,0340	0,0028	0,0112	0,0049	0,0053	0,0024	0,0190	0,0026	0,0048	0,0071	0,0020	0,0027			
BRS	0,0016	0,0220	0,0419	0,0058	0,0149	0,0073	0,0076	0,0032	0,0242	0,0037	0,0071	0,0094	0,0027	0,0030	0,0023		
BRCO	0,0027	0,0156	0,0405	0,0034	0,0099	0,0065	0,0074	0,0030	0,0223	0,0034	0,0061	0,0094	0,0014	0,0023	0,0017	0,0017	-

Figura 2.2. Matriz de distância genética de Reynolds *et al.* (1983) mostrando as distâncias genéticas entre as regiões geográficas da Argentina, Brasil, Colômbia e as populações parentais. EUR (europeu); AFR (africano); AME (ameríndio); Col A (Colômbia A); Col B (Colômbia B); Col C (Colômbia C); Col D (Colômbia D); PAM (região pampa da Argentina); NOR (região noroeste da Argentina); MES (região mesopotâmia da Argentina); GCH (região Gran Chaco da Argentina); PAT (região patagônia da Argentina); BRSE (região sudeste do Brasil); BRNE (região nordeste do Brasil); BRN (região norte do Brasil); BRS (região sul do Brasil); BRCO (região centro-oeste do Brasil).

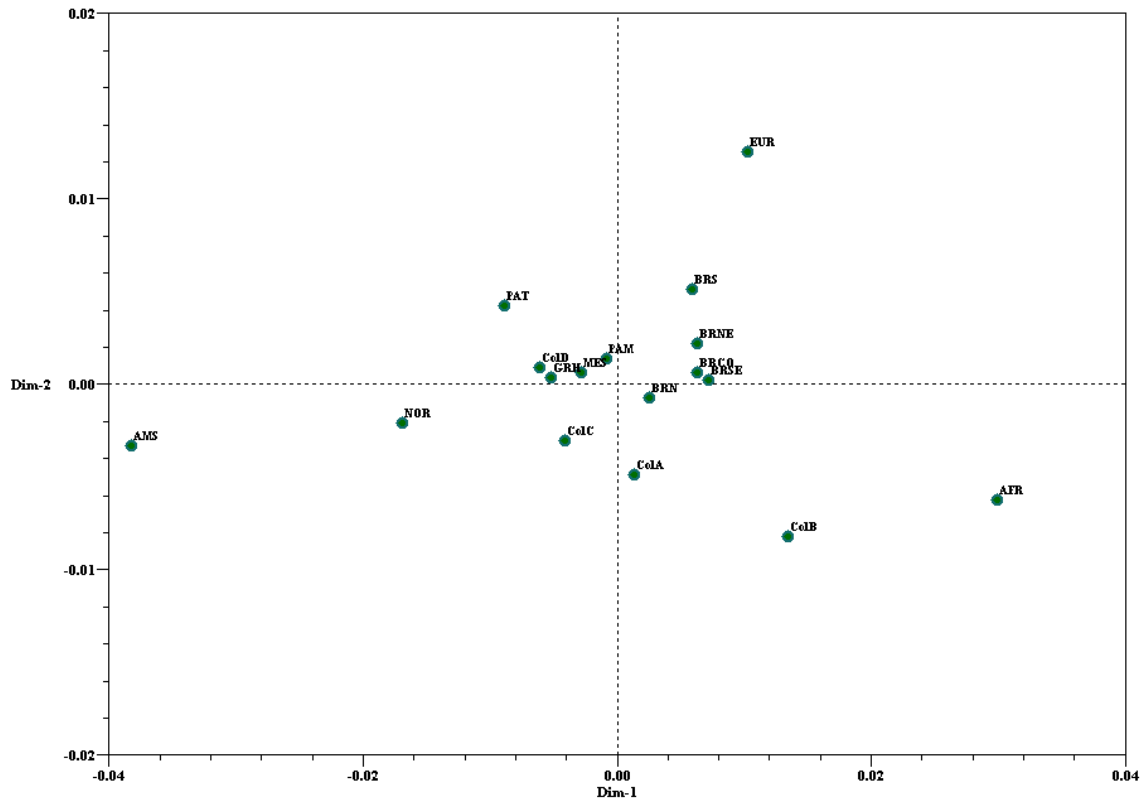


Figura 2.3. Primeira e segunda coordenadas principais obtidas a partir da análise de 13 marcadores STRs mostrando o relacionamento entre a distribuição das freqüências alélicas nas regiões de Argentina, Brasil e Colômbia e as populações parentais. Legenda: Parentais (EUR = Europeu, AFR = Africano, MAS = Ameríndio da América do Sul); Colômbia (ColA = região do Caribe, ColB = Chocó e Ilha de Santo André, ColC = região à sudoeste dos Andes; ColD = região dos Andes, bacias dos rios Amazônia e Orinoco); Argentina (PAM = Pampa, NOR = Noroeste, MES = Mesopotâmia, GRH = Gran Chaco, PAT = Patagônia); Brasil (BRSE = Sudeste, BRNE = Nordeste, BRN = Norte, BRS = Sul, BRCO = Centro-Oeste).

2.4. DISCUSSÃO

Os três países apresentaram como maior contribuição na constituição genética atual a parental européia. O processo de colonização pelos europeus da América, já previamente habitada por ameríndios, iniciou em 1492 com a vinda de Cristóvão Colombo, sendo que em 1498 chegou a América do Sul pelo atual litoral da Venezuela. Em decorrência do Tratado de Tordesilhas, Argentina e Colômbia foram colonizadas pelos espanhóis, enquanto que o Brasil pelos portugueses. O processo de povoamento iniciou pela costa atlântica e posteriormente avançou para o interior do continente. Em um segundo momento chegaram ao continente escravos advindos da África subsaariana.

Apesar de o processo de colonização ter sido parecido, o povoamento de cada país teve suas particularidades, o que refletiu na constituição atual das populações desses países. Os resultados de mistura gênica aqui obtidos sugerem que as diferenças entre as regiões geográficas são decorrentes do histórico de formação populacional de cada país e da forma com que os conquistadores europeus desenvolveram suas atividades econômicas.

A presença européia

O Brasil foi o país que apresentou contribuição européia estimada maior variando de 61 a 82% em todas as regiões geográficas. A região sul do país foi a que apresentou a maior contribuição dessa parental e, em decorrência disso, encontra-se agrupada na árvore com a parental européia. De acordo com a história do Brasil, a região sul recebeu pouco contingente escravo e, além disso, recebeu um grande contingente de migrantes alemães, italianos e de outras regiões da Europa por dois motivos distintos: ocupar as terras praticamente desocupadas de colonizadores e assim protegê-las da invasão de outros povos e fornecer mão-de-obra para os trabalhos agrícolas (Gregory, 2000; Gomes, 2000).

Foi também o Brasil o país que apresentou menor variação entre as regiões geográficas no percentual de contribuição das parentais. Esse resultado pode ser visto refletido nas estimativas de distâncias genéticas entre as regiões brasileiras que foram baixas, variando de 0,004 a 0,012, quando comparado às distâncias calculadas entre as regiões dos demais países. A migração européia, especialmente portuguesa, nos primeiros trezentos anos de colonização foi intensa. Além disso, a migração européia para o Brasil foi muito incentivada, especialmente entre o final do século XIX e início do século XX, visando o processo de “branqueamento” do povo brasileiro sob a cortina de necessidade de mão-de-obra em substituição da mão-de-obra escrava (Ramos, 2002). Fato similar ocorreu

na Argentina onde houve incentivo para migração europeia, incluindo o subsídio de passagens por um curto período, do governo espanhol para povoar a região. A imigração na Argentina foi constante, mas houve períodos em que o fluxo de indivíduos advindos, principalmente da Espanha e da Itália, foi intenso (Silberstein, 2000; Bernasconi, 2000).

A Argentina apresentou a segunda maior contribuição europeia, cerca de 60%, considerando a média ponderada do país como um todo. Entretanto, diferentemente do observado no Brasil, há diferenças importantes entre as regiões argentinas. Por exemplo, na região noroeste a contribuição europeia foi pouco mais de 1/3 da média nacional, ou seja, 23%, enquanto que na região do pampa foi superior à média, chegando a quase 70%. Essas diferenças refletem na estimativa total quando a frequência relativa populacional de cada região é usada para ponderar as contribuições parentais. A região noroeste é pouco povoada, enquanto que Pampa é bastante populosa, representando cerca de 63% da população daquele país (National Geographic, 2005). Quando se faz as estimativas de mistura geral corrigido pelo percentual relativo populacional de cada região, observou-se que a região noroeste contribuiu apenas com 3% para o componente europeu, e a região de Pampa, por ser mais populosa e apresentar maior composição genética europeia, contribuiu com mais de 44%. Assim, como existe uma grande variação na contribuição de cada região para compor a parental europeia o desvio foi grande ($0,601 \pm 0,179$). As estimativas aqui apresentadas são similares às estimadas utilizando cinco marcadores do tipo AIMs para a população de La Plata, cidade localizada no Pampa (Martinez-Marignac *et al.*, 2004), mostrando também altos valores de contribuição parental europeia para essa região.

A diferença entre as regiões também pôde ser observada na matriz de distância, onde a região noroeste foi a população que apresentou menor distância genética da parental ameríndia enquanto que a do Pampa foi mais próxima do grupo parental europeu. A análise de coordenada principal corroborou o achado, pois o noroeste da Argentina ficou posicionado no quadrante da parental ameríndia, mais distante geneticamente do grupo europeu, enquanto que o Pampa ficou próxima ao quadrante da parental europeia. Diferenças entre diversas províncias da Argentina quanto às frequências gênicas para vários *loci* STRs já haviam sido relatadas em um trabalho anterior. Com o objetivo de verificar se um banco de dados geral poderia ser usado para estimar parâmetros forenses nesse país, foi observado que existem diferenças e que essas ocorrem em maior intensidade entre as populações com maior componente ameríndio e aquelas com maior influência europeia (Toscanini *et al.*, 2007).

Por último, a Colômbia apresentou o componente europeu mais baixo na composição da sua população quando comparada com Argentina e Brasil. Notou-se, pelos dados obtidos na matriz de distância genética e nas coordenadas principais, que a Colômbia foi bastante heterogênea, com regiões apresentando altos índices de contribuição européia, como foi o caso da região D (região dos Andes, bacias dos rios Amazônia e Orinoco), localizada mais para o interior e é a segunda região mais populosa do país, onde se encontra a capital. A região que menos contribuiu para elevar as estimativas dessa parental foi a região B (Chocó e Ilha de Santo André), com menos de um por cento em termos de contribuição relativa. Esse fato será mais bem detalhado à frente.

O africano e o ameríndio

Quando analisados simultaneamente, os três países mostraram que, embora a presença européia seja alta na maioria das regiões, o processo de assimilação dos outros dois grupos parentais foi bastante heterogêneo, principalmente na Argentina e na Colômbia.

A contribuição africana na Argentina foi pequena, observando-se apenas pouco mais de 10% na região do Pampa, que é a mais habitada do país, e sua ausência na Patagônia. Na análise de componente principal a Patagônia ficou localizada no quadrante superior à esquerda, entre a parental européia e a ameríndia, portanto oposto à africana. As demais regiões ficaram posicionadas um pouco abaixo no *plot* devido à pequena contribuição africana.

De acordo com a história desse país, entre os séculos XVIII e XIX, existiram províncias em que a população afroderivada constituía pelo menos metade da população. Entretanto, a partir do século XIX houve intensa mortalidade de pessoas descendentes de escravos africanos em decorrência da participação em guerras, com a promessa de emancipação no caso de alistamento, e de doenças, como as epidemias de febre amarela, decorrentes do acentuado crescimento populacional nas cidades (Rock, 1985). Além disso, a imigração de europeus foi bem maior quando comparada a de africanos e a elite social, que era formada por pessoas de origem européia, repudiava a idéia de mistura genética com outras etnias, defendendo a "pureza do sangue europeu". Por isso, atualmente a contribuição da parental africana na população argentina é pequena.

No Brasil, a participação do grupo parental africano na população é mais evidente. De acordo com a história americana, o Brasil foi o país que mais recebeu escravos

advindos da África subsaariana. A maioria foi concentrada naquelas regiões onde havia, naturalmente, o maior desenvolvimento econômico, como em Recife, Salvador, Rio de Janeiro e São Paulo (Reis, 2000). Os resultados da análise de mistura mostraram que a região com maior contribuição parental africana foi a região sudeste do país, consequência da intensa utilização de mão-de-obra escrava nessa região até o século XIX. Por outro lado, a estimativa de contribuição africana na constituição da população atual da região sul foi muito pequena. Do mesmo modo, na matriz de distância genética e na coordenada principal, especialmente no eixo Y, fica claro que o sul do país não teve muita contribuição africana. A história de povoamento da região sul é diferente e bem posterior ao sudeste e nordeste como relatado anteriormente.

A Colômbia apresentou um quadro mais contrastante quanto às contribuições das populações parentais africana e ameríndia. Considerando apenas a contribuição do grupo parental africano, verificou-se presença marcante nas regiões da Costa Caribenha e na Costa Pacífica ao norte, onde se localiza o Estado de Chocó região em que houve a utilização de mão-de-obra escrava. Os escravos foram levados para essa região para trabalharem nas minas de ouro. Quando se faz a análise de mistura incluindo dados genéticos dessa região, a estimativa da contribuição africana na Colômbia é de cerca de 20%. Os resultados obtidos por esse trabalho estão de acordo com o descrito por Paredes *et al.* (2003) que ressaltaram que o processo de miscigenação entre os três grupos étnicos teve diferentes proporções no espaço e no tempo, o que tornou a população colombiana muito heterogênea. A região de Chocó e Ilhas de Santo André se assemelha a grupos isolados no Brasil que possuem elevada composição africana na constituição dos mesmos. Um estudo envolvendo três comunidades remanescentes de quilombo (São Gonçalo, Barra e Valongo), localizadas na Bahia e a última em Santa Catarina, por meio da análise de oito marcadores informativos de ancestralidade, mostrou índices próximos a 70% em São Gonçalo e Valongo (Luizon, 2007), similar ao observado na região de Chocó e Ilha de Santo André. Quando se considera a contribuição relativa de cada região na composição da parental africana, entretanto, a região que mais contribuiu significativamente para o valor encontrado para o país foi a costa caribenha, pois além de ser a região mais populosa apresentou índice de contribuição africana grande (35%).

Similarmente, quando se considera a contribuição parental ameríndia na composição dos três países, os maiores contrastes foram observados na Argentina e na Colômbia. Na Argentina, embora tenha havido a diminuição da população ameríndia durante o processo de colonização, essa parental ainda persiste como fruto da

miscigenação com o europeu principalmente, e em regiões mais isoladas, como a Patagônia e no interior do país onde a concentração populacional é menor. A região noroeste, como pode ser observado na análise da primeira coordenada principal, encontrou-se posicionada mais próxima da parental ameríndia do que todas as outras regiões analisadas. A região da Patagônia foi a segunda região mais próxima dessa parental. Nas regiões mais povoadas, especialmente Pampa onde se localiza a capital Buenos Aires, a contribuição ameríndia foi menor do que nas demais regiões, reflexo da concentração maior de grupos europeus na região e da dizimação provocada pelas doenças e pelas guerras à época da colonização daquela região. Como já foi colocado anteriormente, o contato do europeu com o ameríndio gerou guerras e trouxe doenças para essas comunidades resultando em morte e fuga para o interior.

Na Colômbia, duas regiões (região a sudoeste dos Andes e na região do Andes, Amazonas e Orinoco) tiveram contribuição ameríndia em torno de 40%, bem mais marcante que as regiões brasileiras. Entretanto, nas regiões mais costeiras (Costa do Caribe e Costa oeste ao norte do Caribe e Ilha de Santo André), onde houve os primeiros contatos com os espanhóis, os valores estimados foram quase metade das duas outras regiões. A árvore tipo estrela mostrou claramente esta diferença, sendo que as regiões C e D, com maior componente ameríndio, se agruparam com o ramo da parental ameríndia.

No Brasil ficou evidente a diminuição mais acentuada do povo ameríndio que vivia à época da colonização. Apenas a região norte do país, onde se encontra a maioria dos grupos ameríndios atuais, apresentou um percentual de contribuição mais elevado (17%). Observou-se que a região sudeste, onde ocorreu um processo de colonização mais intenso e desenvolvimento econômico mais acentuado, a proporção de contribuição dessa parental foi muito pequena (6%). Trabalhos anteriores utilizando os mesmos marcadores genéticos (Callegari-Jacques *et al.*, 2003), já haviam mostrado essa realidade, onde a presença ameríndia no interior do país é superior à encontrada na costa atlântica.

Devido a diferenças na história de povoamento e conseqüentemente de mistura entre os diferentes grupos étnicos, na matriz de distância verificou-se duas situações distintas: a primeira foi que as duas regiões colombianas que apresentaram maior contribuição africana contrastaram com a maior presença ameríndia na Argentina; a segunda foi que, de modo inverso, o Brasil, por apresentar a menor contribuição ameríndia na formação de sua população, se distanciou mais das regiões argentinas, que apresentaram contribuição desta parental elevada. O padrão de distribuição de contribuição ameríndia, observado nos três países, indicou que nas regiões onde ocorreram os primeiros contatos e

o desenvolvimento econômico, sua presença foi menor, ao contrário da africana que mostrou maior presença principalmente nas regiões costeiras.

Baseado nos resultados de distância genética entre as diferentes regiões geográficas desses países, pode-se concluir que devido a diferenças no processo de colonização, bem como no processo de migrações internas sucessivas, ocorridas durante e após o processo de colonização pelos europeus, percebe-se que a contribuição genética das populações parentais variou entre as regiões.

A chegada de novos grupos étnicos no subcontinente sul-americano, associada à fuga para o interior do continente dos grupos étnicos existentes, deu origem a um novo panorama populacional onde se percebe praticamente a ausência de povos ameríndios na costa atlântica e a presença mais marcante dos ameríndios no interior. Além disso, o europeu está presente em todas as regiões analisadas em menor ou maior grau, chegando em algumas regiões a alcançar níveis de contribuição muito altos, como é o caso da região sul do Brasil, com quase 82%. A Colômbia e a Argentina possuem ainda uma marcante contribuição da parental ameríndia (31% e 34%, respectivamente), mesmo após a redução da população após o contato com os colonizadores europeus, provavelmente, reflexo dos cruzamentos preferenciais entre o europeu e mulheres nativas. Os cruzamentos preferenciais entre esses grupos étnicos foram decorrentes da migração em massa nos primeiros trezentos anos de colonização de europeus, principalmente, do sexo masculino. Pelo fato da presença feminina européia ter sido escassa nessa época, a prole descendente dos europeus era miscigenada, isto é, filhos e filhas de mulheres africanas ou ameríndias.

Mesmo tendo se encerrado o processo migratório de africanos para a América no final do século XIX, a migração forçada durante 300 anos de tráfico transatlântico foi suficiente para deixar um reflexo importante na constituição populacional da maioria das regiões analisadas.

CAPÍTULO 3

**Mistura gênica na região centro-oeste brasileira estimada
utilizando dados de marcadores microssatélites autossômicos**

3.1. INTRODUÇÃO

A região centro-oeste brasileira, localizada no planalto central, ocupa 18,86% do território nacional. É composto por três estados – Goiás, Mato Grosso e Mato Grosso do Sul – e pelo Distrito Federal, onde se localiza a capital federal, Brasília. Embora seja a segunda região brasileira em extensão territorial, é a menos populosa, com apenas 6% da população total. Em termos de densidade populacional, é a região com a segunda menor densidade, sendo superior a apenas a região norte do país.

Por ser uma região interiorana, o povoamento por não ameríndios iniciou a partir do século XVII com movimentos migratórios promovidos pelo governo (entradas) ou pelas expedições particulares (bandeiras). Estas expedições tinham o objetivo de explorar metais preciosos, promover a colonização de regiões interioranas, capturar indígenas e escravos negros fugitivos. Grande parte da região centro-oeste pertencia à coroa de Castela e Espanha, mas o avanço dos bandeirantes para o oeste tornou-a de domínio de Portugal.

A descoberta de minas de ouro trouxe migrantes para esta região, principalmente de Minas Gerais e São Paulo, que promoveram, dentre outras modificações na região, a fundação de vilas como Vila Real do Bom Jesus de Cuiabá, atual capital do Estado de Mato Grosso, e Vila Boa, hoje a cidade de Goiás (Bertran, 2000; Palacín *et al.*, 1995). Com a decadência da mineração no século XIX, a pecuária e a agricultura (cana-de-açúcar e erva-mate), além da extração da borracha tiveram um avanço. Em decorrência disso ocorreu o segundo evento migratório. Os migrantes desta segunda fase eram advindos principalmente do atual estado do Maranhão e dos estados de Minas Gerais e São Paulo. Isso viabilizou a construção de estradas e a consolidação do desenvolvimento local, além da construção de novas cidades que atraíram mais migrantes (Palacín & Moraes, 1994).

Um terceiro evento migratório ocorreu a partir de 1956 incentivado pela construção da capital federal, Brasília. A migração promoveu uma alteração da população local e culminou com a formação de uma nova população constituída de migrantes das diferentes regiões do país (Bertran, 2000). Para se ter uma noção das mudanças provocadas pela construção de Brasília, em 1956 a população do Distrito Federal era composta por menos de 2000 habitantes. Já a contagem populacional de 2007 estimou a população em 2.455.903 milhões de habitantes (Instituto Brasileiro de Geografia e Estatística – IBGE, 2008), ou seja, houve crescimento de mais de 1200 vezes. Os migrantes desse terceiro evento migratório eram originados, principalmente, a partir dos Estados da própria região centro-oeste, especialmente Goiás, de estados do sudeste, como São Paulo, Minas Gerais e Rio de Janeiro, e do nordeste, como Piauí, Bahia e Paraíba (CODEPLAN, 2008).

A exemplo das demais populações do Brasil, a população da região centro-oeste é derivada da miscigenação de três grupos parentais principais: ameríndio, europeu, especialmente portugueses, e africano, oriundo da África subsaariana. Porém, em decorrência do povoamento tardio, os migrantes provavelmente já apresentavam miscigenação prévia. Por outro lado, encontraram tribos indígenas que possivelmente ainda não eram miscigenadas.

Com relação às tribos indígenas da região centro-oeste, estas sofreram mudanças na sua composição, bem como no número de indivíduos, após a chegada dos migrantes, havendo redução acentuada da população, a exemplo do que aconteceu com a ocupação da costa brasileira (Ribeiro, 1996). Esses primeiros habitantes locais, os ameríndios, pertenciam, em sua maioria, à família Jê, grupo que se localizou apenas no Brasil e, em menor proporção, à família Tupi-Guarani. A maioria das tribos era seminômade e vivia da caça e da pesca, porém algumas utilizavam técnicas rudimentares de cultivo de alimentos, como por exemplo, milho e batata (Bertram, 2000; Palacín *et al.*, 1995).

Devido ao intenso fluxo migratório dos portugueses para o interior do país, os ameríndios migraram ainda mais para o oeste. Entretanto, muitos grupos tribais resistiram ao avanço europeu. Esse contato os expôs a doenças para as quais eles não apresentavam resistência imunológica, como a varíola e o sarampo, de maneira que as epidemias foram as formas mais contundentes de eliminação de muitas tribos indígenas locais. Além disso, os portugueses empregaram diversos mecanismos para implementar a eliminação de indígenas, como a chacina de grupos tribais, o envenenamento de fontes de água, a destruição de plantações e a expulsão de suas terras devido às guerras (Ribeiro, 1996).

Após décadas de contato, a população ameríndia sobrevivente se viu obrigada a se inserir na economia como mão-de-obra barata, ou a sobreviver em reservas indígenas demarcadas pelo governo federal, como os parques indígenas do Xingu, do Araguaia, da Ilha do Bananal, Xavante e Parecis. A tabela 3.1 apresenta as tribos indígenas e o tamanho da população ameríndia em cada um dos três estados da região centro-oeste. Atualmente, como se pode notar, a população ameríndia de Goiás é a menor dessa região, sugerindo que a mesma diminuiu substancialmente ao longo da história. Relatos históricos mostram que, em 1808, existiam 20 tribos indígenas (Palacín & Moraes, 1994), enquanto que hoje existem apenas três. Vale ressaltar que nessa época os estados de Goiás e Tocantins formavam um único estado, sendo que a separação em duas unidades federativas se deu em 1988. Mesmo considerando as tribos indígenas no estado do Tocantins, que são oito na atualidade, observa-se uma redução significativa.

Esses acontecimentos são semelhantes ao que ocorreu em todo o Brasil durante o período de colonização e povoamento. Embora não existam dados precisos sobre a quantidade de indivíduos que compunham a população ameríndia no Brasil à época da colonização, é consenso, entretanto, que houve diminuição acentuada dessa população (Ribeiro, 1996). De acordo com o censo demográfico de 2000, 734.127 pessoas se autodeclararam indígenas no Brasil (IBGE), enquanto que na região centro-oeste esse número foi de 104.360 pessoas, valor esse muito superior ao apresentado na tabela 3.1. Isso sugere que a boa parte dos indígenas da região centro-oeste encontra-se inserido na população urbana e rural, fora das reservas indígenas.

Concomitante à vinda de portugueses e seus descendentes para a região, escravos foram trazidos, inicialmente para trabalharem nas minas e na captura de índios. A população migrante desse período era semelhante àquelas situadas em regiões de mineração, a maioria constituída por homens e solteiros. Outra característica importante da população de Goiás e Mato Grosso é a intensa mobilidade, tendo em vista que os indivíduos se deslocavam constantemente em busca de novas minas de ouro. Entretanto, a quantidade de migrantes (indivíduos de origem europeia e africana) era menor do que a população indígena local (Palacín *et al.*, 1995).

Apesar de não haver documentos históricos que apresentem dados exatos até 1804, é possível afirmar que em determinados momentos a população escrava superou a branca ou miscigenada. Exemplo disso é a estimativa baseada em dados de capitação (imposto cobrado sobre o número de escravos empregados na mineração) de 1736, que revelou uma população de escravos de mais de 10.000 indivíduos em uma população total de migrantes de 15.000 pessoas (Palacín *et al.*, 1995).

Com o objetivo de melhor compreender a história e formação da população da região centro-oeste, foram conduzidos levantamentos de dados biológicos focando no contraste entre a contribuição genética masculina (cromossomo Y) e feminina (DNA mitocondrial) nas populações de Goiás e do Distrito Federal (Barcelos *et al.*, 2006a; Barcelos *et al.*, 2006b). De acordo com esses estudos, a linhagem patrilinear é constituída principalmente por haplogrupos europeu (acima de 45%), com uma contribuição africana intermediária (aproximadamente 14%), e uma ameríndia muito pequena (menor que 2,8%). Por outro lado, a origem étnica da linhagem materna mostrou um contraste com a paterna, pois as contribuições ameríndia e africana são superiores à europeia. Dessa forma, fica evidente a ocorrência de cruzamento preferencial entre o homem europeu e mulheres ameríndias e africanas.

Tabela 3.1. Tribos indígenas localizadas nos Estados da região centro-oeste (Goiás, Mato Grosso e Mato Grosso do Sul) e número total de indivíduos por estado.

População	Grupos indígenas	Total (indivíduos)
Goiás	Avá-Canoeiro, Karajá e Tapuya.	346
Mato Grosso	Apiaká, Arara, Aweti, Bakairi, Bororo, Cinta Larga, Enawené-Nawê, Hahaintsú, Ikpeng, Irantxe, Juruna, Kalapalo, Kamayurá, Karajá, Katitaulú, Kayabí, Kayapó, Kreen-Akarôre, Kuikuro, Matipu, Mehináko, Metuktire, Munduruku, Mynky, Nafukuá, Nambikwara, Naravute, Panará Waurá, Pareci, Parintintin, Rikbaktsa, Suyá, Tapayuna, Tapirapé, Terena, Trumai, Umutina, Xavante, Xiquitano, Yawalapiti e Zoró.	25.123
Mato Grosso do Sul	Atikum, Guarany (Kaiwá e Nhandéwa), Guató, Kadiwéu, Kamba, Kinikinawa, Ofaié, Terena e Xiquitano.	32.519

Fonte: Fundação Nacional do Índio (Funai) – www.funai.gov.br (acessado em 15/04/2008)

O volume de dados referentes à avaliação genética para populações dessa região ainda é escasso, o que restringe a reconstrução histórica do povoamento da região com base nesse tipo de informação. Por isso, este capítulo teve por objetivo traçar um panorama da composição genética da população da região centro-oeste brasileira com relação à porcentagem de contribuição das três principais populações parentais - europeus, africanos subsaarianos e ameríndios – e as similaridades e diferenças genéticas entre as populações dos estados que compõem a região. Para a realização das análises, que incluíram também estimativas de distância genética, foram utilizados dados de frequências alélicas de 14 marcadores genéticos do tipo STR autossômicos para os estados de Goiás e Mato Grosso do Sul e para o Distrito Federal.

3.2. MATERIAL E MÉTODOS

Amostras populacionais

Para análise do perfil genético da população do Estado de Goiás, foram selecionadas amostras de sangue de 153 indivíduos voluntários saudáveis, que pertenciam a um banco de amostras de DNA com 500 indivíduos no total, da população de Goiás. O banco, que teve como objetivo obter uma ampla amostragem da população do estado, foi formado levando em consideração a densidade demográfica das 18 microrregiões goianas, sendo que o número de amostras coletadas foi proporcional ao número de habitantes. A coleta foi realizada nas três cidades mais populosas de cada microrregião entre os meses de fevereiro a junho de 2000. O sorteio das 153 amostras foi realizado buscando a representação proporcional ao tamanho de cada microrregião (Tabela 3.2).

A coleta de sangue foi realizada por punção venosa utilizando EDTA como anticoagulante. Os leucócitos foram separados do sangue total coletado e armazenados em freezer a 20°C negativos até o momento da extração de DNA. O DNA genômico foi extraído utilizando o conjunto de reagentes “GFX™ *Genomic Blood Purification Kit*” – Amersham Pharmacia Biotech, e seguiu as recomendações do fabricante (anexo 2).

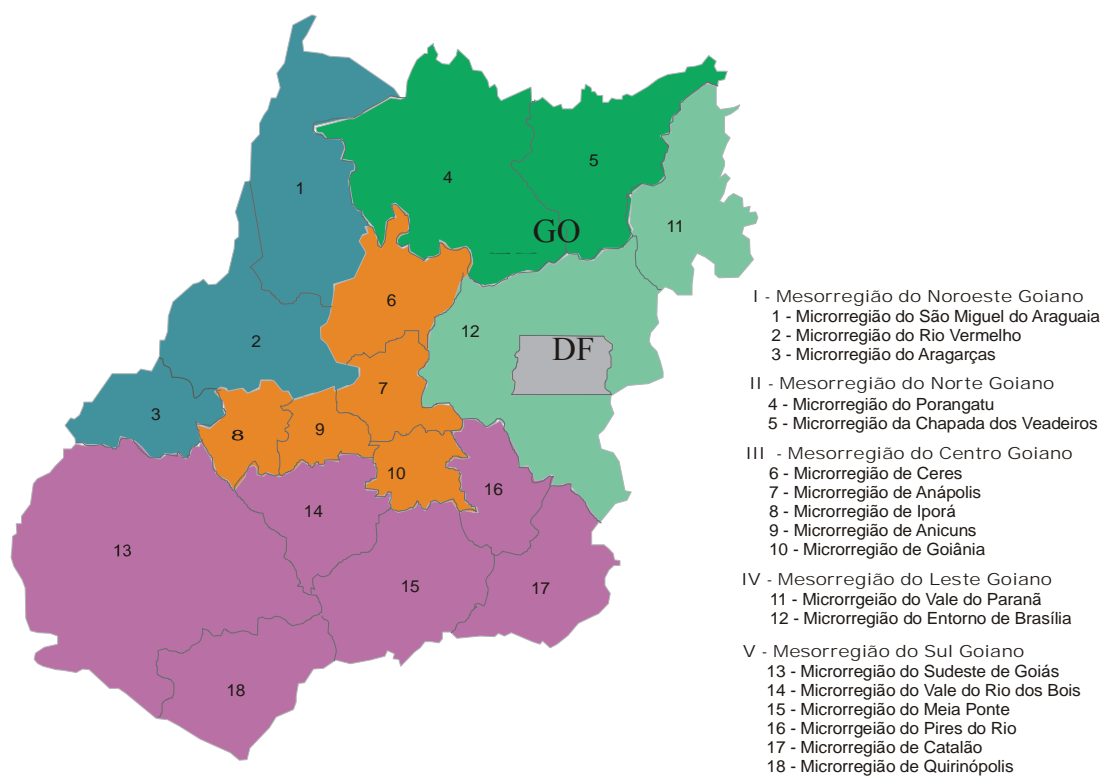


Figura 3.1. Mapa do estado de Goiás destacando as dezoito microrregiões, de acordo com classificação do sistema de cartografia do IBGE (2000).

Tabela 3.2. Discriminação das mesorregiões, microrregiões, do número de habitantes de cada microrregião à época da coleta das amostras de sangue e do número de indivíduos coletados em cada região do Estado de Goiás.

Mesorregião	Microrregião	População	Quantidade de pessoas analisadas
Noroeste Goiano	São Miguel do Araguaia	68.888	2
	Rio Vermelho	91.254	4
	Aragarças	55.263	2
Norte Goiano	Porangatu	226.766	8
	Chapada dos Veadeiros	55.002	2
Centro Goiano	Ceres	212.661	8
	Anápolis	451.474	16
	Iporá	64.030	2
	Anicuns	103.276	2
	Goiânia	1.633.899	40
Leste Goiano	Vão do Paranã	88.913	4
	Entorno de Brasília	766.535	29
Sul Goiano	Sudoeste de Goiás	330.490	13
	Vale do Rio dos Bois	102.491	3
	Meia Ponte	305.427	9
	Pires do Rio	76.856	2
	Catalão	114.686	3
	Quirinópolis	91.190	4
	Total		4.839.101

Fonte: contagem populacional IBGE – 1996.

Após a extração, as amostras de DNA foram quantificadas em gel de agarose a 1% contendo brometo de etídio (Anexo 2). O brometo de etídio é um corante químico que se liga à cadeia dupla do DNA e que, quando exposto à luz ultravioleta, emite fluorescência. Desse modo a leitura dos dados era feita com auxílio do padrão de marcador de peso de molecular de 500 pares de base de DNA pela visualização das bandas em um transiluminador com luz ultravioleta. Após a quantificação, as amostras foram diluídas para 1 ng/μl com água Mili-Q autoclavada.

Os marcadores genéticos analisados foram microssatélites do sistema *PowerPlex®16* (*Promega Corporation*), composto pelos marcadores CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, Penta D, Penta E, TH01, TPOX e vWA e a amelogenina (que identifica o sexo do indivíduo). A amplificação dos fragmentos específicos foi realizada utilizando a técnica de reação em cadeia da polimerase (PCR) e a concentração dos reagentes foi realizada seguindo recomendações do fornecedor (anexo 2).

Controle positivo (0,5ng/μl de DNA 9947A) e negativo (água Milli-Q no lugar da amostra) foi utilizado para todos os conjuntos de reações de amplificação. A PCR foi realizada no termociclador 9700 *Applied Biosystems Thermal Cycler*, de acordo com o protocolo da *Promega Corporation* (anexo 2). A eletroforese foi feita em um seqüenciador ABI 3100 *Avant* (*Applied Biosystems*). Os genótipos foram determinados pela comparação com escadas alélicas para cada marcador usando o programa GeneMapper ID (V3.2).

Levantamento de dados para realização das análises estatísticas

Os dados relativos à região centro-oeste foram complementados com dados publicados para as populações do Distrito Federal (dados gentilmente cedidos por Gustavo G. Dalton) e Mato Grosso do Sul (Silva *et al.*, 2004). O banco de dados utilizado para a população do Distrito Federal foi proveniente de um banco de dados público, isto é, do Laboratório de DNA da Polícia Civil do Distrito Federal, que faz análises em investigações de paternidade, de pessoas que não possuem condições financeiras de arcar com os custos dos exames, e criminais.

Para fins de comparação, foram também levantados dados de frequências de marcadores STRs para populações brasileiras a partir da literatura (anexo 1). Com isso, foi possível obter dados dos seguintes estados brasileiros: Amazonas (Corte Real *et al.*, 2000) e Pará (Rodrigues *et al.*, 2006) da região norte; Alagoas (Silva *et al.*, 2002), Pernambuco (Dellalibera *et al.*, 2004), Piauí (Silva *et al.*, 2002), Rio Grande do Norte (Silva *et al.*, 2003) e Sergipe (Pimentel *et al.*, 2004) da região nordeste; São Paulo (Bydlowski *et al.*, 2003; Ferreira *et al.*, 2006) e Rio de Janeiro (Góes *et al.*, 2004; Silva *et al.*, 2004; Silva & Moura-Neto, 2004) da região sudeste; Rio Grande do Sul (Leite *et al.*, 2003) e Santa Catarina (Caine *et al.*, 2003) da região sul; e dois bancos de dados gerais pertinentes à população brasileira (Grattapaglia *et al.*, 2001; Whittle *et al.*, 2004). Pelo fato de um dos bancos de dados de frequência de São Paulo apresentar população dividida em quatro

diferentes grupos étnicos (Bydlowski *et al.*, 2003), para fins de análise, foi considerada a média ponderada entre os mesmos.

Análise estatística

Dos dados gerados para a população de Goiás

A análise intrapopulacional para a população de Goiás objetivou estimar as frequências alélicas, heterozigose observada e esperada, teste de aderência ao equilíbrio de Hardy-Weinberg, desequilíbrio de ligação entre os marcadores estudados e do conteúdo informativo de polimorfismo (PIC). As análises de frequências alélicas e heterozigose foram realizadas utilizando o programa PowerMarker v3.0 (Liu & Muse, 2005). As demais análises foram realizadas utilizando o programa GDA (Lewis & Zaykin, 2002).

A frequência para cada alelo foi estimada com base no método de contagem direta, que é definido como o número de alelos encontrados na amostra, isto é,

$$p_i = n_i/2n,$$

sendo n_i é número de ocorrências do alelo i na amostra e n é o número total de indivíduos amostrados.

A heterozigose observada, uma medida de variabilidade genética da população, é a quantidade de indivíduos heterozigotos observados na amostra, sendo:

$$\hat{H}_l = 1 - \sum_{i=1}^k P_{l_{ii}}$$

sendo, ii a frequência dos homozigotos i para o *locus* l .

A heterozigose esperada corresponde à probabilidade de que dois alelos escolhidos aleatoriamente em uma população, para um determinado *locus*, sejam diferentes. Tanto a frequência alélica quanto a genotípica foram estimativas baseadas no método da máxima verossimilhança não viesada das frequências populacionais. Os intervalos de confiança de ambas frequências foram formados por 1000 reamostragens dos indivíduos no conjunto de dados.

O teste para equilíbrio de Hardy-Weinberg foi realizado com objetivo de verificar se as proporções de heterozigotos observados estão de acordo com o esperado, conforme o seguinte polinômio:

$$(x_i + x_j + \dots + x_n)^2$$

x_i corresponde à frequência esperada do alelo i ; x_j é a frequência esperada para o alelo j ; x_n é a frequência esperada para o alelo n em um determinado *locus*.

O teste exato foi utilizado para testar a hipótese de equilíbrio, sendo que 10.000 permutações foram realizadas em cada teste. A medida *chi-square* foi utilizada com objetivo de verificar se uma combinação em particular resultou em um valor de *chi-square* mais extremo do que aqueles produzidos pelos dados originais.

O PIC é uma estimativa relacionada à diversidade do *locus*, ou seja, mede o grau de polimorfismo de um determinado marcador.

Análises interpopulacionais

Após a obtenção dos dados de frequência alélica dos marcadores STRs analisados para a população de Goiás, 14 marcadores foram selecionados para análise comparativa com as demais populações, principalmente da região centro-oeste. O marcador Penta D foi retirado das análises descritas a seguir em decorrência da ausência de dados para a população do Mato Grosso do Sul.

A análise de variância molecular (AMOVA - Excoffier *et al.*, 1992), o índice de estruturação populacional *F_{st}* (Weir e Cockerham, 1984) e o teste de significância do mesmo (Excoffier *et al.*, 1992) foram utilizados para comparar os estados de Goiás, Mato Grosso do Sul e Distrito Federal, assim como para comparar as populações dessa região com populações de outros estados da Federação sendo: Alagoas, Amazonas, Pará, Pernambuco, Piauí, Rio de Janeiro, Rio Grande do Norte, Rio Grande do Sul, São Paulo, Santa Catarina, Sergipe e dois bancos de dados gerais pertinentes à população brasileira. O objetivo, ao se realizar tais análises, foi verificar se as populações da região centro-oeste são subestruturadas e, em um segundo momento, avaliar se as populações do Distrito Federal, Goiás e Mato Grosso do Sul são diferentes de outras populações brasileiras. Foi realizada também análise de diversidade gênica intrapopulacional *h_{sk}* (Nei, 1987), estimada para cada *locus* para as populações da região centro-oeste (Distrito Federal, Goiás e Mato Grosso do Sul). Em todas essas análises foi utilizado o programa Arlequin versão 2000 (Schneider *et al.*, 2000).

Para a estimativa da diversidade gênica, foi calculada a média ponderada das mesmas para os marcadores e para as populações do centro-oeste, sendo que nesse último caso foram considerados todos os marcadores simultaneamente.

As distâncias genéticas entre os pares de populações da região centro-oeste e as parentais foram estimadas como proposto por Reynolds *et al.* (1983), utilizando o software PHYLIP versão 3.65 (Felsenstein, 1989). Para as populações parentais foi considerada a média ponderada das estimativas de frequência alélica de cada marcador.

Análise de mistura foi realizada para dois estados da região centro-oeste (o estado de Mato Grosso não apresenta dados disponíveis na literatura) e para o Distrito Federal com o objetivo de estimar o percentual de contribuição das parentais européias, africanas e ameríndias na formação da população dessa região. As populações parentais africanas e européias utilizadas para as análises de mistura foram as mesmas utilizadas no capítulo 1, sendo que para a parental ameríndia foram considerados os dados de frequências alélicas de ameríndios da América do Sul. As estimativas foram calculadas pelo método de identidade gênica (Chakraborty, 1985) utilizando o programa ADMIX 3.

3.3. RESULTADOS

Análise intrapopulacional

A tabela 3.3 apresenta os dados de estatística descritiva para os 15 marcadores STRs avaliados na população do estado de Goiás, como frequência alélica, heterozigose observada e esperada, valor do teste exato para equilíbrio de Hardy-Weinberg. Todos os marcadores encontravam-se em equilíbrio de Hardy-Weinberg, considerando nível de significância de 5%. A estimativa do conteúdo de polimorfismo para os 15 *loci* variou de 0,61 a 0,89. O marcador que apresentou menor conteúdo informativo de polimorfismo foi o TPOX, com sete alelos detectados, sendo que um deles, o alelo 8, teve frequência superior a 50%. O marcador com maior conteúdo informativo de polimorfismo foi o PentaE, com 18 alelos detectados, sendo que o alelo 12 foi o mais frequente (16,5%).

Não foi detectado desequilíbrio de ligação na análise considerando todos os 15 *loci* simultaneamente. Na análise par a par dos 105 pares possíveis, apenas cinco pares apresentaram-se em desequilíbrio, sendo eles: CSF1PO/PentaD, D18S51/FGA, D18S51/vWA, D7S820/PentaD e D7S820/TPOX. Entretanto, após a correção de Bonferroni ($\alpha=0,0005$), todos os pares de *loci* ficam em equilíbrio.

Análise interpopulacional

A tabela 3.4 apresenta os resultados de comparação (*Fst*), que apresentaram valores estatisticamente significativos, entre os marcadores que compõem os bancos de dados das três populações da região centro-oeste e destes com outros bancos de dados do país. Não houve diferença para os *loci* analisados entre os bancos de dados da região centro-oeste.

Tabela 3.3. Frequências alélicas e parâmetros estatísticos estimados para a população do estado de Goiás, correspondente aos 15 marcadores STRs do sistema PowerPlex 16.

Alelo	D8S1179	D16S539	D18S51	vWA	D3S1358	FGA	D21S11	Penta - E
5								0,0888
7								0,1151
8	0,0065	0,0196						0,0822
9	0,0131	0,1046						0,0066
10	0,0915	0,0784	0,0132					0,0526
11	0,0752	0,3627	0,0066	0,0033				0,1349
12	0,1307	0,2353	0,1612	-	0,0033			0,1645
13	0,2941	0,1797	0,0822	-	0,0033			0,1086
14	0,2549	0,0196	0,1678	0,0752	0,0882			0,0691
15	0,1176		0,1382	0,1732	0,2614			0,0395
16	0,0131		0,1447	0,2255	0,2745			0,0395
17	0,0033		0,1184	0,2614	0,2418			0,0461
18			0,0493	0,1863	0,1176	0,0033		0,0066
19			0,0625	0,0588	0,0098	0,1111		0,0132
20			0,0362	0,0098		0,0980		0,0099
21			0,0132	0,0065		0,1471		0,0099
22			0,0033			0,1961		0,0099
22.2			-			0,0033		-
23			-			0,1307		0,0033
24			0,0033			0,1438		
24.2						-	0,0033	

Alelo	D8S1179	D16S539	D18S51	vWA	D3S1358	FGA	D21S11	Penta - E
25						0,0948	-	
26						0,0686	-	
27						-	0,0327	
28						0,0033	0,1732	
29							0,2092	
30							0,2516	
30.2							0,0261	
31							0,0556	
31.2							0,0915	
32							0,0065	
32.2							0,0915	
33							0,0033	
33.2							0,0392	
34.2							0,0131	
38							0,0033	
N	153	153	152	153	153	153	153	152
H _{obs}	0,856	0,739	0,849	0,784	0,758	0,843	0,895	0,895
H _{exp}	0,806	0,765	0,880	0,809	0,779	0,869	0,842	0,905
p	0,611	0,185	0,119	0,721	0,620	0,349	0,065	0,897
PIC	0,78	0,73	0,86	0,78	0,74	0,85	0,82	0,89

N – número de indivíduos, H_{obs} – heterozigose observada, H_{exp} – heterozigose esperada, p – valor de p para o teste exato de equilíbrio de Hardy-Weinberg baseado em 10.000 shufflings (valor menor que 0,05 foi considerado estatisticamente significativo), PIC – conteúdo informativo de polimorfismo.

Continuação da Tabela 3.3

Alelo	Penta D	TH01	TPOX	D5S818	D7S820	CSF1PO	D13S317
2.2	0,0331						
3.2	0,0033						
4	-						
5	0,0066	0,0033					
6	0,0033	0,1928	0,0132				
7	0,0066	0,2222	0,0033	0,0229	0,0196	0,0197	
8	0,0430	0,1373	0,5099	0,0131	0,1601	0,0132	0,1275
9	0,1887	0,1732	0,1020	0,0327	0,1307	0,0263	0,0523
9.3	-	0,2614	-	-	-	-	-
10	0,1258	0,0098	0,0658	0,0588	0,2908	0,2763	0,0621
11	0,1589		0,2599	0,2876	0,2222	0,3322	0,3268
12	0,1755		0,0461	0,3693	0,1471	0,2796	0,2745
13	0,1523			0,2026	0,0261	0,0493	0,1242
14	0,0629			0,0098	0,0033	0,0033	0,0327
15	0,0331			-			
16	0,0066			0,0033			
N	151	153	152	153	153	152	153
H _{obs}	0,874	0,699	0,664	0,778	0,856	0,757	0,725
H _{exp}	0,864	0,799	0,658	0,737	0,803	0,734	0,781
p	0,383	0,518	0,377	0,890	0,269	0,901	0,156
PIC	0,85	0,76	0,61	0,69	0,77	0,68	0,75

Tabela 3.4. Pares de populações que apresentaram diferenças estatisticamente significativas por marcador STR considerando 14 marcadores e 13 bancos de dados da população brasileira.

Marcadores	Pares de populações	Valores de p
CSF1PO	DF x PE	0,0007 ± 0,0005
D3S1358	GO x SE	< 0,0001
	MS x SE	< 0,0001
D7S820	DF x PI	< 0,0001
	GO x PI	< 0,0001
	MS x PI	< 0,0001
D8S1179	DF x PE	0,0291 ± 0,0003
	DF x BR2	0,0298 ± 0,0003
D13S317	DF x PI	< 0,0001
	GO x PI	< 0,0001
	MS x PI	< 0,0001
D16S539	DF x SP*	0,0142 ± 0,0019
	GO x SP*	0,0056 ± 0,0014
	MS x SP*	0,0248 ± 0,0028
D21S11	DF x PE	< 0,0001
	GO x PE	< 0,0001
	MS x PE	< 0,0001
FGA	DF x PA	< 0,0001
	GO x PA	0,0119 ± 0,0018
	MS x PA	0,0004 ± 0,0003
TH01	DF x BR 1	0,0073 ± 0,0015
	DF x BR 2	< 0,0001
	DF x SP*	0,0198 ± 0,0026
TPOX	DF x SP*	< 0,0001
	GO x SP*	0,0089 ± 0,0018
	MS x SP*	0,0004 ± 0,0003
vWA	DF x RS	< 0,0001
	MS x RS	< 0,0001
	GO x RS	0,0004 ± 0,0003

Legenda: BR1 (banco de dados brasileiro publicado por Grattapaglia *et al.*, 2001); BR2 (banco de dados brasileiro publicado por Whittle *et al.*, 2004); DF (Distrito Federal); GO (Goiás); MS (Mato Grosso do Sul); PA (Pará); PE (Pernambuco); PI (Piauí); SE (Sergipe); SP (São Paulo); e RS (Rio Grande do Sul). * Banco de dados referente a Bydlowski *et al.*, 2003.

Na comparação das frequências alélicas, considerando 14 STRs, entre a população da região centro-oeste (Distrito Federal, Goiás, Mato Grosso do Sul) e outros bancos de dados da população brasileira, foram encontradas diferenças entre 29 pares. Dos quatorze *loci* STRs analisados, 11 apresentaram divergência entre as populações analisadas, especialmente envolvendo as populações de São Paulo e Piauí. Essas diferenças entre os pares de populações se refletiram nos resultados de *Fst* quando todas as populações foram analisadas simultaneamente (Tabela 3.5). Dez marcadores apresentaram valores de *p* significativos para *Fst*. Três marcadores (D5S818, D18S51, e PentaE) não apresentaram nenhum tipo de diferenciação no conjunto das populações, enquanto que o marcador D8S1179 diferenciou apenas a população do Distrito Federal da de Pernambuco e um banco de dados da população brasileira (Whittle *et al.*, 2004).

Com relação às estimativas de diversidade gênica intrapopulacional (anexo1), observou-se uma similaridade entre as três populações da região centro-oeste para todos os marcadores analisados. A diversidade gênica média para as três populações da região centro-oeste variou de 0,68 (TPOX) a 0,91 (PentaE). Quanto à média da diversidade gênica estimada para todos os marcadores, verifica-se que não houve diferença entre as populações, sendo que os três estados apresentaram diversidade gênica média de 0,80.

A matriz de distância de Reynolds *et al.* (1983) está apresentada na Figura 3.2. Quando se considerou apenas as distâncias entre as populações parentais com as três populações analisadas da região centro-oeste, observou-se que as menores distâncias ocorreram entre as três populações miscigenadas e a parental européia, sendo que a população de Goiás foi a que mostrou a menor distância. De modo inverso, as três populações miscigenadas se distanciaram mais do componente ameríndio, sendo que a população de Goiás foi o que apresentou a maior distância em relação a essa parental. A distância genética em relação à parental africana apresentou valores intermediários para as três populações analisadas.

A tabela 3.6 apresenta os resultados de estimativa de contribuição das populações parentais européia, africana e ameríndia na constituição atual do Distrito Federal, Goiás e Mato Grosso do Sul. Os resultados obtidos corroboram os dados gerados na análise de distância. As três populações apresentaram maior contribuição genética da parental européia. A contribuição da parental africana foi intermediária em relação às outras duas, sendo que a população do Distrito Federal apresentou o maior percentual (26,2%) desta parental quando comparado com as outras duas populações. A contribuição da parental ameríndia foi baixa nas três populações, sendo que a menor estimativa foi observada na

população de Goiás (3%), enquanto que o Distrito Federal foi o que apresentou contribuição ameríndia (13,6%) superior, em relação às outras duas populações da região centro-oeste.

Tabela 3.5. Teste de significância das análises de *Fst* considerando bancos de dados da população brasileira ($\alpha=5\%$) referente à frequência para 14 marcadores STRs.

<i>Locus</i>	Valor <i>p</i>	Populações analisadas
CSF1PO	< 0,0001	AL, AM, BR1, BR2, DF, GO, MS, PE, PI, RJ, RN, SE, SP
D3S1358	< 0,0001	AM, BR1, BR2, DF, GO, MS, PA, RJ, RS, SC, SE
D5S818	0,3021	AM, BR1, BR2, DF, GO, MS, PA, RJ, RS, SC, SP
D7S820	< 0,0001	AL, AM, BR1, BR2, DF, GO, MS, PE, PI, RJ, RS, SP
D8S1179	0,5249	AM, BR1, BR2, DF, GO, MS, PA, PE, RJ, RS, SC
D13S317	< 0,0001	AL, AM, BR1, BR2, DF, GO, MS, PE, PI, RJ, RN, SC, RS, SE, SP
D16S539	< 0,0001	AL, AM, BR1, BR2, DF, GO, MS, PE, PI, RJ, RN, SP
D18S51	0,9677	AM, BR1, BR2, DF, GO, MS, PA, PE, RJ, SC
D21S11	< 0,0001	AM, BR1, BR2, DF, GO, MS, PE, RJ, RS, SC
FGA	0,0020	AM, BR1, BR2, DF, GO, MS, PA, PE, RJ, RS, SC
Penta-E	0,2082	BR, DF, GO, MS, RJ
TH01	< 0,0001	AL, AM, BR1, BR2, DF, GO, MS, PE, PI, RJ, SP
TPOX	< 0,0001	AL, AM, BR1, BR2, DF, GO, MS, PE, PI, RJ, SP
vWA	< 0,0001	AL, AM, BR1, BR2, DF, GO, MS, PE, RJ, RS, SP

Legenda: AL (Alagoas), AM (Amazonas), BR1 (banco de dados da população brasileira publicado por Grattapaglia *et al.*, 2001), BR2 (banco de dados da população brasileira publicado por Whittle *et al.*, 2004), DF (Distrito Federal), GO (Goiás), MS (Mato Grosso do Sul), PA (Pará), PE (Pernambuco), PI (Piauí), RJ1 (Rio de Janeiro), RN (Rio Grande do Norte), SC (Santa Catarina), SE (Sergipe), SP (São Paulo), RS (Rio Grande do Sul).

	EUR	AFR	AME	DF	GO	MS	CO
EUR	-						
AFR	0,0260	-					
AME	0,0500	0,0665	-				
DF	0,0038	0,0141	0,0394	-			
GO	0,0025	0,0197	0,0487	0,0032	-		
MS	0,040	0,0209	0,0404	0,0028	0,0055	-	
CO	0,0027	0,0158	0,0403	0,0003	0,0022	0,0018	-

Figura 3.2. Matriz de distância genética de Reynolds *et al.* (1983) para as populações do Distrito Federal, Goiás, Mato Grosso do Sul, da região centro-oeste como um todo, bem como das populações parentais utilizando 14 marcadores genéticos do tipo STRs. Legenda: AFR (africano); AME (ameríndio); DF (Distrito Federal); GO (Goiás); MS (Mato Grosso do Sul).

Tabela 3.6. Estimativas de contribuição dos três principais grupos parentais na composição genética atual de dois estados brasileiros da região Centro-oeste – Goiás e Mato Grosso do Sul – e do Distrito Federal, utilizando 14 marcadores microssatélites.

Populações	Populações parentais		
	Africano	Ameríndio	Europeu
Distrito Federal	0,262 ± 0,005	0,136 ± 0,007	0,603 ± 0,008
Goiás	0,133 ± 0,005	0,030 ± 0,006	0,837 ± 0,008
Mato Grosso do Sul	0,139 ± 0,001	0,124 ± 0,002	0,736 ± 0,002
Centro-oeste	0,208 ± 0,004	0,117 ± 0,006	0,675 ± 0,007

3.4. DISCUSSÃO

O processo de colonização da região centro-oeste brasileira foi decisivo para a composição da população local. Os migrantes vieram de todas as regiões brasileiras, principalmente das regiões sudeste e nordeste. Esses migrantes, muitas vezes já eram produtos de miscigenação entre pelo menos dois grupos étnicos, mas tinham em sua constituição, no início da colonização, principalmente o europeu e o africano. É o que pode ser inferido de estudos com marcadores do cromossomo Y para duas populações da região centro-oeste (Barcelos *et al.*, 2006). Além disso, o fluxo migratório intenso modificou a realidade daquela época, em que a presença ameríndia era dominante.

Os resultados da análise estatística para os 15 marcadores STRs para a população do estado de Goiás apresentaram diversidade gênica alta pela análise de três parâmetros estatísticos. Primeiro, os marcadores apresentaram heterozigose entre 0,664 a 0,895. Na amostra foi possível detectar a presença de alelos raros, isto é, com valores de frequência menores que 0,5%, para todos os marcadores analisados. A heterozigose elevada indica que esta população apresenta uma variabilidade elevada, provavelmente mantida pelo cruzamento aleatório entre os indivíduos e pela migração contínua na região.

Os dois outros parâmetros analisados, conteúdo informativo de polimorfismo e a diversidade gênica de Nei (1972), refletem a diversidade de alelos para cada marcador e a frequência desses alelos na população. Os índices encontrados foram similares aos encontrados em outras populações (Gomes *et al.*, 2007; Rodrigues *et al.*, 2007; Morales *et al.*, 2004; Tofanelli *et al.*, 2003). Além disso, as populações parentais, principalmente europeus e africanos, apresentam alta diversidade gênica para os marcadores utilizados nessa análise, sugerindo panmixia na população.

Os resultados obtidos com a análise de equilíbrio de Hardy-Weinberg indicam que a população do estado de Goiás é estruturada. O mesmo foi detectado no Mato Grosso do Sul (Silva *et al.*, 2004) e Distrito Federal (Dalton *et al.*, comunicação pessoal), considerando apenas os 14 STRs em comum. Embora a miscigenação da população seja um fator que altere as proporções genóticas, o tempo decorrido desde o início do processo de povoamento não ameríndio permitiu a redistribuição dos genótipos em frequências mais estáveis. Associações não aleatórias entre pares de genótipos, localizados em cromossomos diferentes ou separados longe o suficiente no cromossomo para permitir recombinação meiótica freqüente, podem ocorrer, em maior ou menor grau, do que o previsto por suas frequências individuais (Destro-Bisol *et al.*, 2000). Isso pode ocorrer devido à subestruturação populacional, ocasionada por fatores evolutivos tais como

migração e deriva genética. Na análise par a par entre os *loci* analisados não foi observado desequilíbrio de ligação após aplicação da correção de Bonferroni, mesmo sendo essa uma população com intenso fluxo migratório, e que teve na composição inicial a contribuição de três grupos parentais divergentes, isso é, povos de origem européia, africana e ameríndia.

Na comparação par a par dos valores de *Fst*, verificou-se que as populações da região centro-oeste apresentaram-se homogêneas, fato explicado pelo processo de formação das populações locais. À época da colonização existiam na região grupos ameríndios que migraram entre os territórios dos estados. Muitos grupos indígenas situados à época no estado de Goiás, por exemplo, foram obrigados a deixarem seus territórios e a migrarem para o oeste, onde atualmente se encontram os estados de Mato Grosso e Mato Grosso do Sul. Além disso, os europeus que para cá vieram, descendentes de portugueses em sua maioria, ocuparam amplamente quase todo o território e trouxeram escravos que trabalharam inicialmente em minas de mineração. Essa ocupação em busca de riquezas abrangeu ao mesmo tempo os três estados da região centro-oeste (Distrito Federal nessa época era parte integrante de Goiás). Como exposto anteriormente, a maioria dos migrantes era composta por indivíduos do sexo masculino e possuíam extrema mobilidade. A cada nova descoberta de metais preciosos, nova onda migratória ocorria integrando as regiões. Assim, o fluxo gênico entre as três populações homogeneizou a distribuição dos genótipos, mantendo a população em equilíbrio.

Quando os dados de frequência das populações da região centro-oeste foram comparados a outros estados brasileiros, observou-se um quadro heterogêneo, distinto do ocorrido na região centro-oeste. As diferenças foram observadas principalmente entre aqueles estados que apresentaram história de povoamento diferente da observada na região centro-oeste. Em metade das comparações, os três bancos de dados do centro-oeste apresentaram simultaneamente as mesmas diferenças. Os estados que apresentaram diferenças em pelo menos três marcadores com as três populações do centro-oeste foram São Paulo, Piauí, Pernambuco, Pará e Rio Grande do Sul. Provavelmente tais diferenças tenham ocorrido devido às peculiaridades no povoamento de cada um deles. Tomando um exemplo, a população da cidade de São Paulo é especialmente complexa quanto ao seu processo de formação. Além dos portugueses, africanos da África subsaariana e ameríndios, populações de outras regiões geográficas contribuíram para a composição atual da mesma, como italianos, alemães e japoneses (Bydlowski *et al.*, 2003).

Com relação à distância genética entre as populações da região centro-oeste e as parentais foram observadas algumas diferenças. A população do Estado de Goiás apresentou distância genética menor em relação ao grupo parental europeu (0,0097). Já o Distrito Federal apresentou menor distância em relação ao africano, fato decorrente do intenso fluxo migratório ocorrido nas últimas décadas, especialmente de duas regiões brasileiras, sudeste e nordeste. Essas duas regiões, nas análises prévias, apresentaram influência genética marcante quanto à parental africana. O Mato Grosso do Sul é o estado que mais se aproximou do grupo parental ameríndio. Dados históricos e de levantamento populacional mostram que a presença ameríndia nesse estado é mais expressiva que no estado de Goiás (Ribeiro, 1996).

A mistura genética reflete os resultados apresentados nas análises de distância. Todos os estados tiveram o predomínio da parental européia na sua composição, refletindo o que a história tem mostrado. Das três populações analisadas, Goiás foi o estado que apresentou maior contribuição européia e o Distrito Federal foi a população com a menor contribuição. Com relação à contribuição parental africana, verifica-se que a população do Distrito Federal apresentou o dobro quando comparado com Goiás ou Mato Grosso do Sul, devido, provavelmente, ao crescimento populacional acentuado nos últimos anos, recebendo migrantes principalmente das regiões sudeste e nordeste. As populações do sudeste e nordeste (capítulo 2), apresentam percentual de contribuição africana superior às demais regiões brasileiras. Portanto, a diferença encontrada não foi ocasionada pelo início da colonização da região centro-oeste, tendo em vista que o Distrito Federal pertencia ao Estado de Goiás inicialmente e até a década de 50 era uma região pouco populosa. Essa diferença também pode ser devido ao fato do banco de dados da população do Distrito Federal, utilizado nessa análise, ser proveniente de um banco de dados público.

Quanto ao estado do Mato Grosso do Sul, a contribuição ameríndia é similar à do Distrito Federal, mas superior à de Goiás que apresentou apenas 3%. Esse resultado era esperado tendo em vista que os grupos ameríndios que existiam à época da colonização se viram obrigados a migrarem mais para o oeste para fugir das guerras e da fome, face ao avanço inicialmente de garimpeiros e posteriormente de criadores de gado (Palacín e Moraes, 1994; Ribeiro, 1996). Como consequência, Mato Grosso do Sul, além de apresentar maior quantidade de comunidades indígenas do que Goiás apresenta, conseqüentemente, maior contribuição ameríndia na população miscigenada.

Os resultados de mistura obtidos para as populações da região centro-oeste se assemelham a algumas populações brasileiras, mas em relação a outras existem contraste

(Arpini-Sampaio *et al.*, 1999; Guerreiro & Chautard-Freire-Maia, 1988; Santos *et al.*, 1987), como é o caso da população do Piauí, Amazonas e Pará. Essas populações apresentam maior contribuição ameríndia do que africana, sendo que a contribuição parental europeia estimada foi menor do que a encontrada nesse trabalho. Por outro lado, estimativas de mistura obtidas em uma população do interior do estado de São Paulo mostraram semelhanças com os dados aqui obtidos, ou seja, alta contribuição europeia, com intermediária contribuição africana e baixa contribuição ameríndia (Ferreira *et al.*, 2006).

Como se pode notar, a população brasileira é complexa quanto à sua formação étnica. Cada região geográfica possui peculiaridades que as tornam modelos em estudos de estimativas de mistura étnica. Esse tipo de análise ajuda a compreender o processo de formação histórica de cada população e com isso, se evita generalizações que podem mascarar informações pertinentes à sua constituição.

CAPÍTULO 4

Constituição genética de Goiás e Distrito Federal, duas populações da região Centro-Oeste Brasileira, com relação a marcadores do tipo AIMS autossômicos

4. 1. INTRODUÇÃO

A utilização de marcadores de DNA para estudos populacionais tem aumentado muito nos últimos tempos, principalmente em decorrência da descrição da seqüência do genoma humano (Venter *et al.*, 2001). Esse fato permitiu maior facilidade para o estudo de novos marcadores e abriu um leque para a sua utilização não só em estudos populacionais, mas também para aplicações jurídicas, que envolvem a identificação de indivíduos. Os marcadores do tipo microssatélites, já comentados nos capítulos anteriores, constituem uma das categorias de marcadores largamente utilizado na área forense, com objetivo de auxiliar a polícia a desvendar crimes e a identificar pessoas desaparecidas, além de serem amplamente utilizados nos estudos de genética de populações.

Por isso, a maioria dos microssatélites disponíveis na literatura são altamente polimórficos, ou seja, apresentam muitos alelos tanto nas populações parentais quanto naquelas em que já houve miscigenação entre dois ou mais grupos populacionais. Pode ocorrer, entretanto, que um alelo seja mais freqüente em uma dada população do que em outra (Marjanovic *et al.*, 2004), mas de um modo geral os microssatélites utilizados nas análises de identificação humana apresentam distribuição de freqüências alélicas pouco discrepantes (Ruitberg *et al.*, 2001). Embora os microssatélites sejam bastante conhecidos e utilizados, existem inúmeras outras classes de marcadores que também são úteis em estudos populacionais, como é o caso das inserções *Alu*, inserções/deleções (indels), e os polimorfismos únicos de nucleotídeos (SNPs).

A identificação de marcadores que permitem uma melhor definição de grupos populacionais e que mostrem diferenças entre eles é fundamental para estudos comparativos, tendo em vista que as diferenças em populações humanas são pequenas quando comparadas com outras espécies animais (Barbujani, 2005). Identificar marcadores que possibilitem acessar essas diferenças, mesmo que mínimas, é um passo importante na busca de explicações para as variações existentes entre os diferentes grupos populacionais humanos. Sabe-se que tais diferenças são decorrentes de fatores evolutivos como a seleção, que pode favorecer um alelo em detrimento do outro, o isolamento devido a grandes distâncias geográficas, do cruzamento preferencial (gênero) entre grupos étnicos, ou em virtude de fatores culturais, gerando, com isso, variações nas freqüências dos alelos (Benn-Torres *et al.*, 2008; Cavalli-Sforza, 1996).

Os marcadores que apresentam essas características, isso é, que apresentam alelos com grandes diferenciais de freqüência entre populações, são chamados marcadores

informativos de ancestralidade - AIMs (Shriver *et al.*, 2003), também conhecidos como alelos específicos de população - PSAs (Shriver *et al.*, 1997). A característica marcante dos AIMs é que eles possuem um diferencial de frequência, em geral, maior que 50%, entre duas populações definidas geográfica ou etnicamente. A medida usada para estimar esta diferença é o δ , que é o valor absoluto de cada diferença de frequência alélica entre duas populações, mais a soma total de todos os alelos do marcador dividido por dois (Collins-Schramm *et al.*, 2002). Vários marcadores que apresentam essa característica já foram identificados, sendo que os primeiros categorizados como tal foram três marcadores - FyNull, Rb2300 e LPL - com altos diferenciais de frequências entre africanos e europeus e africanos e ameríndios e três *loci* - CKMM, PV92 e DRD2 - que apresentaram altos diferenciais de frequências entre ameríndios e dois outros grupos étnicos, europeus e africanos (Shriver *et al.* 1997). Os marcadores do tipo AIMs pertencem a qualquer classe de marcadores, isto é, qualquer marcador genético pode, a princípio, apresentar diferenças de frequências entre populações, e isso depende apenas do comportamento do marcador em uma população específica. Como exemplo de AIM podem ser citados os marcadores protéicos, inserções *Alu*, inserções/deleções e SNPs.

Um exemplo da utilização desse tipo de marcador genético é o mapeamento de genes associados a alguma característica como, por exemplo, a relação entre ancestralidade individual e a pigmentação da pele (Bonilla *et al.*, 2004a). Estudos são realizados também com objetivo de relacionar características próprias de determinados grupos parentais, como a cor da pele, e traços relacionados a doenças complexas como a obesidade (Bonilla *et al.*, 2004b). Outro exemplo é estimar mistura para inferir subestruturação populacional decorrente de fluxo gênico entre grupos geográficos distintos, como é o caso da população americana (Parra *et al.*, 1998). Caracterizações de populações sul-americanas utilizando esses marcadores também são encontradas na literatura, sendo, porém, ainda infreqüentes, como, por exemplo, um estudo realizado na Antioquia, localizada na Colômbia (Bedoya *et al.*, 2006). Esse estudo utilizou dados de oito AIMs (AT3, DRD2 TaqID, LPL, ICAM1, Sb19.3, APOA1 e FyNull) para estimar a contribuição das populações parentais que contribuíram para formar aquela população.

No Brasil os principais alvos de estudo desses marcadores tem sido o relato de frequência em ameríndios, uma das populações parentais brasileiras, e na caracterização de populações não ameríndias. Com relação a populações ameríndias, foi caracterizada a distribuição das frequências alélicas de sete *loci* AIMs bialélicos em uma amostra de 309 indígenas não aparentados, pertencentes a quatro tribos da Amazônia Brasileira (*Tikuna*,

Kaxináwa, Baniwa e Kanamarí). Os resultados indicaram que a distribuição das frequências alélicas, para esses marcadores, era muito similar às observadas para os indígenas da América do Norte e muito diferente a dos africanos (Luizon *et al.*, 2008).

A análise desses marcadores em populações não ameríndias tem sido realizada na busca de uma melhor descrição da contribuição das populações parentais na constituição genética atual de populações urbanas principalmente. A avaliação de uma população urbana da região sudeste do Brasil quanto a três inserções *Alu* (TPA25, PV92 e APO) indicou que esta apresenta alto percentual de contribuição européia (78%), baixo de africana (22%) e ausência de ameríndia (Mendes-Jr *et al.*, 2001). A incidência de ancestralidade africana em indivíduos auto-declarados brancos de quatro regiões brasileiras - Norte, Nordeste, Sudeste e Sul - foi verificada utilizando AIMs mostrando que o processo de miscigenação nessas quatro regiões geográficas não foi homogêneo e que o índice de ancestralidade africana (AAI) varia de 13% na região sul a 32% na sudeste (Parra *et al.*, 2003). Com a análise dos nove alelos específicos de população, a cidade de Salvador, localizada no estado da Bahia, apresentou uma maior contribuição africana (49,2%), seguida de européia (36,3%) e ameríndia (14,5%), sendo que as regiões da cidade com classe socioeconômica mais baixa apresentaram uma maior contribuição africana para os dados de estimativa de mistura (Machado, 2008).

Esses marcadores foram também utilizados na busca de correlação entre ancestralidade genômica e classificação fenotípica, em especial cor da pele. O estudo foi desenvolvido em uma comunidade do Vale do Jequitinhonha em Minas Gerais. Os indivíduos foram classificados de acordo com suas características fenotípicas em negros e mulatos, em contraste com indivíduos autodeclarados como brancos. De acordo com os resultados, há uma proporção similar de ancestralidade não africana dentre os negros e dentre os intermediários (mulatos): 45% e 48% respectivamente (Parra *et al.*, 2003), isso é, não foi possível discriminar os indivíduos de acordo com suas características de cor da pele. Em suma, de acordo com esse estudo a cor da pele é um pobre indicador de ancestralidade genômica.

Por outro lado, em Salvador o quadro é distinto. Utilizando os mesmos marcadores, foi observada uma grande coerência entre ancestralidade genômica e características fenotípicas. Os resultados indicaram que os indivíduos classificados como negros apresentam uma ancestralidade genômica africana muito maior do que os classificados como brancos e que os de fenótipo intermediários apresentam uma gradação na contribuição africana (Bomfim, 2008).

Todos os estudos acima apresentados mostram que os marcadores genéticos informativos de ancestralidade são eficientes para identificar populações definidas geográfica e etnicamente e, com isso, permitem estimar o percentual de contribuição de grupos parentais na composição de populações miscigenadas.

Este capítulo teve o objetivo de ampliar os conhecimentos acerca do processo de formação da população da região centro-oeste brasileira. Para tanto, duas populações desta região, Goiás e Distrito Federal, foram caracterizadas utilizando marcadores genéticos, dos tipos inserção *Alu*, indel e SNP, que apresentaram diferencial de frequência superior a 30% entre pelo menos duas populações parentais que contribuíram para a formação atual dessas populações.

4. 2 MATERIAL E MÉTODOS

Caracterização da população e coleta das amostras

A amostra analisada nesse trabalho foi coletada no Estado de Goiás e no Distrito Federal. A coleta e utilização do material foram aprovadas pelo CEP da Universidade Federal de Goiás, da Faculdade de Medicina da Universidade de Brasília e CONEP, no caso de Goiás, e pelo CEP da Faculdade de Saúde da Universidade de Brasília e CONEP para o Distrito Federal.

Como material biológico foi utilizado sangue venoso coletado com sistema à vácuo com EDTA como anticoagulante. Os critérios básicos para a escolha dos voluntários foram: aleatoriedade respeitando-se os critérios estabelecidos para a participação do presente estudo, isto é, idade acima de 18 (dezoito) anos ou com autorização paterna/materna, e que tivessem residência fixa na cidade/região administrativa de coleta no momento do nascimento.

Distrito Federal

O Distrito Federal, localizado no Planalto Central do Brasil, tem uma população estimada em 2.455.903 habitantes (Censo 2007 – IBGE), sendo que aproximadamente metade dos habitantes é nascida na própria região e metade são pessoas advindas de todas as regiões do Brasil, além de uma pequena participação de estrangeiros, o que torna o brasileiro um caso único quanto a sua formação. Administrativamente, é composto por 27 regiões (considerando Lago Sul e Lago Norte como regiões distintas). Entretanto, à época da coleta eram 19 regiões, sendo que oito outras regiões foram criadas desde então: Águas

Claras (2003), Varjão (2003), Sudoeste/Octogonal (2003), Park Way (2003), Setor Complementar de Indústria e Abastecimento (2004), Sobradinho II (2004), Jardim Botânico (2004) e Itapoã (2005) (sítio oficial do Governo do Distrito Federal – www.districtofederal.df.gov.br acessado em 21 de abril de 2008).

Para este trabalho foram amostradas 12 das 19 regiões (Tabela 4.2) existentes em 2002, ano da coleta das amostras, sendo que dois indivíduos eram oriundos do entorno do Distrito Federal. A informação concernente à região administrativa onde habitavam os pais na época da coleta foi obtida durante a entrevista, dado esse que permitiu o agrupamento dos indivíduos por região. Pouco mais de 10% dos indivíduos não forneceram a informação, principalmente por desconhecimento. Por outro lado, seis das regiões não amostradas não existiam e não eram povoadas antes de 1986, como Samambaia (oficialmente fundada em 1989) e 1993, como Santa Maria, Recanto das Emas, Riacho Fundo e São Sebastião. Com relação a Brazlândia, essa é a região administrativa mais distante do Plano Piloto. Candangolândia é a região administrativa menos povoada, representando aproximadamente 0,7% da população. O mapa da Figura 4.1 representa as regiões administrativas existentes na época da coleta.

As amostras de material biológico e os dados foram coletados no Laboratório de Genética da Universidade de Brasília e no curso preparatório para o Vestibular “Alunos da UnB” (Alunb) no ano de 2002. Paralelamente à coleta de material biológico, foi feito um levantamento de dados demográficos utilizando um questionário (Anexo 3) respondido por todos os indivíduos que tiveram seu sangue coletado. O objetivo do questionário foi obter informações referentes a sexo, idade, ancestralidade, incluindo no mínimo uma geração ascendente, dentre outros.

A amostra foi composta por 73 homens (41,1%) e 105 mulheres (58,9%) e a média da idade dos participantes foi de 21,1 anos (desvio padrão = 3,9 anos). Com relação à autodeclaração de etnia, do total de 178 indivíduos, 122 (68,5%) se autodeclararam como mestiço entre pelo menos duas das parentais que formaram a população, isto é, branco, negro e índio, 44 (24,7%) como brancos, três (1,7%) como negro e nove (5,1%) não informaram sua etnia.

Tabela 4.2. Distribuição dos indivíduos componentes da amostra do Distrito Federal com relação à região administrativa da residência dos pais na época do nascimento.

Região administrativa	Número de indivíduos
Brasília	59
Ceilândia	9
Cruzeiro	13
Entorno	2
Gama	7
Guará	10
Lago Norte	7
Lago Sul	3
Núcleo Bandeirante	5
Paranoá	1
Planaltina	14
Sobradinho	12
Taguatinga	17
Não especificou	19
TOTAL	178



Figura 4.1. Regiões Administrativas que compunham o Distrito Federal em 2002 (Fonte: Codeplan, 1997), época da coleta de material biológico.

Goiás

O Estado de Goiás está situado na Região centro-oeste e está subdividido em 18 microrregiões administrativas, apresentando uma população estimada em 5.644.460 (Censo, 2007 – IBGE). O planejamento da coleta das amostras utilizadas nesse estudo e a forma da coleta foram discutidos no capítulo 3. Para esse estudo foram sorteadas amostras de sangue provenientes de 202 indivíduos do banco de amostras citado no capítulo 3, buscando a representação proporcional ao tamanho de cada microrregião (Tabela 4.2).

Tabela 4.2. Caracterização das mesorregiões, microrregiões quanto ao número de habitantes de cada microrregião a época da coleta das amostras de sangue e do número de indivíduos coletados em cada microrregião do Estado de Goiás.

Mesorregião	Microrregião	População	Número indivíduos analisados
Noroeste Goiano	São Miguel do Araguaia	68.888	3
	Rio Vermelho	91.254	4
	Aragarças	55.263	2
Norte Goiano	Porangatu	226.766	7
	Chapada dos Veadeiros	55.002	2
Centro Goiano	Ceres	212.661	10
	Anápolis	451.474	19
	Iporá	64.030	3
	Anicuns	103.276	4
	Goiânia	1.633.899	64
Leste Goiano	Vão do Paranã	88.913	4
	Entorno de Brasília	766.535	34
Sul Goiano	Sudoeste de Goiás	330.490	14
	Vale do Rio dos Bois	102.491	5
	Meia Ponte	305.427	14
	Pires do Rio	76.856	3
	Catalão	114.686	6
	Quirinópolis	91.190	4
Total		4.839.101	202

Fonte: contagem da população IBGE – 1996.

A amostra foi composta por 126 homens (62,4%) e 76 mulheres (37,6%) e a média da idade dos participantes foi de 32,3 anos (desvio padrão = 9,7 anos). Com relação à autodeclaração de etnia, do total de 202 indivíduos, 141 (69,8%) se autodeclararam como mestiços entre pelo menos duas das parentais que formaram a população, isto é, branco, negro e índio, 54 (26,7%) como brancos, quatro (2%) como negro e 3 (1,5%) não informaram sua etnia.

Procedimentos metodológicos

Seleção dos marcadores genéticos

Foram selecionados 14 marcadores genéticos que podem ser categorizados como AIM por apresentarem diferencial de frequência entre as populações parentais acima de 30% (tabela 4.3), sendo: sete inserções Alu (APO, D1, ECA, FXIII, PV92, Sb19.3, TPA25), uma inserção/deleção (AT3) e seis SNPs (DRD2-A, FyNull, GC, LPL, OCA2 e Rb2300). Onze desses apresentaram diferencial de frequência entre os grupos parentais (europeu, africano e ameríndio) maior que 50%. Os demais marcadores apresentaram diferencial de frequência menor que 50% (DRD2-A, Sb19.3 e TPA25), (tabela 4.3). Sete dos marcadores apresentam alto diferencial de frequência entre africano e os dois outros grupos parentais (APO, AT3, D1, FyNull, GC, LPL e Rb2300), dois entre africano e europeu (OCA2 e Sb19.3), três marcadores com diferencial de frequência entre o grupo parental ameríndio e os outros dois grupos parentais (ECA, FXIII e PV92), um entre ameríndio e europeu (DRD2-A) e um entre ameríndio e africano (TPA25).

Extração de DNA e verificação em gel de agarose

À época da coleta, as amostras do estado de Goiás foram submetidas à extração de DNA por meio do kit de extração GFX™ *Genomic Blood DNA Purification Kit* (Pharmacia), seguindo as recomendações do fabricante – (Anexo 2), enquanto que as amostras do Distrito Federal foram submetidas à extração de DNA pela técnica *Salting out* (Miller *et al.*, 1988) com modificações (Anexo 2). Para o desenvolvimento do presente estudo as amostras que necessitaram ser re-extraídas, foram submetidas à extração pelo método de *Salting out*. Após extração, as amostras de DNA foram conservadas em freezer a 20°C negativos até o momento do uso.

Para verificação da qualidade e quantidade de DNA por amostra, foi utilizado gel de agarose a 1% com brometo de etídio para visualização, utilizando um transiluminador

de luz ultravioleta. O preparo das amostras para quantificação foi realizado de acordo com o descrito no anexo 2. Após a quantificação, as amostras de DNA foram diluídas para 50ng/ μ l e armazenadas em freezer a 20°C negativos até o momento do uso.

Tabela 4.3: Diferenças nas frequências alélicas (δ) entre os grupos parentais (africano, europeu e ameríndio), dos 14 marcadores autossômicos analisados.

<i>Locus</i>	Diferencial de frequência (δ)		
	AFR-EUR	AFR-AME	EUR-AME
APO	0,496	0,534	0,038
AT3	0,586	0,804	0,218
D1	0,389	0,544	0,155
DRD2-A	0,275	0,146	0,421
ECA	0,090	0,465	0,375
FXIII B	0,335	0,841	0,506
FyNull	0,978	0,983	0,005
GC1-F	0,673	0,486	0,187
GC1-S	0,509	0,453	0,056
LPL	0,473	0,507	0,034
OCA2	0,620	0,317	0,303
PV92	0,001	0,662	0,663
Rb2300	0,569	0,749	0,180
Sb19.3	0,395	0,206	0,189
TPA25	0,229	0,300	0,071

Legenda: AFR (africano), AME (ameríndio), EUR (europeu).

Reação em Cadeia da Polimerase (PCR) e visualização dos produtos de amplificação

A tabela 4.4 apresenta os marcadores autossômicos utilizados nesse estudo. Foram ao todo 14 marcadores. As reações de PCR foram preparadas para um volume final de 12,5 μ l para cada marcador. A cada conjunto de reações foi incluído controle negativo, isto é, no lugar da amostra de DNA foi acrescentada água milli-Q. As concentrações dos

reagentes foram levemente distintas entre os marcadores (Tabela 1, Anexo 2) e para cada marcador genético foi utilizado um conjunto de iniciadores específicos (Tabela 4.4).

Na mistura da reação preparada foi adicionado o DNA correspondente a cada amostra, seguido de ligeira homogeneização e depois acrescentado uma gota de óleo mineral para prevenir evaporação dos reagentes. Os microtubos contendo todos os reagentes foram colocados em um termociclador MJ Research – PTC 100, programado para as seguintes condições de reação: desnaturação inicial de 5 minutos a 94°C; ciclos de desnaturação/pareamento/extensão das amostras de DNA por 30 ciclos, de acordo com as temperaturas especificadas para cada marcador (tabela 4.4); extensão final por 10 minutos a 72°.

Reação de restrição

Conforme apresentado na tabela 4.4, seis dos marcadores genéticos estudados - GC, FYNull, LPL, OCA2, Rb2300 e DRD2-A - foram identificados pela avaliação do produto da digestão com enzima de restrição após PCR, ou seja, tais SNPs foram detectados utilizando-se a técnica de PCR-RFLP. Esta técnica consistiu na amplificação específica dos *loci* de interesse, seguida pela digestão por endonucleases de restrição específica. Assim, um indivíduo pode apresentar presença do sítio de restrição em um cromossomo apenas (heterozigoto), nos dois (homozigoto para presença do sítio) ou ainda ausência do sítio de restrição (homozigoto para ausência) no fragmento que foi amplificado pela PCR.

A mistura de reagentes para a reação de restrição dos produtos de PCR está apresentada na Tabela 2, do Anexo 2. As reações foram levadas ao banho-maria a uma temperatura apropriada para a atividade da enzima por tempos variáveis, conforme a necessidade de cada marcador. Passado este período, as amostras foram submetidas a choque térmico, ou seja, colocadas em gelo com objetivo de inativar a enzima, interrompendo, assim, o processo de restrição. Até o momento de realizar a eletroforese, as amostras foram conservadas em geladeira a uma temperatura de 4°C. Condições de restrição, enzimas, temperatura e tempo de restrição para os marcadores SNPs analisados estão descritos na tabela 4.6.

Tabela 4.4. Características dos marcadores genéticos do tipo AIMs analisados: tipo de marcador, modo de identificação, localização cromossômica, seqüência dos iniciadores, temperatura de pareamento na PCR e referência.

<i>Locus</i>	Tipo	Identificação	Localização cromossômica	Seqüência	Temp. anelamento*	Referência
APO	Ins Alu	tamanho	11q23	5' AAGTGCTGTAGGCCATTTAGATTAG 3' 5' AGTCTTCGATGACAGCGTATACAGA 3'	50	Karathanasis, 1985; Batzer <i>et al.</i> , 1996;
AT3- I/D	68pb ins/del	tamanho	1q25.1	5' CCACAGGGTGTAACATTGTGT 3' 5' GAGATAGTGTGATCTGAGGC 3'	55	Liu <i>et al.</i> , 1995
D1	Ins Alu	tamanho	3	5' TGC TGA TGC CCA GGG TTA GTA AA 3' 5' TTT CTG CTA TGC TCT TCC CTC TC 3'	58	Batzer <i>et al.</i> , 1995; Batzer <i>et al.</i> , 1996
DRD2	SNP	RFLP -TaqI A	11q22-q23	5' CCG TCG ACG GCT GGC CAA GTT GTC TA 3' 5' CCG TCG ACC CTT CCT GAG TGT CAT CA 3'	64	Granty <i>et al.</i> , 1989; Gelernter <i>et al.</i> , 1998
ECA	Ins Alu	tamanho	17q23	5' CTG GAG ACC ACT CCC ATC CTT TCT 3' 5' GAT GTG GCC ATC ACA TTC GTC AGA T 3'	58	Jeunemaitre <i>et al.</i> , (1992); Batzer <i>et al.</i> , 1996
FXIIIB	Ins Alu	tamanho	1q31-32	5' TCA ACT CCA TGA GAT TTT CAG AAG T 3' 5' CTG GAA AAA ATG TAT TCA GGT GAG T 3'	60	Webb <i>et al.</i> , 1989; Batzer <i>et al.</i> , 1996
FyNull	SNP	RFLP -StyI	1q23.2	5' AGGCTTGTGCAGGCAGTG 3' 5' GGCATAGGGATAAGGGACT 3'	55	Tournamille <i>et al.</i> , 1995
GC	SNP	RFLP -StyI + e RFLP -HaeIII	4q13q21.1	5' AGATCTGAAATGGCTATTATTTTGC 3' 5' GGAAGGTGAGTTTATGGAACAGC 3'	55	McCombs <i>et al.</i> (1986); Parra <i>et al.</i> , 1998

Continuação tabela 4.4. Características dos marcadores genéticos do tipo AIMs analisados: tipo de marcador, modo de identificação, localização cromossômica, seqüência dos iniciadores, temperatura de pareamento na PCR e referência.

<i>Locus</i>	Tipo	Identificação	Localização cromossômica	Seqüência	Temp. anelamento*	Referência
LPL	SNP	RFLP -PvuII	8p21.3	5' AGGCTTCACTCATCCGTGCCTCC 3' 5' TTATGCTGCTTTAGACTCTTGTC 3'	60	Gotoda <i>et al.</i> , 1992
OCA2	SNP	RFLP -HaeIII	15q13.1	5' CTTTCGTGTGTGCTAACTCC 3' 5' ACCTCTAGCATGGTTCTTGGGC 3'	60	Lee <i>et al.</i> , 1995
PV92	Ins Alu	tamanho	16q23.3	5' AACTGGGAAAATTTGAAGAGAAAGT 3' 5' TGAGTTCTCAACTCCTGTGTGTTAG 3'	54	Batzer <i>et al.</i> , 1994; Batzer <i>et al.</i> , 1996
Rb2300	SNP	RFLP -BamHI	13q14.2	5' CAGGACAGCGGCCCGGAG 3' 5' CTGCAGACGCTCCGCCGT 3'	60	Bookstein <i>et al.</i> , 1990
Sb19.3	Ins Alu	tamanho	19p12	5' TCTAGCCCCAGATTTATGGTAACTG 3' 5' AAGCACAATTGGTTATTTTCTGAC 3'	60	Arcot <i>et al.</i> , 1998; Parra <i>et al.</i> , 1998
TPA25	Ins Alu	tamanho	8p12-q11.2	5' GTA AGA GTT CCG TAA CAG AGC T 3' 5' CCC CAC CCT AGG AGA ACT TCT CTT T 3'	58	Yang-Feng <i>et al.</i> , 1986; Batzer <i>et al.</i> , 1996

* Em graus Celsius

Eletroforese e avaliação fenotípica

Os produtos de PCR, no caso dos marcadores do tipo inserção (AT3, APO, PV92, ECA, TPA25, D1, FXIII B e SB19.3), ou das digestões, no caso dos SNPs (DRD2-A, OCA2, FyNull, GC, RB2300 e LPL), foram analisados em negatoscópio após eletroforese vertical em gel de poliacrilamida corado com solução de nitrato de prata a 10%. A concentração do gel de poliacrilamida do tipo não desnaturante utilizado para os marcadores RB2300, LPL, FyNull e AT3 foi 10% e os demais marcadores 6% (anexo 2).

Os produtos da PCR foram preparados para a corrida eletroforética adicionando-se a cada 5 μ l de amostra 2 μ l de tampão de corrida. As amostras foram aplicadas no gel, sendo que no poço central foi colocado padrão de peso molecular de 100pb como referência. Decorrido o tempo de corrida, os géis foram submetidos à coloração com nitrato de prata. As soluções utilizadas para a eletroforese, bem como as soluções para coloração estão descritas no anexo 2. A leitura do gel foi feita em negatoscópio por comparação das bandas obtidas com a do marcador de peso molecular de 100pb.

A distribuição das frequências gênicas e genotípicas das inserções Alu APO, D1, PV92, ECA, FXIII B e TPA25 na população de Goiás foram descritas por Gigonzac (2002) e Vieira (2003) em dissertação de mestrado realizado na Universidade Federal de Goiás. Para tais estudos, os referidos pesquisadores utilizaram 205 indivíduos do banco de amostras de Goiás, porém apenas 100 dessas amostras coincidiram com as amostras por nós selecionadas. Por isso, os resultados descritos por Gigonzac (2002) e Vieira (2003) foram utilizados parcialmente (genótipos de 100 indivíduos) e, para as demais amostras, a genotipagem foi realizada no presente trabalho.

Análise estatística

A frequência alélica dos marcadores foi calculada por meio do método de contagem gênica. Os valores de heterozigose observada e esperada foram calculados para cada *locus*. Teste exato foi utilizado para checar a consistência das frequências genotípicas observadas em cada população para cada marcador com aquelas esperadas segundo a Lei de Hardy-Weinberg, tendo como pressuposto (hipótese nula) a de união aleatória dos gametas. Uma probabilidade exata menor que 0,05 indicou uma quantidade estatisticamente significativa de desequilíbrio. As estimativas dos valores exatos de *p* foram computadas pelo método da Cadeia de Markov, com 10000 dememorizações. Tais análises foram realizadas no programa GENEPOP.

Foi testado também a hipótese nula de independência dos genótipos de um *locus* em relação a outro, isto é, de equilíbrio de ligação. Esse é um indicador importante para avaliar se existe subdivisão dentro da população. A análise de desequilíbrio de ligação foi realizada também considerando todos os *loci* simultaneamente.

Ambas as análises, teste para Equilíbrio de Hardy-Weinberg e desequilíbrio de ligação foi utilizado o teste exato para testar a hipótese de equilíbrio, sendo que 10.000 permutações foram realizadas em cada teste. A medida *chi-square* foi utilizada com objetivo de verificar se uma combinação em particular resultou em um valor de *chi-square* mais extremo do que aqueles produzidos pelos dados originais. Essa abordagem permutacional usa todas as informações genóticas (ou fenóticas) de um indivíduo em uma amostra para determinar a significância do desequilíbrio por permutações independentes. Sob a hipótese nula de não desequilíbrio, qualquer genótipo (ou fenótipo) observado em um *locus* (alelo) é igualmente provável ocorrer com qualquer outro genótipo em um segundo *locus* (alelo). Portanto, qualquer conjunto de genótipo reconstruído ao acaso com o mesmo número de indivíduos e as mesmas frequências genóticas é igualmente provável de ocorrer apenas ao acaso. O desequilíbrio de ligação entre os *loci* par a par e multilocos, para cada população, foram determinados utilizando o programa GDA.

A comparação genética entre as populações de Goiás e Distrito Federal foi feita por meio de teste exato de diferenciação populacional, utilizando o programa GENEPOP, e através da estatística F, realizada no programa GDA.

Para a estimativa de diferenciação populacional dois testes foram realizados: diferenciação gênica, que compara as frequências alélicas entre as duas populações, e a diferenciação genotípica que compara a frequência dos genótipos entre as mesmas. A hipótese nula para a primeira análise é de que a distribuição alélica é idêntica entre as populações. Para testar essa hipótese, para cada *locus* uma estimativa não viesada do valor de P do teste de probabilidade é realizada como descrito por Raymond e Rousset (1995). Considerando a hipótese nula para a segunda análise, isto é, de que a distribuição genotípica é idêntica nas populações, é feita uma estimativa não viesada do valor de P do teste exato baseado no *log* da verossimilhança (Goudet *et al.*, 1996).

Outro parâmetro utilizado para avaliar a existência de diferença entre as duas populações foi o *Fst* (Cockerham & Weir, 1987). A estatística F analisa três parâmetros: *Theta* (ou *Fst*), *F* (ou *Fit*) e *f* (ou *Fis*). O *f* é uma estimativa de coeficiente de endocruzamento dentro da população, indicando se há déficit de heterozigotos, tendo em

vista que numa população endogâmica há uma tendência no aumento do número de homocigotos. Theta é uma estimativa que verifica se existe subestruturação populacional através da análise da redução de heterocigose em uma subpopulação. Esse parâmetro calcula a probabilidade de dois alelos tomados ao acaso em duas populações serem idênticos por descendência. F é o coeficiente de endogamia total, considerando as duas populações, e mede a perda de heterocigose individual em relação à heterocigose total. Ou seja, a estatística F descreve a quantidade de endogamia dentro das subpopulações (*F_{st}* ou *Theta*), dentro da população total (*F_{is}* ou *f*) e entre as subpopulações (*F_{it}* ou *F*). Foi determinado também os componentes da variância Sigma-P, Sigma-I e Sigma-G que representam a variância entre as populações, entre os indivíduos dentro da amostra e no indivíduo, respectivamente. A significância estatística das distâncias genéticas de *F_{st}* foi calculada usando 10000 permutações. A estimativa foi considerada significativa se o intervalo de confiança não incluisse o zero.

A correção de Bonferroni foi utilizada para controle do nível de significância conjunto, quando testes múltiplos foram realizados.

As estimativas das proporções de mistura das populações parentais para cada população estudada foram estimadas pelo uso do método de identidade gênica (Chakraborty, 1985), implementado no programa ADMIX3. As populações parentais utilizadas nas análises foram aquelas que contribuíram para o povoamento da região, isto é, europeu, africano e ameríndio. Os dados de frequência foram compilados da literatura.

Para o grupo parental africano foram consideradas populações do oeste da África subsaariana. Os dados foram obtidos do banco de dados dbSNP NCBI database (<http://www.ncbi.nlm.nih.gov/SNP> - acessado em 03/05/2008) e do banco de dados ALFRED – *Allelic Frequency Database* (<http://alfred.med.yale.edu> – acessado em 03/05/2008) – para as populações de Nigéria, República Central Africana, Serra Leoa, Iorubá da Nigéria e Benin e Hauçás do oeste da África subsaariana - e em artigos científicos para as populações da Nigéria, São Tomé, Serra Leoa, Benin e Ruanda (Luis *et al.*, 2003; Tomás *et al.*, 2002; Parra *et al.*, 2001; Parra *et al.*, 1998; Batzer *et al.*, 1996).

Os dados para a população europeia também foram obtidos do banco de dados dbSNP NCBI database e ALFRED, sendo compilado dados de frequências alélicas da Alemanha, Espanha e Itália, e em artigos científicos para as populações de Portugal, Espanha, Inglaterra, Irlanda, Alemanha, França (Bonilla *et al.*, 2004; Luis *et al.*, 2003; Tomás *et al.*, 2002; Parra *et al.*, 2001; Comas *et al.*, 2000; Parra *et al.*, 1998; Batzer *et al.*, 1996).

Para a população parental ameríndia as frequências foram obtidas do banco de dados dbSNP NCBI database e ALFRED – tribos indígenas de *Karitiana*, *Suruí*, *Tikúna*, *Ayoreo*, *Gavião*, *Suruí*, *Wai-Wai*, *Xavante*, *Zoró*, *Aché*, *Guarani*, *Kaingang* e *Quechua* – e de dados da literatura – tribos indígenas de *Tikúna*, *Kashinawa*, *Baniwa*, *Kanamari*, *Mayas*, *Pima*, *Cheyenne*, *Pueblo*, *Quechua*, *Mataguayo*, *Toba*, *Chorote*, *Mapuche*, *Tehuelche*, *Lengua*, *Ayoreo*, (Luizon *et al.*, 2008; Bonilla *et al.*, 2004; Martinez-Marignac *et al.*, 2004; Batzer *et al.*, 1994). Dados para a inserção *Alu* D1 foram obtidos de uma tese de doutorado que analisou quatro tribos indígenas da região amazônica – *Tikúna*, *Kashinawa*, *Baniwa*, *Kanamari* (Oliveira, 1999).

O índice de ancestralidade foi estimado com objetivo de comparar as proporções de contribuição entre as parentais africana e européia nas populações do Distrito Federal e Goiás. Para determinar o índice de ancestralidade africana (IAA) nessas duas populações foi tomado como referência os dados de frequência genotípica das parentais africana e européia, para os marcadores APO, D1, Sb19.3, LPL, Rb2300, OCA2, GC e FyNull, pois eles apresentam diferencial de frequência grande entre essas duas parentais. O IAA é dado por:

$$IAA = \log_{10} \frac{PA}{PE}$$

onde, PA é a probabilidade de um dado genótipo multilocos ocorrer na população africana e PE é a probabilidade desse mesmo genótipo ocorrer na população européia, como descrito por Parra *et al.* (2003).

Para realizar esta análise, duas informações são necessárias: dados das frequências genotípicas das populações parentais e os genótipos da população a ser analisada. As estimativas das frequências genotípicas da população miscigenada são calculadas pelo uso das frequências genotípicas observadas para todos os *loci* possíveis da primeira população parental (africana). O mesmo é feito considerando a segunda população parental (européia). O valor esperado do *log* da razão de verossimilhança é então calculado para cada indivíduo somando-se os valores encontrados para cada *locus*. Com os índices obtidos é calculado então o índice de ancestralidade africana (IAA), que se baseia no logaritmo da razão entre a verossimilhança dos genótipos multilocos obtidos considerando a parental africana pela verossimilhança do genótipo encontrado na população européia. Análise de estatística descritiva foi utilizada para avaliar os dados de IAAs obtidos para cada indivíduo nas duas populações, utilizando o programa SPSS 15.0 para Windows, versão 2006 (<http://www.spss.com>). Através dessa análise foram calculados a média, mediana,

variância e os índices de assimetria (avalia a assimetria da curva) e curtose (representa o grau de concentração de valores de uma variável contínua em torno da média, tomando-se a curva normal como referência), com seus respectivos desvios padrões. A homogeneidade das variâncias, obtidas para as populações do Distrito Federal e Goiás, foi testada através do teste de significância de Levene. O teste *t* para amostras independentes foi utilizado para comparação das médias dos valores de IAAs obtidos nas duas populações, considerando nível de significância de 5%.

4. 3. RESULTADOS

As frequências alélicas, os valores de heterozigose observada e esperada, assim como os valores de *p* do teste de aderência ao Equilíbrio de Hardy-Weinberg e os respectivos desvios padrões para os 14 marcadores AIMs autossômicos analisados, para as populações de Goiás e Distrito Federal, estão apresentados na Tabela 4.5. Todos os marcadores encontravam-se em equilíbrio segundo o esperado por Hardy-Weinberg, com nível de significância de 0,05. Entretanto, a população do Distrito Federal apresentou quatro marcadores com valor de *p* próximo a 0,05 (ECA, FyNull, OCA2 e PV92) e a do estado de Goiás um marcador (LPL) nessa mesma situação.

Os valores de heterozigose observada para a população do Distrito Federal variaram de 17 a 62% e para a população do estado de Goiás de 18 a 61% (Tabela 4.5). O menor valor de heterozigose foi observado para o marcador APO em ambas as populações e o maior foi o GC por apresentar três alelos. Dez dos 14 marcadores apresentaram heterozigose acima de 40% na população do Distrito Federal e em Goiás foram 11 marcadores. O marcador que diferiu entre as duas populações foi o PV92 que apresentou heterozigose observada de 35% no DF e em Goiás de 47%.

As frequências alélicas ponderadas e o número amostral total analisado para cada um dos 14 marcadores genéticos encontrados nas populações parentais que contribuíram para o povoamento da população da região Centro-Oeste, isto é, europeu, africano e ameríndio, encontram-se na Tabela 4.5. A comparação das frequências descritas para as populações parentais e as observadas nas duas populações miscigenadas revelou valores intermediários nas populações miscigenadas em relação às parentais. Para a maioria dos marcadores estudados, as frequências alélicas encontradas nas populações miscigenadas apresentaram valores mais próximos às descritas para a parental européia. Entretanto, para quatro marcadores (DRD2-A, GC, PV92 e Rb2300), os valores apresentaram-se intermediários aos dois outros grupos parentais.

Não foi observado desequilíbrio de ligação quando todos os *loci* foram analisados simultaneamente tanto na população do Distrito Federal ($p=0,0762$) quanto em Goiás ($p=0,4387$), considerando nível de significância de 0,05. Entretanto, na análise de desequilíbrio de ligação par a par, das 91 combinações possíveis foram observados 11 pares de combinações em desequilíbrio na população do Distrito Federal e cinco pares na de Goiás (Tabela 4.6). Quando se fez a correção Bonferroni para testes múltiplos, o valor ajustado foi estimado em 0,0005 e com isso apenas uma combinação permaneceu com desvio, o par AT3/FyNull para o Distrito Federal.

A Tabela 4.7 apresenta os valores da estatística F (f, F e Theta-P, que correspondem ao Fis, Fit e Fst respectivamente) entre as duas populações para os 14 *loci* AIMs analisados. De um modo geral, os valores da estatística F foram baixos. O valor de Theta-P total foi baixo, e os valores de f e F similares, sendo que sete marcadores apresentaram valores negativos indicando exogamia. Três marcadores apresentaram valores de f mais elevados (FXIII, AT3 e APO). Sigma-P, Sigma-I e Sigma-G, os componentes da variância de cada marcador, mostraram que as maiores variações foram decorrentes das variações individuais. As variâncias decorrentes da diferença entre as populações e entre os indivíduos dentro da amostra foram baixas. Os marcadores que apresentaram a menor variação individual foi a inserção Alu APO (0,173) e o SNP FyNull (0,292), e os marcadores que contribuíram para maior variação entre os indivíduos foram os *loci* GC (0,617) e o LPL (0,538). Não houve diferença entre as populações, de acordo com o teste de significância para *Fst* (valor de Theta-P variou de 0,003466 a - 0,000003), considerando nível de significância de 5%.

Na Tabela 4.8 estão apresentadas as análises de diferenciação das frequências gênicas e genotípicas entre as duas populações. Não foram observadas diferenças entre as duas populações no que se refere à distribuição das frequências alélicas ($p=0,1683$) e na distribuição genotípica ($p=0,1912$) quando todos os marcadores foram considerados simultaneamente, em um nível de significância de 0,05. Na análise para cada *locus*, foram observadas diferenças estatisticamente significativas nas distribuições das frequências gênicas e genotípicas para o marcador DRD2-A, porém após a correção de Bonferroni ($p=0,0035$) os valores não apresentaram significância estatística.

Tabela 4.5: Frequências alélicas, valores de heterozigose observada e esperada, valores de p do teste de aderência ao Equilíbrio de Hardy-Weinberg e respectivos desvios padrões com relação aos 14 marcadores AIMs autossômicos analisados para as populações de Goiás e Distrito Federal e nas três populações parentais.

	APO	AT3	D1	DRD2-A	ECA	FXIIB	FyNull	GC	LPL	OCA2	PV92	Rb2300	Sb19.3	TPA25
Distrito Federal														
f^*1	0,905	0,364	0,380	0,392	0,444	0,398	0,793	0,478	0,553	0,662	0,267	0,509	0,787	0,512
f^*2	0,095	0,636	0,620	0,608	0,556	0,602	0,207	0,308	0,447	0,338	0,733	0,491	0,213	0,488
f^*3								0,214						
N	168	169	166	167	169	166	162	161	159	164	167	167	167	168
Hobs	0,17	0,42	0,43	0,52	0,57	0,45	0,28	0,62	0,52	0,53	0,35	0,46	0,34	0,49
Hesp	0,17	0,46	0,47	0,48	0,50	0,48	0,33	0,63	0,50	0,45	0,39	0,50	0,34	0,50
p	0,646	0,247	0,321	0,258	0,063	0,415	0,055	0,448	0,634	0,053	0,063	0,277	>0,999	0,760
Desvio padrão	0,001	0,002	0,002	0,002	0,002	0,003	0,001	0,004	0,002	0,001	0,002	0,003	<0,001	0,002
Goiás														
f^*1	0,885	0,311	0,350	0,311	0,378	0,458	0,812	0,482	0,570	0,644	0,312	0,468	0,784	0,540
f^*2	0,115	0,689	0,650	0,689	0,622	0,542	0,188	0,290	0,430	0,356	0,688	0,532	0,216	0,460
f^*3								0,228						
N	196	188	193	201	200	189	197	202	192	202	202	201	199	202
Hobs	0,18	0,40	0,51	0,41	0,48	0,44	0,31	0,61	0,56	0,47	0,47	0,52	0,34	0,47
Hesp	0,20	0,43	0,46	0,43	0,47	0,50	0,31	0,63	0,49	0,46	0,43	0,50	0,34	0,50
p	0,145	0,393	0,116	0,623	>0,999	0,109	>0,999	0,937	0,077	0,880	0,257	0,673	>0,999	0,396
Desvio padrão	0,001	0,003	0,002	0,002	<0,001	0,002	<0,001	0,001	0,003	0,001	0,002	0,003	<0,001	0,003

Cont. 4.5: Frequências alélicas, valores de heterozigose observada e esperada, valores de p do teste de aderência ao Equilíbrio de Hardy-Weinberg e respectivos desvios padrões com relação aos 14 marcadores AIMs autossômicos analisados para as populações de Goiás e Distrito Federal e nas três populações parentais.

	APO	AT3	D1	DRD2-A	ECA	FXIIIB	FyNull	GC	LPL	OCA2	PV92	Rb2300	Sb19.3	TPA25
DF e GO														
f^*1	0,894	0,652	0,364	0,348	0,408	0,430	0,804	0,481	0,563	0,652	0,291	0,486	0,786	0,527
f^*2	0,106	0,348	0,636	0,652	0,592	0,570	0,196	0,298	0,437	0,348	0,709	0,514	0,214	0,473
f^*3								0,222						
N	364	357	359	368	369	355	359	363	351	366	369	368	366	370
Hobs	0,16	0,41	0,48	0,46	0,52	0,44	0,29	0,62	0,54	0,46	0,41	0,49	0,34	0,48
Hesp	0,17	0,45	0,46	0,45	0,48	0,49	0,32	0,63	0,49	0,49	0,41	0,50	0,34	0,50
p	0,101	0,122	0,648	0,819	0,200	0,064	0,180	0,642	0,083	0,170	>0,999	0,677	0,876	0,408
Desvio padrão	0,001	0,003	0,003	0,002	0,003	0,002	0,002	0,005	0,002	0,003	<0,001	0,003	0,001	0,004
Europeu														
f^*1	0,951	0,268	0,389	0,146	0,411	0,417	0,993	0,598	0,497	0,737	0,182	0,325	0,853	0,586
f^*2	0,049	0,732	0,611	0,854	0,589	0,583	0,007	0,152	0,503	0,263	0,818	0,675	0,147	0,414
f^*3								0,250						
N	996	542	267	202	958	306	542	560	359	400	371	328	479	633
Africano														
f^*1	0,455	0,854	0,000	0,421	0,321	0,082	0,015	0,089	0,970	0,117	0,183	0,894	0,458	0,357
f^*2	0,545	0,146	1,000	0,579	0,679	0,918	0,985	0,825	0,030	0,883	0,817	0,106	0,542	0,643
f^*3								0,086						
N	1141	1123	11	123	135	11	1375	1162	1353	1162	290	883	1271	187
Ameríndio														
f^*1	0,989	0,050	0,544	0,567	0,786	0,923	0,998	0,542	0,463	0,434	0,845	0,145	0,664	0,657
f^*2	0,011	0,950	0,456	0,433	0,214	0,077	0,002	0,339	0,537	0,566	0,155	0,855	0,336	0,343
f^*3								0,119						
N	1009	846	313	163	699	400	353	184	662	184	501	618	662	263

Legenda: GO (Goiás); DF (Distrito Federal); alelo 1 do marcador GC corresponde ao alelo 1S; O alelo 2 corresponde ao alelo 1F e O alelo 3 corresponde ao alelo 2.

Tabela 4.6. Pares de *loci* que apresentaram desequilíbrio de ligação ($p < 0,05$) nas populações do Distrito Federal e do estado de Goiás.

População	Pares de <i>loci</i>	Valor de p	
Distrito Federal	AT3/FyNull	< 0,0001	
	AT3/LPL	0,0047	
	AT3/PV92	0,0019	
	D1/GC	0,0238	
	FyNull/OCA2	0,0353	
	OCA2/GC	0,0159	
	PV92/D1	0,0415	
	PV92/FyNull	0,0116	
	Rb2300/FyNull	0,0344	
	Rb2300/LPL	0,0103	
	Sb19.3/FyNull	0,0013	
	Goiás	APO/FXIII	0,0294
		APO/LPL	0,0372
APO/Sb19.3		0,0109	
TPA25/APO		0,0334	
TPA25/D1		0,0025	

Tabela 4.7. Estatística F (f, F e Theta-P) e componentes da variância (Sigma-P, Sigma-I e Sigma-G) para os 14 *loci* autossômicos analisados nas populações do Distrito Federal e do estado de Goiás.

<i>Locus</i>	f	F	Theta-P	Sig-P	Sig-I	Sig-G
APO	0,0869	0,0860	-0,0010	-0,000	0,016	0,173
AT3	0,0836	0,0865	0,0032	0,001	0,037	0,409
D1	-0,0275	-0,0284	-0,0008	-0,000	-0,013	0,476
DRD2-A	-0,0229	-0,0108	0,0118	0,005	-0,010	0,462
ECA	-0,0738	-0,0668	0,0065	0,003	-0,036	0,518
FXIII	0,0971	0,1009	0,0042	0,002	0,048	0,442
FyNull	0,0756	0,0739	-0,0019	-0,001	0,024	0,292
GC	0,0250	0,0226	0,0025	-0,002	0,016	0,617
LPL	-0,0916	-0,0938	-0,0020	-0,001	-0,045	0,538
OCA2	-0,0749	-0,0769	-0,0018	-0,001	-0,034	0,489
PV92	-0,0039	-0,0017	0,0023	0,001	-0,002	0,415
Rb2300	0,0221	0,0227	0,0006	0,000	0,011	0,489
Sb19.3	-0,0109	-0,0136	-0,0027	-0,001	-0,004	0,342
TPA25	0,0478	0,0466	-0,0013	-0,001	0,024	0,476
Total	0,0053	0,0065	0,0012	0,007	0,033	6,138

Tabela 4.8. Diferenciação gênica e genotípica entre as populações do Distrito Federal e Goiás considerando 14 marcadores genéticos do tipo AIMs.

<i>Locus</i>	Diferenciação gênica		Diferenciação genotípica	
	Valor p	Desvio padrão	Valor p	Desvio padrão
APO	0,3969	0,0036	0,4068	0,0078
AT3	0,1543	0,0044	0,1778	0,0079
D1	0,4377	0,0047	0,4216	0,0081
DRD2-A	0,0244	0,0015	0,0272	0,0027
ECA	0,0711	0,0026	0,0636	0,0042
FXIII	0,1132	0,0036	0,1237	0,0063
FyNull	0,5666	0,0035	0,5888	0,0075
GC	0,8470	0,0034	0,8431	0,0051
LPL	0,7010	0,0039	0,6858	0,0069
OCA2	0,6389	0,0041	0,6430	0,0071
PV92	0,1878	0,0037	0,2030	0,0084
Rb2300	0,2712	0,0058	0,2851	0,0090
Sb19.3	0,9265	0,0010	0,9311	0,0018
TPA25	0,4574	0,0045	0,4639	0,0091
Total	0,1683		0,1912	

Tabela 4.9. Estimativas de contribuições das parentais africana, ameríndia e europeia na constituição genética atual das populações do Distrito Federal e de Goiás, utilizando 14 marcadores de DNA autossômicos.

Populações	Grupos parentais		
	Africano	Ameríndio	Europeu
Goiás	0,206±0,001	0,167±0,001	0,627±0,001
Distrito Federal	0,229 ±0,008	0,157 ±0,007	0,614±0,008
Total	0,217±0,001	0,163±0,003	0,621± 0,004

A análise de mistura genética com base nos 14 marcadores estudados nas duas populações está apresentada na Tabela 4.9. Nas duas populações houve o predomínio da parental europeia, sendo que no Distrito Federal a contribuição foi de 61,4% e em Goiás de 62,7%. A contribuição africana, com valores próximos nas duas populações, foi intermediária em relação à europeia e ameríndia. A contribuição parental ameríndia, menor nas duas populações quando comparado com as duas outras parentais, apresentou-se maior na população do estado de Goiás (16,7%) do que no Distrito Federal (15,7%). As estimativas de contribuição obtidas a partir da análise das duas populações em conjunto apresentaram resultados similares àqueles observados nas duas populações individualmente, com predomínio de europeu, contribuição intermediária para o componente africano e mais baixo para o ameríndio.

A Tabela 4.10 apresenta os dados de estatística descritiva para os valores IAAs encontrados para as populações do Distrito Federal e de Goiás. Os resultados encontrados para os índices de assimetria (0,545 para Distrito Federal; 0,358 para Goiás) e curtose (-0,156 para Distrito Federal; 0,27 para Goiás) mostraram que os dados se distribuíram em uma curva simétrica normal e que os valores de IAA estão distribuídos de uma forma contínua. O valor p do teste de significância de Levene encontrado foi de 0,457 indicando que as variâncias são iguais entre as duas amostras analisadas. Como os dados de IAA individuais nas duas populações se distribuíram em uma curva normal e as variâncias entre as mesmas foram semelhantes (homogêneas), o teste t foi utilizado para comparar as médias obtidas entre os IAAs do Distrito Federal e Goiás. O resultado da significância do teste foi de 0,835, indicando que não há diferença estatisticamente significativa entre as médias encontradas. O nível de significância considerado em tais análises foi de 5%.

A Figura 4.2 apresenta os histogramas da distribuição dos valores de índice de ancestralidade africana nas populações do Distrito Federal e do Estado de Goiás. Conforme se observa na figura e na tabela 4.12 houve uma grande variação dos índices nas duas populações, com valores de variância altos. O valor máximo de ancestralidade europeia encontrado para o Distrito Federal foi de -12,89 enquanto que para Goiás foi de -13,54. O maior valor de ancestralidade africana encontrado para o Distrito Federal foi de 4,36 e para o estado de Goiás foi de 6,01. Conforme a observação dos dados individuais pode-se constatar que a grande maioria apresentou valores de IAA abaixo de zero, 89,9% para o Distrito Federal e 92,1% para Goiás, indicando que a contribuição europeia foi maior na maioria das estimativas individuais. O valor médio determinado para o Distrito Federal foi de -4,84, enquanto que para o estado de Goiás a média foi de -4,92.

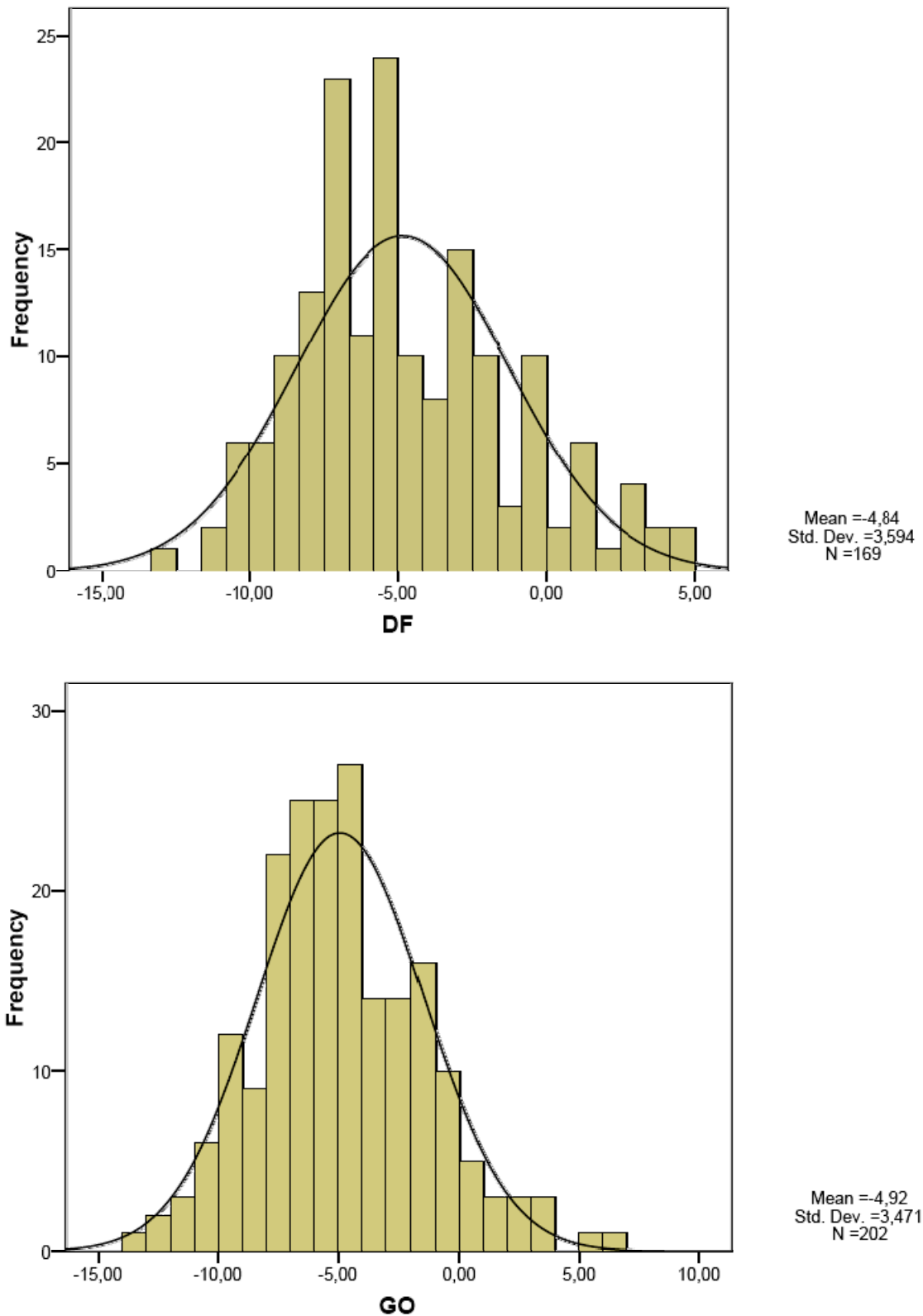


Figura 4.2. Histograma da distribuição dos valores de índice de ancestralidade africana nas populações do Distrito Federal e do Estado de Goiás.

Tabela 4.10. Dados da análise de estatística descritiva para os valores IAAs determinados nos indivíduos amostrados na população do Distrito Federal e Goiás.

Análise descritiva	Populações	
	Distrito Federal	Goiás
Média	-4,8432	-4,9199
Erro padrão da média	0,27648	0,2442
Mediana	-5,5549	-5,3044
Desvio padrão	3,59421	3,47066
Variância	12,918	12,046
Assimetria	0,545	0,358
Erro padrão-assimetria	0,187	0,171
Curtose	-0,156	0,27
Erro padrão-curtose	0,371	0,341
Amplitude	17,25	19,56
Mínimo	-12,89	-13,54
Máximo	4,36	6,01

4. 4. DISCUSSÃO

A história de povoamento e colonização do Distrito Federal e Goiás se confundem e são interligadas. Até o século XVII a região era habitada principalmente por povos ameríndios. O impulso dado pelos movimentos migratórios promovidos pelo governo para povoar o sertão brasileiro permitiu que para essa região viessem povos de diversas regiões do país, especialmente das regiões sudeste e nordeste. Os migrantes por sua vez, muitas vezes já possuíam traços de miscigenação entre o branco e africano, e em menor grau, com o ameríndio. Portanto, além da miscigenação ocorrida após a chegada à região em virtude do encontro de pelo menos três diferentes etnias – europeu, africano e ameríndio – o processo de mistura é mais complexo nesta população tendo em vista a migração de indivíduos já miscigenados (Palacín *et al.*, 1995).

A partir da década de 50 do século XX, a região do Distrito Federal teve sua história transformada em virtude da construção de Brasília, a capital federal. O fluxo de migrantes para essa região foi muito intenso, a ponto de apenas nos primeiros anos a

população ter aumentado mais de 300%. Por causa disso, pergunta-se: houve diferenciação entre essa população e a do estado de Goiás em decorrência o fluxo acentuado de pessoas para o Distrito Federal?

As análises intrapopulacionais e interpopulacionais aqui realizadas para as duas populações revelaram poucas diferenças. Embora a migração seja um fator evolutivo que tende a modificar as frequências alélicas e genotípicas, todos os marcadores analisados nas duas populações mostraram estar em equilíbrio de acordo com o teorema de Hardy-Weinberg.

Entretanto, a população do Distrito Federal apresentou 11 pares em desequilíbrio na análise de desequilíbrio de ligação entre os pares de *loci*. O desequilíbrio de ligação é observado quando há associações não ao acaso entre marcadores genéticos. Não reflete obrigatoriamente a existência de ligação gênica entre os *loci*, em especial em populações miscigenadas. Nestas populações pode-se detectar desequilíbrio de ligação mesmo entre marcadores situados relativamente distantes (Parra *et al.*, 1998; Chakraborty e Weiss, 1988).

Dessa forma, é esperado que em populações oriundas da miscigenação entre parentais com frequências alélicas muito distintas seja observado alto desequilíbrio de ligação e que esse desequilíbrio diminua com o tempo (Chakraborty e Weiss, 1988), caso não exista ligação gênica associada. O desequilíbrio de ligação pode também refletir a existência de subestruturação populacional, porém nesse caso espera-se que também ocorra desequilíbrio da frequência genotípica segundo o teorema de Hardy-Weinberg. Como neste trabalho todos os marcadores, em ambas populações (Distrito Federal e Goiás), encontram-se em equilíbrio de Hardy-Weinberg, é possível propor que este desvio tenha ocorrido em virtude da mistura entre as populações parentais.

Por outro lado, após a utilização da correção de Bonferroni para testes múltiplos, apenas um par de marcadores continuou mostrando desequilíbrio de ligação, o par AT3/FyNull para o Distrito Federal. Nesse caso, o resultado já era esperado, pois esse par de marcadores situa-se próximo fisicamente, no braço longo do cromossomo 1 a aproximadamente 22 cM (centimorgans) de distância um do outro. Resultado similar já foi relatado em outras populações, como é o caso de afro-descendentes de Nova Iorque e Nova Orleans que apresentaram desequilíbrio de ligação entre esses dois *loci* (Parra *et al.*, 1998). Dessa forma, sugere-se que neste estudo o desequilíbrio observado tenha sido ocasionado em virtude de mistura.

A heterozigose é um dos principais parâmetros para avaliar o potencial de discriminação de um determinado marcador em uma dada população, sendo que o valor de heterozigose máximo para *loci* bialélicos é de 50%. Valores de heterozigose baixos indicam que o marcador apresenta baixa diversidade, sendo que a quantidade de heterozigotos na população será uma função direta da frequência de qualquer dos dois alelos (Comas *et al.*, 2000). Como observado na tabela 4.7, os valores de heterozigose para os *loci* analisados, nas duas populações, foram próximos ao máximo esperado, exceto para o marcador APO, cujos valores foram baixos nas duas populações. Esse valor era esperado, tendo em vista que a presença da inserção desse marcador se encontra fixada em populações ameríndias da América do Sul (Oliveira, 1999; Batzer *et al.*, 1994) e próxima de 1 em populações européias (Tabela 4.7). Como a população local apresenta em sua composição o componente ameríndio e, de modo mais expressivo, o europeu, verificou-se que a presença do alelo inserção foi alto, isto é acima de 88%. De um modo geral, observou-se que a diversidade genética nas duas populações foi similar, indicando que as mesmas devem ser homogêneas.

Embora o fluxo migratório tenha sido muito intenso nas últimas décadas para o Distrito Federal, as análises de estatística F , assim como os resultados de diferenciação gênica e genotípica entre Distrito Federal e Goiás, mostraram que as mesmas são similares do ponto de vista genético. Os valores estimados de $f(F_{is})$ foram baixos, indicando que na população há preponderância de acasalamentos aleatórios. Além disso, os baixos valores de $Theta-P (F_{st})$ encontrados mostram que a população não é subestruturada.

Apenas na análise de diferenciação gênica e genotípica observou-se um marcador, o DRD2-A, com valores estatisticamente significativos, indicando que existe diferença na distribuição dos alelos e dos genótipos entre as duas populações para esse marcador. Na análise global, entretanto, esse desvio não ocorreu indicando que as duas populações podem ser consideradas como uma única população. A variância entre as populações pôde ser decomposta em componentes que indicam quais deles mais contribuíram para a variabilidade na amostra. Assim, pôde-se notar que a maior variação nessas duas amostras foi decorrente da variação entre os indivíduos. Quando se faz a relação entre a variância decorrente das diferenças entre as amostras e das diferenças ocorridas dentro da amostra tem-se uma estimativa de significância para determinar se o valor encontrado da variância entre as amostras é significativo ou não. Quanto mais diferentes forem os valores de heterozigose entre as amostras, o resultado da relação entre os dois componentes da variância (entre amostras e dentro das amostras) será maior. Em outras palavras, quanto

menor for o resultado da variância entre as amostras e maior for o da variância dentro das amostras ter-se-á valores baixos indicando que as populações são homogêneas. No presente estudo, o componente da variância entre as amostras foi pequeno quando comparado ao componente de variação dentro das amostras, mostrando, conseqüentemente, mais uma vez que não existe diferença entre as duas populações.

Apesar da fundação do Distrito Federal ser recente, assim como o povoamento da região centro-oeste como um todo, e de que o fluxo migratório continua intenso, em especial para o Distrito Federal, não foi observada subestruturação populacional. As populações encontram-se em equilíbrio segundo o esperado pela Lei de Hardy-Weinberg tanto para os marcadores aqui analisados quanto para os microssatélites, como discutido no capítulo 3. Como explicitado em Material e Métodos, fez-se a opção por incluir na amostra exclusivamente indivíduos nascidos na região buscando justamente obter o produto de ao menos uma geração de casamentos. Conforme o teorema de Hardy-Weinberg, após uma geração de acasalamento ao acaso as frequências genóticas se restabelecem de acordo com as proporções $p^2 + 2pq + q^2$. Como os indivíduos amostrados são de pelo menos uma geração após o estabelecimento da família na região espera-se que, caso tenha ocorrido desvio das proporções esperadas em decorrência da migração, houve restabelecimento do equilíbrio após uma geração. Isso também mostra que as duas populações são, em linhas gerais, panmíticas e com baixa endogamia.

Uma segunda questão a ser abordada é que, de fato, as populações urbanas brasileiras não apresentam grande diferencial genético, como estimado pela análise de microssatélites. Desse modo, o intenso fluxo de migrantes nas últimas décadas em virtude dos atrativos para a região (construção de Brasília, além do desenvolvimento econômico alcançado pela região nos últimos tempos), não foi determinante para que a população se tornasse subestruturada, já que as diferenças entre as regiões brasileiras não são expressivas como ocorre em outros países, como a Argentina e a Colômbia (Toscanini *et al.*, 2007; Paredes *et al.*, 2003).

Além disso, uma questão fundamental é a ocorrência de um intenso fluxo de migrantes da população do estado de Goiás para o Distrito Federal. Conforme dados do IBGE, o povoamento do Distrito Federal teve importante contribuição da população de Goiás, o que torna as populações ainda mais homogêneas. O incremento da migração para o Distrito Federal, de certo modo, fortaleceu a vinda de migrantes também para Goiás. Tanto é assim, que em torno da região do Distrito Federal formaram-se 11 cidades depois que Brasília foi fundada, como Águas Lindas e Valparaíso de Goiás, que apresentam

determinadas características semelhantes à população do Distrito Federal, isso é, os habitantes geralmente são provenientes de outras partes do país como da região sudeste e nordeste. As outras cidades já existentes na região, por sua vez, experimentaram aumento da população nos últimos anos em função de Brasília, como é o caso de Formosa e de Luziânia.

Mistura

Dois fatores são de fundamental importância quando se pretende estimar mistura genética. O primeiro é decorrente da escolha dos marcadores que serão utilizados na análise; o segundo é decorrente da estrutura da população a ser estudada. Quanto à primeira questão, existem marcadores mais indicados para cálculos de mistura. Em teoria, quanto mais diferentes forem as frequências alélicas entre as populações que contribuíram para a formação da população miscigenada, mais precisos serão os resultados obtidos. Os AIMS apresentam essa característica e, por isso, são muito utilizados nesse tipo de análise. Como exposto anteriormente, para delineamento desse estudo, foram escolhidos marcadores que apresentassem diferencial de frequência entre os três principais grupos parentais que contribuíram para a composição atual da população do Distrito Federal e estado de Goiás.

Em relação à segunda questão, isto é, a estrutura da população estudada, convém ressaltar que faz-se necessário um estudo aprofundado da mesma. É necessário saber se as distribuições das frequências alélicas e genotípicas não estão sofrendo alguma pressão evolutiva, decorrente de deriva, migrações ou se a população estudada está sob algum tipo de seleção. As populações aqui estudadas estão sob influência contínua da migração desde o século XVII. Seria de se imaginar, portanto, que o fluxo intenso de migrantes tivesse contribuído para desestabilizar a população na distribuição dos seus alelos visto o alto diferencial de frequências alélicas entre as populações parentais. Assim, o encontro de desequilíbrio na distribuição dos genótipos nas duas populações poderia ter sido encontrado.

Entretanto, as análises estatísticas mostraram que os marcadores utilizados para mistura estão em equilíbrio segundo o proposto por Hardy-Weinberg, inclusive quando os genótipos dos indivíduos analisados no Distrito Federal e em Goiás são considerados um único banco amostral (tabela 4.7). Assim, é pertinente fazer análise de mistura, pois as

estimativas obtidas devem ser representativas da população e não apenas de um grupo dentro da população.

Os resultados de mistura obtidos com os marcadores AIMs indicam, assim como os marcadores microssatélites, que a parental européia é a que mais contribuiu para a composição da população da região. Esse resultado era esperado tendo em vista que a vinda de povos descendentes de europeus foi muito intensa desde o início do povoamento. Além disso, no início do período de colonização houve cruzamento preferencial entre o homem europeu e as mulheres ameríndias e africanas (Barcelos *et al.*, 2006), e com isso, o homem de origem africana e ameríndia tiveram menos sucesso na perpetuação de seus alelos. Essa é uma realidade na maioria da população brasileira (Pena, 2002), sendo que a maioria das linhagens de cromossomo Y do país é de origem portuguesa, havendo contribuição mínima quanto às linhagens do cromossomo Y ameríndia e africana.

Entretanto, um dado novo foi obtido com os marcadores AIMs. A estimativa de contribuição da parental ameríndia teve um aumento significativo, especialmente na população do estado de Goiás. Nas estimativas de mistura com os marcadores microssatélites, foi observado uma baixa contribuição da parental ameríndia (3%), embora a história da região mostre que o contato principalmente com o europeu tenha ocorrido durante os primeiros séculos de povoamento da região. Os dados levantados durante a entrevista com os participantes da pesquisa mostraram que quase 70% se autodeclararam como mestiço entre pelo menos duas parentais. Por isso, seria esperado que houvesse maior participação dessa parental na formação atual da população.

Pelo fato dos contatos entre os colonizadores e a população nativa ameríndia terem sido desastrosos, havendo inclusive a dizimação de muitas tribos assim como a migração forçada em busca de regiões mais isoladas, principalmente para o oeste do país, o resultado de mistura com os marcadores microssatélites não foi de todo estranho. Contudo, os marcadores AIMs mostraram outro quadro. A contribuição dessa parental para a população do estado de Goiás foi maior, cerca de 17%, mostrando claramente que embora a população indígena na região tenha sido praticamente dizimada, o componente ameríndio permanece inserido na população fruto da miscigenação com os demais grupos parentais, principalmente com o europeu.

Para o Distrito Federal, o aumento da contribuição parental ameríndia na constituição da população não foi expressivo quando comparado com os resultados de mistura de microssatélites. Enquanto que as estimativas obtidas com o uso dos marcadores

microsatélites indicaram valores em torno de 13%, com os AIMs os valores foram próximos a 16%.

Por outro lado, convém lembrar que as amostras utilizadas nessas duas análises não foram as mesmas. Para as análises dos microsatélites foi utilizado um banco de dados de um laboratório público (Laboratório de DNA Forense da Polícia Civil do DF), enquanto que para as análises com os AIMs foram usadas amostras de pessoas nascidas em 13 das 19 regiões administrativas do Distrito Federal existentes á época da coleta. Dessa forma, podemos ter um diferencial de origem socioeconômico entre essas amostragens. Pode-se inferir que os indivíduos que compõem o banco da polícia civil ou são pessoas que não possuem condições financeiras de arcar com os custos de um exame de paternidade ou fazem parte de um processo de investigação criminal. De um modo geral, portanto, os indivíduos devem ser provenientes de classes sociais mais baixas.

Em um estudo anterior (Dalton *et al.*, 2007) foram estimadas as contribuições de mistura genética para a população do Distrito Federal utilizando o banco de dados da polícia civil e um outro banco de dados obtido a partir de análise de paternidade em um laboratório particular (Gratapaglia *et al.*, 2001). A conclusão a que se chegou é que o percentual de contribuições parentais africana e ameríndia é maior no banco da polícia civil do que no banco particular. Portanto, os resultados mostraram que a constituição genética de populações miscigenadas com relação à contribuição de populações parentais difere em decorrência da classe social.

Estudo semelhante foi realizado com uma população da Venezuela (Martínez *et al.*, 2007), com objetivo de estimar os componentes de mistura de amostras provenientes de dois diferentes níveis socioeconômicos. As amostras foram coletadas na cidade de Caracas em dois bancos de sangue, sendo um proveniente de uma maternidade pública e o outro de uma maternidade privada. Os resultados obtidos com marcadores de DNA autossômicos mostraram que na maternidade privada os indivíduos apresentam maior contribuição européia, enquanto que na maternidade pública o componente europeu é menor que o ameríndio e o africano. Ambos estudos mostram que as diferenças na composição genética ancestral em diferentes níveis sócioeconômicos na população podem gerar interpretações errôneas sobre a população de um modo geral. A origem da amostra estudada deve ser conhecida para evitar interpretações tendenciosas.

No presente estudo, as análises de mistura realizadas com marcadores microsatélites autossômicos e com marcadores AIMs mostraram resultados similares, embora o esperado fosse que as análises com marcadores AIMs mostrassem um aumento

da contribuição parental ameríndia e africana, ou de modo inverso, que os resultados de mistura com marcadores microssatélites tivessem apresentado as duas contribuições parentais, africana e ameríndia, com valores abaixo dos encontrados, semelhante ao ocorrido com os resultados encontrados para o estado de Goiás. O fato dos resultados de mistura com marcadores microssatélites para a população do Distrito Federal não seguirem a mesma tendência que houve em Goiás, não os invalida, mas as interpretações dos resultados devem ser coerentes com os critérios de coleta da amostra analisada.

Com relação à contribuição parental africana nas duas populações, foram obtidos valores intermediários entre as duas outras parentais, porém mais próxima da parental ameríndia. As considerações feitas quanto à contribuição parental ameríndia também se estendem para a africana. Cabe, entretanto, ressaltar que sete dos marcadores AIMs utilizados para a análise de mistura diferenciam a parental africana das duas outras parentais.

O índice de ancestralidade africana (IAA) estimado para cada indivíduo corrobora os resultados das estimativas de mistura apresentados. De acordo com Parra *et al.* (2003), o maior IAA em uma amostra de portugueses foi de -4,86 enquanto que o menor em uma amostra de São Tomé foi de 2,86. Em 58% dos indivíduos do Distrito Federal e Goiás os valores de IAA apresentaram valores menores do que o maior valor encontrado nos europeus, indicando, portanto, que tais indivíduos apresentam um índice de ancestralidade europeia superior à africana. Pouco mais de 3% dos indivíduos estudados no presente trabalho apresentou valores iguais ou acima do encontrado em São Tomé. Para os demais indivíduos analisados no presente estudo (38% da amostra), os índices estimados estavam no intervalo entre o maior índice determinado para europeu e o menor índice determinado para africano, indicando, nesse caso, que tais indivíduos apresentam na sua composição genética uma contribuição europeia e africana similares.

Pelo fato das populações do Distrito Federal e do estado de Goiás formarem uma população homogênea, foi realizado o teste de mistura considerando as duas populações como sendo uma só amostra. A contribuição das três parentais foi similar ao observado nas duas populações separadamente. Com isso, pode-se concluir que essa região apresenta uma população miscigenada e que os três grupos parentais principais da população brasileira efetivamente colaboraram para a composição genética atual da população. Outro fator que deve ser levantado é que a parental ameríndia embora tenha praticamente desaparecido do estado de Goiás, restando poucos indivíduos vivendo em reservas indígenas, foi em parte absorvida pela população que estava se formando.

Contextualização dos dados apresentados

Fontes históricas e antropológicas, assim como genéticas, consideram a população brasileira híbrida, essencialmente um produto de cruzamentos ou contatos de grupos europeus, africanos e ameríndios. Este tipo de estrutura populacional pode ser, em teoria, muito útil na quantificação do fluxo gênico que culminou com a formação da população atual.

Tem-se buscado estudar a população brasileira, especialmente quanto a sua formação, do ponto de vista genético, pois os traços físicos de um indivíduo, como cor da pele, cor e textura do cabelo, são adaptações para fatores seletivos geográficos como o calor e a radiação (Cavalli-Sforza, 1997). Por isso, desenvolver estudos que permitam compreender melhor a estrutura genética da população brasileira pode auxiliar no esclarecimento de pontos históricos. Exemplo disso foram os estudos genéticos demonstrando que a formação da população brasileira é resultado de cruzamentos preferenciais entre homens europeus e mulheres africanas e indígenas. Foi descrito que a vasta maioria dos cromossomos de homens brasileiros de origem branca tem origem européia, com uma frequência muito baixa de cromossomos Y africanos e completa ausência de contribuição ameríndia (Carvalho-Silva *et al.*, 2001). Por outro lado, o mtDNA (DNA mitocondrial) de tais indivíduos dessa amostra apresentam uma alta contribuição de da parental ameríndia (33%) e africana (28%) (Alves-Silva *et al.*, 2000).

Outro motivo importante para se estudar a população brasileira do ponto de vista genético é que poderá contribuir para entendimento das diferenças regionais. Sabe-se que devido à vasta extensão territorial de nosso país, o processo de colonização não se deu de forma igual, sendo que estudos em genética de populações vêm comprovando essas diferenças (Parra *et al.*, 2003; Callegari-Jacques *et al.*, 2003). Tais estudos mostraram que existe maior predominância de contribuição ameríndia feminina na região norte e uma preponderância da africana feminina no nordeste. Além disso, os níveis mais pronunciados de mistura foram observados no sudeste e no nordeste, com uma tendência norte-sul de contribuição européia.

Um desses estudos realizado na população brasileira teve como objetivo estudar as cinco regiões geográficas brasileiras utilizando 12 *loci* do tipo STRs (Callegari-Jacques *et al.*, 2003). As amostras utilizadas para tal foram provenientes de indivíduos envolvidos em disputa de paternidade analisadas em um laboratório particular entre os anos de 2000 e 2001. Os resultados encontrados por este estudo das proporções acumuladas de

ancestralidade para a região Centro-Oeste (europeu = $0,71 \pm 0,002$; africanos = $0,18 \pm 0,001$ e ameríndio = $0,11 \pm 0,001$) não diferem muito dos resultados obtidos neste trabalho para os marcadores STRs analisados.

Porém, uma questão a ser levantada é se os marcadores STRs seriam tão interessantes para estimar mistura como, em teoria, são os marcadores AIMS. Isso porque, a taxa de mutação das regiões microssatélites é alta, o que acarreta uma maior taxa de mutação recorrente. Pela análise de grupos populacionais maiores, como o estudo desenvolvido com os países da América Latina e Caribe, e com as regiões geográficas de três países, foi possível separar os países ou as regiões em agrupamentos de acordo com a maior proporção de contribuição dos grupos parentais que formaram essas populações. Entretanto, as análises de mistura realizadas nas duas populações da região centro-oeste mostraram que os marcadores AIMS são mais eficazes para discernir o grau de contribuição das populações parentais. Mesmo assim, ambos, microssatélites e AIMS, mostraram o mesmo quadro na composição genética da população: existe predomínio de genes europeus em relação aos outros dois grupos étnicos, resultados esses concordantes com estudos anteriores para populações brasileiras.

CONSIDERAÇÕES FINAIS

A expansão de vários impérios europeus devido às grandes navegações, a partir do século XV, promoveu transformações econômicas, culturais e genéticas nas populações dos vários continentes e em especial na América. Análises de mistura e distância genética entre 13 populações latino-americanas indicaram um quadro heterogêneo. A contribuição européia supera as duas outras parentais que também fizeram parte da história de formação da população do continente em algumas regiões. Por outro lado, embora os africanos tenham migrado para a América em um espaço de tempo determinado, na região do Caribe e em alguns países que utilizaram de modo mais acentuado o trabalho escravo, como a Colômbia e o Brasil, a contribuição dessa parental é alta.

Cada país americano teve uma história diferente quanto ao encontro dos três grupos étnicos. Os resultados aqui apresentados das análises das regiões de três países da América do Sul tomados como modelos (Argentina, Brasil e Colômbia) indicaram que o processo de migrações internas foi decisivo para a estruturação das populações. No Brasil, por exemplo, a colonização européia promoveu uma diminuição acentuada da população autóctone, o que está refletido na baixa estimativa de ameríndios nas populações urbanas em comparação aos dois outros países. Embora existissem conflitos entre europeus e ameríndios na Argentina, a população ameríndia estava mais presente em algumas regiões mais isoladas, como a região noroeste e a Patagônia.

Essas diferenças regionais foram detectadas por diferenças nas frequências alélicas de marcadores de DNA. Pelo fato das parentais serem originárias de três continentes distintos, as frequências alélicas dos marcadores utilizados no presente estudo, especialmente os AIMs, têm um diferencial de frequência entre elas. A miscigenação entre esses grupos parentais gerou, conseqüentemente, substruturação populacional. Por outro lado, a migração é uma força evolutiva que tende a homogeneizar as populações, em especial as migrações internas. O desequilíbrio das distribuições genotípicas causado pelo processo de mistura tende a se estabilizar após algumas gerações. Como se observa pela análise dos STRs e AIMs na região centro-oeste, a população apresentou-se em equilíbrio mesmo após sofrer migrações sucessivas ao longo dos três últimos séculos.

Quando comparamos os dados de distância entre as populações dos diferentes países, e em alguns casos entre regiões dentro de um país, verificou-se um quadro bastante heterogêneo. Entretanto, nas análises de distância envolvendo três estados do centro-oeste brasileiro, verificou-se homogeneidade entre as mesmas indicando que o fluxo gênico entre

grupos populacionais mais próximos favoreceu a migração entre os estados, isto é, a troca de migrantes nesse caso foi proporcional à distância geográfica das populações analisadas.

REFERÊNCIA BIBLIOGRÁFICA

- Abe-Sandes, K; Silva-Jr, W; Zago, MA. (2004). Heterogeneity of the Y chromosome in Afro-Brazilian populations. *Hum Biol* 76(1):77-86.
- Acuña, MP; Llop, ER; Rothhammer, FE. (2000). Composición genética de la población chilena: las comunidades rurales de los valles de Elqui, Limarí y Choapa. *Rev. Méd. Chile* 128(6): 593-600.
- Agrawal, S; Khan, F. (2005). Reconstructing recent human phylogenies with forensic STR *loci*: A statistical approach. *BMC Genetics* 6:47-53.
- Alencastro, LP. (2000). O trato dos viventes: formação do Brasil no Atlântico Sul. Companhia das Letras, São Paulo/SP, 523p.
- Alves, C; Gusmão, L; Damasceno, A; Soares, B; Amorim, A. (2004). Contribution for an African autosomic STR database (AmpF/STR Identifiler and Powerplex 16 System) and a report on genotypic variations. *F Sci Int* 139: 201-205.
- Alves, C; Gusmão, L; López-Parra, AM; Mesa, MS; Amorim, A; Arroyo-Pardo, E. (2005). STR allelic frequencies for an African population sample (Equatorial Guinea) using AmpFISTR Identifiler and Powerplex 16 kits. *F. Sci. Int.* 148: 239-242.
- Alves-Silva, J; Santos, MS; Guimarães, PEM; Ferreira, ACS; Bandelt, HJ; Pena, SDJ; Prado, VF. (2000). The Ancestry of Brazilian mtDNA Lineages. *Am J Hum Genet* 67: 444–461.
- Arcot, SS; Adamson, AW; Risch, GW; LaFleur, J; Robichaux, MB; Lamerdin, JE; Carrano, AV; Batzer, MA. (1998). High-resolution cartography of recently integrated human chromosome 19-specific Alu fossils. *Journal of Molecular Biology* 281(5): 843-56.
- Arpini-Sampaio, Z; Costa, MC; Melo, AA; Carvalho, MF; Deus, MF; Simões, AL. (1999). Genetic polymorphisms and ethnic admixture in African-derived black communities of northeast Brazil. *Human Biology* 71(1):69-85.
- Bamshad, M; Wooding, S; Salisbury, BA; Stephens, JC. (2004). Deconstructing the relationship between genetics and race. *Nature* 5: 598-609.
- Bandelt, HJ; Herrnstadt, C; Yao, YG; Kong, QP; Kivisild, T; Rengo, C; Scozzari, R; Richards, M; Villems, R; Macaulay, V; Howell, N; Torroni, A; Zhang, YP. (2003). Identification of Native American Founder mtDNAs Through the Analysis of Complete mtDNA Sequences: Some Caveats. *Annals of Human Genetics* 67(6):512–524
- Barbujani, G. (2005). Human races: classifying people vs understanding diversity. *Current Genomics* 6:1-12.
- Barcelos, R; Ribeiro, G; Silva Jr, W; Abe-Sandes, K; Godinho, N; Marinho-Neto, F; Gigonzac, M; Klautau-Guimarães, M; Oliveira, S. (2006a). Male contribution in the

- constitution of the Brazilian Centro-Oeste population estimated by Y-chromosome binary markers. *International Congress Series* 1283:228-230.
- Barcelos, RSS; Araújo da Silva-Jr, W; Abe-Sandes, K; Godinho, NMO; Oliveira, SF. (2006b). Contribución parental en la constitución de poblaciones del Brasil central estimada por el cromosoma Y y marcadores del mtADN. In Andreo, A.M.A. *Diversidad biológica y salud humana*. Universidad Católica San Antonio, Murcia, Espanha.
- Batzer, M; Stoneking, M; Alegria-Hartman, M; Bazan, H; Kass, DH; Shaik, H; Novick, GH; Ioannou, PA; Scheer, WD; Herrera, RJ; Deininger, PL. (1994). African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci* 91: 12288-12292.
- Batzer, MA; Knight, A; Stoneking, M; Tiwar, HK; Scheer, WD; Herrera, RJ; Deininger, PL. (1996). DNA sequences of Alu elements indicate a recent replacement of the human autosomal genetic complement. *Proc Natl Acad Sci* 93: 4360-4364.
- Batzer, MA; Arcot, SS; Phinney, JW; Alegria-Hartman, MA; Kass, DH; Miligan, SM; Kimptom, C; Gill, P; Hochmaister, M; Ioannou, PA; Hehhera, RJ; Boudreau, DA; Scheer, WD; Keats, BJ; Deininger, PL; Stoneking, M. (1996). Genetic variation of recent Alu insertions in human populations. *J. Mol.* 42: 22-29.
- Batzer, MA; Rubinb, CM; Hellmann-Blumbergc, U; Alegria-Hartmana, M; Leeflangc, EP; Sternb, JD; Bazand, HA; Shaikhhd, TH; Deininger, PL; Schmid, CW. (1995). Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. *Journal of Molecular Biology.* 247(3):418-27.
- Bedoya, G; Montoya, P; Garcia, J; Soto, I; Bourgeois, S; Carvajal, L; Labuda, D; Alvarez, V; Ospina, J; Hedrick, PW; Ruiz-Linares, A. (2006). Admixture dynamics in Hispanics: A shift in the nuclear genetic ancestry of a South American population isolate. *PNAS* 103(19):7234-7239.
- Behar, DM; VILLEMS R; Soodyall, H; Blue-Smith, J; Pereira, L; Metspalu, E; Scozzari, R; Makkan, H; Tzur, S; Comas, D; Bertranpetit, J; Quintana-Murci, L; Tyler-Smith, C; Wells, RS; Rosset, S; Geographic Consortium. (2008). The dawn of human matrilineal diversity. *Am J Hum Genet* 82:1-11.
- Benn-Torres, J; Bonilla, C; Robbins, CM; Waterman, L; Moses, TY; Hernandez, W; Santos, ER; Bennett, F; Aiken, W; Tullock, T; Coard, K; Hennis, A; Wu, S; Nemesure, B; Leske, MC; Freeman, V; Carpten, J; Kittles, RA. (2007). Admixture and population stratification in African Caribbean populations. *Ann. Hum. Gen.* 71:1-9.
- Benn-Torres, J; Bonilla, C; Robbins, CM; Waterman, L; Moses, TY; Hernandez, W; Santos, ER; Bennett, F; Aiken, W; Tullock, T; Coard, K; Hennis, A; Wu, S; Nemesure, B; Leske, MC; Freeman, V; Carpten, J; Kittles, RA. (2008). Admixture and Population Stratification in African Caribbean Populations. *Ann. Hum. Gen.* 72(1): 90–98.

- Bernasconi, A. (2000). Imigrantes italianos na Argentina (1880-1930) – uma aproximação. In Fausto, B. *Fazer a América*. Ed. Universidade de São Paulo, São Paulo/SP, p 61-92.
- Bertran, P. (2000). História da terra do homem no planalto central: eco-história do Distrito Federal – do indígena ao colonizador. Verano Editora, Brasília/DF, 270p.
- Bomfim, TF. (2008). Ancestralidade genômica em uma amostra de portadores do HIV-1 do Estado da Bahia. PPGBSMI – Programa de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa Salvador –Dissertação de Mestrado.
- Bonilla, C; Parra, EJ; Plaff, CL; Dios, S; Marshall, JA; Hamman, RF; Ferrell, RE; Hoggart, CL; Mckeigue, PM; Shriver, D. (2004a). Admixture in the hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann. Hum. Gen.* 68: 139-153.
- Bonilla, C; Shriver, MD; Parra, EJ; Jones, A; Fernández, JR. (2004b). Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York city. *Hum. Genet.* 115:57-68.
- Bookstein, R; Lai, CC; To, H; Lee, WH. (1990). PCR-based detection of a polymorphic BamHI site in intron 1 of the human retinoblastoma (RB) gene. *Nucleic. Acid. Res.* 18(6):1666.
- Bowcock, AM; Ruiz-Linares, A; Tomfohrde, J; Minch, E; Kidd, JR; Cavalli-Sforza, LL. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature.* 368(6470):455-7.
- Bydlowski, SP; Moura-Neto, RS; Soares, RPS; Silva, R; Debes-Bravo, AA; Morganti, L. (2003). Genetic data on 12 STRs (F13A01, F13B, FESFPS, LPL, CSF1PO, TPOX, TH01, vWA, D16S539, D7S820, D13S317, D5S818) from four ethnic groups of São Paulo, Brazil. *F Sci Int.* 135:67–71.
- Caine, LM; Corte-Real, F; Anjos, MJ; Carvalho, M; Serra, A; Antunes, H; Vide, MC; Vieira, DN. (2003). Allele frequencies of 13 *loci* in the Santa Catarina Population of Southern Brazil. *J Forensic Sci* 48(4): 901-902.
- Callegari-Jacques, SM; Grattapaglia, D; Salzano, FM; Salamoni, SP; Crossetti, SC; Ferreira, ME; Hutz, MH. (2003). Historical genetics: spatiotemporal analysis of formation of the brazilian population. *Am J Hum Biol* 15: 824-834.
- Campos-Sánchez R, Barrantes R, Silva S, Escamilla M, Ontiveros A, Nicolini H, Mendoza R, Munoz R, Raventos H. (2006). Genetic structure analysis of three Hispanic populations from Costa Rica, Mexico, and the southwestern United States using Y-chromosome STR markers and mtDNA sequences. *Hum. Biol.* 78(5):551-63.
- 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; Menozzi, P; Piazza A. (1996). The history and geography of human genes. Princeton University Press, Princeton, New Jersey, 443p.

- Cavalli-Sforza, LL. (1997). Genes, peoples, and languages. *Proc Natl Acad Sci USA*. 94(15):7719-24.
- Cerda-Flores, RM; Budowle, B; Jin, L; Barton, AS; Keka, R; Chakraborty, R. (2002). Maximum likelihood estimates of admixture in Northeastern Mexico using 13 short tandem repeat *loci*. *Am J Hum Biol* 14:429-439.
- Chakraborty, R. (1985). Gene identity in racial hybrids and estimation of admixture rates. In: Y Ahuja and JV Neel (Eds.): *Genetics Microdifferentiation in Human and Other Animal Populations*. Delhi, India: Indian Anthropological Association, Delhi University Anthropology Department, pp. 171-180.
- Chakraborty, R; Weiss, KM. (1988). Admixture as a tool for finding linked genes and detecting that difference from allelic association between *loci*. *Proc Natl Acad Sci USA*. 85(23):9119-23.
- Choudhry, S; Coyle, NE; Tang, H; Salari, K; Lind, D; Clark, SL; Tsai, H; Naqvi, M; Ung, Phong, A; Ung, N; Matallana, H; Avila, PC; Casal, J; Torres, A; Nazario, S; Castro, R; Battle, NC; Perez-Stable, EJ; Kwok, PY; Sheppard, D; Shriver, MD; Rodriguez-Cintron, W; Risch, N; Ziv, E; Burchard, GE. (2006). Population stratification confounds genetic association studies among Latinos. *Hum. Genet.* 118: 652–664.
- Cockerham, CC; Weir, BS. (1987). Correlations, descent measures: Drift with migration and mutation. *Proc Natl Acad Sci* 84: 8512-8514.
- Collins-Schramm, HE; Phillips, CM; Operario, DJ; Lee, JS; Weber, JL; Hanson, RL; Knowler, WC; Cooper, R; Li, H; Seldin, MLF. (2002). Ethnic-Difference Markers for Use in Mapping by Admixture Linkage Disequilibrium. *Am J Hum Genet* 70: 737–750.
- Comas, D; Calafell, F; Benchemsi, N; Helal, A; Lefranc, G; Stoneking, M; Batzer, MA; Bertranpetit, J; Sajantila, A. (2000). Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* 107:312-319.
- Campos-Sánchez, R; Barrantes, R; Silva, S; Escamilla, M; Ontiveros, A; Nicolini, H; Mendoza, R; Munoz, R; Raventos, H. (2006). Genetic analysis of three Hispanic populations from Costa Rica, Mexico, and the Southwestern United States using Y-Chromosome STR markers and mtDNA sequences. *Hum Biol* 78(5):551-563.
- Corte-Real, F; Andrade, L; Anjos, MJ; Carvalho, M; Vieira, DN; Carracedo, A; Vide, MC. (2000). Population genetics of nine STR *loci* in two populations from Brazil. *J Forensic Sci*. 45(2): 432-435.
- Dalton, G; Godinho, NMO; Gontijo, CC; Diniz, MECG; Alencar, GF; Amorim, CEG; Barcelos, RSS; Klautau-Guimarães, MN; Oliveira, SF. (2007). Population substructure by social stratification. In: 22nd Congress of the International Society for Forensic Genetics, Denmark. *Annals of the 22nd Congress of the International Society for Forensic Genetics*.

- Dellalibera, E; Havro, MLB; Souza, M; Kajihara, K; Mauricio-da-Silva, L; Silva, RS. (2004). Genetic analysis of 13 STR *loci* in the population from the State of Pernambuco, northeast Brazil. *F Sci Int* 146:57–59.
- Destro-Bisol, G; Boschi, I; Cagliá, A; Tofanelli, S; Pascali, V; Paoli, G; Spedini, G. (2000). Microsatellite variation in Central Africa: an analysis of intrapopulation and interpopulation genetic diversity. *Am. J. Phy. Anthropology*. 112:319-337.
- Dixon, L; Murray, C; Archer, E; Dobbins, A; Koumi, P; Gill, P. (2005). Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Sci. Int.* 154(1):62-77.
- Drobnic, K; Budowle, B. (2000). The analysis of three short tandem repeat (STR) *loci* in the Slovene population by multiplex PCR. *J. Forensic Sci.* 45(4):893-5.
- Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* 16(12):551-8.
- Excoffier, L; Smouse, P; Quattro, J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-491.
- Fagundes, NJR; Kanitz, R; Eckert, R; Valls, ACS; Bogó, MR; Salzano, FM; Smith, DG; Silva Jr, WA; Zago, MA; Ribeiro-dos-Santos, AKC; 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. *American Journal of Human Genetics*, v. 82, p. 1-10.
- Felsenstein, J. (1989). PHYLIP – phylogeny inference package. *Cladistics* 5:164–166.
- Fernandes, AT; Brehm, A. (2002). Population data of five STRs in three regions from Portugal. *F Sci Int* 129:72–74.
- Ferreira, FL; Leal-Mesquita, ER; Santos, SEB; Ribeiro-dos-Santos, AKC. (2005). Genetic characterization of the population of São Luís, MA, Brazil. *Genet Mol Biol* 28(1):22-31.
- Ferreira, LB; Mendes-Júnior, CT; Wiezel, CEV; Luizon, MR; Simões, AL. (2006). Genomic ancestry of a sample population from the State of São Paulo, Brazil. *Am J Hum Biol* 18:702-705.
- Gadelha, SR; Alcantara, LC; Costa, GC; Rios, DL; Galvão-Castro, B. (2005). Ethnic differences in the distribution of interleukin-6 polymorphisms among three Brazilian ethnic groups. *Hum Biol* 77(4):509-14.
- Gelernter, J; Kranzler, H; Cubells, JF; Ichinose, H; Nagatsu, T. (1998). DRD2 allele frequencies and linkage disequilibria, including the -141Cins/Del promoter polymorphism, in European-American, African-American, and Japanese subjects. *Genomics* 51:21-26.

- Gene, M; Moreno, P; Borrego, N; Pique, E; Brandt, C; Mas, J; Luna, M; Corbella, J; Huguët, E. (2001). The Bubi population of Equatorial Guinea characterized by HUMTH01, HUMVWA31A, HUMCSF1PO, HUMTPOX, D3S1358, D8S1179, D18S51 and D19S253 STR polymorphisms. *Int J Legal Med.* 114:298–300.
- Gigonzac, MAD. (2002). Análise de marcadores de DNA D7S460, TPA25, ACE, PV92 na população do estado de Goiás. Dissertação de mestrado, Universidade Federal de Goiás.
- Góes, ACS; Silva, DA; Gila, EHF; Silva, MTD; Pereira, RW; Carvalho, EF. (2004). Allele frequencies data and statistic parameters for 16 STR *loci*—D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX, D3S1358—in the Rio de Janeiro population, Brazil. *F Sci Int.* 140:131–132.
- Gomes, AC. (2000). Imigrantes italianos: entre a italianità e a brasilidade. In *Brasil 500 anos de povoamento*, IBGE, Rio de Janeiro, p 159-177.
- Gomes, AV; Maurício-da-Silva, L; Raposo, G; Vieira, JRC; Silva, RS. (2007). 13 STR *loci* frequencies in the population from Paraíba, Northeast Brazil. *Forensic Science International.* 173:231-234.
- Gonçalves, R; Jesus, J; Fernandes, AT; Brehm, A. (2002). Genetic profile of a multi-ethnic population from Guiné-Bissau (West African coast) using the new PowerPlex1 16 System kit. *F Sci Int.* 129:78–80.
- González-Andrade, F; Sánchez, K; González-Solórzano, J; Gascón, S; Martínez-Jarreta, B. (2007). Sex-specific genetic admixture of Mestizos, Amerindian Kichwas, and Afro-Ecuadorans from Ecuador. *Hum Biol* 79(1):51-77.
- González-Andrade, F; Sánchez-Q, D; Martínez-Jarreta, B. (2006). Genetic analysis of the Amerindian Kichwas and Afroamerican descendents populations from Ecuador characterised by 15 STR-PCR polymorphisms. *Forensic Sci Int* 160(2-3):231-5.
- Gotoda, T; Yamada, N; Murase, T; Miyake, S; Murakami, R; Kawamura, M; Kozaki, K; Mori, N; Shimano, H; Shimada, M. (1992). A newly identified null allelic mutation in the human lipoprotein lipase (LPL) gene of a compound heterozygote with familial LPL deficiency. *Biochim Biophys Acta.* 1138(4):353-6.
- Goudet, J; Raymond, M; Meeiis, T; Roussett, F. (1996). Testing Differentiation in Diploid Populations. *Genetics.* 144: 1933-1940.
- Grandy, DK ; Litt, M; Allen, L; Bunzow, JR; Marchionni, M; Makam, H; Reed, L; Magenis, RE; Civelle, O. (1989). The human dopamine D2 receptor gene is located in chromosome 11 at q22-q23 and identifies a TaqI polymorphism. *Am J Hum Genet* 45:778-785.
- Gratapaglia, D; Schmidt, AB; Costa e Silva, C; Stringher, C; Fernandes, AP; Ferreira, ME. (2001). Brazilian population database for the 13 STR *loci* of AmpFISTR Profiler Plus TM and Cofiler TM multiplex kits. *Forensic Sci. Int.* 11:891-4.

- Gregory, V. (2000). Imigração alemã: formação de uma comunidade teuto-brasileira. In Brasil 500 anos de povoamento, IBGE, Rio de Janeiro, p 141-157.
- Guerreiro, JF; Chautard-Freire-Maia, EA. (1988). ABO and Rh blood groups, migration and estimates of racial admixture for the population of Belém, State of Pará, Brazil. *Brazil J Genetics* 11:171-186.
- Hancock, JM. (1996). Microsatellites and other simple sequences in the evolution of the human genome. In Jackson, M.; Strachan, T. & Dover, G. *Human Genome Evolution*. BIOS. 306 p.
- Handley, L JL; Manica, A; Goudet, J; Balloux, F. (2007). Going the distance: human population genetics in a clinal world. *Trends in Genetics* 23(9):432-439.
- Herman, J. (2000). Cenário do encontro de povos: a construção do território. In Brasil 500 anos de povoamento, IBGE, Rio de Janeiro, p 17-34.
- Hernández-Gutiérrez, S; Hernández-Franco, P; Martínez-Tripp, S; Ramos-Kuria, M; Rangel-Villalobos, H. (2005). STR data for 15 *loci* in a population sample from the central region of Mexico. *F Sci Int* 151:97-100.
- Jeunemaitre, X; Soubrier, F; Kotelevtsev, YV; Lifton, RP; Willians, CS; Charru, A. (1992). Molecular basis of human hypertension: role of angiotensinogen. *Cell* 71(1):169-80.
- Jorde, LB; Wooding, SP. (2004). Genetic variation, classification and 'race'. *Nat Genet* 36(11 Suppl):S28-33.
- Karathanasis, SK. (1985). Apolipoprotein multigene family: tandem organization of human apolipoprotein AI, CIII, and AIV genes. *Proc Natl Acad Sci USA* 82(19):6374-8.
- Kenneth, KK. (2007). The Allele Frequency Database. A resource of gene frequency data on human populations supported by the U. S. National Science Foundation. Disponível em <http://alfred.med.yale.edu/alfred/>. Yale University.
- Kidd, KK; Morar, B; Castiglione, CM; Zhao, H; Pakstis, AJ; Speed, WC; Bonne-Tamir, B; Lu, RB; Goldman, D; Lee, C; Nam, YS; Grandy, DK; Jenkins, T; Kidd, J. (1998). A global survey of haplotype frequencies, and linkage disequilibrium at the DRD2 locus. *Hum Genet* 103:211-227.
- Klein, HS. (1986). *African Slavery in Latin America and the Caribbean*. Oxford University Press, New York, 316p.
- Klein, HS. (2002). As origens africanas dos escravos brasileiros *in Homo brasilis*. Organizado por Sérgio DJ Pena. Funpec- Editora, São Paulo, p 93-112.
- Klein, H.S. (2000). Migração internacional na história das Américas. In Fausto, B. *Fazer a América*. Ed. Universidade de São Paulo, p 13-31.
- Kruglyak, S; Durrett, RT; Schug, MD; Aquadro, CF. (1998). Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc Natl Acad Sci USA* 95(18):10774-8.

- Lee, ST; Nicholls, RD; Jong, MTC; Fukai, K; Spritz, RA. (1995). Organization and sequence of the human P gene and identification of transport proteins. *Genomics* 26, 354 – 363.
- Leite, FPN; Menegassi, FJ; Schwengber, SP; Raimann, PE; Albuquerque, TK. (2003). STR data for 09 autosomal STR markers from Rio Grande do Sul (southern Brazil). *F Sci Int* 132:223–224.
- Lewis, PO & Zaykin, D. (2002). - Genetic Data Analysis: version 1.1 for Windows 95/NT. <http://www.lewis.eeb.uconn.edu/lewishome/>. 2002. GDA user's manual.
- Lind, JM; Hutcheson-Dilks, HB; Williams, SM; Moore, JH; Essex, M; Ruiz-Pesini, E.; Wallace, DC; Tishkoff, SA; O'Brien, SJ; Smith, MW. (2007). Elevated male European and female African contributions to the genomes of African American individuals. *Human Genetics* 120:713-722.
- Liu, Y; Saha, N; Low, PS; Tay, JS. (1995). Linkage disequilibrium between two *loci* (5'DdeI) of the antithrombin III gene in three ethnic groups in Singapore. *Hum. Hered.* 45:192-198.
- Liu, K; Muse, SZ. (2005). PowerMarker: integrated analysis environment for genetic marker data. *Bioinformatics* 21:2128-9.
- López-Camelo, JS; Cabello, PH; Dutra, MG. (1996). A simple model for the estimation of congenital malformation frequencies in racial mixed populations. *Braz. J. Genet.* 19:659-663.
- Lovo-Gómez, J; Salas, A; Carracedo, A. (2007). Microsatellite autosomal genotyping data in four indigenous populations from El Salvador. *F Sci Int* 170:86-91.
- Loyo, MA; de Guerra, DC; Izaguirre, MH; Rodriguez-Larralde, A. (2004). Admixture estimates for Churuguara, a Venezuelan town in the State of Falcón. *Annals of Hum Biol* 31(6):669-80.
- Luis, JR; Terreros, MC; Martínez, L; Rojas, D; Herrera, RJ. (2003). Two problematic human polymorphic Alu insertions. *Electrophoresis* 24(14):2290-2294.
- Luizon, MR. (2007). Dinâmica da mistura étnica em comunidades remanescentes de quilombo brasileiras. Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.
- Luizon, MR; Mendes-Junior, CT; de Oliveira, SF; Simões, AL. (2008). Ancestry informative markers in Amerindians from Brazilian Amazon. *Am. J. Hum. Biol.* 20(1):86-90.
- Luna-Vazquez, A; Vilchis-Dorantes, G; Paez-Riberos, LA; Muñoz-Vallec, F; González-Martin, A; Rangel-Villalobos, H. (2003). Population data of nine STRs of Mexican-Mestizos from Mexico City. *F Sci Int* 136:96–98.

- Machado, TMBM. (2008). Ancestralidade em Salvador-BA. Dissertação de mestrado em Biotecnologia em Saúde e Medicina Investigativa – Centro de Pesquisa Gonçalo Moniz – FIOCRUZ-BA.
- Manica, A; Prugnolle, F; Balloux, F. (2005). Geography is a better determinant of human genetic differentiation than ethnicity. *Hum. Genet.* 118(3-4):366-71.
- Marjanovic, D; Kapur, L; Drobnic, K; Budowle, B; Pojskic, N; Hadziselimovic, R. (2004). Comparative Study of Genetic Variation at 15 STR *Loci* in Three Isolated Populations of the Bosnian Mountain Area. *Human Biology* 76(1): 15-31.
- Martínez, H; Rodríguez-Larralde, A; Izaguirre, MH; de Guerra, DC. (2007). Admixture estimates for Caracas, Venezuela, based on autosomal, Y-chromosome, and mtDNA markers. *Hum Biol.* 79(2):201-213.
- Martínez-Marignac, VL; Bertoni, B; Parra, EJ; Bianchi, NO. (2004). Characterization of admixture in an urban sample from Buenos Aires, Argentina, using uniparentally and biparentally inherited genetic markers. *Hum Biol.* 76(4):543-557.
- Martinez-Marignac VL, Valladares A, Cameron E, Chan A, Perera A, Globus-Goldberg R, Wacher N, Kumate J, McKeigue P, O'Donnell D, Shriver MD, Cruz M, Parra EJ. (2007). Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. *Hum. Genet.* 120(6):807-19.
- Martinovic, I; Barac, L; Furac, I; Janićijević, B; Kubat, M; Perićić, M; Vidović, B; Rudan, P. (1999). STR polymorphisms in the population of the island of Hvar. *Hum Biol.* 71(3):341-52.
- McCombs, JL; Yang, F; Bowman, BH; McGill, JR; Moore, CM. (1986). Chromosomal localization of group-specific component by in situ hybridization. *Cytogenet Cell Genet.* 42(1-2):62-64.
- Melo, JO. (1996). Historia de Colômbia. La dominación Española. Imprenta Nacional de Colômbia. Presidencia de la República, 358p.
- Mendes-Jr, CT; Simões, AL. (2001). Alu insertions and ethnic composition in a Brazilian population sample. *IJHG* 1(4):249-254.
- Miller, SA; Dykes, DD; Polesky, HF. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research.* 161, 1215.
- Morales, JÁ; Monterrosa, JC; Puente, J. (2004). Population genetic data from El Salvador (Central America) using AmpFISTR Identifiler PCR amplification kit. *International Congress Series* 1261:223-225.
- Morera, B; Barrantes, R; Marin-Rojas, R. (2003). Gene admixture in the Costa Rican population. *Annals of Hum Genet.* 67:71-80.
- Nalls, MA; Wilson, JG; Patterson, NJ; Tandon, A; Zmuda, JM; Huntsman, S; Garcia, M; Hu, D; Li, R; Beamer, BA; Patel, KV; Akylbekova, EL; Files, JC; Hardy, CL; Buxbaum, SG; Taylor, HA; Reich, D; Harris, TB; Ziv, E. (2008). Admixture mapping

- of white cell count: genetic locus responsible for lower white blood cell count in the health ABC and Jackson heart studies. *Am J Hum Genet* 82:81-87.
- National Geographic Society. (2008). *Atlas National Geographic: América do Sul*. Editora Abril, São Paulo/SP, 95p.
- Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press, New York, NY, USA.
- Nei, M. (1972). Genetic distances between populations. *Am. Nat.* 106:283-292.
- Oliveira, SF; Ferreira, LB; Klautau-Guimarães, MN; Ribeiro, GGBL; Simões, AL. (2006). Reconstrucción historica de poblaciones afro-descendientes aisladas de Brasil: el constraste entre la contribuciones masculina y feminina. *Diversidad biologica y saude humana*, Murcia – Sociedad Espanhola de Antropologia, pag 203-210.
- Oliveira, SF. (1999). *Inserções de Alu em populações indígenas da Amazônia brasileira*. Tese apresentada para obtenção do título de doutor, Faculdade de Medicina de Ribeirão Preto. Departamento de Genética, Universidade de São Paulo.
- Oliveira, SF; Ferreira, LB. (2004). Biological view of the inexistence of human races. *Eubios Journal Of Asian And International Bioethics*, Ibaraki, 14(2): 60-63.
- Olson, S. (2002). *Mapping human history: genes, race, and our commom origins*. Mariner Books, New York/NY, 292p.
- Palacín, L; Garciam LF; Amado, J. (1995). *História de Goiás em documentos: Colônia*. Editora UFG, Goiânia/GO, 190p.
- Palacín, L; Moraes, MAS. (1994). *História de Goiás*. Editora UCG, 6a edição, Goiânia/GO, 122p.
- Paredes, M; Crespillo, M; Luque, JÁ; Valverde, JL. (2003a). STR frequencies for the PowerPlex 16 System Kit in a population from Northeast Spain. *F. Sci. Int.* 135:75-78.
- Paredes, M; Galindo, A; Bernala, M; Ávila, S; Andrade, D; Vergara, C; Rincón, M; Romero, RE; Navarrete, M; Cárdenas, M; Ortega, J; Suarez, D; Cifuentes, A; Salas, A; Carracedo, A. (2003). Analysis of the CODIS autosomal STR *loci* in four main Colombian regions. *F. Sci. Int.* 137:67–73.
- Parra, EJ; Kittles, RA; Argyropoulos, G; Pfaff, CL; Hiester, K; Bonilla, C; Sylvester, N; Parrish-Gause, D; Garvey, WT; Jin, L; McKeigue, PM; Kamboh, MI; Ferrell, RE; Pollitzer, WS; Shriver, MD. (2001). Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol* 114(1):18-29.
- 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(6):1839-51.

- Parra, FC; Amado, RC; Lambertucci, JR; Rocha, J; Antunes, CM; Pena, SD. (2003). Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci USA* 100(1):177-82.
- Pena, SDJ. (2002). O retrato molecular do Brasil *in* Homo brasilis: Aspectos Genéticos, Linguísticos, Históricos e Socioantropológicos da Formação do Povo Brasileiro. 1ª ed. Ribeirão Preto: FUNPEC, volume 1, 191 p.
- Pfaff, CL; Parra, EJ; Bonilla, C; Hiester, K; McKeigue, PM; Kamboh, MI; Hutchinson, RG; Ferrell, RE; Boerwinkle, E; Shriver, MD. (2001). Population Structure in Admixed Populations: Effect of Admixture Dynamics on the Pattern of Linkage Disequilibrium. *Am J Hum Genet* 68(1):198-207.
- Pimentel, BJ; De Azevedo, DA; Silva, LA. (2004). Population genetics of eleven STR *loci* in the State of Sergipe, Northeastern Brazil. *J Forensic Sci.* 49(2): 49(2):402-3.
- Ramos, JS. (2002). O Brasil sob o paradigma racial: sociologia histórica de uma representação. In Pena, S.D.J. Homo brasilis. Editora Funpec, São Paulo/SP, p 131-148.
- Raymond, M; Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Heredity* 86:248-249.
- Reis, JJ. (2000). Presença negra: conflitos e encontros. In Brasil 500 anos de povoamento, IBGE, Rio de Janeiro, p 79-100.
- Reynolds, JB; Weir BS; Cockerham, CC. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105: 767-779.
- Ribeiro, D. (1996). Os índios e a civilização: a integração das populações indígenas no Brasil moderno. Companhia das Letras, São Paulo/SP, 559p.
- Rock, D. (1985). Argentina: 1516-1982. University of California Press, Los Angeles/California, 478p.
- Rodrigues, EMR; Palha, TJBF; Santos, SEB. (2006). Allele frequencies data and statistic parameters for 13 STR *loci* in a population of the Brazilian Amazon Region. *Forensic Science International* 168(2-3):244-247.
- Rodríguez, A; Arrieta, G; Sanou, I; Vargas, MC; García, O; Yurrebaso, I; Pérez, JÁ; Villalta, M; Espinoza, M. (2007). Population genetic data for 18 STR *loci* in Costa Rica. *Forensic Science International* 168:85-88.
- Rohlf, FJ. (1992). NTSYS-pc Numerical taxonomy and multivariate analysis system. Version 1.70, New York: Applied Biostatistics.
- Rowold, DJ; Herrera, RJ. (2003). Inferring recent human phylogenies using forensic STR technology. *Forensic Sci. Int.* 133(3):260-5.
- Ruitberg, CM; Reeder, DJ; Butler JM. (2001). STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res.* 29(1): 320-322.

- Saitou, N; Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 4(4):406-25.
- Salzano, MF. (2004). Interethnic variability and admixture in Latin América – social implications. *Rev Biol Tropical.* 52(3): 405-415.
- Salzano, MF; Callegari-Jacques, SM. (1988). South americans indians – a case study in evolution. Oxford University Press, New York.
- Sans, M. (2000). Admixture studies in Latin America: from the 20th to the 21st century. *Hum Biol* 72(1):155-77.
- Santos, SEB; Guerreiro, JF; Salzano, FM; Weimer, TA; Hutz, MH; Franco, MHL. (1987). Mobility, blood genetic traits and race mixture in the Amazonian population of Oriximiná. *Revista Brasileira de Genética.* 4:745-759.
- Schneider, S; Rosslie, D; Excoffier, L. (2000). Arlequin Ver 2.000, A software for population genetics data analysis. University of Geneva, Geneva: Genetics and Biometry Laboratory. Disponível em <http://anthropologie.unige.ch/arlequin>.
- Schröer, K; Schmitt, C; Staak, M. (2000). Analysis of the co-amplified STR *loci* D1S1656, D12S391 and D18S51: population data and validation study for a highly discriminating triplex-PCR. *Forensic Sci. Int.* 113(1-3):17-20.
- Serre, D; Pääbo, S. (2004). Evidence for gradients of human genetic diversity within and among continents. *Genome Research* 14:1679-1685.
- Shriver, MD; Parra, EJ; Dios, S; Bonilla, C; Norton, H; Jovel, C; Pfaff, C; Jones, C; Massac, A; Cameron, N; Baron, A; Jackson, T; Argyropoulos, G; Jin, L; Hoggart, CJ; McKeigue, PM; Kittles, RA. (2003). Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum. Genet.* 112(4):387-99.
- 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(4):957-64.
- Silberstein, CF. (2000). A imigração espanhola na Argentina (1880-1930). In Fausto, B. *Fazer a América.* Ed. Universidade de São Paulo.
- 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, EB; Dellalibera, E; Souza, M; Silva, RS; Maurício-da-Silva, L. (2002). Population genetics of eight STR *loci* - CSF1PO, TPOX, TH01, D16S359, D7S820, D13S317, F13B and LPL in a Brazilian population from the State of Piauí, Northeast Brazil. *Forensic Sci. Int.* 126:90-92.

- Silva, FF; Dellalibera, E; Nigam, P; Mauricio-da-Silva, L; Santos Silva, R. (2003). Microsatellite markers in the population from Rio Grande do Norte, Northeastern Brazil. *J Forensic Sci.* 48(5):1189-90.
- 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, R; Moura-Neto, RS. (2004). Genetic diversity and admixture data on 11 STRs (F13B, TPOX, CSF1PO, F13A01, D7S820, LPL, TH01, vWA, D13S317, FESFPS, and D16S539) in a sample of Rio de Janeiro European-descendants population, Brazil. *Forensic Sci. Int.* 142:51–53.
- Simmons, AD; Rodriguez-Arroyo, G; Rodríguez-Larralde, A. (2007). Admixture estimates base don ABO, Rh and nine STRs in two Venezuelan Regions. *Annals of Hum Biol* 34(1): 56-67.
- Steinlechner, M; Schmidt, K; Kraft, HG; Utermann, G; Parson, W. (2002). Gabon black population data on the ten short tandem repeat *loci* D3S1358, VWA, D16S539, D2S1338, D8S1179, D21S11, D18S51, D19S433, TH01 and FGA. *Int J Legal Med.* 116 :176–178.
- Tang, H; Choudhry, S; Mei, R; Morgan, M; Rodrigues-Cintron, W; Burchard, EG; Risch, NJ. (2007). Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am J Hum Genet* 81:626-633.
- Tofanelli, S; Boschi, I; Bertoneri, S; Coia, V; Taglioli, L; Franceschi, MG; Destro-Bisol, G; Pascali, V; Paoli, G. (2003). Variation at 16 STR *loci* in Rwandans (Hutu) and implications on profile frequency estimation in Bantu-speakers. *Int J Legal Med.* 117:121–126.
- Tomás, G; Seco, L; Seixas, S; Faustino, P; Lavinha, J; Rocha, J. (2002). The peopling of São Tomé (Gulf of Guinea): origins of slave settlers and admixture with the Portuguese. *Hum. Biol.* 74(3):397-411.
- Toscanini, U; Gusmão, L; Berardi, G; Amorim, A; Carracedo, A; Salas, A; Raimondi, E. (2007). Testing for genetic structure in different urban Argentinian populations. *Forensic Sci Int.* 165:35-40.
- Tournamille, C; Le Van Kim, C; Gane, P; Cartron, JP; Colin, Y. (1995). Molecular basis and PCR-DNA typing of the Fya/fyb blood group polymorphism. *Hum Genet.* 95(4):407-10.
- Underhill, PA. (2003) *Inferring Human History: Clues from Y-Chromosome Haplotypes.* Cold Spring Harbor Symposia on Quantitative Biology, Volume LXVIII. Cold Spring Harbor Laboratory Press 0-87969-709-1/04.
- Vainfas, R. (2000). História indígena: 500 anos de despovoamento. In *Brasil 500 anos de povoamento*, IBGE, Rio de Janeiro, p 35-59.

- Venâncio, RP. (2000). Presença portuguesa de colonizadores a imigrantes. In Brasil 500 anos de povoamento, IBGE, Rio de Janeiro, p 61-77.
- Venter, JC; Adams, MD; Myers, EW; Li, PW; Mural, RJ; Sutton, GG; Smith, HO; Yandell, M; Evans, CA; Holt, RA; Gocayne, JD; Amanatides, P; Ballew, RM; Huson, DH; Wortman, JR; Zhang, Q; Sodira, CD; Zheng, XH; Chen, L; Skupski, M; Subramanian, G; Thomas, PD; Zhang, J; Miklos, GLG; Nelson, C; Broder, S; Clark, AG; Nadeau, J; Mckusick, VA; Zinder, N; Levine, AJ; Roberts, RJ; Simon, M; Slayman, C; Hunkapiller, M; Bolanos, R; Delcher, A; Dew, I; Fasulo, D; Flanigan, M; Florea, L; Halpern, A; Hannenhalli, S; Kravitz, S; Levy, S; Mobarry, C; Reinert, K; Remington, K; Abu-Threideh, J; Beasley, E; Biddick, K; Bonazzi, V; Brandon, R; Cargill, M; Chandramouliswaran, I; Charlab, R; Chaturvedi, K; Deng, Z; De Francesco, V; Dunn, P; Eilbeck, K; Evangelista, C; Gabrielian, AE; Gan, W; Ge, W; Gong, F; Gu, Z; Guan, P; Heiman, TJ; Higgins, ME; Ji, RR; Ke, Z; Ketchum, KA; Lai, Z; Lei, Y; Li, Z; Li, J; Liang, Y; Lin, X; Lu, F; Merkulov, GV; Milshina, N; Moore, H.M.; Naik, A.K.; Narayan, V.A.; Neelam, B.; Nusskern, D.; Rusch, D.B.; Salzberg, S; Shao, W; Shue, B; Sun, J; Wang, A; Wang, ZY; Wang, X; Wang, J; Wei, MH; Wides, R; Xiao, C; Yan, C; Yao, A; Ye, J; Zhan, M; Zhang, W; Zhang, H; Zhao, Q; Zheng, L; Zhong, F; Zhong, W; Zhu, SC; Zhao, S; Gilbert, D; Baumhueter, S; Spier, G; Carter, C; Cravchilk, A; Woodage, T; Ali, F; An, H; Awe, A; Baldwin, D; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, ML; Curry, L; Danaher, S; Davenport, L; Desilets, R; Dietz, S; Dodson, K; Doup, L; Ferreira, S; Garg, N; Gluecksmann, A; Hart, B; Haynes, J; Haynes, C; Heiner, C; Hladun, S; Hostin, D; Houck, J; Howland, T; Ibegwam, C; Johnson, F; Kalush, F; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T; McMullen, I; Moy, M; Moy, L; Murphy, B; Nelson, K; Pfannkoch, C; Pratts, E; Puri, V; Qureshi, H; Reardon, M; Rodriguez, R; Rogers, YH; Romblad, D; Rhfel, B; Scott, R; Sitter, C; Smallwood, M; Stewart, E; Strong, R; Suh, EL; Thomas, R; Tint, NN; Tse, S; Vech, C; Wang, G; Wetter, J; Williams, S; Williams, M; Windsor, S; Winn-Deen, E; Wolfe, K; Zaveri, J; Zaveri, K; Abril, JF; Guigó, R; Campbell, MJ; Sjolander, KV; Karlak, B; Kejariwal, A; Mi, H; Lazareva, B; Hatton, T; Narechania, A; Diemer, K; Muruganujan, A; Guo, N; Sato, S; Bafna, V; Istrail, S; Lppert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J; Blick, L; Caminha, M; Carnes-Stine, J; Caulk, P; Chiang, YH; Coyne, M; Dahlke, C; Mays, AD; Dombroski, M; Donnelly, M; Ely, D; Esparham, S; Fosler, C; Gire, H; Glanowski, S; Glasser, K; Glodek, A; Gorkhov, M; Graham, K; Gropman, B; Harris, M; Heil, J; Henderson, S; Hoover, J; Jennings, D; Jordan, C; Jordan, J; Kasha, J; Kagan, L; Kraft, C; Levitsky, A; Lewis, M; Liu, X; Lopez, J; Ma, D; Majoros, W; McDaniel, J; Murphy, S; Newman, M; Nguyen, T; Nguyen, N; Nodell, M; Pan, S; Peck, J; Peterson, M; Rowe, W; Sanders, R; Scott, J; Simpson, M; Smith, T; Sprague, A; Stockwell, T; Turner, R; Venter, E; Wang, M; Wen, M; Wu, D; Wu, M; Xia, A; Zandieh, A; Zhu, X. (2001). The sequence of the Human Genome. *Science* 291:1304-1350.

- Vieira, TC. (2003). Marcadores Alu em banco de dados da população do estado de Goiás. Dissertação de mestrado, Universidade Federal de Goiás.
- Wang, J; Caballero, A. (1999). Developments in predicting the effective size of subdivided populations. *Heredity* 82:212-226.
- Wang, S; Lewis Jr, CM; Jakobsson, M; Ramachandran, S; Ray, N; Bedoya, G; Rojas, W; Parra, MV; Molina, JA; Gallo, C; Mazzotti, G; Poletti, G; Hill, K; Hurtado, AM; Salzano, FM. (2007). Genetic variation and population structure in Native Americans. *PLOS Genetics*, 3:2049-2067.
- 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.doi:10.1371/journal.pgen.1000037
- Webb, GC; Coggan, M; Ichinose, A; Board, PG. (1989). Localization of the coagulation factor XIII B subunit gene (F13B) to chromosome bands 1q31-32.1 and restriction fragment length polymorphism at the locus. *Human Genetics* 81:157-160.
- Weir, BS; Cockerham, CC. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370.
- Whittle, MR; Romano, NL; Negreiros, VAC. (2004). Updated Brazilian genetic data, together with mutation rates, on 19 STR *loci*, including D10S1237. *Forensic Sci. Int.* 13:9207-10.
- Yang-Feng, TL; Opdenak, G; Volckaert, G; Francke, U. (1986). Human tissue-type plasminogen activator gene located near chromosomal breakpoint in myeloproliferative disorder. *Am J Hum Genet* 39(1):79-87.
- Zúñiga, J; Ilzarbe, M; Acunha-Alonzo, V; Rosetti, F; Herbert, Z; Romero, V; Almeciga, I; Clavijo, O; Stern, JNH; Granados, J; Fridkis-Hareli, M; Morrison, P; Azocar, J; Yunis, EJ. (2006). Allele frequencies for 15 autosomal STR *loci* and admixture estimates in Puerto Rican Americans. *F Sci Int* 164:266-270.

Sítios da Internet

CODEPLAN: [http:// www.codeplan.df.gov.br](http://www.codeplan.df.gov.br) - Acessado em janeiro de 2008.

dbSNP NCBI database: <http://www.ncbi.nlm.nih.gov/SNP> - acessado em 03/05/2008.

Funai: <http://www.funai.gov.br> - Acessado em maio de 2006.

Instituto Brasileiro de Geografia e Estatística (IBGE) – PNAD 2007:
<http://www.ibge.gov.br/home/estatística/população> - Acessado em abril de 2008.

Universidade de Düsseldorf:

<http://www.uni-duesseldorf.de/WWW/MedFak/Serology/database.html> - acessado em 03/05/2008

Governo do Distrito Federal: www.districtofederal.df.gov.br.

<http://statgen.ncsu.edu/powermarker/downloads.htm>.

<http://taxonomy.zoology.gla.ac.uk/rod/rod.html>.

NTSYS-pc versão 1.70 (Applied Biostatistics Inc, Setauket, NY, USA)

SPSS 15.0 para Windows, versão 2006 - <http://www.spss.com>.

ANEXO 1: TABELAS SUPLEMENTARES**Tabela 1:** Marcadores utilizados para análise populacional de 13 países da América Latina e Caribe, número de indivíduos amostrados, diversidade gênica (Hsk) de Nei (1987) e seus receptivos desvios padrões e referências bibliográficas utilizadas para levantamento das frequências alélicas dos 13 marcadores STRs.

	Argentina	Bahamas	Brasil	Chile	Colômbia	Costa Rica	El Salvador	Equador	Jamaica	México	Peru	Porto Rico	Venezuela
CSF1PO													
N	1628	158	8228	397	3618	949	296	317	208	354	100	205	1395
Hsk	0,724	0,803	0,749	0,731	0,731	0,715	0,724	0,723	0,777	0,714	0,712	0,705	0,738
Desvio padrão	0,004	0,014	0,002	0,008	0,003	0,006	0,012	0,010	0,013	0,010	0,024	0,017	0,005
Ref.	4, 22, 23, 26, 42	29	5, 7, 11, 12, 13, 15, 32, 35, 37, 38, 39, 40, 41, 44	14, 29	1, 3, 16, 25, 26, 30, 34	27, 29, 36	29	17	29	8, 18	31	43	2, 9, 33
D3S1358													
N	1969	157	3771	80	3995	1028	617	317	194	485	100	205	576
Hsk	0,754	0,736	0,776	0,729	0,745	0,777	0,713	0,694	0,743	0,765	0,661	0,771	0,771
Desvio padrão	0,005	0,016	0,002	0,033	0,004	0,005	0,012	0,017	0,016	0,010	0,031	0,013	0,008
Ref.s	4, 19, 22,	29	6, 7, 11,	29	16, 25, 26,	27, 29, 36	29	17	29	8, 18, 21	31	43	2, 9

	Argentina	Bahamas	Brasil	Chile	Colômbia	Costa Rica	El Salvador	Equador	Jamaica	México	Peru	Porto Rico	Venezuela
	23, 28, 42		10, 15, 20, 32, 37, 44		30, 34								
D5S818													
N	1951	162	4754	81	4112	1028	619	317	244	484	100	205	907
Hsk	0,715	0,748	0,734	0,699	0,737	0,727	0,698	0,726	0,725	0,719	0,714	0,731	0,745
Desvio padrão	0,007	0,017	0,003	0,038	0,004	0,008	0,013	0,017	0,016	0,015	0,033	0,016	0,008
Ref.	4, 19, 22, 23, 28, 42	29	6, 7, 10, 11, 15, 20, 35, 37, 44	29	1, 16, 25, 26, 30, 34	27, 29, 36	29	17	29	8, 18, 21	31	43	2, 9
D7S820													
N	2252	324	6610	193	4148	1228	619	317	244	467	100	205	1409
Hsk	0,779	0,784	0,809	0,779	0,788	0,793	0,779	0,750	0,778	0,787	0,750	0,809	0,799
Desvio padrão	0,004	0,011	0,002	0,016	0,003	0,004	0,007	0,010	0,009	0,008	0,022	0,012	0,004
Ref.	4, 19, 22, 23, 26, 28, 42	29	5, 6, 7, 10, 11, 12, 15, 20, 35, 37, 38, 39, 40, 41, 44	29	1, 3, 16, 25, 26, 30, 34	27, 29, 36	29	17	29	8, 18, 21	31	43	2, 9, 33
D8S1179													
N													
Hsk	0,813	0,795	0,810	0,774	0,792	0,806	0,786	0,795	0,794	0,797	0,814	0,826	0,807

	Argentina	Bahamas	Brasil	Chile	Colômbia	Costa Rica	El Salvador	Equador	Jamaica	México	Peru	Porto Rico	Venezuela
Desvio padrão	0,004	0,017	0,003	0,026	0,004	0,006	0,008	0,011	0,014	0,009	0,019	0,010	0,008
Ref.	4, 19, 22, 23, 28, 42	29	6, 7, 10, 11, 12, 15, 20, 35, 37, 44	29	3, 16, 30, 34	27, 29, 36	29	17	29	8, 18, 21	31	43	2, 9
D13S317													
N	1978	157	4165	81	2682	1028	619	317	194	484	100	205	536
Hsk	0,830	0,721	0,791	0,841	0,811	0,823	0,838	0,836	0,693	0,818	0,824	0,808	0,805
Desvio padrão	0,003	0,021	0,002	0,009	0,003	0,005	0,004	0,007	0,020	0,006	0,016	0,013	0,005
Ref.	4, 19, 22, 23, 26, 28, 42	29	5, 6, 7, 10, 11, 12, 15, 20, 32, 35, 37, 38, 39, 40, 41, 44	29	16, 25, 26, 30, 34	27, 29, 36	29	17	29	8, 18, 21	31	43	2, 9, 33
D16S539													
N	1782	158	6289	193	1686	949	296	317	206	354	100	205	1282
Hsk	0,789	0,802	0,799	0,794	0,797	0,800	0,787	0,797	0,795	0,775	0,802	0,778	0,795
Desvio padrão	0,004	0,014	0,002	0,011	0,003	0,004	0,008	0,007	0,012	0,010	0,011	0,013	0,004
Ref.	4, 22, 23, 26, 42	29	5, 11, 12, 15, 32, 35,	29	30, 34	27, 29, 36	29	17	29	8, 18	31	43	2, 9, 33

	Argentina	Bahamas	Brasil	Chile	Colômbia	Costa Rica	El Salvador	Equador	Jamaica	México	Peru	Porto Rico	Venezuela
			37, 38, 39, 40, 41, 44										
D18S51													
N	1943	157	4096	80	2667	1028	619	317	194	484	100	205	255
Hsk	0,873	0,883	0,882	0,865	0,878	0,900	0,877	0,862	0,881	0,873	0,864	0,873	0,886
Desvio padrão	0,002	0,009	0,001	0,016	0,002	0,003	0,004	0,008	0,008	0,004	0,012	0,006	0,006
Ref.	4, 19, 22, 23, 28, 42	29	6, 7, 10, 11, 12, 15, 20, 32, 35, 37, 44	29	16, 30, 34	27, 29, 36	29	17	29	8, 18, 21	31	43	9
D21S11													
N	1973	157	4148	80	2680	1028	296	317	194	484	100	205	255
Hsk	0,840	0,863	0,858	0,849	0,838	0,835	0,843	0,844	0,854	0,839	0,845	0,831	0,843
Desvio padrão	0,004	0,011	0,002	0,021	0,004	0,006	0,010	0,008	0,013	0,008	0,013	0,013	0,012
Ref.	4, 19, 22, 23, 28, 42	29	6, 7, 10, 11, 12, 15, 20, 35, 37, 44	29	16, 30, 34	27, 29, 36	29	17	29	8, 18, 21	31	43	9

	Argentina	Bahamas	Brasil	Chile	Colômbia	Costa Rica	El Salvador	Equador	Jamaica	México	Peru	Porto Rico	Venezuela
FGA													
N	1972	157	3550	81	3990	1028	619	317	194	484	100	205	255
Hsk	0,882	0,877	0,877	0,892	0,880	0,872	0,872	0,868	0,881	0,883	0,885	0,877	0,891
Desvio padrão	0,002	0,010	0,001	0,011	0,001	0,003	0,004	0,007	0,009	0,004	0,011	0,006	0,004
Ref.	4, 19, 22, 23, 28, 42	29	6, 7, 10, 11, 12, 15, 20, 35, 37, 44	29	25, 26, 30, 34	27, 29, 36	29	17	29	8, 18, 21	31	43	9
TH01													
N	1531	158	7562	397	3936	949	296	317	208	354	100	205	1402
Hsk	0,775	0,760	0,799	0,789	0,760	0,768	0,728	0,754	0,759	0,790	0,694	0,801	0,791
Desvio padrão	0,004	0,019	0,001	0,008	0,003	0,006	0,015	0,011	0,014	0,011	0,032	0,006	0,003
Ref.	4, 22, 23, 26, 42	29	5, 7, 11, 12, 13, 15, 32, 35, 37, 38, 39, 40, 41, 44	14, 29	1, 3, 25, 26, 30, 34	27, 29, 36	29	17	29	8, 18	31	43	2, 9, 33
TPOX													
N	1819	158	8707	397	3641	949	296	317	208	354	100	205	1402
Hsk	0,677	0,782	0,705	0,674	0,684	0,684	0,681	0,603	0,750	0,665	0,662	0,717	0,697
Desvio	0,007	0,015	0,003	0,015	0,006	0,009	0,020	0,021	0,018	0,017	0,035	0,018	0,008

	Argentina	Bahamas	Brasil	Chile	Colômbia	Costa Rica	El Salvador	Equador	Jamaica	México	Peru	Porto Rico	Venezuela
padrão													
Ref.	4, 22, 23, 26, 42	29	5, 7, 11, 12, 13, 15, 32, 35, 37, 38, 39, 40, 41, 44	14, 29	1, 16, 25, 26, 30, 34	27, 29, 36	29	17	29	8, 18	31	43	2, 9, 33
vWA													
N	2253	162	7941	72	4150	949	618	317	244	484	100	205	1398
Hsk	0,780	0,829	0,806	0,789	0,772	0,784	0,753	0,737	0,807	0,770	0,711	0,822	0,800
Desvio	0,004	0,012	0,002	0,024	0,003	0,007	0,009	0,014	0,012	0,010	0,028	0,010	0,005
padrão													
Ref	4, 19, 22, 23, 25, 28, 42	29	5, 6, 7, 10, 11, 12, 13, 15, 20, 32, 35, 37, 40, 41, 44	14	1, 3, 16, 25, 26, 30, 34	27, 29, 36	29	17	29	8, 18, 21	31	43	2, 9, 33
Hsk média	0,790	0,800	0,800	0,780	0,780	0,790	0,780	0,770	0,790	0,780	0,760	0,800	0,800

1. Acosta, MA; Brion, M; Lareu, MV; Carracedo, A. (2002). Genetic data on eight STRs (D5S818, D7S820, F13B, LPL, TH01, TPOX, VWA31, CSF1PO) from a Colombian population. *Forensic Science International* 129:216–218.
2. Alvarez, M; Chiarello, A; Dictamen, A. (2007). Genetic Population Data do 15 Autosomal *Loci* from Central Region of Venezuela. 22º Congress of the International Society for Forensic Genetics, Copenhagen, Denmark, poster 213.
3. Benítez-Páez, A; Reyes, HO. (2003). Allelic frequencies at 12 STR *loci* in Colombian population. *Forensic Science International* 136:86–88.

4. Berardi, G; Toscanini, U; Raimondi, E. (2003). STR data for PowerPlex116 System from Buenos Aires population, Argentina. *Forensic Science International* 134:222–224.
5. Bydlowska, SP; Moura-Neto, RS; Soares, RPS; Silva, R; Debes-Bravo, AA; Morganti, L. (2003). Genetic data on 12 STRs (F13A01, F13B, FESFPS, LPL, CSF1PO, TPOX, TH01, vWA, D16S539, D7S820, D13S317, D5S818) from four ethnic groups of São Paulo, Brazil. *Forensic Science International* 135:67–71.
6. Caine, LM; Corte-Real, F; Anjos, MJ; Carvalho, M; Serra, A; Antunes, H; Vide, MC; Vieira, DN. (2003). Allele frequencies of 13 *loci* in the Santa Catarina Population of Southern Brazil. *Journal of Forensic Sciences* 48(4):901-2.
7. 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.
8. Cerda-Flores, RM; Budowle, B; Jin, L; Barton, AS; Keka, R; Chakraborty, R. (2002). Maximum likelihood estimates of admixture in Northeastern Mexico using 13 short tandem repeat *loci*. *Am Journal of Human Biology* 14:429-439.
9. Chiurillo, MA; Morales, A; Mendes, A;M; Lander, N; Tovar, F; Fuentes, A; Ramírez, JL. (2003). Genetic profiling of a central Venezuelan population using 15 STR markers that may be of forensic importance. *Forensic Science International* 136:99–101.
10. Corte-Real, F; Andrade, L; Anjos, MJ; Carvalho, M; Vieira, DN; Carracedo, A; Vide, MC. (2000). Population genetics of nine STR *loci* in two populations from Brazil. *Journal of Forensic Sciences*. 45(2):432-5.
11. Comunicação pessoal, dados de frequência da população do Distrito Federal.
12. Dellalibera, E; Havro, MLB; Souza, M; Kajihara, K; Mauricio-da-Silva, L; Silva, RS. (2004). Genetic analysis of 13 STR *loci* in the population from the State of Pernambuco, northeast Brazil. *Forensic Science International* 146:57–59.
13. Ferreira, LB; Mendes-Junior, CT; Wiezel, CEV; Luizon, MR; Simões, AL. (2006). Genomic Ancestry of a Sample Population From the State of São Paulo, Brazil. *American Journal of Human Biology* 18:702–705.
14. Figueroa, CC; Lic, S; Acuña, M; Cifuentes, L. (2005). Gene frequencies for six STR *loci* in a Chilean population of mixed ancestry. *Journal of Forensic Science* 50(3):705-6.

-
15. Góes, ACS; Silva, DA; Gila, EHF; Silva, MTD; Pereira, RW; Carvalho, EF. (2004). Allele frequencies data and statistic parameters for 16 STR *loci*—D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX, D3S1358—in the Rio de Janeiro population, Brazil. *Forensic Science International* 140:131–132.
 16. Gomez, MV; Reyes, ME; Cardenas, H; Garcia, O. (2003). Genetic Variation For 12 STRs *loci* in a Colombian Population (Department Of Valle del Cauca). *Forensic Science International* 137:235–237.
 17. González-Andrade, F; Sánchez-Qa, D; Martínez-Jarreta, B. (2003). Genetic profile of the Ecuadorian Mestizo population (Ecuador—South America) by using the Power Plex1 16 System Kit. *Forensic Science International* 135:64–66.
 18. Hernández-Gutiérrez, S; Hernández-Franco, P; Martínez-Tripp, S; Ramos-Kuria, M; Rangel-Villalobos. H. (2005). STR data for 15 *loci* in a population sample from the central region of Mexico. *Forensic Science International* 151:97–100.
 19. Jaime, J; Fondevila, M; Salas, A; Lareu, MV; Carracedo, A. (2004). Nine autosomal STRs genotype profiles in a sample from Córdoba (Argentina). *Forensic Science International* 139:81–83.
 20. Leite, FPN; Menegassi, FJ; Schwengber, SP; Raimann, PE; Albuquerque, TK. (2003). STR data for 09 autosomal STR markers from Rio Grande do Sul (southern Brazil). *Forensic Science International* 132:223–224.
 21. Luna-Vazquez, A; Vilchis-Dorantes, G; Paez-Riberos, LA; Muñoz-Vallec, F; González-Martin, A; Rangel-Villalobos H. (2003). Population data of nine STRs of Mexican-Mestizos from Mexico City. *Forensic Science International* 136:96–98.
 22. Marino, M; Sala, A; Corach, D. (2006). Genetic analysis of the populations from Northern and Mesopotamian provinces of Argentina by means of 15 autosomal STRs. *Forensic Science International* 160:224–230.
 23. Marino, M; Sala, A; Corach, D. (2006). Genetic attributes of 15 autosomal STRs in the population of two patagonian provinces of Argentina. *Forensic Science International* 160:84–88.
 24. Martinez, B; Caraballo, L; Baron, F; Gusmão, L; Amorim, A; Carracedo, A. (2006). Analysis of STR *loci* in Cartagena, a Caribbean city of Colombia. *Forensic Science International* 160:221–223.

-
25. Martínez, B; Caraballo, L; Gusmão, L; Amorim, A; Carracedo, A. (2005). Autosomic STR population data in two Caribbean samples from Colombia. *Forensic Science International* 152:79–81.
 26. Martínez, G; Vázquez, E; Schaller, C.; Quevedo, N. (2003). Genetic data on 11 STRs (CSF1PO, TPOX, TH01, F13A01, FESFPS, vWA, D16S539, D7S820, D13S317, F13B, LPL) in an Argentine northeast population. *Forensic Science International* 133:254–255.
 27. Melendez, E; Martinez-Espin, E; Karlson, IS; Lorente, JA; Budowle, B. (2004). Population data on 15 STR *loci* (Powerplex 16 Kit) in a Costa Rica (Central America) sample population. *Journal of Forensic Science* 49(1):170-173.
 28. Miozzo, MC; Pacharonia, CM; Mutala, SA; Maxzuda, MK; Casanova, AE; Modesti, NM. (2003). STRs data for the *loci* D3S1385, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820 from Córdoba (Argentina). *Forensic Science International* 131:214–217.
 29. <http://alfred.med.yale.edu/> (2008). Kenneth K. Kidd, Yale University. All rights reserved.
 30. Paredes, M; Galindo, A; Bernala, M; Ávila, S; Andrade, D; Vergara, C; Rincón, M; Romero, RE; Navarrete, M; Cárdenas, M; Ortega, J; Suarez, D; Cifuentes, A; Salas, A; Carracedo, A. (2003). Analysis of the CODIS autosomal STR *loci* in four main Colombian regions. *Forensic Science International* 137:67–73.
 31. Pérez, L; Hau, J; Izarra, F; Ochoa, O; Zubiato, U; García, O. (2003). Allele frequencies for the 13 CODIS STR *loci* in Peru. *Forensic Science International* 132:164–165.
 32. Pimentel, BJ; De Azevedo, DA; Silva, LA. (2004). Population genetics of eleven STR *loci* in the State of Sergipe, Northeastern Brazil. *Journal of Forensic Science*. 49(2):1-2.
 33. Pineda-Bernala, L; Borjas-Fajardo, L; Zabala, W; Fernández, E; Delgado, W; Salas, A; Sánchez-Diz, P; Carracedo, A. (2002). Data for nine autosomal STRs markers (CSF1PO, D13S31, D16S539, D7S820, F13A01, FESFPS, TH01, vWA, TPOX) from Venezuela. *Forensic Science International* 125:277–278.
 34. Rey, M; Gutiérrez, A; Schroeder, B; Usaquén, W; Carracedo, A; Bustos, I; Giraldo A. (2003). Allele frequencies for 13 STR's from two Colombian populations: Bogotá and Boyacá. *Forensic Science International* 136:83–85.

-
35. Rodrigues, EMR; Palha, TJBF; Santos, SEB. (2006). Allele frequencies data and statistic parameters for 13 STR *loci* in a population of the Brazilian Amazon Region. *Forensic Science International*. 168(2-3):244-247.
 36. Sanóu, I; Núñez, G; Rodríguez, A; Silva, S; García, O; Uriarte, J; Espinoza, M. (2005). STR data for 11 autosomal STR markers from Costa Rica, Central América. *Forensic Science International* 153:272–273.
 37. 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 Science International* 139:173–176.
 38. Silva, EB; Dellalibera, E; Souza, M; Silva, RS; Maurício-da-Silva, L. (2002). Population genetics of eight STR *loci* - CSF1PO, TPOX, TH01, D16S359, D7S820, D13S317, F13B and LPL in a Brazilian population from the State of Piauí, Northeast Brazil. *Forensic Science International* 126:90-92.
 39. Silva, FF; Dellalibera, E; Nigam, P; Mauricio-da-Silva, L; Santos Silva, R. (2003). Microsatellite markers in the population from Rio Grande do Norte, Northeastern Brazil. *Journal of Forensic Science* 48(5):1189-1190.
 40. 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 Science International* 130:187–188.
 41. Silva, R; Moura-Neto, RS. (2004). Genetic diversity and admixture data on 11 STRs (F13B, TPOX, CSF1PO, F13A01, D7S820, LPL, TH01, vWA, D13S317, FESFPS, and D16S539) in a sample of Rio De Janeiro European-descendants population, Brazil. *Forensic Science International* 142:51–53.
 42. Toscanini, U; Berardi, G; Raimondi, E. (2003). STR data for PowerPlex116 System from Neuquen population, SW Argentina. *Forensic Science International* 134:219–221.
 43. Zúñiga, J; Ilzarbe, M; Acunha-Alonzo, V; Rosetti, F; Herbert, Z; Romero, V; Almeciga, I; Clavijo, O; Stern, JNH; Granados, J; Fridkis-Hareli, M; Morrison, P; Azocar, J; Yunis, EJ. (2006). Allele frequencies for 15 autosomal STR *loci* and admixture estimates in Puerto Rican Americans. *Forensic Science International* 164:266-270.
 44. Presente tese.

Tabela 2. Diversidade gênica (H_{sk}) de Nei (1987) para as populações do Distrito Federal, Goiás, Mato Grosso do Sul e da região centro-oeste, de 14 marcadores STRs autossômicos.

População Marcador	Distrito Federal	Goiás	Mato Grosso do Sul	Média Centro-Oeste
CSF1PO	0,74	0,74	0,74	0,74
D3S1358	0,78	0,78	0,78	0,78
D5S818	0,73	0,74	0,72	0,73
D7S820	0,80	0,81	0,81	0,81
D8S1179	0,81	0,81	0,82	0,81
D13S317	0,79	0,79	0,77	0,78
D16S539	0,79	0,77	0,79	0,78
D18S51	0,88	0,88	0,88	0,88
D21S11	0,85	0,85	0,86	0,85
FGA	0,88	0,87	0,87	0,87
PENTA-E	0,91	0,91	0,90	0,91
TH01	0,79	0,80	0,80	0,80
TPOX	0,70	0,66	0,67	0,68
vWA	0,80	0,81	0,81	0,81
Média	0,80	0,80	0,80	0,80

ANEXO 2. PROTOCOLOS DOS PROCEDIMENTOS LABORATORIAIS

1. Processamento do material biológico

As amostras biológicas foram coletadas em tubos de coleta a vácuo estéreis devidamente identificados com os números das amostras e mantidos em gelo durante o transporte até o Laboratório de Genética da Universidade de Brasília. No caso das amostras de Goiás, as amostras foram encaminhadas para o Laboratório de Genética da Universidade Federal de Goiás. O processamento do sangue coletado consistiu na separação do fluido em 3 frações distintas (plasma, creme leucocitário e hemácias) através de sucessivas centrifugações e lavagens com solução salina.

Cada amostra foi inicialmente centrifugada a 3000 rpm por cinco minutos para garantir a separação do sangue em seus componentes. Primeiramente, o plasma foi retirado e estocado em um microtubo para posterior análise de proteínas séricas e como possível fonte de DNA. Em seguida, o creme leucocitário foi separado e estocado em um microtubo para ser utilizado para extrações de DNA. A fração contendo hemácias, foi lavado três vezes com solução salina 0,9%, e após a última centrifugação foi estocado em tampão glicerol 40% em citrato tripotássio (0,1M, KH₂PO₄ 0,0345M, K₂HPO₄ 0,0344M) para análise de proteínas eritrocitárias e, quando necessário, como fonte de DNA. Todas as frações foram devidamente identificadas e estocadas a -20° C.

2. Extração de DNA

2.1. Método “Salting Out”

Um volume inicial de sangue de 200 µL em 200µL de tampão (*Buffer A*) resfriado foi centrifugado a 5000rpm durante 15 minutos. Em seguida, o sobrenadante foi descartado e o *pellet* foi ressuscitado em 150 µL de tampão (*Buffer B*). À solução, foi adicionado 20 µL de SDS 10% e 2,2 µL de proteinase K (10mg/µL), que então foi incubada a 55°C por 1h30.

Após a incubação, foi adicionada uma quantidade de NaCl saturado (6 M) que variou de 65 a 80 µL, para que então o tubo fosse centrifugado a 9000 rpm durante 15 minutos. O sobrenadante, resultado da centrifugação, foi transferido para outro tubo de 1,5 µL, ao qual foi adicionado 467 µL de isopropanol. A solução foi então agitada até a formação de “medusa” e centrifugada a 11000 rpm durante 4 minutos.

O etanol foi então descartado e a solução foi deixada para secagem à temperatura ambiente. O material depositado no fundo do tubo foi ressuscitado em 50 µL de TE (1X - pH 8,0) e então incubado a 73° C por uma hora.

2.2 “GFX™ Genomic Blood DNA Purification Kit” –Pharmacia

Em um tubo de microcentrífuga, foram adicionados 300µL da amostra de sangue e três volumes da solução de lise pré-aquecido a 70° C, homogeneizando-se, por inversão do tubo. O sistema foi incubado à temperatura ambiente durante cinco minutos e posteriormente submetido à centrifugação em alta rotação (12.000 rpm) por vinte segundos. O sobrenadante foi removido, por aspiração, tomando-se o cuidado para não atingir o aglomerado de células brancas. Um líquido residual de 20-50 µL foi aproveitado para ressuspensão das células.

Após a ressuspensão completa das células brancas, foram adicionados 500 µL de solução de extração, homogeneizando-se vigorosamente. Em seguida, foi realizada uma incubação em temperatura ambiente durante cinco minutos.

O sistema de extração foi transferido para uma coluna GFX e submetido a uma microcentrifugação, durante 60 segundos, descartado-se o volume do tubo coletor.

Novo volume de 500 µL de tampão de extração foi aplicado à coluna GFX, sendo que, após centrifugação, o volume do tubo coletor foi mais uma vez descartado. Em seguida, foram adicionados à coluna GFX 500 µL de solução de lavagem para nova centrifugação em alta rotação (12000 rpm), durante 3 minutos. O fluxo foi descartado e a coluna transferida para um novo tubo coletor.

Após a adição de 100 – 200 µL de água milli-Q (pré-aquecida) à coluna GFX, os sistemas foram incubados em temperatura ambiente durante 60 segundos e centrifugados a 8000 rpm, por igual tempo, para a recuperação do DNA purificado.

3. Reação em Cadeia da Polimerase

Reagentes e soluções:

- Amostras de DNA genômico (50 ng/µL)
- Tampão de reação (Tris/HCl 0,2 M pH 8,5; KCl 0,5 M; MgCl₂ 0,02 M)
- Solução de dNTP 20 mM;
- Taq DNA Polimerase;
- Solução de trabalho dos iniciadores (10 µM):

Os iniciadores foram diluídos com água mili-Q esterilizada para uma concentração de 50 µM utilizando-se para base de cálculo a quantidade de nano moles obtidos na síntese destes (dados fornecidos pelo fabricante). Uma alíquota de 100 µL de

iniciadores diluídos foi então depositada em tubo, para serem estocados em freezer (-20°C) até o momento do uso.

Para uso durante os procedimentos de PCR, prepararam-se soluções de trabalho, com os dois iniciadores necessários para a amplificação de cada região genômica selecionada, na concentração de 10 µM de cada iniciador.

** Observação: Todas as reações foram realizadas tomando-se o cuidado com o controle de qualidade interno, utilizando controles positivo e negativo para cada sistema.

3.2 Procedimentos e Realização da PCR

As amostras de DNA genômico a serem analisadas foram descongeladas e aliqüotadas em microtubos de polipropileno de 0,5 µL devidamente identificados com o número da amostra, marcador analisado e número de registro da PCR (uso interno do laboratório). Uma mistura de reagentes da PCR suficientes para o número de indivíduos em análise (conforme item 2.1 deste anexo) foi adicionada ao microtubo que continha o DNA do indivíduo a ser analisado para cada sistema (Tabela 1, anexo 2). Óleo mineral foi adicionado aos microtubos no final do procedimento, para que então fossem levados à máquina de PCR com o programa correspondente a cada sistema.

4. Procedimentos para reação de restrição

Os sistemas GC, Rb2300, DRD2-A, DRD2-B, DRD2-D, LPL, FYNull e OCA2 foram analisados por RFLP (*Restriction fragment length polymorphism*), técnica que consiste em amplificar as regiões de interesse, que são então submetidas à digestão enzimática, utilizando enzimas específicas para cada marcador (Tabela 2, Anexo 2). O preparo do sistema, no geral, foi da seguinte forma:

- a) Preparava-se um mix contendo tampão da enzima diluído a 1X e 1,0 U da enzima;
- b) Adicionava-se 5,0 µL do produto de PCR;
- c) Após a vedação da placa, incubava-se em banho-maria (temperatura e tempo específicos de cada enzima);
- d) Após a incubação os produtos de digestão eram visualizados em gel de poli-acrilamida a 6 ou 10%, conforme o marcador.

Tabela 1. Condições da mistura para a reação de PCR para os marcadores genéticos analisados.

Reagentes Marcador	Água mili-Q (µl)	Tampão¹(µl)	dNTP²(µl)	Iniciador³(µl)	MgCl⁴	Taq polimerase (unidade)	DNA (50ng)
APO	6,60	1,25	1,25	2,00	0,25	1,0	1,0
AT3	7,15	1,25	0,63	1,50	0,50	2,0	1,0
D1*	5,70	1,25	1,25	2,00	0,25	0,6	0,5
DRD2-A	6,70	1,25	1,25	1,60	-	0,5	1,5
ECA*	6,90	1,25	1,25	1,00	0,50	0,6	1,0
FXIIIB	7,40	1,25	1,25	1,25	0,50	1,0	1,0
FyNull	7,80	1,10	1,40	0,60	-	0,6	1,0
GC	6,90	1,25	1,25	1,50	-	0,6	1,0
LPL	8,10	1,25	0,63	1,50	-	0,5	1,0
OCA2	7,70	1,20	0,63	1,50	-	0,5	1,5
PV92*	7,40	2,00	1,25	0,50	-	1,0	1,0
Rb2300*	5,25	1,75	1,25	1,25	-	0,5	1,5
Sb19.3	6,60	1,40	1,25	1,50	-	0,5	1,5
TPA25	7,00	1,50	1,25	1,25	-	1,0	1,0

* Na mistura de reação desses marcadores foi acrescentado 0,315µl de DMSO. 1 – solução composta por Tris/HCl 75 mM pH 9,0; 50 mM KCl; 20 mM (NH₄)₂SO₄; 2 mM de MgCl. 2 – Solução com 10 mM de cada dNTP: dATP, dCTP, dGTP, dTTP. 3 – 2,5 µM específico para cada *locus*. Para o marcador FyNull a concentração foi de 1,25 µM. 4 – Solução de cloreto de magnésio a 2,0 mM.

Tabela 2. Concentração dos reagentes utilizados nas reações de restrição dos produtos de PCR dos marcadores do tipo SNPs, temperatura ótima da enzima e tempo de restrição.

Marcador	DRD2-A (TaqI)	LPL (PvuII)	OCA2 (HaeIII)	Rb2300 (BamHI)	FyNull (StyI)	GC – 1S (HaeIII)	GC – 2 e 1F (StyI)
Reagentes							
Água Mili-Q	5,4µl	6,0µl	5,5µl	5,7µl	5,6µl	5,6µl	5,5µl
Tampão¹	1,2µl	1,0µl	1,2µl	1,0µl	1,0µl	1,0µl	1,2µl
BSA²	0,1µl	-	-	-	0,1µl	0,1µl	-
Enzima³	0,3µl	0,04µl	0,3µl	0,3µl	0,3µl	0,3µl	0,3µl
Temperatura	65°C	37°C	37°C	37°C	37°C	37°C	37°C
Tempo	4 h	4 h	4 h	18 h	4 h	4 h	4 h

1 - Tampão da enzima composto por: 50 mM Tris-HCl (pH 8.0), 10 mM MgCl₂, 50 mM NaCl para as enzimas TaqI e HaeIII; 50 mM Tris-HCl (pH 7.4), 6 mM MgCl₂, 50 mM NaCl e 50 mM KCl para a enzima Pvu II; 100 mM Tris-HCl (pH 8.5), 8 mM MgCl₂, 80 mM NaCl e 0,8 mM DTT (dithiothreitol) para a enzima StyI; e 33 mM Tris-acetato, 66 mM K-acetato, 10 mM Mg-acetato e 0,5 mM DTT para a enzima Bam HI. 2 - Soroalbumina Bovina (500µg/ml). 3 - Concentração = 10U/µl.

5. Procedimentos para Eletroforese - Análise do produto de PCR e de Restrição

5.1 Gel de Poliacrilamida – Eletroforese vertical

5.1.1 Soluções para gel

Tampão TBE10X pH8,0 (Tris/HCl 0,89M, EDTA 0,08M, Ácido Bórico 0,89M);

Solução acrilamida 29% e bis-acrilamida 1%;

Glicerol;

TEMED (N,N,N',N', tetrametiletilenodiamina);

Persulfato de potássio ou amônia (APS) a 0,1%.

5.2 Preparo das amostras para aplicação no sistema de eletroforese

As amostras (produto da PCR e da restrição) foram preparadas para eletroforese adicionando-se tampão de corrida, acrescido do corante bromofenol, na proporção de 2 µL de corante para 5 µL do produto de PCR ou restrição.

5.3 Procedimento

Após a assepsia, as placas de vidro foram montadas com espaçadores e prendidas com grampos de aço. A solução de poliacrilamida foi preparada nas concentrações de 6% e 10% de acordo com protocolo descrito na Tabela 3 deste anexo.

Em seguida os catalisadores TEMED e Persulfato de Amônia foram adicionados na proporção descrita na tabela 1 para que então a solução fosse entornada na placa preparada.

Tempo médio de polimerização: 30 minutos.

À cuba de eletroforese foi adicionada solução de TBE prepara pela diluição de solução estoque de TBE concentrada, que, por sua vez, foi preparada segundo este protocolo: 108,0 g de TRIS; 55,0 g de Ac. Bórico; 9,3 g de EDTA e 1L de água destilada.

Após a montagem das placas na cuba, a voltagem da eletroforese foi programada para ser 50 volts durante os primeiros 20 minutos, para que então fosse aumentada para 150 a 200V, de acordo com o sistema. A corrida variava de 3 a 4 horas conforme o marcador. Para os géis de poliacrilamida a 6% foi aplicada voltagem de 200 V por 3 horas e, para os géis a 10%, 150V por 4 horas.

Tabela 3. Preparo da solução para géis de poliacrilamida a 6% e a 10%

Reagentes	6%*	10%**
Solução de acrilamida	4 ml	6 ml
Glicerol	1,4 ml	1,4 ml
Água	12,4 ml	9,6 ml
TBE 10X	2 ml	2 ml
TEMED	15 µL	15 µL
APS	300 µL	300 µL
Volume Final	20 ml	20 ml

* marcadores: PV-92, TPA-25, APO, ACE, FXIII B, Rb2300, LPL, OCA2, Sb19.3 e D1.

** marcadores: DRD2-A, DRD2-B, DRD2-D, AT3, FYNull

5.1.4 Coloração

a) Solução Fixadora

750 ml de H₂O destilada.

144 ml de Álcool Etilíco.

6,0 ml de Ácido Acético.

b) Solução Reveladora.

22,5 g de NaOH.

Completar para 1L de H₂O destilada.

c) Solução de prata 10%

1 g nitrato de prata

10 mL de água destilada

Para a coloração utilizou primeiramente solução fixadora por 5 minutos. Em seguida, foi acrescentado 1mL de nitrato de prata e manteve-se em agitação por 5 minutos. O gel foi lavado com água destilada e mergulhado em solução reveladora previamente aquecida com 600µl de formaldeído até que as bandas se tornassem visíveis.

5.1.5 Leitura das placas de gel

As determinações dos perfis alélicos e genotípicos foram realizadas por comparações visuais entre os alelos padrões e o alelo presente em cada amostra ou então com o auxílio de uma escada alélica. As leituras dos géis foram realizadas com auxílio de um transiluminador – Macro VUE UVIS-20 - com luz branca (visível). Uma segunda leitura dos resultados foi confirmada posteriormente utilizando-se a escolha de amostras ao acaso.

5.2 Análise do sistema PowerPlex®16

5.2.1 Amplificação dos fragmentos

Foram analisados 16 marcadores STRs (Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Amelogenina, Penta D, CSF1PO, D16S539, D7S820, D13S317, D5S818) separados em 3 grupos de fluorescência (fluoresceína, carboxy-tetramethylrhodamine -TMR, 6-carboxy-4',5'-dichloro-2',7'-dimethoxy-fluorescein - JOE). Primeiramente foi realizada uma reação de PCR utilizando os seguintes reagentes:

Água livre de nuclease (ultra-pura): suficiente para completar 25,0 µl

Tampão de reação Gold STR X: 2,5µl

Conjunto de primers (PowerPlex® 16) concentrados a 10X: 2,5µl

Enzima Taq DNA polimerase (5U): 0,8µl

DNA: 1 a 2ng

Todas as reações apresentavam controle positivo (amostra fornecida pelo kit) e negativo (água ultra-pura). Todos os 16 *loci* foram amplificados simultaneamente em um único tubo e analisados em uma única injeção de sequenciamento.

Foi utilizado o termociclador System 9700 Thermal Cycler¹ e o programa utilizado constituía em:

95°C por 11 minutos

96°C por 1 minuto seguido de 10 ciclos de:

rampa 100% a 94°C por 30 segundos

rampa 29% a 60°C por 30 segundos

rampa 23% a 70°C por 45 segundos

22 ciclos de:

rampa 100% a 90°C por 30 segundos

rampa 29% a 60°C por 30 segundos

rampa 23% a 70°C por 45 segundos

60°C por 30 minutos

4°C

5.2.2 Detecção dos fragmentos

A detecção dos fragmentos amplificados foi realizada através do sequenciador 3100-Avant Genetic Analyzer. As amostras foram preparadas utilizando-se uma mistura de 0,5µl ILS 600 (Internal Lane Standard 600) e 9,5µl Hi-DiTM formamida. O ILS apresenta fragmentos de 60, 80, 100, 120, 140, 160, 180, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550 e 600 bases com carboxy-X-rhodamine (CXR) o que permite sua detecção separada dos fragmentos amplificados em PCR. A solução era misturada de 10 a 15 segundos com o auxílio de vortex. O volume total da solução era calculado de acordo com a quantidade de reações de PCR mais 10%. Eram adicionado nos poços das placas de sequenciamento 1 µL de produto de PCR e 10 µL da mistura formamida/internal lane standard. Todos os experimentos apresentavam um poço contendo escada alélica dos marcadores em estudo.

A aplicação das amostras seguiu o indicado no manual do equipamento. Foi utilizado o modo “HIDFragmentAnalysis36_POP4”, o tempo de injeção das amostras foi de 5 segundos e a voltagem de 3kV. O tempo de corrida foi de 2000 segundos.

5.2.3 Análise de dados

O software utilizado para leitura foi o GeneMapper® ID versão 3.2 fornecido pelo fabricante. Padrões gerados pelas escadas alélicas foram utilizados como padrões para as leituras no programa.

ANEXO 3. TERMOS DE CONSENTIMENTO LIVRE E ESCLARECIDOS

Termo de Consentimento utilizando nas coletas de amostras biológicas na população do estado de Goiás.

TERMO DE CONSENTIMENTO

Eu, _____, natural de _____, nascido em ____/____/____, portador do R.G. n.º _____ SSP _____, residente em _____

filho de _____ (natural de _____)
e de _____ (natural de _____)
Etnia: _____

Telefone: _____

Declaro para os devidos fins:

- I. que fui convidado (a) a participar de uma pesquisa científica que tem por objetivo estabelecer um banco de dados de DNA da população do Estado de Goiás, onde serei submetido a uma coleta sanguínea venosa com um volume de 5,0 ml (cinco mililitros);
- II. que me foi explicado que o material coletado será utilizado exclusivamente para análise de frequência de marcadores genéticos da população;
- III. que me foi explicado a finalidade do estudo e os benefícios que dele a ciência possa vir a ter;
- IV. que me foi explicado e garantido que posso recusar a participar do presente estudo;
- V. que me foi garantido sigilo dos dados pessoais acima discriminados;
- VI. que os resultados obtidos nessa pesquisa poderão ser publicados em periódicos e eventos científicos ou quaisquer outros veículos de comunicação;
- VII. que não fui submetido à transfusão sanguínea e/ou transplante de medula óssea;
- VIII. que, por livre e espontânea vontade, concordo em participar deste estudo e após ler e sem ter quaisquer dúvidas, assino.

_____, ____/____/20____

Doador

Testemunha com R.G.

Pesquisador responsável

Pesquisador responsável

**Termo de consentimento livre e esclarecido e do questionário demográfico
utilizado na coleta do material biológico dos indivíduos do do Distrito Federal**

UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE GENÉTICA E MORFOLOGIA
LABORATÓRIO DE GENÉTICA

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO, PÓS- INFORMAÇÃO

O abaixo assinado, _____ declara ter sido plenamente esclarecido, lido e ouvido o presente termo de responsabilidade que lhe informa estar ciente do seguinte:

- a) Que pelo presente documento concorda em participar da pesquisa “Brasília: reflexo da constituição genética brasileira?” visando a caracterização genética da população brasiliense e a comparação desta com populações urbanas brasileiras.
- b) Que esta participação implicará na retirada de aproximadamente 5ml de sangue de uma das veias do antebraço sendo que este material será utilizado com finalidade de tipagem sanguínea, análise hemoglobina, extração de DNA, análise de marcadores moleculares e marcadores protéicos e isoenzimáticos.
- c) Que esse procedimento é método usual na pesquisa genética, não implicando risco para a saúde, podendo, porém, provocar um pouco de dor.
- d) Que a recusa em participar do projeto poderá ser feita a qualquer tempo mesmo após a assinatura do presente termo de consentimento e ficará livre para abandonar a pesquisa a qualquer momento bastando a comunicação aos responsáveis do projeto.

Brasília, ____ de _____ de 2001

Assinatura : _____

Documento de identidade: _____

Responsável pela pesquisa: _____

UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE GENÉTICA E MORFOLOGIA
LABORATÓRIO DE GENÉTICA
PROJETO BRASÍLIA

Nome:		No.
Endereço:		Telefone:
E-mail:		Estado civil:
Data de nascimento:		Sexo: <input type="checkbox"/> F <input type="checkbox"/> M
Classificação fenotípica : () Br () Ng () Am (Ind) () Am (Asia) () Misto		
Se misto		
() Traços de Caucasoíde		Foto 3 X 4
() Traços de Negróide		
() Traços de Indígena		
() Traços de Asiático		
() Outros (especificar)		
Ascendência materna:		
Ascendência paterna:		
Peso :		Pressão arterial 1 :
Altura :		Pressão arterial 2 :
Grau de Instrução:		Profissão:
Local de residência dos pais na data do nascimento:		
<u>Filiação</u>		
Pai :		Mãe :
Local de nascimento :		LN :
Avó paterna :		Avó materna :
LN :		LN :
Avô paterno :		Avô materno :
LN:		LN:
Tempo em Brasília (dos pais) :		Motivo da mudança :
Irmãos ? () Masc () Fem.		
Gêmeos ? () Não		
() Gêmeo Uni. () Masc. () Fem.		
() Gêmeo Bi. () Masc. () Fem.		
() Irmão Gêmeo Uni. () Masc. () Fem.		
() Irmão Gêmeo Bi. () Masc. () Fem. Outro:		
Alguma doença genética na família:		Qual (is):
Afetados:		
Nome do cônjuge:		Data de nascimento:
Consangüinidade entre os cônjuges		
LN:		Obs : Se nascido em Brasília, fazer ficha à parte
No. de filhos :		Gêmeos?
No. de natimortos :		No. de abortos espontâneos:
No. de filhos mortos:		Causa:
Nome:		Causa:
LN:		Sexo: ()masc. ()fem.
Nome:		Idade:
LN:		Sexo: ()masc. ()fem.
Nome:		Idade:
LN:		Sexo: ()masc. ()fem.
Nome:		Idade:
LN:		Sexo: ()masc. ()fem.
LN:		Idade:

OBS: