

Tese de Doutorado
Programa de Pós Graduação em Genética
Universidade Federal de Minas Gerais

Padrões geográficos de ancestralidade genômica em Minas Gerais: o caso da doença
falciforme

Maria Clara Fernandes da Silva
Orientador: Prof. Dr. Eduardo Martin Tarazona Santos
Co-Orientadora: Profa Dra. Marina Lobato Martins

Belo Horizonte
Setembro / 2010

Maria Clara Fernandes da Silva

Padrões geográficos de ancestralidade genômica em Minas Gerais: o caso da doença
falciforme

Tese apresentada ao curso de Pós Graduação em Genética do
Departamento de Biologia Geral do Instituto de Ciências
Biológicas da Universidade Federal de Minas Gerais, como
requisito parcial para a obtenção do título de Doutor em
Genética.

Área de concentração: Genética de Populações

Orientador: Prof. Dr. Eduardo Martin Tarazona Santos

Co-Orientadora: Profa. Dra. Marina Lobato Martins

Belo Horizonte

Departamento de Biologia Geral

Instituto de Ciências Biológicas

Universidade Federal de Minas Gerais

2010

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Eduardo Martin Tarazona Santos, por possibilitar meu contato com um universo de novos conhecimentos; por idealizar e possibilitar a realização desse projeto; pela compreensão em momentos difíceis e pelo exemplo de dedicação e competência.

A minha co-orientadora, Profa. Dra. Marina Lobato Martins, pela parceria, amizade e colaboração; pelo grande exemplo de profissionalismo e competência; e pelo incentivo em minha trajetória ao longo da execução desse trabalho.

As grandes amigas e colaboradoras Luciana Werneck Zucheratto e Zilma Martins Vieira, pela ajuda imprescindível para a viabilidade deste trabalho.

A todos os amigos e colegas do Laboratório de Diversidade Genética Humana (ICB – UFMG), em especial Giordano Bruno e Wagner Magalhães pela colaboração e companheirismo.

Às estagiárias Flávia Campos Lucena e Julia Barbosa pelo auxílio. Espero que vocês tenham aprendido tanto comigo quanto eu aprendi com vocês!

Aos colaboradores do Laboratório de Biodiversidade e Evolução Molecular e Prof. Fabrício Santos, Núcleo de Análise de Genomas e Expressão Gênica (NAGE) e Profa. Santuza Teixeira, e Laboratório de Genética Bioquímica e Profa. Gloria Franco pelo acesso aos seqüenciadores de DNA.

Aos colaboradores da Universidade Católica de Brasília, em especial ao Prof. Rinaldo Wellerson Pereira e ao Tulio Lins pela colaboração com os protocolos de minisequenciamento policromático, permissão do uso do painel de 10 marcadores informativos de ancestralidade ameríndia e acesso ao analisador de DNA.

Aos colaboradores do Laboratório de Genética Bioquímica, em especial ao Prof. Dr. Sergio D. J. Pena, Fernanda Kehdi e Vanessa Gonçalves por permitirem e possibilitarem o uso do painel de 40 marcadores empregados no presente estudo.

Aos colegas de diversos laboratórios, em especial ao Juliano Leal e Daniela Lacerda pela colaboração e disponibilidade em ajudar.

Aos amigos e colegas dos Serviços de Pesquisa e de Imuno-Hematologia da Fundação Hemominas, em especial a Luciana, Maria Estela, Marta e Elisângela pelo apoio e compreensão.

Aos colegas das diferentes unidade regionais da Fundação Hemominas: José Carlos M. Chaves do Núcleo Regional de Ituiutaba, Heloísa das Graças de Souza do Hemocentro de Uberaba, Tereza Cristina de Oliveira e Yara Maria Costa Pinheiro do Núcleo Regional de Patos de Minas, Vânia M. Oliveira Nascimento do Núcleo Regional de Uberlândia, Rosimeire das Dores Esteves do Núcleo Regional de Divinópolis, Telma Regina Guedes Machado do Hemocentro de Montes Claros, Cristian de Freitas Guimarães do Núcleo Regional de Ponte Nova, Izabel Christina França da Luz do Núcleo Regional de Diamantina, Simone Avelino T. Rodrigues do Núcleo Regional de Manhuaçu, Antônio Geraldo de Carvalho do Núcleo Regional de São João Del Rei, Maria Tereza Rodrigues Amaro do Hemocentro de Juiz de Fora, Lígia Martins Costa do Hemocentro de Governador Valadares e Alcione Maria de Pinho do Hemocentro de Belo Horizonte, pela contribuição valiosa no processo de amostragem. Sem esta colaboração este estudo não seria possível.

As instituições que contribuíram no financiamento do presente trabalho: National Institutes of Health (NIH), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação para o Apoio à Pesquisa (FAPEMIG) e Fundação Hemominas.

Aos demais amigos e colegas não mencionados, mas não esquecidos, que contribuíram das mais diversas formas para que fosse possível a conclusão desse estudo.

Aos meus familiares, em especial meu marido Frederico, minha filha Clarissa e minha mãe Iracy, pelo amor e principalmente por me fazerem ver o grande valor das coisas mais simples.

A Deus, por permitir que eu pudesse adquirir conhecimentos que me possibilitaram admirar ainda mais sua grande obra: a vida humana.

SUMÁRIO

1 INTRODUÇÃO	3
1.1 Constituição da população brasileira	3
1.2 Estimativas de ancestralidade biogeográfica e miscigenação.....	7
1.2.1 Miscigenação no nível populacional.....	9
1.2.2 Miscigenação nos níveis individual e genômico.....	11
1.3 Implicações da miscigenação no desenho de estudos epidemiológicos.....	13
1.3.1 Impacto da estratificação gerada pela miscigenação em estudos de epidemiologia genética do tipo caso-controle.....	13
1.3.2 Controle dos efeitos da estratificação em estudos de epidemiologia genética do tipo caso-controle.....	16
1.4 A doença falciforme	19
1.4.1 Aspectos moleculares e manifestações clínicas.....	19
1.4.2 A doença Falciforme no Brasil e em Minas Gerais.....	22
2 OBJETIVOS.....	25
2.1 Objetivo Geral.....	25
2.2 Objetivos Específicos.....	25
2.2.1 Criar uma coleção de amostras de DNA de indivíduos sadios, doadores de sangue, e pacientes com doença falciforme provenientes de diversas regiões de Minas Gerais.....	25
2.2.2 Padronizar a genotipagem de painéis de marcadores informativos de ancestralidade (MIAs) europeia, africana e ameríndia pelo método de minisequenciamento.....	25
2.2.3 Genotipar os doadores de sangue e pacientes com doença falciforme para 54 marcadores moleculares informativos de ancestralidade (14 SNPs e 40 INDELS).....	25
2.2.4 Estimar a miscigenação ameríndia, europeia e africana nos níveis populacional e individual para os doadores de sangue e pacientes com doença falciforme.....	25
2.2.5 Determinar, a partir dos marcadores genotipados, a estrutura genética das populações de pacientes com doença falciforme e de doadores de sangue em Minas Gerais em função dos níveis de miscigenação.....	25
2.2.6 Testar a hipótese nula que os padrões de ancestralidade genômica em Minas Gerais determinam uma estrutura populacional que se adapta ao modelo de isolamento por distância.....	25
2.2.7 Verificar a ocorrência de padrões geográficos de ancestralidade genômica para doadores de sangue e pacientes com doença falciforme.....	25
3 MATERIAL E MÉTODOS.....	26
3.1 Populações e amostras.....	26
3.1.1 Populações parentais.....	26
3.1.2 Populações miscigenadas.....	26
3.2 Genotipagem.....	28
3.2.1 Construção do painel de 14 SNPs informativos de ancestralidade.....	29
3.2.2 Painéis de 40 marcadores de inserção-deleção (Bastos-Rodrigues <i>et al.</i> , 2006).....	34
3.2.3 Painel de 10 marcadores de inserção-deleção.....	34

3.3. Análises estatísticas.....	35
3.3.1 Equilíbrio de Hardy-Weinberg (EHW)	35
3.3.2 Estrutura Genética	36
3.3.3 Isolamento por Distância (IBD).....	37
3.3.4. Estimativas de miscigenação no nível populacional.....	38
3.3.5 Estimativas de miscigenação no nível individual	38
3.3.6. Identificação de padrões geográficos de miscigenação.....	39
4 RESULTADOS.....	40
4.1. Coleta das amostras.....	40
4.2. Seleção de MIAs para desenvolvimento de painéis de genotipagem.....	42
4.3. Genotipagem.....	44
4.3.1 Padronização das Reações de PCR.....	44
4.3.2 Padronização da genotipagem de SNPs pela técnica de minisequenciamento.....	45
4.4 Controle da qualidade dos painéis de SNPs.....	50
4.5 Genotipagem dos painéis de SNPs nas populações parentais.....	51
4.6 Análise Genética da população.....	51
4.7 Isolamento por distância.....	64
4.8 Precisão da estimativa de miscigenação utilizando o painel de 14 SNPs.....	67
4.8.1 Análises das populações parentais.....	67
4.8.2 Miscigenação populacional.....	67
4.8.3 Miscigenação Individual.....	71
4.9 Análises de miscigenação dos doadores de sangue e pacientes com doença falciforme.....	74
4.9.1 Miscigenação populacional.....	74
4.9.2 Miscigenação individual.....	83
5. DISCUSSÃO.....	89
6. CONCLUSÃO.....	108
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	111

LISTA DE FIGURAS

Figura 1. Representação da miscigenação e do efeito da estratificação de populações em estudos de epidemiologia genética	18
Figura 2. Incidência da Doença Falciforme em nascidos vivos, por diretoria Regional de saúde – Programa Estadual de Triagem Neonatal (1998-2001)	24
Figura 3: Distribuição geográfica das Unidades da Fundação Hemominas onde houve coleta de amostras	27
Figura 4: Minisequenciamento Monocromático e Policromático para um indivíduo heterozigoto (T/G)	31
Figura 5: Reações de PCR multiplex para os painéis AFR3, AFR10 e AMR4, utilizando-se o Quiagen Multiplex kit e o Tampão STR Buffer (Pomega)	45
Figura 6: Flexibilidade de genotipagem por minisequenciamento monocromático em um sequenciador <i>Megabace</i> (a), e por <i>Snapshot</i> em um sequenciador <i>Applied Biosystems 3130</i> (b), evidenciada pelos eletroferogramas da genotipagem do SNP rs1372115, do painel AFR3 em um indivíduo heterozigoto, nas duas plataformas	47
Figura 7: Eletroferogramas da genotipagem por minisequenciamento monocromático em reações multiplex dos Painéis AFR3 (a) e AMR4 (b), utilizando-se DNA de um pool de amostras (10 indivíduos), apresentando a maioria dos loci em heterozigose (2 picos para cada marcador)	50
Figura 8. Representação gráfica bidimensional dos valores de F_{st} par a par para doadores de sangue amostrados em diferentes localidades, usando a abordagem MDS (Multidimensional scaling). Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas	63
Figura 9. Representação gráfica bidimensional dos valores de F_{st} par a par para pacientes com doença falciforme amostrados em diferentes localidades, usando a abordagem MDS (Multidimensional scaling). Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas	64
Figura 10. Análise do isolamento por distância para doadores de sangue amostrados em diferentes regiões do Estado de Minas Gerais utilizando dados genéticos de 14 SNPs e 40 INDELS	66
Figura 11. Análise do isolamento por distância para pacientes com doença falciforme amostrados em diferentes regiões do Estado de Minas Gerais utilizando dados genéticos de 14 SNPs e 40 INDELS	66
Figura 12. Análise de miscigenação para os indivíduos representantes das populações parentais (AFR = africanos, EUR = europeus e AMR = ameríndios) utilizando o programa STRUCTURE, a partir dos dados de genotipagem do painel de 14 MIAs. Cada barra vertical representa um indivíduo e as cores representam a ancestralidade.....	67
Figura 13. Estimativas de miscigenação africana para os Hispânicos considerando diferentes subconjuntos de marcadores.....	70

Figura 14. Estimativas de miscigenação européia para os Hispânicos considerando diferentes subconjuntos de marcadores.....	70
Figura 15. Estimativas de miscigenação ameríndia para os Hispânicos considerando diferentes subconjuntos de marcadores.....	71
Figura 16. Variação da amplitude dos intervalos de probabilidade de 90% das estimativas de miscigenação individuais em função das estimativas pontuais de miscigenação fornecidas pelo programa STRUCTURE, com a utilização de diferentes conjuntos de marcadores	73
Figura 17. Proporções de contribuição das populações parentais africana, européia e ameríndia para os grupos de doadores de amostrados em diferentes localidades	77
Figura 18. Proporções de contribuição das populações parentais africana, européia e ameríndia para os grupos de pacientes com doença falciforme amostrados em diferentes localidades	77
Figura 19. Estimativa de miscigenação africana para doadores de sangue amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente	78
Figura 20. Estimativa de miscigenação européia para doadores de sangue amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente	78
Figura 21. Estimativa de miscigenação ameríndia para doadores de sangue amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente	79
Figura 22. Estimativa de miscigenação africana para pacientes com doença falciforme amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente	79
Figura 23. Estimativa de miscigenação européia para pacientes com doença falciforme amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente.	80
Figura 24. Estimativa de miscigenação européia para pacientes com doença falciforme amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente.....	80
Figura 25. Mapa da distribuição geográfica da miscigenação populacional estimada pelo método de identidade genética para os pacientes com doença falciforme de Minas Gerais	81
Figura 26. Mapa da distribuição geográfica da miscigenação populacional estimada pelo método de identidade genética para os doadores de sangue de Minas Gerais.....	81
Figura 27. Análise de miscigenação individual dos doadores de sangue e pacientes com doença falciforme amostrados em diferentes localidades, utilizando o programa STRUCTURE. Cada barra vertical representa um indivíduo.....	84
Figura 28. Proporção de doadores de sangue de diferentes regiões de Minas Gerais, com alta ancestralidade africana, alta ancestralidade européia e com proporções intermediárias de miscigenação.....	86

Figura 29. Proporção de pacientes com doença falciforme de diferentes regiões de Minas Gerais, com alta ancestralidade africana, alta ancestralidade europeia e com proporções intermediárias de miscigenação	86
Figura 30. Miscigenação populacional africana (AFR IG) e percentual de indivíduos com ancestralidade africana superior a 85% (>85% AFR) em doadores de sangue amostrados em diferentes localidades	87
Figura 31. Miscigenação populacional europeia (EUR IG) e percentual de indivíduos com ancestralidade europeia superior a 85% (>85% EUR) em doadores de sangue amostrados em diferentes localidades	87
Figura 32. Miscigenação populacional africana (AFR IG) e percentual de indivíduos com ancestralidade africana superior a 85% (>85% AFR) em pacientes com doença falciforme amostrados em diferentes localidades	88
Figura 33. Miscigenação populacional europeia (EUR IG) e percentual de indivíduos com ancestralidade europeia superior a 85% (>85% EUR) em pacientes com doença falciforme amostrados em diferentes localidades	88

LISTA DE TABELAS

Tabela 1. Amostragem de doadores de sangue e pacientes com doença falciforme de Minas Gerais. Os números em parêntese indicam os indivíduos considerados nas análises genéticas e de miscigenação.....	41
Tabela 2. Marcadores selecionados a construção dos painéis de SNPs informativos de ancestralidade	43
Tabela 3. Tamanhos esperados e observados dos picos produzidos pela genotipagem por minisequenciamento monocromático para o painel de 14 SNPs informativos de ancestralidade	48
Tabela 4. Seqüência dos iniciadores de PCR e Minisequenciamento para o painel de 14 SNPs. Os trechos em letras minúsculas representam as seqüências de plasmídio PUC 18 inseridas na extremidade 5' dos iniciadores com o propósito de alterar seus tamanhos....	49
Tabela 5. Frequências dos SNPs genotipados em diferentes populações	52
Tabela 6. Teste exato de Fisher para o equilíbrio de Hardy –Weinberg, para as populações parentais. Valores em negrito indicam loci fora do equilíbrio após correção de Bonferroni	53
Tabela 7. Teste exato de Fisher para o equilíbrio de Hardy -Weinberg para as amostras de doadores de sangue e pacientes com doença falciforme. Valores em negrito indicam loci fora do equilíbrio após correção de Bonferroni. Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas	55
Tabela 8. Matriz de significância para F_{ST} par a par para doadores de sangue genotipados para 14 SNPs e 40 INDELS. Valores negativos correspondem a $F_{ST} = 0$. A significância de cada estimativa foi obtida pela avaliação do intervalo de confiança de 95% e os valores significativos estão em negrito. Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.....	61
Tabela 9. Matriz de significância para F_{ST} par a par para pacientes com doença falciforme genotipados para 14 SNPs e 40 INDELS. Valores negativos correspondem a $F_{ST} = 0$. A significância de cada estimativa foi obtida pela avaliação do intervalo de confiança de 95% e os valores significativos estão em negrito. Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.....	62
Tabela 10. Análise de isolamento por distância para doadores de sangue e pacientes com doença falciforme amostrados em diferentes regiões do Estado de Minas Gerais.....	65
Tabela 11. Estimativas de miscigenação populacional obtidas por três métodos de análise para a amostra de Hispânicos utilizando diferentes conjuntos de marcadores.....	69

Tabela 12. Coeficientes de correlação entre estimativas de miscigenação para a amostra de hispânicos utilizando 108 SNPs, e as estimativas utilizando subconjuntos de 80, 50, 20 MIAs, bem como os conjunto de 13 SNPs e 40 INDELS genotipados no presente estudo.....	72
Tabela 13. Estimativa de miscigenação populacional para doadores de sangue e pacientes com doença falciforme amostrados em cada localidade, utilizando diferentes métodos de estimativa. AFR = Africanos, EUR = Europeus, AMR = Ameríndios. Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.....	75
Tabela 14. Correlação entre as estimativas de miscigenação africana, européia e ameríndia obtidos com o método IG para doadores de sangue e pacientes com doença falciforme e a latitude ou longitude da localidade onde os indivíduos foram amostrados.	82
Tabela 15 Correlação entre as estimativas de miscigenação obtidas para doadores e pacientes amostrados em diferentes regiões do Estado de Minas Gerais utilizando o programa STRUCTURE empregando diferentes conjuntos de marcadores.....	83
Tabela 16. Proporção de indivíduos amostrados em diferentes localidades, que apresentam ancestralidade africana ou européia superior a 85%	85

LISTA DE ABREVIATURAS

HBH	Belo Horizonte
DIA	Diamantina
DIV	Divinópolis
GOV	Governador Valadares
ITU	Ituiutaba
JFO	Juiz de Fora
MCU	Manhuaçu
MOC	Montes Claros
PMI	Patos de Minas
PNO	Ponte Nova
SJR	São João del Rei
URA	Uberaba
UDI	Uberlândia
AFR	Africano
EUR	Europeu
AMR	Ameríndio
SNP	Polimorfismo de uma única base
INDEL	Polimorfismo de inserção-deleção
IBD	Isolamento por distância
TCLE	Termo de consentimento livre e esclarecido
EWH	Equilíbrio de Hardy-Weinberg
MCMC	Cadeias de Markov Monte Carlo
EXO I	Exonuclease I
SAP	Fosfatase alcalina de camarão

RESUMO

A miscigenação é um evento que ocorre quando indivíduos pertencentes a populações parentais que estiveram isoladas por várias gerações, formam uma nova população híbrida. O interesse em se estimar a miscigenação é crítico em estudos de epidemiologia genética do tipo caso-controle envolvendo populações miscigenadas, já que a ocorrência de estratificação gerada pela miscigenação pode levar a resultados falso-positivos. Marcadores que apresentam frequências alélicas altamente diferenciadas entre populações humanas são especialmente informativos para se estimar a miscigenação e são conhecidos como Marcadores Informativos de Ancestralidade (MIAs). No presente estudo são apresentados dois painéis de MIAs (total de 14 SNPs) desenvolvidos para serem genotipados em duas reações de minisequenciamento. O desempenho desses painéis foi avaliado por meio da comparação dos resultados obtidos usando os 14 SNPs em relação aos resultados obtidos com um painel de 108 MIAs genotipados para os mesmos indivíduos. No nível populacional, o painel de 14 MIAs mostrou-se útil para se estimar a miscigenação européia mas pode superestimar a miscigenação africana e subestimar a miscigenação nativo-americana. Nos testes de desempenho, a adição dos resultados de genotipagem de 40 INDELS, previamente validado para se capturar a estrutura de populações humanas, foi capaz de melhorar as estimativas de miscigenação tanto no nível individual quanto no populacional. Esses painéis de 14 SNPs e 40 INDELS foram utilizados para se estimar a miscigenação em indivíduos portadores de doença falciforme (n=200) e indivíduos supostamente saudáveis, doadores de sangue (n= 291) amostrados em diferentes regiões do estado de Minas Gerais. Os resultados obtidos apontam para uma intensa miscigenação em ambos os grupos, embora uma miscigenação africana mais elevada tenha sido observada entre os pacientes. Variações nos níveis de miscigenação entre as diferentes regiões de Estado de Minas Gerais, bem como entre os indivíduos amostrados em cada região, foram observadas tanto para os pacientes quanto para os doadores. Esses resultados advertem para o risco de associações espúrias em estudos de epidemiologia genética com desenho caso-controle envolvendo esses grupos. Por outro lado, a constatação de que os pacientes portadores de doença falciforme possuem níveis substanciais de miscigenação abre a perspectiva de que se possa empregar a técnica de mapeamento por miscigenação na investigação de polimorfismos genéticos associados às diferentes manifestações clínicas dessa doença.

ABSTRACT

Admixture occurs when individuals from parental populations that have been isolated for hundreds of generations, form a new hybrid population. The interest in estimating admixture is critical in genetic epidemiology studies in admixed populations, since stratification generated by admixture can lead to false-positive associations. Markers with highly differentiated frequencies among human populations are especially informative for estimating admixture and are called Ancestry Informative Markers (AIMs). We present two multiplex panels of AIMs (for a total of 14 SNPs) developed to be genotyped by two mini-sequencing reactions. We tested the performance of these AIMs by comparing results obtained using our 14 AIMs with those obtained using 108 AIMs genotyped in the same individuals. At population level, our 14 AIMs are useful to estimate European admixture, but overestimate African admixture and underestimate Native American admixture. In performance tests, the addition of genotyping results for 40 INDELS, previously validated to capture the structure of human populations, improved the estimative of admixture both at individual and population levels. These panels of 14 SNPs and 40 INDELS were used to estimate admixture in subjects with sickle cell disease ($n = 200$) and healthy individuals, blood donors ($n = 291$), sampled in different regions of the state of Minas Gerais. The results point to an intense admixture in both groups, although a higher African admixture was observed among patients. Variations in levels of admixture between the different regions of Minas Gerais, as well as among individuals sampled in each region, were observed for both patients and donors. These results warn of the risk of spurious associations in genetic epidemiology studies with case-control design, involving these groups. Moreover, the finding that patients with sickle cell disease have substantial levels of admixture opens the perspective for the use of admixture mapping strategy in the investigation of genetic polymorphisms associated with different clinical manifestations of this disease

Key words: Mini-sequencing, admixture, sickle cell disease, blood donor.

1 INTRODUÇÃO

1.1 Constituição da população brasileira

Os brasileiros constituem uma população altamente miscigenada (Callegari-Jacques *et al.*, 2003; Parra, *et al.*, 2003; Santos *et al.*, 2010). Três grupos parentais principais contribuíram na formação da população brasileira: Africanos, Europeus e Nativos Americanos. Cinco séculos de um processo intenso de migração e miscigenação envolvendo estes três grupos continentais moldaram no Brasil uma população heterogênea, cujos cromossomos são verdadeiros mosaicos compostos de fragmentos genômicos de origens biogeográficas diversas.

O território que viria a tornar-se o Brasil era originalmente povoado por um número incerto de indivíduos. Estimativas mais aceitas indicam que a população nativa do Brasil seria de aproximadamente 2,5 milhões de pessoas, por volta do ano de 1500 (Vainfas, 2000). A maior parte desta população foi dizimada ao longo do processo de colonização, a ponto de ter sido reduzida para pouco mais de 300.000 indivíduos, conforme estimativas realizadas no ano de 1998.

Os europeus que imigraram para o Brasil provinham em sua maioria de Portugal. Acredita-se que cerca de 100 mil portugueses desembarcaram no Brasil nos primeiros dois séculos de colonização, tendo esse número aumentado para 600 mil no século seguinte. Entretanto o auge da imigração portuguesa se deu entre 1901 e 1930, quando a média anual de imigrantes foi superior a 25 mil indivíduos (Venâncio, 2000).

A miscigenação entre europeus e ameríndios iniciou-se logo após a chegada dos primeiros colonizadores, tendo sido, inclusive, encorajada como política colonial de ocupação do território. Este aspecto torna-se posteriormente explícito no Alvará de 4 de abril do ano de 1755, emitido pelo rei português Dom José, estimulando a miscigenação de portugueses com os nativos americanos: “Faço saber aos que este meu Alvará de Lei virem, considerando o quanto convém, que os meus reais domínios da América se povoem e para este fim pode concorrer muito a comunicação com os índios por meio de casamento.” O incentivo ao casamento de colonos brancos com indígenas foi em seguida reforçado pelo “Diretório dos Índios”, documento que se tornou público em 1757 (Almeida, 1997).

A imigração forçada dos africanos, para o trabalho escravo, principalmente nas lavouras e na mineração iniciou-se por volta de 1530. Os africanos que foram trazidos para o Brasil provinham de três grandes regiões: Costa Ocidental da África (principalmente Costa da Mina), África Centro-Ocidental (Angola-Congo) e Costa Oriental (Moçambique), de acordo com registros históricos (Florentino, 1997; Mamigonian, 2004) corroborados por estudos genéticos (Hunemeier *et al.*, 2007). Estudos indicam que aproximadamente quatro milhões de pessoas foram trazidas da África para o Brasil, entre os séculos XVI e meados do XIX, o que faz com que o Brasil seja, entre os países do continente americano, o que recebeu o maior número de escravos africanos (Reis, 2000).

Outros povos, além de ameríndios, portugueses e africanos, contribuíram para a formação da população brasileira, como italianos, espanhóis, japoneses, alemães e sírio-libaneses. Esses imigrantes chegaram ao país em sua maior parte nos séculos XIX e XX. O número estimado de europeus que veio para o Brasil nesta época é de cerca de 5,5 milhões de pessoas (Venâncio, 2000). A imigração européia para o Brasil foi também influenciada por uma política social e ideológica, de caráter racista, conhecida como “branqueamento” ou “europeização” da população brasileira, a qual objetivava a diminuição dos traços africanos e ameríndios, considerados então indesejáveis (Seyferth, 2002).

O estado de Minas Gerais teve um papel fundamental na colonização do interior do Brasil. Diversas cidades foram fundadas durante o chamado “ciclo do ouro” (século XVIII), e durante este período o estado tornou-se o centro econômico da colônia, com rápido povoamento.

Do ponto de vista colonial, o território que veio a se tornar o estado de Minas Gerais foi primeiramente desbravado por bandeirantes paulistas, em 1596, mas sua colonização efetiva somente se deu com a descoberta do ouro em 1693, e então sua exploração sistemática a partir do início do século XVIII (IBGE, 2000). A descoberta do ouro abundante desencadeou um processo de imigração maciça e desordenada. Estima-se que cerca de 300 mil portugueses vieram para o estado na primeira metade do século XVII, além de um grande número de brasileiros de outros estados como Rio de Janeiro e Bahia (Bacci, 2002). Um grande contingente de africanos foi trazido para trabalhar como escravos nas minas e fazendas, a partir do Rio de Janeiro, vindos da África, e a partir da Bahia, vindos da África, mas também das plantações açucareiras do nordeste (Souza, 1994). Isso fez com que Minas Gerais recebesse escravos de várias regiões africanas, pois os escravos que chegavam pelo porto do Rio de Janeiro vinham principalmente de Angola; já os traficantes da Bahia tinham

se especializado em escravos da região da Costa da Mina (Reis, 2000). Os índios, por sua vez, foram massacrados, escravizados e transformados em minoria populacional já no século XVIII (De Souza, 2009).

A miscigenação que se deu nos últimos 500 anos no Brasil foi intensa, porém não ocorreu de maneira uniforme. Uma das características marcantes deste processo é a miscigenação direcional, envolvendo homens europeus e mulheres ameríndias e africanas. Tal padrão resulta de uma combinação entre fatores históricos, princípios administrativos coloniais e valores religiosos e sociais (Stolke, 2006, Gonçalves *et al.*, 2007).

Além dos relatos históricos, a miscigenação direcional no Brasil é corroborada por dados genéticos. Estudos envolvendo a população “branca” do Brasil demonstraram que a maioria dos cromossomos Y (linhagem patrilinear) possui origem européia (Carvalho-Silva *et al.*, 2001; Carvalho-Silva *et al.*, 2006). Por outro lado, estudos utilizando DNA mitocondrial (mtDNA, linhagem matrilinear), revelaram que 33% desses marcadores são de origem ameríndia, 39% européia e 28% africana. A diferença observada para o mtDNA entre as regiões brasileiras estudadas refletem aspectos históricos da constituição da população: predominância das linhagens ameríndias na região amazônica, africanas no nordeste e européias no sul do país (Alves-Silva *et al.*, 2000).

Diversos estudos já foram conduzidos com o intuito de estimar os níveis de miscigenação na população brasileira, por meio de dados genéticos obtidos a partir de marcadores de diferentes tipos (Bortolini *et al.*, 1999; Alves-Silva *et al.*, 2000; Carvalho-Silva *et al.*, 2001; Salzano e Bortolini, 2002; Callegari-Jacques *et al.*, 2003; Parra *et al.*, 2003; Abe-Sandes *et al.*, 2004; Ferreira *et al.*, 2005; Marrero *et al.*, 2005; Pimenta *et al.*, 2006; Lins *et al.*, 2010). De acordo com estudos moleculares utilizando marcadores autossômicos (microssatélites e polimorfismos de nucleotídeos únicos - SNPs), os brasileiros constituem uma população heterogênea, tendo sido relatada uma predominância de ancestralidade européia (70% a 80% de contribuição européia, 10 a 20% de contribuição africana e entre 7 a 18% de contribuição ameríndia) (Callegari-Jacques *et al.*, 2003; Lins *et al.*, 2010). Entretanto, outros estudos, utilizando marcadores autossômicos do tipo inserção-deleção (INDELs), relacionaram proporções de ancestralidade semelhantes a estas apenas no grupo de brasileiros categorizados como “brancos”, independente da região de amostragem (Pena *et al.*, 2009) ou mesmo a indivíduos especificamente amostrados no sul do país (Santos *et al.*, 2010), onde a ancestralidade européia é sabidamente mais elevada. Entre indivíduos pertencentes a grupos identificados como “pretos” ou “afro-descendentes” as estimativas de miscigenação estariam

entre 9 e 37% de ancestralidade européia, 50 a 81% de ancestralidade africana e 14 a 16% de ancestralidade ameríndia, embora diferenças nos processos de amostragem, genotipagem e análise entre os diferentes estudos devam ser levados em consideração (Pena *et al.*, 2009; Santos *et al.*, 2010).

Atualmente, o Instituto Brasileiro de Geografia e Estatística (IBGE), por meio do questionário de Censo Demográfico, categoriza os brasileiros em cinco grupos: branco, preto, pardo, amarelo e indígena, com base em auto-classificação, de acordo com a cor da pele. A Pesquisa Nacional por Amostra de Domicílios (PNAD) realizada em 2008 estimou que a população brasileira é constituída por 48,8% de brancos, 6,8% de pretos, 43,8% de pardos e 0,9% de amarelos ou indígenas (IBGE). Ainda de acordo com o PNAD, o percentual de grupos classificados por auto-declaração de cor de pele na população de Minas Gerais foi similar à estimativa nacional: 46,2% de brancos, 45,0% de pardos, 8,4% de pretos e 0,3% de amarelos ou indígenas.

É importante ressaltar que estudos genéticos evidenciaram uma grande variabilidade nos níveis de ancestralidade de indivíduos brasileiros, mesmo dentro das categorias fenotípicas tradicionais como branco, preto e pardo (Parra *et al.*, 2003), indicando que a auto-classificação ou mesmo averiguação de características fenotípicas como cor da pele, cor e textura do cabelo, constituição do nariz e da boca não são uma medida adequada da ancestralidade genômica individual dos brasileiros, pelo menos quando esta é estimada a partir de um número de marcadores da ordem de poucas dezenas. Em Minas Gerais, já foi demonstrado que 89% dos brancos têm mais de 10% de ancestralidade africana e 87% dos pretos têm mais de 10% de ancestralidade européia (Pena e Bortolini, 2004).

Desta forma, estudos sobre a miscigenação em Minas Gerais, incluindo informações sobre como a ancestralidade genômica individual se distribui ao longo do espaço geográfico são necessários para que seja possível avaliar as implicações da estrutura genética desta população no desenho de estudos de epidemiologia genética. Isto é especialmente importante em grupos para os quais existe interesse emergente quanto a realização de estudos de associação entre genótipos e fenótipos (por exemplo, doenças complexas) do tipo caso-controle.

1.2 Estimativas de ancestralidade biogeográfica e miscigenação

Do ponto de vista evolutivo, a miscigenação pode ser considerada como uma espécie de fluxo gênico, que ocorre quando indivíduos de duas ou mais populações parentais, as quais estiveram isoladas durante algum tempo, unem-se formando uma nova população híbrida (Bernstein, 1931; Chakraborty, 1986; Tarazona-Santos *et al.*, 2007). Neste contexto, as populações latino-americanas, incluindo a brasileira, são bons exemplos de populações miscigenadas.

O homem anatomicamente moderno originou-se na África a cerca de 200.000 anos, de onde se expandiu vindo a povoar todos os continentes. Estudos genéticos e arqueológicos indicam que o povoamento da Europa se deu a aproximadamente 40.000 anos, tendo o *Homo sapiens* substituído outros hominídeos que então habitavam aquele território (DeSalle e Tattersall, 2008; Currat e Excoffier, 2004). O continente Americano, por sua vez, foi povoado há cerca de 14.000 a 22.000 anos (Carvalho-Silva *et al.*, 1999; Underhill *et al.*, 2001; Hammer e Zegura, 2002; Bortolini *et al.*, 2003). A partir do ano de 1.500, quando os primeiros portugueses aportaram no Brasil, iniciou-se um intenso processo de miscigenação, com participação de três grupos biogeográficos principais: ameríndios, europeus e africanos.

A categorização tipológica dos seres humanos em “raças”, ainda hoje utilizada em diversos âmbitos de nossa sociedade, incluindo o meio biomédico, possui origem imprecisa. Entretanto, uma obra em particular parece ter marcado de forma relevante esta classificação: trata-se do livro “*De Generis Humani Varietate Nativa*”, do antropólogo alemão Johan Friedrich Blumenbach publicado em 1795 (Bamshad *et al.*, 2004; Pena, 2005; Barnholtz-Sloan *et al.*, 2008). Nesta obra, o autor descreveu cinco principais “raças” humanas: caucasóide, mongolóide, etiópica, americana e malaia tendo como base a origem geográfica bem como parâmetros morfológicos.

Estudos têm demonstrado que a diversidade genética humana é baixa, quando comparada a outras espécies (Li e Sadler, 1991; Bamshad *et al.*, 2004) e que a maior parte desta variação está dentro das populações (85 a 90%) e não entre elas (10 a 15%) (Barbujani *et al.*, 1997; Rosenberg *et al.*, 2002). Esses achados têm sido usados como argumento, por diferentes grupos de pesquisa, no sentido de questionar o conceito de “raças humanas” como entidades biológicas. Independente desta discussão está o fato de que variação genética humana (ainda que pequena) tende a seguir padrões geográficos, de forma que indivíduos

geograficamente mais próximos tendem a ser geneticamente mais similares do que indivíduos geograficamente distantes (Pena, 2005; Ramachandran *et al.*, 2005; Rosenberg *et al.*, 2002). Isto é o que torna possível a separação de indivíduos em grupos de ancestralidade biogeográfica distintos com base em marcadores biológicos, quer sejam fenotípicos, bioquímicos ou genéticos.

Diversos estudos relativos a estimativas de proporções de ancestralidade biogeográfica em populações miscigenadas, como a brasileira, já foram conduzidos. Para isto, tais estudos empregaram diferentes tipos de marcadores, como os antígenos de grupos sanguíneos, marcadores moleculares uniparentais como os do mtDNA e do cromossomo Y (Bortolini *et al.*, 1999; Alves-Silva *et al.*, 2000; Carvalho-Silva *et al.*, 2001; Abe-Sandes *et al.*, 2004; Marrero *et al.*, 2005), e mais recentemente, marcadores autossômicos (Parra *et al.*, 2003; Callegari-Jacques *et al.*, 2003; Pimenta *et al.*, 2006; Ferreira *et al.*, 2006; Lins *et al.*, 2010; Santos *et al.*, 2010).

Embora os estudos usando marcadores genéticos sejam realizados há muito tempo, os métodos que utilizam genótipos multiloci para atribuir indivíduos a populações e estimar miscigenação são relativamente recentes (Beaumont e Rannala, 2004). Marcadores com grandes diferenças de frequência entre as populações parentais (Marcadores Informativos de Ancestralidade, MIAs) são especialmente úteis para estimar a miscigenação. O uso de marcadores altamente informativos é desejável para a redução dos esforços de genotipagem, por permitir a diminuição do número de marcadores necessários para estimar a miscigenação, quando comparado ao uso de marcadores escolhidos ao acaso (Rosenberg *et al.*, 2003; Pfaff *et al.*, 2004).

Para se estimar a miscigenação de populações latino-americanas, que são geralmente tri-híbridas, é preciso acessar a informação de ancestralidade das três populações parentais: africana, européia e ameríndia. Entretanto, não há um único painel de marcadores que pode ser considerado ideal e que possa ser aplicado a todas as populações latino-americanas. Isto ocorre porque a informatividade de um painel depende não só das frequências alélicas dos marcadores nas populações parentais, mas também das proporções de contribuição dessas populações para a formação da população miscigenada alvo do estudo (Pfaff *et al.*, 2004), e estas proporções podem variar substancialmente conforme a população estudada (Sans, 2000). Em geral, para populações latino-americanas tri-híbridas, os melhores MIAs são aqueles que apresentam uma frequência muito diferente em uma das três populações parentais e frequências similares nas outras duas.

Outra questão importante acerca da construção de painéis de marcadores com o intuito de estimar a miscigenação consiste no número de marcadores necessários para se alcançar uma estimativa adequada. Este número não é fixo, mas depende da informatividade dos marcadores que compõem o painel, bem como no nível de precisão que se deseja alcançar. Apesar da possibilidade técnica atual de se genotipar centenas de milhares de marcadores a custos cada vez mais acessíveis (Chung *et al.*, 2010), dependendo da composição de marcadores e da população alvo do estudo, painéis relativamente pequenos (por exemplo, 24 MIAs) podem ser úteis para se determinar a origem continental de indivíduos e para corrigir a estratificação devido à miscigenação (Kosoy *et al.*, 2008).

Paralelamente aos avanços técnicos que tornaram possível hoje genotipar um grande número de marcadores, permitindo assim uma melhor representação do genoma, está o desenvolvimento dos métodos estatísticos com o objetivo de estimar a origem biogeográfica. Os diferentes métodos já propostos são capazes de estimar a miscigenação nos níveis populacional, individual e até mesmo genômico.

As aplicações dos métodos de estimativa de miscigenação incluem uma relevante contribuição na compreensão da história evolutiva das populações (Chikhi *et al.*, 2002) e investigações de epidemiologia genética (Chakraborty e Weiss, 1988). Além destas aplicações “clássicas”, as estimativas de origem biogeográficas com base em dados genéticos atualmente extrapolam os estudos antropológicos e biomédicos, influenciando as ciências forenses, a genômica personalizada como serviço direto ao consumidor com interesses comerciais, bem como questões envolvendo os direitos civis de minorias (Soo-Jin Lee *et al.*, 2009).

1.2.1 Miscigenação no nível populacional

Os métodos estatísticos para estimar proporções de miscigenação com base em dados genéticos começaram a ser desenvolvidos a partir da década de 30. Desde então, inúmeros foram os esforços para desenvolver estimadores eficientes (Wang, 2003).

Dentre os métodos desenvolvidos para se estimar a miscigenação no nível populacional, os que utilizam as abordagens de *Identidade Genética (IG)* e *Tempos de Coalescência (TC)* estão entre os mais citados na literatura nos últimos 20 anos (Choisy *et al.*, 2004).

A equação formulada em 1932 por Bernstein foi um dos passos inaugurais para o desenvolvimento de estimadores de proporções de miscigenação. A equação de Bernstein expressa a frequência $p_i^{(h)}$ do alelo i na população híbrida como uma combinação linear das frequências alélicas $p_i^{(1)}$ e $p_i^{(2)}$ de duas populações parentais, onde λ é a contribuição das populações parentais na formação da população miscigenada (Bernstein, 1932 *apud* Choisy *et al.*, 2004):

$$p_i^{(h)} = \lambda p_i^{(1)} + (1 - \lambda) p_i^{(2)}.$$

Desta forma, a probabilidade de um alelo de frequência $p_i^{(1)}$ na população 1 ser encontrado na população híbrida h é dependente da fração de genes da população 1 encontrado na população híbrida (Frudakis, 2008).

O método de Identidade Genética (Chakraborty, 1985), baseia-se no coeficiente de Identidade Genética de Nei (1971), tendo como ponto de partida a equação de Bernstein. O Coeficiente de Identidade Genética é definido como a probabilidade de que dois genes tirados ao acaso, um da população 1 e outro da população 2 sejam idênticos (Elston *et al.*, 2002; Frudakis, 2008). A abordagem IG baseia-se na lógica de que a miscigenação causa a mesma mudança na frequência alélica para qualquer locus do genoma e, contanto que não haja seleção favorecendo ou desfavorecendo alelos, estimativas de miscigenação em populações híbridas podem ser expressas em termos da probabilidade de selecionar genes idênticos por estado, a partir das populações parentais e da população híbrida (Frudakis, 2008).

Os métodos de estimativa de miscigenação geralmente assumem que as proporções de miscigenação estimadas a partir das frequências alélicas atuais são representativas das contribuições das populações parentais para a formação da população miscigenada, ocorridas na época da hibridização.

Um problema comum aos métodos que, para estimar a miscigenação, baseiam-se na comparação das frequências alélicas entre as populações parentais e as miscigenadas é que como estas frequências não são conhecidas e sim estimadas a partir de amostras das populações atuais, fatores como os erros de amostragem, a deriva genética e a ocorrência de mutações desde o evento de miscigenação podem tornar os resultados desses estimadores imprecisos, caso não sejam levados em consideração, sobretudo se o evento de miscigenação for antigo (Wang, 2003; Wang, 2006).

O método de identidade genética desenvolvido por Chakraborty (1985) e

implementado no programa ADMIX95 leva em consideração possíveis erros de amostragem, bem como o efeito da deriva genética, tanto nas populações parentais quanto nas miscigenadas (Chakraborty, 1986).

Uma estratégia alternativa e mais recente para estimar a miscigenação populacional consiste no uso da Teoria do Coalescente (Bertorelle e Excoffier, 1998). Esta metodologia leva em consideração o tempo médio de coalescência para um par de genes, um deles amostrado na população híbrida e outro amostrado em uma população parental. As linhagens ancestrais de cada um dos pares de genes amostrados podem coalescer de duas formas: 1) A linhagem ancestral do gene amostrado na população miscigenada muda para a mesma população parental do outro gene do par amostrado, onde irão coalescer ou, 2) o gene amostrado na população híbrida muda para a outra população parental e eles irão coalescer na população ancestral (população a partir da qual se originaram as populações parentais). Tendo como base a distribuição esperada desses tempos de coalescência, Bertorelle e Excoffier (1998) desenvolveram um estimador de miscigenação no nível populacional que apresenta eficiência mesmo em condições mais extremas (Bertorelle e Excoffier, 1998; Choisy *et al.*, 2004). Posteriormente, este método foi estendido para levar em consideração qualquer número de populações parentais (Dupanloup e Bertorelle, 2001).

Estudos comparativos apontam que em condições ideais como populações parentais altamente diferenciadas e evento de miscigenação recente, métodos como IG e TC possuem desempenho igualmente bom. No entanto, naqueles casos em que as condições não são ótimas, os métodos apresentam desempenhos diferentes. Ainda assim, não há um único estimador que mostre superioridade em relação aos outros em toda e qualquer circunstância (Choisy *et al.*, 2004; Excoffier *et al.*, 2005).

1.2.2 Miscigenação nos níveis individual e genômico

A consideração de fatores evolutivos como a ação da deriva nas frequências alélicas desde o evento de miscigenação até o evento da amostragem representou uma melhoria nos métodos de estimativa de miscigenação no nível populacional (Excoffier *et al.*, 2005). Entretanto existe um reconhecimento crescente da importância de se estimar a miscigenação no nível individual (Tang *et al.*, 2005), sobretudo em estudos de associação genética envolvendo doenças complexas (Barnholtz-Sloan *et al.*, 2008).

A diferença fundamental entre as estimativas de miscigenação nos níveis populacional e individual é que no primeiro caso será estimada a fração do “pool” gênico da população miscigenada que se originou de uma ou outra, entre as populações parentais, enquanto no segundo caso, serão estimadas proporções de loci do genoma de um único indivíduo miscigenado, originadas de cada uma das populações parentais sob análise.

Uma das abordagens mais utilizadas para estimar a miscigenação no nível individual por meio de genótipos multiloci consiste no método Bayesiano implementado no programa STRUCTURE (Pritchard *et al.*, 2000). Embora o foco original desta abordagem tenha sido a identificação de sub-populações, o STRUCTURE tornou-se extremamente popular em estudos de genética de populações com o intuito de medir as proporções de contribuição de populações parentais para o genoma de um indivíduo miscigenado (Tang *et al.*, 2005).

O procedimento implementado no STRUCTURE consiste de um método de agrupamento (do inglês, “*clustering*”) para inferência de estrutura populacional, usando dados de genótipos de marcadores não ligados. O modelo assumido pelo método considera que: 1) existem “k” populações, caracterizadas por conjuntos de frequências alélicas em cada locus e 2) os loci estão em equilíbrio de Hardy-Weinberg e equilíbrio de ligação dentro das populações. Indivíduos da amostra são então atribuídos probabilisticamente a uma dessas “k” populações (ou a mais de uma população, caso sejam miscigenados) de forma que, dentro das populações sejam alcançados os equilíbrios de Hardy-Weinberg e de ligação, tanto quanto possível (Pritchard *et al.*, 2000).

Embora este método possibilite que indivíduos sejam atribuídos a populações de forma apropriada, os pesquisadores devem estar atentos ao fato de que a precisão das estimativas de miscigenação obtidas com o uso do STRUCTURE depende de vários fatores como o número de indivíduos analisados, o número de loci, a quantidade de miscigenação e a extensão da diferença das frequências alélicas entre as populações parentais (Pritchard *et al.*, 2000).

Com o avanço dos estudos genéticos, emergiram algumas situações nas quais a necessidade de se estimar a origem biogeográfica ou a miscigenação extrapola até mesmo o nível do indivíduo, emergindo assim o interesse de se determinar a fonte de partes específica de um dado genoma. Neste contexto, o modelo originalmente implementado no programa STRUCTURE foi posteriormente estendido (Falush *et al.*, 2003), tendo sido então proposto um novo modelo, capaz de levar em consideração a correlação entre marcadores ligados que surge como resultado da miscigenação (desequilíbrio de ligação gerado por miscigenação). Esta extensão metodológica permitiu a utilização do programa STRUCTURE para a

realização de estimativas da origem de regiões cromossômicas individuais.

Um exemplo de situação onde torna-se importante estimar a origem biogeográfica de regiões específicas do genoma é a estratégia conhecida como mapeamento por miscigenação – MM. O Mapeamento por miscigenação parte do princípio de que os efeitos da miscigenação podem ser utilizados para mapear genes responsáveis por fenótipos complexos (por exemplo, doenças) que apresentam diferenças na incidência entre as populações parentais que deram origem à população miscigenada, alvo do estudo. No MM, o interesse consiste em determinar se certas regiões do genoma dos indivíduos miscigenados e portadores da doença são herdadas de uma das populações parentais, em uma taxa maior que a esperada, o que pode ser alcançado por meio da genotipagem de marcadores informativos de ancestralidade ao longo dos cromossomos. Desta forma, espera-se que na região genômica que contém a mutação envolvida na patologia, a contribuição da população parental onde esta mutação é mais frequente será maior, em comparação com a mesma região genômica de indivíduos não afetados ou mesmo em comparação com outras regiões genômicas dos próprios indivíduos afetados (Seldin *et al.*, 2004; Reich e Patterson, 2005; Smith e O'Brien, 2005).

Sendo grande a diversidade de métodos estatísticos desenvolvidos com o intuito de estimar a miscigenação em diferentes níveis, a escolha do método a ser empregado estará fundamentalmente relacionada aos objetivos específicos da investigação. Desta forma, o pesquisador deverá sempre levar em consideração as aplicações e também limitações inerentes a cada método.

1.3 Implicações da miscigenação no desenho de estudos epidemiológicos

1.3.1 Impacto da estratificação gerada pela miscigenação em estudos de epidemiologia genética do tipo caso-controle

Estudos epidemiológicos com desenho do tipo caso-controle constituem uma ferramenta importante para mapear genes envolvidos em doenças. Este tipo de estudo baseia-se na premissa de que se um determinado polimorfismo genético aumenta a susceptibilidade a uma doença, espera-se que sua frequência seja mais elevada entre os indivíduos afetados

(casos) do que entre os não afetados (controles). Contudo, a estratificação populacional pode comprometer a eficácia desta abordagem, por levar a associações falso-positivas entre marcadores genéticos e fenótipos de interesse. Este efeito pode ocorrer tanto em populações estruturadas (uma população maior, composta por diversas sub-populações, as quais apresentam diferentes frequências alélicas e diferentes predisposições à doença) ou em populações miscigenadas, caso o risco de desenvolver a doença dependa do grau de miscigenação (Pritchard *et al.*, 2001).

A população brasileira, sendo altamente miscigenada, está sujeita aos efeitos adversos da estratificação populacional em estudos de epidemiologia genética. Se uma doença apresenta uma maior incidência em um dos três grupos étnicos principais (europeus, ameríndios ou africanos) que contribuíram na formação da população brasileira, este grupo étnico pode estar sobre-representado na amostra de casos em estudos caso-controle. Nesta situação, qualquer variante gênica associada estatisticamente com o grupo étnico de maior incidência, pode apresentar associação com a doença, independentemente do seu envolvimento na patogênese. Esta situação pode levar a resultados falso-positivos, conhecidos como associações espúrias, em estudos do tipo caso-controle (**Figura 1**). Desta forma, para evitar as associações espúrias em estudos envolvendo populações miscigenadas é necessário conhecer e controlar o efeito da estratificação (Schlesselman, 1982; Tarazona-Santos *et al.*, 2007).

O risco de obter resultados falso-positivos em estudos de associação genótipo-fenótipo depende do nível de estruturação ou estratificação da população. Numa população na qual todos os indivíduos têm o mesmo nível de miscigenação, o problema não existe. Contrariamente, o problema aumenta se a miscigenação é variável entre indivíduos, de modo que casos e controles podem ter níveis de miscigenação diferentes (**Figura 1**).

Em algumas regiões do Brasil, como é o caso do Rio Grande do Sul, foi evidenciado uma ancestralidade genômica homogênea em um estudo epidemiológico casos-controles de doenças cardiovasculares (Zembruski *et al.*, 2006). Porém, outras regiões, como Minas Gerais, parecem apresentar outra realidade.

Estudos já conduzidos em populações mineiras apontam para a existência de níveis substanciais de miscigenação. Entre indivíduos mineiros auto-declarados como “brancos” (n=142) foi demonstrada, por meio da genotipagem de 40 polimorfismos de inserção-deleção, a ocorrência de 70,8% de ancestralidade européia, 16,1% de ancestralidade africana e 13,1% de ancestralidade ameríndia (Pena *et al.*, 2009). Além disso, em análises de miscigenação no nível individual para esse mesmo grupo, utilizando o programa STRUCUTRE, foi observado

que embora a maioria dos indivíduos apresentasse ancestralidade predominantemente europeia, vários deles possuíam diferentes níveis de miscigenação africana e ameríndia. Estes dados sugerem que a miscigenação individual em Minas Gerais pode ser variável, mesmo entre indivíduos agrupados em categorias fenotípicas específicas (no caso, “brancos”)

Outro estudo, conduzido em indivíduos da comunidade rural de Queixadinha, município de Carai, Vale do Jequitinhonha (n=173), categorizados em três grupos (brancos, pardos e preto), com base em características fenotípicas, também apontam para uma variação na miscigenação inter-individual. Os resultados obtidos a partir da análise de 10 MIAs mostraram que, no nível populacional, a proporção de ancestralidade africana foi de 51% para o grupo dos indivíduos classificados como pretos, 44% para o grupo dos pardos, 32% para o grupo dos brancos e 44% para a população total. Uma análise do Índice de Ancestralidade Africana (uma estimativa individual da ancestralidade africana) foi também realizada para esses indivíduos. Os resultados dessa análise mostraram a ocorrência de um largo espectro de valores, mesmo dentro das categorias fenotípicas branco, pardo ou preto, o que indica que há uma substancial variação na miscigenação africana entre esses indivíduos (Parra *et al.*, 2003).

Já na cidade de Belo Horizonte, capital do estado, com base na genotipagem de 13 microsátélites de uso forense em 234 indivíduos não relacionados e oriundos do banco de dados de testes de paternidade de um laboratório particular, foi estimada uma contribuição europeia de 65%, africana de 34%, e ameríndia de 1% (Scliar *et al.*, 2009). Embora os resultados desse estudo forneçam informações apenas no nível populacional, mais uma vez foram evidenciados altos níveis de miscigenação na população mineira.

Desta forma, fica evidente que para planejar estudos de epidemiologia genética de doenças complexas em Minas Gerais, é importante conhecer os níveis de miscigenação individual dos habitantes do estado de modo geral, e também de grupos específicos, para os quais existe interesse de se realizar estudos de associação genética. É importante identificar padrões espaciais, tais como a existência de gradientes de miscigenação ou regiões geográficas onde os padrões de miscigenação mudam abruptamente. Estes conhecimentos contribuirão para definir e conhecer melhor os níveis de estratificação populacional em grupos a serem alvos de estudos epidemiológicos e conseqüentemente, poderão permitir aos pesquisadores planejar melhor esses estudos.

1.3.2 Controle dos efeitos da estratificação em estudos de epidemiologia genética do tipo caso-controle

Considerando que a estratificação entre casos e controles pode levar a efeitos indesejáveis em estudos epidemiológicos, diversas estratégias foram desenvolvidas com o intuito não apenas de detectar, mas também de atenuar ou corrigir este problema.

Uma forma tradicional de lidar com o problema da estratificação é equiparar, da melhor forma possível, o componente étnico de casos e controles. Entretanto, atualmente há dúvidas a respeito da efetividade dessa estratégia, já que na maior parte dos casos, indivíduos podem não relatar sua ancestralidade de modo preciso. Além disso, grupos étnicos culturalmente definidos podem não representar a real estrutura genética da população (Pritchard *et al.*, 2001). Assim, uma quantidade considerável de “estratificação críptica” pode permanecer mesmo após a equiparação cuidadosa de casos e controles (Reich e Goldstein, 2001; Barnholtz-Sloan *et al.*, 2008). No Brasil, onde a população é miscigenada, já foi demonstrado que características fenotípicas associadas a determinados grupos étnicos, como cor da pele, textura do cabelo e conformação dos lábios e nariz, não necessariamente descrevem a ancestralidade genética individual (Parra *et al.*, 2003), o que aumenta a complexidade deste cenário.

Embora não exista um consenso a respeito de qual seria a metodologia mais adequada para remover ou reduzir o viés causado pela estratificação, métodos que utilizam resultados da genotipagem, quer sejam de marcadores selecionados de forma aleatória, quer sejam de marcadores informativos de ancestralidade, têm sido ferramentas úteis em estudos de epidemiologia genética. Dentre esses métodos, duas categorias se destacam: o controle genômico e a associação estruturada (Barnholtz-Sloan *et al.*, 2008).

Uma das primeiras abordagens utilizadas para lidar com a estratificação foi o método de controle genômico. Este método considera a estratificação no nível populacional e pode ajudar a prevenir associações falso-positivas em estudos caso-controle realizados em populações nas quais existe uma quantidade desconhecida de estratificação (Barnholtz-Sloan *et al.*, 2008). Nesta abordagem, para averiguar se a associação estatística observada para um marcador candidato em relação a uma determinada doença é verdadeira, é avaliada também a associação de marcadores não ligados para os quais se espera não haver relação com a doença (marcadores nulos). Se a associação estatística encontrada para os marcadores nulos, em relação à doença em estudo não for muito diferente da encontrada para o marcador candidato,

conclui-se que aquela associação pode ser atribuída apenas à estratificação (Reich e Goldstein, 2001; Barnholtz-Sloan *et al.*, 2008).

Na prática, para verificar o quanto a existência de uma eventual estratificação está afetando os resultados de um estudo caso-controle, é necessário acessar o grau de elevação dos valores do teste estatístico de associação padrão (X^2) causado pela estratificação (Barnholtz-Sloan *et al.*, 2008). Para isto, o conjunto de marcadores nulos é usado para estimar um fator de inflação do valor de X^2 obtido a partir da análise do marcador candidato. O método assume que esta inflação pode ser atribuída à estratificação, e procura então corrigir o teste estatístico de associação X^2 , por meio da razão entre o teste de associação padrão e o fator de inflação (Reich e Goldstein, 2001).

A outra classe de métodos desenvolvidos com o intuito de controlar os efeitos adversos da estratificação populacional em estudos de epidemiologia genética com desenho caso-controle consiste da abordagem de Associação Estruturada. Métodos de associação estruturada utilizam um conjunto de marcadores não ligados para atribuir indivíduos a grupos ou sub-populações. Uma das implementações mais comumente usadas para este fim é o programa STRUCTURE, capaz de estimar a miscigenação no nível individual e atribuir indivíduos a populações a partir de genótipos multiloci. O conhecimento detalhado da estruturação da população pode então ser levado em consideração nos estudos de epidemiologia genética. Nesta abordagem, conjuntos de marcadores selecionados ao acaso ou MIAs podem ser utilizados para definir as sub-populações (Pritchard *et al.*, 2000; Pritchard *et al.*, 2001; Barnholtz-Sloan *et al.*, 2008).

Enquanto nos testes de associação tradicionais a hipótese nula é que a frequência alélica do gene candidato independe do fenótipo sob análise, no caso da abordagem de associação estruturada, a hipótese nula passa a ser que esta independência existe especificamente dentro das sub-populações que constituem a população alvo do estudo. Sendo assim, os testes de associação são conduzidos dentro das sub-populações definidas pelo método de agrupamento. Um problema inerente ao método de associação estruturada é que se houver falha na definição das sub-populações, a eficácia do método poderá ser comprometida (Pritchard *et al.*, 2001; Barnholtz-Sloan *et al.*, 2008).

Como não existe uma única forma de se controlar a estratificação em estudos epidemiológicos, a decisão sobre a escolha do método deverá levar em conta a convicção por parte do pesquisador de que as premissas do método escolhido não estão sendo violadas e que o método escolhido é realmente adequado para testar a hipótese em questão (Barnholtz-Sloan *et al.*, 2008).

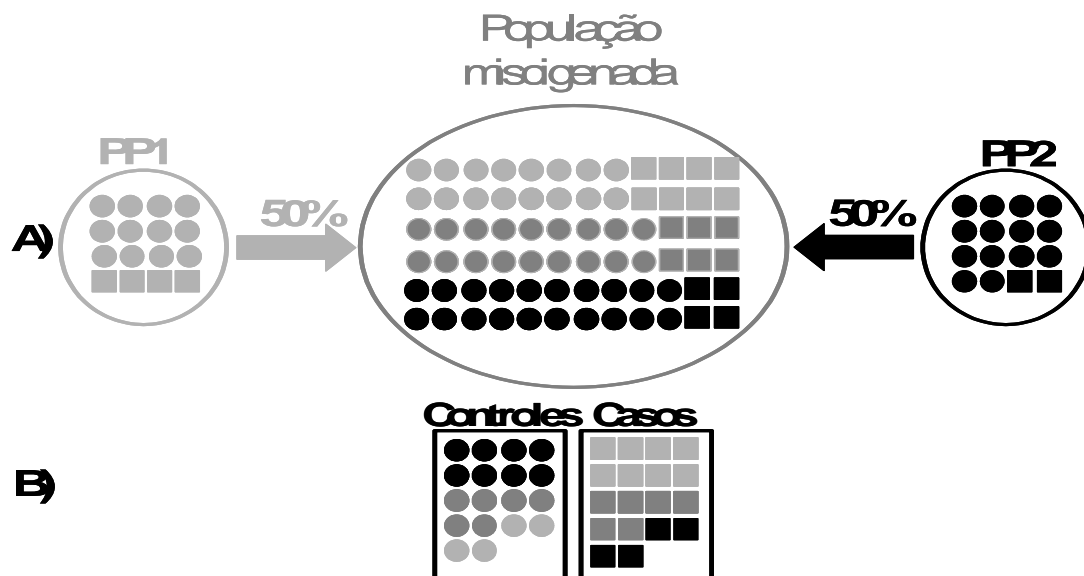


Figura 1. Representação da miscigenação e do efeito da estratificação de populações em estudos de epidemiologia genética. **A)** Cada uma das populações parentais PP1 (cinza claro) e PP2 (preto) contribuíram com 50% do “pool” gênico da população miscigenada (cinza escuro). Indivíduos da população miscigenada possuem diferentes níveis de miscigenação. Na figura, esta situação foi simplificada, sendo apresentados apenas três níveis de miscigenação individual: indivíduos cujos genomas derivam completamente de PP1 (cinza claro), de PP2 (preto) e indivíduos miscigenados (cinza escuro). Indivíduos saudáveis e doentes são representados por círculos e quadrados, respectivamente. Note que a incidência da doença é maior em PP1 do que em PP2. Na população miscigenada, a incidência da doença é maior entre os indivíduos cinza claros, intermediária entre os cinza escuros e menor entre os pretos. **B)** Um estudo caso-controle: casos são representados por quadrados e controles por círculos. Como a incidência da doença é maior entre indivíduos com maior ancestralidade cinza claro (PP1), estes estarão super-representados entre os casos. Como consequência, qualquer alelo que tenha frequência mais alta em PP1 do que em PP2 pode apresentar associação com a doença, independentemente se ele é o alelo causal.

Fonte: Tarazona-Santos et al, 2007.

1.4 A doença falciforme

1.4.1 Aspectos moleculares e manifestações clínicas

A hemoglobina (Hb) é uma proteína globular presente no interior dos eritrócitos cuja principal função é o transporte de oxigênio para os tecidos. Sua estrutura é composta por quatro cadeias polipeptídicas, cujos centros apresentam um complexo denominado heme, onde localiza-se um íon ferro.

Ao longo do desenvolvimento humano, diferentes grupos de genes do agrupamento da alfa ou da beta globina são ativados ou suprimidos e diferentes cadeias globínicas são sintetizadas. A combinação destas diferentes cadeias origina os diferentes tipos de hemoglobinas que se sucedem ao longo da vida humana, da fase embrionária até a adulta (Ashley-Koch *et al.*, 2000). Nos adultos, a molécula de hemoglobina predominantemente é a HbA (96 a 98%), composta por duas globinas do tipo alfa e duas do tipo beta, embora a HbA2, composta por cadeias alfa e delta e a HbF (hemoglobina fetal), composta por cadeias alfa e gama, também possam ser normalmente encontradas (Neto e Pitombeira, 2003).

A ocorrência da substituição de uma valina (aminoácido neutro) por um ácido glutâmico (aminoácido com carga negativa) na sexta posição da cadeia polipeptídica da globina beta (Ingram, 1956; Ingram, 1957), origina uma hemoglobina anormal denominada hemoglobina S (HbS). As propriedades eletroforéticas distintivas da HbS foram demonstradas pela primeira vez em 1949, por Linus Paulin e sua base genética consiste em uma mutação de ponto (GAG → GTG). A alteração molecular descrita provoca mudanças nas características físico-químicas da hemoglobina tais como carga elétrica e solubilidade, conferindo à HbS uma tendência à formação de polímeros na fase desoxigenada (Ferrone, 2004; Pace, 2007).

A polimerização da HbS que ocorre sob condições de desoxigenação é o evento primário, indispensável para a patogênese molecular da doença falciforme, sendo a causa do fenômeno de falcização, na qual os eritrócitos perdem sua forma habitual de disco bicôncavo, assumindo uma forma de foice. Neste estado, as hemácias perdem sua maleabilidade, tornando-se mais rígidas, o que dificulta o fluxo dessas células na micro-circulação e encurta suas sobrevidas (Frenette e Atweh, 2007; Pace, 2007). As modificações celulares ocasionadas pela polimerização da HbS causam ainda a ativação da coagulação e a depleção de óxido nítrico (NO), o que contribui para a vasoconstrição e ativação da inflamação (Zago e Pinto, 2007).

A ocorrência de vaso-oclusões, principalmente dos pequenos vasos, é favorecida pela maior adesão dos eritrócitos falcizados ao endotélio, o que desencadeia fenômenos inflamatórios (Hoover *et al.*, 1979; Frenette e Atweh, 2007). Estas vaso-oclusões são responsáveis pela maioria dos sinais e sintomas observados pelos pacientes portadores de doença falciforme (Schnog *et al.*, 2004; Stuart e Nagel, 2004; Zago e Pinto, 2007). O retardo da circulação dos eritrócitos, por sua vez, resulta em uma amplificação do fenômeno de falcização, agravando as alterações celulares, teciduais e sistêmicas observadas na doença falciforme.

Além do estado de oxigenação, outros fatores como porcentagem de HbS intracelular, o grau de desidratação celular, a concentração de hemoglobina corpuscular média, o tempo de trânsito dos glóbulos vermelhos na micro-circulação, a composição das hemoglobinas dentro das células, e o pH, influenciam na polimerização da HbS (Noguchi, 1984; Schnog *et al.*, 2004; Frenette e Atweh, 2007; Pace, 2007).

A ocorrência da mutação responsável pela produção da HbS em sua forma homozigótica caracteriza a doença denominada Anemia Falciforme, que é a forma mais comum e mais severa da doença (Zago, 2002; Frenette e Atweh, 2007). Entretanto, a HbS pode ser encontrada juntamente com outras hemoglobinas anormais como a HbC, HbD, etc, resultando em formas que também são sintomáticas. O termo Doenças Falciformes abrange todas estas formas sintomáticas onde a HbS está presente e que apresentam graus variados de gravidade (Zago, 2002). Por outro lado, indivíduos que apresentam a HbS em heterozigose com a HbA são portadores assintomáticos em condições fisiológicas normais e não apresentam alterações na expectativa de vida (Alberto e Costa, 2001).

Do ponto de vista da severidade dos sintomas, a doença falciforme caracteriza-se como uma doença complexa, apresentando diversas manifestações clínicas (Stuart e Nagel, 2004; Zago e Pinto, 2007; Steinberg, 2009). Alguns pacientes apresentam sintomas de grande gravidade, gerando alta morbidade e mortalidade, redução da capacidade laboral e da expectativa de vida, enquanto outros indivíduos apresentam evolução benigna chegando em alguns casos a serem quase assintomáticos. As principais intercorrências relacionadas às doenças falciformes são: anemia crônica, crises álgicas, síndrome torácica aguda, úlceras de membros inferiores, sequestro esplênico, priapismo, necrose óssea, retinopatia, insuficiência renal crônica e acidente vascular cerebral (Schnog *et al.*, 2004).

Fatores adquiridos e herdados estão envolvidos na variabilidade clínica encontrada nas doenças falciformes. O fator ambiental mais importante parece ser a condição sócio-econômica do paciente, embora fatores climáticos e desidratação também tenham sido

associados a crises vaso-oclusivas (Nolan *et al.*, 2008). Entre os fatores genéticos destacam-se os haplótipos associados à HbS, por influenciarem o nível de HbF, a concomitância da alfa-talassemia, por provocar redução da HbS intracelular, bem como uma série de polimorfismos genéticos capazes de modular a fisiopatologia da doença (Nagel *et al.*, 1985; Nagel e Steinberg, 2001; Nagel, 2005).

Cinco haplótipos relacionados à HbS, definidos pela presença de sítios polimórficos ao longo do grupamento de genes da beta-globina, já foram descritos. Cada haplótipo recebe denominações de acordo com os locais de sua origem: Benin, República Centro-Africana - CAR (ou Banto), Senegal, Camarões e Árabe-Indiano (Pagnier *et al.*, 1984; Nagel *et al.*, 1985; Oner *et al.*, 1992; Powar *et al.*, 1994). Estudos moleculares evidenciaram origens independentes para cada haplótipo da HbS (Wainscoat *et al.*, 1983; Pagnier *et al.*, 1984). A partir de suas origens, as mutações aumentaram suas frequências em populações expostas de forma endêmica à malária falciparum, por efeito da seleção natural, já que a presença da HbS em heterozigose com a HbA torna os indivíduos resistentes às formas mais graves da doença (Allison *et al.* 1954; Allison, 1964; Hanchard *et al.*, 2007).

Os haplótipos associados à HbS possuem relação com a severidade dos sintomas apresentados pelos pacientes. De um modo geral, portadores dos haplótipos Senegal ou Árabe-Indiano têm maiores níveis de HbF e apresentam formas menos graves da doença, enquanto os portadores do haplótipo CAR apresentam menor nível de HbF e manifestações mais severas. Já os haplótipos Benin e Camarões estariam associados a formas intermediárias da doença. (Powars, 1991; Powars e Hiti, 1993; Ashley-Koch *et al.*, 2000; Schnog *et al.*, 2004; Steinberg, 2006; Rund e Fucharoen, 2008; Steinberg, 2009).

Embora diversos polimorfismos do agrupamento dos genes da beta-globina tenham sido descritos, não se sabe precisamente como os haplótipos da HbS interferem nas expressões fenotípicas da anemia falciforme. Além disso, outros fatores como sexo e idade estão relacionados à persistência da HbF, de forma que não é possível aplicar a ocorrência dos haplótipos, isoladamente, como marcadores de prognóstico no acompanhamento clínico individual dos pacientes (Belisario, 2010).

Com o avanço da compreensão da patologia, aumenta o número de potenciais genes modificadores associados às diferentes manifestações clínicas da doença falciforme. Polimorfismos que apresentam impacto na adesão celular, trombose, desidratação dos eritrócitos, inflamação e persistência de HbF são candidatos a apresentarem envolvimento na heterogeneidade dos sintomas observados (Adewoye *et al.*, 2006; Steinberg, 2006; Frenette e Atweh, 2007). Diversos estudos têm sido conduzidos no intuito de relacionar polimorfismos

genéticos com a predisposição a determinados sintomas (ou sub-fenótipos) da doença. Resultados desses estudos incluem a associação de diversos polimorfismos (SNPs) com complicações clínicas da doença falciforme como acidente vascular cerebral, (Sebastiani *et al.*, 2005), priapismo (Nolan *et al.*, 2004), osteonecrose (Baldwin *et al.*, 2005), crises álgicas (Lettre *et al.*, 2008), e hipertensão pulmonar (Allison *et al.*, 2008).

A maior parte dos estudos de associação entre SNPs e sub-fenótipos da doença falciforme já realizados não adotaram estratégias para evitar a associação espúria causada pela estratificação entre casos e controles, relacionada à miscigenação. Alguns autores alegam que a estratificação devido à miscigenação não é um problema em estudos envolvendo a anemia falciforme nos Estados Unidos, país onde a maior parte dos estudos foi conduzida, já que esse grupo de pacientes não seria heterogêneo no que concerne à miscigenação (Adams *et al.*, 2003; Hoppe *et al.*, 2007). No entanto, estudos recentes demonstraram que os níveis de miscigenação do grupo conhecido como Afro-Americanos podem ser não só elevados, mas também bastante heterogêneos (Bryc *et al.*, 2010).

No Brasil, e mais especificamente no estado de Minas Gerais, não existem estudos acerca dos níveis de miscigenação dos portadores de doença falciforme. Entretanto, como se trata de um país altamente miscigenado (Callegari-Jacques *et al.*, 2003; Parra *et al.*, 2003; Santos *et al.*, 2009), o risco de associação espúria entre polimorfismos genéticos e manifestações clínicas devido à estratificação relacionada à miscigenação não pode ser negligenciado.

Desta forma, acreditamos que uma maior compreensão da estrutura genética da população de pacientes portadores de doenças falciformes de Minas Gerais contribuirá para o desenho de futuros estudos de epidemiologia genética neste grupo.

1.4.2 A doença Falciforme no Brasil e em Minas Gerais

O polimorfismo associado às doenças falciformes foi introduzido no Continente Americano por meio da imigração dos africanos, por ocasião do tráfico de escravos, uma vez que a HbS nunca foi relatada em populações ameríndias e sua contribuição direta por migrantes europeus pode ser considerada insignificante (Rogers *et al.*, 1989; Zago *et al.*, 1992).

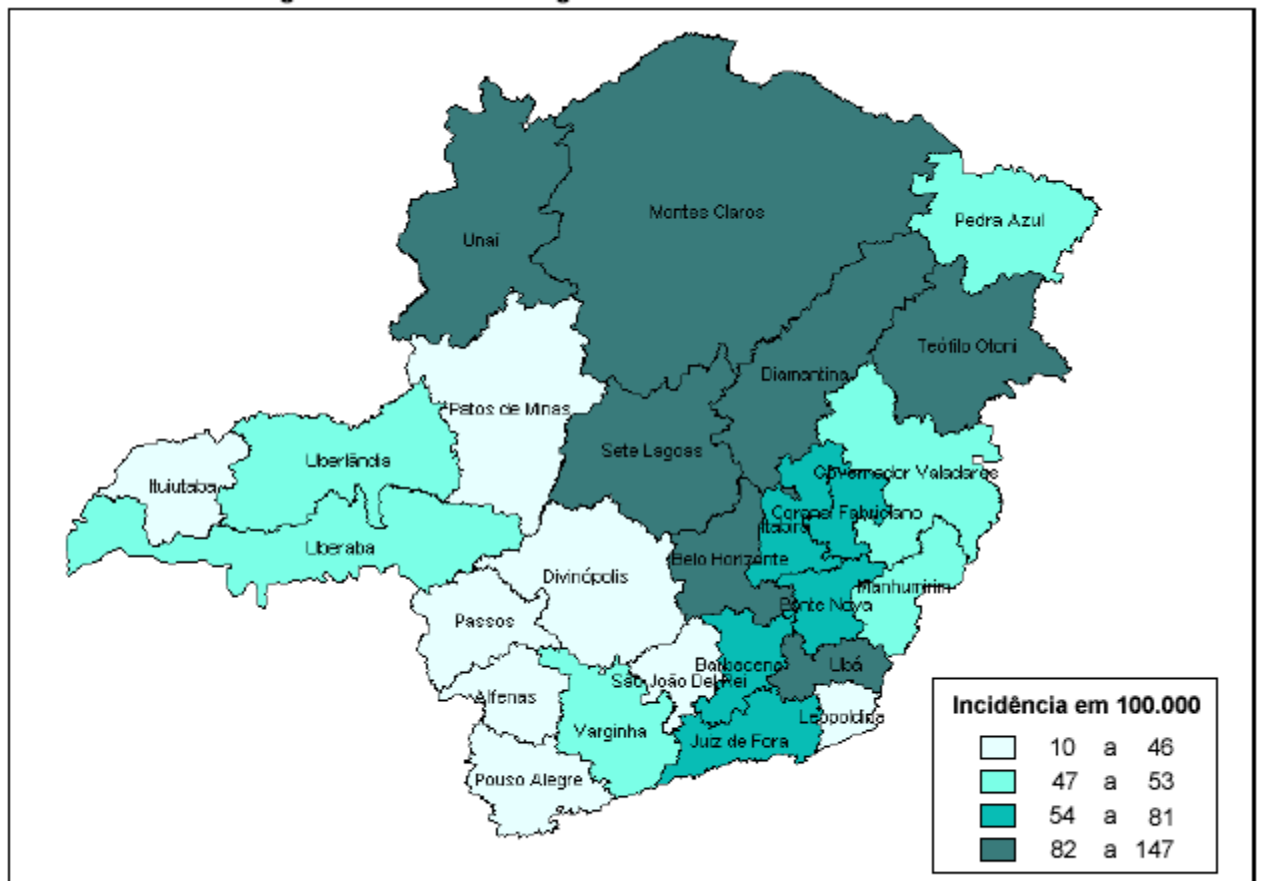
No Brasil, a anemia falciforme é a doença hereditária monogênica mais comum. O

Ministério da Saúde estima a prevalência de 25.000 a 30.000 indivíduos com anemia falciforme no país, e a incidência de 3.500 novos casos a cada ano. A distribuição da HbS no território brasileiro é heterogênea, sendo mais alta no nordeste do país. A prevalência de portadores (heterozigotos AS) é de 2% no sudeste do Brasil, mas pode subir para 6 a 10% entre indivíduos “afro-descendentes”. No estado de Minas Gerais estima-se o nascimento de uma criança com anemia falciforme para cada 2.500 recém-nascidos vivos (Januario, 2002). Estes números, em conjunto com as características clínicas da doença, fazem da anemia falciforme um problema de saúde pública, conforme explicitado pela própria Agência Nacional de Vigilância Sanitária (Anvisa 2002, Cançado e Jesus, 2007).

Dentre os haplótipos associados à HbS, o CAR é o mais frequente no Brasil, seguido pelos haplótipos Benin e Senegal (Pante de Souza *et al.*, 1998; Zago *et al.*, 1992, Zago e Pinto, 2007). Em Minas Gerais, um estudo realizado em crianças nascidas entre 1999 e 2006 e acompanhadas no Ambulatório do Hemocentro de Belo Horizonte da Fundação Hemominas demonstrou que os haplótipos mais freqüentes foram o CAR e o Benin, embora o haplotipo Árabe-Indiano também tenha sido observado. Entretanto, não foi encontrada associação entre os haplótipos e a ocorrência de eventos clínicos e hematológicos nessa população (Belisario, 2010).

A doença falciforme é uma das enfermidades diagnosticadas pelo Programa Nacional de Triagem Neonatal (PNTN), implantado no Brasil em 2001 pelo Sistema Único de Saúde (SUS). O objetivo deste programa é promover a detecção de doenças congênitas em fase pré-sintomática em todos os nascidos vivos, permitindo o tratamento precoce. Em Minas Gerais, o Programa Estadual de Triagem Neonatal (PETN) foi criado em 1993, tendo sido a triagem para a doença falciforme implantada em 1998, através do Núcleo de Ações e Pesquisa em Apoio Diagnóstico (NUPAD), órgão complementar da Faculdade de Medicina da Universidade Federal de Minas Gerais.

No Estado de Minas Gerais, o recém-nascido diagnosticado com doença falciforme é encaminhado para a **Fundação Hemominas**, que é o centro de referência responsável pelo seguimento clínico-hematológico desses pacientes, sendo este acompanhamento realizado nas unidades regionais da instituição, por equipes multi-disciplinares. A figura 2 apresenta a incidência da doença falciforme em Minas Gerais, conforme registros do Programa Estadual de Triagem Neonatal.



(*) Inclui fenótipos: Hb SS, Hb SC, Hb S/ β^0 tal, Hb S/ β^+ tal, Hb S/D_{Punjab}

Figura 2. Incidência da Doença Falciforme em nascidos vivos, por diretoria Regional de saúde – Programa Estadual de Triagem Neonatal (1998-2001)

Fonte: Januário (2002)

2 OBJETIVOS

2.1 Objetivo Geral

Identificar padrões geográficos de ancestralidade genômica Ameríndia, Européia e Africana em indivíduos supostamente saudáveis, doadores de sangue, e pacientes com doença falciforme do Estado de Minas Gerais.

2.2 Objetivos Específicos

2.2.1 Criar uma coleção de amostras de DNA de doadores de sangue, e pacientes com doença falciforme provenientes de diversas regiões de Minas Gerais.

2.2.2 Padronizar a genotipagem de painéis de marcadores informativos de ancestralidade (MIAs) européia, africana e ameríndia pelo método de minisequenciamento.

2.2.3 Genotipar os doadores de sangue e pacientes com doença falciforme para 54 marcadores moleculares informativos de ancestralidade (14 SNPs e 40 INDELS).

2.2.4 Estimar a miscigenação ameríndia, européia e africana nos níveis populacional e individual para os doadores de sangue e pacientes com doença falciforme.

2.2.5 Determinar, a partir dos marcadores genotipados, a estrutura genética das populações de pacientes com doença falciforme e de doadores de sangue em Minas Gerais em função dos níveis de miscigenação.

2.2.6 Testar a hipótese nula que os padrões de ancestralidade genômica em Minas Gerais determinam uma estrutura populacional que se adapta ao *modelo de isolamento por distância*.

2.2.7 Verificar a ocorrência de padrões geográficos de ancestralidade genômica para doadores de sangue e pacientes com doença falciforme.

3 MATERIAL E MÉTODOS

3.1 Populações e amostras

3.1.1 Populações parentais

Foram utilizados como representantes das populações parentais que deram origem à população brasileira três conjuntos de indivíduos: 31 europeus e 24 africanos componentes do painel do SNP500Cancer (http://snp500cancer.nci.nih.gov/home_1.cfm, Packer *et al.*, 2006), além de 85 Nativos americanos, provenientes do Peru e amostrados na região entre a encosta oriental dos Andes e a floresta Amazônica (região chamada Selva Alta). Parte desses ameríndios é da região de Cuzco e pertence às comunidades de Shimaa (n=30) e Monte Carmelo (n=15), do grupo linguístico Matsiguenga e parte deles (n = 40) reside nas aldeias Ashaninka situadas ao longo do rio Tambo (região de Junin).

3.1.2 Populações miscigenadas

As populações alvo do presente estudo são constituídas pelos doadores de sangue e pelos pacientes com doença falciforme do Estado de Minas Gerais, atendidos pela Fundação Hemominas.

A Hemominas é uma instituição pública que atua na organização do sistema Hematológico e Hemoterápico de Minas Gerais, atendendo mais de 87% da demanda transfusional do Estado. Além do atendimento aos doadores de sangue, a Hemominas também é responsável pelo atendimento de pacientes. Os Ambulatórios localizados nas diferentes Unidades da Hemominas são referência no Estado de Minas Gerais quanto ao atendimento dos pacientes com coagulopatias e hemoglobinopatias (incluindo a doença falciforme), realizando o diagnóstico, acompanhamento multidisciplinar, profilaxia e tratamento.

Utilizando o suporte logístico da Fundação Hemominas em todo o Estado, foi realizada a coleta de sangue de duas amostras populacionais: (a) doadores de sangue e (b)

pacientes com doença falciforme. A Caixa 1 define os critérios para selecionar as amostras (a) e (b). A **Figura 3** apresenta as unidades da Fundação Hemominas que participaram do processo de amostragem.

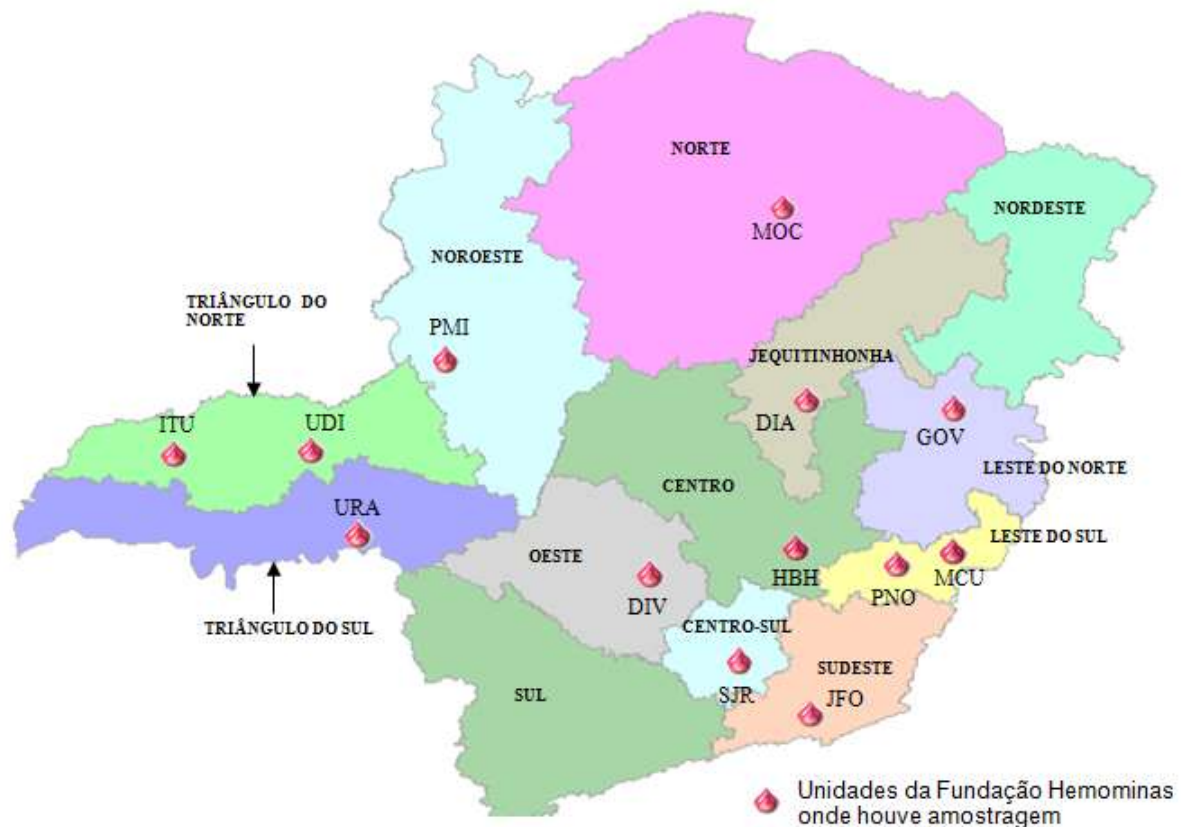


Figura 3: Distribuição geográfica das Unidades da Fundação Hemominas onde houve coleta de amostras. Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.

Em uma primeira etapa do presente estudo, iniciou-se a amostragem dos indivíduos com doença falciforme, e dada a distribuição geográfica destas amostras, prosseguiu-se a coleta de indivíduos sadios com distribuição geográfica similar, procurando garantir uma distribuição adequada das amostras no espaço geográfico.

Uma amostra de 5 ml de sangue de cada indivíduo foi coletada após a explicação dos objetivos do estudo e da obtenção da assinatura do termo de consentimento livre e

esclarecido. Para cada sujeito da pesquisa, foi registrado o município de nascimento e de residência, o sexo e a idade. O protocolo deste projeto de pesquisa tem aprovação dos Comitês de Ética em Pesquisa da UFMG e da Fundação Hemominas.

Para cada amostra, foi realizada a purificação do DNA genômico, utilizando o kit comercial *QiaAmp DNA Blood Mini Kit (Qiagen)*. O DNA extraído foi então quantificado em espectrofotômetro, diluído para uma concentração padronizada de 50 ng/μl e estocado em freezer -20°C.

CAIXA 1

Definição de doadores de sangue: Pessoas consideradas aptas na triagem clínica (aptos clínicos), e que chegam a colher bolsa de sangue, independentemente do posterior resultado na triagem sorológica, ou seja, serão incluídos aptos clínicos e aptos e/ou inaptos sorológicos. Os principais critérios para doação de sangue são:

- Ter idade entre 18 e 64 anos;
- Ter peso acima de 50 Kg;
- Não ter comportamento de risco para doenças sexualmente transmissíveis;
- Não ter tido hepatite após os 10 anos de idade;
- Não estar gripado, resfriado, com febre ou diarreia;
- Não ter ingerido bebida alcoólica nas últimas 12 horas;
- Não ser usuário de drogas;
- Não apresentar ferimento ainda não cicatrizado;
- Não ter sido submetido a exame de endoscopia ou broncoscopia nos últimos 12 meses;
- Não estar grávida ou em período de amamentação. Após o parto normal é necessário aguardar três meses e após cesárea, seis meses;
- Não ter feito tatuagem nos últimos 12 meses;
- O prazo mínimo entre uma doação de sangue total e outra é de 60 dias para os homens e de 90 dias para mulheres. Para doadores com idade entre 60 e 64 anos, o intervalo é de seis meses. Homens podem realizar até 4 quatro doações de sangue total em um período de 12 meses e as mulheres três doações.

Definição de pacientes: São indivíduos com doença falciforme, atendidos nos Ambulatórios da Fundação Hemominas e diagnosticados no Laboratório de Hematologia da Fundação, através dos seguintes exames laboratoriais:

- 1) Hemograma e análise morfológica de esfregaço de sangue em lâmina;
- 2) Teste de solubilização usando ditonito de sódio;
- 3) Eletroforese em pH alcalino.

Se o diagnóstico permanecer em dúvida, são feitos os seguintes testes complementares: teste de falcização (método de dissulfeto de sódio) e eletroforese em pH ácido. Serão incluídos pacientes com doença falciforme (presença do gene HbS em homozigose (SS) ou em associação com outra hemoglobina anormal, de qualquer sexo e idade.

3.2 Genotipagem

Após extração do DNA, foram genotipados 54 marcadores em cada um dos indivíduos amostrados, visando capturar a estrutura genética das populações em função dos níveis de miscigenação. Destes 54 marcadores, 14 são do tipo SNPs e tiveram a genotipagem

padronizada por nós, pelo método de mini-sequenciamento em duas reações e 40 são do tipo INDELs e tiveram a genotipagem padronizada por Bastos-Rodrigues *et al.* (2006) em quatro reações.

Os indivíduos das populações parentais (africanos, europeus e ameríndios), bem como 23 indivíduos miscigenados denominados Hispânicos, componentes do painel do projeto SNP500cancer também foram genotipados para os 54 marcadores acima relacionados, assim como para outros 10 marcadores do tipo INDELs. Os marcadores do painel de 10 INDELs tiveram sua genotipagem padronizada por Tulio Lins, do grupo do Prof. Rinaldo W. Pereira, da Universidade Católica de Brasília em uma única reação com o kit SNaPshot (Applied Biosystems), a qual foi modificada para que fosse possível sua realização em um sequenciador Megabace (*GE Healthcare*) (dados não publicados).

3.2.1 Construção do painel de 14 SNPs informativos de ancestralidade

Os SNPs que compõem os painéis padronizados por nosso grupo foram selecionados com base em sua informatividade de ancestralidade (índice I_a de Rosenberg *et al.* 2003) ou na diferença das frequências alélicas entre as populações parentais - δ , conforme detalhado na **Caixa 2**. Além disso, questões operacionais como a viabilidade de genotipagem em sistemas multiplex e flexibilidade para a utilização das plataformas de minisequenciamento monocromático (em sequenciador *Megabace*, *GE Healthcare*) e policromático (*SNaPshot* em sequenciador *3130 Applied Biosystems*) também foram levadas em consideração na elaboração dos painéis de MIAs.

As reações de minisequenciamento empregadas para a genotipagem dos SNPs utilizaram como molde os produtos de PCR das regiões flanqueadoras do SNP de interesse, co-amplificadas em reações múltiplas (*multiplex*). Após a amplificação das regiões flanqueadoras e visualização dos produtos amplificados em gel de poliacrilamida 15%, foi realizada uma purificação enzimática para degradação do excesso de iniciadores e de dNTPs (EXO-SAP). Posteriormente, a genotipagem dos SNPs foi realizada utilizando o mini-sequenciamento monocromático (Carvalho e Pena 2004, modificado) ou o sistema policromático (*SNaPshot™ Multiplex System*, *Applied Biosystems*). Finalmente, uma nova purificação foi realizada para remoção dos grupos fosforil das extremidades 5' dos ddNTPs fluorescentes, com a utilização da enzima SAP. A eletroforese capilar foi realizada em um

sequenciador *Megabace* (*GE Healthcare*) ou no sequenciador automático ABI 3100 (*Applied Biosystems*). A **Figura 4** traz um resumo das principais etapas envolvidas nas reações de mini-sequenciamento. Informações detalhadas do processo de genotipagem são apresentadas na **Caixas 3**.

A análise dos resultados foi realizada com o auxílio dos programas computacionais *Fragment Profiler* para o mini-sequenciamento monocromático ou *GeneMapper* para o mini-sequenciamento policromático.

Após a padronização da genotipagem dos SNPs, foram realizados testes de controle de qualidade visando assegurar a confiabilidade dos experimentos. Para isto, os marcadores foram genotipados em nove indivíduos (três africanos, três europeus e três ameríndios) por três vezes (três reações independentes de minisequenciamento monocromático). Um segundo controle consistiu na corrida (eletroforese capilar) dos produtos de minisequenciamentos em dois sequenciadores *Megabace* distintos (localizados no Laboratório de Biodiversidade e Evolução Molecular - LBEM e no Núcleo de Análise de Genoma e Expressão Gênica - NAGE), o que foi feito para duas reações de minisequenciamento.

A etapa seguinte de avaliação dos painéis padronizados consistiu na verificação da acurácia das estimativas de miscigenação fornecidas por esses painéis. Para isto, foi realizada a genotipagem dos 14 SNPs em indivíduos para os quais já havia resultados de genótipos individuais disponíveis em um banco de dados público para um número maior de MIAs. Estes experimentos foram conduzidos para que pudéssemos comparar os resultados de estimativa de miscigenação obtidos com o painel de 14 SNPs em relação a um painel maior, o qual assumimos prover estimativas mais precisas. Para isto foram utilizadas amostras de DNA de 23 indivíduos latino-americanos miscigenados identificados como Hispânicos no painel do SNP500cancer. Para esses indivíduos foram obtidos os dados de genotipagem individual para um painel de 108 SNPs MIAs não ligados que atendiam aos seguintes critérios: 1) $F_{ST} > 0.20$ entre populações africanas, européias e ameríndias, do painel do CEPH (Cann *et al.*, 2002), 2) $F_{ST} < 0.10$ entre populações dentro desses grupos.

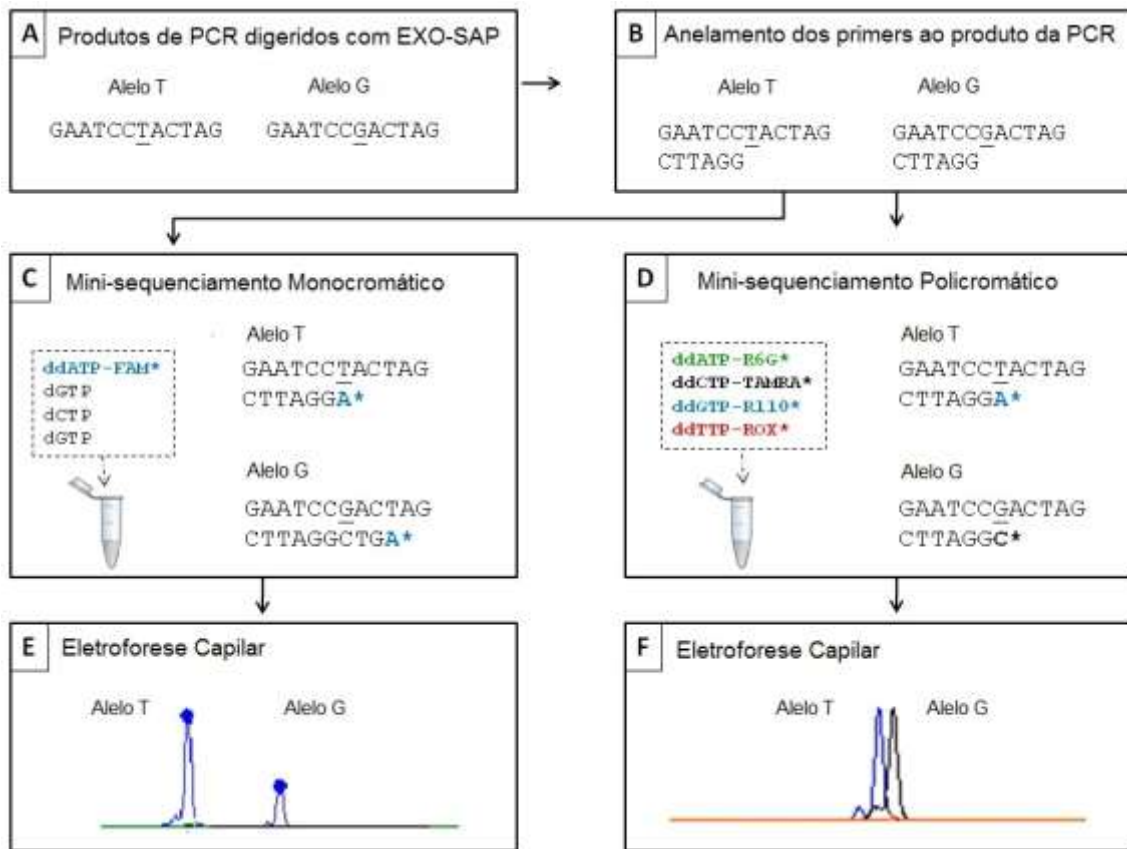


Figura 4: Minisequenciamento Monocromático e Policromático para um indivíduo heterozigoto (T/G). As etapas de amplificação por PCR da região flanqueadora do polimorfismo e purificação EXO-SAP são comuns às duas abordagens (A). Na etapa seguinte o iniciador de minisequenciamento anela-se ao produto da PCR em posição adjacente ao polimorfismo de interesse (B). No mini-sequenciamento Monocromático é utilizado apenas um ddNTP fluorescente (no caso, ddATP) juntamente com os outros três dNTPs não fluorescentes (dTTP, dCTP e dGTP) (C), o que faz que sejam produzidos fragmentos de dois tamanhos: quando na posição do polimorfismo temos um “T”, um ddATP é incorporado ao iniciador e a sequência é interrompida, sendo o iniciador alongado em apenas um nucleotídeo, mas quando temos um “G” na mesma posição, um dCTP é incorporado e a reação prossegue, incorporando dNTPs até que um ddATP seja adicionado (E). Nesse caso, os alelos são distinguidos por meio da diferença de tamanho dos produtos de mini-sequenciamento. No mini-sequenciamento policromático, a reação contém os quatro ddNTPs marcados com fluorescências distintas (D), o que faz com que o iniciador seja sempre alongado em apenas uma base e permite a distinção dos alelos por meio da visualização de picos de cores diferentes (a cor do pico depende do ddNTP incorporado) (F).

Caixa 2: Seleção e Genotipagem de 14 SNPs Marcadores Informativos de Ancestralidade (MIAs)

A escolha dos marcadores de ancestralidade europeia e africana a serem usados, baseou-se na **Informatividade de ancestralidade (Ia)** (Rosemberg *et al.* 2003), que quantifica a informação sobre a ancestralidade de um indivíduo fornecida pelo genótipo em um determinado *locus*. Com base nesta medida, Tian *et al.* (2006) selecionaram mais de 4000 marcadores informativos de ancestralidade Africana/Européia a partir de SNPs genotipados no projeto HapMap (www.hapmap.org). Esse conjunto de marcadores foi a base para a seleção dos MIAs a serem utilizados no presente estudo para estimar a miscigenação europeia-africana e, a partir deles, foram pré-selecionados os 250 SNPs mais informativos. Para a elaboração de um painel informativo para miscigenação nativo-americana, a seleção de MIAs foi baseada na diferença de frequência alélica (δ) entre Europeus, Africanos e Ameríndios, a partir de um conjunto de 150 SNPs obtidos a partir do banco de dados do *SNP500Cancer* (<http://snp500cancer.nci.nih.gov/>).

As seqüências flanqueadoras de todos os SNPs pré-selecionados foram obtidas no banco de dados *dbSNPs* (<http://www.ncbi.nlm.nih.gov/SNP/>). Estas seqüências foram submetidas *on-line* ao programa *Multiplex* (<http://genomics14.bu.edu:8080/MuPlex/MuPlex.html>, Rachlin *et al.* 2005), para o desenho de iniciadores visando a amplificação de fragmentos contendo os polimorfismos, a partir de reações de PCR multiplex. A partir de um conjunto de SNPs pré-selecionados, *Multiplex* identifica um grande número de pares de iniciadores candidatos viáveis para cada *locus* de interesse. Em seguida, o programa seleciona um par de iniciadores para cada locus e os distribui em conjuntos multiplex-compatíveis, cada um correspondendo a uma reação de PCR multiplex potencial. *Multiplex* permite impor condições como a quantidade desejada de loci por tubo multiplex, o intervalo de tamanho desejado dos produtos de PCR e a diferença entre eles, conteúdo de GC, variação máxima de temperatura de anelamento para os iniciadores contidos em um único tubo, etc.

A partir do resultado do *Multiplex*, foram eliminados os tubos que não alcançaram o nível de multiplexidade desejado (10 SNPs) e os que apresentavam marcadores ligados. Desta forma, foram selecionados os três potencialmente melhores tubos de PCR multiplex de MIAs Africana-Européia (AFR3 e AFR10) e Ameríndia (AMR4). Os iniciadores desenhados pelo *Multiplex* foram então submetidos à ferramenta *eletronic PCR* (<http://www.ncbi.nlm.nih.gov/sutils/e-pcr/reverse.cgi>) para análise de especificidade. Os amplicons obtidos com essas reações PCR multiplex são utilizados como molde para a genotipagem dos SNPs.

Na etapa seguinte foram desenhados os iniciadores de minisequenciamento, para a realização da genotipagem propriamente dita. Estes foram desenhados para se anelar à fita molde até um nucleotídeo antes do SNP alvo e para apresentar uma temperatura de anelamento próxima de 55°C. No desenho desses iniciadores procurou-se flexibilidade, para que os mesmos oligonucleotídeos pudessem ser usados na genotipagem pelos sistemas *policromático* (sequenciador *Applied Biosystem 3130*), e *monocromático* (sequenciador *Megabace*, GE Healthcare). Esta flexibilidade pode ser alcançada evitando-se polimorfismos A/T e G/C na mesma reação, já que não é possível realizar estas genotipagens simultaneamente utilizando o sistema monocromático.

No sistema *policromático* são utilizados unicamente ddNTPs fluorescentes, que se ligam na posição do polimorfismo, alongando o iniciador em apenas uma base. Nesta metodologia, na eletroforese capilar, iniciadores de tamanhos diferentes são utilizados para diferenciar os loci e fluorescências diferentes (cada ddNTP é marcado com uma fluorescência distinta) para diferenciar os alelos.

Diferentemente, o minisequenciamento monocromático (*Megabace*) utiliza apenas um tipo de fluorescência e o tamanho do fragmento é o critério para diferenciar tanto os loci quanto os alelos. Adquirindo unicamente um ddNTP (ddATP e ddGTP) marcados com a mesma fluorescência, é genotipar todos os SNPs possíveis. No minisequenciamento monocromático, ddATP e ddGTP são utilizados apenas um por vez (mas nunca no mesmo tubo de reação já que são marcados com a mesma fluorescência), juntamente com os 3 dNTPs restantes. Sendo assim, considerando, por exemplo, um polimorfismo T/C, caso se esteja utilizando ddATP fluorescente na reação e o alelo em questão for um T, o iniciador é alongado em apenas um nucleotídeo, mas se houver um C é adicionado um dGTP e a reação de sequenciamento prosseguirá até o próximo T, gerando fragmentos de tamanhos diferentes para cada alelo de um marcador.

Para evitar sobreposição no tamanho dos produtos finais, diferenciando assim os loci a serem genotipados em uma mesma reação, a diferença mínima de tamanho entre os iniciadores internos foi de 5bp, o que se conseguiu adicionando caudas com a seqüência do plasmídeo PUC18 na extremidade 5' dos iniciadores.

Caixa 3. Detalhes Metodológicos das Reações de Genotipagem

1. Reações de PCR multiplex para co-amplificação das regiões flanqueadoras dos MIAs

As reações de PCR foram padronizadas visando a co-amplificação das regiões flanqueadoras de 10 marcadores. O programa de termociclagem consistiu de: fase inicial de 95°C por 5 min, seguida por 30 ciclos de 94°C por 30 s, 57°C por 90 s, 72°C por 90 s com uma fase final de 72°C por 10 min. O mesmo protocolo aplica-se aos três painéis (AMR4, AFR3 e AFR10).

Reação utilizando o Quiagen Multiplex kit (Qiagen): Reação com volume final de 25 µL contendo 100 ng de DNA molde, 0,2 µM de cada iniciador, Qiagen Multiplex PCR Master Mix 1X (o reagente “2X Qiagen Multiplex PCR Master Mix” contém concentrações pré-otimizadas de HotStarTaq DNA Polymerase, um PCR Buffer contendo 6 mM MgCl₂ e pH 8,7 a 20°C, além de dNTPs).

Reação utilizando STR Buffer (Promega): Reação com volume final de 25 µL contendo 100 ng de DNA molde, 0,2 µM de cada iniciador, 1,5 U de Platinum Taq DNA Polimerase (Invitrogen) e STR Buffer 1X (o reagente “10X STR Buffer” contém 500mM KCl, 100mM Tris-HCl pH 9,0, 15mM MgCl₂, 1% Triton® X-100, 2mM de cada dNTP).

2. Purificação enzimática dos produtos de PCR

A purificação de 3,0 µL dos produtos de PCR foi realizada numa reação contendo 1 U da enzima exonuclease 1 (EXO1, 10 U/µL), 0,90 U de fosfatase alcalina de camarão (SAP, 1 U/µL) e 0,2 µL de tampão 10X da enzima SAP. As reações foram incubadas por 90 min a 37°C e por 20 min a 80°C para desnaturação das enzimas.

3. Reação de Genotipagem

3.1 Genotipagem Monocromática: Foi realizada por meio de reação com volume final de 13,5 µL contendo 2,0 µL do produto de PCR purificado, 0,37 µM de cada iniciador, 0,46 µM de ddNTP (ddA ou ddG) marcado com a fluoresceína (Perkin Elmer Life Sciences), 0,46 µM dos 3 dNTPs restantes (GE Healthcare), 1 unidade de enzima Thermo Sequenase DNA Polymerase (GE Healthcare), 1,3 µL de Tampão da enzima e H₂O Milli-Q qsp. A termociclagem adotou o seguinte protocolo: 5 min de desnaturação a 80°C, seguido de 30 ciclos de 95°C por 30 s, anelamento a 55°C durante 30 s e extensão a 72°C durante 20 s. Esta termociclagem foi realizada como “hot start”, sendo que a enzima Thermo Sequenase DNA Polymerase foi adicionada 2 min após a desnaturação inicial.

3.2 SnaPshot Multiplex System: Foi realizada por meio de uma reação com volume final de 5,0 µL contendo 1,0 µL do produto de PCR purificado, 1 µL de SNaPshot™ Kit Reaction Mix, 2 µL de MIX de iniciadores (em concentrações de 1µM cada) e H₂O Milli-Q qsp. A termociclagem foi feita adotando-se o protocolo: 2 min de desnaturação a 96°C, seguido de 25 ciclos de 96°C por 10 seg, anelamento a 55°C durante 5 seg e extensão a 60°C durante 30 seg.

4. Purificação enzimática dos produtos de minisequenciamento

Para os produtos do SnaPshot esta purificação foi feita com 0,5U de SAP e 0,5 µL de tampão 10X da SAP adicionados a diretamente a 5µL de produto SNaPshot. No caso do minisequenciamento monocromático utilizou-se 0,3 U de SAP, 0,2 µL de tampão 10X da SAP e 1,6 µL de H₂O Milli-Q adicionados a 5 µL do produto de minisequenciamento. Em ambos os casos as reações foram incubadas durante 60 min a 37°C e 15 min a 75°C para desnaturação da enzima.

5. Eletroforese capilar

A eletroforese dos produtos de SnaPshot purificados foi realizada no sequenciador ABI 3100 (Applied Biosystems) com 1,0 µL do amplicon final, 8,9 µL de formamida Hi-Di e 0,1 µL de Liz120 Size Standard (Applied Biosystems). Os parâmetros críticos de corrida foram: Capilar de 36 cm, voltagem da injeção de 1,2 Kv, tempo de injeção de 18 s, voltagem da corrida de 15 Kv e tempo de corrida de 800 s. Os resultados foram analisados com os programas GeneScan Analysis 3.7 e Genotyper 3.7 (Applied Biosystems). Os produtos do minisequenciamento monocromático foram diluídos na proporção 1:2 em água Mili-Q e submetidos à eletroforese em um sequenciador Megabace (GE Healthcare) com 7,75 µL de Tween 20 a 0,1% e 0,25 µL de ET-ROX 550 (GE Healthcare). Os parâmetros críticos da corrida foram: voltagem de injeção de 3 Kv, tempo de injeção de 80 s, voltagem da corrida de 10 Kv e tempo de corrida de 75 min, e filtros de genotipagem Dye Set II.

3.2.2 Painéis de 40 marcadores de inserção-deleção (Bastos-Rodrigues *et al.*, 2006)

Os 40 marcadores bialélicos do tipo inserção-deleção (INDELs) cuja fenotipagem foi padronizada por Bastos-Rodrigues *et al.* (2006) foram selecionados a partir de um conjunto maior de INDELs descrito por Weber *et al.* (2002) e disponíveis em <http://research.marshfieldclinic.org/genetics/indels/default.asp>. A amplificação dos fragmentos de DNA correspondente aos marcadores consistiu de quatro reações de PCR multiplex contendo uma combinação de 10 pares de iniciadores. Para cada iniciador direto de PCR (“*primer forward*”) uma cauda de M13-40 de 17 mer foi adicionada. As PCRs multiplex foram realizadas em um volume final de 10 μl contendo tampão de PCR 1X (10 mM Tris-HCl pH 8.3 ou pH 9, 2,75 mM KCl e 3,5 mM MgCl_2), 200 μM de dNTPs, 1,0 U de Platinum *Taq* DNA polymerase (Invitrogen), 20 ng de DNA genômico, 1,5 μM de iniciador direto M13-40 marcados com a fluorescência FAM, 1,5 μM de cada iniciador reverso não fluorescente, e 0,1 μM de cada iniciador direto não fluorescente. 2 μl dos produtos de PCR diluídos na proporção 1:3 em água Mili-Q foram submetidos à eletroforese capilar em um sequenciador Megabace (GE Healthcare) com 7,75 μL de Tween 20 a 0,1% e 0,25 μL de ET-ROX 550 (GE Healthcare).

Esses 40 INDELs, além de serem capazes de caracterizar a estrutura de populações humanas no nível global (Bastos-Rodrigues *et al.*, 2006), também já foram utilizados em estimativas de miscigenação na população brasileira (Pena *et al.*, 2009).

3.2.3 Paineis de 10 marcadores de inserção-deleção

Os 10 marcadores bialélicos do tipo INDELs cuja fenotipagem foi padronizada por Tulio Cesar de Lima Lins, do grupo do Prof. Rinaldo W. Pereira (Universidade Católica de Brasília-UCB) também foram selecionados a partir do conjunto descrito por Weber *et al.* (2002).

A genotipagem desses marcadores foi realizada por meio de uma reação de PCR com um volume final de 25 μL , contendo 10 ng de DNA genômico, tampão de PCR 1X (10 mM Tris-HCl pH 8.3 e 2,75 mM KCl), 1,0 U de Platinum *Taq* DNA polimerase, 0,25 mM de dNTPs, 0,2 mg/mL de BSA (albumina sérica bovina), MgCl_2 4 mM e 0,5 μL de um mix de

iniciadores em concentrações descritas na **tabela S1**. 2 µL dos produtos de PCR diluídos na proporção 1:10 em água Mili-Q foram submetidos à eletroforese em um sequenciador Megabace (GE Healthcare) com 7,75 µL de Tween 20 a 0,1% e 0,25 µL de ET-ROX 550 (GE Healthcare).

Estes marcadores são especialmente informativos para discriminar a miscigenação ameríndia, uma vez que possuem frequências diferenciadas nessa população, em relação aos africanos e europeus. Informações adicionais a respeito desses painéis estão na **tabela S1**.

3.3. Análises estatísticas

3.3.1 Equilíbrio de Hardy-Weinberg (EHW)

Os cálculos das frequências alélicas foram realizados a partir dos dados genotípicos gerados, utilizando o programa Convert (Glaubitz, 2004), obtido no endereço eletrônico <http://www.agriculture.purdue.edu/fnr/html/faculty/Rhodes/Students%20and%20Staff/glaubitz/software.htm>. As frequências genotípicas foram testadas quanto ao equilíbrio de Hardy-Weinberg, por meio do teste exato de Fisher, implementado no programa GDA (Lewis e Zaykin, 2001), obtido a partir do endereço eletrônico <http://lewis.eeb.uconn.edu/lewishome/software.html>. A princípio, foram considerados como significativos (loci fora do EHW), os resultados associados a valores p menores que 0,05.

A correção de Bonferroni foi realizada em seguida, para avaliar a real significância dos resultados obtidos. Esta correção foi realizada com o intuito de evitar que locus sejam indevidamente considerados como estando fora do equilíbrio, devido ao efeito das comparações múltiplas. A correção de Bonferroni é um procedimento estatístico que altera o nível de significância (p valor) a fim de evitar este tipo de erro.

3.3.2 Estrutura Genética

Os estudos de genética de populações geralmente se concentram em dois aspectos fundamentais: a descrição da estrutura genética das populações e o estudo dos fatores evolutivos que moldam esta estrutura.

Segundo Templeton (2006), a estrutura genética pode ser conceituada como a quantidade de variabilidade genética e sua distribuição dentro e entre populações e indivíduos, dentro de uma espécie. Esta estrutura é resultante da ação de três fatores principais: sistema de acasalamento, deriva genética e fluxo gênico.

Estudos da estrutura genética de populações são possibilitados pela aplicação de métodos estatísticos a dados de polimorfismos genéticos levantados a partir de uma amostra de indivíduos que compõem uma dada população. Este tipo de análise permite investigar a existência de variação genética dentro e entre populações.

Freqüentemente, a estrutura genética das populações é estudada a partir de uma abordagem hierárquica (F_{IT} , F_{IS} e F_{ST}) (Wright, 1951). Estas abordagens quantitativas são conhecidas como “Estatísticas F de Wright” e podem ser empregadas na avaliação do nível de heterozigosidade de uma população e da ocorrência de uma eventual redução desta heterozigosidade, em comparação àquela esperada, segundo o equilíbrio de Hardy-Weinberg (Hamilton, 2009). Assim sendo, esta abordagem pode ser usada para quantificar as diferenças genéticas entre subgrupos (F_{ST}) a partir do estudo dos níveis de heterozigosidade.

Existe uma extensa discussão na literatura não apenas em relação à interpretação dos valores de F_{ST} , mas sobretudo sobre a metodologia usada para estimá-lo (Weir e Cockerham, 1984, Templeton, 2006). Diversos são os estimadores desenvolvidos com o intuito de avaliar a estrutura genética das populações.

A partição da variação genética em níveis hierárquicos (indivíduo, sub-populações, meta-população) permite a aplicação de uma abordagem de análise de variância no estudo da estrutura genética, na qual se deseja saber quanto da variação genética total encontra-se em cada nível hierárquico. Neste contexto, estimador θ , um análogo do F_{ST} proposto por Weir e Cockerham (1984) mede a quantidade de variação genética entre sub-populações.

Embora o F_{ST} clássico seja amplamente empregado para avaliar a divergência genética entre sub-populações, ele não leva em conta os erros de amostragem. Por outro lado, o estimador θ , leva em consideração este tipo de erro, sendo, portanto, considerado um avanço

metodológico (Weicker *et al.* 2001). Simulações mostram que o θ é independente do número de grupos amostrados, bem como do número de indivíduos amostrados em cada grupo (Weir e Cockerham 1984; Cockerham e Weir 1993). Esta característica, aliada ao fato de que este método foi implementado em um programa computacional de fácil acesso (GDA) faz com que este estimador seja amplamente utilizado em estudos de genética de populações.

No presente estudo, os valores de F_{st} (θ) par a par foram calculados para as diferentes populações de doadores e pacientes, e uma matriz de distâncias genéticas foi computada por meio do programas GDA. Para isto, cada local de coleta foi considerado como sendo uma população distinta. Doadores de sangue e pacientes com doença falciforme foram analisados separadamente.

As matrizes de F_{st} obtidas foram então analisadas por meio da ferramenta estatística “**Multidimensional scaling**” (MDS), que permite a visualização gráfica de uma matriz de distâncias entre pares de objetos, permitindo desta forma que as populações fossem exploradas quanto a similaridades e dissimilaridades genéticas. O MDS é um método capaz de transformar uma matriz de similaridades/dissimilaridades em distâncias representadas em um espaço Euclidiano n-dimensional (Kruskal 1964, Tarazona-Santos *et al.* 2001).

O MDS foi realizado por meio pacote “MASS package”, desenvolvido em linguagem R, obtido a partir do endereço eletrônico <http://cran.r-project.org/>.

3.3.3 Isolamento por Distância (IBD)

O teste para o modelo de isolamento por distância (IBD - Wright, 1943) foi empregado, visando a investigação de um eventual decréscimo da similaridade genética entre populações à medida que as distâncias geográficas entre elas aumentam. O modelo IBD descreve uma dinâmica de fluxo gênico na qual a probabilidade de migração diminui com o aumento da distância geográfica entre populações.

O teste para o modelo IBD foi realizado por meio do “*Isolation by Distance Web Service*” (IBDWS), disponível no endereço eletrônico <http://ibdws.sdsu.edu/~ibdws/> (Bohonak, 2002; Jensen *et al.*, 2005).

O IBDWS foi utilizado para avaliar se o padrão de variabilidade genética de doadores de sangue e pacientes com doença falciforme amostrados nas diferentes localidades se adapta ao modelo de IBD por meio do teste de Mantel, com 1000 permutações, a partir da análise de

matrizes de distâncias genéticas (F_{st} par a par) e distâncias geográficas (distância linear entre os locais de amostragem).

3.3.4. Estimativas de miscigenação no nível populacional

A miscigenação populacional foi estimada utilizando-se dois métodos: 1) método de Identidade Genética desenvolvido por Chakraborty (1985) e implementado no programa ADMIX 95 (desenvolvido por Bernardo Bertoni e disponível no endereço eletrônico <http://www.genetica.fmed.edu.uy/software.htm> e 2) método baseado nos Tempos de Coalescência desenvolvido por Dupanloup e Bertorelle (2001) e implementado no programa ADMIX2.

Uma ferramenta que facilitou imensamente essas análises foi o programa Adfit (Gourjon e Degioanni, 2009), o qual é capaz de formatar os arquivos de entrada exigidos para diferentes programas computacionais para estimativa de miscigenação. O uso do Adfit representou uma grande economia de tempo e esforço, além de prevenir erros que podem ser eventualmente introduzidos ao se formatar este tipo de arquivo manualmente.

3.3.5 Estimativas de miscigenação no nível individual

A miscigenação individual foi estimada utilizando o método de agrupamento bayesiano desenvolvido por Pritchard (2000) e implementado no software STRUCTURE v2.2.3 (Hubisz *et al.*, 2009; Pritchard *et al.*, 2000). As análises adotaram os seguintes parâmetros: $K=3$, “burn-in Period” = 100 000, repetições de MCMC após “burning”= 100 000, informação a priori para os indivíduos das populações parentais para auxiliar no agrupamento (USEPOPINFO=1), modelo = ADMIXTURE para os indivíduos miscigenados, parâmetro alfa inferido para cada população, GENSBACK = 2 e MIGRPRIOR = 0,05, e frequências alélicas correlacionadas. Em alguns casos específicos, análises sem informação a priori sobre a origem dos indivíduos das populações parentais também foram realizadas.

3.3.6. Identificação de padrões geográficos de miscigenação

Para a investigação de padrões geográficos de miscigenação, cada ponto de coleta foi associado com suas localizações geográficas (latitude e longitude), bem como com os componentes de miscigenação Nativo Americana, Européia e Africana. Gráficos representando as estimativa de miscigenação populacional foram plotados no mapa do Estado de Minas Gerais para facilitar a busca de possíveis padrões. As estimativas de miscigenação individual de doadores de sangue e pacientes com doença falciforme também foram avaliadas quanto à correlação em relação às coordenadas geográficas (latitude ou longitude) do local onde as amostras foram coletadas.

4 RESULTADOS

4.1. Coleta das amostras

Após aprovação do projeto nos Comitês de Ética da UFMG e da Fundação Hemominas (**anexo 1**), foi solicitada a colaboração das diferentes Unidades da Hemominas, distribuídas em todo o Estado, para que as mesmas contribuíssem no processo de amostragem. Um conjunto de instruções padronizadas para a coleta das amostras foi elaborado para que todas elas fossem coletadas sob os mesmos critérios. Tanto doadores quanto pacientes amostrados são não aparentados e nascidos no estado de Minas Gerais (**anexo 2**).

A coleta das amostras de pacientes iniciou em novembro do ano 2006. Ao fim do processo de amostragem, foram obtidas 225 amostras de pacientes portadores de doença falciforme, das quais 200 foram utilizadas para as análises genéticas apresentadas no presente estudo. A coleta das amostras de doadores foi iniciada posteriormente para obtermos uma distribuição similar à da dos pacientes, tendo sido iniciada em janeiro de 2008. Um total de 336 amostras de doadores de sangue foram coletadas, das quais 291 foram utilizadas no presente estudo.

A **Tabela 1** evidencia a distribuição geográfica das amostras obtidas. Como é possível observar, tanto as amostras de pacientes quanto as de doadores, apresentam uma ampla distribuição, ao longo do território de Minas Gerais.

Tabela 1. Amostragem de doadores de sangue e pacientes com doença falciforme de Minas Gerais. Os números em parêntese indicam os indivíduos considerados nas análises genéticas e de miscigenação.

Cidade de Coleta	Número de Indivíduos
Doadores	
Uberlândia	19 (13)
Uberaba	30 (26)
São João del Rei	18 (15)
Ponte Nova	29 (25)
Patos de Minas	24 (21)
Montes Claros	29 (24)
Manhuaçu	32 (30)
Juiz de Fora	36 (30)
Ituitaba	28 (23)
Governador Valadares	7 (7)
Divinópolis	25 (22)
Diamantina	30 (24)
Belo Horizonte	31 (31)
Pacientes	
Uberlândia	22 (22)
Uberaba	20 (16)
São João del Rei	3 (3)
Patos de Minas	38 (32)
Montes Claros	23 (23)
Manhuaçu	8 (8)
Juiz de Fora	14 (11)
Ituitaba	7 (7)
Governador Valadares	30 (26)
Divinópolis	27 (21)
Diamantina	1 (1)
Belo Horizonte	31 (28)

Para todos os sujeitos da pesquisa, foi coletada uma amostra de 5 ml de sangue, após a concordância do sujeito de pesquisa e obtenção da assinatura do termo de consentimento livre e esclarecido (**anexos 3 e 4**). No caso dos pacientes menores de 18 anos, um termo específico foi assinado pelo responsável (**anexo 5**). Para cada amostra foi realizada a extração do DNA utilizando o kit *QiaAmp DNA Blood Mini Kit (Qiagen)*, conforme orientação do fabricante. O DNA extraído foi então quantificado em espectrofotômetro, diluído para uma concentração padronizada de 50 ng/μl e estocado em freezer -20°C. Alíquotas de sangue total foram preservadas em freezer a -80°C para a eventual necessidade de uma nova extração de DNA. Nos casos em que os participantes autorizaram, o DNA será conservado após o término desse projeto, o que possibilitará sua utilização em futuros estudos, após devida aprovação pelos

Comitês de Ética em Pesquisa.

Para alguns pacientes com doença falciforme, já existe parecer favorável emitido pelo Comitê Nacional de Ética em Pesquisa (CONEP), que autoriza a genotipagem de 1.000.000 de SNPs, a ser realizada nos Estados Unidos (**anexo 6**). Esta autorização abre perspectiva para a continuidade e aprofundamento dos estudos de ancestralidade genômica e epidemiologia genética nessas populações.

4.2. Seleção de MIAs para desenvolvimento de painéis de genotipagem

Para construir painéis de marcadores informativos de ancestralidade, partindo de um subconjunto de 250 marcadores altamente informativos (Tian *et al.*, 2006) foram selecionados **20 MIAs (SNPs)** não ligados (caixa 2). Estes marcadores foram organizados em dois painéis multiplex denominados AFR3 e AFR10 (informativos para estimar ancestralidade africana e europeia), cada um contendo 10 MIAs. A seleção dos marcadores baseou-se no índice Ia (informatividade de ancestralidade, Rosenberg *et al.*, 2003) e o agrupamento dos marcadores em cada painel baseou-se na viabilidade de genotipagem simultânea em reações multiplex pelo método de minisequenciamento.

Um terceiro painel, AMR4 (informativo para ancestralidade Ameríndia), também composto por 10 marcadores, foi selecionado, usando como base um conjunto de 150 SNPs informativos. Para este painel, a diferença mínima de frequências alélicas entre ameríndios e europeus ou africanos demandada foi de 0,2.

Os marcadores selecionados, bem como os correspondentes alelos e cromossomos onde estão localizados são apresentados na **tabela 2**.

Tabela 2. Marcadores selecionados para a construção dos painéis de SNPs informativos de ancestralidade

Marcador	Cromossomo	Alelo 1	Alelo 2
Painel AFR3			
rs7512316	1	C	A
rs1372115	2	T	C
rs2697520	3	T	G
rs1443985	5	T	C
rs2789823	9	G	A
rs2416791	12	G	A
rs735480	15	T	C
rs8035530	15	T	C
rs6046024	20	T	C
rs9626698	22	G	A
Painel AFR10			
rs6698938	1	G	A
rs768392	1	G	C
rs7575147	2	T	C
rs2697520	3	T	G
rs10254729	7	T	G
rs12003714	9	G	A
rs527709	11	G	A
rs4891825	18	G	A
rs6510895	19	T	C
rs11698339	20	G	A
Painel AM4			
rs800292	1	G	A
rs2088102	1	C	T
rs691968	2	A	C
rs2569190	5	C	T
rs2234636	14	C	T
rs2619681	15	C	T
rs2518967	15	A	G
rs700518	15	A	G
rs8058694	16	C	A
rs3760657	19	T	C

4.3. Genotipagem

4.3.1 Padronização das Reações de PCR

Inicialmente, as reações de PCR para cada marcador (30 reações) foram testadas individualmente, para verificar a qualidade dos iniciadores e para a obtenção dos produtos de PCR necessários para padronização da genotipagem por mini-sequenciamento. Os produtos amplificados individualmente foram visualizados em gel de agarose 2%. Dos 30 loci avaliados, apenas três não apresentaram resultados satisfatórios nesta etapa (ausência de amplificação da banda específica): rs7512316 (Painel AFR3), rs2518967 e rs2619681 (Painel AMR4).

As reações de PCR Multiplex foram padronizadas em seguida, utilizando-se o *Quiagen Multiplex kit (Qiagen)* ou, como alternativa, o tampão *STR Buffer (Promega)* usado em conjunto com a *Platinum Taq DNA Polymerase (Invitrogen)*, visando a co-amplificação de 10 fragmentos em uma única reação. Os produtos amplificados foram visualizados em gel de poliacrilamida 15% corado com brometo de etídeo (**Figura 5**). As duas metodologias empregadas para a co-amplificação dos marcadores apresentaram resultados semelhantes, o que pôde ser constatado tanto na eletroforese quanto nos resultados das reações de mini-sequenciamento.

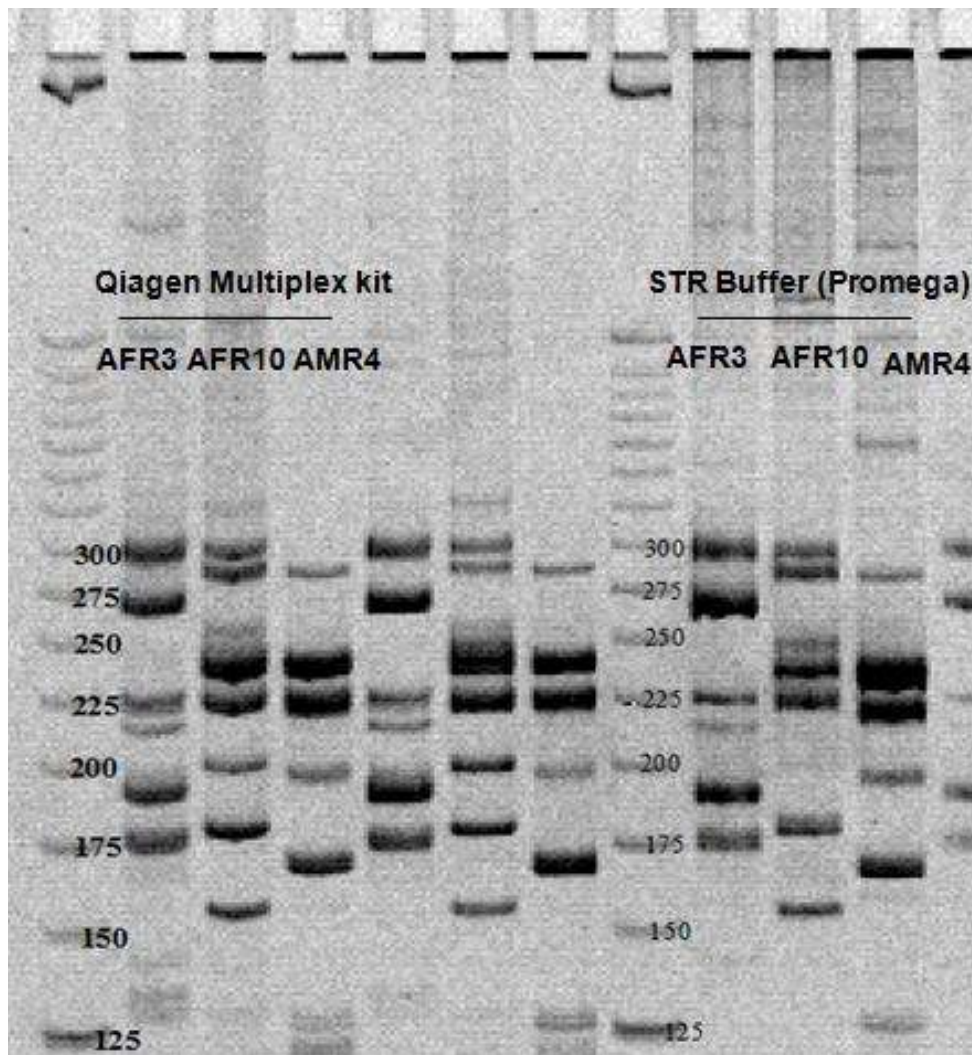


Figura 5: Reações de PCR multiplex para os painéis AFR3, AFR10 e AMR4, utilizando-se o *Qiagen Multiplex kit* e o *Tampão STR (Promega)*. Gel de poliacrilamida 15% corado com brometo de etídeo, sob luz UV. Padrão de peso molecular 25pb *DNA ladder (Invitrogen)*.

4.3.2 Padronização da genotipagem de SNPs pela técnica de minisequenciamento

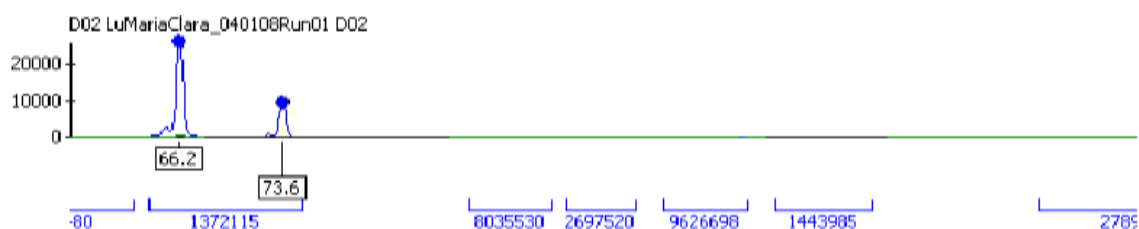
A padronização das reações de minisequenciamento monocromático em um sequenciador Megabace (GE Healthcare) foi realizada utilizando-se como molde para a genotipagem, os produtos de PCR amplificados a partir de um pool do DNA extraído de 10 indivíduos. O uso de um pool de DNA permite alcançar a heteroziguidade do maior número possível de loci, maximizando assim a possibilidade de visualização dos dois alelos de cada

marcador. A visualização dos alelos na etapa de padronização é importante por permitir a observação do tamanho dos fragmentos na eletroforese capilar, uma vez que esse tamanho pode divergir ligeiramente do esperado.

Inicialmente, a padronização da genotipagem de 26 marcadores foi realizada com sucesso, em três reações multiplex, sendo que apenas quatro marcadores (rs768392 do painel AFR10, rs2789823 e rs6046024 do painel AFR3 e rs700518 do painel AMR4) não puderam ser genotipados nem individualmente e nem em reações multiplex.

Foram conduzidos também experimentos para avaliar a flexibilidade de genotipagem em diferentes plataformas (Mini-sequenciamento monocromático e policromático - *SNaPshot*) utilizando os mesmos iniciadores. Um exemplo desta flexibilidade é apresentado na **Figura 6**, que mostra os eletroferogramas obtidos nas duas plataformas, a partir da genotipagem do SNP rs1372115, do painel AFR3 em um indivíduo heterozigoto. O iniciador interno, antes da reação de extensão apresentava 65 nucleotídeos. Após a reação de minisequenciamento monocromático (**Figura 6-a**), é possível observar a presença de um pico de 66 nucleotídeos (iniciador alongado apenas com um ddATP fluorescente) e um segundo pico de 73 nucleotídeos (a reação de minisequenciamento prosseguiu inserindo dNTPs até que um ddATP fosse incorporado na sequência provocando sua interrupção). Esses picos correspondem aos dois alelos do marcador genotipado, que podem ser distinguidos por meio da diferença de tamanho, e apresentam a mesma cor, pois apenas um tipo de fluorescência é utilizado (ddATP marcado com fluoresceína). Já no *SNaPshot* (**Figura 6-b**), observa-se dois picos praticamente do mesmo tamanho, porém de cores diferentes. Neste caso para ambos os alelos o iniciador foi alongado em apenas uma base e, como a reação conta com quatro ddNTPs marcados com fluorescências distintas, os alelos podem ser distinguidos por meio da cor dos picos. Os experimentos de genotipagem na plataforma *SNaPshot* foram conduzidos na Universidade Católica de Brasília, com a colaboração dos pesquisadores Ms.Tulio César de Lima Lins e Dr. Rinaldo Wellerson Pereira.

a) Minisequenciamento monocromático em um sequenciador Megabace



b) Minisequenciamento policromático em um sequenciador ABI 3130 (kit Snapshot. Applied Biosystems)

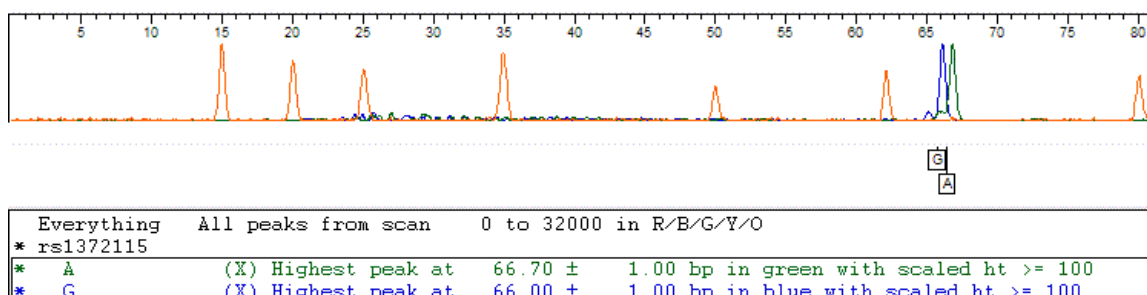


Figura 6: Flexibilidade de genotipagem por minisequenciamento monocromático em um sequenciador *Megabace* (a), e por *Snapshot* em um sequenciador *Applied Biosystems 3130* (b), evidenciada pelos eletroferogramas da genotipagem do SNP rs1372115, do painel AFR3 em um indivíduo heterozigoto, nas duas plataformas. O iniciador de interno, antes da reação de extensão do iniciador apresentava 65 nucleotídeos.

Embora a genotipagem por minisequenciamento monocromático em multiplex tenha sido a princípio padronizada para 26 marcadores, posteriormente foram eliminados três marcadores (rs7512316 do painel AFR3; rs2569190 e rs2518967 do painel AMR4) os quais apresentavam muitas falhas de genotipagem (*missing data*). Além disso, no decorrer dos experimentos, optou-se por eliminar todo o painel AFR10 devido a inconsistências observadas nos resultados (falhas de genotipagem e picos inespecíficos que geravam dúvidas na genotipagem), restando assim um total de 14 SNPs.

Os tamanhos esperados e observados para os picos correspondentes aos alelos de cada locus genotipado por mini-sequenciamento monocromático são apresentados na **Tabela 3**. As sequências dos iniciadores que compõem o painel de 14 SNPs, cuja genotipagem foi padronizada com sucesso, são apresentadas na **tabela 4**. A **figura 7** apresenta os

eletroferogramas de genotipagem por minisequenciamento monocromático dos SNPs cuja genotipagem foi padronizada e validada.

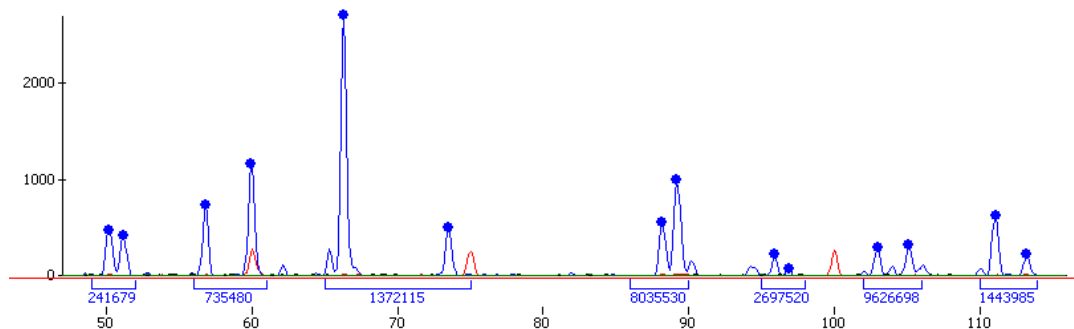
Tabela 3. Tamanhos esperados e observados dos picos produzidos pela genotipagem por minisequenciamento monocromático para o painel de 14 SNPs informativos de ancestralidade

PAINEL	SNP	Tamanho esperado (pb)	Tamanho observado (pb)
AFR3	rs2697520	96/97	96,0/97,0
	rs8035530	89/90	88,2/89,2
	rs1372115	66/73	66,2/73,6
	rs241679	49/50	50,0/51,0
	rs9626698	103/105	103,0/104,0
	rs1443985	111/113	111,2/113,5
	rs735480	56/60	56,6/59,5
AMR4	rs8058694	63/64	62,9/64,0
	rs691968	89/91	89,1/90,1
	rs2234636	57/58	56,6/57,8
	rs3760657	48/51	49,2/52,0
	rs2619681	83/84	82,6/83,6
	rs800292	69/71	69,3/71,5
	rs2088102	106/108	105,9/107,0

Tabela 4. Sequência dos iniciadores de PCR e Minisequenciamento para o painel de 14 SNPs. Os trechos em letras minúsculas representam as seqüências de plasmídeo PUC 18 inseridas na extremidade 5' dos iniciadores com o propósito de alterar seus tamanhos. Os números entre parênteses indicam o tamanho dos iniciadores de misequenciamento.

SNP	Iniciadores de PCR		Iniciadores de Mini-sequenciamento
	Diretos	Reversos	
AFR3_rs2697520	TGCCTGTGTGTGTGTGT ATGT	CGTATTCTTTAAATGGTGCAA CA	tcgcgctttcggatgacgggtgaaacctctgacacatgcagctcccggagacggtcacagctgtctTTTGTCT AAAAGTATCCCCTGTTA (95)
AFR3_rs8035530	CCTCTTCAGTACCTCTTC CAGTG	CAACAGAAGGGAGAGCAAGA C	tcgcgctttcggatgacgggtgaaacctctgacacatgcagctcccggagacggtcacagcGCAATACAAGA AAAGCACCTTCAT (88)
AFR3_rs137211	GCTTGCCTATTGATAGA AGGGT	CCCTCTTTCACGATACTTAGC AT	tcgcgctttcggatgacgggtgaaacctctgacacatgcCACCACAGTTTTGATTTTCAGTCC (65)
AFR3_rs2416791	CCCAGCTTGTGGTCATT TATTA	GATTCTGGCAACGCTGACT	tcgcgctttcggatgacgggtgaaCATCTACCATCAGCCCAATTC (48)
AFR3_rs9626698	GGCAACGCTCCTAGTGA TC	CAACAGCCAAGAAATATGGT G	tcgcgctttcggatgacgggtgaaacctctgacacatgcagctcccggagacggtcacagctgtctgtaagcgGAG AGGGGTTTTTGAATAACATTA (102)
AFR3_rs1443985	CGGGTGCTATGTGTACT TTCAT	CAAATTCTGCTTCCTGAAATA GA	tcgcgctttcggatgacgggtgaaacctctgacacatgcagctcccggagacggtcacagctgtctgtaagcggCTT GAGAAATGCTCTATACAAATGCCCTATTAA (110)
AFR3_rs735480	TTAGTTCGCTGCTGCTCT AAT	GATGGAGGCCTCAGAGAGA	tcgcgctttcggatgacgggtgaaacctctGCTTAAAAGTCATGCGCCTAAG (55)
AMR4_rs8058694	GATTTCTTCCTGCAGCCT C	TGGAAACCTACACCACCTCTC	tcgcgctttcggatgacgggtgaaacctctgacacatGGAAGGATTGCATCACGATACA (62)
AMR4_rs2234636	TTCCATGTCCATCCAAT GTA	ATGGCTGTGGAGTTCGAA	tcgcgctttcggatgacgggtgaaacctctgaGCGACTCCAAAAAGAAGACAA (56)
AMR4_rs3760657	AGGAGCAGACAAACAG TCCTAT	GGATGGCAACAATACCTCAC T	tcgcgctttcggatgacgggtgaaaaGTGTCAGATGACACAGCAC (47)
AMR4_rs2619681	CATGCTTGCCAACACGA T	AACATATGAACAGATGTCCA ACA	tcgcgctttcggatgacgggtgaaacctctgacacatgcagcCAACATCATTATTAATCAGGAAA ATTAAAATCACATT (82)
AMR4_rs800292	CCAAACATATCCAGAAG GCAC	AATTCAGGCATAATTGCTACA AT	tcgcgctttcggatgacgggtgaaacctctgacacatgcaCCCTGGATATAGATCTCTTGGAAAT (68)
AMR4_rs2088102	GGATCTTAGCCTCTTCT GCTACA	CCTTCTCCAGCACAGGG	tcgcgctttcggatgacgggtgaaacctctgacacatgcagctcccggagacggtcacagctgtctgtaagcggatgcc gggaCCCAGGTGCTCAACAGAA (105)
AMR4_rs691968	TTTCACTCCTCCAAGGA CC	TGATGGCTACACAAGAGCAA A	tcgcgctttcggatgacgggtgaaacctctgacacatgcagctcccggagacggtcacagctgtctgtGCCATCC TGGTGCCTA (88)

a) AFR3



b) AMR 4

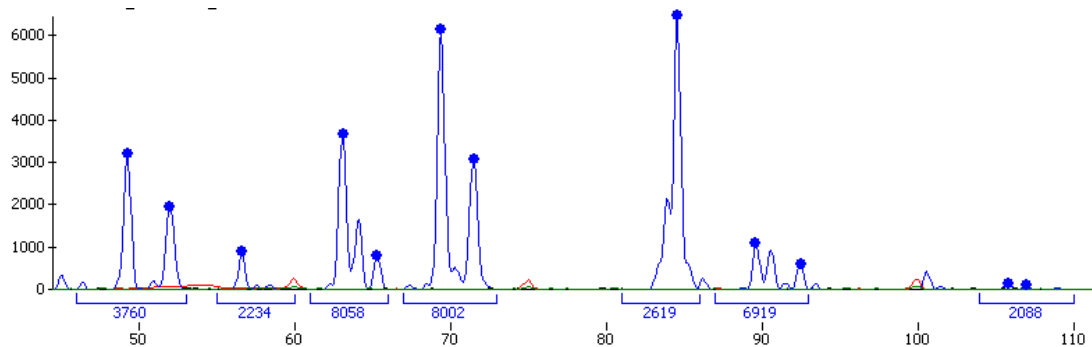


Figura 7: Eletroferogramas da genotipagem por minisequenciamento monocromático em reações multiplex dos Painéis AFR3 (a) e AMR4 (b), utilizando-se DNA de um pool de amostras (10 indivíduos), apresentando a maioria dos loci em heterozigose (2 picos para cada marcador).

4.4 Controle da qualidade dos painéis de SNPs

Os resultados dos testes de controle de qualidade da genotipagem dos 14 SNPs foram considerados satisfatórios, tendo sido alcançado 87% de sucesso de maneira geral, sendo que 92,5% dos genótipos esperados (loci genotipados com sucesso) foram alcançados para o painel AFR3 e 82% para o painel AMR4. O sucesso de genotipagem foi considerado quando todos os resultados obtidos para a genotipagem de um dado indivíduo com um determinado marcador foram idênticos (tanto para reações distintas de minisequenciamento quanto para eletroforeses distintas), sem ocorrência de falhas de genotipagem (“*missing data*”) em três reações independentes de minisequenciamento. Apenas alguns casos isolados de discordância foram verificados e, a maioria deles deveu-se a casos onde ocorreu baixa intensidade do sinal (tamanho do pico no eletroferograma). Esta situação que fez com que um dos picos ficasse abaixo do tamanho mínimo necessário para ser considerado como específico, gerando assim a discordância. Esse tipo de falha pode ser controlada mediante a inspeção cuidadosa dos picos gerados pelo programa de análise e eliminação dos loci apresentando picos de baixa

intensidade.

4.5 Genotipagem dos painéis de SNPs nas populações parentais

As frequências alélicas obtidas a partir da genotipagem dos 14 SNPs nos indivíduos africanos (AFR1), europeus (EUR1) e nativos americanos (AMR, MCA, ASU e SHI) são apresentadas na **tabela 5**. Esta tabela apresenta também os dados de frequências obtidas a partir de bancos de dados para outras populações africanas (AFR, YRI NiloS) e européias (Edo, EUR, CEU, DEU) que não foram utilizadas no presente estudo, com o intuito de possibilitar a comparação dos dados de frequência dos marcadores dentro e entre continentes.

Os dados apresentados na **tabela 5** mostram que, de modo geral, foram detectadas grandes diferenças nas frequências alélicas entre grupos continentais e relativa homogeneidade dentro desses grupos, o que confirma que os marcadores selecionados são de fato informativos de ancestralidade.

4.6 Análise genética da população

As frequências alélicas para todos os loci genotipados para os indivíduos parentais e miscigenados foram estimadas. Esses dados foram utilizados para análises de Equilíbrio de Hardy-Weinberg (EHW). A **tabela 6** mostra os resultados dos testes para o EHW para todos os loci genotipados nas populações parentais e a **tabela 7** mostra os resultados para os painéis de 14 SNPs e 40 INDELS nas diferentes populações de doadores e pacientes. Os resultados obtidos mostram que, após a correção de Bonferroni, apenas os seguintes loci encontraram-se fora do EHW: locus rs1372115 do painel de 14 SNPs em doadores de sangue amostrados em Diamantina, locus 20 do painel de 40 INDELS em pacientes com doença falciforme amostrados em Montes Claros e em ameríndios e locus 2937 do painel de 10 INDELS em ameríndios. Esses loci estão destacados em negrito nas **tabelas 6 e 7**.

Tabela 5. Frequências dos SNPs genotipados em diferentes populações

SNP	Populações ¹													
	Africanas					Européias					Ameríndias			
	AFR1	AFR	YRI	NiloS	Edo	AFA	EUR1	EUR	CEU	DEU	AMR	MCA	ASH	SHI
Painel AFR3														
rs1443985	0,15	0,09	0,1	0,11	0,1	0,21	0,85	0,91	0,90	0,94	0,84	0,85	0,76	0,9
rs9626698	0,50	0,79	0,8	0,76	0,8	0,57	0,00	0,04	0,03	0,04	0,00	0,00	0,00	0,0
rs2416791	0,76	0,05	0,0	0,10	0,0	0,18	0,88	0,92	0,92	0,92	0,53	0,64	0,44	0,5
rs2697520	0,17	0,07	0,1	0,10	0,1	0,28	0,80	0,88	0,89	0,85	0,31	0,31	0,26	0,3
rs8035530	0,70	0,82	0,8	0,86	0,8	0,67	0,08	0,02	0,02	0,02	0,48	0,44	0,53	0,5
rs735480	0,09	0,03	0,0	0,06	0,0	0,21	0,90	0,95	0,95	0,96	0,88	0,82	0,99	0,8
rs1372115	1,00	0,99	1,0	0,97	1,0	0,79	0,37	0,07	0,17	0,23	0,41	0,07	0,62	0,5
Painel AMR4														
rs2234636	0,20	0,00	-	-	-	-	0,28	0,28	-	-	0,78	0,77	0,73	0,8
rs3760657	0,09	0,02	-	-	-	-	0,10	0,07	-	-	0,26	0,12	0,40	0,3
rs8058694	0,59	0,23	-	-	-	-	0,88	0,61	-	-	0,70	0,62	0,75	0,7
rs800292	0,63	0,63	-	-	-	-	0,25	0,18	-	-	0,93	1,00	0,87	0,9
rs2619681	0,17	0,06	-	-	-	-	0,13	0,17	-	-	0,86	0,88	0,90	0,8
rs691968	0,41	0,42	-	-	-	-	0,04	0,00	-	-	0,01	0,00	0,0	0,0
rs2088102	0,71	0,65	-	-	-	-	0,54	0,50	-	-	0,33	0,27	0,40	0,3

¹ Populações: AFR1 = Africanos (painel do SNP500Cancer, genotipados no presente estudo), AFR = Africanos (Tian *et al.*, 2006), YRI= Yoruban (Africanos ocidentais do projeto HapMap), NiloS = Kanuri (Nilo-Saharan speakers - Nigeria), Edo = Bini (grupo Niger-Congo - Bantu), AFA = Afro-Americanos (Coriell Institute for Medical Research), EUR1= Europeus (painel do SNP500Cancer, genotipados no presente estudo), EUR = Europeus (Tian *et al.*, 2006), CEU = Europeus do painel CEPH, DEU = Europeus-Americanos da cidade de New York, AMR = Média para todos os Ameríndios, MCA= Ameríndios Monte Carmelo, ASH = Ameríndios Ashaninka, SHI= Ameríndios Shima. Os dados apresentados para africanos (AFR, YRI, NiloS, Edo, AFA) e europeus (EUR, CEU, DEU) foram obtidos de Tian *et al.*, 2006 (painel AFR3) ou do banco de dados do SNP500Cancer (painel AMR4).

Tabela 6. Teste exato de Fisher para o equilíbrio de Hardy –Weinberg, para as populações parentais. Valores em negrito indicam loci fora do equilíbrio após correção de Bonferroni

	Parentais		
	Africanos	Europeus	Ameríndios
40 INDELS			
locus 1	0,047188	0,645000	0,812813
locus 2	0,648125	0,539062	0,622812
locus 3	0,326875	1,000000	1,000000
locus 4	0,543750	0,633750	0,724375
locus 5	0,155312	1,000000	1,000000
locus 6	0,618125	0,211875	1,000000
locus 7	1,000000	1,000000	0,116250
locus 8	0,015938	1,000000	0,206250
locus 9	0,330000	0,031563	0,226250
locus 10	1,000000	0,672188	1,000000
locus 11	0,073750	1,000000	0,418750
locus 12	0,082500	1,000000	1,000000
locus 13	1,000000	0,210000	0,336875
locus 14	0,098750	0,657500	1,000000
locus 15	0,315625	1,000000	0,479375
locus 16	0,163125	1,000000	1,000000
locus 17	0,098125	0,685937	0,635312
locus 18	1,000000	0,194062	0,500625
locus 19	0,307188	0,120625	0,604062
locus 20	1,000000	0,010000	<0,00001
locus 21	1,000000	1,000000	0,560937
locus 22	0,607500	1,000000	1,000000
locus 23	1,000000	0,407187	0,680625
locus 24	1,000000	0,369688	0,528125
locus 25	1,000000	0,067500	0,205313
locus 26	0,397813	0,104375	0,077188
locus 27	1,000000	1,000000	0,204687
locus 28	1,000000	1,000000	0,757188
locus 29	0,427500	0,220938	1,000000
locus 30	0,387500	0,276562	0,593437
locus 31	0,624375	1,000000	1,000000
locus 32	1,000000	1,000000	0,032813
locus 33	0,357812	0,655937	0,789687
locus 34	1,000000	1,000000	1,000000
locus 35	0,230313	1,000000	0,525000
locus 36	1,000000	1,000000	0,186250
locus 37	0,392188	1,000000	0,536563
locus 38	1,000000	0,670625	1,000000
locus 39	0,515938	0,375312	1,000000
locus 40	0,207500	0,415312	0,609375

Tabela 6. Continuação

	Parentais		
	Africanos	Europeus	Ameríndios
14 SNPs			
locus rs1443985	0,417812	1,000000	1,000000
locus rs9626698	0,675312	1,000000	1,000000
locus rs2416791	0,000625	0,338125	0,634375
locus rs2697520	0,422812	0,532813	0,047188
locus rs8035530	1,000000	1,000000	0,819063
locus rs735480	1,000000	1,000000	0,012188
locus rs1372115	1,000000	0,220000	0,504375
locus rs2234636	1,000000	0,589375	0,505625
locus rs3760657	1,000000	1,000000	0,255312
locus rs8058694	0,071875	0,025313	0,761250
locus rs800292	0,397813	0,282813	0,002188
locus rs2619681	0,432500	1,000000	0,604688
locus rs691968	1,000000	1,000000	1,000000
locus rs2088102	0,590313	0,669375	0,246562
10 INDELS			
locus 777	1,000000	1,000000	0,060312
locus 2937	1,000000	1,000000	<0,00001
locus 2650	1,000000	0,217188	0,506250
locus 216	0,540625	1,000000	0,455313
locus 1652	1,000000	0,365000	1,000000
locus 1689	0,649687	1,000000	0,035000
locus 735	1,000000	0,371875	0,528438
locus 52	0,068750	0,003125	0,011875
locus 145	1,000000	1,000000	0,805000
locus 2750	1,000000	1,000000	0,007812
locus 2062	0,689063	0,272813	1,000000

Tabela 7. Teste exato de Fisher para o equilíbrio de Hardy -Weinberg para as amostras de doadores de sangue e pacientes com doença falciforme. Valores em negrito indicam loci fora do equilíbrio após correção de Bonferroni. Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.

	DOADORES												
	HBH	DIA	DIV	GOV	ITU	JFO	MCU	MOC	PMI	PNO	SJR	URA	UDI
locus 1	1,0000	0,6841	0,6328	1,0000	1,0000	0,0778	0,4672	1,0000	0,6647	1,0000	1,0000	0,6388	1,0000
locus 2	1,0000	0,4053	0,5434	1,0000	1,0000	0,6509	1,0000	0,6325	0,2834	1,0000	1,0000	1,0000	1,0000
locus 3	0,2741	0,3613	0,6159	1,0000	1,0000	1,0000	1,0000	0,3513	0,6656	0,4206	1,0000	0,6825	0,2503
locus 4	0,7038	0,6028	0,3522	1,0000	1,0000	0,1019	0,6959	1,0000	0,6569	1,0000	0,2841	0,1844	1,0000
locus 5	0,5597	1,0000	1,0000	1,0000	1,0000	0,3784	0,6163	1,0000	0,1259	1,0000	0,5847	0,2906	1,0000
locus 6	1,0000	1,0000	0,6363	1,0000	0,6369	1,0000	0,7119	0,3747	1,0000	0,2041	1,0000	0,6831	0,5838
locus 7	0,7147	0,6394	0,6616	1,0000	0,0463	1,0000	0,7094	1,0000	1,0000	0,0434	0,2834	1,0000	0,5291
locus 8	0,1247	0,6503	0,3484	0,4159	0,5809	0,4416	0,2303	1,0000	1,0000	1,0000	0,6169	1,0000	0,4859
locus 9	1,0000	1,0000	1,0000	0,1403	1,0000	1,0000	0,0263	0,0759	1,0000	0,0897	0,5844	0,4403	1,0000
locus 10	0,3800	1,0000	0,3878	0,2666	0,1288	0,7150	0,2575	0,6663	0,4106	0,1784	1,0000	1,0000	0,5456
locus 11	1,0000	0,1806	0,6231	1,0000	1,0000	1,0000	0,4294	0,6219	1,0000	0,5475	0,5938	1,0000	1,0000
locus 12	1,0000	1,0000	0,6578	1,0000	1,0000	0,4656	1,0000	1,0000	1,0000	0,0128	0,5944	0,0209	1,0000
locus 13	0,4538	0,6744	0,3244	0,3984	0,0747	0,2866	0,7097	0,6453	0,2150	0,4313	0,6303	0,4163	1,0000
locus 14	0,1347	0,6775	1,0000	1,0000	0,3306	1,0000	1,0000	0,0500	0,1859	0,4000	0,6009	0,1259	1,0000
locus 15	1,0000	0,4038	0,6484	1,0000	1,0000	0,4463	0,7066	0,6772	1,0000	0,6716	1,0000	0,6719	1,0000
locus 16	0,6784	0,1853	0,1800	0,4688	1,0000	1,0000	0,3653	1,0000	0,3044	1,0000	1,0000	1,0000	1,0000
locus 17	1,0000	0,6847	1,0000	0,0994	0,4234	0,6925	0,4672	1,0000	1,0000	1,0000	0,6031	0,6975	0,5897
locus 18	0,4572	0,2428	1,0000	0,4347	0,0947	0,4497	1,0000	1,0000	0,6613	0,4113	1,0000	1,0000	1,0000
locus 19	0,2647	1,0000	1,0000	1,0000	1,0000	0,0438	0,6856	1,0000	0,6416	0,1034	1,0000	1,0000	0,2122
locus 20	1,0000	1,0000	1,0000	0,4256	1,0000	0,0544	0,5444	1,0000	1,0000	0,4031	1,0000	1,0000	0,1728

Tabela 7. Continuação

	DOADORES												
	HBH	DIA	DIV	GOV	ITU	JFO	MCU	MOC	PMI	PNO	SJR	URA	UDI
locus 21	0,2891	1,0000	1,0000	1,0000	0,1725	0,6763	1,0000	0,5381	1,0000	1,0000	1,0000	0,4141	1,0000
locus 22	0,6684	0,6534	1,0000	1,0000	1,0000	1,0000	0,6847	0,6038	0,3516	0,2006	1,0000	0,6038	0,1684
locus 23	0,6306	0,2097	0,6638	1,0000	1,0000	0,6872	0,3716	0,1441	1,0000	0,0594	1,0000	0,6616	1,0000
locus 24	0,0903	0,6931	0,6700	1,0000	0,6881	0,2706	0,4503	1,0000	0,0931	1,0000	1,0000	0,6256	0,5519
locus 25	1,0000	0,5206	0,6356	0,0081	1,0000	0,0275	1,0000	0,3291	1,0000	1,0000	0,5488	1,0000	0,4709
locus 26	0,6941	0,6925	1,0000	1,0000	0,2278	0,7050	0,4853	0,6813	0,3794	0,2209	0,5925	0,2359	1,0000
locus 27	1,0000	1,0000	0,3541	1,0000	0,6203	1,0000	1,0000	0,3572	1,0000	0,3431	1,0000	1,0000	1,0000
locus 28	0,3975	0,1828	0,6575	1,0000	1,0000	0,2388	0,1447	0,3553	0,6491	0,3728	0,1338	1,0000	0,5634
locus 29	0,3978	0,3838	0,4053	0,4553	0,2113	0,4669	0,2781	0,3797	0,6491	1,0000	0,3013	0,0128	0,5941
locus 30	1,0000	0,6278	0,5409	1,0000	0,0878	0,0953	1,0000	0,3628	0,6066	0,6616	1,0000	1,0000	1,0000
locus 31	0,3981	0,2059	0,0853	1,0000	0,0563	1,0000	1,0000	0,6544	1,0000	0,0391	0,3222	1,0000	0,1947
locus 32	0,0800	1,0000	0,6588	1,0000	1,0000	1,0000	1,0000	0,6459	1,0000	0,0666	0,1388	1,0000	0,5506
locus 33	0,1803	0,1219	1,0000	1,0000	0,0244	0,4425	1,0000	1,0000	1,0000	0,6331	1,0000	1,0000	1,0000
locus 34	0,1234	1,0000	1,0000	1,0000	0,6528	0,6738	0,4750	0,6913	0,6584	1,0000	0,0244	0,6934	1,0000
locus 35	0,0141	1,0000	0,6550	1,0000	1,0000	0,6438	1,0000	0,6044	1,0000	0,2328	1,0000	0,0800	0,5959
locus 36	0,5534	0,5925	0,5231	1,0000	0,6513	0,0581	0,7034	0,2869	0,5441	0,1134	0,2263	0,3959	0,3838
locus 37	0,6925	0,2103	0,5341	0,1425	1,0000	0,0716	0,4238	1,0000	1,0000	0,6063	1,0000	1,0000	0,2272
locus 38	0,6941	0,2028	0,0716	1,0000	0,3150	1,0000	0,7338	0,1641	1,0000	1,0000	0,3006	0,3994	0,2828
locus 39	1,0000	0,0894	0,3728	1,0000	0,3269	1,0000	0,3000	1,0000	0,6359	1,0000	0,0588	0,4228	1,0000
locus 40	0,4897	1,0000	0,3216	1,0000	1,0000	0,6741	0,0113	1,0000	0,6384	0,6419	0,0363	0,1456	0,0494
locus rs1443985	1,0000	1,0000	1,0000	0,4453	0,2894	1,0000	0,6666	1,0000	0,2603	1,0000	0,4666	0,3356	1,0000
locus rs9626698	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,0859	0,5297	1,0000	1,0000	1,0000	0,3728
locus rs2416791	1,0000	0,6125	0,1869	0,4416	1,0000	0,1563	0,3850	0,6766	0,0006	0,6634	0,1694	0,0853	1,0000
locus rs2697520	1,0000	0,2313	0,5175	1,0000	1,0000	1,0000	0,1316	0,3963	0,1041	0,6200	1,0000	0,2813	1,0000

Tabela 7. Continuação

	DOADORES												
	HBH	DIA	DIV	GOV	ITU	JFO	MCU	MOC	PMI	PNO	SJR	URA	UDI
locus rs8035530	1,0000	0,5981	1,0000	1,0000	0,5675	1,0000	0,5450	0,1094	1,0000	0,5478	1,0000	1,0000	1,0000
locus rs735480	1,0000	0,2181	0,5394	0,4466	0,6478	0,4566	0,2219	1,0000	0,3256	0,1747	0,5159	0,6309	0,0950
locus rs1372115	1,0000	<0,0001	0,0059	0,4272	0,1156	0,0022	0,2569	0,2456	0,1850	0,0978	1,0000	0,4275	0,0022
locus rs2234636	1,0000	0,5347	0,6263	1,0000	0,3641	0,6313	0,6944	1,0000	0,5291	0,5441	0,6009	1,0000	0,5622
locus rs3760657	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,2088	1,0000	1,0000
locus rs8058694	1,0000	0,6719	0,1450	1,0000	1,0000	1,0000	0,4278	1,0000	0,6791	0,6734	1,0000	0,3878	1,0000
locus rs800292	1,0000	0,4413	0,1125	1,0000	0,1763	0,7131	0,1022	0,4319	1,0000	0,7169	1,0000	1,0000	0,5806
locus rs2619981	1,0000	0,4872	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,2559	1,0000	0,4650	0,5469	0,2131
locus rs691968	1,0000	0,5547	1,0000	1,0000	1,0000	1,0000	0,2491	1,0000	1,0000	0,5975	0,3241	0,1894	1,0000
locus rs2088102	1,0000	1,0000	1,0000	1,0000	0,2275	0,0563	0,0003	1,0000	1,0000	0,0028	1,0000	0,5816	0,0344
	PACIENTES												
	HBH	DIV	GOV	ITU	JFO	MCU	MOC	PMI	SJR	URA	UDI		
locus 1	0,3272	0,3228	0,1059	1,0000	0,0769	0,4409	0,6409	0,3781	1,0000	0,0544	1,0000		
locus 2	0,1744	0,5931	1,0000	0,4353	0,0534	1,0000	0,3697	0,4659	1,0000	0,6109	1,0000		
locus 3	0,0319	1,0000	0,4322	1,0000	1,0000	0,2138	1,0000	0,7181	1,0000	1,0000	0,2547		
locus 4	1,0000	1,0000	0,4034	1,0000	0,2272	1,0000	1,0000	0,2369	1,0000	0,5431	0,5197		
locus 5	1,0000	1,0000	0,0625	1,0000	1,0000	0,0794	1,0000	1,0000	1,0000	1,0000	0,2794		
locus 6	1,0000	0,3859	0,0625	0,3866	1,0000	1,0000	1,0000	0,7244	0,3275	1,0000	1,0000		
locus 7	0,0597	0,0256	1,0000	1,0000	0,5413	1,0000	0,2500	0,0769	1,0000	0,0494	1,0000		
locus 8	0,3609	0,6063	0,2238	1,0000	1,0000	1,0000	0,5219	0,4000	1,0000	0,6000	1,0000		
locus 9	0,1103	0,3694	0,6856	1,0000	0,5331	1,0000	0,1488	1,0000	1,0000	1,0000	1,0000		
locus 10	0,0078	0,6028	1,0000	0,3406	1,0000	0,3850	0,6138	0,2566	0,2075	0,1059	0,2803		
locus 11	0,3309	0,3341	0,6641	0,3141	1,0000	1,0000	0,0606	0,0809	1,0000	0,6000	0,2859		
locus 12	1,0000	1,0000	0,1328	1,0000	0,4725	0,5484	1,0000	0,1775	1,0000	1,0000	1,0000		

Tabela 7. Continuação

	PACIENTES										
	HBH	DIV	GOV	ITU	JFO	MCU	MOC	PMI	SJR	URA	UDI
locus 13	0,0372	1,0000	0,1953	1,0000	1,0000	1,0000	1,0000	0,4088	1,0000	1,0000	0,1863
locus 14	0,6322	0,0925	0,2184	1,0000	0,2306	0,4906	1,0000	1,0000	1,0000	0,6369	1,0000
locus 15	1,0000	0,6144	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,6506
locus 16	1,0000	0,1694	0,5688	0,4306	0,3181	1,0000	1,0000	0,6269	1,0000	0,2356	0,6206
locus 17	1,0000	0,6422	0,6716	1,0000	1,0000	0,4903	0,2078	0,2616	1,0000	0,2841	0,3463
locus 18	0,7009	1,0000	0,6256	1,0000	0,0422	0,1484	0,0381	1,0000	1,0000	1,0000	0,6863
locus 19	0,0784	1,0000	1,0000	1,0000	0,1459	0,0888	0,1266	0,6478	0,3325	1,0000	0,5456
locus 20	0,0097	0,1797	0,0244	1,0000	1,0000	1,0000	<0,0001	0,0013	1,0000	0,0119	1,0000
locus 21	0,5394	1,0000	0,5588	1,0000	1,0000	1,0000	0,1384	0,1381	0,2003	1,0000	0,5666
locus 22	1,0000	0,2531	0,6281	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,2906	1,0000
locus 23	0,2916	0,0578	1,0000	1,0000	1,0000	1,0000	0,6238	0,0650	1,0000	1,0000	0,6481
locus 24	0,6678	0,0822	1,0000	1,0000	0,2172	1,0000	0,6888	0,2622	1,0000	0,6081	0,6822
locus 25	0,1247	1,0000	0,6534	1,0000	1,0000	0,5353	0,3528	0,6788	1,0000	0,5716	0,6553
locus 26	1,0000	0,1384	0,7009	0,4394	1,0000	0,4816	0,4041	1,0000	1,0000	1,0000	0,0744
locus 27	1,0000	1,0000	0,1331	0,3250	0,4897	1,0000	0,6531	0,2356	1,0000	1,0000	0,6178
locus 28	1,0000	0,6344	1,0000	1,0000	1,0000	1,0000	0,3838	1,0000	1,0000	0,6103	0,6025
locus 29	0,2463	0,4019	1,0000	1,0000	0,4278	0,3897	0,2159	0,7328	1,0000	1,0000	0,6725
locus 30	1,0000	0,4003	0,0638	1,0000	1,0000	0,5369	1,0000	1,0000	1,0000	0,5906	1,0000
locus 31	0,6959	0,3700	0,4003	1,0000	1,0000	0,4719	0,0088	0,4484	0,2084	0,6116	1,0000
locus 32	0,6522	1,0000	0,1750	1,0000	1,0000	1,0000	0,4106	0,1013	1,0000	0,5866	0,3759
locus 33	0,6891	1,0000	0,6619	0,4425	1,0000	1,0000	0,3884	0,4575	1,0000	1,0000	0,6350
locus 34	1,0000	0,6422	0,1709	0,4825	1,0000	1,0000	1,0000	1,0000	1,0000	0,6256	0,5891

Tabela 7. Continuação

	PACIENTES										
	HBH	DIV	GOV	ITU	JFO	MCU	MOC	PMI	SJR	URA	UDI
locus 35	0,1009	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,4681	1,0000	0,6075	0,1091
locus 36	0,6028	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,2363	1,0000	1,0000	1,0000
locus 37	0,3850	0,3625	1,0000	0,4369	1,0000	0,1069	0,5963	1,0000	1,0000	0,1034	1,0000
locus 38	0,6775	1,0000	0,3047	0,4575	1,0000	1,0000	0,4238	0,3069	1,0000	0,5834	0,1622
locus 39	1,0000	0,6347	0,6681	1,0000	0,5594	1,0000	1,0000	0,7303	1,0000	1,0000	0,5706
locus 40	0,1750	0,6091	0,0759	0,4244	1,0000	0,4922	0,6656	1,0000	1,0000	0,5719	0,0366
locus rs1443985	0,1266	1,0000	0,2434	1,0000	0,4450	0,4356	0,1313	0,3022	1,0000	0,6525	0,3541
locus rs9626698	0,2700	0,1884	1,0000	0,1075	0,5809	0,4234	1,0000	0,6797	1,0000	1,0000	0,6313
locus rs2416791	0,7063	0,5863	0,7088	1,0000	1,0000	0,5481	0,6488	0,7147	1,0000	0,1500	0,6678
locus rs2697520	1,0000	0,2597	0,2200	1,0000	1,0000	0,2144	0,3900	0,0544	1,0000	0,3200	0,3459
locus rs8035530	1,0000	0,5891	0,1869	1,0000	1,0000	0,5025	1,0000	0,3594	1,0000	1,0000	0,5675
locus rs735480	0,0006	0,1297	0,7122	1,0000	0,2216	1,0000	0,6844	0,4675	1,0000	1,0000	0,2619
locus rs1372115	0,0703	0,1159	0,0925	0,4300	1,0000	0,3903	0,4241	0,2438	1,0000	0,5859	0,0072
locus rs2234636	0,5866	0,6550	0,6516	1,0000	1,0000	1,0000	0,3163	0,0031	0,1991	0,3641	1,0000
locus rs3760657	1,0000	1,0000	1,0000	0,0878	1,0000	1,0000	1,0000	0,4828	1,0000	1,0000	0,0706
locus rs8058694	0,4556	0,3931	1,0000	1,0000	1,0000	1,0000	0,6672	0,7106	1,0000	0,1513	0,1016
locus rs8002292	0,2597	0,0856	1,0000	1,0000	1,0000	0,0800	0,2097	0,7250	1,0000	0,3578	0,4228
locus rs2619681	1,0000	0,1353	1,0000	1,0000	1,0000	0,2275	1,0000	0,6197	1,0000	0,5259	1,0000
locus rs691968	0,1419	0,3244	0,1594	1,0000	0,2394	1,0000	1,0000	1,0000	1,0000	0,0559	0,0775
locus rs2088102	0,7056	0,0225	0,0456	0,2750	0,2522	1,0000	1,0000	0,0672	1,0000	0,2881	0,1625

Os valores de F_{st} par a par para doadores de sangue e pacientes com doença falciforme amostrados em diferentes localidades foram estimados e utilizados para computar uma matriz de distâncias genéticas (**tabelas 8 e 9**). A significância de cada estimativa foi obtida pela avaliação do intervalo de confiança de 95% fornecidos pelo programa GDA e os valores significativos estão apresentados em negrito.

As representação gráficas das matrizes de F_{st} obtidas aplicando o “*Multidimensional scaling*” (MDS) permitiram que as populações fossem exploradas quanto a similaridades genéticas (**figuras 8 e 9**). Nessas análises, devido ao tamanho reduzido da amostra, foram excluídas as populações de doadores de sangue de Governador Valadares e pacientes com doença falciforme de Diamantina, Ituiutaba e São João del Rei.

Embora as análises tenham sido realizadas também para os subconjuntos de marcadores (14 SNPs e 40 INDELS) separadamente, são apresentados somente os resultados correspondentes à análise da totalidade dos loci já que não foram evidenciados padrões particulares ou significativos nas análises utilizando subconjuntos de marcadores (exceto para o grupo de doadores de sangue amostrados em Belo Horizonte, para os quais foi utilizado apenas o painel de 40 INDELS).

Tabela 8. Matriz de significância para F_{ST} par a par para doadores de sangue genotipados para 14 SNPs e 40 INDELS. Valores negativos correspondem a $F_{ST} = 0$. A significância de cada estimativa foi obtida pela avaliação do intervalo de confiança de 95% e os valores significativos estão em negrito. Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.

	HBH	DIA	DIV	ITU	JFO	MCU	MOC	PMI	PNO	SJR	URA	UDI
HBH	0,00000											
DIA	0,008256	0,00000										
DIV	0,002055	0,009450	0,00000									
ITU	0,014326	0,009955	-0,002353	0,00000								
JFO	0,014877	0,005401	-0,002140	-0,001967	0,00000							
MCU	0,001529	0,004188	-0,005593	0,000178	-0,000962	0,00000						
MOC	0,004778	0,008412	-0,007684	-0,002554	-0,002471	0,000845	0,00000					
PMI	0,009359	0,019238	-0,002576	-0,005911	0,004816	-0,002970	-0,002887	0,00000				
PNO	0,012323	0,003077	0,003315	-0,004182	-0,003387	0,004082	-0,000221	0,006115	0,00000			
SJR	0,005227	0,010826	-0,000714	-0,005495	0,003794	0,001059	0,000505	0,001050	0,000665	0,00000		
URA	0,004371	0,015934	-0,003806	-0,004625	0,008859	0,002413	0,004114	0,001277	0,003593	-0,006513	0,00000	
UDI	0,019920	0,017029	0,005257	0,008921	-0,001546	0,001699	0,006855	0,005236	0,001748	0,009265	0,016201	0,00000

Tabela 9. Matriz de significância para F_{ST} par a par para pacientes com doença falciforme genotipados para 14 SNPs e 40 INDELS. Valores negativos correspondem a $F_{ST} = 0$. A significância de cada estimativa foi obtida pela avaliação do intervalo de confiança de 95% e os valores significativos estão em negrito. Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.

	HBH	DIV	GOV	JFO	MCU	MOC	PMI	URA	UDI
HBH	0,00000								
DIV	0,007712	0,00000							
GOV	-0,002349	0,004135	0,00000						
JFO	0,042657	0,026993	0,029557	0,00000					
MCU	0,011535	0,011743	0,001061	0,027904	0,00000				
MOC	0,007339	0,011788	0,001624	0,051954	-0,005319	0,00000			
PMI	0,000594	0,006290	-0,002353	0,040239	-0,005550	0,001165	0,00000		
URA	-0,006524	0,006133	-0,004215	0,029242	-0,012157	-0,003369	-0,004326	0,00000	
UDI	0,000599	0,010470	0,001887	0,045537	-0,002846	0,001151	-0,001788	-0,008728	0,00000

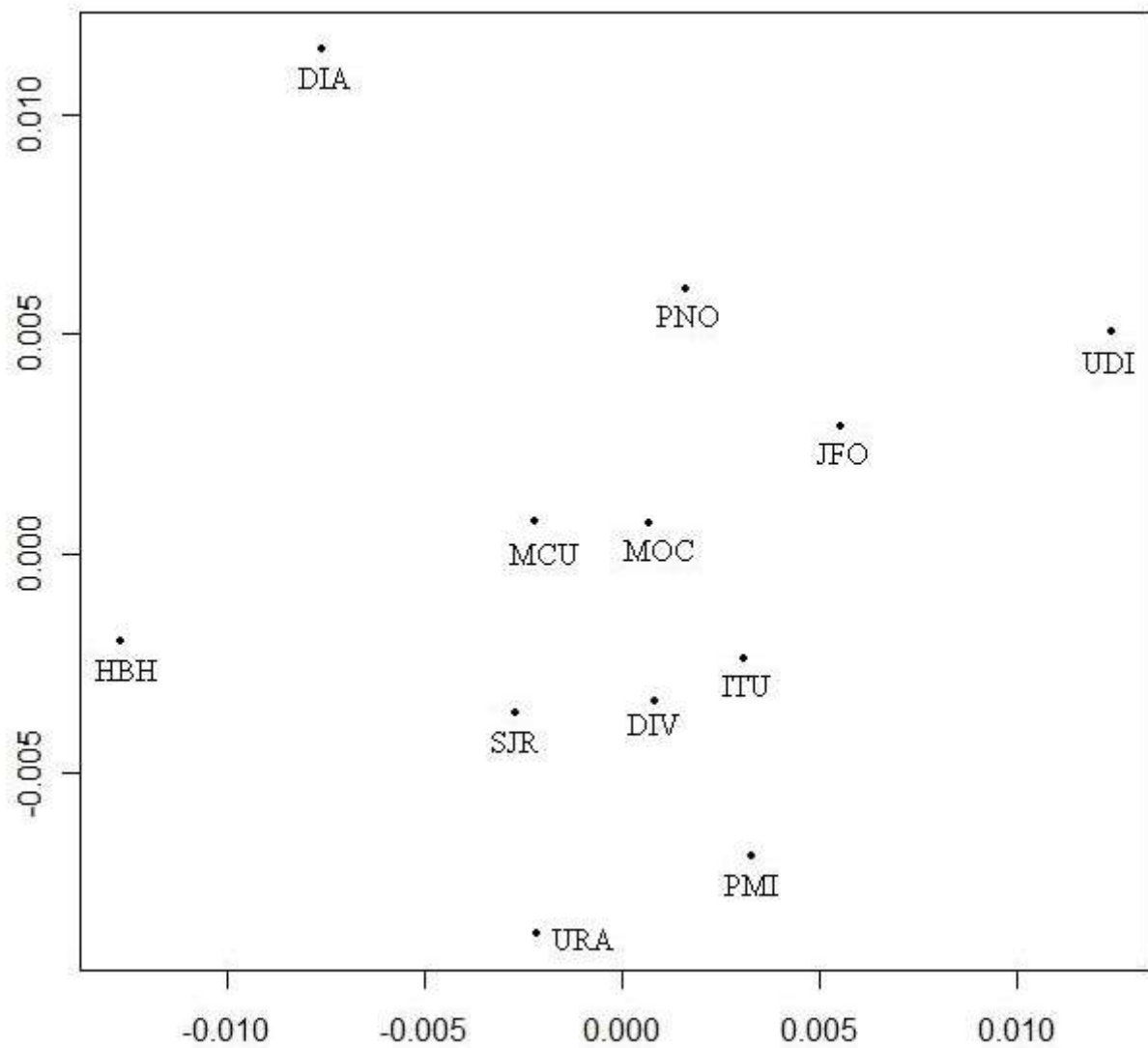


Figura 8. Representação gráfica bidimensional dos valores de F_{st} par a par para doadores de sangue amostrados em diferentes localidades, usando a abordagem MDS (Multidimensional scaling). Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.

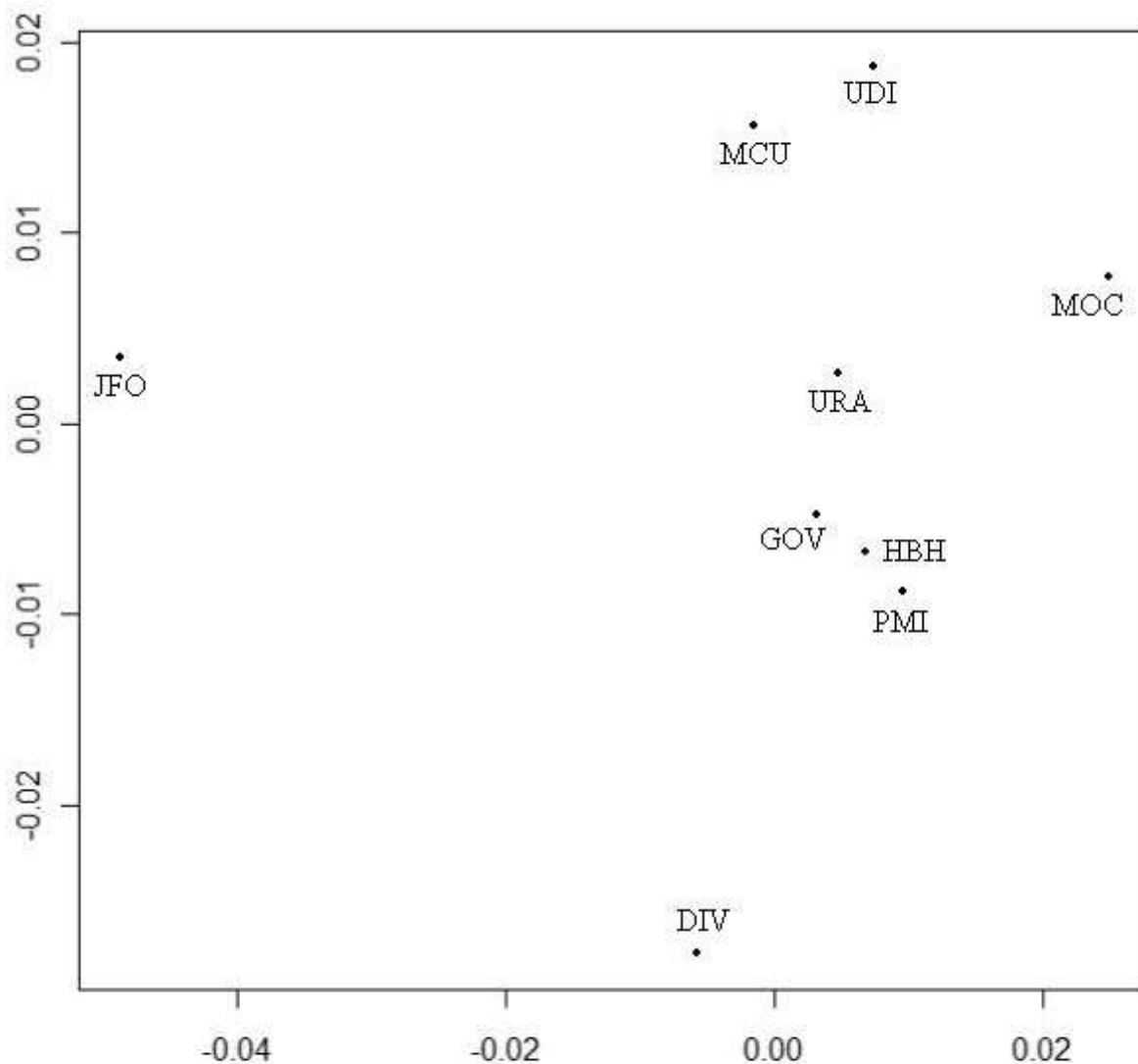


Figura 9. Representação gráfica bidimensional dos valores de F_{st} par a par para pacientes com doença falciforme amostrados em diferentes localidades, usando a abordagem MDS (Multidimensional scaling). Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.

4.7 Isolamento por distância

O teste do modelo de IBD foi empregado para acessar a significância da associação entre as matrizes de distâncias genéticas (F_{st}) e geográficas (distância linear entre os locais de amostragem) para os doadores de sangue e pacientes com doença falciforme amostrados em diferentes regiões do estado de Minas Gerais, por meio do teste de Mantel, com 1000 permutações. Nessas análises, devido ao tamanho reduzido da amostra, foram excluídas as populações de doadores de sangue de Governador

Valadares e pacientes com doença falciforme de Diamantina, Ituiutaba e São João del Rei. Resultados destas análises são apresentados na **tabela 10 e figuras 10 e 11**. As coordenadas geográficas das diferentes localidades estão apresentadas na tabela S2. Análises utilizando o logaritmo das distâncias genéticas e/ou geográficas, bem como os testes realizados com subconjuntos de marcadores (14 SNPs e 40 INDELS separadamente) também foram realizadas e os resultados obtidos não foram dotados de significância estatística.

Tabela 10. Análise de isolamento por distância para doadores de sangue e pacientes com doença falciforme amostrados em diferentes regiões do Estado de Minas Gerais

	Teste de Mantel		
	Z	r	p-valor
Doadores de sangue			
	Distâncias genéticas X Distâncias geográficas		
14 SNPs	1.059.457	-0,0974	0,7440
40 INDELS	1.507.431	-0,0788	0,6860
14SNPs + 40 INDELS	1.059.457	-0,0974	0,7400
	Distâncias genéticas X log(Distâncias geográficas)		
14 SNPs	0,7505	-0,0900	0,7540
40 INDELS	10.575	-0,0776	0,6890
14SNPs + 40 INDELS	0,7505	-0,0900	0,7330
	log (Distâncias genéticas) X Distâncias geográficas		
14 SNPs	-722491742	-0,1728	0,8680
40 INDELS	-697.378.814	-0,1527	0,8500
14SNPs + 40 INDELS	-722.491.742	-0,1728	0,8710
	log (Distâncias genéticas) X log(Distâncias geográficas)		
14 SNPs	-4.787.222	0,1691	0,8950
40 INDELS	-4.622.781	-0,1531	0,1330
14SNPs + 40 INDELS	-4.787.222	-0,1691	0,8900
Pacientes com doença falciforme			
	Distâncias genéticas X Distâncias geográficas		
14 SNPs	2.818.722	0,1399	0,2480
40 INDELS	1.130.868	-0,0546	0,6240
14SNPs + 40 INDELS	0,9704	0,0857	0,3280
	Distâncias genéticas X log(Distâncias geográficas)		
14 SNPs	17.427	0,1478	0,1430
40 INDELS	0,7745	-0,0184	0,5760
14SNPs + 40 INDELS	1.497.549	0,0671	0,4450
	log (Distâncias genéticas) X Distâncias geográficas		
14 SNPs	-406.933.888	0,0386	0,4600
40 INDELS	-390.197.112	-0,2025	0,8790
14SNPs + 40 INDELS	-384.352.739	-0,1364	0,7450
	log (Distâncias genéticas) X log(Distâncias geográficas)		
14 SNPs	-2.730.093	0,0479	0,4300
40 INDELS	-2.535.309	-0,1415	0,7890
14SNPs + 40 INDELS	-2.517.757	-0,0745	0,6440

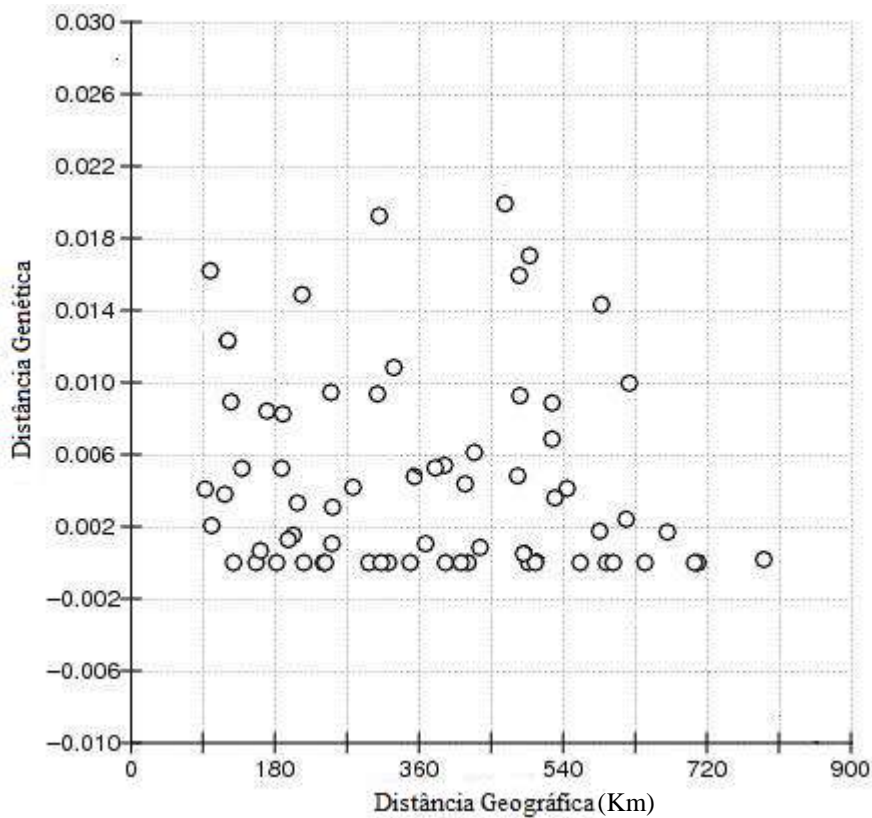


Figura 10. Análise do isolamento por distância para doadores de sangue amostrados em diferentes regiões do Estado de Minas Gerais utilizando dados genéticos de 14 SNPs e 40 INDELS

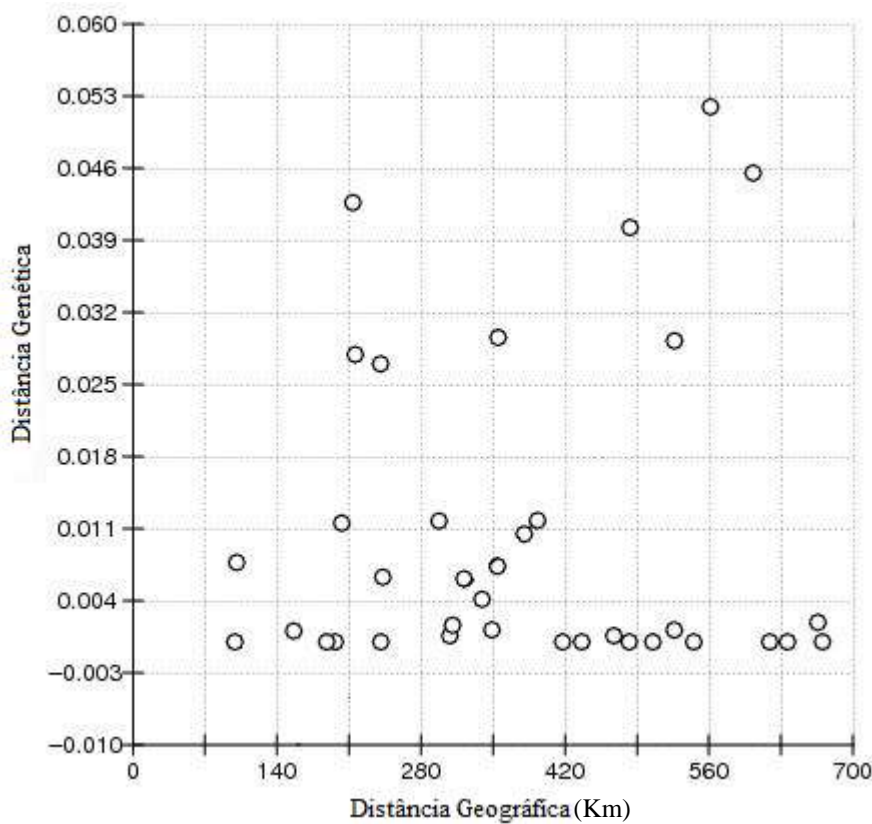


Figura 11. Análise do isolamento por distância para pacientes com doença falciforme amostrados em diferentes regiões do Estado de Minas Gerais utilizando dados genéticos de 14 SNPs e 40 INDELS

4.8 Precisão da estimativa de miscigenação utilizando o painel de 14 SNPs

4.8.1 Análises das populações parentais

Os dados de genótipos dos 14 SNPs para os indivíduos pertencentes às populações parentais putativas foram analisados por meio do método bayesiano de agrupamento implementado no programa STRUCTURE. Nessas análises, o STRUCTURE foi capaz de atribuir corretamente quase todos os indivíduos a seus verdadeiros grupos, mesmo sem informação a priori das populações de origem, ou seja, utilizando exclusivamente os dados genéticos (**figura 12**).

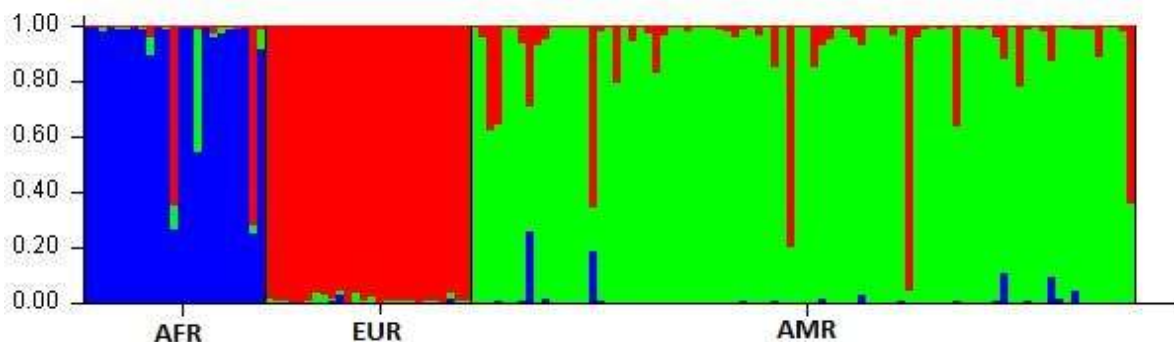


Figura 12. Análise de miscigenação para os indivíduos representantes das populações parentais (AFR = africanos, EUR = europeus e AMR = ameríndios) utilizando o programa STRUCTURE, a partir dos dados de genotipagem do painel de 14 MIAs. Cada barra vertical representa um indivíduo e as cores representam a ancestralidade.

4.8.2 Miscigenação populacional

Para avaliar a precisão das estimativas de miscigenação obtidas com o conjunto de 14 MIAs, os resultados obtidos com esses marcadores foram comparados com as estimativas obtidas para um conjunto de 108 MIAs, considerando-se os mesmos indivíduos tanto nas populações parentais quanto na miscigenada.

Esses 108 SNPs foram selecionados dentre os milhares de marcadores genotipados no projeto SNP500Cancer, por apresentarem $F_{ST} > 0.20$ entre populações africanas, europeias e ameríndias, do painel do CEPH (Cann *et al.*, 2002), 2) e $F_{ST} < 0.10$ entre populações dentro desses grupos.

Os indivíduos utilizados para esta comparação foram os 23 latino-americanos do projeto

SNP500Cancer, denominados como Hispânicos. Estudos prévios (Hughes *et al.*, 2008), usando aproximadamente 350.000 SNPs haviam mostrado que este grupo apresenta ancestralidade predominantemente européia. Devido ao elevado número de falhas de genotipagem para o locus rs241679 especificamente na amostra de hispânicos, esse marcador foi excluído das análises.

A **tabela 11** mostra os resultados comparativos obtidos utilizando-se três métodos estatísticos para estimativa de miscigenação. Esta tabela evidencia que nas análises realizadas considerando como população parental ameríndia apenas os indivíduos provenientes da localidade de Monte Carmelo: 1) tanto o painel de 13 SNPs quanto o de 108 confirmam a predominância de miscigenação européia nos Hispânicos; 2) utilizando o conjunto de 13 SNPs na amostra estudada, estimativas adequadas da miscigenação européia foram obtidas, mas por outro lado o painel de 13 SNPs parece ter superestimado a miscigenação africana e subestimada a ameríndia; 3) Os métodos descritos por Chakraborty (1985) e Dupanloup and Bertorelle (2001), desenvolvidos para estimar a miscigenação no nível populacional apresentaram resultados mais concordantes entre si do que em relação ao programa STRUCTURE, desenvolvido para estimar a miscigenação no nível individual.

As análises considerando apenas os nativos americanos de Monte Carmelo foram realizadas porque são esses os indivíduos ameríndios genotipados no projeto SNP500Cancer. Sendo assim, há dados de genótipos para um grande número de loci disponíveis para esses indivíduos. Isto facilita a comparação das estimativas de miscigenação obtidas para os Hispânicos com diferentes subconjuntos de marcadores, mantendo-se os mesmos indivíduos tanto na composição da população miscigenada, quanto das parentais.

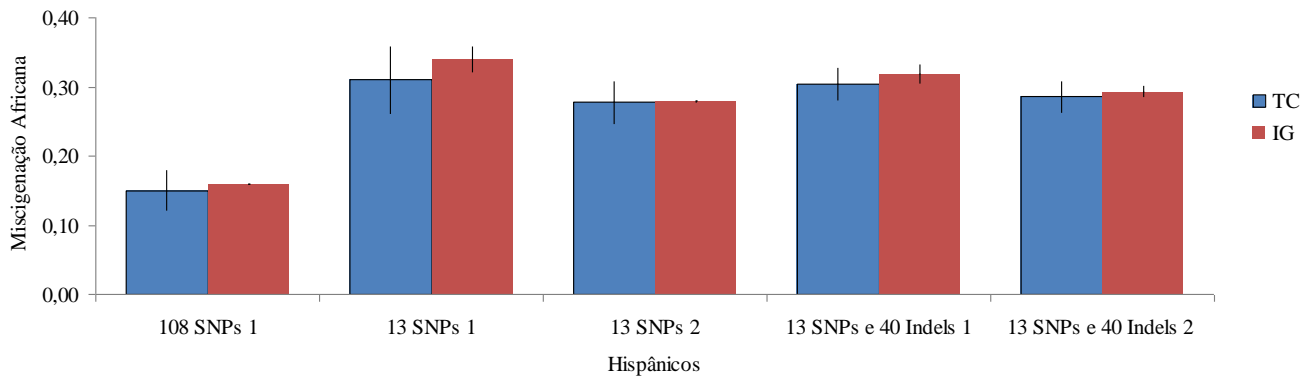
Tabela 11. Estimativas de miscigenação populacional obtidas por três métodos de análise para a amostra de Hispânicos, utilizando diferentes conjuntos de marcadores.

Conjunto de marcadores	Populações Parentais					
	Africana		Européia		Ameríndia	
	Estimativa de Miscigenação	95% CI ou SD	Estimativa de Miscigenação	95% CI ou SD	Estimativa de Miscigenação	95% CI ou SD
Dupanloup e Bertorelle (2001)						
108 SNPs ¹	0,15	0,030	0,66	0,030	0,20	0,020
13 SNPs ¹	0,31	0,050	0,62	0,060	0,07	0,050
13 SNPs ²	0,28	0,032	0,53	0,047	0,19	0,048
13 SNPs e 40 INDELS ¹	0,30	0,024	0,53	0,038	0,16	0,025
13 SNPs e 40 INDELS ²	0,29	0,024	0,50	0,036	0,22	0,027
Chakraborty (1985)						
108 SNPs ¹	0,16	0,000	0,64	0,000	0,20	0,000
13 SNPs ¹	0,34	0,020	0,58	0,020	0,08	0,020
13 SNPs ²	0,28	0,003	0,53	0,004	0,19	0,003
13 SNPs e 40 INDELS ¹	0,32	0,014	0,53	0,017	0,15	0,011
13 SNPs e 40 INDELS ²	0,29	0,010	0,50	0,013	0,21	0,009
Pritchard <i>et al.</i> (2000)						
108 SNPs ¹	0,09	-	0,75	-	0,16	-
13 SNPs ¹	0,23	-	0,69	-	0,08	-
13 SNPs ²	0,22	-	0,63	-	0,16	-
13 SNPs e 40 INDELS	0,14	-	0,80	-	0,06	-

1: Análise considerando como população parental ameríndia apenas o grupo Monte Carmelo

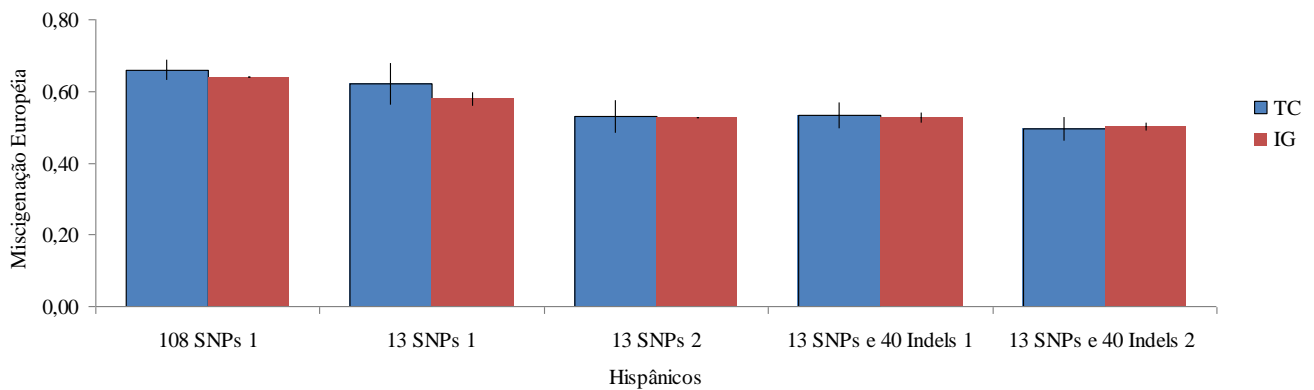
2: Análise considerando como população parental ameríndia os grupos Monte Carmelo, Shimaá e Ashaninkas.

Posteriormente, os Hispânicos foram genotipados para o painel de 40 INDELS e as estimativas de miscigenação populacional foram realizadas, considerando-se a totalidade dos marcadores genotipados (13 SNPs + 40 INDELS). Análises adicionais foram conduzidas, considerando como população parental nativo-americana a totalidade dos indivíduos ameríndios genotipados (Monte Carmelo, Shimaá e Ashaninkas) (**tabela 11**). As **figuras 13, 14 e 15** mostram as estimativas de miscigenação africana, européia e ameríndia obtidas para os Hispânicos com diferentes conjuntos de marcadores e diferentes composições de população parental ameríndia. Nesses gráficos, estimativas obtidas com a utilização dos métodos de Tempos de Coalescência (TC, Dupanloup e Bertorelle, 2001) e Identidade genética (IG, Chakraborty, 1985) são apresentadas.



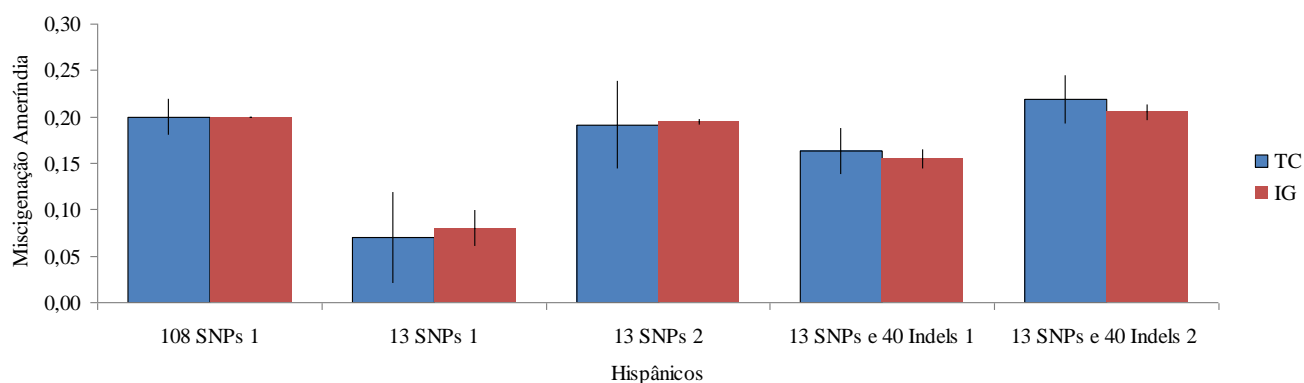
- 1: Análise considerando como população parental ameríndia apenas o grupo Monte Carmelo
- 2: Análise considerando como população parental ameríndia os grupos Monte Carmelo, Shimaá e Ashaninkas.

Figura 13. Estimativas de miscigenação africana para os Hispânicos considerando diferentes subconjuntos de marcadores



- 1: Análise considerando como população parental ameríndia apenas o grupo Monte Carmelo
- 2: Análise considerando como população parental ameríndia os grupos Monte Carmelo, Shimaá e Ashaninkas.

Figura 14. Estimativas de miscigenação européia para os Hispânicos considerando diferentes subconjuntos de marcadores



- 1: Análise considerando como população parental ameríndia apenas o grupo Monte Carmelo
- 2: Análise considerando como população parental ameríndia os grupos Monte Carmelo, Shimaa e Ashaninkas.

Figura 15. Estimativas de miscigenação ameríndia para os Hispânicos considerando diferentes subconjuntos de marcadores

4.8.3 Miscigenação Individual

Para avaliar a precisão das estimativas de miscigenação individual usando diferentes conjuntos de marcadores, a amostra de Hispânicos foi analisada por meio do programa STRUCTURE. Estas análises foram realizadas considerando o conjunto de 108 MIAs do SNP500Cancer, bem como subconjuntos de 80, 50, e 20 SNPs selecionados ao acaso, a partir do conjunto inicial de 108 marcadores.

As estimativas de miscigenação individual obtidas para os Hispânicos considerando os 108 MIAs foram também comparadas com as estimativas obtidas para os painéis de 13 SNPs, 40 INDELS e 10 INDELS previamente descritos.

A **tabela 12** mostra que, utilizando subconjuntos de 50 ou 80 SNPs, a correlação das estimativas de miscigenação individual em relação às obtidas para os 108 MIAs são maiores que 88%, enquanto com menos de 20 MIAs, esta correlação é menor que 66%. Além disso, é possível observar que a adição dos resultados de genotipagem do painel de 40 INDELS (Bastos-Rodrigues *et al.*, 2006) aos dados de 13 SNPs aumentaram a correlação das estimativas de miscigenação, em comparação com os resultados obtidos com os 108 SNPs. Por outro lado, a adição dos 10 INDELS informativos para a miscigenação ameríndia provocaram a diminuição dos valores desta correlação, apontando para a introdução de um possível viés.

A queda da precisão das estimativas de miscigenação no nível individual devido à redução do número de marcadores fica também evidente quando os intervalos de probabilidade bayesianos de 90% (90%-PI) para a miscigenação individual, gerados pelo STRUCTURE, são plotados contra as estimativas de miscigenação individual (**Figura 16**). Esta figura mostra que, para os indivíduos mais miscigenados, 90%-PI se eleva dramaticamente pela redução do número de marcadores, abaixo de 50.

Tabela 12. Coeficientes de correlação entre estimativas de miscigenação para a amostra de Hispânicos utilizando 108 SNPs, e as estimativas utilizando subconjuntos de 80, 50, 20 MIAs, bem como os conjuntos de 13 SNPs e 40 INDELS genotipados no presente estudo.

Marcadores	Populações parentais					
	Africanos		Europeus		Ameríndios	
	Coeficiente de correlação	p valor	Coeficientes de correlação	p valor	Coeficientes de correlação	p valor
80 SNPs	0,93	<0,001	0,93	<0,001	0,96	<0,001
50 SNPs	0,96	<0,001	0,93	<0,001	0,88	<0,001
20 SNPs	0,58	0,004	0,66	0,001	0,64	0,001
13 SNPs	0,60	0,003	0,49	0,019	0,52	0,011
40 INDELS	0,74	<0,001	0,52	0,011	0,55	0,007
13SNPs+40INDELS	0,78	<0,001	0,66	0,001	0,73	<0,001
13SNPs+40INDELS+10 MIAs	0,73	<0,001	0,59	0,003	0,68	<0,001

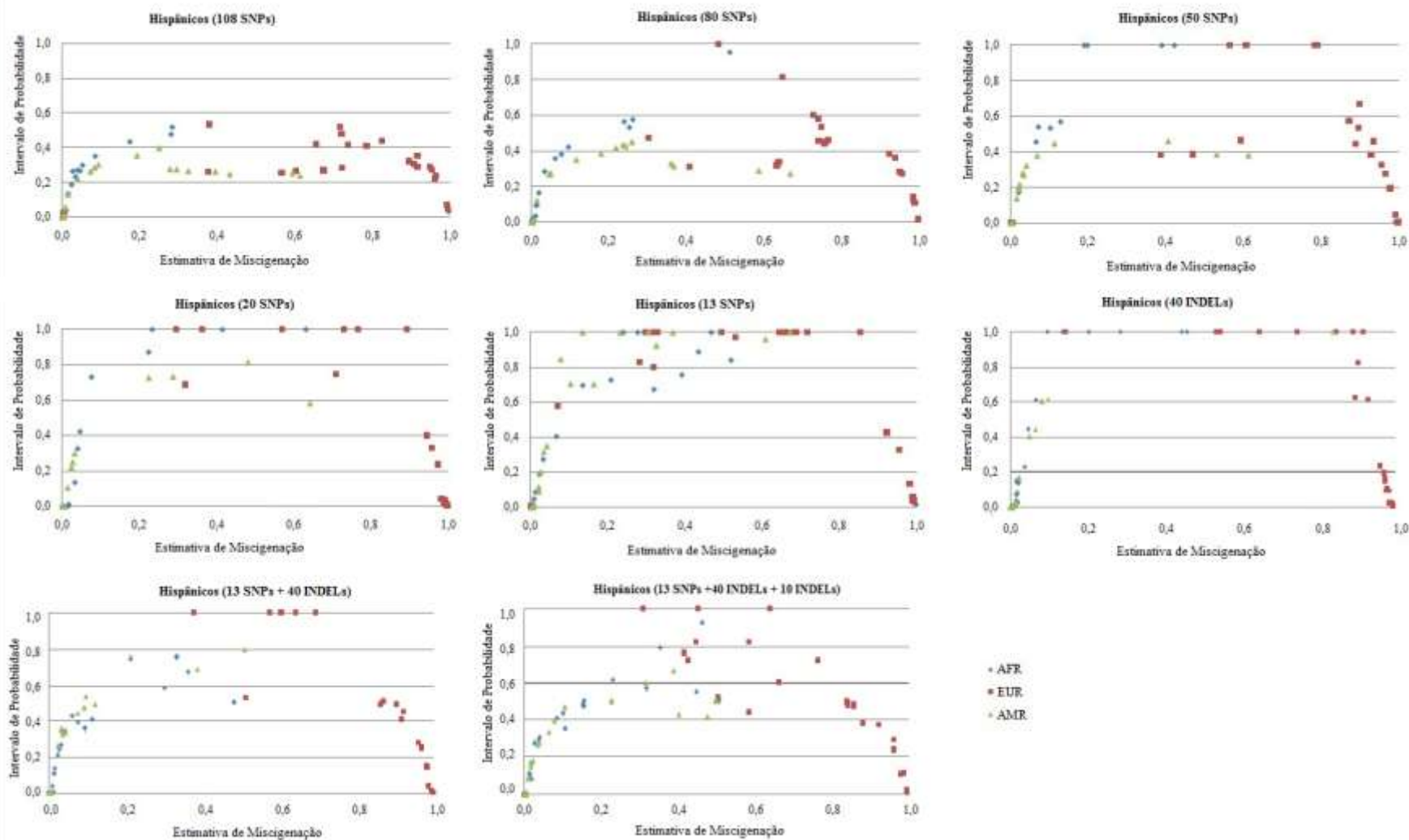


Figura 16. Variação da amplitude dos intervalos de probabilidade de 90% das estimativas de miscigenação individuais em função das estimativas pontuais de miscigenação fornecidas pelo programa STRUCTURE, com a utilização de diferentes conjuntos de marcadores

4.9 Análises de miscigenação dos doadores de sangue e pacientes com doença falciforme

4.9.1 Miscigenação populacional

A miscigenação no nível populacional para os doadores de sangue e os pacientes portadores de doença falciforme amostrados em diferentes regiões do Estado de Minas Gerais foi estimada por meio dos métodos de Identidade Genética (Chakraborty, 1985) e Tempos de Coalescência (Dupanloup e Bertorelle, 2001), com o auxílio dos programas ADMIX95 e ADMIX2, respectivamente, assim como por meio do método Bayesiano implementado no programa STRUCTURE (Pritchard *et al.*, 2000) (**tabela 13**). No caso do STRUCTURE, a estimativa populacional foi computada pela média das estimativas de miscigenação individual. Essas análises foram realizadas considerando-se o conjunto de 54 marcadores (14 SNPs e 40 INDELS), com exceção do grupo de doadores amostrados em Belo Horizonte, para os quais foi considerado apenas o painel de 40 INDELS.

Tanto os doadores quanto os pacientes foram analisados em conjunto, bem como separadamente, em grupos formados por indivíduos amostrados em cada localidade. Para os pacientes com doença falciforme, foram estimadas as contribuições das populações parentais para todo o grupo, e também separadamente para os portadores da anemia falciforme (HOM), que representam 71,2% dos pacientes e para os demais pacientes (HET), que representam 26,0% deste grupo (**tabela 13**). Para seis pacientes, não foi possível acessar a informação a respeito do perfil de hemoglobina, não sendo possível identificar se eram portadores de anemia falciforme ou outra forma heterozigótica da doença.

Tabela 13. Estimativa de miscigenação populacional para doadores de sangue e pacientes com doença falciforme amostrados em cada localidade, utilizando diferentes métodos de estimativa. AFR = Africanos, EUR = Europeus, AMR = Ameríndios. Os códigos de três letras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.

Populações	Estimativa de miscigenação populacional														
	Pritchard et al (2000)			Chakraborty (1985)						Dupanloup e Bertorelle (2001)					
	AFR	EUR	AMR	AFR	CI/SD	EUR	CI/SD	AMR	CI/SD	AFR	CI/SD	EUR	CI/SD	AMR	CI/SD
Doadores															
HBH	0,260	0,730	0,020	0,364	0,019	0,608	0,025	0,028	0,016	0,274	0,040	0,725	0,066	0,010	0,004
DIA	0,390	0,570	0,050	0,455	0,013	0,472	0,015	0,073	0,014	0,460	0,024	0,422	0,035	0,118	0,025
DIV	0,210	0,760	0,030	0,295	0,012	0,632	0,018	0,073	0,012	0,309	0,025	0,599	0,037	0,091	0,027
ITU	0,230	0,720	0,060	0,315	0,010	0,542	0,015	0,143	0,010	0,297	0,024	0,574	0,035	0,128	0,026
JFO	0,330	0,640	0,020	0,378	0,009	0,566	0,014	0,056	0,011	0,380	0,024	0,567	0,035	0,053	0,026
MCU	0,170	0,800	0,030	0,258	0,009	0,650	0,013	0,092	0,010	0,260	0,024	0,653	0,034	0,086	0,026
MOC	0,340	0,640	0,020	0,415	0,011	0,544	0,016	0,041	0,012	0,399	0,024	0,537	0,035	0,064	0,026
PMI	0,170	0,800	0,030	0,248	0,011	0,654	0,016	0,098	0,011	0,241	0,024	0,671	0,036	0,089	0,026
PNO	0,380	0,590	0,030	0,451	0,017	0,500	0,026	0,049	0,021	0,449	0,025	0,468	0,037	0,083	0,027
SJR	0,250	0,720	0,040	0,307	0,016	0,619	0,023	0,074	0,016	0,300	0,025	0,628	0,037	0,072	0,027
URA	0,220	0,750	0,030	0,264	0,014	0,641	0,022	0,095	0,014	0,275	0,025	0,630	0,037	0,095	0,027
UDI	0,270	0,710	0,020	0,369	0,012	0,459	0,020	0,172	0,014	0,344	0,025	0,517	0,039	0,139	0,028
Total	0,286	0,679	0,034	0,341	0,012	0,585	0,018	0,075	0,013	0,338	0,024	0,577	0,035	0,084	0,026

Tabela 13. Continuação

Populações	Estimativa de miscigenação populacional														
	Pritchard et al (2000)			Chakraborty (1985)					Dupanloup e Bertorelle (2001)						
	AFR	EUR	AMR	AFR	CI/SD	EUR	CI/SD	AMR	CI/SD	AFR	CI/SD	EUR	CI/SD	AMR	CI/SD
Pacientes															
HBH	0,410	0,560	0,030	0,451	0,012	0,475	0,018	0,075	0,013	0,440	0,023	0,455	0,034	0,105	0,025
DIV	0,480	0,490	0,030	0,497	0,017	0,455	0,023	0,049	0,018	0,502	0,024	0,410	0,037	0,088	0,026
GOV	0,460	0,460	0,080	0,489	0,012	0,422	0,014	0,089	0,012	0,503	0,024	0,385	0,035	0,111	0,025
JFO	0,820	0,150	0,040	0,779	0,006	0,174	0,007	0,047	0,006	0,796	0,024	0,145	0,039	0,058	0,027
MCU	0,530	0,410	0,070	0,562	0,008	0,316	0,011	0,122	0,010	0,576	0,024	0,285	0,038	0,139	0,027
MOC	0,360	0,580	0,060	0,378	0,018	0,522	0,029	0,100	0,021	0,398	0,024	0,472	0,036	0,129	0,026
PMI	0,360	0,530	0,120	0,384	0,018	0,466	0,024	0,151	0,019	0,414	0,025	0,370	0,039	0,216	0,027
URA	0,410	0,540	0,060	0,430	0,015	0,490	0,021	0,080	0,017	0,450	0,024	0,426	0,036	0,124	0,026
UDI	0,350	0,580	0,070	0,388	0,008	0,463	0,012	0,149	0,011	0,397	0,024	0,431	0,033	0,172	0,024
HOM	0,461	0,469	0,070	0,473	0,012	0,441	0,015	0,087	0,013	0,487	0,024	0,384	0,034	0,129	0,025
HET	0,383	0,573	0,044	0,407	0,016	0,539	0,021	0,055	0,018	0,430	0,024	0,455	0,036	0,115	0,026
Total	0,437	0,504	0,059	0,455	0,012	0,464	0,016	0,081	0,013	0,473	0,024	0,397	0,034	0,130	0,025

CI = Intervalo de confiança, SD = desvio padrão.

Os resultados dessas análises mostram que houve variação nas contribuições africana, europeia e ameríndia estimadas para as diferentes regiões do estado, tanto para doadores de sangue quanto para os pacientes com doença falciforme. Para facilitar a visualização dos resultados de cada região, as estimativas de miscigenação obtidos pelo método TC foram representadas nas **figuras 17 e 18**.

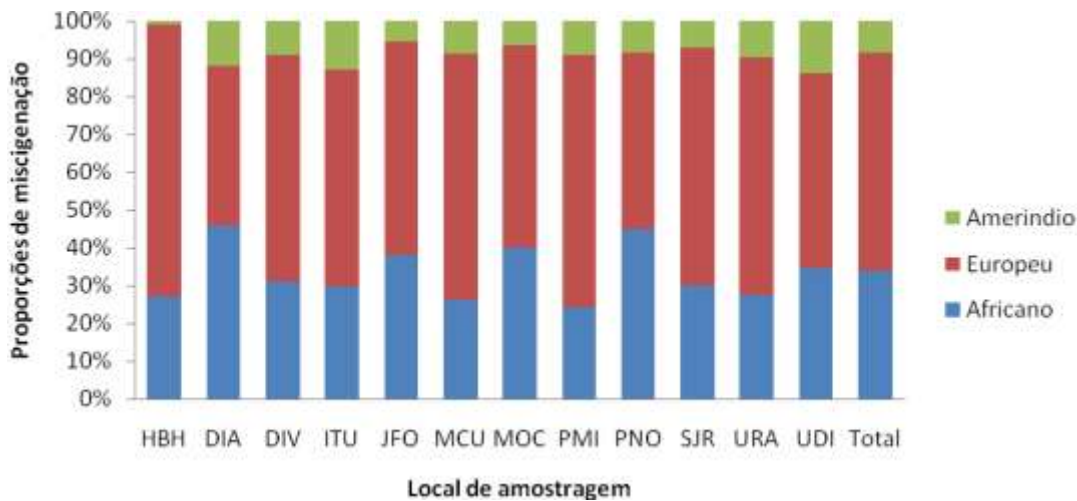


Figura 17. Proporções de contribuição das populações parentais africana, europeia e ameríndia para os grupos de doadores de sangue amostrados em diferentes localidades.

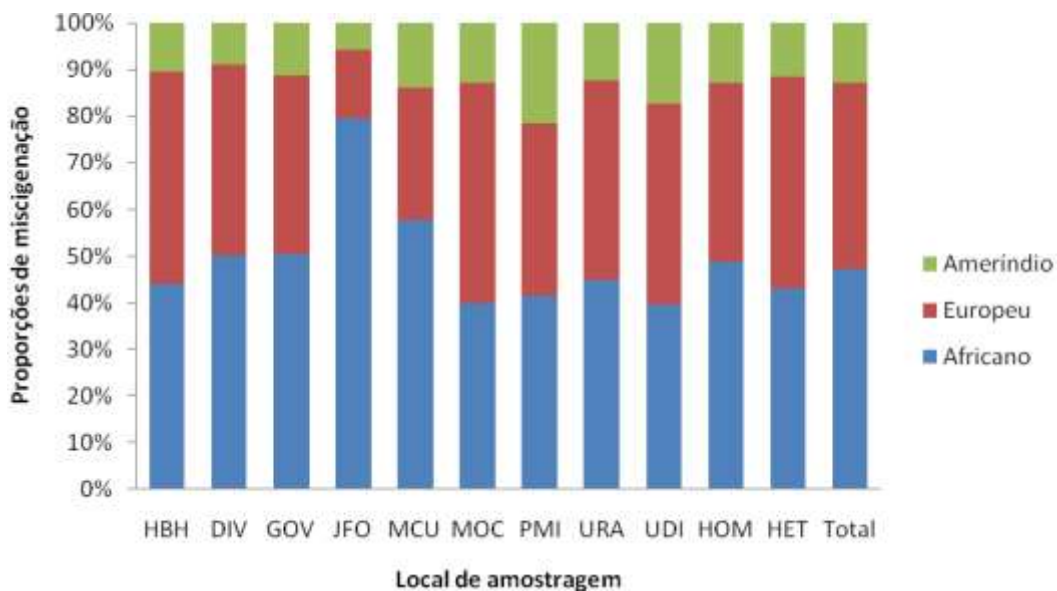


Figura 18. Proporções de contribuição das populações parentais africana, europeia e ameríndia para os grupos de pacientes com doença falciforme amostrados em diferentes localidades.

As **figuras 19 a 24** apresentam as estimativas de miscigenação africana, europeia e ameríndia para doadores de sangue e pacientes com doença falciforme amostrados em diferentes regiões do Estado de Minas Gerais, obtidas com os métodos de TC e IG. Estes gráficos permitem a visualização e comparação dos resultados obtidos com os dois métodos estatísticos. Os códigos de três letras apresentados nessas figuras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas.

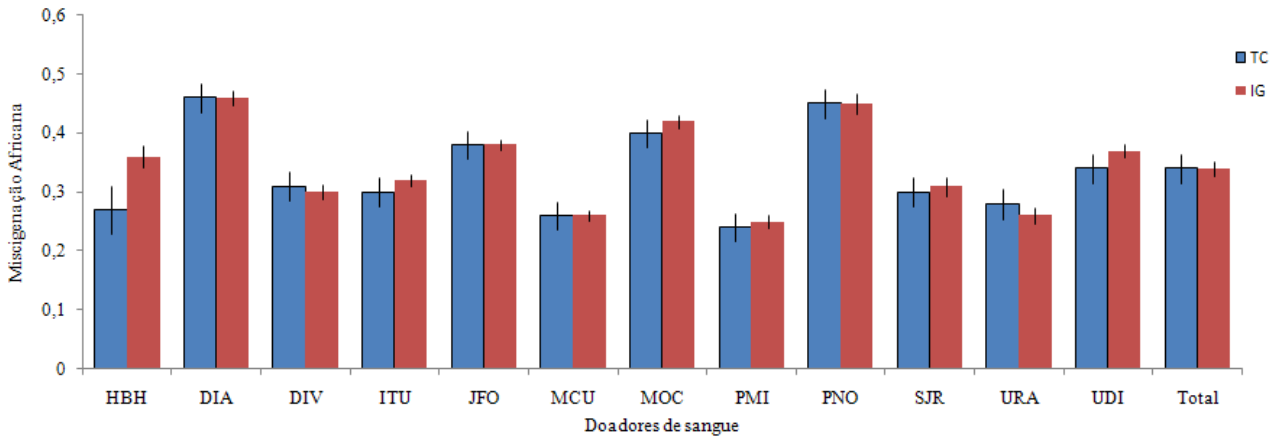


Figura 19. Estimativas de miscigenação africana para doadores de sangue amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente.

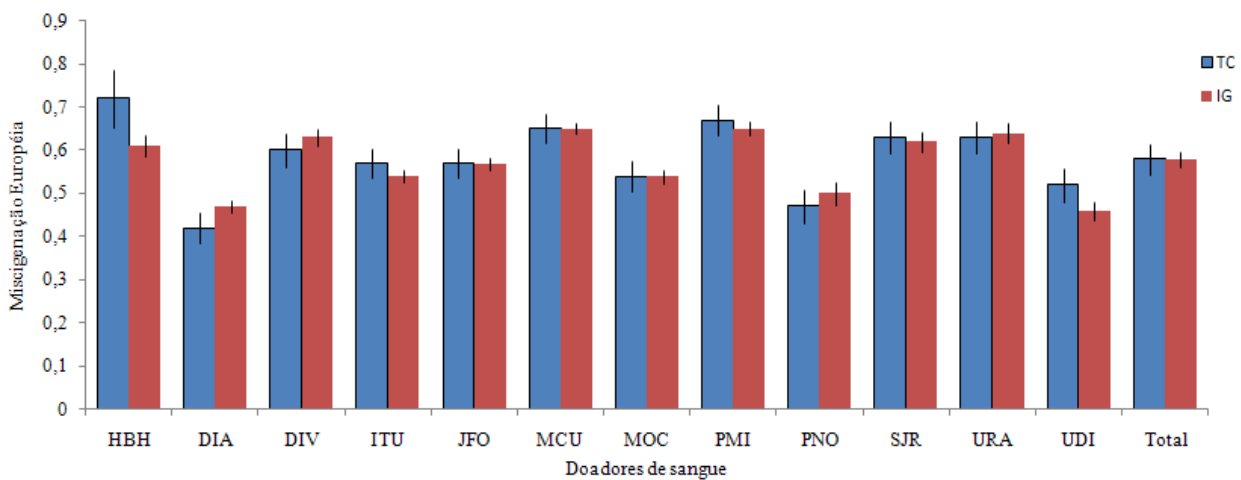


Figura 20. Estimativas de miscigenação europeia para doadores de sangue amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente.

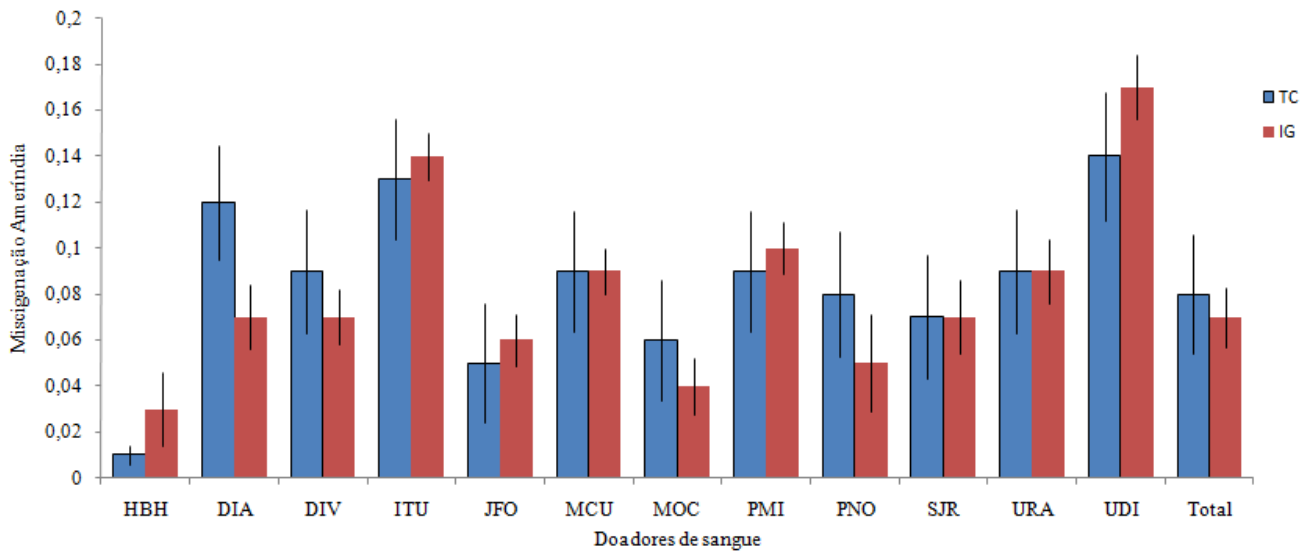


Figura 21. Estimativas de miscigenação ameríndia para doadores de sangue amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente.

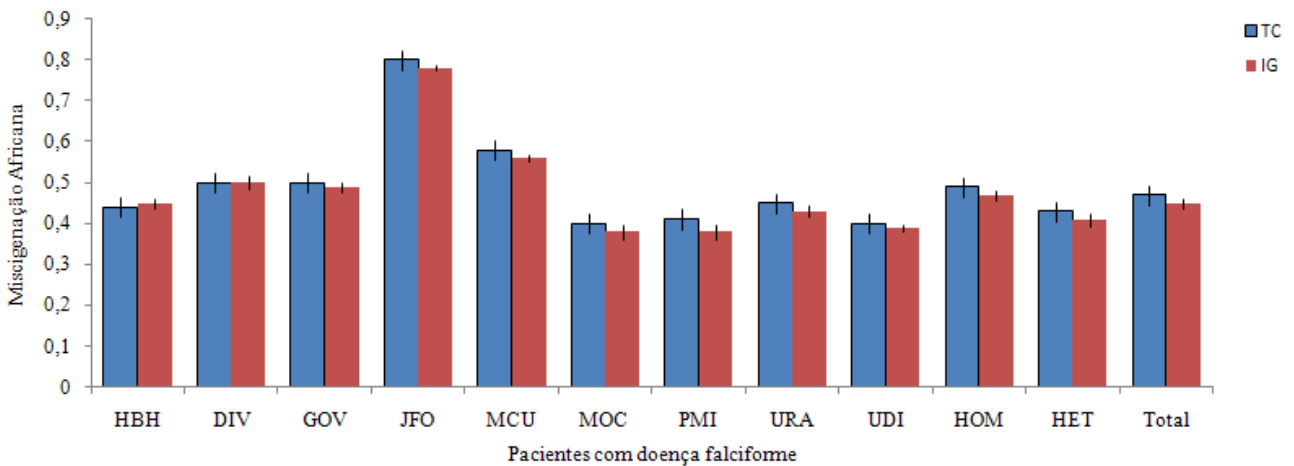


Figura 22. Estimativas de miscigenação africana para pacientes com doença falciforme amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente.

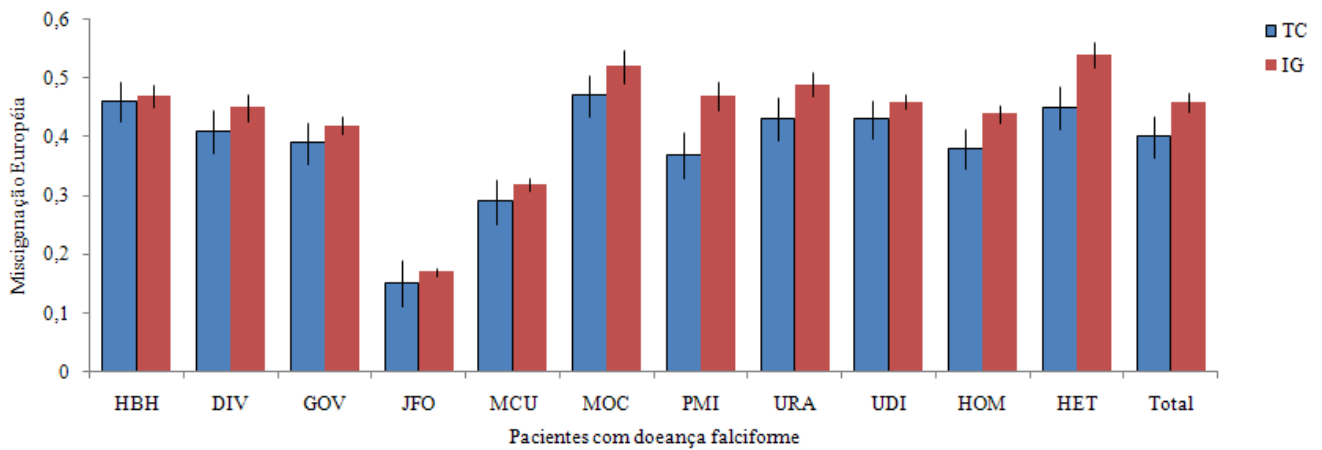


Figura 23. Estimativas de miscigenação europeia para pacientes com doença falciforme amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente.

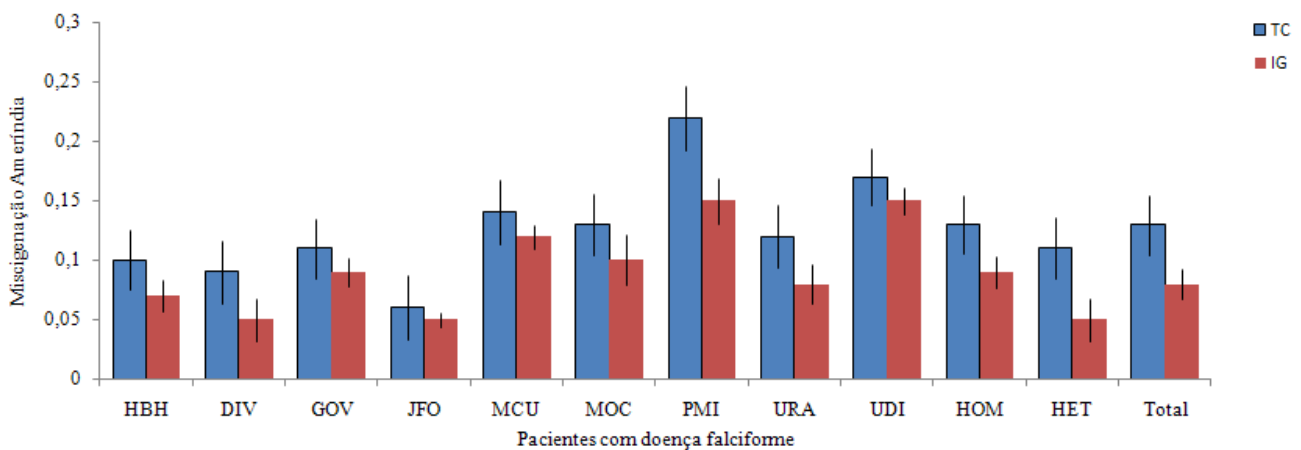


Figura 24. Estimativas de miscigenação ameríndia para pacientes com doença falciforme amostrados em diferentes localidades, obtidas pelos métodos de tempos de coalescência (TC) e identidade genética (IG) com o auxílio dos programas ADMIX2 e ADMIX95, respectivamente.

Com o intuito de investigar a existência de eventuais padrões geográficos de miscigenação para doadores de sangue e pacientes com doença falciforme, os resultados obtidos por meio do método de IG para esses grupos foram representados sobre o mapa do Estado de Minas Gerais (**figuras 25 e 26**),

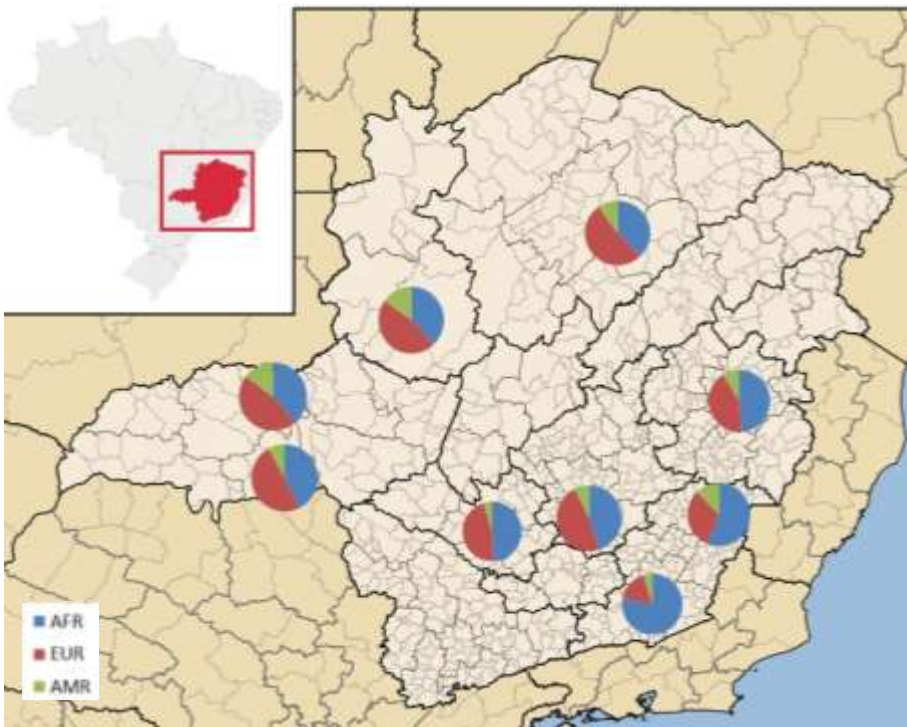


Figura 25. Mapa da distribuição geográfica da miscigenação populacional estimada pelo método de identidade genética para os pacientes com doença falciforme de Minas Gerais.

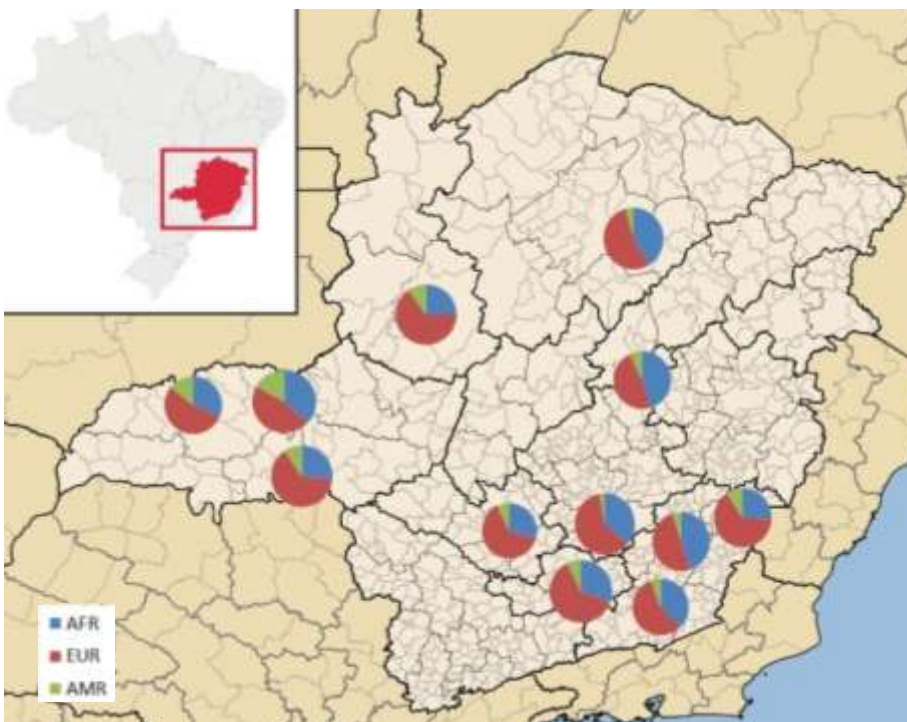


Figura 26. Mapa da distribuição geográfica da miscigenação populacional estimada pelo método de identidade genética para os doadores de sangue de Minas Gerais.

Ainda objetivando a pesquisa de padrões geográficos nos perfis de miscigenação, os valores de estimativa de miscigenação africana, européia e ameríndia obtidos com o método IG foram correlacionados à latitude e à longitude das localidades onde os indivíduos foram amostrados. Os resultados dessas análises (coeficientes de correlação de Pearson) são apresentados na **tabela 14** e mostram que a latitude se correlaciona negativamente com a ancestralidade africana e positivamente com a ancestralidade européia para os pacientes enquanto a longitude se correlaciona negativamente com a miscigenação ameríndia apenas para os doadores.

Tabela 14. Correlação entre as estimativas de miscigenação africana, européia e ameríndia obtidas com o método IG para doadores de sangue e pacientes com doença falciforme e a latitude ou longitude da localidade onde os indivíduos foram amostrados.

	Miscigenação Africana		Miscigenação Européia		Miscigenação Ameríndia	
	Coefficiente de correlação	p valor	Coefficiente de correlação	p valor	Coefficiente de correlação	p valor
Latitude						
doadores	0,233	0,465	-0,347	0,269	0,145	0,653
pacientes	-0,810	0,008	0,766	0,016	0,513	0,157
Longitude						
doadores	0,354	0,259	0,095	0,771	-0,763	0,004
pacientes	0,504	0,166	-0,463	0,211	-0,365	0,334

Análises adicionais realizadas com o programa STRUCTURE considerando os painéis de 14 SNPs e 40 INDELS separadamente e em conjunto revelaram uma alta correlação entre as estimativas de miscigenação obtidas com esses dois painéis tanto para doadores quanto para pacientes (**tabela 15**). As análises adotaram os seguintes parâmetros: K=3, “burn-in Period” = 100 000, repetições de MCMC após “burning”= 100 000, modelo = ADMIXTURE, parâmetro alfa inferido para cada população, GENSBACK = 2 e MIGRPRIOR = 0.05, e frequências alélicas correlacionadas. Os resultados apresentados na **tabela 15** foram obtidos sem informação a priori para os indivíduos das populações parentais para auxiliar no agrupamento. Desta forma, os resultados obtidos para os coeficientes de correlação de Pearson apresentados nesta tabela incluem as estimativas também para os indivíduos das populações parentais.

4.9.2 Miscigenação individual

A miscigenação no nível individual para doadores de sangue e pacientes com doença falciforme amostrados em diferentes regiões do Estado de Minas Gerais foi estimada por meio do programa computacional STRUCTURE. As análises adotaram os seguintes parâmetros: $K=3$, “burn-in Period” = 100 000, repetições de MCMC após “burning” = 100 000, informação a priori para os indivíduos das populações parentais para auxiliar no agrupamento (USEPOPINFO=1), modelo = ADMIXTURE para os doadores e pacientes, parâmetro alfa inferido para cada população, GENSBACK = 2 e MIGRPRIOR = 0.05, e frequências alélicas correlacionadas. Também nessas análises foram considerados os resultados dos 14 SNPs e 40 INDELS. A **figura 27** mostra uma representação gráfica dos resultados da análise, onde cada barra vertical representa um indivíduo.

Tabela 15 Correlação entre as estimativas de miscigenação obtidas para doadores e pacientes amostrados em diferentes regiões do Estado de Minas Gerais utilizando o programa STRUCTURE empregando diferentes conjuntos de marcadores

Conjuntos de marcadores	Estimativas de miscigenação					
	AFR		EUR		AMR	
	CC	p valor	CC	p valor	CC	p valor
Doadores e Parentais						
14 SNPs X 40 INDELS	0,96	<0,001	0,95	<0,001	1,00	<0,001
14 SNPs X 14 SNPs + 40 INDELS	0,97	<0,001	0,97	<0,001	1,00	<0,001
40 SNPs X 14 SNPs+40 INDELS	0,98	<0,001	0,99	<0,001	1,00	<0,001
Pacientes e Parentais						
14 SNPs X 40 INDELS	0,84	<0,001	0,86	<0,001	0,95	<0,001
14 SNPs X 14 SNPs+40 INDELS	0,87	<0,001	0,89	<0,001	0,96	<0,001
40 SNPs X 14 SNPs+40 INDELS	0,98	<0,001	0,99	<0,001	1,00	<0,001

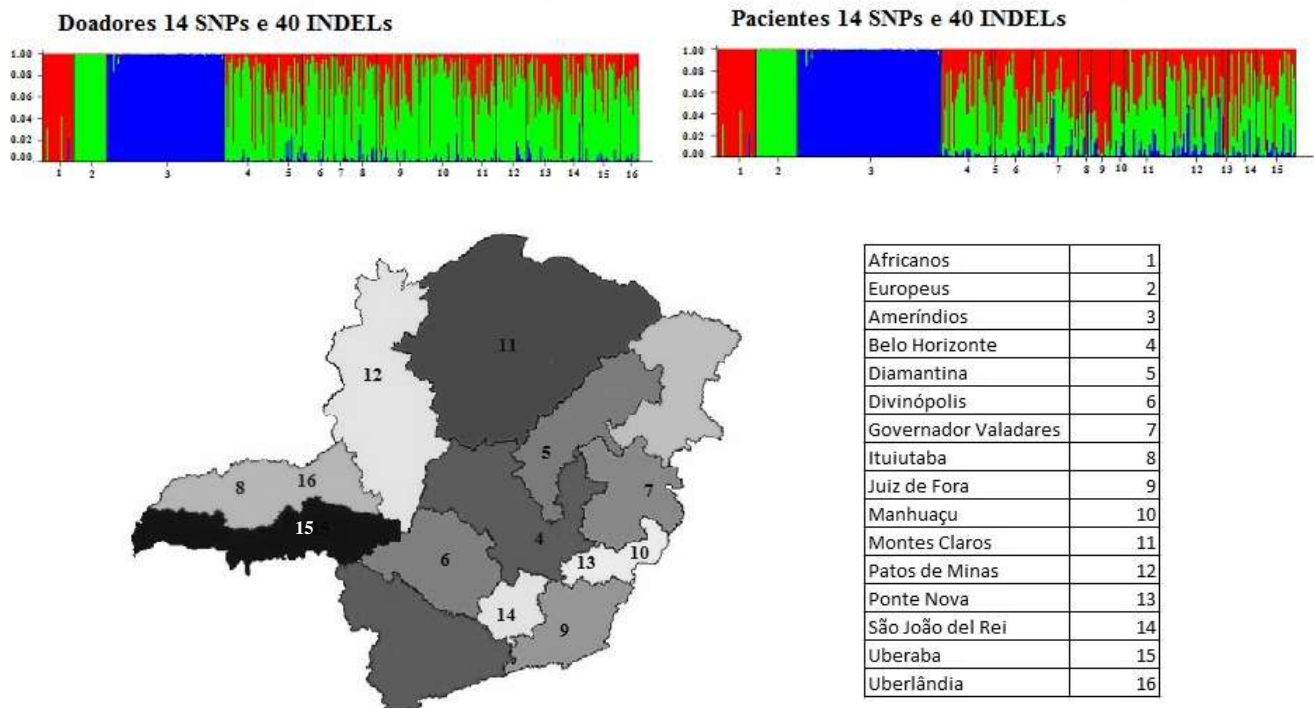


Figura 27. Análise de miscigenação individual dos doadores de sangue e pacientes com doença falciforme amostrados em diferentes localidades, utilizando o programa STRUCTURE. Cada barra vertical representa um indivíduo.

Os resultados de miscigenação no nível individual apontaram para a ocorrência de variação dos níveis de ancestralidade para doadores e pacientes dentro das diferentes regiões do estado e também entre elas, conforme apresentado na **figura 27**.

Nas análises subsequentes, os doadores e pacientes foram classificados em três grupos: 1) indivíduos com ancestralidade africana superior a 85%; 2) Indivíduos com ancestralidade europeia superior a 85%; e 3) indivíduos com proporções intermediárias de miscigenação (**tabela 16**). Esta classificação foi feita porque os indivíduos com valores mais extremos de miscigenação (superior a 85%) apresentam intervalos de probabilidade menores, produzindo estimativas mais precisas, enquanto os indivíduos com valores intermediários de miscigenação apresentam amplos intervalos de probabilidade, produzindo, portanto, estimativas menos acuradas. Assim sendo, evitou-se analisar separadamente os indivíduos, sobretudo aqueles com estimativas intermediárias de miscigenação, com o intuito de se evitar inferências inadequadas. As **figuras 28 e 29** apresentam o percentual de indivíduos em cada região do Estado de Minas Gerais, pertencentes a cada um desses três grupos. Nesses gráficos, indivíduos com proporções intermediárias de ancestralidade europeia e africana são designados como “outros”.

Tabela 16. Proporção de indivíduos amostrados em diferentes localidades, que apresentam ancestralidade africana e/ou europeia superior a 85%

	Ancestralidade europeia maior que 85%	Ancestralidade africana maior que 85%	Ancestralidade europeia ou africana maior que 85%
Doadores			
Belo Horizonte	51,6	6,5	58,1
Diamantina	12,5	4,2	16,7
Divinópolis	45,5	0,0	45,5
Ituiutaba	34,8	0,0	34,8
Juiz de Fora	30,0	6,7	36,7
Manhuaçu	56,7	3,3	60,0
Montes Claros	25,0	4,2	29,2
Patos de Minas	42,9	0,0	42,9
Ponte Nova	24,0	8,0	32,0
São João Del-Rei	40,0	0,0	40,0
Uberaba	38,5	0,0	38,5
Uberlândia	23,1	0,0	23,1
Pacientes			
Belo Horizonte	14,3	14,3	28,6
Divinópolis	14,3	9,5	23,8
Gov. Valadares	7,7	7,7	15,4
Ituiutaba	14,3	14,3	28,6
Juiz de Fora	0,0	54,5	54,5
Manhuaçu	0,0	12,5	12,5
Montes Claros	26,1	0,0	26,1
Patos de Minas	15,6	0,0	15,6
Uberaba	19,2	7,7	43,8
Uberlândia	22,7	9,1	31,8

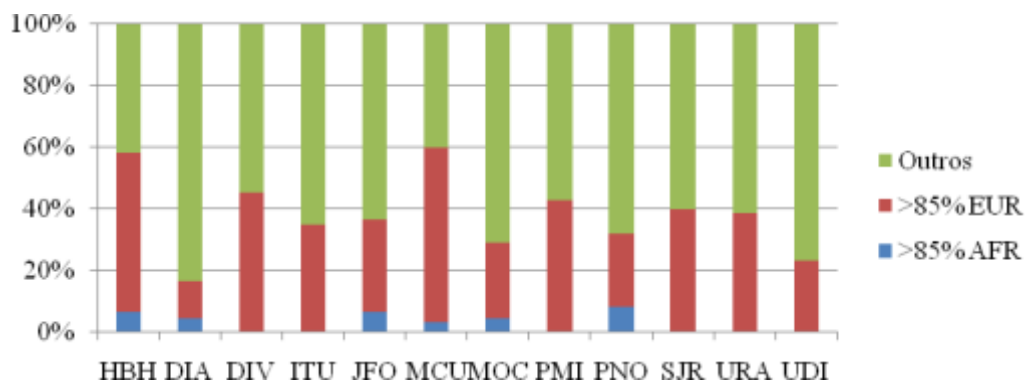


Figura 28. Proporção de doadores de sangue de diferentes regiões de Minas Gerais, com alta ancestralidade africana (>85% AFR), alta ancestralidade europeia (>85% EUR) e com proporções intermediárias de miscigenação (outros)

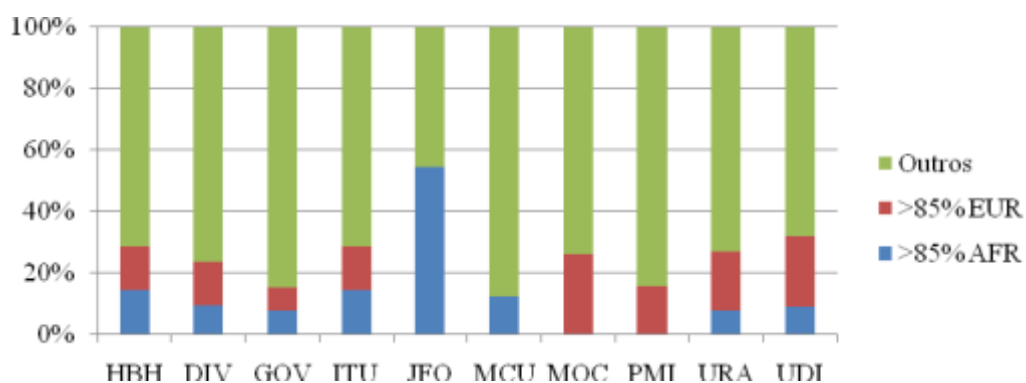


Figura 29. Proporção de pacientes com doença falciforme de diferentes regiões de Minas Gerais, com alta ancestralidade africana (> 85% AFR), alta ancestralidade europeia (> 85% EUR) e com proporções intermediárias de miscigenação (outros)

Por fim, visando a identificação de particularidades regionais relacionadas ao processo de miscigenação no nível individual, a miscigenação populacional africana e europeia, estimada com o método IG, foi comparada com o percentual de indivíduos com miscigenação africana ou europeia superior a 85% amostrados em diferentes localidades. As figuras 30 a 33 apresentam os resultados destas análises e mostram que mesmo para localidades com miscigenação populacional semelhantes, a miscigenação individual pode variar consideravelmente. Os códigos de três letras apresentados nessas figuras representam as diferentes localidades onde os sujeitos da pesquisa foram amostrados, conforme descrito na lista de abreviaturas

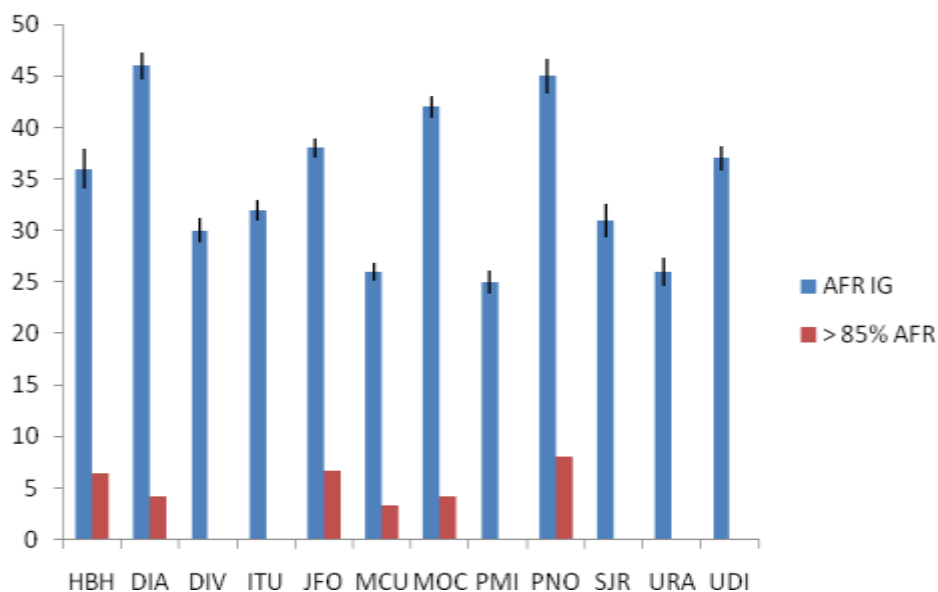


Figura 30. Miscigenação populacional africana (AFR IG) e percentual de indivíduos com ancestralidade africana superior a 85% (>85% AFR) em doadores de sangue amostrados em diferentes localidades

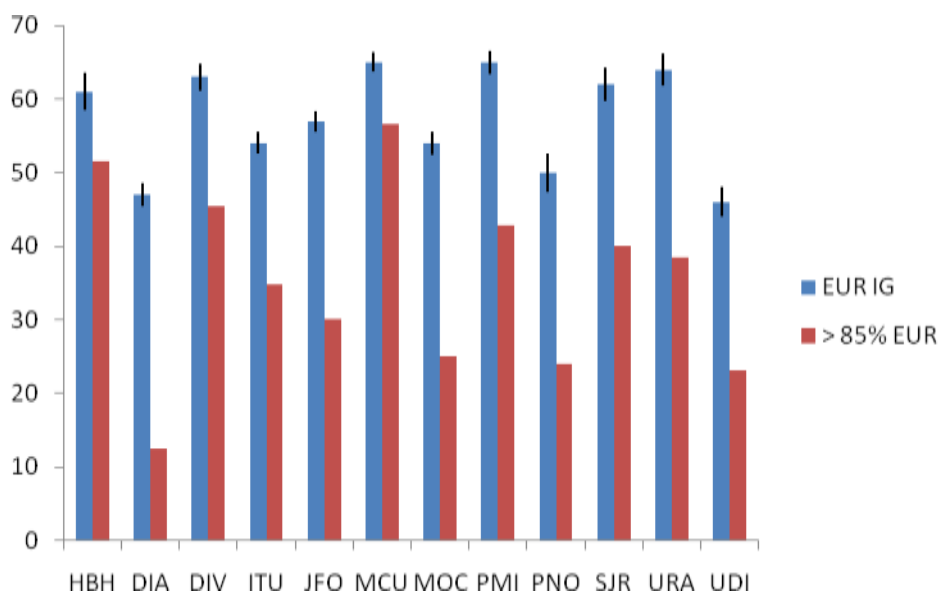


Figura 31. Miscigenação populacional europeia (EUR IG) e percentual de indivíduos com ancestralidade europeia superior a 85% (>85% EUR) em doadores de sangue amostrados em diferentes localidades

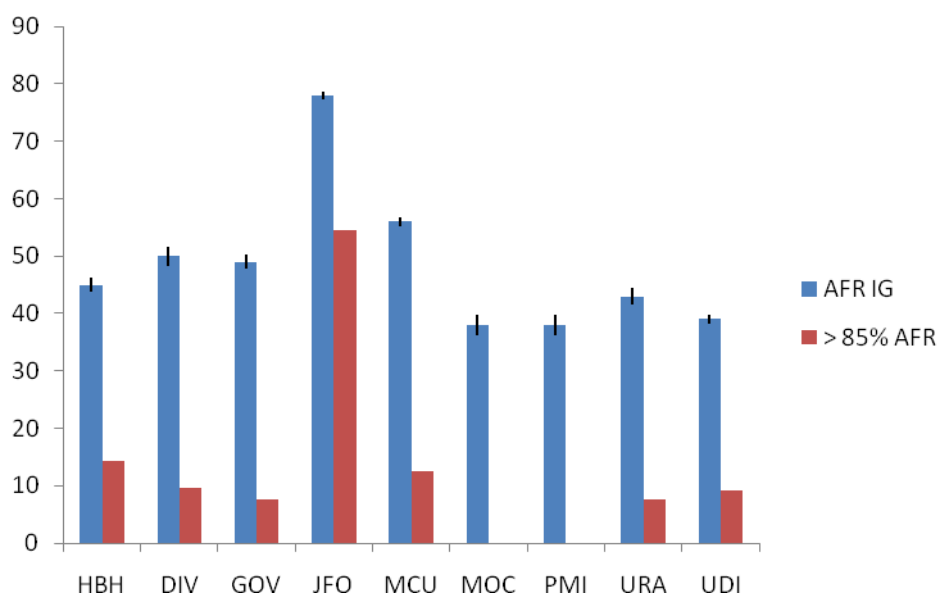


Figura 32. Miscigenação populacional africana (AFR IG) e percentual de indivíduos com ancestralidade africana superior a 85% (>85% AFR) em pacientes com doença falciforme amostrados em diferentes localidades

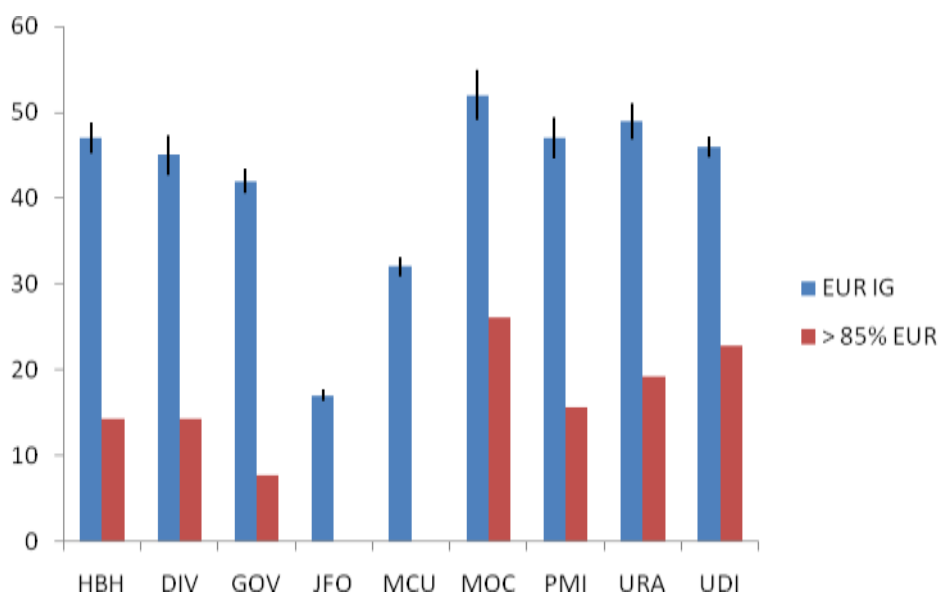


Figura 33. Miscigenação populacional europeia (EUR IG) e percentual de indivíduos com ancestralidade europeia superior a 85% (>85% EUR) em pacientes com doença falciforme amostrados em diferentes localidades

5. DISCUSSÃO

O ponto de partida para se estimar a miscigenação, quer seja no nível populacional, individual ou mesmo genômico é a seleção dos marcadores. Os primeiros estudos desenvolvidos com o intuito de estimar a miscigenação na população do sudeste brasileiro datam de 1933 e utilizaram marcadores fenotípicos como os antígenos de sistemas de grupos sanguíneos (Salzano e Bortolini, 2002). Em uma segunda fase, esses estudos fizeram ampla utilização de marcadores genéticos uniparentais como mtDNA e marcadores do cromossomo Y (Alves-Silva *et al.*, 2000; Carvalho-Silva *et al.*, 2001; Carvalho-Silva *et al.*, 2006). Mais recentemente, tais estudos passaram a adotar marcadores moleculares autossômicos como microssatélites, INDELs e SNPs (Bastos Rodrigues *et al.*, 2006; Scliar *et al.*, 2009; Pena *et al.*, 2009). Os marcadores autossômicos têm sido atualmente preferidos quando o objetivo do estudo é estimar as proporções de miscigenação, sobretudo no nível individual, já que esses marcadores podem permitir uma melhor representação do genoma.

Se a alguns anos atrás a escassez de marcadores polimórficos era um fator limitante nos estudos de miscigenação em populações humanas, atualmente, com a abundância de polimorfismos cujos dados de frequências alélicas e genotípicas em diferentes populações estão disponíveis em bancos como HapMap e dbSNP, o desafio passou a ser a escolha dos conjuntos mais adequados para atender a objetivos específicos.

Neste contexto, critérios econômicos, estatísticos e técnicos devem ser levados em consideração na construção de painéis de marcadores a serem genotipados para estudos de inferência de ancestralidade e estimativa de miscigenação (Rosemberg *et al.*, 2003).

Do ponto de vista econômico, a utilização de marcadores informativos de ancestralidade (MIAs), em relação aos selecionados aleatoriamente, é desejável para a redução dos esforços de genotipagem por permitir a diminuição do número de marcadores necessários para estimar a miscigenação (Rosemberg *et al.*, 2003; Pffaf *et al.*, 2004). Outra estratégia desejável para promover a viabilidade econômica desse tipo de estudo é a genotipagem simultânea do maior número possível de marcadores em uma mesma reação.

Um aspecto adicional, no que diz respeito aos critérios econômicos da genotipagem, é a possibilidade de redução dos custos por meio da padronização de reações “*in house*”. No caso específico das reações de minisequenciamento para a genotipagem de SNPs realizadas no presente estudo, foi alcançado menor custo por meio da estratégia de minisequenciamento monocromático (padronizada no Laboratório de Diversidade Genética Humana - LDGH), em relação aos kits comerciais. A opção por genotipar os marcadores altamente informativos em sistemas multiplex por meio de reações padronizadas “*in house*” representou uma economia de tempo e recursos, fatores que são fundamentais para a viabilidade de qualquer procedimento.

Além do aspecto financeiro, a seleção de marcadores informativos de ancestralidade envolve o aspecto estatístico, no que concerne à escolha do critério de informatividade a ser adotado. O tipo ideal de marcador para estimativa de miscigenação é aquele que apresenta diferentes alelos fixados em cada população parental. Entretanto marcadores desse tipo são escassos, o que leva à utilização de outros que, ainda que não sejam ideais, possuam um grau de informatividade adequado (Pfaff *et al.*, 2004; Rosenberg *et al.*, 2003).

No presente estudo, optou-se pela utilização da medida de informatividade de ancestralidade (Ia), descrito por Rosenberg *et al.* (2003) como um dos critérios para a seleção dos marcadores integrantes dos Painéis AFR3 e AFR10. Esses dois conjuntos foram elaborados para distinguir a miscigenação africana e europeia. No entanto, devido a problemas técnicos, o painel AFR10 foi posteriormente abandonado.

Na construção do Painel AMR4, desenhado para a avaliação de miscigenação ameríndia, a seleção de marcadores tornou-se mais limitada devido à escassez de dados publicados acerca de SNPs MIAs em populações de nativos americanos, sobretudo do Brasil. Desta forma, a seleção dos marcadores para a composição deste painel foi menos restritiva e a diferença das frequências alélicas de SNPs entre as populações parentais (δ) foi o critério de informatividade adotado (δ mínimo de 0,2). As informações relativas aos ameríndios aqui consideradas referem-se às populações Maya-Pyma (banco de dados do *SNP500Cancer*).

Na construção de painéis para estimar miscigenação, uma questão também relevante a ser observada, é o uso de marcadores bialélicos (que é o caso da maioria dos SNPs) para estimar a ancestralidade de populações miscigenadas que se originaram a partir de mais de duas populações parentais. No caso de uma população di-híbrida um bom marcador de ancestralidade é aquele que apresenta as maiores diferenças possíveis de frequências alélicas (δ) entre as duas populações parentais. Contudo, quando a população miscigenada originou-se de mais duas populações parentais (por exemplo, população miscigenada tri-híbrida), existe uma maior complexidade na escolha de marcadores. Ao empregar marcadores informativos de ancestralidades (MIAs) bialélicos para medir miscigenação tri-híbrida, tais marcadores poderão, na melhor das hipóteses, indicar a origem de um determinado alelo, discriminando uma população das duas demais, mas não poderá resolver a ambiguidade restante. Considerando por exemplo um marcador bialélico cujo alelo “a” tenha frequência igual a 1 na população A e frequência zero nas populações B e C, a genotipagem do mesmo será suficiente apenas para distinguir a ancestralidade oriunda de A, restando a ambiguidade quanto a ancestralidade B e C. Esta ambiguidade deverá ser resolvida à medida que são acrescentados à análise marcadores informativos para as outras populações parentais. No caso do presente estudo, foi construído um painel de marcadores com o intuito de distinguir a ancestralidade africana da europeia (AFR3), e um segundo painel, para contribuir na resolução da ambiguidade quanto à ancestralidade

ameríndia (AMR4), por possuir marcadores com frequências alélicas mais diferenciadas em ameríndios (**tabela 5**).

Além da informatividade, um critério estatístico adicional e bastante relevante a ser considerado na seleção de marcadores foi a independência (escolha de marcadores não ligados) objetivando evitar a redundância de informação, que tornaria a genotipagem menos eficiente.

Do ponto de vista técnico, quando se deseja genotipar conjuntamente diversos marcadores, é importante considerar a compatibilidade para genotipagem em uma mesma reação. Além disso, é desejável dispor de algum nível de flexibilidade, para evitar uma total dependência em relação a um kit comercial ou um determinado tipo de equipamento. No presente estudo, esta flexibilidade foi buscada por meio do desenho dos iniciadores de forma a permitir a genotipagem dos painéis por meio dos sistemas de minisequenciamento monocromático e policromático (kit SNaPshot, Applied Biosystems) (**figura 6**). A compatibilidade de marcadores a serem genotipados em uma mesma reação multiplex monocromática (com apenas um ddNTP fluorescente) implica principalmente em não haver polimorfismos do tipo G/C e A/T no mesmo painel. A flexibilidade possibilitada com o uso desta estratégia compensa a limitação na escolha dos marcadores.

A utilização da ferramenta bioinformática Muplex para o desenho dos iniciadores de PCR para reações conjuntas (multiplex) revelou-se uma ótima opção. Apenas três dos 30 marcadores inicialmente selecionados não foram eficientemente amplificados, em reações individuais ou conjuntas, por meio dos iniciadores desenhados por esse programa, conforme visualizado em eletroforese em géis de agarose ou poliacrilamida. Além disso, esta deficiência foi apenas aparente uma vez que, na etapa de minisequenciamento, foi possível genotipar esses loci, ou seja, os produtos amplificados, embora não estivessem em quantidade suficiente para visualização em gel, estavam presentes e aptos a servirem de molde nas reações de minisequenciamento. Estes resultados sugerem que mesmo na ausência da visualização dos produtos de PCR em géis, é conveniente testar a reação de mini-sequenciamento.

Ao final da padronização das reações de genotipagem, 14 SNPs foram genotipados com sucesso, em duas reações multiplex (**tabela 3**). O principal fator limitante observado na etapa de padronização da genotipagem por minisequenciamento foi o fato de que fragmentos pequenos (menores que 45 a 50 nucleotídeos) não podem ser visualizados no eletroferograma da genotipagem monocromática (sequenciador Megabace), embora seja possível genotipá-los com a estratégia de minisequenciamento policromático (*SNaPshot*). Este fator exigiu que alguns iniciadores fossem redesenhados, de modo a aumentar as caudas adicionadas na extremidade 5'. Entretanto, a eficiência da genotipagem ficou bastante comprometida com a utilização de fragmentos muito grandes (acima de 120 a 125 nucleotídeos) devido ao baixo sinal no eletroferograma (picos demasiadamente baixos ou inexistentes), além de exigir a utilização de iniciadores purificados por HPLC, o que aumenta o custo

das reações. A utilização de iniciadores muito longos e não purificados nesta etapa pode levar à ocorrência de picos inespecíficos gerados pela heterogeneidade de tamanho dos iniciadores durante o processo de síntese.

Conforme já relatado em estudos semelhantes (Vallone *et al.*, 2004; Carvalho e Pena, 2004), diferenças nos sinais (magnitude dos picos) para diferentes loci genotipados foram observados durante o procedimento de genotipagem (**figura 7**). As tentativas de uniformização do sinal por meio da alteração da concentração dos iniciadores não apresentaram resultados satisfatórios uma vez que levavam a melhora da genotipagem de alguns marcadores à custa da piora de outros.

A taxa média de recuperação de genótipos por locus no presente estudo foi de 90% para a totalidade de doadores, pacientes e indivíduos das populações parentais genotipados. Os únicos loci para os quais a taxa de recuperação não foi superior a 80% foram os marcadores INDELS 4, 19 e 20, para os quais a taxa de falha de genotipagem foi de 21,2, 26,1 e 32,2%, respectivamente.

Os testes para o Equilíbrio de Hardy-Weinberg realizados com o auxílio do programa GDA revelaram que, após correção de Bonferroni, o locus rs1372115 do painel de 14 SNPs em doadores de sangue amostrados em Diamantina e o locus 20 do painel de 40 INDELS em pacientes com doença falciforme amostrados em Montes Claros e na população parental Ameríndia estavam fora do equilíbrio.

Os dados apresentados na **tabela 5** mostram que os marcadores que compuseram o painel de 14 SNPs atendem razoavelmente bem os requisitos de bons marcadores informativos de ancestralidade: frequências alélicas diferentes entre os grupos parentais, porém relativamente homogêneas dentro deles. Estas características podem amenizar a escolha de populações parentais que não são iguais, embora sejam semelhantes àquelas que deram origem à população miscigenada sob investigação. Este é um fator importante já que muitas vezes as populações parentais na realidade podem consistir em uma amálgama de populações (como é o caso os africanos, que vieram para o Brasil a partir de diversas regiões da África). Em outros casos, as populações parentais sequer existem na atualidade, para que possam ser estudadas (como é o caso de várias populações de Nativos Americanos).

Em estudos de genética de populações, a estrutura genética é um termo usado para indicar a existência de diferenças nas frequências alélicas e/ou genotípicas entre duas ou mais populações. Esta estrutura é moldada ao longo do tempo pela interação de diversos fatores como seleção natural, deriva genética, mutação, migração, endogamia, efeito fundador, entre outros. Uma das formas de se avaliar a estrutura genética de uma população é o uso dos índices de fixação (estatística F de Wright, ou F_{st}) como medidas de distâncias genéticas (Lewis e Zaykin, 2002).

Os valores de F_{st} par a par obtidos para os grupos de doadores e pacientes amostrados em diferentes localidades apresentaram valores baixos (**tabelas 8 e 9**), sendo que apenas 10 de 66 valores computados para doadores e 12 de 36 valores computados para pacientes apresentaram significância

estatística (**tabelas 8 e 9**). Os maiores valores encontrados foram **0,019920** entre doadores, para o par Belo Horizonte - Uberlândia e **0,051954** entre pacientes, para o par Juiz de Fora- Montes Claros. O maior número de valores de Fst significativos foi observado para os doadores de sangue amostrados em Diamantina (7 valores de Fst significativos) e para os pacientes com doença falciforme amostrados em Juiz de Fora (8 valores de Fst significativos). Comparando estas análises com os resultados de miscigenação, pode-se observar que são justamente estes os grupos de doadores e pacientes com maior miscigenação africana.

Os baixos valores de Fst obtidos para os grupos de doadores e pacientes mostram que não há uma redução substancial da heterozigosidade nos grupos estudados. Esses achados sugerem baixa diferenciação genética entre as regiões do Estado de Minas Gerais, para doadores de sangue e pacientes com doença falciforme. Resultados de Fst baixos ou sem significância estatística entre regiões do Brasil já foram relatados (Lins, 2007), não sendo, portanto, inesperada a sua ocorrência dentro do estado de Minas Gerais.

Estes pequenos valores de Fst podem estar relacionados aos fatores históricos, sociais e culturais, que influenciaram o processo de ocupação do território no estado. Relatos históricos sugerem que o processo inicial de povoamento não foi uniforme ao longo do espaço geográfico mineiro, com a participação de diferentes populações nativas e influxo de imigrantes europeus e africanos que variaram em número e origem, dependendo da região. No entanto, eventos posteriores como migrações entre as diferentes regiões podem ter contribuído para diminuir as distâncias genéticas inter-regionais. Além disso, embora se espere que indivíduos que residem mais próximos tenham uma maior probabilidade de se reproduzirem entre si, do que com aqueles que residem em pontos mais afastados, os grupos constituídos por indivíduos (doadores e pacientes) amostrados em diferentes regiões não constituem populações naturais, mas grupos artificialmente formados com o objetivo de viabilizar as análises.

Adicionalmente, embora a totalidade dos doadores de sangue e pacientes portadores de doença falciforme amostrados no presente estudo tenha declarado que nasceu em Minas Gerais, não se pode excluir a possibilidade de que muitos deles sejam filhos ou netos de migrantes. Esta situação é particularmente provável em municípios para os quais se observou grande influxo migratório, como é o caso de Belo Horizonte, que viu sua população subir de 13.472, no ano de 1900 para 2.452.617 em 2009.

Os pequenos valores de Fst computados para doadores e pacientes possivelmente contribuíram para obtenção de resultados negativos para a existência de um padrão geográfico compatível com o modelo de isolamento por distância (teste de Mantel) em Minas Gerais (**tabela 10; figuras 10 e 11**). Desta forma, a ausência de adaptação dos dados genéticos ao modelo IBD indica que não há um decréscimo da similaridade genética entre populações à medida que as distâncias geográficas entre elas

umentam, embora a ocorrência de outros padrões mais complexos de estruturação geográfica da diversidade genética não tenha sido descartada.

Um procedimento muito importante a ser adotado quando se pretende estudar a miscigenação ou estrutura populacional utilizando um número limitado de marcadores (no caso, o painel de 14 SNPs) é testar previamente a precisão dos resultados fornecidos por esse painel. Isto pode ser obtido por meio da genotipagem dos marcadores sob análise em um conjunto de amostras de referência, para as quais existem dados de genótipos disponíveis para um grande número de marcadores. Esta mesma estratégia foi usada por Bastos-Rodrigues (2006) para avaliar o painel de 40 INDELS em estudos de estrutura genética de populações humanas em nível mundial, por meio das amostras de DNA do painel de referência do HGDP-CEPH (Cann *et al.*, 2002).

No presente estudo o desempenho do painel de 14 SNPs foi avaliado por meio da comparação entre as estimativas de miscigenação individual e populacional obtidas com este conjunto de marcadores, e os resultados obtidos com a utilização de um painel de 108 MIAs (os quais assumimos fornecerem estimativas mais precisas). Para isto, foram empregadas as amostras de indivíduos miscigenados denominados como Hispânicos do painel do projeto SNP500Cancer. Estas amostras, assim como as dos africanos, europeus e ameríndios considerados como populações parentais, são uma referência apropriada para a finalidade descrita, uma vez que estão disponíveis como células imortalizadas no repositório Coriell (www.coriell.org). Este repositório pode fornecer amostras de DNA de boa qualidade, para reproduzir ou mesmo estender os resultados obtidos no presente estudo. Células imortalizadas, que podem ser usadas com propósito semelhante, são encontradas em painéis como os do CEPH-HGDP (Cann *et al.*, 2002), o SNP500Cancer (Packer *et al.*, 2006) e o HapMap (The International HapMap Consortium 2007).

A análise dos indivíduos das populações parentais (africanos, europeus e ameríndios) genotipados para o painel de 14 SNPs por meio do programa STRUCTURE revelou que quase todos os indivíduos foram corretamente atribuídos a suas populações de origem (**figura 12**), com base exclusivamente em dados genéticos. No entanto, conforme foi evidenciado em análises subseqüentes, este não é um preditor eficiente para a acurácia de um painel em estimar a miscigenação individual, principalmente naqueles casos em que as taxas de miscigenação são elevadas.

Ao testar o painel de 14 SNPs, em relação a um maior conjunto de marcadores (108 SNPs), utilizando exatamente os mesmos indivíduos tanto das populações parentais, quanto da miscigenada foi possível verificar que este painel é adequado para medir a miscigenação européia, mas pode superestimar a miscigenação africana e subestimar a miscigenação ameríndia (**tabela 11**). A introdução de mais indivíduos na composição da população parental ameríndia (Ashaninkas e Shimaa) foi capaz de alterar as estimativas de miscigenação populacional para os hispânicos, obtidas com o painel de 14 SNPs, de modo a tornar a miscigenação ameríndia mais precisa. Entretanto, esta mesma

alteração fez com que a miscigenação européia passasse a ser subestimada, enquanto a miscigenação africana continuou a ser superestimada, só que agora em menor escala. Esta tendência de superestimativa da miscigenação africana, em comparação com os resultados obtidos com os 108 SNPs, foi uma característica observada nas análises utilizando o painel de 14 SNPs e que persistiu mesmo após a ampliação do número de marcadores, com a incorporação do painel de 40 INDELS (**figuras 13, 14 e 15**).

Em geral, os métodos estatísticos para estimativa de miscigenação assumem que não há erro na escolha dos grupos parentais ou em suas frequências gênicas. No caso dos estudos conduzidos no Brasil, a adequação a esta suposição é particularmente difícil no que diz respeito à população parental ameríndia (Scliar, 2009). Os grupos nativos da América do Sul constituem as populações mais heterogêneas de todos os continentes, apresentando uma grande diferenciação nas frequências alélicas, em decorrência de fatores como a deriva genética causada pelos pequenos tamanhos efetivos dos grupos, e a ocorrência de efeitos fundadores ao longo da história dessas populações (Mesa *et al.*, 2000). No Brasil, e especificamente em Minas Gerais, a obtenção de amostras representativas da população parental ameríndia é dificultada, já que a maior parte dos grupos está extinto ou apresenta miscigenação com outras populações. Desta forma, a utilização das amostras de nativos americanos do Peru como população parental ameríndia, embora não seja em teoria uma escolha perfeita, foi a melhor opção disponível.

As estimativas de miscigenação populacional africana e européia para os Hispânicos, conduzidas com os métodos de identidade genética e tempos de coalescência, apresentaram resultados bastante concordantes entre si (**tabela 11, figuras 13 e 14**). Dentre os métodos desenvolvidos para se estimar a miscigenação no nível populacional, os que utilizam as abordagens de Identidade Genética e Tempos de Coalescência estão entre os mais citados na literatura nos últimos 20 anos (Choisy *et al.*, 2004).

Segundo testes comparativos entre métodos desenvolvidos com o intuito de se estimar a miscigenação (Choisy *et al.*, 2004), o método de IG apresenta bom desempenho para populações miscigenadas apresentando proporções centrais de miscigenação (ou seja próximos de 50% de contribuição de cada uma entre duas populações parentais), produzindo, nesses casos, estimativas bastante precisas. Quando as proporções de miscigenação são mais baixas (menos de 10% de contribuição de uma das populações parentais) ou o evento de hibridização tem idade superior a 100 gerações, observa-se uma queda da qualidade das estimativas obtidas com esta abordagem. No caso dos doadores de sangue e pacientes, para os quais a miscigenação iniciou-se a não mais que 15 gerações, as populações parentais são relativamente diferenciadas, e as taxas de miscigenação são altas, este método é uma boa escolha, podendo fornecer estimativas confiáveis. Esta opção é

corroborada por outros trabalhos que ponderaram acerca da melhor escolha do método estatístico para estimativa de miscigenação (Salzano e Bortolini, 2002).

Ainda de acordo com os resultados de comparação entre métodos (Choisy *et al.*, 2004), a abordagem TC apresenta um dos melhores desempenhos em diferentes condições de análise (mesmo as adversas), o que torna também esta uma ótima opção. Por fim, além dos métodos IG e TC serem amplamente utilizados em estudos em todo o mundo, testes comparativos de desempenho não foram capazes de demonstrar superioridade de um em detrimento do outro em toda e qualquer circunstância (Choisy *et al.*, 2004, Excoffier *et al.*, 2005).

Já o método implementado no programa STRUCTURE, para o qual a miscigenação populacional é calculada pela média das estimativas individuais, observou-se superestimativa da miscigenação europeia e subestimativa da miscigenação africana, quando comparado aos outros métodos (**tabela 11**).

As estimativas de miscigenação no nível individual realizadas para os Hispânicos com os 108 MIAs e subconjuntos menores de SNPs selecionados aleatoriamente a partir desses 108 marcadores, evidenciaram que utilizando subconjuntos de 50 ou 80 SNPs, a correlação das estimativas de miscigenação individual em relação às obtidas com os 108 MIAs é maior que 88%, enquanto com menos de 20 MIAs, a correlação é menor que 66% (**tabela 12**). Desta forma, ainda que o conjunto de 14 MIAs contenha informação para estimar a miscigenação no nível populacional, as inferências de miscigenação no nível individual não são precisas com menos de 20 marcadores, sobretudo para os indivíduos mais miscigenados. Podemos afirmar portanto que, conforme já previsto para pequenos painéis de marcadores, o painel de 14 SNPs padronizado no presente estudo não fornece estimativas adequadas da miscigenação no nível individual.

A ampliação do número de marcadores genotipados, com a introdução do painel de 40 INDELs padronizado por Bastos-Rodrigues *et al.* (2006) resultou na melhoria das estimativas de miscigenação individuais, quando comparado com as estimativas obtidas com 108 MIAs e, portanto, as análises subsequentes de miscigenação foram realizadas com os conjuntos de 14 SNPs e 40 INDELs (**tabela 12**).

Na expectativa de melhorar ainda mais estes resultados, um terceiro painel, composto por 10 INDELs específicos para detectar a miscigenação ameríndia foi incorporado nas análises. Contudo o que se observou foi a diminuição dos valores para a correlação em relação aos resultados obtidos com 108 MIAs (que assumimos serem mais precisos) (**tabela 12**). Uma possível explicação para este acontecimento é a introdução de um viés. Por esse motivo, as análises de estrutura populacional e estimativas de miscigenação ulteriores foram realizadas apenas com os conjuntos de 14 SNPs e 40 INDELs.

Outra evidência da baixa precisão das estimativas de miscigenação individual com a utilização de um número reduzido de marcadores emergiu quando os intervalos bayesianos de probabilidade de 90% (90%-PI) para a miscigenação individual gerados pelo STRUCTURE foram plotados contra as estimativas de miscigenação individual (**Figura 16**). Esta análise sugere que, sobretudo para indivíduos com altos níveis de miscigenação (contribuição de uma das populações parentais superior a 20%), os intervalos de probabilidade se elevam de forma pronunciada quando o número de marcadores utilizados para a inferência da miscigenação individual é inferior a 50.

Resultados semelhantes, ou seja, amplos intervalos de probabilidade, resultando em estimativas de miscigenação individual pouco precisas para indivíduos miscigenados ou híbridos foram relatados em estudos envolvendo diversas outras espécies como o peixe lúcio *Esox lucius* (Larsen *et al.*, 2005), o bacalhau do Atlântico *Gadus morhua* (Nielsen *et al.*, 2003), gatos silvestres da Europa *Felis silvestris* (Pierpaoli *et al.*, 2003), pequenos felídeos do Brasil como *Leopardus geoffroyi* e *L. tigrinus* (Trigo *et al.*, 2008), e a espécie de cobra *Sistrurus catenatus*.

Uma vez concluída a etapa de testes para avaliação das potencialidades e limitações dos painéis de marcadores utilizados, partiu-se para a análise dos sujeitos de pesquisa especificamente amostrados para o presente projeto: doadores de sangue e pacientes com doença falciforme.

A miscigenação no nível populacional para doadores e pacientes amostrados em diferentes regiões do Estado de Minas Gerais, considerando-se o conjunto de 54 marcadores (14 SNPs e 40 INDELs), foi estimada por meio dos métodos de identidade genética (Chakraborty, 1985) tempos de coalescência (Dupanloup e Bertorelle, 2001), com o auxílio dos programas ADMIX95 e ADMIX2, respectivamente, assim como por meio do método Bayesiano implementado no programa STRUCTURE (Pritchard *et al.*, 2000).

Os resultados das análises de miscigenação populacional pelo método IG (**tabela 13**) evidenciaram que, de modo geral, com a utilização de 14 SNP e 40 INDELs, doadores apresentaram menor miscigenação africana (34,1%) que os pacientes (45,5%). Esta diferença era esperada uma vez que, embora o alelo associado à doença falciforme esteja presente na África sub-Saariana, África do Norte, Grécia, Itália, Oriente Médio, Península Arábica, Índia e China (Pena, 2008), a origem deste alelo no Brasil tem origem quase exclusivamente africana, tendo sido introduzido pela imigração forçada dos escravos no período colonial. Desta forma, aquele trecho específico do genoma onde se localiza a mutação possui ancestralidade africana com grande probabilidade para todos os pacientes, o que pode não ocorrer para os doadores. Entre os pacientes, os homozigotos (portadores de anemia falciforme) apresentaram miscigenação africana (47,3%) mais elevada do que os heterozigotos (40,0%) para a mutação (portadores de outras doenças falciformes), o que era igualmente esperado.

Estudos prévios, porém relativamente recentes, acerca da miscigenação em Minas Gerais apresentam resultados diversificados:

- 1) Entre indivíduos mineiros auto-declarados como “brancos” (n=142) foi demonstrada, por meio da genotipagem de 40 polimorfismos de inserção-deleção (os mesmos utilizados no presente estudo), a ocorrência de 16,1% de ancestralidade africana, 70,8% de ancestralidade europeia, e 13,1% de ancestralidade ameríndia com a utilização do programa STRUCTURE (Pena *et al.*, 2009).
- 2) Em indivíduos da comunidade rural de Queixadinha, município de Carai, Vale do Jequitinhonha (n=173), categorizados em três grupos (brancos, pardos e preto), com base em características fenotípicas, os resultados obtidos a partir da análise de 10 MIAs mostraram que, no nível populacional, a proporção de ancestralidade africana foi de 51% para o grupo dos indivíduos classificados como pretos, 44% para o grupo dos pardos, 32% para o grupo dos brancos e 44% para a população total (Parra *et al.*, 2003).
- 3) Na cidade de Belo Horizonte, capital do estado, com base na genotipagem de 13 microsátélites de uso forense em 234 indivíduos não relacionados e oriundos do banco de dados de testes de paternidade de um laboratório particular, foi estimada uma contribuição africana de 34%, europeia de 65%, e ameríndia de 1%, com a utilização do método estatístico IG (Scliar *et al.*, 2009).

Os doadores de sangue da região metropolitana de Belo Horizonte possuem um perfil sócio-econômico conhecido, sendo este grupo composto predominantemente por indivíduos de menor poder aquisitivo e predominância de “cor” morena, mulata ou parda, em consonância com as características gerais da população brasileira (Souza e Silva, 2000; Brener *et al.*, 2008, Vertchenko, 2005). Ainda assim, os resultados de miscigenação encontrados para doadores amostrados em Belo Horizonte (36% de contribuição africana, 61% de contribuição europeia e 3% de contribuição ameríndia) não foram discrepantes dos descritos por Scliar *et al.* (2009) para amostras deste mesmo município, oriundos do banco de dados de testes de paternidade de uma instituição privada, e também obtidos pelo método de IG. Estes achados chamam a atenção uma vez que a princípio se esperaria uma maior ascendência europeia para a amostra proveniente de testes de paternidade, a qual abrigaria indivíduos com poder aquisitivo mais elevado. Esta conjuntura, a qual na realidade não foi constatada, seria esperada devido a fatores históricos e socio-econômicos já que no Brasil, as classes com maior poder econômico são predominantemente constituídas por indivíduos “brancos”, que supostamente possuem uma maior ancestralidade europeia, enquanto nas classes de menor poder econômico, predominam “pardos” e “pretos” (IBGE). Além disso, embora não seja possível a realização de uma comparação rigorosa, os resultados encontrados para miscigenação por meio de marcadores moleculares não apresenta discordância, em relação aos dados de composição da população obtidos por auto-declaração no ano de 2006 (IBGE) para a região metropolitana de Belo Horizonte: 42,3% de brancos, 10,3% de pretos, 46,8% de pardos e 0,6% de amarelos ou indígenas.

Para o grupo de doadores amostrados em Diamantina, os resultados de ancestralidade africana (46%) foram semelhantes àqueles observados em Queixadinha (44%) tanto para a população geral

quanto para o grupo de indivíduos classificados como pardos. Esta comparação é pertinente uma vez que tanto Diamantina quanto Queixadinha localizam-se na mesma região do estado (Vale do Jequitinhonha).

Quanto às estimativas de miscigenação descritas por Pena *et al.* (2009) para os mineiros classificados como brancos por meio de auto-declaração, utilizando o programa STRUCTURE, os resultados mais próximos encontrados no presente estudo fora aqueles obtidos para os doadores amostrados em Manhuaçu e Patos de Minas (ancestralidade africana de 17%, europeia de 80% e ameríndia de 3%), embora nenhuma das amostras de doadores tenha apresentado miscigenação ameríndia tão elevada (13%) quanto à descrita por Pena *et al.* (2009), com esta metodologia de análise (método bayesiano implementado no programa STRUCTURE). Entretanto, possíveis inferências elaboradas a partir desta comparação possuem valor limitado, já que o perfil das amostras utilizadas nesses estudos é bastante distinto.

Para as demais localidades do Estado de Minas Gerais onde houve amostragem de doadores e pacientes, não há dados de literatura específicos, referentes a estimativas de miscigenação, que possibilitem comparações adicionais.

Por outro lado, outras fontes de dados que permitem comparações suplementares são os trabalhos que estimaram a miscigenação em amostras da região sudeste do Brasil, onde está situado o estado de Minas Gerais. Callegari-Jacques *et al.* (2003), analisando um grupo de indivíduos amostrados em Campinas, estado de São Paulo com a utilização de 12 marcadores STR e o método estatístico IG, encontraram ancestralidade africana de 20% ($\pm 5,1$), europeia de 73% ($\pm 4,5$) e ameríndia de 7% ($\pm 2,8$). Lins (2007), analisando uma amostra de 40 indivíduos da região sudeste com 28 SNPs marcadores informativos de ancestralidade com o programa STRUCTURE encontrou ancestralidade africana de 12,9% e europeia de 78,8% considerando três populações parentais ($K=3$) e miscigenação africana de 18,8% e europeia de 81,8% considerando duas populações parentais ($K=2$).

A comparação desses resultados de miscigenação com os obtidos no presente estudo indicam que as estimativas apresentadas por Callegari-Jacques *et al.* (2003) apresentam maior componente europeu e menor componente africano do que os observados para os doadores de sangue de Minas Gerais também obtidos pelo método de análise IG. A maior ancestralidade africana encontrada para os doadores de sangue de Minas Gerais pode indicar que os resultados obtidos apresentam um viés em direção à superestimativa da miscigenação africana e subestimativa da miscigenação europeia. Esta hipótese está em concordância com os resultados das análises realizadas para os Hispânicos, já que também nessas análises a miscigenação africana foi superestimada e a europeia subestimada com o uso do conjunto de 14 SNPs e 40 INDELS, em relação aos resultados obtidos com 108 MIAs. Por outro lado, os resultados obtidos por Callegari-Jacques *et al.* (2003) podem também apresentar um viés em direção oposta: superestimativa da miscigenação europeia e subestimativa da miscigenação africana,

conforme discutido pelos próprios autores, já que comparações com estudos anteriores (Salzano e Borrtolini, 2002), conduzidos com dados de grupos sanguíneos e proteínas estimaram menores contribuições européias e maiores contribuições africanas. Além disso, vale ressaltar que 1) as amostras usadas no estudo de Callegari-Jacques *et al.* (2003) são provenientes de bancos de testes de paternidade, podendo, portanto apresentar viés em direção a um maior componente europeu, devido ao fator sócio-econômico, conforme anteriormente discutido; 2) outros estados importantes da região sudeste possuem população auto-declarada como branca mais elevada do que o estado de Minas Gerais (46,2%), como São Paulo (67,9%) e Rio de Janeiro (54,6%), o que pode contribuir para o maior componente europeu detectado por Callegari-Jacques *et al.* (2003) para o sudeste brasileiro; 3) os estudos de Callegari-Jacques *et al.* (2003) e Scliar *et al.* (2009) foram publicados com seis anos de intervalo e, durante este período, o perfil dos clientes de testes de paternidade mudou consideravelmente, devido ao barateamento do exame e facilitação das condições de pagamento (Malta, 2010, comunicação pessoal).

Em vista de todas essas comparações, podemos dizer que não é possível descartar a existência de um viés para os resultados das estimativas de miscigenação dos doadores de sangue em direção à superestimativa da miscigenação africana. Entretanto, a comparação dos resultados obtidos no presente estudo com aqueles conduzidos com amostras coletadas em regiões específicas de Minas Gerais como Belo Horizonte (Scliar *et al.*, 2009) e Vale do Jequitinhonha (Parra *et al.*, 2003) apresentaram resultados bastante concordantes. Este fato pode ser um indício de que os valores reportados por nós podem na verdade estar próximos da realidade. Vale ainda ressaltar que as variações entre os resultados dos diferentes estudos podem refletir outros aspectos envolvidos nos processos de estimativa de miscigenação como a heterogeneidade das amostras analisadas e as diferentes naturezas dos painéis de marcadores empregados.

No que diz respeito às proporções de miscigenação populacional, obtidas para os pacientes com doença falciforme, não há dados de literatura que permitam uma comparação aceitável, já que não há relatos para a miscigenação deste tipo de pacientes no Brasil, utilizando análises genéticas.

Os resultados obtidos para a totalidade dos sujeitos de pesquisa investigados nos permitem dizer que foram constatadas diferenças significativas para as proporções de miscigenação entre os grupos, tanto de doadores, quanto de pacientes, amostrados em diferentes regiões do estado de Minas Gerais (**figuras 19 a 24**).

A partir desta observação, é possível deduzir que estudos com amostragem restrita a um único município ou região do estado podem não fornecer resultados passíveis de serem extrapolados para todo o território mineiro e tampouco para a região sudeste do país. Isto se dá porque o Estado de Minas Gerais possui grande extensão territorial com área equivalente à de países como a França, e porque sua ocupação foi bastante heterogênea.

A análise dos resultados de miscigenação de doadores e pacientes em relação às coordenadas geográficas do local da amostragem (latitude e longitude) mostraram que a latitude se correlaciona negativamente com a ancestralidade africana ($cc=-0,810$, $p=0,008$) e positivamente com a ancestralidade europeia ($cc=0,766$, $p=0,016$) para os pacientes enquanto a longitude se correlaciona negativamente com a miscigenação ameríndia ($cc= -0,763$, $p= 0,004$) apenas para os doadores (**tabela 14**).

Como os valores de latitude são negativos, decrescendo em direção ao sul, os resultados dessas análises significam que quanto mais ao sul está o local de amostragem (menor latitude) maior é a ancestralidade populacional africana e menor é a europeia, para os pacientes com doença falciforme (**tabela 14, figura 25**). Esta tendência é oposta à descrita para a miscigenação no âmbito do território brasileiro, em sua totalidade, onde em geral, se observa um incremento em direção Norte → Sul da contribuição europeia (Alves-Silva *et al.*, 2000; Parra *et al.*, 2003; Callegari-Jacques *et al.*, 2003). Isto sugere que os pacientes portadores de doença falciforme possuem características peculiares no que diz respeito aos padrões geográficos de miscigenação, e que resultados obtidos com outras populações não podem ser extrapolados para esse grupo.

A correlação entre longitude e miscigenação ameríndia encontrada para os doadores de sangue, com maior concentração de ancestralidade ameríndia no triângulo mineiro (**tabela14; figura 26**), não tem paralelo na literatura. Os povos indígenas que habitam atualmente o estado de Minas Gerais localizam-se principalmente nas regiões norte e nordeste Os estados fronteiriços do triângulo mineiro (São Paulo e Goiás) também não apresentam população ameríndia expressiva a ponto de se esperar alguma influência sobre a região (IBGE). Desta forma, é possível que a maior proporção ameríndia encontrada nesta região seja um reflexo de eventos de miscigenação mais antigos, entre os povos nativos e os colonizadores.

O estabelecimento de uma relação entre os resultados de estimativas de miscigenação encontrados no presente estudo e os aspectos históricos envolvidos na ocupação do território mineiro não é uma tarefa simples. O processo de povoamento do estado de Minas Gerais foi extremamente heterogêneo, envolvendo a participação de povos nativos e estrangeiros, cujos padrões demográficos variaram substancialmente ao longo do tempo e do espaço geográfico (Souza, 2009).

A descoberta da existência de ouro em abundância no estado de Minas Gerais atraiu centenas de milhares de imigrantes europeus e brasileiros na primeira metade do século XVIII. Com o declínio da mineração, na segunda metade do século XVIII e a conseqüente desestabilização da economia local, a entrada de africanos diminuiu, e a diversidade regional tornou-se mais acentuada. As regiões sul e sudeste (fronteira com São Paulo e Rio de Janeiro) tornaram-se importantes áreas cafeicultoras. Os antigos centros mineradores, onde se localizavam os maiores e mais antigos centros urbanos da província, permaneceram como centros de artesanato, comércio e agricultura. Já no norte e no nordeste

onde se observavam baixas densidades populacionais, predominou a pecuária extensiva (Luna e Klein, 2004).

As particularidades envolvidas no povoamento de regiões, e mesmo de municípios, resulta em um cenário extremamente complexo, no qual generalizações e simplificações podem implicar em suposições inadequadas. Assim sendo, optou-se por destacar apenas aspectos específicos do processo de miscigenação envolvido na formação da população atual do estado, que são ilustrados pelos resultados obtidos.

Uma das observações mais interessantes, que emergiu da análise dos resultados obtidos no presente estudo frente aos aspectos históricos envolvidos no povoamento de estado é a aparente falta de correlação entre a proporção de ancestralidade africana das populações regionais atuais e número de escravos africanos que habitavam estas regiões no passado.

No período colonial brasileiro o estado de Minas Gerais foi um dos que mais recebeu escravos africanos. Por volta de 1830, a população de Minas Gerais era formada por 30% de escravos. Na região da Zona da Mata mineira, o contingente de escravos era particularmente alto, compondo cerca de 40% da população total (Luna e Klein, 2004).

A imigração europeia em Minas Gerais nos séculos XIX e XX não foi tão expressiva quanto à observada em estados como São Paulo e Rio de Janeiro. O censo de 1900 mostra que, da população total do estado, apenas 4% era formada por estrangeiros, (Monteiro, 1973). Entretanto, os fluxos migratórios europeus atingiram proporções significativas em algumas regiões específicas de Minas Gerais. Boa parte dos imigrantes europeus foi direcionada à Zona da Mata, graças à expansão da lavoura cafeeira (Botelho *et al.*, 2007), o que teria contribuído para a elevação da proporção de ancestralidade europeia na região. Os dados de miscigenação para doadores de sangue amostrados nesta região mostram uma elevada contribuição europeia para os doadores de Manhuaçu (65,0%), não obstante Juiz de Fora e Ponte Nova, também localizados na Zona da Mata, tenham apresentado miscigenação europeia mais modesta (56,6% e 50,0%, respectivamente) (**tabela 13**).

Montes Claros, por sua vez, é um município situado no norte do estado de Minas Gerais. O povoamento desta região se deu inicialmente por meio dos criadores de gado da região nordeste do Brasil (principalmente da Bahia) e dos bandeirantes de São Paulo, a partir do final do século XVII (Mata-Machado, 1991). Levantamentos realizados em 1830 mostram que a região norte-nordeste apresentava a menor proporção de escravos dentre as regiões de Minas Gerais, equivalente a 22% (Luna e Klein, 2004). Em Montes Claros, entre 1833 e 1835 esta proporção era de 15% (Botelho, 1994). Apesar deste contingente relativamente reduzido de escravos, a miscigenação africana entre os doadores de sangue amostrados em Montes Claros é a terceira maior dentre as localidades do estado estudadas, provavelmente devido à proximidade geográfica e influência historicamente documentada

sobre esta região, do estado da Bahia (Prado, 2006), que é atualmente o estado brasileiro com maior proporção de pessoas que se auto-declaram como “pretos” (IBGE, 2007).

Desta forma, os resultados obtidos sugerem que, no estado de Minas Gerais, dados históricos e demográficos a respeito da população escrava Africana podem não serem bons preditivos na ancestralidade africana das populações regionais urbanas atuais. Múltiplos fatores podem ter contribuído para o delineamento deste cenário.

Além da ocorrência do influxo migratório posterior de europeus, bem como de brasileiros de ascendência européia, a redução do componente africano ao longo do tempo em certas regiões pode estar relacionado ao baixo índice de reprodução natural e à alta taxa de mortalidade, observados para os escravos do Brasil, decorrentes das más condições de vida às quais estavam submetidos. Embora haja indícios de crescimento natural positivo da população escrava em algumas áreas e períodos específicos, é provável que, dos 4 milhões de escravos que foram trazidos para o Brasil, grande parte tenha morrido sem deixar descendentes.

No que diz respeito à população ameríndia, o documento “Marco de Referência – Povos indígenas em Minas Gerais”, publicado em 2008 pela Secretaria de Estado de Planejamento e Gestão relata que os povos indígenas do Estado de Minas Gerais oficialmente reconhecidos encontram-se divididos em 08 grupos: Xakriabá, Pankararu, Aranã, Maxakali, Kaxixó, Pataxó, Krenak, e Xukuru-Kariri, distribuídos pelo território mineiro. A população ameríndia atual residente em Minas Gerais varia conforme fonte consultada. De acordo com o IBGE, 48.720 pessoas residentes em Minas Gerais identificavam-se como indígenas no ano de 2002 (0,27% da população total do estado e 6,68% da população indígena do Brasil). Já de acordo com a FUNAI, que considera apenas índios que residem em reservas, a população indígena de Minas Gerais seria em torno de 7.500 pessoas.

Sendo assim, a pequena contribuição ameríndia observada para praticamente todas as regiões do estado vai de encontro aos registros demográficos e históricos. As populações nativo-americanas foram dizimadas a partir da chegada dos primeiros colonizadores, tanto pela violência (escravização, conflitos e expulsão das terras onde originalmente viviam), quanto em decorrência do contágio de doenças infecciosas.

Embora uma substancial miscigenação entre colonizadores e ameríndios, ocorrida nos primórdios da ocupação, possa ser inferida a partir da observação da presença de linhagens do DNA mitocondrial de origem indígena em cerca de um terço dos indivíduos do sudeste brasileiro (Alves-Silva *et al.*, 2000; Pena *et al.*, 2009), o processo subsequente de extermínio provavelmente fez com que a contribuição ameríndia fosse perdendo importância no decorrer do processo de miscigenação que se deu em Minas Gerais.

A etapa subsequente das análises realizadas no presente estudo incluiu as estimativas de miscigenação no nível individual, obtidas com o programa STRUCTRE, empregando os conjuntos de

14 SNPs e 40 INDELS (**Figura 27, 28 e 29**).

Com o intuito de evitar inferências inadequadas, análises envolvendo os sujeitos de pesquisa com proporções intermediárias de miscigenação foram evitadas nesta etapa das análises. Esta precaução se deu em decorrência das grandes amplitudes dos intervalos de probabilidade associados às estimativas de miscigenação individuais calculadas pelo STRUCTURE.

O método bayesiano implementado no programa STRUCTURE, que atribui indivíduos a populações usando informação de loci não ligados ganhou grande popularidade nos estudos que visam estimar a miscigenação no nível individual (Manel *et al.*, 2005). As estimativas individuais de ancestralidade possuem grande relevância em estudos evolutivos, de conservação da vida silvestre e de epidemiologia genética (Chikhi *et al.*, 2002; Chakraborty e Weiss, 1986; Wang, 2006). Entretanto, a precisão das estimativas de miscigenação obtidas com o uso do STRUCTURE depende de vários fatores como o número de indivíduos analisados, a quantidade de miscigenação e a extensão da diferença das frequências alélicas entre as populações parentais (Pritchard *et al.*, 2000).

Outras limitações das estimativas de ancestralidade individual advêm de fatores como o uso de números limitados de marcadores e painéis não validados, as limitações estatísticas dos métodos de estimativa e a suposição de que as populações atuais de fato representam as populações parentais.

Dentre os fatores limitantes mencionados, a construção do painel de marcadores é aquele que está mais ao alcance dos pesquisadores. Embora já tenha sido demonstrado que pequenos painéis de marcadores (24 SNPs) podem ser usados para averiguar a origem continental de indivíduos e para corrigir a estratificação em populações miscigenadas (Kosoy *et al.*, 2009), o número de marcadores necessários para se inferir a ancestralidade e estimar a miscigenação não é fixo, dependendo inclusive de características da própria população que está sendo estudada.

No presente estudo, optou-se por construir um pequeno painel de marcadores informativos de ancestralidade e testar suas limitações a partir de amostras para as quais havia dados disponíveis de um grande número de marcadores. Adicionalmente, foi utilizado um painel de 40 INDELS, o qual já havia sido validado para atribuir indivíduos a diferentes populações continentais (Bastos-Rodrigues *et al.*, 2006) e também já tinha sido utilizado em estimativas de miscigenação na população brasileira (Pena *et al.*, 2009).

A escolha dos parâmetros utilizados no programa STRUCTURE para as estimativas individuais de miscigenação foi realizada empiricamente, usando como uma referência as condições empregadas por Kosoy *et al.* (2009), já que este estudo também envolveu populações miscigenadas latino americanas e também avaliou o desempenho de pequenos painéis de marcadores (24 SNPs). Ademais, a maioria dos artigos publicados omite detalhes dos parâmetros utilizados nesse tipo de análise.

As comparações entre estimativas de miscigenação nos níveis populacional e individual realizadas no presente estudo levaram em conta as estimativas populacionais obtidas com o método IG por se tratar de um método próprio para lidar com dados desta natureza, enquanto o STRUCTURE possui um foco maior nas estimativas individuais.

Essas análises evidenciaram que em alguns casos, regiões com estimativas de miscigenação populacional semelhantes variaram quanto aos valores de miscigenação individual dos sujeitos que compõem aquela amostra populacional. Este fenômeno pode ser ilustrado pela comparação entre os doadores amostrados em Juiz de Fora e Uberlândia (**figura 30**). Embora doadores de ambas as localidades apresentem proporções de miscigenação populacional africana similares (JFO = 0,38 ±0,009; UDI = 0,37±0,012), Juiz de Fora apresenta 6,7% de indivíduos com alta proporção de ancestralidade africana, sendo este o segundo maior valor observado no estado (perdendo apenas para Ponte Nova) enquanto Uberlândia não apresentou nenhum indivíduo com miscigenação africana superior a 85%.

Entre os pacientes, também há casos onde se observa o fenômeno acima descrito (**figura 32**). Esta situação pode ser exemplificada pela comparação entre pacientes de Uberlândia, Montes Claros e Patos de Minas. Embora estas três localidades apresentem miscigenação populacional africana semelhante (UDI = 0,39 ±0,01, MOC e PMI = 0,38 ±0,018), em Uberlândia 9,1% dos indivíduos apresentaram alta proporção de ancestralidade africana, enquanto Montes Claros ou Patos de Minas não apresentaram indivíduos com ancestralidade africana superior a 85%.

As variações descritas em termos de comparação de miscigenação individual e populacional para as diferentes regiões do estado podem ter origem em fatores históricos e culturais. A espécie humana possui comportamento complexo no que diz respeito ao aspecto reprodutivo. A escolha de um parceiro envolve variáveis psicológicas, culturais, sócio-econômicas e biológicas. Assim sendo, os casamentos não ocorrem ao acaso em relação aos grupos étnicos, mesmo porque indivíduos pertencentes a esses diferentes grupos diferem não apenas quanto às características físicas, mas também quanto à classe-social e aos costumes (Salzano e Bortolini, 2002; Risch *et al.* 2009). A observação da existência de diferentes proporções de indivíduos com origem predominante africana ou europeia em localidades cujas estimativas de miscigenação populacional foram equivalente pode indicar que o próprio processo de miscigenação se deu de forma heterogênea no estado de Minas Gerais.

Algumas regiões de Minas Gerais parecem ter experimentado uma miscigenação mais intensa, produzindo grandes parcelas de indivíduos que apresentam proporções equilibradas de miscigenação africana e europeia. Diamantina é um exemplo desta situação, já que nesta localidade a proporção de doadores de sangue com miscigenação africana ou europeia maior de 85% não chegou a 17%. No extremo oposto, temos localidades como Manhuaçu, que apresentou 60% de indivíduos pouco

miscigenados (com ancestralidade africana ou europeia maior que 85%) (**figura 28, tabela 16**) .

Considerando todo o estado de Minas Gerais, observou-se que, entre os doadores, 36,8% têm alto nível de ancestralidade europeia e 3,09% tem alto nível de ancestralidade africana. Os demais 60,1% da população teriam níveis intermediários de miscigenação. Nenhum doador com alto nível de miscigenação ameríndia foi detectado.

Por outro lado, a análise global dos pacientes portadores de doença falciforme revelou que enquanto do ponto de vista populacional os pacientes apresentem miscigenação africana entre 43,7% e 47,3%, conforme o método estatístico de análise empregado, do ponto de vista da miscigenação individual, 11,05% tem alto nível de miscigenação africana e 15,58% deles tem alto nível de miscigenação europeia. Entre os pacientes, chamam a atenção casos particulares como os de Juiz de Fora, onde 54,4% dos pacientes apresentaram grande proporção de ancestralidade africana e o de Montes Claros, onde 26,1% dos indivíduos possuem grande proporção de ancestralidade europeia. Os resultados obtidos para os pacientes também apontam para uma marcante heterogeneidade nos padrões de miscigenação individual entre as regiões (**figura 29, tabela 16**).

Estes achados são extremamente interessantes, uma vez que demonstram que a doença falciforme, que já foi considerada uma doença de “negros” pode atingir indivíduos com grandes proporções de ancestralidade europeia.

Em países como os Estados Unidos a anemia falciforme tem sido amplamente associada aos “afro-americanos” que, segundo algumas publicações, não seriam um grupo heterogêneo no que concerne à miscigenação (Adams *et al.*, 2003; Hoppe *et al.*, 2007). Contudo, estudos recentes demonstraram que a miscigenação do grupo conhecido como Afro-Americanos é não só elevada, mas também bastante heterogênea (Bryc *et al.*, 2010).

No Brasil, o fenômeno da miscigenação, envolvendo todas as regiões e classes sociais, é amplamente conhecido (Salzano e Bortolini, 2002). A concepção de que, em decorrência da intensa miscigenação vivenciada em nosso país a doença falciforme não é exclusiva de “negros” vem sendo gradativamente reconhecida. Já em 1985, estimava-se que 1% dos brasileiros identificados como brancos seriam portadores do alelo causal da doença, em heterozigose (genótipo AS) (Salzano, 1985). Já a frequência do alelo “S” no sudeste do Brasil foi estimada como sendo 0,2% entre indivíduos classificados como de ascendência majoritariamente europeia, 0,8% entre miscigenados e 3,9% entre os indivíduos com ascendência majoritariamente africana (Salzano e Bortoloni, 2002).

Vale ainda ressaltar que a própria concepção da doença falciforme como “doença racial” é totalmente equivocada, já que a prevalência do alelo “S” na África sub-Saariana está relacionada não à etnia dos indivíduos que lá habitam, mas sim a uma estratégia evolutiva decorrente da distribuição geográfica da malária causada pelo *Plasmodium falciparum* (Pena, 2008).

Por fim, tendo em vista a totalidade das análises realizadas, podemos dizer que embora o presente estudo apresente limitações, trata-se de um ensaio inédito acerca da miscigenação dos pacientes portadores de doença falciforme e doadores de sangue do Estado de Minas Gerais.

A amostragem com ampla distribuição geográfica permitiu apreciar as características desses grupos e reconhecer suas particularidades.

As diferenças descritas para a miscigenação entre as diferentes regiões do estado e também entre os indivíduos de cada região alertam para o risco da ocorrência de resultados falso positivos em estudos de epidemiologia genética envolvendo tanto o grupo dos pacientes com doença falciforme quanto o grupo dos doadores de sangue.

Por outro lado, a constatação de que os pacientes portadores de doença falciforme possuem níveis substanciais de miscigenação abre a perspectiva de que se possa empregar a técnica de mapeamento por miscigenação na investigação de polimorfismos genéticos associados às diferentes manifestações clínicas dessa doença. Uma varredura genômica com o intuito de aplicar esta estratégia requer a genotipagem de marcadores em um número suficiente apenas para identificar a ancestralidade de segmentos cromossômicos. Para populações com miscigenação recente, como é o caso tanto dos afro-americanos dos Estados Unidos quanto dos pacientes com doença falciforme de Minas Gerais, um número de marcadores informativos de ancestralidade entre 1.500 e 2.000 poderia ser suficiente (Winkler *et al.*, 2010). Tendo em vista os avanços técnicos que possibilitam a genotipagem de um grande número de marcadores a custos cada vez mais baixos, a aplicação do mapeamento por miscigenação para identificar genes envolvidos em sub-fenótipos da doença falciforme no Estado de Minas Gerais poderá ser possível em um futuro próximo.

Assim sendo, acreditamos que este estudo cumpriu seu principal objetivo, ao fornecer subsídios para que possam ser desenvolvidos estudos de epidemiologia genética com desenho adequado no âmbito tanto da Fundação Hemominas quanto do próprio estado de Minas Gerais.

6. CONCLUSÃO

No presente estudo, foi padronizada a genotipagem de dois painéis de marcadores informativos de ancestralidade (MIAs) do tipo SNPs, pelo método de minisequenciamento, específicos para estimar a miscigenação africana, europeia e ameríndia. Este pequeno conjunto, que totalizou 14 marcadores, foi submetido a testes de controle de qualidade, quanto à reprodutibilidade e acurácia para estimativas de miscigenação, frente a um painel maior, composto por 108 MIAs. Em seguida, o painel de 14 SNPs foi utilizado para estimar a miscigenação de duas amostras populacionais: doadores de sangue e pacientes com doença falciforme de Minas Gerais, juntamente com outros 40 marcadores do tipo inserção-deleção padronizados por Bastos-Rodrigues *et al.* (2006). Indivíduos das populações parentais putativas da população brasileira (africanos, europeus e ameríndios) também foram genotipados.

Tendo como referência as amostras de 23 indivíduos miscigenados do painel do projeto SNP500cancer, denominados Hispânicos e genotipados para 108 MIAs, constatou-se que o painel de 14 SNPs foi capaz de estimar adequadamente a miscigenação europeia no nível populacional, mas subestimou a miscigenação ameríndia e superestimou a africana. A adição de mais indivíduos para compor a população parental ameríndia, levou a uma alteração dos resultados: a miscigenação ameríndia passou a ser corretamente estimada enquanto a miscigenação europeia passou a ser subestimada e a africana continuou a ser superestimada, porém agora em menor escala. Esta tendência de superestimativa da miscigenação africana, em comparação com os resultados obtidos para os hispânicos com 108 MIAs, permaneceu mesmo após a introdução da genotipagem de mais 40 marcadores do tipo inserção-deleção.

Embora a superestimativa da miscigenação africana tenha sido corroborada pela comparação com resultados anteriormente obtidos para a região sudeste (Callegari-Jacques, 2003, Lins, 2007), as estimativas obtidas para os doadores de sangue, os quais se espera que se assemelhem mais com a população geral do estado, foram bastante concordantes com as obtidas por estudos que consideraram regiões específicas do estado de Minas Gerais como Belo Horizonte (Scliar *et al.*, 2009) e Vale do Jequitinhonha (Parra *et al.*, 2003). Isto pode indicar que os resultados apresentados no presente estudo podem na verdade estar próximos da realidade.

Para as estimativas de miscigenação no nível individual obtidas com o painel de 14 SNPs, foram evidenciados amplos intervalos de probabilidade (análogos bayesianos dos intervalos de confiança). Isto significa que, sobretudo para indivíduos muito miscigenados as estimativas obtidas com pequenos painéis de marcadores não são muito precisas. A adição de 40 marcadores do tipo inserção-deleção possibilitou melhoria das estimativas de miscigenação no nível individual. No

entanto, ainda assim continuou-se a observar amplos intervalos de probabilidade para indivíduos mais miscigenados.

A genotipagem dos 14 SNPs e 40 INDELS, realizada para doadores de sangue e pacientes com doença falciforme amostrados em diferentes localidades do estado de Minas Gerais, permitiu constatar a ocorrência de variações nas contribuições das três populações parentais para a constituição das populações das diferentes regiões mineiras. Esta variação foi observada tanto para os doadores quanto para os pacientes.

De modo geral, os pacientes com doença falciforme apresentaram maior miscigenação africana, em relação aos doadores de sangue, o que era esperado devido à origem quase exclusivamente africana da mutação causal da doença na população brasileira. Ainda assim, foram observados pacientes portadores de anemia falciforme (forma homozigótica da doença), apresentando ancestralidade genômica majoritariamente europeia. Este achado reforça a idéia de que, especialmente em um país miscigenado como o Brasil, a idéia da existência de doenças “étnicas” ou “raciais” não é cientificamente adequada.

Em relação aos padrões geográficos de miscigenação, os pacientes apresentaram maior miscigenação africana ao sul e maior miscigenação europeia ao norte, contrariando o padrão descrito para a população geral do Brasil. Isto indica que os portadores de doença falciforme constituem um grupo distinto e que inferências acerca da miscigenação, obtidas para a população geral não são aplicáveis a esse grupo, pelo menos no Estado de Minas Gerais. Já os doadores, não apresentaram padrões geográficos discerníveis para a miscigenação africana e europeia, embora tenham apresentado maior miscigenação ameríndia a oeste, na região conhecida como Triângulo Mineiro.

Grande variabilidade foi detectada também para a miscigenação no nível individual, inclusive para indivíduos amostrados em uma mesma localidade. As análises de ancestralidade genômica mostraram que mesmo para regiões com níveis de miscigenação populacional semelhantes, foram observadas composições distintas quanto à miscigenação individual. Além disso, observou-se que enquanto em algumas regiões do estado predominam indivíduos com proporções intermediárias de miscigenação, em outras são muitos os indivíduos com ancestralidade predominantemente africana ou europeia (pouco miscigenados).

Os resultados obtidos no presente estudo alertam para o risco de obtenção de resultados falso-positivos nos estudos de associação genética envolvendo os pacientes com doença falciforme, ou mesmo os doadores de sangue, uma vez que ambos os grupos mostraram-se bastante heterogêneos no que diz respeito à miscigenação. Assim sendo, cuidados especiais devem ser tomados no desenho de estudos do tipo caso-controle realizados no âmbito do Estado de Minas Gerais.

Por outro lado, a constatação de que os pacientes com doença falciforme apresentam níveis substanciais de miscigenação abrem a perspectiva para a aplicação da abordagem de mapeamento por

miscigenação na investigação de polimorfismos genéticos associados às diferentes manifestações clínicas dessa doença.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- Abe-sandes, K. *et al.* 2004. Heterogeneity of the Y chromosome in Afro-Brazilian populations. *Hum. Biol.* 76 (1): 77-86.
- Adams, G. T. *et al.* 2003. Genetic risk factors for cerebrovascular disease in children with sickle cell disease: design of a case-control association study and genomewide screen. *BMC Med. Genet.* 4:6.
- Adams, G. T. *et al.* 2003. Genetic risk factors for cerebrovascular disease in children with sickle cell disease: design of a case-control association study and genomewide screen. *BMC Med. Genet.* 4 (6):1-10.
- Adewoye, A. H. *et al.* 2006. Association of polymorphisms of IGF1R and genes in the transforming growth factor-beta/bone morphogenetic protein pathway with bacteremia in sickle cell anemia. *Clin Infect Dis.* 43: 593-598.
- Alberto, F. L. e Costa, F. F. 2002. Heterozigose para hemoglobina S. In: Ministério da Saúde. Agência Nacional de Vigilância Sanitária. Manual de diagnóstico e tratamento de doenças falciformes. Brasília: ANVISA.
- Allison A. C. (1954) Protection afforded by sickle cell traits against subtertian malarial infection. *Br. Med. J.* 1: 290-294.
- Allison, A. C. 1964. Polymorphism and Natural Selection in Human Populations. *Cold Spring Harb. Symp. Quant. Biol.* 29: 137-49.
- Allison, A. C. 2009. Genetic control of resistance to human malaria. *Curr. Opin. Immunol.* 21 (5): 499-505.
- Allison, E. *et al.* 2008. Identification of genetic polymorphisms associated with risk for pulmonary hypertension in sickle cell disease. *Blood.* 111 (12): 5721-5726.
- Alves-Silva, J. *et al.* 2000. The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.* 67: 444-461.
- Almeida. R. H. 1997. O Diretório dos Índios: Um projeto de "civilização" no Brasil do século XVIII. Brasília: UNB.
- Ashley-Koch, A. *et al.* 2000. Sickle Hemoglobin (Hb S) Allele and Sickle Cell Disease: A Huge Review. *Am. J. Epidemiol.* 151 (8): 839-845.
- Baldwin. C. *et al.* 2005. Association of klotho, bone morphogenic protein 6, and annexin A2 polymorphisms with sickle cell osteonecrosis. *Blood.* 106 (1): 372-375.
- Bamshad, M. *et al.*, 2004. Deconstructing the relationship between genetics and race. *Nature Rev. Genet.* 5 (8): 598-609.
- Barbujani, G. *et al.* 1997. An apportionment of human DNA diversity. *PNAS.* 94: 4516-4519.

- Barnholtz-Sloan, J. S. *et al.*, 2008. Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. *Cancer Epidemiol. Biomarkers Prev.* 17 (3): 471-477.
- Bastos-Rodrigues L, *et al.* (2006). The Genetic Structure of Human Populations Studied Through Short Insertion-Deletion Polymorphisms. *Ann. Hum. Genet.* 70: 658-665.
- Beaumont, M. A e Rannala, B. 2004. The Bayesian revolution in genetics. *Nature Rev. Genet.* 5: 251-261.
- Belisário, A. R. 2010. Genótipos da alfa-talassemia e haplótipos do agrupamento de genes da beta-globina como moduladores da gravidade na anemia falciforme em crianças do programa estadual de triagem neonatal de MG e matriculados no Hemocentro de Belo Horizonte. Dissertação. (Mestrado). Saúde da Criança e do Adolescente. Faculdade de Medicina da Universidade Federal de Minas Gerais. Belo Horizonte.
- Bertorelle, G. e Excoffier, L. 1998. Inferring admixture proportions from molecular data. *Mol. Biol. Evol.* 15 (10): 1298-1311.
- Bortolini, M. C. *et al.* 1999. African-derived South American populations: A history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. *Am. J. Hum. Biol.* 11: 551-563.
- Bortolini, M. C. *et al.* 2003, Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am. J. Hum. Genet.* 73: 524-539.
- Botelho, T. R. 1994. Famílias e escravarias: demografia e família escrava no Norte de Minas Gerais no século XIX. Dissertação (Mestrado). Universidade de São Paulo – USP. São Paulo.
- Botelho, T. R. *et al.* 2007. Imigração e família em Minas Gerais no final do século XIX. *Rev. Bras. Hist.* 27: 155-176.
- Brener S. *et al.* 2008. Factors associated to clinical aptness for blood donation – demographic and socioeconomic determinants. *Rev. Bras. Hematol. Hemoter.* 30 (2): 108-113.
- Bohonak, A. J. 2002. IBD (isolation by distance): A program for analyses of isolation by distance. *Journal of Heredity.* 93 (2): 153-154.
- Bryc, K. *et al.* 2010. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *PNAS* 107 (2): 786-791.
- Callegari-Jacques, S. *et al.* 2003. Historical genetics: spatiotemporal analysis of the formation of the brazilian population. *Am. J. Hum. Biol.* 15: 824-834.
- Cançado, R. D. e Jesus, J. A. 2007. A doença falciforme no Brasil. *Rev. Bras. Hematol. Hemoter.* 29 (3): 203-206.
- Cang, Y. C. *et al.* 1995. An Analysis of Fetal Hemoglobin Variation in Sickle Cell Disease: The Relative Contributions of the X-Linked Factor, α -Globin Haplotypes, α -Globin Gene Number,

- Gender, and Age. *Blood*, 85 (4): 1111-1117.
- Cann, H. M. *et al.* 2002. A human genome diversity cell line panel. *Science*. 296: 261-262.
- Carvalho-Silva D. R. *et al.* 1999. Divergent human Y-chromosome microsatellite evolution rates. *Mol. Evol.* 49: 204-214.
- Carvalho-Silva, D. R. *et al.* 2001. The phylogeography of brazilian Y-chromosome lineages. *Am. J. Hum. Genet.* 68: 281-286.
- Carvalho-Silva, D. R. *et al.* 2006. Y chromosome diversity in Brazilians: switching perspectives from slow to fast evolving markers. *Genetica*. 126: 251-260.
- Chakraborty, R. 1986. Gene admixture in human populations: Models and predictions. *Yearb Phys Anthropol.* 29: 1-43.
- Chakraborty, R. 1985. Gene identity in racial hybrids and estimation of admixture rates. In: Genetic microdifferentiation in man and other animals. Neel, J. V. & Ahuja, Y. (eds.). Delhi: Indian Anthropological Association.
- Chakraborty, R. e Weiss, K. M. 1988. Admixture as a toll for finding linked genes and detecting that difference from allelic association between loci. *PNAS*. 85: 9119–9123.
- Chikhi, L. *et al.* 2002. Y genetic data support the Neolithic demic diffusion model. *PNAS*. 99 (17): 11008–11013.
- Choisy, M. *et al.* 2004. Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. *Mol. Ecol.* 13: 955-968.
- Chung, C. C. *et al.* 2010. Genome-wide association studies in cancer--current and future directions. *Carcinogenesis*. 31: 111-120.
- Cockerham, C. C. e Weir, B. S. 1993. Estimation of geneflow from F-statistics. *Evolution* 47: 855-863.
- Currat, M. e Excoffier, L. 2004. Modern Humans Did Not Admix with Neanderthals during Their Range Expansion into Europe. *PLoS Biol.* 2 (12): 2264-2274.
- DeSalle, R. e Tattersall, I. 2008. Human Origins: What Bones And Genomes Tell Us About Ourselves. College Station: Texas A&M University Anthropology
- De Souza, R. 2009. Os sentidos do povoamento: aspectos da ocupação territorial de Minas Gerais. *Rev. Cienc. Hum.* 1 (1) 1-19.
- Dupanloup, I. e Bertorelle, G. 2001. Inferring admixture proportions from molecular data: Extension to any number of parental populations. *Mol. Biol. Evol.* 18: 672-675.
- Elson, R. *et al.* 2002. Biostatistical Genetics and Genetic Epidemiology. New York: John Wiley & Sons.
- Excoffier, L. G. *et al.* 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1: 47-50.

- Excoffier, L. *et al.* 2005. Bayesian Analysis of an Admixture Model with Mutations and Arbitrarily Linked Markers. *Genetics*. 169: 1727-1738.
- Falush, D. *et al.* 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 164 (4): 1567-1587.
- Ferreira, L. B. *et al.* 2006. Genomic ancestry of a sample population from the state of Sao Paulo, Brazil. *Am. J. Hum. Biol.* 18 (5): 702-705.
- Ferreira, L.B. *et al.* Genomic ancestry of a sample population from the state of Sao Paulo, Brazil. *Am J Hum Biol.* 18(5):702-5.
- Ferrone, F. A. 2004. Polymerization and sickle cell disease: a molecular view. *Microcirculation*. 11 (2): 115-128.
- Florentino, M. G. 1997. Em costas negras: uma história do tráfico de escravos entre a África e o Rio de Janeiro (séculos XVIII e XIX). São Paulo: Companhia das Letras.
- Frederico Scott Varella Mata, 10 de junho de 2010. Comunicação pessoal.
- Frenette, P. S. e Atweh, J. G. F. 2007 Sickle cell disease: old discoveries, new concepts, and future promise. *J. Clin. Invest.* 117: 850-858.
- Frudakis, T. N. 2008. Molecular Photofitting: Predicting Ancestry and Phenotype Using DNA. Burlington: Academic Press.
- FUNAI – Fundação Nacional do Índio. www.funai.gov.br. As consultas na página da FUNAI foram feitas em 2010.
- Glaubitz, J. C. 2004. CONVERT: a user friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Mol. Ecol. Notes* 4: 309-310.
- Gonçalves, V. F. *et al.* 2007. Sex-biased gene flow in African Americans but not in American Caucasians. *Genet. Mol. Res.* 6:256-261.
- Hammer, M. F. e Zegura, S. L. 2002. The human Y chromosome haplogroup tree: nomenclature and phylogeny of its major subdivisions. *Ann. Rev. Anthropol.* 31: 303-321.
- Hamilton, B. M. 2009. Population Genetics. Chinchester: Jonh Wiley-Blackwell.
- Hanchard, N. *et al.* 2007. Classical sickle beta-globin haplotypes exhibit a high degree of long-range haplotype similarity in African and Afro-Caribbean populations. *BMC Genet.* 8: 52.
- Hoover, R. *et al.* 1979. Adhesion of normal and sickle erythrocytes to endothelial monolayer cultures. *Blood*. 54: 872-876.
- Hoppe, C. *et al.* 2007. Confirmation of an Association Between the TNF(-308) Promoter Polymorphism and Stroke Risk in Children With Sickle Cell Anemia. *Stroke*. 38 (8): 2241-2246.
- Hubisz, M. *et al.* 2009. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Res.* 9 (5): 1322-1332.

- Hünemeier, T. *et al.* 2007. Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. *Am J Phys Anthropol.* 133: 854-867.
- IBGE - Instituto Brasileiro de Geografia e Estatística. 2000. Brasil: 500 anos de povoamento. Rio de Janeiro: IBGE.
- IBGE – Instituto Brasileiro de Geografia e Estatística. 2010. <http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2008/default.shtm>
- Ingram, V. M. 1956. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature.* 178 (4537): 792-794.
- Ingram, V. M. 1957. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature.* 180 (4581): 326-328.
- Januário, J. N. 2002. Incidência da doença falciforme em um milhão de nascidos vivos em Minas Gerais (1998 – 2001). Dissertação (Mestrado). Faculdade de Medicina da Universidade Federal de Minas Gerais. Belo Horizonte.
- Jensen, J. L. *et al.* 2005. Isolation by distance, web service. *BMC Genetics* 6: 13. v.3.16 <http://ibdws.sdsu.edu/>
- Gourjon, G e Degioanni, A. 2009. AdFiT v1.7 (Admixture Files Tool): input files creating tool for population genetic admixture estimation software. *Bulletins et mémoires de la Société d'Anthropologie de Paris* , 21: 3-4.
- Kosoy, R *et al.* 2008. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum. Mutat.* 30: 69-78.
- Larsen, P. F. *et al.*, 2005 Stocking impact and temporal stability of genetic composition in a brackish northern pike population (*Esox lucius L.*), assessed using microsatellite DNA analysis of historical and contemporary samples *Heredity* 95: 136–143.
- Lettre, G. *et al.* 2008. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *PNAS.* 105 (33): 11869-11874.
- Lewis, P. e Zaykin, D. 2001. Genetic data analysis: computer program for the analyses of allelic data (software). Version 1.0 (d16). Disponível em: <http://lewis.eeb.uconn.edu/lewishome/software.html> Acesso em: junho de 2010.
- Li, W. H. e Sadler, L. A. 1991. Low nucleotide diversity in man. *Genetics.* 129 (2): 513-523.
- Lins, T. C. *et al.*, 2010. Genetic composition of Brazilian population samples based on a set of twenty eight ancestry informative SNPs. *Am J Hum Biol.* 22: 187-192.
- Lins, T. C. L. 2007. Impacto da miscigenação na aplicação do HapMap para a população brasileira avaliados nos genes PTPN22 e VDR. Dissertação (Mestrado) – Universidade Católica de Brasília. Brasília.

- Luna, F. V. e Klein, H. S. 2004. Economia e sociedade escravista: Minas Gerais e São Paulo. *Rev. Bras. Est. Pop.* 21: 173-193.
- Mamigonian, B. G. 2004. África no Brasil: mapa de uma área em expansão. *Topoi.* 9 (5) :33-53.
- Manual de Diagnóstico e Tratamento de Doença Falciformes. - Brasília : ANVISA, 2001. <http://bvsms.saude.gov.br/bvs/publicacoes/anvisa/diagnostico>.
- Marrero, A. R. *et al.* 2005. Heterogeneity of the genome ancestry of individuals classified as white in the state of Rio Grande do Sul, Brazil. *Am. J. Hum. Biol.* 17: 496-506.
- Mata-Machado, B. 1991. História do Sertão Noroeste de Minas Gerais 1690-1930. Belo Horizonte: Imprensa Oficial.
- Mesa, N. R. *et al.* 2000. Autosomal, mtDNA, and Y-chromosome diversity in amerinds: pre- and post-columbian patterns of gene flow in South America. *Am. J. Hum. Genet.* 67: 1277-1286.
- Monteiro, N. G. 1973. Imigração e colonização em Minas – 1889-1930. Belo Horizonte: Imprensa Oficial Belo Horizonte.
- Nagel, R. L. 2005. Epistasis and the genetics of human diseases. *C. R. Biol.* 328 (7): 606-615.
- Nagel, R. L. e Steinberg, M. H. 2001. Role of epistatic (modifier) genes in the modulation of the phenotypic diversity of sickle cell anemia. *Pediatr. Pathol. Mol. Med.* 20 (2): 123-136.
- Nagel, R.L. *et al.* 1985. Hematologically and genetically distinct forms of sickle cell anemia in Africa. *N. Engl. J. Med.* 312: 880-884.
- Neto G.C. e Pitombeira, M. S. 2003. Aspectos moleculares da anemia falciforme. *J. Bras. Patol. Med. Lab.* 39 (1):51-56.
- Nielsen, E. E. *et al.* 2003. Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis. *Mol. Ecol.* 12: 1497–1508.
- Noguchi, C. T. 1984. Polymerization in erythrocytes containing S and non-S hemoglobins. *Biophys. J.* 45 (6): 1153–1158.
- Nolan, V. G. *et al.* 2004. Association of single nucleotide polymorphisms in klotho with priapism in sickle cell anaemia. *Br. J. Haematol.* 128: 266-272.
- Nolan, V. G. *et al.* 2008. Association between wind speed and the occurrence of sickle cell acute painful episodes: results of a case-crossover study. *Br. J. Haematol.* 143 (3): 433-438.
- Oner, C. *et al.* 1992. Beta S haplotypes in various world populations. *Hum. Genet.* 89: 99-104.
- Pace, B. (editor). 2007. Renaissance of sickle cell disease research in the genome Era. London: Imperial College Press.
- Packer, B. R. *et al.* 2006. SNP500Cancer: a public resource for sequence validation, assay

- development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Res.* 34: 617-621.
- Pagnier, J. *et al.* 1984. Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. *PNAS.* 81: 1771-1773.
- Pante-De-Sousa, G. *et al.* 1999. Beta-globin haplotype analysis in Afro-Brazilians from the Amazon region: evidence for a significant gene flow from Atlantic West Africa. *Am. J. Hum. Biol.* 26: 365-373.
- Parra, F. C. *et al.* 2003. Color and genomic ancestry in Brazilians. *PNAS.* 100(1): 177-182.
- Pena, S. D. 2008. Anemia falciforme: uma doença geográfica. <http://cienciahoje.uol.com.br/colunas/deriva-genetica/anemia-falciforme-uma-doenca-geografica>.
- Pena, S. D. J. 2005. Razões para banir o conceito de raça da medicina brasileira. *Hist. cienc. saude-Manguinhos.* 12 (1): 321-46
- Pena, S. D. J. e Bortolini, M. C. 2004. Pode a genética definir quem deve se beneficiar das cotas universitárias e demais ações afirmativas? *Estud. Av.* 18 (50): 31-50.
- Pena, S. D. J. *et al.* 2009. DNA tests probe the genomic ancestry of Brazilians. *Braz. J. Med. Biol. Res.* 42 (10): 870-876.
- Pfaff, C. L. *et al.* 2004. Information on ancestry from genetic markers. *Genet Epidemiol.* 26: 305-315.
- Pfaff, C. L. *et al.* 2004. Information on ancestry from genetic markers. *Genet. Epidemiol.* 26: 305-315.
- Pierpaoli, M. *et al.* Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Mol. Ecol.* 12 (10): 2585-2598.
- Pimenta, J. R. *et al.* 2006. Color and genomic ancestry in Brazilians: a study with forensic microsatellites. *Hum. Hered.* 62: 190-195.
- Powar, D. e Hiti, A. 1993. Sickle cell anemia: beta S gene cluster haplotypes as genetic markers for severe disease expression. *Am. J. Dis. Child,* 147: 1197-1202.
- Powars, D. *et al.* 1990. The variable expression of sickle cell disease is genetically determined. *Sem Hematol.* 27: 360-376.
- Powars, D. R. 1991. β S-gene-cluster haplotypes in sickle cell anemia: clinical and hematologic features. *Hematol. Oncol. Clin. North Am.* 5 (3): 475-493.
- Prado, C. J. 2006. Formação do Brasil Contemporâneo. 23 ed. São Paulo: Brasiliense.
- Pritchard, J. K. e Donnelly, P. 2001. Case-control studies of association in structured or admixed populations. *Theor. Popul. Biol.* 60: 227-237.
- Pritchard, J.K. *et al.*, 2000. Inference of population structure using multilocus genotype data.

Genetics. 155: 945-959.

Ramachandran, S *et al.* 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *PNAS*. 102 (44): 15942–15947.

Reich, D e Patterson, N. 2005. Will admixture mapping work to find disease genes? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360: 1605–1607.

Reich, E. D. e Goldstein, D. B. 2001. Detecting Association in a Case-Control Study While Correcting for Population Stratification. *Genet. Epidemiol.* 20: 4–16.

Reis, J. J. 2000. A presença negra: encontros e conflitos. In: IBGE (Editor), Brasil: 500 anos de povoamento. Rio de Janeiro: Instituto Brasileiro de Geografia Estatística – IBGE.

Risch, N. *et al.* 2009. Ancestry-related assortative mating in Latino populations. *Genome Biol.* 10: R132.

Rogers, Z. R. *et al.* 1989. Nonblack patients with sickle cell disease have African beta S gene cluster haplotypes. *JAMA*. 261 (20): 2991-2994.

Rosenberg N. A *et al.* 2002. Genetic Structure of Human Populations. *Science*. 298 (5602): 2381-2385

Rosenberg, N. A. *et al.* Informativeness of genetic markers for inference of ancestry. *Am. J. Hum. Genet.* 73 (6): 1402-1422.

Rund, D. e Fucharoen, S. 2008. Genetic Modifiers in Hemoglobinopathies. *Curr. Mol. Medicine*, 8 (7): 600-608.

Salzano, F. M. e Bortolini, M. C. 2002. The evolution and genetics of Latin American populations. Cambridge: Cambridge University Press.

Salzano, F. M. e Tondo, C. V. 1985. Incidence, effects and management of sickle cell disease in Brazil. *Am. J. Pediatr. Hematol. Oncol.* 7: 240-44.

Sans, M. 2000. Admixture studies in Latin America: from the 20th to the 21st century. *Hum. Biol.* 72: 155–177.

Santos, N. P. C. *et al.* 2010. Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum. Mut.* 31:184-90.

Schlesselman, J. J. 1982. Case-control studies-design, conduct, analysis. New York, Oxford University Press.

Schnog, J. B. *et al.* 2004. Sickle cell disease; a general overview. *Neth. J. Med.* 62 (10): 364-374.

Scliar, M.O. *et al.* 2009. Brief communication: Admixture analysis with forensic microsatellites in Minas Gerais, Brazil: The ongoing evolution of the capital and of an African-derived

- community. *Am. J. Phys. Antropol.* 139: 591-595.
- Sebastiani, P. *et al.* 2004. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nature Genetics.* 37: 435-440.
- Seldin, M. F. *et al.* 2004. Putative Ancestral Origins of Chromosomal Segments in Individual African Americans: Implications for Admixture Mapping. *Genome Res.* 14: 1076-1084.
- Seyferth, G. 2002. Colonização, imigração e a questão racial no Brasil. *Rev. USP*, 53: 117-149.
- Smith, M. W. e O'Brien, S. J. 2005. Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nat. Rev. Genet.* 6: 623-662.
- Soo-Jin Lee S. *et al.* 2009. The Illusive Gold Standard in Genetic Ancestry Testing. *Science.* 325: 38-39.
- Souza, L. M. 1994. *Opulência e miséria das Minas Gerais.* São Paulo: Brasiliense.
- Souza, M. C. S. e Silva V. C. 2002. Caracterização socioeconômica e demográfica de doadores e não-doadores de sangue da Região Metropolitana de Belo Horizonte. XII Encontro Nacional de Estudos Populacionais. 23 a 27 de outubro de 2000. Caxambu, Minas Gerais
- Steinberg, M. H, e Adewoye, A. H. 2006. Modifier genes and sickle cell anemia. *Curr. Opin. Hematol.* 13 (3): 131-136.
- Steinberg, M. H. 2009. Genetic etiologies for phenotypic diversity in sickle cell anemia. *ScientificWorldJournal.* 18 (9): 46-67.
- Stolke, V. 2006. O enigma das interseções: classe, "raça", sexo, sexualidade: a formação dos impérios transatlânticos do século XVI ao XIX. *Rev. Estud. Fem.*14(1) 15-42.
- Stuart, M. e Nagel, R. 2004. Sickle-cell disease. *The Lancet.* 364: 1343-60.
- Tang, H. *et al.* 2005. Estimation of individual admixture: Analytical and study design considerations. *Genet. Epidemiol.* 28 (4): 289-301.
- Tarazona-Santos, E. *et al.* 2001. Genetic Differentiation in South Amerindians Is Related to Environmental and Cultural Diversity: Evidence from the Y Chromosome. *Am. J. Hum. Genet.* 68:1485–1496,
- Tarazona-Santos *et al.* 2007. Controlling the effects of population stratification by admixture in pharmacogenetics. In: Guilherme Suarez-Kurtz. (Editor): *Pharmacogenomics in Admixed populations.* Austin: Landes Bioscience.
- Templeton, A. 2006. *Population genetics and microevolutionary theory.* New Jersey: John Wiley and Sons.
- Trigo, T. C. *et al.*, 2008. Inter-species hybridization among Neotropical cats of the genus *Leopardus*, and evidence for an introgressive hybridzone between *L. geoffroyi* and *L. tigrinus* in southern Brazil. *Mol. Ecol.* 17: 4317-4333.
- Underhill, P. A. *et al.* The phylogeography of Y chromosome binary haplotypes and the origins

- of modern human populations. *Ann. Hum. Genet.* 65: 43-62.
- Vainfas, R. 2000, História indígena: 500 anos de despovoamento. In: IBGE (Editor), Brasil: 500 anos de povoamento. Rio de Janeiro: Instituto Brasileiro de Geografia Estatística – IBGE.
- Venâncio, R. P. 2000. Presença Portuguesa: de colonizadores a imigrantes. In: IBGE (Editor), Brasil: 500 anos de povoamento. Rio de Janeiro: Instituto Brasileiro de Geografia Estatística – IBGE.
- Vertchenko, S. B. 2005. Doação de sangue: aspectos sócio-econômicos, demográficos e culturais na região metropolitana de Belo Horizonte. Dissertação. (Mestrado). Saúde Pública. Faculdade de Medicina Universidade Federal de Minas Gerais. Belo Horizonte.
- Wainscoat, J. S. *et al.* 1983. Multiple origins of the sickle mutation: evidence from β S globin gene cluster polymorphism. *Mol. Biol. Med.* 1:191-197.
- Wang, J. 2003. Maximum-Likelihood Estimation of Admixture Proportions from Genetic Data. *Genetics.* 164: 747-765.
- Wang, J. 2006. A Coalescent-Based Estimator of Admixture from DNA Sequences. *Genetics.* 173: 1679-1692.
- Weber, J. L. *et al.* 2002. Human Diallelic Insertion/Deletion Polymorphisms. *Am. J. Hum. Genet.* 71: 854–862,
- Weicker, J. J. *et al.* 2001. Estimating the unbiased estimator θ for population genetic survey data. *Evolution.* 55 (12): 2601–2605.
- Weir, B. S. e Cockerham, C. C. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution.* 38 (6): 1358-1370
- Winkler CA, *et al.* 2010. Admixture Mapping Comes of Age. *Annu Rev Genomics Hum Genet.* 11 (18): 44-54.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97–159.
- Wright, S. 1943. Isolation by Distance. *Genetics.* 28 (2): 114-138.
- Zago, M. A. e Pinto, A. C. S. 2007. Fisiopatologia das doenças falciformes: da mutação genética à insuficiência de múltiplos órgãos. *Rev. Bras. Hematol. Hemoter.* 29 (3): 207-214.
- Zago, M. A. *et al.* 1992. Bantu β S-cluster haplotype predominates among Brazilian blacks. *Am. J. Phys. Anthropol.* 88: 295-298.
- Zago, M. 2002. Considerações Gerais. In: Ministério da Saúde. Agência Nacional de Vigilância Sanitária. Manual de diagnóstico e tratamento de doenças falciformes. Brasília: ANVISA.
- Zembruski, V. M. *et al.* 2006. Application of an African Ancestry Index as a genomic control approach in a Brazilian population. *Ann. Hum. Genet.* 70 (6): 822-828.

Tabela S1. Informações complementares para o painel de 10 INDELS

Marcador	Cromossomo	Concentração dos Iniciadores na reação de PCR (μM)	Fragmento INDEL	Frequências alélicas				
				Tamanho	Europeus	Africanos	Ameríndios	Japoneses
MID-777	16	0,08	GAA	3	0,41	0,33	0,94	0,42
MID-2937	8	0,2	GATTATA	7	0,075	0,15	0,735	0,431
MID-2350	3	0,2	TGC	3	0,869	0,708	0,172	0,419
MID-216	2	0,08	TATT	4	0,79	0,87	0,16	0,55
MID-1952	15	0,08	TCCA	4	0,34	0,26	0,92	0,52
MID-1689	11	0,1	AG	2	0,37	0,37	0,85	0,22
MID-52	4	0,5	TTG	3	0,15	0,21	0,86	0,79
MID-145	19	0,5	TAAT	4	0,83	0,96	0,34	0,97
MID-2750	20	0,2	TGGA	4	0,639	0,769	0,106	0,393
MID-2062	6	0,5	TTCATCTGTCATGGATGGTA G	21	0,7	0,58	0,07	0,72

Tabela S2. Coordenadas geográficas das localidades onde houve amostragem de doadores de sangue e/ou pacientes com doença falciforme

Localidade	Latitude	Longitude
Belo Horizonte	-19,923	-43,945
Diamantina	-18,245	-43,6
Divinópolis	-20,139	-44,884
Governador Valadares	-18,853	-41,945
Ituiutaba	-18,969	-49,465
Juiz de Fora	-21,764	-43,35
Manhuaçu	-20,258	-42,033
Montes Claros	-16,735	-43,862
Patos de Minas	-18,592	-46,521
Ponte Nova	-20,416	-42,909
São João Del-Rei	-21,136	-44,261
Uberaba	-19,751	-47,937
Uberlândia	-18,919	-48,277

ANEXO 1



MINISTÉRIO DA SAÚDE
 Conselho Nacional de Saúde
 Comissão Nacional de Ética em Pesquisa - CONEP
 Comitê de Ética em Pesquisa da Fundação Hemominas

PARECER CONSUBSTÂNCIADO

1. Título do Projeto de Pesquisa "Padrões geográficos de ancestralidade genômica em Minas Gerais e perspectivas do uso do desequilíbrio-de-ligação-gerado-por-miscigenação (DLM) para mapeamento de doenças complexas"		
SUJEITOS DA PESQUISA		
2. Número de sujeitos No Centro: 1000 Total: 1000	3. Grupos Especiais: <18 anos () Portador de deficiência mental () Embrião/feto () Relação de dependência (estudantes, militares, presidiários...) () Outros () Não se aplica ()	
PESQUISADOR RESPONSÁVEL		
4. Nome: Eduardo Martin Tarazona Santos		
5. Instituição a que pertence: Universidade Federal de Minas Gerais		
INSTITUIÇÃO(S) ONDE SERÁ REALIZADO		
6. Nome: Fundação Hemominas		
7. Unidade/Órgão: Fundação Hemominas		
8. Participação Estrangeira: () Sim (X) Não		
9. Projeto Multicêntrico: Sim (X) Não () Nacional (X) Internacional ()		
PATROCINADOR Não se aplica (X)		
10. Nome:		
COMITÊ DE ÉTICA EM PESQUISA - CEP		
11. Data de Entrada: 03/04/2006	12. Registro no CEP 140	
15. Objetivos: 1) Estimar a ancestralidade genômica individual Ameríndia, Européia e Africana em amostras de indivíduos saudáveis e pacientes com doença falciforme das diversas regiões de Minas Gerais. 2) Identificar padrões geográficos de miscigenação Ameríndia, Européia e Africana em indivíduos saudáveis e pacientes com doença falciforme de Minas Gerais. 3) Investigar a viabilidade de utilizar o Desequilíbrio-de-Ligação-gerado-por-Miscigenação recente (DLM), para mapear doenças complexas na população de Minas Gerais.		
16. Resumo: A população alvo deste estudo é a população do Estado de Minas Gerais e os pacientes com doença falciforme. Será utilizado o suporte logístico da Fundação Hemominas em todo o Estado de Minas Gerais. Serão coletadas amostras de 500 indivíduos saudáveis doadores de sangue e 500 de pacientes com doença falciforme. Após extração do DNA, serão genotipadas para 40 INDELS que são capazes de capturar a estrutura genética das populações humanas. Para testar a viabilidade de futuros estudos de DLM, serão genotipados 20 SNPs/INDELS marcadores de ancestralidade da região 6q24, que tem sido associada com hipertensão em indivíduos africanos e será investigado se os haplótipos definidos por estes marcadores, podem identificar a população parental da região 6q24 nos indivíduos estudados. Estes estudos de miscigenação ajudam no planejamento de estudos epidemiológicos casos-controles, contribuindo a reduzir resultados falsos positivos devidos a estruturação populacional entre casos e controles. Adicionalmente, em populações miscigenadas como a brasileira, o Desequilíbrio-de-ligação-gerado-por-miscigenação recente (DLM) pode ser usado para mapear mutações que modificam o risco de desenvolver doenças complexas. O cronograma apresentado está adequado com a proposta do estudo, o orçamento está apresentado de acordo com as recomendações da Res. CNS 196/96, assim como os demais itens solicitados na referida resolução.		
17. Comentário dos Relatores: Após terem sido feitas as modificações solicitadas por este CEP, o projeto se encontra de acordo com as recomendações das Resoluções CNS/MS 196/96 e suas complementares. Diante do exposto este projeto tem aprovação para execução de acordo com a avaliação ética.		
18. Parecer: APROVADO		
19. Conclusão: Aprovado (X)	Pendência ()	Não Aprovado ()
Data: 05/06/06	Data: / /	Data: / /
20. Encaminhamento a CONEP: Os dados acima para registro () O projeto para apreciação () Data: / /	21. Coordenador <i>Susie Dutra</i> Assinatura Susie Dutra Coordenadora do Comitê de Ética em Pesquisas Fundação Hemominas	


Parecer nº. ETIC 055/06

Interessado: Prof. Eduardo Martin Tarazona Santos
Departamento de Biologia Geral
Instituto de Ciências Biológicas-UFMG

DECISÃO

O Comitê de Ética em Pesquisa da UFMG – COEP aprovou, no dia 09 de agosto de 2006, o projeto de pesquisa intitulado **"Padrões geográficos de ancestralidade genômicas em Minas Gerais e perspectivas do uso do desequilíbrio-de-ligação-gerado-por-miscigenação (DLM) para mapeamento de doenças complexas"** bem como o Termo de Consentimento Livre e Esclarecido do referido projeto.

O relatório final ou parcial deverá ser encaminhado ao COEP um ano após o início do projeto.


Profa. Dra. Maria Elena de Lima Perez Garcia
Presidente do COEP/UFMG

ANEXO 2

Instruções de Amostragem

Prezado Colaborador,

Estamos muito gratos por sua decisão em colaborar com o projeto Padrões geográficos de ancestralidade genômica em Minas Gerais e perspectivas do uso do desequilíbrio-de-ligação-gerado-por-miscigenação (DLM) para mapeamento de doenças complexas que, além de gerar dados relevantes do ponto de vista científico, permitirá a criação de uma nova linha de pesquisa na Fundação Hemominas, relacionada à Genética de Populações Humanas. Sem a colaboração dos colegas das unidades espalhadas por todo o Estado, este estudo não seria possível.

A sua colaboração se dará por meio da coleta de assinaturas dos Pacientes com Doenças Falciformes que aceitarem participar do Estudo. O paciente deverá assinar um documento chamado “Termo de Consentimento Livre e Esclarecido” (TCLE) e colher uma amostra de sangue com anti-coagulante. O TCLE é um documento exigido pelo Comitê de Ética em Pesquisa, em atendimento à resolução 196/96, que trata de Pesquisa envolvendo Seres Humanos.

Estão descritos a seguir todos os procedimentos relativos à amostragem. Encaminho também os TCLEs para serem assinados e uma cópia do projeto. Embora a uma primeira leitura o projeto possa parecer complexo, por usar muitos termos técnicos, a leitura do próprio TCLE poderá ajudar em sua compreensão uma vez que está apresentado em linguagem mais simples e direta. Estarei sempre pronta para responder qualquer dúvida que possa surgir.

Atenciosamente,

Maria Clara Fernandes da Silva Malta

Instruções para a amostragem:

- No primeiro momento da pesquisa, coletaremos, em cada Unidade da Fundação Hemominas, amostras de sangue colhidas com anti-coagulante (EDTA) de 20 pacientes portadores de doenças falciformes.
- Estes pacientes deverão ser amostrados aleatoriamente, de preferência na ordem em que aparecerem para atendimento na Unidade e não poderão ser aparentados.
- Deverão ser coletadas as seguintes informações adicionais, para cada paciente: cidade de nascimento, cidade de residência, sexo, idade e perfil de hemoglobina (por exemplo, SS ou SC etc). Estas informações deverão ser anotadas na ficha encaminhada e enviadas junto com a amostra de sangue e o TCLE.
- Todos os pacientes que aceitarem participar da pesquisa deverão assinar o Termo de Consentimento Livre e Esclarecido (TCLE) e colher uma amostra de sangue. Pacientes que já forem colher sangue para exames Imuno-Hematológicos (painel, fenotipagem, etc) não precisarão colher outra amostra específica para a pesquisa mas apenas assinar o TCLE e encaminhá-lo junto com o protocolo de atendimento.
- Menores de idade (inclusive as crianças que forem encaminhadas para o exame de fenotipagem eritrocitária) poderão participar, desde que o responsável assine o TCLE.
- O TCLE deverá ser assinado em duas vias, sendo que uma delas ficará em poder do sujeito de pesquisa.
- No momento da assinatura, o paciente deverá optar, marcando no campo próprio para isto, se concorda que sua amostra possa ser usada em estudos futuros. O paciente deve ser informado de que, mesmo que ele aceite agora e desista depois, poderá entrar em contato com os pesquisadores e solicitar que sua amostra seja descartada.
- É fundamental que o paciente entenda que sua colaboração é indispensável para a realização da pesquisa e que, caso tenha qualquer dúvida, mesmo no futuro, poderá fazer contato com os pesquisadores por meio dos telefones que constam no TCLE.
- Cada Unidade da Hemominas irá amostrar preferencialmente pacientes que nasceram na sua área de cobertura Hemoterápica, para que todo o Estado de Minas Gerais seja representado.

Coloco-me à disposição para esclarecimentos adicionais.

Atenciosamente,

Maria Clara Fernandes da Silva Malta

ANEXO 3

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA DOADOR DE SANGUE

Conselho Nacional de Saúde, Resolução 196/96

Pesquisa: Padrões geográficos de ancestralidade genômica em Minas Gerais e perspectivas do uso do desequilíbrio-de-ligação-gerado-por-miscigenação (DLM) para mapeamento de doenças complexas

O objetivo deste estudo é contribuir para que, no futuro, pesquisadores dos problemas de saúde pública possam realizar estudos de epidemiologia genética melhor planejados no Estado de Minas Gerais. Os estudos de epidemiologia genética são aqueles que tentam determinar se alguma característica do material genético do indivíduo (denominado DNA) aumenta ou diminui o risco de desenvolver alguma doença.

A maior parte da população brasileira é, em alguma medida, produto da miscigenação (mistura) de nativos Americanos, Europeus e Africanos. A proporção destes componentes em cada indivíduo pode variar. Medir as diferenças nos níveis de miscigenação entre indivíduos da mesma população ou entre indivíduos de populações diferentes do Estado de Minas Gerais pode ajudar os pesquisadores a selecionar conjuntos de indivíduos com composição genética mais parecida entre si para seus estudos. Isto não significa que estes indivíduos não devam ser miscigenados (quase todos os brasileiros são miscigenados), mas que aqueles que participam de um estudo não tenham muitas diferenças entre si quanto aos níveis de miscigenação. Um objetivo deste estudo é medir quanto variam os níveis de miscigenação entre indivíduos de Minas Gerais. Para cada indivíduo participante, nós vamos analisar o material genético (DNA) para medir que percentagem de Nativo Americano, Europeu e Africano o indivíduo recebeu dos seus antepassados.

Por causa da miscigenação, as diferentes partes do material genético dos brasileiros podem ter origem diferente: Nativo Americana, Africana ou Européia. Para regiões específicas do seu material genético, nós queremos verificar se fazendo análises de tipo genético, nós conseguimos saber se essas pequenas porções de cada indivíduo foram herdadas de ancestrais Africanos, Europeus ou Nativos Americanos. Se isso for possível, será mais fácil no futuro desenhar estudos em populações miscigenadas, como a brasileira, para identificar características do material genético que aumentam ou diminuem o risco de desenvolver doenças, como a hipertensão (pressão arterial alta).

Nós estamos interessados no estudo da hipertensão porque a Fundação Hemominas, além de receber sangue de doadores, presta atenção médica a pacientes com doença falciforme (uma doença do sangue). Estes pacientes podem ter um risco maior de desenvolver hipertensão que os indivíduos

sadios. Por este motivo, nós estamos realizando o mesmo estudo em pacientes com doença falciforme, para comparar os de doadores sadios como você.

Procedimento

Nós o convidamos a participar desta pesquisa como doador de sangue. Caso você aceite participar deste projeto, você não terá nenhum custo. Iremos colher 5 mL (mililitros) de seu sangue, utilizando uma agulha por meio de uma punção (picada) na veia. A partir desta amostra, será extraído o material que conserva a informação herdada dos seus pais (o material genético das células, denominado DNA). Este material será estocado no Laboratório de Pesquisa no Hemocentro de Belo Horizonte e no Laboratório de Diversidade Genômica Humana do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, para uso na análise do grau de miscigenação utilizando análise genética. O resultado do seu teste não terá valor clínico e não será repassado para você, a menos que você o solicite.

Riscos e Benefícios

Sua saúde pessoal não receberá nenhum benefício deste estudo. Porém, os pesquisadores responsáveis acreditam que os resultados do estudo contribuirão para desenvolver melhores estudos epidemiológicos e entender melhor as causas de algumas doenças no Estado de Minas Gerais e no Brasil. Você também não receberá nenhuma retribuição em dinheiro ou outro benefício material por participar desta pesquisa.

Informação sobre privacidade

A amostra de sangue e o DNA (o material genético) extraído serão etiquetados com um código, não com seu nome, para garantia da sua privacidade. Este código não incluirá nenhuma informação que permita descobrir que a amostra é sua. Apenas os três responsáveis pelo estudo conservarão a correspondência entre o código e seu nome e terão acesso a esta informação. Estas pessoas são os pesquisadores Marina Lobato Martins e Maria Clara da Silva, da Fundação Hemominas, e Eduardo Tarazona Santos, da Universidade Federal de Minas Gerais. Eles conservarão essa informação em lugar seguro, sob chave. Quaisquer outros pesquisadores envolvidos no projeto não poderão saber que o DNA é seu.

O material genético (DNA) recolhido neste projeto poderá ser usado pelos mesmos pesquisadores envolvidos nesta pesquisa em futuras investigações com os mesmos objetivos deste estudo, caso você autorize. Neste caso, sua amostra será conservada sem prazo para descarte. Entretanto, o uso futuro deste material nunca será ligado a nenhum dado pessoal. Você pode negar a permissão para o uso de

sua amostra em estudos futuros, e ainda participar deste projeto. Se você autorizar o uso futuro da sua amostra, qualquer novo estudo utilizando sua amostra tem que ser aprovado pelos Comitês de Ética em Pesquisa da Fundação Hemominas e da Universidade Federal de Minas Gerais, e você não será informado sobre seu uso, se a nova pesquisa não trouxer benefícios ou riscos para você.

Na divulgação dos resultados, seu nome nem seus dados nunca serão mostrados, garantindo a você sigilo e privacidade.

Participação voluntária no estudo

Você tem toda a liberdade para decidir se quer ou não participar desta pesquisa, e não será prejudicado em seu atendimento, caso não queira participar. Você também pode decidir participar agora, mas desistir depois, que não sofrerá nenhum prejuízo.

Perguntas sobre o estudo

Se você não entender alguma coisa, pergunte. Se você não tem perguntas agora, sempre pode perguntar no futuro. Qualquer dúvida que você tiver, você poderá entrar em contato com os pesquisadores Maria Clara da Silva (3248-4563) ou Marina Lobato Martins (3248-4587) da Fundação Hemominas em Belo Horizonte, ou com o Prof. Eduardo Tarazona Santos do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais (3499-2572), para esclarecimentos.

Eu, _____, após leitura das duas páginas anteriores e esclarecimentos, aceito participar da pesquisa. (Coloque também as iniciais do seu nome e sobrenome nas páginas 1 e 2 deste Termo de Consentimento).

Aceito também que os pesquisadores usem minha amostra em estudos futuros.

Não aceito que os pesquisadores usem minha amostra em estudos futuros.

Carteira de Identidade: _____

Endereço

Rua: _____ Bairro: _____

Cidade: _____ Telefone: _____

Assinatura do participante (ou seu responsável, em caso de menor):

Data: ____/____/____

Assinatura do pesquisador:

Data: ____/____/____

Laboratório de Imunohematologia – Fundação Hemominas: 3248-4563

Comitê de Ética em Pesquisa da Fundação Hemominas: 3248-4587

ANEXO 4

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA PACIENTE COM DOENÇA FALCIFORME

Conselho Nacional de Saúde, Resolução 196/96

Pesquisa: Padrões geográficos de ancestralidade genômica em Minas Gerais e perspectivas do uso do desequilíbrio-de-ligação-gerado-por-miscigenação (DLM) para mapeamento de doenças complexas

O objetivo deste estudo é contribuir para que, no futuro, pesquisadores dos problemas de saúde pública possam realizar estudos de epidemiologia genética melhor planejados no Estado de Minas Gerais. Os estudos de epidemiologia genética são aqueles que tentam determinar se alguma característica do material genético do indivíduo (denominado DNA) aumenta ou diminui o risco de desenvolver alguma doença.

A maior parte da população brasileira é, em alguma medida, produto da miscigenação (mistura) de nativos Americanos, Europeus e Africanos. A proporção destes componentes em cada indivíduo pode variar. Medir as diferenças nos níveis de miscigenação entre indivíduos da mesma população ou entre indivíduos de populações diferentes do Estado de Minas Gerais pode ajudar os pesquisadores a selecionar conjuntos de indivíduos com composição genética mais parecida entre si para seus estudos. Isto não significa que estes indivíduos não devam ser miscigenados (quase todos os brasileiros são miscigenados), mas que aqueles que participam de um estudo não tenham muitas diferenças entre si quanto aos níveis de miscigenação. Um objetivo deste estudo é medir quanto variam os níveis de miscigenação entre indivíduos de Minas Gerais. Para cada indivíduo participante, nós vamos analisar o material genético (DNA) para medir que percentagem de Nativo Americano, Europeu e Africano o indivíduo recebeu dos seus antepassados.

Por causa da miscigenação, as diferentes partes do material genético dos brasileiros podem ter origem diferente: Nativo Americana, Africana ou Européia. Para regiões específicas do seu material genético, nós queremos verificar se fazendo análises de tipo genético, nós conseguimos saber se essas pequenas porções de cada indivíduo foram herdadas de ancestrais Africanos, Europeus ou Nativos Americanos. Se isso for possível, será mais fácil no futuro desenhar estudos em populações miscigenadas, como a brasileira, para identificar características do material genético que aumentam ou diminuem o risco de desenvolver doenças, como a hipertensão (pressão arterial alta).

Nós estamos interessados no estudo da hipertensão porque a Fundação Hemominas, além de receber sangue de doadores, presta atenção médica a pacientes com doença falciforme (uma doença do

sangue). Estes pacientes podem ter um risco maior de desenvolver hipertensão que os indivíduos saudáveis. Por este motivo, nós estamos realizando o mesmo estudo em pessoas sem doença falciforme, para comparar os resultados.

Procedimento

Nós o convidamos a participar desta pesquisa como paciente com doença falciforme. Caso você aceite participar deste projeto, você não terá nenhum custo. Iremos colher 5 mL (mililitros) de seu sangue, utilizando uma agulha por meio de uma punção (picada) na veia. A partir desta amostra, será extraído o material que conserva a informação herdada dos seus pais (o material genético das células, denominado DNA). Este material será estocado no Laboratório de Pesquisa no Hemocentro de Belo Horizonte e no Laboratório de Diversidade Genômica Humana do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, para uso na análise do grau de miscigenação utilizando análise genética. O resultado do seu teste não terá valor clínico e não será repassado para você, a menos que você o solicite.

Riscos e Benefícios

Como resultado da coleta de sangue, existe um risco mínimo que se forme um hematoma (cor roxa) na região do braço onde a agulha foi introduzida. Para diminuir este risco, seu sangue será coletado por profissional treinado.

Sua saúde pessoal não receberá nenhum benefício deste estudo. Porém, os pesquisadores responsáveis acreditam que os resultados do estudo contribuirão para desenvolver melhores estudos epidemiológicos e entender melhor as causas de algumas doenças no Estado de Minas Gerais e no Brasil, como a hipertensão. Você também não receberá nenhuma retribuição em dinheiro ou outro benefício material por participar desta pesquisa.

Informação sobre privacidade

A amostra de sangue e o DNA (o material genético) extraído serão etiquetados com um código, não com seu nome, para garantia da sua privacidade. Este código não incluirá nenhuma informação que permita descobrir que a amostra é sua. Apenas os três responsáveis pelo estudo conservarão a correspondência entre o código e seu nome e terão acesso a esta informação. Estas pessoas são os pesquisadores Marina Lobato Martins e Maria Clara da Silva, da Fundação Hemominas, e Eduardo Tarazona Santos, da Universidade Federal de Minas Gerais. Eles conservarão essa informação em lugar seguro, sob chave. Quaisquer outros pesquisadores envolvidos no projeto não poderão saber se o DNA é seu.

O material genético (DNA) recolhido neste projeto poderá ser usado pelos mesmos pesquisadores envolvidos nesta pesquisa em futuras investigações com os mesmos objetivos deste estudo, caso você autorize. Neste caso, sua amostra será conservada sem prazo para descarte. Entretanto, o uso futuro deste material nunca será ligado a nenhum dado pessoal. Você pode negar a permissão para o uso de sua amostra em estudos futuros, e ainda participar deste projeto. Se você autorizar o uso futuro da sua amostra, qualquer novo estudo utilizando sua amostra tem que ser aprovado pelos Comitês de Ética em Pesquisa da Fundação Hemominas e da Universidade Federal de Minas Gerais, e você não será informado sobre seu uso, se a nova pesquisa não trouxer benefícios ou riscos para você.

Na divulgação dos resultados, seu nome nem seus dados nunca serão mostrados, garantindo a você sigilo e privacidade.

Participação voluntária no estudo

Você tem toda a liberdade para decidir se quer ou não participar desta pesquisa, e não será prejudicado em seu atendimento, caso não queira participar. Você também pode decidir participar agora, mas desistir depois, que não sofrerá nenhum prejuízo.

Perguntas sobre o estudo

Se você não entender alguma coisa, pergunte. Se você não tem perguntas agora, sempre pode perguntar no futuro. Qualquer dúvida que você tiver, você poderá entrar em contato com os pesquisadores Maria Clara da Silva (3248-4563) ou Marina Lobato Martins (3248-4587) da Fundação Hemominas em Belo Horizonte, ou com o Prof. Eduardo Tarazona Santos do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais (3499-2572), para esclarecimentos.

Eu, _____, após leitura das duas páginas anteriores e esclarecimentos, aceito participar da pesquisa. (Coloque também as iniciais do seu nome e sobrenome nas páginas 1 e 2 deste Termo de Consentimento).

Aceito também que os pesquisadores usem minha amostra em estudos futuros.

Não aceito que os pesquisadores usem minha amostra em estudos futuros.

Carteira de Identidade: _____

Endereço

Rua: _____ Bairro: _____

Cidade: _____ Telefone: _____

Assinatura do participante (ou seu responsável, em caso de menor):

Data: ____/____/____

Assinatura do pesquisador:

Data: ____/____/____

Laboratório de Imunohematologia – Fundação Hemominas: 3248-4563

Comitê de Ética em Pesquisa da Fundação Hemominas: 3248-4587

ANEXO 5

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA PACIENTE COM DOENÇA FALCIFORME COM IDADE INFERIOR A 18 ANOS

Conselho Nacional de Saúde, Resolução 196/96

Pesquisa: Padrões geográficos de ancestralidade genômica em Minas Gerais e perspectivas do uso do desequilíbrio-de-ligação-gerado-por-miscigenação (DLM) para mapeamento de doenças complexas

O objetivo deste estudo é contribuir para que, no futuro, pesquisadores dos problemas de saúde pública possam realizar estudos de epidemiologia genética melhor planejados no Estado de Minas Gerais. Os estudos de epidemiologia genética são aqueles que tentam determinar se alguma característica do material genético do indivíduo (denominado DNA) aumenta ou diminui o risco de desenvolver alguma doença.

A maior parte da população brasileira é, em alguma medida, produto da miscigenação (mistura) de nativos Americanos, Europeus e Africanos. A proporção destes componentes em cada indivíduo pode variar. Medir as diferenças nos níveis de miscigenação entre indivíduos da mesma população ou entre indivíduos de populações diferentes do Estado de Minas Gerais pode ajudar os pesquisadores a selecionar conjuntos de indivíduos com composição genética mais parecida entre si para seus estudos. Isto não significa que estes indivíduos não devam ser miscigenados (quase todos os brasileiros são miscigenados), mas que aqueles que participam de um estudo não tenham muitas diferenças entre si quanto aos níveis de miscigenação. Um objetivo deste estudo é medir quanto variam os níveis de miscigenação entre indivíduos de Minas Gerais. Para cada indivíduo participante, nós vamos analisar o material genético (DNA) para medir que percentagem de Nativo Americano, Europeu e Africano o indivíduo recebeu dos seus antepassados.

Por causa da miscigenação, as diferentes partes do material genético dos brasileiros podem ter origem diferente: Nativo Americana, Africana ou Européia. Para regiões específicas do seu material genético, nós queremos verificar se fazendo análises de tipo genético, nós conseguimos saber se essas pequenas porções de cada indivíduo foram herdadas de ancestrais Africanos, Europeus ou Nativos Americanos. Se isso for possível, será mais fácil no futuro desenhar estudos em populações miscigenadas, como a brasileira, para identificar características do material genético que aumentam ou diminuem o risco de desenvolver doenças, como a hipertensão (pressão arterial alta).

Nós estamos interessados no estudo da hipertensão porque a Fundação Hemominas, além de receber

sangue de doadores, presta atenção médica a pacientes com doença falciforme (uma doença do sangue). Estes pacientes podem ter um risco maior de desenvolver hipertensão que os indivíduos saudáveis. Por este motivo, nós estamos realizando o mesmo estudo em pessoas sem doença falciforme, para comparar os resultados.

Procedimento

Nós o convidamos a participar desta pesquisa como paciente com doença falciforme. Caso você aceite participar deste projeto, você não terá nenhum custo. Iremos colher 5 mL (mililitros) de seu sangue, utilizando uma agulha por meio de uma punção (picada) na veia. A partir desta amostra, será extraído o material que conserva a informação herdada dos seus pais (o material genético das células, denominado DNA). Este material será estocado no Laboratório de Pesquisa no Hemocentro de Belo Horizonte e no Laboratório de Diversidade Genômica Humana do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, para uso na análise do grau de miscigenação utilizando análise genética. O resultado do seu teste não terá valor clínico e não será repassado para você, a menos que você o solicite.

Riscos e Benefícios

Como resultado da coleta de sangue, existe um risco mínimo que se forme um hematoma (cor roxa) na região do braço onde a agulha foi introduzida. Para diminuir este risco, seu sangue será coletado por profissional treinado.

Sua saúde pessoal não receberá nenhum benefício deste estudo. Porém, os pesquisadores responsáveis acreditam que os resultados do estudo contribuirão para desenvolver melhores estudos epidemiológicos e entender melhor as causas de algumas doenças no Estado de Minas Gerais e no Brasil, como a hipertensão. Você também não receberá nenhuma retribuição em dinheiro ou outro benefício material por participar desta pesquisa.

Informação sobre privacidade

A amostra de sangue e o DNA (o material genético) extraído serão etiquetados com um código, não com seu nome, para garantia da sua privacidade. Este código não incluirá nenhuma informação que permita descobrir que a amostra é sua. Apenas os três responsáveis pelo estudo conservarão a correspondência entre o código e seu nome e terão acesso a esta informação. Estas pessoas são os pesquisadores Marina Lobato Martins e Maria Clara da Silva, da Fundação Hemominas, e Eduardo Tarazona Santos, da Universidade Federal de Minas Gerais. Eles conservarão essa informação em lugar seguro, sob chave. Quaisquer outros pesquisadores envolvidos no projeto não poderão saber se o

DNA é seu.

O material genético (DNA) recolhido neste projeto poderá ser usado pelos mesmos pesquisadores envolvidos nesta pesquisa em futuras investigações com os mesmos objetivos deste estudo, caso você autorize. Neste caso, sua amostra será conservada sem prazo para descarte. Entretanto, o uso futuro deste material nunca será ligado a nenhum dado pessoal. Você pode negar a permissão para o uso de sua amostra em estudos futuros, e ainda participar deste projeto. Se você autorizar o uso futuro da sua amostra, qualquer novo estudo utilizando sua amostra tem que ser aprovado pelos Comitês de Ética em Pesquisa da Fundação Hemominas e da Universidade Federal de Minas Gerais, e você não será informado sobre seu uso, se a nova pesquisa não trouxer benefícios ou riscos para você.

Na divulgação dos resultados, seu nome nem seus dados nunca serão mostrados, garantindo a você sigilo e privacidade.

Participação voluntária no estudo

Você tem toda a liberdade para decidir se quer ou não participar desta pesquisa, e não será prejudicado em seu atendimento, caso não queira participar. Você também pode decidir participar agora, mas desistir depois, que não sofrerá nenhum prejuízo.

Perguntas sobre o estudo

Se você não entender alguma coisa, pergunte. Se você não tem perguntas agora, sempre pode perguntar no futuro. Qualquer dúvida que você tiver, você poderá entrar em contato com os pesquisadores Maria Clara da Silva (3248-4563) ou Marina Lobato Martins (3248-4587) da Fundação Hemominas em Belo Horizonte, ou com o Prof. Eduardo Tarazona Santos do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais (3499-2572), para esclarecimentos.

Eu, _____ responsável pelo menor de idade _____, após leitura das duas páginas anteriores, na presença do referido menor, e esclarecimentos, aceito que o menor sob minha responsabilidade participe da pesquisa. (Coloque também as iniciais do seu nome e sobrenome nas páginas 1 e 2 deste Termo de Consentimento).

Aceito também que os pesquisadores utilizem a amostra em estudos futuros.

Não aceito que os pesquisadores utilizem a amostra em estudos futuros.

Carteira de Identidade: _____

Endereço

Rua: _____ Bairro: _____

Cidade: _____ Telefone: _____

Assinatura do responsável:

Data: ____/____/____

Assinatura do pesquisador:

Data: ____/____/____

Laboratório de Imunohematologia – Fundação Hemominas: 3248-4563

Comitê de Ética em Pesquisa da Fundação Hemominas: 3248-4587

ANEXO 6



MINISTÉRIO DA SAÚDE
Conselho Nacional de Saúde
Comissão Nacional de Ética em Pesquisa - CONEP

PARECER Nº 576/09

Registro CONEP: 15479 (Este nº deve ser citado nas correspondências referentes a este projeto)

CAAE – 0005.0.215.203-09

Processo nº 25000.543470/2009-31

Projeto de Pesquisa: Emenda ao protocolo de pesquisa "Padrões geográficos de ancestralidade genômica em Minas Gerais e perspectivas do uso do desequilíbrio de ligação gerado por miscigenação (DLM) para o mapeamento de doenças complexas".

Pesquisador Responsável: Dr. Eduardo Martin Tarazona Santos

Instituição: Fundação Centro de Hemoterapia e Hematologia de Minas Gerais - HEMOMINAS

CEP de origem: Fundação Centro de Hemoterapia e Hematologia de Minas Gerais - HEMOMINAS

Área Temática Especial: Genética humana, Pesquisa com cooperação estrangeira.

Patrocinador: FAPEMIG e CNPq

Sumário geral do protocolo

Três grupos étnicos principais têm contribuído na formação do "pool" gênico da população brasileira: Nativos Americanos, Europeus e Africanos. Estudos em populações miscigenadas têm utilizando marcadores protéicos e, mais recentemente, marcadores moleculares uniparentais: DNA mitocondrial, cromossomo Y e polimorfismos autossômicos.

Em termos geográficos, tem sido observado que as populações miscigenadas do Norte apresentam maior contribuição Ameríndia, as do Nordeste maior contribuição Africana e as do Sul maior contribuição Européia, embora sejam observáveis níveis significativos de miscigenação tri-híbrida em todas as regiões e em todos os extratos sócio-econômicos. Uma grande variabilidade tem sido evidenciada nos níveis de ancestralidade genômica de indivíduos brasileiros, mesmo dentro de categorias fenotípicas como *branco*, *intermédio* e *pardo*. Se uma doença apresenta uma maior incidência em um dos três grupos étnicos que têm contribuído na formação da população brasileira, na realização de um estudo caso-controle, o grupo étnico que apresenta maior incidência estará mais representado nos casos que nos controles. Neste caso, qualquer variante gênica associada com o grupo étnico de maior incidência vai apresentar associação com a doença, independentemente do seu envolvimento na patogênese, podendo apresentar falsos positivos. Para evitar estas associações espúrias, em estudos casos-controles envolvendo populações miscigenadas é necessário controlar o efeito da etnicidade.

A população brasileira, predominantemente miscigenada, é considerada excelente candidata para realização de estudos de Mapeamento por Miscigenação (MM) de doenças complexas. Uma doença candidata a ser objeto de MM é a hipertensão, que tem um forte componente genético e apresenta frequências maiores em populações de origem africana. Em particular, a hipertensão é comum em indivíduos com anemia falciforme, embora não seja claro se a patogênese nestes pacientes e em indivíduos sem esta doença seja a mesma. Porém, a viabilidade de identificar mutações responsáveis pela hipertensão na população de pacientes, com anemia falciforme ou na população brasileira em geral, depende do sucesso na identificação da origem dos fragmentos cromossômicos utilizando haplótipos de MIAs (Marcadores Informativos de Ancestralidade).

A maior parte dos estudos de miscigenação desenvolvidos em populações brasileiras tem estimado a miscigenação no nível populacional. Métodos recentes e a possibilidade de genotipar um grande número de marcadores permitem atualmente estimar componentes de miscigenação no nível individual.

Os objetivos do estudo são: (1) Estimar a ancestralidade genômica individual Ameríndia, Européia e Africana em amostras de indivíduos saudáveis e pacientes com doença falciforme das diversas regiões de Minas Gerais; (2) Identificar padrões geográficos de miscigenação Ameríndia, Européia e Africana em indivíduos saudáveis e pacientes com doença falciforme de Minas Gerais; (3)

Investigar a viabilidade de utilizar o Desequilíbrio-de-Ligação-gerado-por-Miscigenação recente (DLM) para mapear doenças complexas na população de Minas Gerais.

O alvo do estudo é a população do Estado de Minas Gerais e os pacientes com doença falciforme. Utilizando o suporte logístico da Fundação HEMOMINAS serão realizadas coletas de 500 indivíduos sadios, doadores de sangue (a), e 500 pacientes com doença falciforme (b). As amostras serão obtidas nas 19 unidades da Fundação HEMOMINAS distribuídas em todo o Estado. Será coletada uma amostra de 5 cc de sangue de cada indivíduo. As amostras (a) e (b) terão uma distribuição geográfica similar, que compreenda todo o Estado.

Após extração do DNA, serão genotipadas 40 inserções-deleções (INDELS) que são capazes de capturar a estrutura genética das populações humanas. Estes marcadores serão genotipados por mini-sequenciamento em 4 reações multiplex. Os produtos de mini-sequenciamento serão analisados num sequenciador MEGABACE. Para testar a viabilidade de MM, 20 SNPs escolhidos do "database" do projeto HapMap na região 6q24, associados com hipertensão em indivíduos de origem Africana, serão genotipados por mini-sequenciamento. Estes SNPs serão escolhidos entre aqueles que evidenciem maiores diferenças de frequência entre as amostras de Africanos Yoruba e Europeus.

As informações relativas a doadores sadios e pacientes ficarão estocadas no HEMOMINAS conforme especificado no protocolo de pesquisa.

Emenda: (para avaliação da CONEP)

Ao projeto de pesquisa original foram acrescentados os seguintes objetivos, conforme especificado em carta do pesquisador responsável, de 01 de dezembro de 2008, endereçada ao CEP HEMOMINAS: "1. Genotipar de 500.000 a 1 milhão de marcadores selecionados como tag-SNPs do genoma humano a partir dos dados do projeto internacional HapMap (www.hapmap.org), nos indivíduos sadios (doadores de sangue) e nos pacientes com doença falciforme que aceitaram participar do projeto de pesquisa acima relacionado. Na primeira parte deste objetivo modificado, propomos a genotipagem de 60 indivíduos (30 pacientes com doença falciforme e 30 doadores sadios). Estamos procurando financiamento para ampliar a genotipagem a toda a amostra coletada. 2. Utilizando estes dados e outros, de databases públicos, selecionar um painel de 4000 a 6000 Marcadores Informativos de Ancestralidade (MIAs) e verificar a sua informatividade para estudos de mapeamento por miscigenação, avaliando a viabilidade de estimar a ancestralidade de fragmentos cromossômicos de ancestralidade contínua e a distribuição do comprimento destes fragmentos ao longo do genoma dos brasileiros".

A justificativa para esta alteração consiste no fato de que a viabilidade de identificar mutações responsáveis por doenças complexas na população de pacientes com doença falciforme ou na população brasileira em geral, utilizando o desequilíbrio de ligação gerado por miscigenação, depende do sucesso na identificação da origem dos fragmentos de cromossomos de diferentes origens formados no processo de miscigenação. Neste estudo, esta possibilidade será testada genotipando de 500.000 a 1 milhão de SNPs distribuídos em todo o genoma, selecionando entre esses SNPs um conjunto de 4000-6000 MIAs e verificando se é possível inferir a população de origem de fragmentos cromossômicos em função do número de MIAs utilizados. Isto permitirá verificar como a dinâmica da recombinação nos últimos cinco séculos influencia a capacidade de inferir a ancestralidade de regiões cromossômicas a partir de haplótipos compostos por MIAs distribuídos em todo o genoma.

Segundo o pesquisador, a confidencialidade dos dados referentes aos sujeitos de pesquisa será preservada, sendo os resultados publicados unicamente como frequências agregadas por marcador e região geográfica, sem tornar público genótipos multilocus individuais. Esta precaução constitui o padrão ouro mais recente em matéria de proteção da confidencialidade de indivíduos participantes em estudos genômicos.

Sendo assim, a alteração proposta consiste em uma oportunidade de melhoria metodológica do protocolo de pesquisa original mantendo-se, entretanto, a essência dos objetivos iniciais.

Local de realização

Trata-se de estudo nacional, multicêntrico, com participação estrangeira.

Segundo lê-se no projeto original, "As amostras de DNA serão armazenadas em geladeiras ou freezers em ambientes sob chave no Laboratório de Pesquisa no Hemocentro de

Belo Horizonte e no Laboratório de Diversidade Genômica Humana do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais até o final deste estudo. No final deste estudo, todas as amostras de DNA de sujeitos de pesquisa que não tenham autorizado o uso das amostras para estudos futuros serão descartadas. As amostras dos participantes que autorizarem o uso das amostras para futuros estudos com objetivos similares a este serão conservadas”.

Os laboratórios candidatos para a contratação das genotipagens são: “The Center for Applied Genomics” (Toronto, Canadá, do “The Hospital for Sick Children”), “Vanderbilt University” (EUA) e “The Translational Genomics Research Institute” (TGen, Arizona, EUA). Foi ressaltado pelo pesquisador que a decisão de realizar as genotipagens no exterior baseia-se no fato de que não existem atualmente no Brasil laboratórios com a suficiente experiência para a realização destas (página 5 da numeração do CEP). A genotipagem dos marcadores será realizada através do emprego dos arranjos desenvolvidos pelas empresas *Illumina* ou *Affymetrix*.

Apresentação do protocolo

Uma nova Folha de Rosto (FR - 242793) foi apresentada quando da submissão da emenda à análise ética pelo CEP. Encontra-se assinada pelo pesquisador proponente e pelo responsável pela Instituição Onde Será Realizado (Júnia Guimarães Mourão Cioffi, Diretora Técnico Científica da Fundação Hemominas) e informa a previsão de constituição de banco de materiais biológicos.

O estudo está sendo realizado em uma parceria entre a Fundação HEMOMINAS e a UFMG, constituindo parte da tese de doutoramento de Maria Clara da Silva, funcionária do HEMOMINAS. Também na página 5 (numeração do CEP) consta que, como parte do programa de pesquisa do grupo, está sendo desenvolvida uma plataforma de bioinformática denominada DIVERGENOMA, o que facilitará a análise dos dados na UFMG e na Fundação HEMOMINAS.

No projeto original consta que “As amostras de DNA serão armazenadas em geladeiras ou freezers sob chave no Laboratório de Pesquisa no Hemocentro de Belo Horizonte e no Laboratório de Diversidade Genômica Humana do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais. Os resultados das análises genéticas serão conservadas indefinidamente no database de dados genéticos do grupo de pesquisa do Dr. Tarazona, sem nenhum referimento com os dados pessoais do participante” (página 29 da numeração do CEP).

Foram anexadas as três versões de Termo de Consentimento Livre e Esclarecido (TCLE) do projeto original, destinados aos (a) doadores de sangue, (b) pacientes com doença falciforme e (c) pacientes com doença falciforme e idade inferior a 18 anos.

Na página 21 (numeração do CEP) foi apresentado o cronograma do projeto original, com duração prevista de 28 meses a partir da aprovação do CEP. Na página 6 é informado que o cronograma inicial será prorrogado em dois anos, visando a execução dos procedimentos adicionais.

Na página 7 (numeração do CEP) consta que os salários dos pesquisadores envolvidos são pagos pela UFMG e pela Fundação Hemominas, e que o pesquisador responsável é bolsista de produtividade em Pesquisa nível II do CNPq. É informado também que o projeto conta atualmente com os seguintes financiamentos: 1) Projeto “Diversidade Genética em Minas Gerais no Contexto da Diversidade Genômica Humana: Inferências Históricas e Implicações Epidemiológicas” - Programa Pesquisador Mineiro R\$48.000,00 (coordenador Eduardo T. Santos); 2) Projeto “Estrutura Genômica dos Latinoamericanos: Ancestralidade Indígena, Miscigenação e Ferramentas Bioinformáticas para a Análise” - Edital CNPq Universal 2008 R\$40.000,00 (01/12/2008 a 30/11/2010); e financiamento do Instituto Milênio (coordenação do Dr. Francisco Salzano) no valor de R\$83.000,00 para o grupo do pesquisador responsável. Ainda nesta página consta que: a) a coleta e extração de DNA é financiada pela Fundação Hemominas; b) genotipagem de 25 MIAs em 600 indivíduos: R\$25 por indivíduo x 600 indivíduos = R\$15.000; c) genotipagem de 60 indivíduos para 500.000 SNPs: US\$300/indivíduo x 60 indivíduos = US\$18.000 (aproximadamente R\$42.000).

Foi apresentado (página 40 da numeração do CEP) o parecer consubstanciado do CEP da Fundação HEMOMINAS quanto à aprovação, em 05/06/06, do projeto original. A carta de aprovação do estudo (09/08/06) pelo Comitê de Ética em Pesquisa da UFMG encontra-se na página 43.

Os currículos dos pesquisadores, disponíveis na Plataforma Lattes/CNPq, foram apresentados, observando-se adequação à proposta.

Comentário/ Consideração

1. Segundo informado nos TCLEs do projeto original, o material genético recolhido, quando autorizado, pode ser usado em investigações futuras. Na página 5 (emenda) lê-se "Após o experimento de genotipagem as amostras serão devolvidas para o nosso grupo no Brasil". Da mesma forma, a FR relativa à emenda em análise informa a utilização/constituição de banco de materiais biológicos. Neste sentido, solicita-se a comprovação do cumprimento do disposto nos itens 1.1 (Justificativa quanto a necessidade e oportunidade para usos futuros) e 1.4 (Norma ou regulamento elaborado pela instituição depositária para armazenamento de materiais biológicos humanos) da Resolução CNS 347/2005.

Diante do exposto, a Comissão Nacional de Ética em Pesquisa – CONEP, de acordo com as atribuições definidas na Resolução CNS 196/96, manifesta-se pela aprovação do projeto de pesquisa proposto, devendo o CEP verificar o cumprimento das questões acima, antes do início do estudo.

Situação: **Protocolo aprovado com recomendação.**

Brasília, 25 de novembro de 2009.


Gyselle Saddi Tannous
Coordenadora da CONEP/CNS/MS