

Kelly Nunes

Populações Ameríndias da América do Sul:
Estudo *multi-locus* e inferência histórico-
demográfica e seletiva

Native population in South American: a multi-
locus study of demographic and selective
history

Versão Corrigida

São Paulo

2011

Kelly Nunes

Populações Ameríndias da América do Sul:
Estudo *multi-locus* e inferência histórico-
demográfica e seletiva

Native population in South American: a multi-
locus study of demographic and selective
history

Tese apresentada ao Instituto de
Biotecnologia da Universidade de São
Paulo, para a obtenção de Título de
Doutor em Ciências, na Área de
Genética.

Orientador: Diogo Meyer

São Paulo

2011

Ficha Catalográfica

Nunes, Kelly

Populações Ameríndias da América do Sul: Estudo *multi-locus* e inferência histórico-demográfica e seletiva

203 páginas

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Populações Ameríndias 2. História Demográfica 3. Seleção Natural
Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Cópia Corrigida a tese original encontra-se disponível no Instituto de Biociências da USP.

Comissão Julgadora:

Profa. Dra. Maria Luiza Petzl-Erler

Profa. Dra. Tatiana Teixeira

Prof. Dr. Nelson Fagundes

Prof. Dr. Gabriel Marroig

Prof. Dr. Diogo Meyer

Orientador

Agradecimentos

Aos meus pais, Elisa e Edgar, pelo amor incondicional, apoio e incentivo constante.

À minha irmã Carla, meu porto seguro.

Ao professor Diogo Meyer, pela confiança, pelos desafios impostos e pela paciência.

Aos colegas do Laboratório de Genética Evolutiva Bárbara, Fábio, Gustavo, Márcia, Maria Helena e Ramalho, pelas discussões que sempre contribuíram para embasar esse estudo. Em especial ao Rodrigo Francisco por apresentar o mundo do “*HLA* e seus colegas” e pela ajuda ímpar na discussão dos dados de *HLA-B*.

Ao professor João Morgante, por gentilmente abrir as portas do LABEC, para a realização da parte laboratorial do presente estudo.

À todos os colegas do LABEC, pela convivência diária e, principalmente, pela compreensão nos momentos de monopólio dos termocicladores.

Ao professor Eduardo dos Santos, pelas amostras Amazônicas e colaboração com as discussões.

Ao professor Sidney dos Santos, pelo auxílio com o desenho dos oligonucleotídeos iniciadores e com a padronização dos protocolos de genotipagens. Agradeço em especial à Elzemar, à Ana Paula e à Maria Helena pela ajuda.

Ao professor Jorge Rocha, pela acolhida em Portugal e por me contagiar com seu entusiasmo pela genética de populações humanas e pelo uso de microssatélites.

À Dora e ao Bruno pela amizade e por tornarem a estadia em Portugal mais familiar.

Ao professor Rui Murrieta, pelo tempo disponibilizado aos estudos em grupo sobre aspectos antropológicos das populações Amazônicas.

À professora Maria Luiza Petzl-Erlar, ao professor Andrés Ruíz Linares e ao professor João Guerreiro pelas amostras cedidas.

À querida professora Eleidi Chauthard Freire Maia, pelas conversas e sábios conselhos.

Ao Duca, pelo companheirismo, pela compreensão, paciência, força nos momentos difíceis, a ajuda com as figuras, configurações e impressão da tese.

A FAPESP, pela bolsa concedida e por todo o suporte financeiro.

A todos que contribuíram de forma direta ou indireta, nesta longa jornada.

Índice

Lista de tabelas	vii
Lista de figuras	ix
Resumo	xii
Abstract	xv
Introdução Geral	1
Contextualização	4
O povoamento da América: implicações demográficas	4
O povoamento da América: evidências não genéticas	4
O povoamento da América: evidências genéticas	6
América do Sul: povoamento, padrões de diversidade e questões em aberto	8
O povoamento da América: implicações seletivas	9
O contato com os Europeus: o gargalo populacional mais recente	11
História demográfica versus história seletiva	12
Justificativa	13
Capítulo I	15
Introdução	15
Material e métodos	17
Amostras	17
Amostras para genotipagem dos microssatélites autossômicos espalhados pelo genoma	17
Bancos públicos de dados	18
Marcadores Moleculares	20
Microssatélites autossômicos espalhados pelo Genoma	20
Metodologia	23
Extração de DNA	23
Amplificação do genoma inteiro	24
Amplificação de fragmentos os microssatélites	24
Confirmação de amplificação dos produtos de PCR	24
Reação de genotipagem dos microssatélites	27
Reação de desnaturação	27
Eletroforese	27
Análise das genotipagens	27
Obtenção do banco de dados final utilizado nas análises	27
Calibração dos dados	29
Análise de dados	29
Estimativa de erro de genotipagem	29
Frequências alélicas	29
Equilíbrio de Hardy-Weinberg	29
Descrição da variabilidade genética	31
Diferenciação Inter-populacional	32
Análise de estruturação populacional	32
Teste de populações ancestrais	33
Estimativa de miscigenação	33
Estimativa de gargalo populacional	34
Resultados	35
Avaliação do banco de dados	35
Quantificação das falhas de amplificação	35

Estimativa de erro de genotipagem	37
Estimativas de desvios de Hardy-Weinberg	37
Calibração dos dados	39
Estimativa de miscigenação	41
Descrição da variabilidade genética	44
Descrição da diferenciação e estruturação populacional	47
Relação das populações do oeste e leste da América do Sul: possíveis rotas migratórias	57
Cenário Demográfico	57
Discussão	59
Avaliação do banco de dados	59
Estimativa de miscigenação	61
Diversidade populacional e gargalos populacionais	65
Diferenciação e estrutura populacional	67
A diferenciação entre Sudoeste e Leste da Amazônia	71
Relação das populações do oeste e leste da América do Sul: possíveis rotas migratórias	72
Capítulo 2	74
Introdução	74
O MHC e genes HLA: características e evidências de seleção	76
O gene <i>HLA-B</i> e seu perfil nas Américas	80
Material e métodos	82
Material	82
Amostras	82
Marcadores Moleculares	82
Microsatélites autossômicos da região do MHC	82
Microsatélites autossômicos espalhados pelo genoma	82
Metodologia	87
Extração de DNA	87
Amplificação do genoma inteiro	87
Amplificação de fragmentos os microsatélites	87
Reação de genotipagem dos microsatélites	88
Agrupamento dos produtos de PCR na eletroforese	88
Reação de desnaturação	88
Eletroforese	89
Análise das genotipagens	89
Obtenção do banco de dados final utilizado nas análises	89
Análise dos dados	91
Frequências alélicas	91
Equilíbrio de Hardy-Weinberg	91
Descrição da variabilidade genética	91
Diferenciação inter-populacional	92
Detecção de loci sob seleção	92
Determinação da fase gamética	93
Desequilíbrio de ligação	93
Heterozigose específica para os alelos de <i>HLA-B</i>	93
Diferenciação inter e intra linhagem de <i>HLA-B</i>	94
Associação de alelos de microsatélites e alelos e linhagens de <i>HLA-B</i>	94
Resultados	95
Avaliação dos dados	95
Quantificação das falhas de amplificação	95
Estimativas de desvios de Hardy-Weinberg	96
Frequências alélicas	97
Perfil dos microsatélites da região do MHC	97

Comparação de índices de diversidade	97
Diferenciação populacional: microssatélites restritos à região do MHC versus microssatélites espalhados pelo genoma	104
Teste de seleção na região do MHC por F_{ST} extremos	112
Desequilíbrio de ligação	116
Desequilíbrio de ligação entre HLA-B e microssatélites adjacentes	123
Heterozigose específica dos alelos de microssatélites em função do alelo <i>HLA-B</i>	123
Diferenciação inter e intra-linhagem	130
Associação de alelos de microssatélites com linhagens e alelos de <i>HLA-B</i>	130
Discussão	136
Falha de amplificação	136
Desequilíbrio de Hardy-Weinberg e detecção de alelos nulos	137
Perfil dos microssatélites da região do MHC	138
Averiguação de indícios de seleção na região do MHC	138
Perfil dos desvios de neutralidade para a região do MHC	142
Desequilíbrio de ligação	144
A heterozigose associada aos alelos de <i>HLA-B</i>	145
Diferenciação inter e intra-linhagem	147
Discussão Geral e Conclusões	149
Referências Bibliográficas	153
ANEXO 1	164
Resultados complementares	164
ANEXO 2	183
Manuscrito capítulo de livro	183
ANEXO 3	202
Script em linguagem R para a função HSH	202

Lista de tabelas

Tabela 1. Populações genotipadas pelo presente estudo, para 68 microssatélites localizados em cromossomos autossômicos espalhados pelo genoma. _____	18
Tabela 2. Lista dos 68 microssatélites selecionados, com informações sobre o cromossomo localizado, nome do <i>locus</i> , marcador e local onde foi realizada a genotipagem. _____	22
Tabela 3. Microssatélites agrupados nos conjuntos de multiplexagem utilizados para as genotipagens, com informações sobre a variação do tamanho do fragmento contendo o microssatélite em pares de base (pb) e fluorescência utilizada. _____	25
Tabela 4. Número de amostras utilizadas após a eliminação de indivíduos com excesso de falha de amplificação. _____	28
Tabela 5. Número de pares de base (pb) acrescidos ou diminuídos para calibração dos dados gerados no presente estudo com o banco de dados de Wang <i>et al.</i> (2007). _____	30
Tabela 6. Estimativa em percentagem (%) do erro de genotipagem do presente estudo, com base em sete <i>loci</i> . _____	37
Tabela 7. Estimativa da proporção de ancestralidade africana, europeia e ameríndia nas populações analisadas no presente estudo. _____	43
Tabela 8. Estatísticas sumárias para os índices de diversidade nas populações nativas americanas genotipadas no presente estudo. _____	45
Tabela 9. Índices de diferenciação populacional (F_{ST} e R_{ST}) em diferentes regiões do mundo. _____	48
Tabela 10. Índices de diferenciação populacional (F_{ST} e R_{ST}) ao longo das regiões da América do Sul. _____	54
Tabela 11. Comparação da contribuição de cada população na diferenciação populacional. _____	55
Tabela 12. Resumo do perfil das populações investigada no presente estudo e seu contato com europeus e euro-brasileiros segundo os registros oficiais do governo. _____	62
Tabela 13. Populações genotipadas pelo presente estudo, para 20 microssatélites localizados na região do MHC. _____	84
Tabela 14. Microssatélites da região do MHC analisados no presente estudo. _____	85
Tabela 15. Informações sobre microssatélites do MHC, reação de PCR, conjuntos de multiplexagem, e agrupamentos para a eletroforese. _____	88
Tabela 16. Número de amostras utilizadas nas análises após a exclusão de indivíduos com excesso de falha de amplificação. _____	90
Tabela 17. Número médio de alelos e heterozigose média dos microssatélites restritos a região do mhc nas 30 populações analisadas. _____	102

Tabela 18. Comparação dos índices de diversidade entre os microssatélites espalhados pelo genoma e os microssatélites da região do MHC. _____	103
Tabela 19. Comparação da diferenciação entre os microssatélites do genoma e restritos a região do MHC. _____	105
Tabela 20. Comparação da contribuição de cada população na diferenciação populacional. _____	109
Tabela 21. Heterozigose específica para os alelos de microssatélites em função dos alelos do gene <i>HLA-B</i> . _____	129
Tabela 22. Índice de diferenciação (F_{ST}) por locus de microssatélite intra e inter-linhagens de <i>HLA-B</i> . _____	130
Tabela 23. Associação do alelo mais comum do microssatélite com as linhagens de <i>HLA-B</i> . _____	131
Tabela 24. Associação dos alelos mais frequentes dos microssatélites com os alelos de <i>HLA-B</i> . _____	133

Lista de figuras

Figura 1. Mapa da localização geográfica das populações analisadas pelo presente estudo. _____	19
Figura 2. Gel de agarose 3% para verificação da amplificação dos produtos de PCR para posterior aplicação na eletroforese em ABI 3031. _____	26
Figura 3. Percentagem de falha de amplificação ao longo das populações nativa americanas nos 61 <i>loci</i> de microssatélites. _____	36
Figura 4. Percentagem de falha de amplificação ao longo dos 61 <i>loci</i> de microssatélites nas populações nativas americanas. _____	36
Figura 5. Distribuição dos desvios significativos de Hardy-Weinberg ($p < 0,05$) nas populações Ameríndias para o conjunto de 61 <i>loci</i> de microssatélites. _____	38
Figura 6. Distribuição dos desvios significativos de Hardy-Weinberg ($p < 0,05$) para cada um dos 61 <i>loci</i> das populações Ameríndias. _____	38
Figura 7. Averiguação de calibração dos dados para dados HGDP-CEPH, WANG <i>et al.</i> , 2007 e presente estudo. _____	40
Figura 8. Perfil da estruturação populacional quando introduzido erro proposital de calibração. _____	42
Figura 9. Agrupamento dos indivíduos (círculos) de acordo com a contribuição do grupo étnico parental. _____	44
Figura 10. Representação gráfica da estimativa de ancestralidade nas populações ameríndias genotipadas no presente estudo. _____	44
Figura 11. Estatísticas sumárias de diversidade genética ao longo das populações do mundo. _____	46
Figura 12. Índice de diferenciação populacional (F_{ST}) par a par das populações. _____	49
Figura 13. Estruturação populacional nas populações mundiais, mostrando a formação de grupos para as principais regiões. _____	51
Figura 14. Padrão de estruturação populacional nas Américas. _____	52
Figura 15. Influência das populações localizadas nas regiões Noroeste da América do Sul e Andinas na constituição das populações do Leste da América do Sul. _____	57
Figura 16. Estimativa de gargalo populacional através do Índice de Desequilíbrio β . _____	58
Figura 17. Rotas migratórias para o povoamento da América do Sul segundo o modelo de SCHMITZ, 1983. _____	73
Figura 18. Representação gráfica do processamento e apresentação do peptídeo antigênicos pelas moléculas HLA. _____	78
Figura 19. Mapa com a localização das populações analisadas no presente estudo. _____	83

Figura 20. Localização dos microssatélites genotipados na região do MHC. _____	86
Figura 21. Percentagem de falha de amplificação ao longo das populações nos conjuntos de 16 <i>loci</i> de microssatélites da região do MHC. _____	95
Figura 22. Percentagem de falha de amplificação ao longo dos 16 <i>loci</i> de microssatélites da região do MHC nas populações. _____	95
Figura 23. Distribuição dos desvios significativos de Hardy-Weinberg ($p < 0,05$) nas populações para o conjunto de 16 <i>loci</i> de microssatélites da região do MHC. _____	96
Figura 24. Distribuição dos desvios significativos de Hardy-Weinberg ($p < 0,05$) para cada um dos 16 <i>loci</i> da região do MHC nas populações. _____	96
Figura 25. Distribuição das frequências alélicas dos microssatélites <i>D6S276, D6S105, D6S2972, D6S265</i> . _____	98
Figura 26. Distribuição das frequências alélicas dos microssatélites <i>D6S2885, HLA-BC, D6S2928, D6S2927</i> . _____	99
Figura 27. Distribuição das frequências alélicas dos microssatélites <i>D6S2926, MICA, BAT2, D6S273</i> . _____	100
Figura 28. Distribuição das frequências alélicas dos microssatélites <i>G51152, D6S2874, D6S439, D6S291</i> . _____	101
Figura 29. Gráfico de dispersão contrastando microssatélites restritos a região do MHC com microssatélites genômicos para os índices de diversidade Theta (h), Heterozigose (He) e o número de alelos (K). _____	102
Figura 30. Diferenciação populacional par a par para os microssatélites espalhados pelo genoma. _____	106
Figura 31. Diferenciação populacional par a par para os microssatélites do MHC. _____	107
Figura 32. Regressão linear dos valores de F_{ST} par a par entre as 30 populações. _____	108
Figura 33. Regressão linear dos valores de F_{ST} par a par entre as 30 populações, destacando as populações mais diferenciadas. _____	110
Figura 34. Teste de permutação com a proporção de valores de F_{ST} para o grupo de 16 microssatélites que são maiores do que para o grupo de 61 microssatélites. _____	111
Figura 35. Regressão linear dos valores de F_{ST} par a par entre os microssatélites do tipo dinucleotídeos e tetranucleotídeos da base de dados de WANG <i>et al.</i> , (2007). _____	112
Figura 36. Comparação do F_{ST} e heterozigose (He) para identificação de <i>loci</i> candidatos à seleção, nas 30 populações analisadas no presente estudo. _____	113
Figura 37. Comparação do F_{ST} e heterozigose (He) para identificação de <i>loci</i> candidatos à seleção, nas 28 populações nativas da América analisadas no presente estudo. _____	113
Figura 38. Comparação do F_{ST} e heterozigose (He) para identificação de <i>loci</i> candidatos à seleção, nas tres populações da América do Norte analisadas no presente estudo. _____	114

- Figura 39.** Comparação do F_{ST} e heterozigose (H_e) para identificação de *loci* candidatos à seleção, nas seis populações do Noroeste da América do Sul analisadas no presente estudo. _____ **114**
- Figura 40.** Comparação do F_{ST} e heterozigose (H_e) para identificação de *loci* candidatos à seleção, nas três populações Andinas analisadas no presente estudo. _____ **115**
- Figura 41.** Comparação do F_{ST} e heterozigose (H_e) para identificação de *loci* candidatos à seleção, nas 13 populações do Leste da América do Sul analisadas no presente estudo. _____ **115**
- Figura 42.** Padrão de desequilíbrio de ligação para os microssatélites restritos à região do MHC em cada uma das populações analisadas no presente estudo. ___ **122**
- Figura 43.** Desequilíbrio de ligação por distância física (em pares de bases) dos microssatélites em cada região geográfica. _____ **122**
- Figura 44.** Desequilíbrio de ligação entre os microssatélites *MICA*, *D6S2926*, *D6S2927*, *D6S2928*, *HLA-BC* e *D6S2885* com o gene *HLA-B*. em cada uma das 30 populações analisadas no presente estudo. _____ **128**

Resumo

O presente estudo foca em dois temas principais: a) história demográfica das populações nativas do continente americano e b) como a história demográfica e seletiva moldam a diversidade e diferenciação dos genes *HLA*, nessas populações.

As populações nativas americanas apresentam uma história evolutiva peculiar, com padrões demográficos distintos das outras populações do mundo. Estudos anteriores sugerem que elas possuem: a) baixa diversidade genética e alta diversidade inter-populacional em relação às demais populações do mundo; b) um gradiente de diminuição de diversidade no sentido norte-sul do continente americano c) altos níveis de variabilidade intra-populacional, e baixos níveis de variabilidade genética inter-populacional nas populações que vivem na região oeste da América do Sul, em comparação às populações do leste. Contudo, esses achados são baseados em estudos que apresentam uma deficiência amostral para populações ameríndias das terras do baixo rio Amazonas. No presente estudo nós suprimos essa deficiência ao analisar 11 populações das terras baixas do rio Amazonas e 3 populações do centro-sul do Brasil. Constatamos que: a) as populações do leste e oeste da América do Sul são altamente diferenciadas, corroborando estudos anteriores; b) a diferenciação entre as populações do leste da América do Sul é maior que entre as populações do oeste (região andina e noroeste da América do Sul) em acordo com estudos anteriores; c) a maior diferenciação entre as populações do leste da América do Sul é causada por populações específicas (Ache, Arara do Iriri, Araweté, Surui e Ticuna Tarapaca), que apresentam histórias evolutivas peculiares; d) análises excluindo essas populações com alto grau de diferenciação mostram que para o conjunto restante de populações o nível de diferenciação é similar dentro das regiões leste com o encontrado para a diferenciação na região noroeste do continente sul americano corroborando os estudos morfológicos; e) o leste da América do Sul forma dois grupos populacionais distintos (sudoeste da Amazônia e leste da Amazônia+centro –sul do leste da América do Sul) e com diferente ancestralidade genética (o oeste da Amazônia com maior componente genético do Noroeste da América do Sul e o leste da

Amazônia+centro-sul com maior componente Andino), em acordo com o modelo de rotas migratórias proposto por SCHMITZ, 1983.

Os genes *HLA*, localizados na região do MHC, estão envolvidos na resposta imune adaptativa e tem como função apresentar peptídeos na superfície celular. Diversos estudos mostram que os genes *HLA* estão evoluindo sob regime de seleção balanceadora. O perfil da variabilidade genética dos genes *HLA* nas populações nativas da América (com grande número de alelos e alguns com frequência muito distinta de outras regiões do mundo) diferem dos demais genes até então analisados nessas populações. No presente estudo investigamos as contribuições da história demográfica e seletiva para moldar a variabilidade genética nas regiões adjacentes aos genes *HLA* em populações nativo-americanas. Para tanto, comparamos o perfil de 16 microssatélites na região do MHC com 61 microssatélites espalhados pelo genoma (controle demográfico) em 28 populações nativas da América, uma população africana (Ovimbundu) e uma europeia (Portugueses). Verificamos que: a) os microssatélites da região do MHC apresentam alto nível de desequilíbrio de ligação entre si corroborando estudos anteriores; b) apresentam maior diferenciação inter-populacional do que os microssatélites espalhados pelo genoma. Este sinal é oposto ao que esperávamos verificar para genes evoluindo sob seleção balanceadora. Estudos anteriores mostram que mesmo populações que são compostas por conjuntos de alelos *HLA* distintos apresentam baixos níveis de diferenciação. Sugerimos que a seleção poderia estar favorecendo alelo (ou conjunto de alelos) específico em diferentes regiões geográficas e que como os índices de diferenciação populacional como o F_{ST} estimam a variância das frequências alélicas, ele seria “cego” para a diferença na composição alélica das populações.

Como estudo de caso olhamos o gene *HLA-B*, que nas populações nativas da América, apresenta um conjunto de alelos endêmicos, ou que ocorrem em alta frequência no continente americano e em baixa frequência fora dele. Com intuito de verificar se há sinais de seleção nas regiões adjacentes a este gene, analisamos seis microssatélites flanqueadores a ele em 28 populações nativas. Estimamos a heterozigose dos microssatélites associada a um determinado alelo de *HLA-B* bem com o grau de associação dos alelos de microssatélites com as linhagens e os alelos de *HLA-B*. As análises mostram que: a) de modo geral, a heterozigose associada aos

alelos endêmicos é semelhante à observada em alelos cosmopolitas (com exceção dos alelos *HLA-B*3909* e *B*3543*); b) a diferenciação observada a partir dos microssatélites adjacentes é maior entre as nas linhagens de *HLA-B* do que dentro das linhagens c) haplótipos específicos de microssatélites apresentam forte associação com os as linhagens de *HLA-B*. Esses resultados são surpreendentes visto que as linhagens de *HLA-B* existem mesmo antes da especiação humana. Sugerimos que a baixa heterozigose associada as linhagens pode estar relacionada a dois fatores: a) o gargalo populacional ocorrido durante a entrada do homem moderno no continente americano; b) seleção atuando no favorecimento não apenas os alelos mais também as linhagens de *HLA-B*.

Desta forma concluímos que apesar da intensa história demográfica, as populações nativas americanas apresentam sinais da atuação de forças seletivas na região do MHC.

Abstract

The present study addresses two main themes: a) demographic history of the Native American populations and b) how the demographic and selective history shapes the diversity and differentiation of the *HLA* genes on those populations.

The Native American populations present a peculiar evolutive history, with demographic pattern that are distinct from other populations in the world. Previous studies suggest that they have: a) low genetic diversity and high inter-populational diversity compared to the other world populations; b) a diversity decreasing gradient on the north-south direction of the American continent; c) high levels of genetic variability within populations and low levels of genetic variability between populations that live in the western South American region, compared with the eastern ones. However, these findings are based on studies that present a sampling deficiency for the Amerindian populations located on the Amazon River lowlands. On the present study we suppress this deficiency by analyzing 11 populations of the Amazon River lowlands and 3 populations of the Brazilian South-Center. We have observed that: a) the populations of the East and West of the South America are highly differentiated, in accordance with previous studies; b) the differentiation among the Eastern South America is greater than among the Western ones (Andean region and Northwestern South America) agreeing with previous studies; c) the larger differentiation among the Eastern South American populations is caused by specific groups (Ache, Arara do Iriri, Araweté, Surui and Ticuna Tarapaca), which present peculiar evolutive histories; d) analysis that exclude these high differentiation level populations shows that for the remaining group, the differentiation level of the Eastern regions is similar to the differentiation levels found on the Western regions of the South American continent, corroborating the morphological studies; e) the east of the South America is divided in two distinct populational groups (Amazon Southwest and the Amazon East/South Central of South America) and with different genetic ancestry (the Amazon West with greater genetic component from the Northwestern South America and the Amazon

East/South Central with greater Andean genetic component), agreeing with the migration routes model proposed by SCHMITZ, 1983.

The *HLA* genes, located on the MHC region, are involved on the adaptative immune response and have the function of presenting peptides on the cellular surface. Various studies show that the *HLA* genes are evolving under balancing selection. The genetic variability profile of the *HLA* genes on the Native American populations (with great number of alleles and some with a frequency very distinct from other world regions) differs from the other genes so far analyzed in these populations. On the present study we investigate the contributions of the demographic history on shaping the genetic variability of the regions adjacent to the *HLA* genes on Native-American populations. To accomplish that, we compared the profile of 16 microsatellites of the MHC region with 61 microsatellites spread through the genome (demographic control) in 28 Native American populations, 1 African population (Ovimbundu) and 1 European population (Portuguese). We observed that: a) the microsatellites of the MHC region present a high linkage disequilibrium among themselves, corroborating previous studies; b) present higher inter-population differentiation than the microsatellites spread through the genome. This signal is opposed to the ones we expected from genes that are evolving under balancing selection. Previous studies show that even populations composed by groups of distinct *HLA* alleles present low levels of differentiation. We suggest that the selection could be favoring specific allele (or allele group) in different geographic regions, and since the population differentiation indexes such as F_{ST} estimate the variance of the allelic frequencies, it would be “blind” to the difference on the allelic composition of the populations.

As a case study, we investigate the *HLA-B* gene, which, in the American native populations, present a group composed by endemic allele and alleles that occur in high frequency on the American continent and in low frequency outside of it. With the intention of verifying if there are signs of selection on the regions adjacent to this gene, we have analyzed 6 microsatellites flanking to the *HLA-B* in 28 native populations. We estimated the heterozygosity of the microsatellites associated to a determined *HLA-B* allele as well as the degree of association of the microsatellite alleles with the *HLA-B* lineage and alleles. The analysis showed that: a) in general, the

heterozygosity associated to the endemic alleles is similar to the one observed in cosmopolitan alleles (with exception of the *HLA-B*3909* and *B*3543*); b) the differentiation observed from the adjacent microsatellites is greater between the *HLA-B* lineages than inside the lineages; c) specific microsatellite haplotypes present strong association with the *HLA-B* lineages. These results are surprising since the *HLA-B* lineages have existed even before the human speciation. We suggest that the low heterozygosity associated to the lineages could be related to two factors: a) the populational bottleneck occurred during the modern human entrance on the American continent; b) selection acting on the favoring, not only of the alleles, but also on the *HLA-B* lineages.

In conclusion, despite the intense demographic history, the Native American populations present signs of selective forces acting on the MHC region.

Introdução Geral

As populações humanas apresentam uma ampla diversidade cultural, linguística e genética. A descrição e compreensão desse padrão é algo que instiga a nossa curiosidade desde os primórdios da história. No último século, várias disciplinas como arqueologia, antropologia, linguística e genética têm se esforçado para compreender o nosso passado, reconstruir a nossa história e identificar quais fatores são responsáveis pela variabilidade e diferenciação entre as populações humanas.

A diversidade e diferenciação genética nas populações são moldadas por fatores genéticos e eventos demográficos (mutação, recombinação, deriva, seleção natural, migração, expansões e gargalos populacionais) (BALARESQUE *et al.*, 2007). Esses processos deixam marcas características no genoma humano, que são transmitidas às gerações seguintes e podem ser detectadas através de alterações que provocam nas frequências alélicas. Portanto, para reconstruir a história, é necessário partir de modelos pré-estabelecidos para realizar inferências sobre quais processos moldaram a diversidade genética das populações sob estudo (JOBLING *et al.*, 2004). Os resultados de estudos genéticos são particularmente valiosos se interpretados à luz de esforços interdisciplinares, unindo a Genética de Populações e a Evolução Molecular com Antropologia, Arqueologia, Imunologia, entre outras.

Com base em diferentes estudos, hoje é reconhecida a existência de um gradiente de perda de diversidade genética nas populações humanas, à medida que elas se distanciam do Leste da África (INGMAN *et al.*, 2000; UNDERHILL *et al.*, 2000; WANG *et al.*, 2007). Esses dados corroboram a ideia de que o homem moderno tem sua origem no continente africano e que sucessivas migrações sequenciais envolvendo apenas um subconjunto da população (e, portanto da variabilidade genética) marcaram o povoamento das diferentes regiões do mundo (CAVALLI-SFORZA e FELDMAN, 2003; RAMACHANDRAN *et al.*, 2005). Esse processo primeiramente leva à perda de diversidade genética inicial e ao longo do tempo cada população ficará sujeita aos eventos de deriva genética, mutação, recombinação e seleção natural que podem resultar na diferenciação e estruturação populacional.

A compreensão de como está distribuída a variação genética, e quais os níveis de diferenciação entre populações, é fundamental para entender a história dos movimentos das populações humanas, e também a forma como esses eventos impactam as suas características genéticas.

Diversos estudos com diferentes tipos de marcadores indicam que a maior parcela da variação genética (85 – 94%) é encontrada entre indivíduos da mesma população e não entre populações (2,5 – 7%) ou continentes (8 – 3,5%) (LEWONTIN, 1974; BARBUJANI *et al.*, 1997; ROSENBERG *et al.*, 2005; WANG *et al.*, 2007). Essa pequena proporção de diferenças genéticas entre as populações é reflexo tanto de eventos demográficos (migrações sequenciais, gargalos, expansões populacionais) como seletivos (adaptação a novos habitats). É reconhecido que a saída do homem moderno da África implica na conquista de novos habitats, o que acarreta em desafios como adaptação a mudanças climáticas (ex. diferença de radiação UV – mudança de pigmentação da pele (JABLONSKI, 2004)), alimentares (ex. domesticação de animais e consumo de leite na vida adulta – persistência da expressão do gene da lactase (TISHKOFF *et al.*, 2007)) e de patógenos, os quais podem ter exercido pressão seletiva. As consequências funcionais dessas variantes genéticas que facilitaram a sobrevivência das populações ancestrais é uma pequena, mas importante, contribuição para a diferença entre os grupos populacionais.

No presente estudo nós buscamos explorar a história evolutiva (demográfica e seletiva) das populações com menores índices de diversidade e maior diferenciação populacional no mundo – as populações Nativas Americanas, e em especial as populações Sul-Americanas. Essas populações apresentam uma história evolutiva peculiar: foram fundadas por um pequeno grupo de indivíduos, expandiram e povoaram o continente com maior gradiente latitudinal (portanto com ampla variação climática, de biomas e patógenos endêmicos), ficaram isoladas das demais populações do mundo por cerca de 10 mil anos e à cerca de 500 anos tiveram um encontro catastrófico com os europeus, o que dizimou cerca de 90% da população nativa. Portanto, elas apresentam uma das histórias mais complexas entre as populações humanas e explorá-la é um desafio.

Nosso interesse inclui dois temas gerais: (a) estudar a história demográfica e de movimentos populacionais que resultaram na ocupação das Américas, (b) usar

essa informação para compreender de que forma a seleção natural e os processos demográficos moldaram as características do sistema imune, em especial a região do Complexo Principal de Histocompatibilidade (MHC) nas populações nativas. Ambas são questões a respeito das quais há vários aspectos que permanecem controversos.

A presente tese está estruturada da seguinte maneira:

- 1) Revisão das principais características da história evolutiva das populações nativas da América e a contextualização da problemática e justificativa do presente estudo;
- 2) Capítulo 1: estudo da história demográfica das populações nativas da América do Sul;
- 3) Capítulo 2: estudo da história seletiva sobre a região do Complexo Principal de Histocompatibilidade (MHC) em populações nativas da América do Sul;
- 4) Discussão Geral, Conclusões Gerais e Considerações Finais.

Contextualização

O povoamento da América: implicações demográficas

A história do povoamento do continente americano é intrigante. Apesar de diversas áreas do conhecimento se aventurarem nesses estudos, muitas vezes os achados levam a interpretações controversas e, portanto há muito mais questionamentos do que respostas consensuais a respeito desse processo. Desta maneira, questões primordiais ainda são amplamente discutidas como: a) de onde vieram os primeiros nativos americanos? b) quando eles chegaram? c) em quantas ondas migratórias? d) quais rotas migratórias foram utilizadas ao longo do continente? (revisões em: SALZANO, 2002; 2007; SCHURR e SHERRY, 2004).

Nas próximas seções faremos uma breve síntese desse processo sob o ponto de vista de evidências não genéticas e genéticas.

O povoamento da América: evidências não genéticas

Atualmente, o único consenso a respeito do povoamento das Américas é que ele ocorreu via Estreito de Bering. Apesar do atual isolamento geográfico, dados geológicos e paleo-climáticos evidenciam que, durante a última era glacial, a América do Norte estava unida à Ásia através de uma ponte natural, formada na região do Estreito de Bering. Essa ponte provavelmente serviu de passagem para os primeiros humanos modernos migrarem para as Américas (ROGERS *et al.*, 1995; FIEDEL, 2000).

Os primeiros registros arqueológicos irrefutáveis da presença humana no continente americano consistem de ferramentas bi-faciais de pedras, ossos e marfim. Um dos primeiros sítios descobertos, compostos por esse conjunto de ferramentas, foi encontrado em uma localidade próxima a cidade de Clovis no Novo México (EUA) e datada entre 13.250 a 12.800 anos A.P. (Antes do Presente – sendo o presente considerado o ano de 1950) (revisão em FIEDEL, 2000; WALTERS e STAFFORD, 2007). O descobrimento de sítios arqueológicos como o de Monte Verde no Chile, com 14.220 a 13.980 anos A.P., geraram grande debate após a constatação de que estes sítios na América do Sul são mais antigos porém contemporâneos com os da América do Norte (DILLEHAY, 1997; DILLEHAY *et al.*, 2008). Apesar das revisões e reavaliações dessas datações, não houve grandes alterações neste cenário, confirmando que os sítios arqueológicos da América do Norte e do Sul

apresentam datações contemporâneas, o que indica que o povoamento das Américas ocorreu de modo rápido (WALTERS e STAFFORD, 2007; DILLEHAY *et al.*, 2008).

A ideia de povoamento que predominou por muitos anos foi a de que os primeiros habitantes do Novo Mundo atravessaram a ponte natural na região do Estreito de Bering e no leste do atual estado do Alasca e depararam-se com duas geleiras: Cordelerana (oeste) e Laurentidea (leste), que momentaneamente os impediu de adentrarem no continente. Com o aquecimento global ocorrido no final do Pleistoceno (11.500 – 12.300 anos A.P.), essas geleiras teriam cedido de tal forma que um corredor livre de gelo se abriu entre elas (Corredor de Alberta). Esse corredor possuía em torno de 1.200 km de extensão e serviu de passagem entre o Alasca e as demais regiões dos Estados Unidos (FIEDEL, 2000).

Devido as datações arqueológicas serem mais antigas que a data sugerida para a abertura do corredor de Alberta, alguns pesquisadores questionaram se o corredor já estaria disponível para a passagem humana nesta época (15.000-13.000 A.P.). Em 1979, FLADMARK sugeriu que a costa do Pacífico poderia ter sido uma rota alternativa utilizada pelos imigrantes. Essa teoria tem ganhado força nos últimos anos, pois os dados geológicos evidenciam que a deglaciação das costas noroeste e norte da América ocorreu entre 16.800 – 14.850 anos A.P.. Dessa maneira, a ocupação humana na costa oeste do continente americano durante o final da última glaciação era possível e o uso de pequenas embarcações poderia ter acelerado o processo de povoamento (revisado em FIEDEL, 2000; DILLEHAY *et al.*, 2008).

Endossando essa ideia, há registros paleoparasitológicos de helmintos (*Necator sp.*, *Ancylostoma duodenale* e *Tichuris trichiura*) encontrados em múmias e coprólitos na América do Norte e do Sul. Esses parasitas apresentam um ciclo de vida que necessita que seus ovos permaneçam em solo por algumas semanas. O clima frio da região norte não favoreceria a sobrevivência desses parasitas em solo, a não ser que a migração tenha sido rápida e por rotas não tão gélidas (CONFALONIERI *et al.*, 1991; ARAÚJO *et al.*, 2008). É importante salientar que as teorias de rota terrestre (via Corredor de Alberta) e rota costeira (via litoral do Pacífico) não são mutuamente exclusivas.

Várias hipóteses a respeito do número de migrações necessárias para explicar o atual padrão de diversidade nos povos nativos americanos foram propostas. Uma das primeiras, formulada por GREENBERG, TURNER e ZEGURA (1986), propõe que três ondas migratórias distintas, vindas do nordeste da Ásia para as Américas, deram origem aos três

grupos linguísticos atualmente encontrados nas populações nativas americanas: Ameríndio, Na-Dene e Esquimó-Aleuta. Essa teoria está caindo em desuso uma vez que apresenta problemas tanto de classificação linguística como de morfologia dentária.

Em 2003, NEVES e colaboradores, propuseram um modelo segundo o qual a ocupação do Novo Mundo se deu a partir da chegada de dois componentes biológicos distintos oriundos da Ásia: a) Os Paleoíndios, representantes da primeira leva de colonizadores com características craniométricas similares à observada em populações Australo-Melanesias e Africanas atuais; b) segunda leva, com morfologia similar à encontrada nos Asiáticos atuais.

Contudo, em 2008, GONZÁLEZ-JOSÉ e colaboradores sugeriram que as morfologias craniofaciais dos Paleoíndios e dos Nativos Americanos atuais poderiam ser extremos de um gradiente de variação morfológica. De acordo com esses autores, os primeiros habitantes do continente americano exibiam uma grande diversidade crânio-facial e apenas uma onda migratória teria ocorrido.

O que aconteceu com os Paleoíndios e por que esse componente biológico desapareceu ainda é um mistério. Os registros mais recentes desta morfologia foram encontrados em grupos indígenas da Terra do Fogo (América do Sul) e em *Pericus* (América do Norte), extintos durante o contato com os Europeus no século XVI (GONZALEZ-JOSÉ *et al.*, 2003). Esse tema é alvo de várias discussões e ainda não há consenso a respeito.

O povoamento da América: evidências genéticas

Inúmeros estudos têm analisado a diversidade genética nas populações ameríndias e a maioria dos marcadores moleculares investigados indicam um padrão de diversidade genética diferente daquele de populações dos demais continentes. Análises em 52 populações ao redor do mundo, por exemplo, com 377 marcadores do tipo microssatélites, que em geral são marcadores genéticos neutros, mostraram um decréscimo na variação genética intra-populacional e na proporção de alelos exclusivos, além de aumento da diferenciação inter-populacional no sentido que parte da África Sub-Saariana, passando por Eurásia, Leste da Ásia e Oceania até a América (ROSENBERG *et al.*, 2005; ZHIVOTOVSKY *et al.*, 2003; PRUGNOLLE *et al.*, 2005). Além disso, os alelos observados em nativos americanos geralmente são subconjuntos dos encontrados no leste asiático, de acordo com a hipótese

de que essas populações foram fundadas a partir de indivíduos vindos da Ásia (revisado em SALZANO, 2002).

Várias hipóteses a respeito do número de migrações para explicar o padrão de diversidade genética no continente americano foram formuladas (ver detalhes em TORRINI *et al.*, 1994; LELL *et al.*, 2002; WALLACE 1995; KARAFET *et al.*, 1999). Atualmente, a hipótese mais aceita é a de apenas uma onda migratória. Ou seja, o padrão de variação genética observada nos nativos americanos atuais pode ser atribuído à diferenciação *in situ* e movimentos populacionais que ocorreram após o povoamento do Novo Mundo. Essa teoria tem encontrado suporte principalmente nos dados de haplogrupos de DNAm (MERRIWETHER *et al.*, 1995; FOSTER *et al.*, 1996; BONATTO e SALZANO, 1997; STONE e STONEKING, 1993), cromossomo Y (PENA *et al.*, 1995; UNDERHILL *et al.*, 1996; BIANCHI *et al.*, 1998; KARAFET *et al.*, 1999; LELL *et al.*, 1997; ZEGURA *et al.*, 2004) e em estudo com microssatélites em cromossomos autossômicos (WANG *et al.*, 2007; SCHROEDER *et al.*, 2007; SCHROEDER *et al.*, 2009).

Os dados também indicam que antes de adentrarem no continente americano os indivíduos viveram por um tempo na Beríngia (região de terra firme com cerca de 1.600km a qual unia a atual Sibéria ao Alasca durante a última era glacial). Essa hipótese se apóia na constatação de que há mutações endêmicas tanto em haplogrupos de DNA mitocondrial (A2, B2, C1b, C1c, C1d, D1, D4h3 e X2a), cromossomo Y (Q1a3a) e microssatélites (D9S1120 com nove unidades repetidas). Esses alelos são compartilhados por indivíduos de praticamente todas as populações nativo-americanas, mas são inexistentes em outras populações do mundo. Assim, acredita-se que os primeiros habitantes permaneceram na Beríngia por um longo período, tempo suficiente para acumular essas mutações (BONATTO e SALZANO, 1997; TAMM *et al.*, 2007; SCHROEDER *et al.*, 2007). Independente das discussões a respeito dos processos migratórios é amplamente aceito pela comunidade científica que a ocupação de todo o continente americano aconteceu por pequenas migrações sequenciais. Sugere-se que um subgrupo da população parental deslocou-se para regiões ainda não habitadas, fundaram novas pequenas populações que se expandiram, das quais partiram outros subgrupos que fundaram outras populações sucessivamente. Como consequência, pequenos gargalos populacionais sequenciais, além do forte efeito fundador, contribuíram para a diminuição da diversidade genética. Esse cenário é corroborado pela observação da

existência de um gradiente de perda de diversidade genética na América do Sul em relação à América do Norte (DIXON, 2001; WANG *et al.*, 2007). Esses dados também ilustram que o sentido do povoamento foi do Norte para o Sul do continente americano.

América do Sul: povoamento, padrões de diversidade e questões em aberto

Apesar de um grande número de populações da América do Sul ter sido alvo de estudos genéticos, a maior parte dos trabalhos busca responder as questões em aberto sobre a chegada do homem na América ou sobre temas microrregionais, sendo que poucos estudos têm investigado como se deu o povoamento específico da América do Sul (TARAZONA-SANTOS *et al.*, 2001).

Acredita-se que a América do Sul tenha sido povoada a partir de grupos de caçadores-coletores vindos através do Istmo do Panamá (KEYEUX *et al.*, 2002). Contudo, alguns autores argumentam que a Selva de Darien, na divisa entre o Panamá e a Colômbia, e o Rio Antrato, na Colômbia, poderiam ter dificultado a travessia.

SCHMITZ (1983) apud SALZANO e CALLEGARI-JACQUES (1988) elaborou um modelo de rotas de migração para a América do Sul, baseado em dados da distribuição dos sítios arqueológicos e análise de marcadores genéticos polimórficos como grupos sanguíneos, antígenos de histocompatibilidade e proteínas séricas e eritrocitárias. O modelo consiste de duas rotas principais: uma rota ao longo da costa do Pacífico e da Cordilheira dos Andes em direção ao extremo sul do continente, e outra rota ao norte da América do Sul, pela Venezuela e Guianas em direção ao leste até o norte da Amazônia. Sugerem-se ainda rotas secundárias pelas Bacias do Rio Amazonas, pelo Rio da Prata e costa Atlântica.

Em 1994 CAVALLI-SFORZA e colaboradores realizaram análise de componente principal de variação a partir de 72 marcadores clássicos em populações desde a América do Norte à América do Sul. Os autores identificaram que 52,9% da variância total observada era explicada pelos três primeiros componentes. Os dois primeiros componentes revelam claras diferenças entre as porções leste e oeste da América do Sul. A partir do segundo componente principal também é possível identificar um gradiente ao longo do norte da região amazônica em direção ao leste, com valores se alterando gradativamente a partir do Panamá em direção à Colômbia, Guianas, Amapá, Pará, até chegar à região central do Brasil (CAVALLI-SFORZA *et al.*, 1994).

Abordagens alternativas utilizando o algoritmo de Monmonnier para identificação de áreas de descontinuidade abruptas identificaram claras diferenças entre o leste e oeste da América do Sul (CAVALLI-SFORZA *et al.*, 1994). Esses resultados foram reproduzidos ao analisar dados de DNA mitocondrial, cromossomo Y e microssatélites autossômicos (FUSELLI *et al.*, 2003; TARAZONA-SANTOS *et al.*, 2001; WANG *et al.*, 2007).

Além dessas diferenças alguns autores sugerem maiores níveis de variabilidade intra-populacional e menores níveis de variabilidade genética inter-populacional nas populações que vivem na porção oeste da América do Sul, em comparação com as populações do leste do continente (TARAZONA-SANTOS *et al.*, 2001; WANG *et al.*, 2007). Contudo, esses achados são baseados em estudos que apresentam uma deficiência amostral para populações ameríndias na região das terras do baixo rio Amazonas (entendem-se por terras baixas as regiões ao leste, a partir do sopé, da Cordilheira dos Andes), as quais podem acrescentar informações e contribuir para melhor esclarecer esse cenário demográfico.

Esse tema será abordado no Capítulo 1, uma vez que buscamos montar um banco de dados com microssatélites autossômicos para onze populações nativas do baixo rio Amazonas, dessa forma buscando realizar uma análise sem as deficiências amostrais dos estudos anteriores. Além disso, realizamos análise para determinar a influência e proximidade genética das populações Andinas e do Noroeste da América do Sul nas populações do Leste da América do Sul.

O povoamento da América: implicações seletivas

Ao povoar o continente americano, as populações encontraram um amplo gradiente latitudinal com grande variação de biomas e climas. O clima e os biomas no final do Pleistoceno eram diferentes dos atuais. As florestas da região neotropical estavam retraídas e as savanas predominavam na paisagem. As mudanças climáticas e ambientais ocorridas no Holoceno (entre 10.000 e 3.000 anos A.P.) levaram ao aumento da temperatura e umidade do planeta e a expansão das florestas tropicais (VUILLEUMIER, 1971).

Sabemos que o ambiente neotropical apresenta uma grande diversidade biológica, em especial da sua fauna. Sendo que quanto maior a diversidade de espécies animais em uma região, maior os riscos de doenças infecciosas e

parasitárias. Somente na região Amazônica, por exemplo, já foram identificados 187 tipos de arbovírus endêmicos, sendo 32 destes capazes de infectar humanos (revisão em CONFALONIERI, 2005). Desta forma, supomos que a ocupação de novos ambientes e o contato com patógenos endêmicos como vírus, bactérias e parasitas tenham exercido significativa pressão seletiva, principalmente sobre genes do sistema imunológico (CONFALONIERI, 2000).

O desencadeamento da resposta imune é extremamente complexo. Um dos mecanismos de ativação da resposta imune, se dá através das proteínas codificadas pelos genes *HLA* (Human Leucocyte Antigen – família genica localizada no Complexo Principal de Histocompatibilidade (MHC) no braço curto do cromossomo 6 – 6p21.3). Esses genes codificam proteínas responsáveis pela apresentação de peptídeos (antígenos próprios ou de origem patogênica) na superfície celular, que são reconhecidos pelos linfócitos que ativam a resposta imunológica (ABBAS *et al.*, 2008). A apresentação dos peptídeos na superfície celular depende da capacidade que cada alelo de *HLA* tem de se ligar e apresentar os epítomos específicos. Assim, indivíduos que possuem alelos com alta capacidade de apresentar uma grande gama de epítomos, em especial dos patógenos aos quais está sendo exposto, terão uma maior aptidão (probabilidade de sobrevivência e reprodução) e seus alelos serão transmitidos às gerações seguintes.

Os genes *HLA* são os mais polimórficos do genoma humano e apresentam evidências de estarem evoluindo sobre regime de seleção balanceadora (mecanismo que aumenta a diversidade alélica em relação ao esperado sobre neutralidade) (HEDRICK, 1999). Vários estudos têm utilizado os genes *HLA* para caracterização genética de diferentes populações humanas (SOLBERG *et al.*, 2008, MEYER *et al.*, 2006), com o intuito de caracterizar o modo como a seleção interage com a história demográfica em diferentes regiões. Os estudos desses genes em Ameríndios revelam um padrão distinto do documentado para outros genes. Os genes *HLA* nas Américas apresentam um significativo número de alelos endêmicos e/ou alelos com frequência muito diferenciada de outras regiões do mundo (PARHAM *et al.*, 1997, FRANCISCO, 2009). Além disso, um estudo desenvolvido em nosso grupo de pesquisa detectou sinais de seleção balanceadora para o gene *HLA-B* nessas populações (FRANCISCO, 2009).

Esse tema será o foco do capítulo 2, no qual faremos uma análise comparativa da variação genética presumivelmente neutra (caracterizada no capítulo 1 através de marcadores do tipo microssatélites, espalhados ao longo do genoma) com a que está supostamente sobre seleção (inferida com base em microssatélites localizados próximos a genes *HLA* clássicos).

O contato com os Europeus: o gargalo populacional mais recente

O contato com os europeus a partir do século XV mudou drasticamente a vida dos nativos americanos. A história conta que algumas populações nativas tiveram 90% dos indivíduos dizimados após a chegada dos europeus. Embora uma pequena parcela de nativos tenha sido de fato morta em guerra direta com os estrangeiros, a maioria morreu devido a doenças trazidas pelos colonizadores (como varíola, sarampo, tifo, gripe, tuberculose entre outras) e a fome (BLACK, 1992; DIAMOND, 2006).

Ao contrário do imaginário popular, até mesmo as populações da região leste do continente Sul Americano, que habitavam a floresta Amazônica não viviam isoladas umas das outras, mas eram conectadas por uma intensa rede de comércio (HECKENBERGER, 2008). Desse modo, há relatos que muitas vezes as doenças chegaram às tribos indígenas antes do contato direto com os europeus, levando muitos nativos à morte mesmo sem sequer ter visto o “homem branco” (DIAMOND, 2006; GURGEL, 2010).

A menor capacidade de resposta imune a essas doenças trazidas pelos colonizadores europeus não foi necessariamente uma consequência de um sistema imune deficiente. A maior suscetibilidade das populações nativas americanas está provavelmente associada ao isolamento de 10 mil anos e ao contato abrupto com um conjunto de patógenos completamente estranhos, trazido pelos europeus. Populações europeias, africanas e asiáticas domesticaram animais como bois e aves, que possuem agentes patológicos que são transmitidos ao homem (ex. varíola, sarampo, gripe), e essa convivência selecionou indivíduos, não totalmente imunes, mas capazes de reconhecer e combater essas doenças. A interligação entre Europa, Ásia e África, facilitou o relacionamento entre povos de diferentes continentes possibilitando o intercâmbio gradual e progressivo dos microrganismos e seu

consequente estímulo imunológico. Já os nativos americanos praticamente não tiveram animais para domesticar e seu sistema imune foi estimulado a responder aos patógenos do continente americano. Essa é uma das principais hipóteses de porque o contato com os patógenos europeus foi tão devastador aos nativos (GURGEL, 2010).

As doenças não pouparam as principais lideranças indígenas, o que provocou uma grande desestruturação social. Sem condições para o cultivo da terra, a fome e desnutrição atingiram as populações. Como consequência, os sobreviventes famintos vendiam-se como escravos ou abandonavam suas terras (muitas às margens dos rios) e fugiam para o interior da mata (DIAMOND, 2006; GURGEL, 2010).

Em resumo, o encontro com os europeus levou a um brusco gargalo populacional, e ao deslocamento de muitas populações (ou o que sobrou delas) para outras localidades, criando assim o falso estigma de que as populações nativas sempre foram formadas por pequenos grupos de indivíduos que vivem isolados.

História demográfica versus história seletiva

A partir do exposto acima, é possível constatar que as populações nativas da América possuem uma história demográfica complexa, e separar o efeito da história demográfica da seletiva é um desafio.

É sabido que alguns tipos de eventos demográficos deixam assinaturas idênticas às da seleção natural. Quando a população sofre uma rápida expansão, por exemplo, são encontrados alelos raros nas amostras, um efeito também observado quando houve seleção positiva (a favor de um alelo). Por outro lado, em um evento de gargalo populacional, alelos em baixas frequências são preferencialmente eliminados, deixando os remanescentes em frequências parecidas, um efeito semelhante ao observado em seleção balanceadora (BAMSHAD e WOODING, 2003; NIELSEN, 2005). Consequentemente, a análise de *loci* únicos não permite estabelecer se foram processos demográficos ou seletivos que operam sobre a população.

Para distinguir entre os efeitos seletivos e demográficos sobre a variação genética, utiliza-se do fato que a seleção deixa assinaturas em *loci* específicos e a história demográfica deixa sinais em todo o genoma. Uma forma de abordagem é

comparar, nos mesmos indivíduos, *loci* neutro e *loci* candidatos a estarem sob seleção e verificar se os padrões apresentados são semelhantes ou distintos (AKEY *et al.*, 2004).

No presente estudo nós utilizamos essa abordagem ao analisar marcadores do tipo microsatélites (pequenas sequencias de DNA compostas por motivos de 1 a 6 nucleotídeos repetidos em série, também denominados de STR – *Sort Tandem Repeats*) espalhados pelo genoma e concentrados na região do MHC (*Major Histocompatibility Complex* – Complexo Principal de Histocompatibilidade). Os microsatélites espalhados pelo genoma contaram a história demográfica dessas populações enquanto os microsatélites do MHC são também moldados pela história específica dessa região genômica.

Justificativa

Como exposto na seção anterior, a diversidade genética nas populações é moldada pela ação conjunta de fatores demográficos e seletivos. Um grande desafio nos estudos de genética de populações é caracterizar essa variação e inferir a importância desses dois fatores na origem e manutenção da diversidade. Uma abordagem que permite distinguir entre esses dois fatores é a análise de marcadores neutros e selecionados nos mesmos indivíduos de um conjunto de populações. No presente trabalho, apresentamos um estudo de caso sobre como forças seletivas e demográficas interagem no contexto do estudo da região do MHC, onde se encontram os genes *HLA* (presumivelmente selecionados) e marcadores genômicos (presumivelmente neutros, informativos da história demográfica) em populações nativas do continente americano.

Como a história que estamos tentando interpretar é recente, em termos evolutivos, necessitamos de um tipo de marcador molecular que tenha uma alta taxa evolutiva. Desta forma, optamos por utilizar microsatélites, que possuem alta taxa de mutação (10^{-3} a 10^{-6}), altos níveis de polimorfismo e possibilitam empreender estudos multi-locus, que permitem inferências histórico-demográfica e seletiva de alta resolução. Perguntas sobre a micro-evolução das populações nativas americanas, assim como a inferência de eventos demográficos recentes, são viáveis através dessa abordagem.

Sabemos que as populações nativas americanas apresentam padrões demográficos distintos das demais populações do mundo, apresentando menor variabilidade intra-populacional e maior diferenciação inter-populacional. No que diz respeito à variação genética intra-continental, há menos informação, mas estudos anteriores mostram que as populações sul-americanas da região oeste são genética e morfológicamente distintas das populações da região leste. Entretanto, tais resultados são baseados em bancos de dados extremamente limitados para dados de populações do leste do continente. O presente estudo pretende suprir a deficiência de amostragem da região das terras baixas da Amazônia e contribuir para entender os processos de ocupação e diferenciação das populações humanas na América do Sul.

A informação sobre a história e demografia dessas populações servirá para contextualizar nosso estudo sobre como a seleção natural atuou sobre os genes do sistema imune adaptativo, em especial na região do MHC.

No presente estudo analisamos 68 microssatélites autossômicos espalhados pelo genoma em um total de 333 indivíduos pertencentes a 16 populações, que serviram de base para a construção de cenários demográficos apropriados para testar as hipóteses de como a seleção atua sobre esses genes da região do MHC. Os dados para microssatélites da região do MHC, que consistem em 20 *loci* em 590 indivíduos de 30 populações, ajudarão a entender como a seleção atua ou atuou sobre o MHC, em especial sobre o gene *HLA-B*.

Capítulo I

Estrutura populacional e história demográfica de populações nativas da América do Sul

Introdução

A América do Sul e os movimentos populacionais que levaram a povoamento do continente têm sido estudados através de comparações entre os padrões de variação genética, linguística e a proximidade geográfica entre as populações.

Contudo, algumas dificuldades têm sido encontradas visto que grande parte dos estudos é realizada numa escala regional e poucos comparam populações de diferentes áreas do continente sul-americano. Além disso, muitas das análises são baseadas em poucos marcadores genéticos, que nem sempre são compartilhados por outros estudos, comprometendo a comparação entre os resultados dos dados populacionais.

Abordagens recentes com maior número de marcadores compartilhados com os de outros estudos populacionais têm sido mais informativas sobre padrões demográficos da América do Sul. Tanto dados moleculares como morfológicos evidenciam uma grande diferença entre o oeste e leste da América do Sul (WANG *et al.*, 2007; TARAZONA-SANTOS *et al.*, 2001; FUSELLI *et al.*, 2003; PUCCIARELLI *et al.*, 2006).

Em 2001, TARAZONA-SANTOS e colaboradores estudaram seis microssatélites ao longo do cromossomo Y em 12 populações nativas sul-americanas, e constataram que as seis populações localizadas na parte oeste do continente (Cayapá, Tayacajá, Arequipa, Susque, Mapuche e Toba) têm maior diversidade e menor diferenciação populacional que as seis populações do leste (Gavião, Xavante, Wai-wai, Karitiana, Ticuna e Lengua) ($F_{ST} = 0,024$ e $0,312$ respectivamente). O mesmo padrão foi observado em um estudo em 24 populações nativas, distribuídas ao longo do continente americano, com 678 microssatélites localizados em cromossomos autossômicos ($F_{ST} = 0,057$ e $0,147$ para as 11 populações do oeste e para 6 populações do leste da América do Sul respectivamente) (WANG *et al.*, 2007). Esses resultados sugerem que as populações que vivem na porção oeste da América do Sul exibem altos níveis de variabilidade intra-populacional e baixos níveis de variabilidade inter-populacional, comparando com as populações do leste (TARAZONA-SANTOS *et al.*, 2001; WANG *et al.*, 2007).

Com base nessa informação sobre variação genética intra e inter-populacional e considerações geográficas e paleo-ecológicas, TARAZONA-SANTOS *et al.* (2001) propuseram um modelo qualitativo para a evolução das linhagens masculinas na América do Sul. O modelo sugere que na parte oeste do continente (região Andina) as populações exibem grande tamanho efetivo e altos níveis de fluxo gênico entre elas, o que implica na homogeneização do componente genético. Em contraste, as populações do leste (Amazônia, Brasil Central e Chaco) exibem menor tamanho populacional, altos níveis de deriva genética e baixos níveis de fluxo gênico, que resulta numa maior diferenciação genética. Esse modelo sugere que as populações nativas sul americanas deveriam ser consideradas como dois grupos que evoluem de modos diferentes.

Ao observar as evidências de diferenças genéticas entre as populações nativas do oeste e leste da América do Sul, PUCCIARELLI *et al.* (2006) analisaram se o mesmo padrão era revelado com dados craniométricos, partindo do pressuposto que a morfologia facial é seletivamente neutra. Os autores encontraram fortes evidências de que o oeste e leste são dois universos com micro-evolução independente. Contudo, em contraste com TARAZONA-SANTOS *et al.* (2001) os dados de craniometria revelaram similar grau de variação inter-populacional para cada uma das regiões do continente sul americano. Os autores argumentam que a diferença encontrada por TARAZONA-SANTOS e colaboradores deve-se a características da amostragem utilizada, pois as populações andinas foram representadas por um pequeno número de amostras vindas de localidades geograficamente próximas, enquanto as populações do leste eram largamente distribuídas na região de terras baixas do Brasil a Argentina. PUCCIARELLI *et al.* (2006) sugerem que estudos genéticos com melhores amostragens da região dos Andes poderiam revelar padrões similares de variação inter-populacionais nos dois lados da cordilheira. Essa análise foi realizada por WANG *et al.* (2007), que encontraram o mesmo padrão descrito por TARAZONA-SANTOS *et al.* (2001). Contudo, esse trabalho apresenta baixa representação de tribos das terras baixas: setes populações, sendo três delas (Karitiana, Surui e Ache) formadas por indivíduos com alto grau de consanguinidade.

O presente estudo têm como objetivo analisar um conjunto de microssatélites em populações nativas da região do baixo rio Amazonas e Centro-Sul do Brasil, para suprir a deficiência amostral dos estudos anteriores e contribuir para esclarecer essa questão. Além disso, pretendemos investigar qual a contribuição genética das populações Andinas e do

Noroeste da América do Sul na formação das populações do Leste do continente sul americano.

Material e métodos

Amostras

Amostras para genotipagem dos microssatélites autossômicos espalhados pelo genoma

Foram genotipados 333 indivíduos pertencentes a 14 populações Ameríndias, sendo 11 delas localizadas nas terras baixas do rio Amazonas (Tiriyó, Zo'é, Arara do Laranjal, Arara do Iriri, Urubu-Kaapor, Awá-Guajá, Parakanã, Asurini, Araweté, Kayapó-Krokraimoro e Kayapó-Xikrin) e três na região Centro-Sul do Brasil (Guarani Kaiowá, Guarani M'bya e Kaingang). Amostramos também uma população Africana oriunda de Angola (Ovimbundu) e uma população Europeia (Portugueses) (tabela 1, figura 1), que servirão para contextualizar a variação genética nas populações nativas da América num contexto global, junto com dados previamente publicados.

As amostras de DNA utilizadas no presente estudo foram obtidas através de doação de bancos de DNA, sob tutela dos Professores Dr. Eduardo José Melo dos Santos e João Guerreiro do Laboratório de Genética Humana e Médica da Universidade Federal do Pará – LGHM UFPA (amostras indígenas da Amazônia), da Professora Dra. Maria Luiza Petzl-Erlor do Laboratório de Genética Molecular Humana da Universidade Federal do Paraná – LGMH UFPR (Guarani e Kaingang), do Professor Dr. Jorge Rocha da Universidade do Porto (Ovimbundu e Portugueses).

Todas as amostras utilizadas pelo presente estudo atendem aos requisitos das Resoluções CNS 196/96 e 304/00 sobre Diretrizes e Normas Regulamentadoras de Pesquisa envolvendo Seres Humanos, tendo aprovação do Comitê Nacional de Ética em Pesquisa (CONEP 2046, parecer 332/2001).

Bancos públicos de dados

Além das amostras genotipadas no presente estudo, utilizamos dados publicamente disponíveis para microssatélites genotipados no painel de diversidade do HGDP-CEPH (Human Genome Diversity Cell Line Panel do Centre d'Etude du Polymorphisme Humain), composto por 1064 indivíduos pertencentes a 54 populações, e os dados de WANG *et al.*, 2007, composto por 403 indivíduos pertencentes a 25 populações ameríndias e uma siberiana. A localização geográfica de todas as populações amostradas está apresentada na figura 1.

Tabela 1. Populações genotipadas pelo presente estudo, para 68 microssatélites localizados em cromossomos autossômicos espalhados pelo genoma.

População	Localização	Tamanho amostral (N)	Procedência
Tiryó	Pará	20	LGHM – UFPA
Zo'é	Pará	20	LGHM – UFPA
Arara do Laranjal	Pará	20	LGHM – UFPA
Arara do Iriri	Pará	18	LGHM – UFPA
Urubu-Kaapor	Maranhão	20	LGHM – UFPA
Awá-Guajá	Maranhão	14	LGHM – UFPA
Parakanã	Pará	11	LGHM – UFPA
Asurini	Pará	20	LGHM – UFPA
Awereté	Pará	20	LGHM – UFPA
Kayapó Krokraimoro	Pará	20	LGHM – UFPA
Kayapó Xikrin	Pará	20	LGHM – UFPA
Guarani Kaiowá	Mato Grosso do Sul	25	LGMH – UFPR
Guarani M'bya	Paraná	15	LGMH – UFPR
Kaingang	Paraná	25	LGMH – UFPR
Ovimbundu	Angola	25	CTM – UP
Portugueses	Portugal	40	CTM – UP
Total		333	

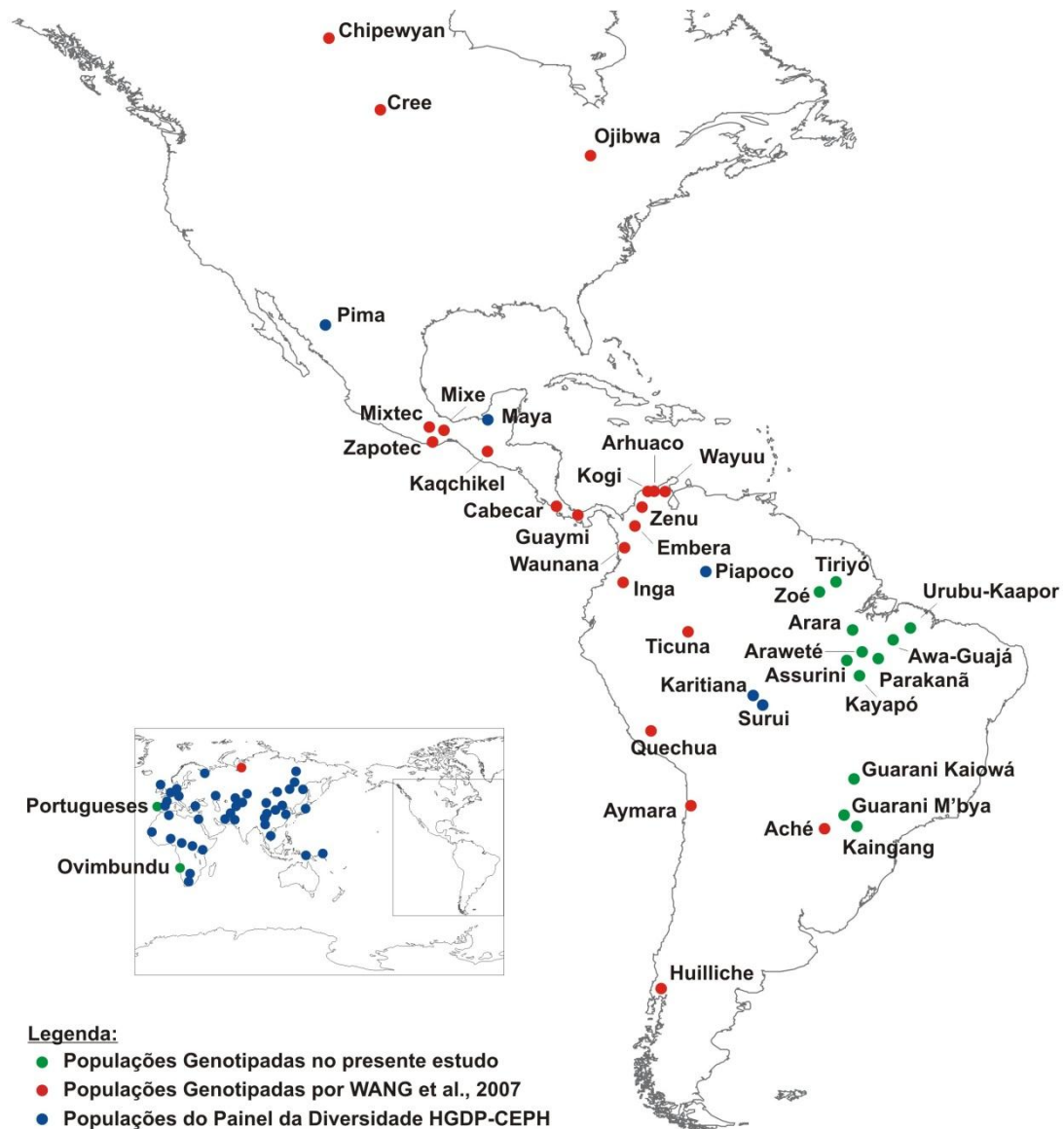


Figura 1. Mapa da localização geográfica das populações analisadas pelo presente estudo. Em verde estão aquelas genotipadas no presente estudo, em vermelho aquelas genotipadas por WANG *et al.* (2007), e em azul as do painel de diversidade HGDP-CEPH genotipadas por ROSENBERG *et al.* (2005). O ponto Ticuna representa as populações Ticuna Arara e Ticuna Tarapaca, Kayapó representa as populações Kayapó Krokraimoro e Kayapó Xikrin e Arara representa as populações Arara do Laranjal e Arara do Iriri.

Marcadores Moleculares

Microssatélites autossômicos espalhados pelo Genoma

Com o intuito de gerar dados que seriam comparáveis com aqueles previamente publicados por WANG *et al.* (2007) e com os do HGDP-CEPH (ROSENBERG *et al.*, 2005), genotipamos um subconjunto de 68 microssatélites, oriundos do conjunto de 678 microssatélites analisados por esses autores. Esses microssatélites fazem parte dos painéis set 16 e set 54, desenvolvidos pela empresa de genotipagem *Marshfield Screening* e cujas informações estão disponíveis no site <http://research.marshfieldclinic.org/genetics>.

Originalmente escolhemos microssatélites, com base nos seguintes critérios:

- 1) Apresentam unidades de repetições perfeitas;
- 2) Apresentam em média mais de quatro alelos nas populações nativas americanas;
- 3) Possuem coeficiente de endocruzamento (F_{IS}) com $p > 0,01$. Esse critério exclui possíveis alelos nulos gerados pela ausência de amplificação do produto de PCR, que podem gerar um excesso de homozigotos “falsos” na amostra, algo evidenciado no F_{IS} ;
- 4) A quantidade de dados faltantes por população é inferior a 3,3%. Dessa forma, eliminamos os microssatélites que apresentam maiores proporções de problemas na amplificação e maximizamos a quantidade de informação partilhada entre os marcadores do presente estudo e os de WANG *et al.* (2007).

Após filtrar os dados com esses critérios, 110 marcadores foram selecionados. Com o objetivo de agilizar o processo de genotipagem e possibilitar a multiplexagem dos microssatélites nas reações de PCR (Reação em Cadeia da Polimerase), os oligonucleotídeos iniciadores foram redesenhados da seguinte maneira: 1) obtivemos as sequências nucleotídicas dos 110 microssatélites e de suas regiões flanqueadoras, através do banco público de dados do NCBI (www.ncbi.nlm.nih.gov); 2) eliminamos sequências com possíveis sobreposições parciais com outras regiões do genoma humano, usando o programa BLAT (<http://genome.brc.mcw.edu/cgi-bin/hgBlat>); 3) utilizando o programa Primer 3 (<http://frodo.wi.mit.edu/>), para desenhar oligonucleotídeos iniciadores para reações de PCR multiplex. Após essas etapas obtivemos 32 microssatélites que foram genotipados no presente estudo

(tabela 2). Toda essa etapa foi realizada com a supervisão do Prof. Dr. Sidney dos Santos, da Universidade Federal do Pará, com auxílio da mestranda Maria Helena T. Maia.

Com base em análise preliminar realizada nesse conjunto de 32 microssatélites, constatamos que seria necessário aumentar o número de marcadores analisados para melhorar a resolução dos nossos resultados de estruturação populacional. Para esse fim, foi estabelecida colaboração com o Professor Dr. Jorge Rocha, da Universidade do Porto. Através dessa parceria genotipamos 45 microssatélites, dos quais 43 fazem parte dos 678 microssatélites genotipados no trabalho de WANG *et al.*, 2007 (tabela 2), sendo que sete microssatélites também fazem parte do conjunto de 32 previamente genotipados por nós. Dessa forma, obtivemos um conjunto final de 68 marcadores presentes nos nossos dados nos dados de WANG *et al.* (2007). e ROSENBERG *et al.* (2005).

Os microssatélites do grupo da UP foram selecionados com base nos seguintes critérios:

- 1) São tetranucleotídeos;
- 2) Distância de no mínimo 2Mb entre marcadores;
- 3) Taxa de heterozigose entre 0,6 e 0,8;
- 4) No mínimo 6 alelos nas populações africanas.

A genotipagem desses microssatélites foi realizada no Centro de Testagem Molecular (CTM) do Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO) da Universidade do Porto (UP), como parte de estágio realizado em Portugal.

Tabela 2. Lista dos 68 microssatélites selecionados, com informações sobre o cromossomo localizado, nome do *locus*, marcador e local onde foi realizada a genotipagem.

Cromossomo	<i>Locus</i>	Marcador	Local Genotipagem
1	D1S551	GATA6A05	USP/UP
1	NA-D1S-3	GATA133A08Q	UP
1	D1S1660	GATA48B01	USP
1	D1S1665	GATA61A06	UP
2	GATA148G10P_2	GATA148G10P	UP
2	D2S1328	GATA27A12	UP
2	D2S1391	GATA65C03M	UP
2	D2S2968	GATA178G09M	UP
3	D3S3038	GATA73D01	USP
3	GATA146B10_3	GATA146B10	UP
3	D3S3039	GATA7F05	USP
3	D3S1763	GATA3H01	USP
4	D4S2366	GATA22G05M	USP
4	D4S3248	GATA28F03	USP/UP
4	D4S2368	GATA27G03	USP
4	D4S1652	GATA5B02M	UP
5	D5S2500	GATA67D03	USP
5	D5S816	GATA2H09	USP
5	D5S1456	GATA11A11P	USP
6	GATA112F02P_6	GATA112F02P	UP
6	D6S474	GATA31	UP
6	D6S1040	GATA23F08	USP/UP
6	D6S1277	GATA81B01	UP
7	D7S3056	GATA24F03ZP	UP
7	D7S3047	GATA119B03	UP
7	D7S2846	GATA31A10	USP
7	D7S821	GATA5D08	USP
7	D7S3061	GGAA6D03N	USP
7	TATG002P_7	TATG002P	UP
8	D8S1136	GATA41A01	UP
8	D8S2324	GATA14E09	USP
8	D8S1128	GATA21C12	UP
9	D9S1121	GATA87E02N	UP
9	D9S1122	GATA89A11	UP
9	D9S301	GATA7D12	USP
9	D9S922	GATA21F05	USP
9	D9S934	GATA64G07	UP
10	D10S1435	GATA88F09	UP
10	D10S1426	GATA73E11	USP
10	D10S2327	GGAT1A4	USP
10	D10S2470	GATA115E01N	USP
10	D10S1248	GGAA23C05N	UP
11	D11S2365	GATA63F09	UP

Tabela 2 (continuação). Lista dos 68 microssatélites selecionados, com informações sobre o cromossomo localizado, nome do *locus*, marcador e local onde foi realizada a genotipagem.

Cromossomo	Locus	Marcador	Local Genotipagem
11	NEW01	GATA71E06	USP
12	D12S395	GATA4H01	USP
12	D12S1300	GATA85A04M	USP/UP
12	D12S373	GATA6C01	UP
13	D13S894	GATA86H01	UP
13	D13S1807	GATA11C08P	UP
13	D13S800	GATA64F08	UP
13	D13S317	GATA7G10	USP/UP
14	D14S1280	GATA31B09	UP
14	D14S588	GGAA4A12	USP
14	D14S1434	GATA168F06	UP
15	D15S1515	GATA197B10P	UP
16	D16S3253	GATA22F09P	UP
16	D16S539	GATA11C06N	USP
16	D16S764	GATA42E11	UP
17	GATA64B04P_17	GATA64B04P	UP
17	D17S1299	GATA25A04	USP
17	GATA63G01_17	GATA63G01	UP
18	D18S877	GATA64H04	UP
18	D18S535	GATA13	USP/UP
19	D19S559	UT7544	UP
20	D20S482	GATA51D03	USP/UP
21	D21S1432	GATA11C12	UP
21	GATA163G03_21	GATA163G03	UP
21	D21S1437	GGAA3C07	USP
22	D22S689	GATA21F03	UP

Metodologia

Extração de DNA

As amostras analisadas tiveram seu DNA extraído pelos métodos de fenol-clorofórmio (SAMBROOK *et al.*, 1989) ou *salting out* (LAHIRI e NURENBERG, 1991) nos respectivos laboratórios que as doaram.

Amplificação do genoma inteiro

As amostras de DNA utilizadas passaram pelo processo de amplificação inteira do genoma realizado com o kit *GenomiPhi V2 DNA Amplification Kit (GE Healthcare)* seguindo instruções do fabricante.

Amplificação de fragmentos os microsatélites

As amplificações dos fragmentos de DNA contendo microsatélites foram realizadas através de multiplexagem dos oligonucleotídeos iniciadores na reação de PCR (tabela 3).

Para os microsatélites genotipados na USP, utilizamos 3mM de MgCl₂, 0,2mM DNTPs, tampão 1x, 0,06pmol de cada iniciador, 0,07U/uL de taq polimerase e cerca de 10ng de DNA. O seguinte protocolo de amplificação foi realizado: 94°C por 11 minutos seguidos por 30 ciclos de 94°C por 1 minuto, 60°C por 1 minuto, 72°C por 1 minuto e um ciclo de extensão final de 60°C por 1 hora. Essa metodologia foi desenvolvida com supervisão do prof. Dr. Sidney dos Santos da UFPA, com auxílio da aluna Elzemar dos Santos.

Para os microsatélites genotipados no CTM da UP, utilizamos o kit de amplificação Taq PRC Master Mix da Qiagen. O protocolo de amplificação foi: 95°C por 15 minutos seguidos por 11 ciclos de 95 °C por 30 segundos, touch down de 61 a 56°C por 30 segundos, e 72°C por 30 segundos, seguidos de 31 ciclos de 95°C por 30 segundos, 56°C por 1 minuto e 72°C por 39 segundos seguidos de extensão final de 60°C por 30 minutos. Essa metodologia foi desenvolvida com supervisão da técnica em biologia molecular Susana Lopes e com auxílio da mestrandia Dora Henriques.

Confirmação de amplificação dos produtos de PCR

A confirmação de amplificação dos produtos de PCR foi realizada através da visualização em gel de agarose. Essa visualização foi feita homogenizando 2µL de corante GelRed (*GelRedTM Nucleic Acid Gel Stain - Biotium*) em 2µL de produto de PCR e aplicados em gel de agarose 3%. A eletroforese foi realizada com 90V e 45mA por 35 minutos e em seguida visualizada no transiluminador (Figura 2).

Tabela 3. Microsatélites agrupados nos conjuntos de multiplexagem utilizados para as genotipagens, com informações sobre a variação do tamanho do fragmento contendo o microsatélite em pares de base (pb) e fluorescência utilizada.

MULTIPLEX 1			MULTIPLEX 2		
<i>Locus</i>	Tamanho (pb)	Fluorescência	<i>Locus</i>	Tamanho (pb)	Fluorescência
<i>D5S816</i>	122-154	FAM	<i>D13S317</i>	116-156	FAM
<i>D1S551</i>	167-187	FAM	<i>D8S2324</i>	198-218	FAM
<i>D7S3061</i>	204-240	FAM	<i>D19S559</i>	229-249	FAM
<i>D4S3248</i>	268-296	FAM	<i>D6S1040</i>	285-309	FAM
<i>D16S539</i>	137-169	HEX	<i>D20S482</i>	106-146	HEX
<i>D9S922</i>	180-204	HEX	<i>D9S301</i>	197-233	HEX
<i>D10S1426</i>	221-237	HEX	<i>D18S535</i>	242-270	HEX
<i>D7S821</i>	265-297	HEX	<i>D21S1437</i>	208-244	HEX
MULTIPLEX 3			MULTIPLEX 4		
<i>Locus</i>	Tamanho (pb)	Fluorescência	<i>Locus</i>	Tamanho (pb)	Fluorescência
<i>D14S588</i>	203-243	FAM	<i>D12S395</i>	207-235	FAM
<i>D5S1456</i>	234-258	FAM	<i>D17S1299</i>	248-276	FAM
<i>D4S2368</i>	288-312	FAM	<i>D1S1660</i>	320-348	FAM
<i>D3S1763</i>	315-335	FAM	<i>D3S3039</i>	351-391	FAM
<i>D5S2500</i>	327-355	FAM	<i>D10S2470</i>	220-248	HEX
<i>D7S2846</i>	203-237	HEX	<i>D4S2366</i>	326-350	HEX
<i>D12S1300</i>	233-259	HEX	<i>New</i>	355-379	HEX
<i>D10S2327</i>	259-287	HEX	<i>D3S3038</i>	380-416	HEX
MULTIPLEX 5			MULTIPLEX 6		
<i>Locus</i>	Tamanho (pb)	Fluorescência	<i>Locus</i>	Tamanho (pb)	Fluorescência
<i>D18S877</i>	109-141	VIC	<i>D2S1328</i>	138-170	VIC
<i>D16S764</i>	92-124	FAM	<i>D6S474</i>	139-179	FAM
<i>D2S2968</i>	167-207	NED	<i>D20S482</i>	139-171	NED
<i>D13S1807</i>	182-214	ROX	<i>D7S3047</i>	156-188	ROX
<i>GATA22F09P</i>	163-203	VIC	<i>GATA21F03</i>	194-234	VIC
<i>GATA6A05</i>	158-190	FAM	<i>GATA61A06</i>	211-255	FAM
<i>GATA41A01</i>	229-269	NED	<i>GATA197B10</i>	196-224	NED
<i>GATA88F09</i>	244-288	ROX	<i>GATA168F06</i>	208-236	ROX
<i>GATA63F09</i>	226-258	VIC	<i>GATA81B01</i>	278-318	VIC
<i>GGAA23C05</i>	221-269	FAM	<i>GATA31B09</i>	277-312	FAM

Continua

Tabela 3 (continuação). Microsatélites agrupados de acordo com os conjuntos de multiplexagem utilizados para as genotipagens, com informações sobre a variação do tamanho do fragmento de microsatélite em pares de base (pb) e a marcação fluorescente utilizada.

MULTIPLEX 7			MULTIPLEX 8		
<i>Locus</i>	Tamanho (pb)	Fluorescência	<i>Locus</i>	Tamanho (pb)	Fluorescência
<i>D4S1652</i>	126-154	VIC	<i>GATA133A08</i>	102-167	VIC
<i>D12S1300</i>	111-143	FAM	<i>D6S1270</i>	112-144	FAM
<i>D9S1122</i>	186-214	NED	<i>D2S1391</i>	102-154	NED
<i>D9S1121</i>	172-228	ROX	<i>GATA148G10</i>	147-207	ROX
<i>D13S894</i>	176-212	VIC	<i>D17S1534</i>	206-250	VIC
<i>D7S3056</i>	164-192	FAM	<i>D13S317</i>	167-207	FAM
<i>D6S1040</i>	257-289	NED	<i>D12S373</i>	192-232	NED
<i>D8S1128</i>	240-268	ROX	<i>GATA146B10</i>	230-270	ROX
<i>D21S2053</i>	255-263	VIC	<i>D13S800</i>	283-331	VIC
<i>D4S3248</i>	229-265	FAM	<i>D17S1303</i>	232-252	FAM

MULTIPLEX 9		
<i>Locus</i>	Tamanho (pb)	Fluorescência
<i>D21S1432</i>	131-159	FAM
<i>TATG002P</i>	147-187	FAM
<i>D18S535</i>	123-167	NED
<i>D9S934</i>	198-242	FAM

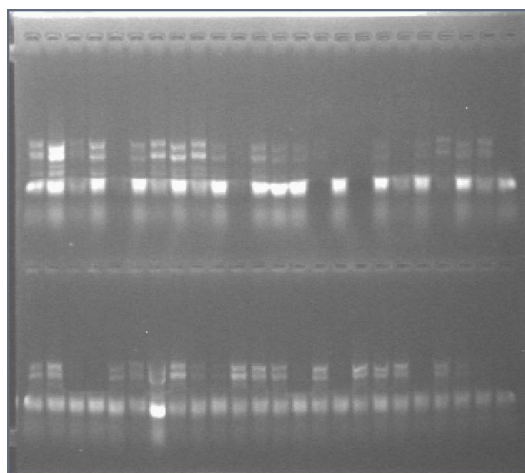


Figura 2. Gel de agarose 3% para verificação da amplificação dos produtos de PCR para posterior aplicação na eletroforese em ABI 3031.

Reação de genotipagem dos microssatélites

Reação de desnaturação

Para a desnaturação foram misturados 0,5µL do produto de PCR acrescido de 0,5µL do marcador de tamanho molecular (*GeneScanTM – 500 ROXTM – Applied Biosystems*) e 9µL de formamida (*HI-DITM formamide – Applied Biosystems*). Essa solução foi homogeneizada e aquecida a 95°C por 3 minutos e em seguida resfriada em gelo por 20 minutos.

Eletroforese

A eletroforese para separação dos diferentes tamanhos de fragmentos de DNA contendo os microssatélites foi realizada no *ABI 3130 Genetic Analyzer* do Centro de Estudos do Genoma Humano da Universidade de São Paulo, ou do Centro de Testagem Molecular da Universidade do Porto.

Análise das genotipagens

Os microssatélites genotipados no Centro de Estudos do Genoma Humano da Universidade de São Paulo foram analisados com o auxílio do programa *Peak ScannerTM Software v.1.0* da *Applied Biosystem* e os genotipados no Centro de Testagem Molecular da Universidade do Porto, no programa *Gene Mapper[®] Software v.4.1* da *Applied Biosystem*.

Obtenção do banco de dados final utilizado nas análises

Usamos os seguintes critérios de exclusão para garantir que as amostras e os *loci* a serem usados nas análises introduzissem o mínimo de erro experimental:

- 1) Indivíduos com menos de 50% de amplificação dos *locus* selecionados (24 indivíduos excluídos).
- 2) *loci* com menos de 50% de sucesso de amplificação nas amostras (dois *loci* excluídos: *D5S1456* e *D14S588*).
- 3) *loci* com falha total de amplificação em alguma população (um *loci* excluído *D21S1437* eliminado).

- 4) *loci* com unidades de repetição imperfeita (dois *loci* excluídos *GATA81B01* e *GATA31B09*).
- 5) *loci* sem correspondência no banco de dados de WANG *et al.* (2007) (dois *loci* excluídos: *GATA61A* e *GATA42E*).

Dessa maneira o banco de dados final é composto por 307 amostras (tabela 4) e 61 marcadores moleculares do tipo microsatélites. Sendo esses os dados utilizados nas análises.

Tabela 4. Número de amostras utilizadas após a eliminação de indivíduos com excesso de falha de amplificação.

População	Tamanho amostral inicial	Tamanho amostral final após eliminação
Tiryó	20	17
Zo'é	20	13
Arara do Laranjal	20	16
Arara do Iriri	18	18
Urubu-Kaapor	20	20
Awá-Guajá	14	11
Parakanã	11	9
Asurini	20	20
Awereté	20	20
Kayapó Krokraimoro	20	18
Kayapó Xikrin	20	20
Guarani Kaiowá	25	23
Guarani M'bya	15	15
Kaingang	25	22
Ovimbundu	25	25
Portugueses	40	40
Total	333	307

Calibração dos dados

Para possibilitar a comparação dos dados deste estudo com os previamente publicados, cinco amostras genotipadas por WANG *et al.* (2007) foram regenotipadas no presente estudo. A tabela 5 mostra os marcadores e os ajustes realizados no nosso banco de dados para ele ser integrado ao banco de dados pré-existente. A calibração dos dados se fez necessária uma vez que usamos oligonucleotídeos iniciadores e equipamento de genotipagem distintos dos utilizados por WANG e colaboradores (2007).

Análise de dados

Estimativa de erro de genotipagem

Sete microssatélites (*D1S551*, *D4S3248*, *D6S1040*, *D12S1300*, *D13S317*, *D18S535*, *D20S482*) foram genotipados duas vezes em todas as amostras, uma vez com cada par de oligonucleotídeos iniciadores distintos (aqueles desenhados pelo grupo da USP e aqueles pelo grupo da UP). A comparação dos genótipos dos mesmos indivíduos (~300) permitiu estimar, através de contagem direta, o erro de genotipagem.

Frequências alélicas

As frequências alélicas foram obtidas através da contagem direta do número de cromossomos e de indivíduos, respectivamente.

Equilíbrio de Hardy-Weinberg

As frequências observadas dos genótipos foram comparadas com aquelas esperadas de acordo com o princípio de equilíbrio de Hardy-Weinberg, através do teste exato de GUO e THOMPSON (1992), implementado no pacote de programas do Arlequin 3.5 (EXCOFFIER e LISCHER, 2010). Essa análise permite verificar a confiabilidade da metodologia de genotipagem dos microssatélites, uma vez que a existência de alelos nulos pode resultar em dados com excesso de homozigotos.

Tabela 5. Número de pares de base (pb) acrescentados ou diminuídos para calibração dos dados gerados no presente estudo com o banco de dados de WANG *et al.*(2007).

<i>Locus</i>	Calibração	<i>Locus</i>	Calibração
<i>D1S551</i>	-1	<i>D9S1122</i>	3
<i>NA-D1S-3</i>	2	<i>D9S922</i>	71
<i>D1S1660</i>	-98	<i>D9S934</i>	4
<i>GATA148G10P_2</i>	1	<i>D10S1435</i>	4
<i>GATA27A12</i>	0	<i>D10S1426</i>	-65
<i>D2S1391</i>	0	<i>D10S2327</i>	-59
<i>D2S2968</i>	1	<i>D10S2470</i>	23
<i>D3S3038</i>	-189	<i>D10S1248</i>	2
<i>GATA146B10_3</i>	13	<i>D11S2365</i>	3
<i>D3S3039</i>	-83	<i>GATA71E06_11</i>	-180
<i>D3S1763</i>	-55	<i>D12S373</i>	1
<i>D4S2366</i>	-156	<i>D12S1300</i>	-122
<i>D4S3248</i>	-35	<i>D12S395</i>	12
<i>D4S2368</i>	20	<i>D13S894</i>	1
<i>D4S1652</i>	2	<i>D13S1807</i>	-8
<i>D5S2500</i>	-178	<i>D13S800</i>	5
<i>D5S816</i>	103	<i>D13S317</i>	55
<i>GATA112F02P_6</i>	4	<i>D14S1434</i>	1
<i>D6S474</i>	2	<i>D15S1515</i>	-5
<i>D6S1040</i>	-20	<i>D16S3253</i>	-6
<i>D7S3056</i>	31	<i>D16S539</i>	7
<i>D7S3047</i>	-1	<i>GATA64B04P_17</i>	2
<i>D7S2846</i>	-35	<i>D17S1299</i>	-64
<i>D7S821</i>	-27	<i>GATA63G01_17</i>	1
<i>D7S3061</i>	-90	<i>D18S877</i>	3
<i>TATG002P_7</i>	0	<i>D18S535</i>	-111
<i>D8S1136</i>	1	<i>D19S559</i>	-57
<i>D8S2324</i>	-2	<i>D20S482</i>	33
<i>D8S1128</i>	2	<i>GATA163G03_21</i>	0
<i>D9S1121</i>	4	<i>D22S689</i>	2
<i>D9S301</i>	8		

Descrição da variabilidade genética

Os índices de diversidade genética foram estimados pelo do pacote computacional Arlequin 3.5 (EXCOFFIER e LISCHER, 2010).

A diversidade gênica corresponde à heterozigose esperada (probabilidade de amostrar ao acaso dois alelos em uma dada população e eles serem diferentes). A diversidade gênica (\hat{H}) e a sua variância amostral ($V(\hat{H})$) foram calculadas da seguinte maneira:

$$\hat{H} = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$$

$$V(\hat{H}) = \frac{2}{n(n-1)} \left\{ 2(n-2) \left[\sum_{i=1}^k p_i^3 - \left(\sum_{i=1}^k p_i^2 \right)^2 \right] + \sum_{i=1}^k p_i^2 - \left(\sum_{i=1}^k p_i^2 \right)^2 \right\}$$

Onde n é o número de cópias gênicas na amostra, k o número de alelos e p_i a frequência dos i alelos.

O theta-H é um estimador de theta (4Nu). Ele é específico para o caso de microssatélites, supondo o "*stepwise mutation model*" (modelo em que cada mutação implica na perda ou ganho de uma unidade de repetição, permitindo apenas um passo mutacional). Essas qualificações são importantes, pois a expressão acima não se aplica a outras classes de marcadores. Essa estimativa foi obtida a partir de:

$$\hat{\theta}_H = \frac{1}{(1-\hat{H})^2} - 1,$$

Onde \hat{H} é a heterozigose esperada.

O número de alelos (K) foi estimado por contagem direta do número de alelos observados em cada *locus* e população

A amplitude no tamanho dos alelos (R) foi obtida através da diferença entre os alelos com maior e menor número de repetições observado nas populações.

Diferenciação Inter-populacional

Matrizes de diferenciação entre os pares de populações foram estimadas utilizando o programa Arlequin 3.5 (EXCOFFIER e LISCHER, 2010) para os índices de diferenciação populacional F_{ST} (estima a diferença de probabilidade de identidade por descendência de dois alelos de uma mesma população em relação a dois alelos amostrados em populações diferentes, baseado no modelo de infinitos alelos) e R_{ST} (análogo ao F_{ST} mas baseado no modelo mutacional *stepwise*). As significâncias foram obtidas através 10.000 permutações dos indivíduos entre as populações.

A diferenciação populacional para F_{ST} e R_{ST} entre as regiões e ou grupos de populações foi estimada através do particionamento hierárquico da variação genética ao longo dos *loci* (AMOVA). O intervalo de confiança foi obtido realizando reamostragens ao longo dos *loci* através de 10.000 *bootstrap*.

O teste de Mantel foi utilizado para testar a significância da correlação entre as matrizes de diferenciação de F_{ST} e R_{ST} entre os pares de população.

Análise de estruturação populacional

As análises de estruturação populacional foram realizadas usando agrupamento Bayesiano implementado no programa *STRUCTURE* v.2.3.3 (PRITCHARD, WEN e FALUSH, 2010). Essa abordagem identifica grupos de indivíduos que apresentam frequência alélica similar e os agrupam, não necessitando de classificação *a priori* dos indivíduos nas populações.

As corridas foram realizadas assumindo o modelo de miscigenação, em que cada indivíduo pode ter ancestralidade em vários grupos, usando o modelo F de correlação das frequências alélicas ao longo dos grupos. Cada corrida foi realizada utilizando Cadeia de Markov, com *burn-in* de 20.000 seguidos por 10.000 interações.

Para as análises na América, 100 repetições foram efetuadas para cada cluster (denominado de K) (K=2 a K=20). Nas análises com as populações mundiais (HGDP-CEPH – ROSENBERG *et al.* (2005) e nativas da América (WANG *et al.*, 2007) dez repetições foram efetuadas para cada K (2 a 20).

Para resolver os problemas de multimodalidade (cenários alternativos para o padrão de estruturação populacional) e escolher o melhor cenário, utilizamos o programa *CLUMPP* (JAKOBSSON e ROSENBERG, 2007). Assumimos que os pares de corridas com algoritmo de Greedy $\geq 0,9$ apresentam a mesma moda e, portanto correspondem ao mesmo padrão de estruturação.

Os gráficos com os resultados foram gerados no programa *DISTRUCT* (ROSENBERG, 2004).

Teste de populações ancestrais

Através de análise supervisionada realizada no programa *STRUCTURE* v.2.3.3 (PRITCHARD, WEN e FALUSH, 2010), fixamos os grupos Noroeste da América do Sul e Andes para averiguar qual deles apresente maior proximidade e influência nas populações do Leste da América do Sul. Cada corrida foi realizada utilizando Cadeia de Markov, com 20.000 *burn-in* e 10.000 interações.

Estimativa de miscigenação

Os indivíduos do banco do HGDP-CEPH (ROSENBERG *et al.*, 2005) e WANG *et al.* (2007) foram utilizados como populações parentais para estimar a proporção de miscigenação, nos indivíduos genotipados no presente estudo. Para tanto utilizamos a abordagem Bayseana implementada no programa computacional *STRUCTURE* v.2.3.3 (PRITCHARD, WEN e FALUSH, 2010). A informação sobre a origem populacional dos indivíduos previamente genotipados foi utilizada para forçar as populações africanas, europeias e nativas americanas a formarem três grupos separados. Em seguida estimamos a probabilidade dos indivíduos pertencerem aos grupos em que foram inseridos. Todos os indivíduos que apresentam probabilidade inferior a 50% de pertencer aquele grupo foram excluídos das análises (ver tabela A.1 e figura A.1 – ANEXO 1).

Em seguida estimamos o grau de miscigenação nas populações Ameríndias do presente estudo, assumindo desconhecer sua origem populacional e admitindo o modelo de miscigenação em que cada indivíduo pode ter ancestralidade em qualquer grupo pré-estabelecido. Para isso, assumimos diferentes probabilidades (1 a 10% - como sugerido pelos autores do *STRUCTURE*) de haver na população indivíduos com ancestralidade em

outro grupo nesta ou até quatro gerações anteriores. Cada corrida foi realizada utilizando Cadeia de Markov (com *burn-in* de 20.000 seguidos por 10.000 interações), baseada na amostragem de Gibbs para gerar a distribuição da proporção de miscigenação a partir dos genótipos observados.

Estimativa de gargalo populacional

Usamos o Índice de Desequilíbrio β ($\beta = \Theta (V) / \Theta (He)$) descrito por KIMMEL *et al.* (1998) e KING *et al.* (2000) para estimar eventos de gargalos e expansões populacionais. O índice leva em consideração a relação entre o parâmetro mutacional Θ ($4Ne\mu$) da variância do número de repetições dos alelos de microssatélite (V) e da heterozigose esperada (He). Em populações que estão em equilíbrio entre deriva e mutação espera-se que a razão entre $\Theta (V) / \Theta (He)$ seja igual a 1. Populações que passam por um gargalo populacional apresentam um desequilíbrio entre a variância no tamanho dos alelos de microssatélites e heterozigose fazendo com que a razão seja >1 .

Assumindo o modelo mutacional *stepwise*, o $\Theta (He)$ foi estimado por:

$$\theta = \frac{1}{2} \left(\frac{1}{(He - 1)^2} - 1 \right)$$

E o Θ da variância no número de repetições dos alelos de microssatélites (V) por:

$$\theta = 2V$$

Resultados

Avaliação do banco de dados

Basicamente temos duas fontes primárias de erros que podem comprometer o sucesso das nossas análises: a) erros de genotipagem (causadas por vários fatores como qualidade das amostras de DNA, interação dos produtos de PCR no processo de multiplexagem e erros humanos (anotação e troca de amostras); b) erros de calibração que tornariam a base de dados gerados no presente estudo incompatível para a comparação com os bancos de dados previamente publicados (HGDP-CEPH – ROSENBERG *et al.*, 2005 e WANG *et al.*, 2007).

Para certificar que nosso banco de dados apresenta características semelhantes ao banco de dados ao qual será integrado, realizamos algumas análises, cujos resultados serão apresentados a seguir.

Quantificação das falhas de amplificação

Após a exclusão de indivíduos e *loci* com excesso de falha de amplificação (ver item “Obtenção do banco de dados final utilizado nas análises” da Seção Materiais e Métodos), reanalisamos o banco de dados. Obtivemos 5,86% de falha de amplificação ao longo dos 61 *loci* de microssatélites das 14 populações nativas sul-americanas genotipadas. Esse valor é ligeiramente maior que o encontrado no banco de dados utilizado por WANG e colaboradores para esses 61 microssatélites (3,7%), contudo ainda dentro de uma margem de falha aceitável.

Para verificar se a falha de amplificação está concentrada em algum *locus* ou população em particular, verificamos a percentagem de falha de amplificação para cada população nativa americana analisada (figura 3) e para cada um dos 61 *loci* de microssatélites (figura 4). Apesar das percentagens de falha de amplificação serem mais elevadas no nosso banco de dados, nenhum valor excede a linha de corte utilizada no estudo de WANG *et al.*, 2007.

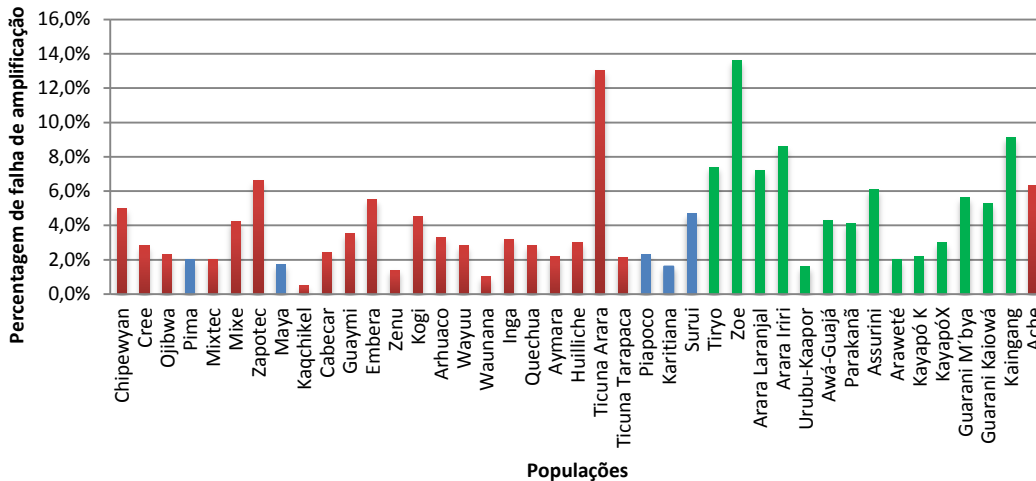


Figura 3. Percentagem de falha de amplificação ao longo das populações nativa americanas nos 61 *loci* de microssatélites. Em vermelho populações analisadas por WANG *et al.*, 2007, em azul populações do HGDP-CEPH e em verde populações analisadas no presente estudo.

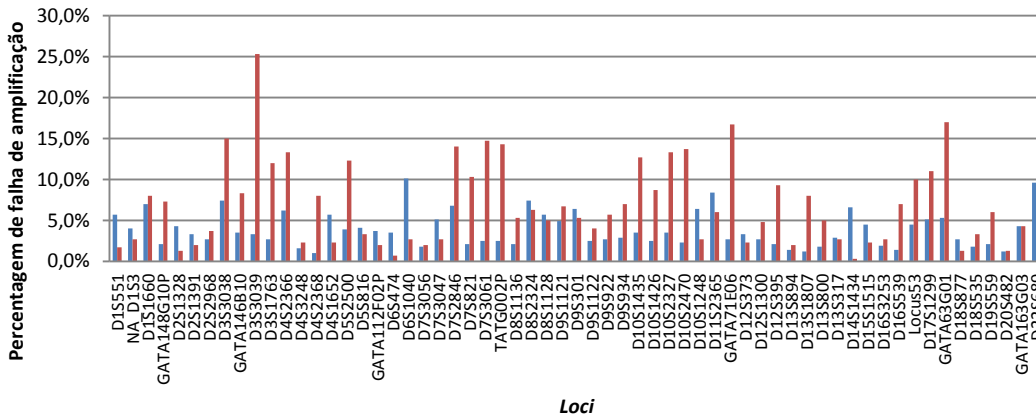


Figura 4. Percentagem de falha de amplificação ao longo dos 61 *loci* de microssatélites nas populações nativas americanas. Em azul populações analisadas por WANG *et al.*, 2007 e HGDP-CEPH, em vermelho populações analisadas no presente estudo.

Estimativa de erro de genotipagem

Sete microssatélites (*D1S551*, *D4S3248*, *D6S1040*, *D12S1300*, *D13S317*, *D18S535*, *D20S482*) foram genotipados duas vezes em todas as amostras e a comparação das duas genotipagens nos mesmos indivíduos permitiu estimar a percentagem de erro para cada *locus*. Os resultados dessa análise são apresentados na tabela 6.

A taxa média de erro de genotipagem estimada foi de 3,59% (\pm 0,97%). Desses, 0,048% são compostos por alelos falsos (alelos distintos identificados nas duas genotipagens) e 3,54% consistem em indivíduos genotipados como heterozigotos em uma das genotipagem e homozigotos na outra.

Tabela 6. Estimativa em percentagem (%) do erro de genotipagem do presente estudo, com base em sete *loci*.

<i>Loci</i>	Erro genotipagem (%)
<i>D1S551</i>	2,33%
<i>D4S3248</i>	4,66%
<i>D6S1040</i>	3,88%
<i>D12S1300</i>	3,00%
<i>D13S317</i>	4,33%
<i>D18S535</i>	2,50%
<i>D20S482</i>	4,50%
Média	3,59%
Desvio	0,97%

Estimativas de desvios de Hardy-Weinberg

Observamos 5,9% de desvios significativos de Hardy-Weinberg ($p < 0,05$) em nosso conjunto total de dados. Essa estimativa de desvio é semelhante à encontrada no banco de dados de WANG e colaboradores para o conjunto de 61 microssatélites (6,9%).

As figuras 5 e 6 mostram a distribuição dos desvios por população e por *locus* respectivamente. É possível observar que a distribuição dos desvios significativos ao longo das populações é homogênea nos dois estudos. Por outro lado, a distribuição

dos desvios ao longo dos *loci* mostra variação entre os dois estudos. Apesar disso, nenhuma percentagem excede a linha de corte adotada por WANG *et al.*, 2007.

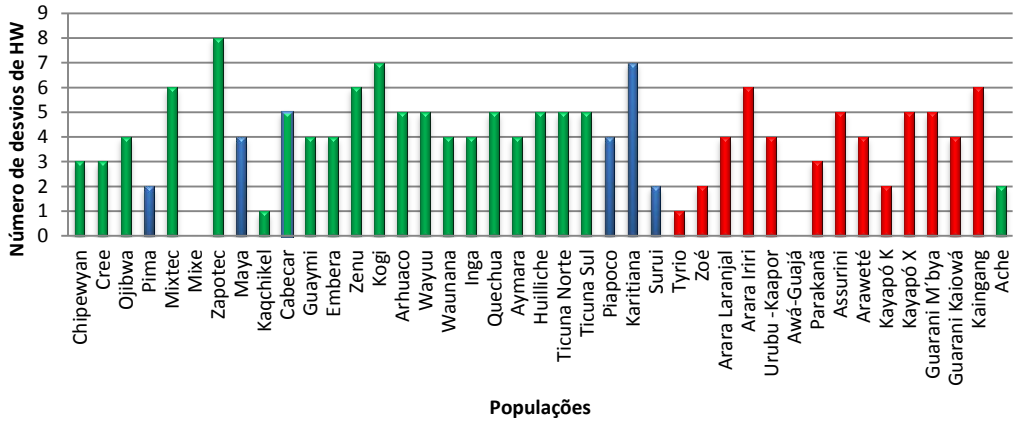


Figura 5. Distribuição dos desvios significativos de Hardy-Weinberg ($p < 0,05$) nas populações Ameríndias para o conjunto de 61 *loci* de microssatélites. Em vermelho, populações genotipadas no presente estudo, em azul populações genotipadas no estudo de WANG *et al.* 2007.

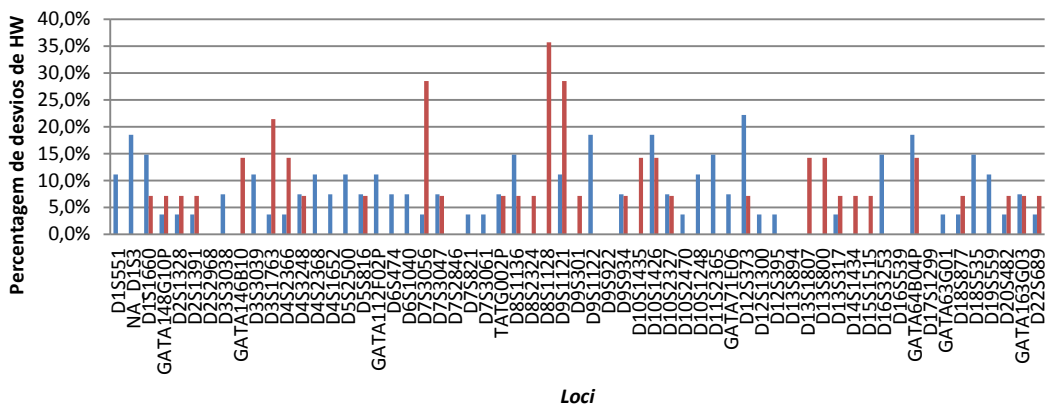


Figura 6. Distribuição dos desvios significativos de Hardy-Weinberg ($p < 0,05$) para cada um dos 61 *loci* das populações Ameríndias. Em azul populações genotipadas por WANG *et al.*, 2007 em vermelho populações genotipadas no presente estudo.

Para minimizar erros de interpretação dos resultados, nós realizamos todas as análises de estrutura e diferenciação populacional eliminando os três microssatélites *outliers* (*D7S3056*, *D8S1128* e *D9S1121*). Nenhuma mudança de resultados foi observada, portanto todos os resultados apresentados no presente estudo serão referentes às análises realizadas com o banco de dados incluindo esses três microssatélites.

Em seguida, comparamos as estimativas de heterozigose observada e esperada de acordo com o princípio de equilíbrio de Hardy-Weinberg. Verificamos que no presente estudo, 66,67% dos desvios são por excesso de homozigose (sendo 82,57% encontrado no estudo de WANG e colaboradores).

Após efetuarmos todas essas análises, nós aplicamos o método de Bonferroni para corrigir o valor de significância (*p*-valor) e verificamos que nenhum valor se manteve significativo ($p_{\text{Bonferroni}} = 0,00006$).

Calibração dos dados

Para possibilitar a comparação dos nossos dados com o banco de dados do HGDP-CEPH e WANG e colaboradores, cinco indivíduos, pertencentes à população Guaymi (Panamá), genotipados por esses autores, foram regenotipados no presente estudo (ver item Calibração dos dados da seção Material e Métodos). Para nos certificarmos que a calibração dos dados foi executada de modo correto utilizamos a análise de estruturação populacional.

Foi possível observar que para todos os K testados (K=2 a 22) sempre ocorreu o agrupamento dos indivíduos Guaymi genotipados no presente estudo com os indivíduos Guaymi genotipados por WANG *et al.* (2007) (figura 7). Sendo que a comparação dos genótipos dos indivíduos sobrepostos entre os dois estudos mostrou ser idêntica.

Também verificamos que, para todos os K testados (K=2 a 22), as populações Ovimbundu (Angola) e Portugueses, genotipadas no presente estudo, sempre agrupam com as demais populações Africanas e Europeias do banco público do HGDP-CEPH (figura 7).

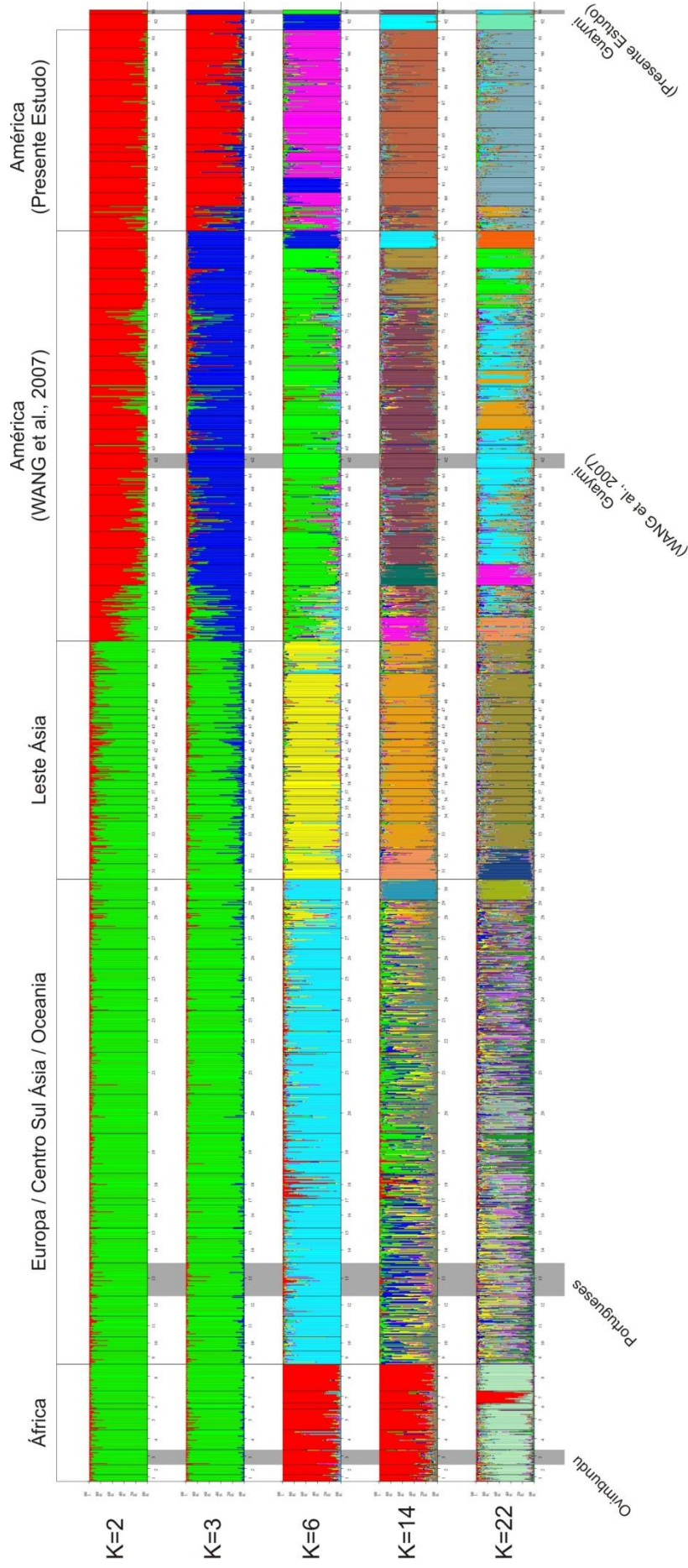


Figura 7. Averiguação de calibração dos dados para dados HGDP-CEPH, WANG *et al.*, 2007 e presente estudo. Para todos os K testados (2 a 22) sempre há agrupamento de Ovimbundu e Portugueses com África e Europa respectivamente. E correspondência de agrupamento de Guaymi genotipados presente estudo com os Guaymi genotipados por WANG *et al.*, 2007.

WANG e colaboradores genotiparam em seu estudo um pequeno número de indivíduos das populações Guarani e Kaingang (10 e 7 indivíduos respectivamente). No presente estudo, nós aumentamos a amostragem de indivíduos nessas populações (25 e 23 respectivamente). Novamente, por análise de estruturação populacional, foi possível observar que as populações Guarani e Kaingang genotipadas no presente estudo agrupam com as populações Guarani e Kaingang genotipadas por WANG e colaboradores (figura A.2 – ANEXO 1). Tendo em vista que para essas duas populações nós não temos a informação exata de quais foram os indivíduos utilizados por WANG (há ambiguidade no código utilizado pelos autores) e, portanto para evitar que o mesmo indivíduo fosse amostrado duas vezes, optamos por realizar todas as análises de Guarani e Kaingang utilizando apenas os indivíduos genotipados no presente estudo.

Para averiguar a sensibilidade do método de agrupamento implementado no programa STRUCTURE a um eventual erro de calibração, fizemos um teste onde dois *loci* foram propositalmente não calibrados. Isso foi feito criando um banco de dados em que as populações Ovimbundu, Portuguesas e Nativas americanas genotipadas no presente estudo não foram calibradas para dois *loci* em relação às demais populações (ou seja, a calibração correta de +4 pb no *locus D9S934* e -8pb no *locus D13S1807*, não foi efetuado nessas populações). A análise de estruturação populacional para todo o banco de dados revelou que se houvesse erro de calibração as populações Ovimbundu, Portugueses apresentariam componentes tanto de sua região de origem como das populações nativas americanas do presente estudo (figura 8).

Com base nessas análises nos asseguramos que não há erros de calibração em nossa base de dados.

Estimativa de miscigenação

A miscigenação com outros povos não indígenas também pode distorcer a interpretação dos nossos resultados. Por essa razão achamos prudente conhecer o grau de miscigenação nos indivíduos analisados no presente estudo.

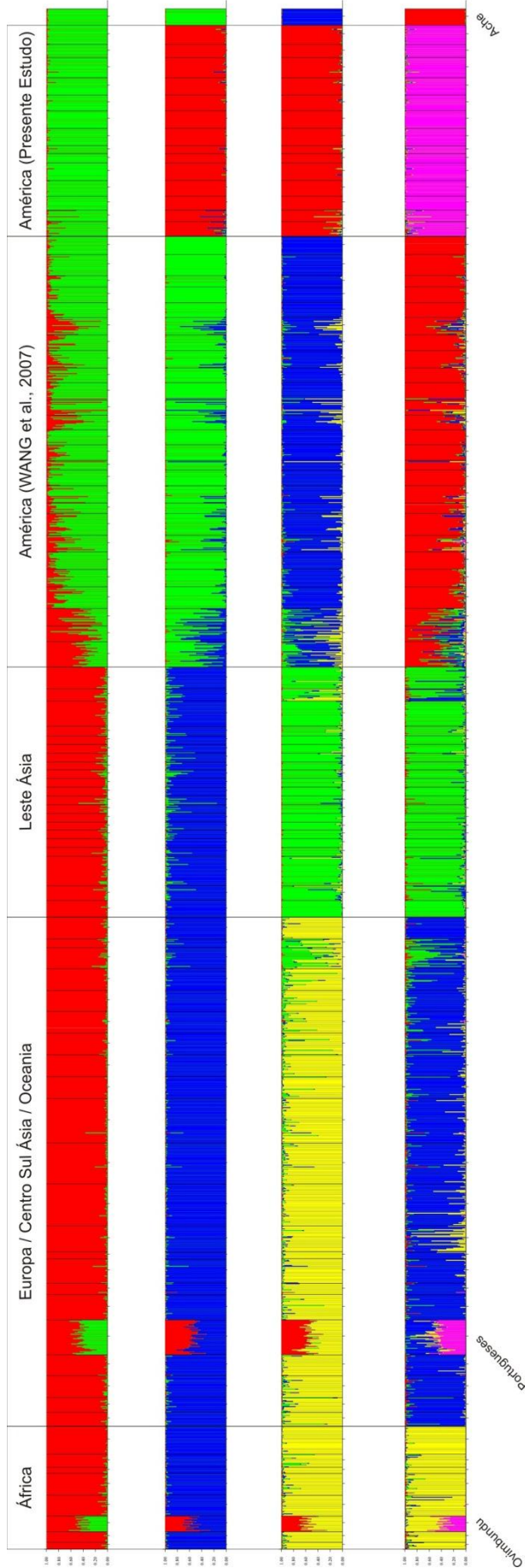


Figura 8. Perfil da estruturação populacional quando introduzido erro proposital de calibração. As populações Ovimbundu e Portugueses e Ameríndias foram calibradas de modo errado para os mesmos dois microsatélites..

As estimativas apontam para baixos níveis de miscigenação não nativa nas nossas populações (tabela 7). É possível observar que as populações Tiriyo e Zo'é são as que apresentam as maiores proporções de miscigenação, principalmente com populações africanas (tabela 7 e figuras 9 e 10).

Há dois indivíduos da população Zo'é, um com 68% componente africano e outro com 39% de componente africano e 53% de componente europeu (figuras 9 e 10). Três indivíduos Tiriyo que apresentam 46%, 26% de componente africano e 26% europeu (figuras 9 e 10). Essas amostras foram excluídas das análises seguintes.

Tabela 7. Estimativa da proporção de ancestralidade africana, europeia e ameríndia nas populações analisadas no presente estudo.

População	Proporção de Ancestralidade (%)		
	Africana	Europeia	Ameríndia
Tiriyo	8,35% ± 0,35%	2,95% ± 0,49%	88,85% ± 0,35%
Zo'é	8,80% ± 0,28%	4,50% ± 0,42%	86,60% ± 0,28%
Arara Laranjal	0,30% ± 0,00%	0,25% ± 0,07%	99,40% ± 0,00%
Arara Iriri	0,20% ± 0,00%	0,25% ± 0,07%	99,55% ± 0,07%
Urubu Kaapor	0,20% ± 0,00%	0,80% ± 0,14%	99,00% ± 0,14%
Awá Guajá	0,15% ± 0,07%	0,15% ± 0,07%	99,70% ± 0,00%
Parakanã	2,05% ± 0,21%	1,00% ± 0,28%	96,75% ± 0,21%
Asurini	0,50% ± 0,00%	0,30% ± 0,00%	99,20% ± 0,00%
Araweté	0,60% ± 0,14%	0,25% ± 0,07%	99,15% ± 0,07%
Kayapó Krokraimoro	1,20% ± 0,14%	1,10% ± 0,28%	97,75% ± 0,35%
Kayapó Xikrin	0,35% ± 0,07%	0,20% ± 0,00%	99,45% ± 0,07%
Guarani Kaiowá	2,05% ± 0,21%	0,65% ± 0,07%	97,25% ± 0,21%
Guarani M'bya	0,50% ± 0,00%	0,30% ± 0,00%	99,20% ± 0,00%
Kaingang	0,40% ± 0,00%	0,35% ± 0,07%	99,25% ± 0,07%

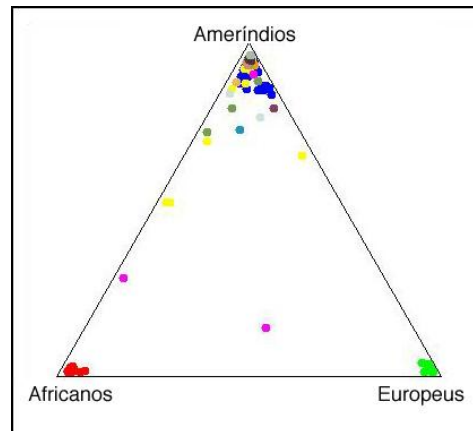


Figura 9. Agrupamento dos indivíduos (círculos) de acordo com a contribuição do grupo étnico parental. Em vermelho populações africanas, em verde populações europeias, em azul populações ameríndias demais cores populações do presente estudo.

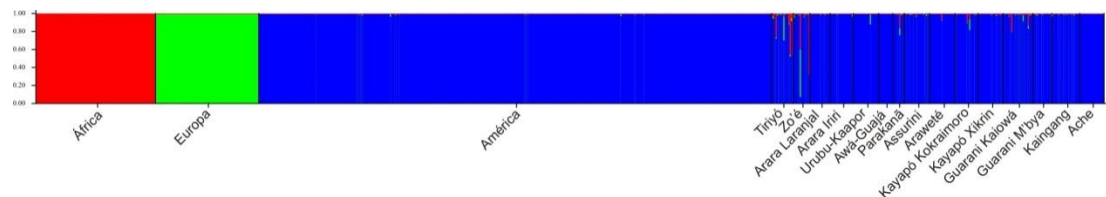


Figura 10. Representação gráfica da estimativa de ancestralidade nas populações ameríndias genotipadas no presente estudo.

Descrição da variabilidade genética

As estatísticas sumárias para estimativas de diversidade nas populações nativas genotipadas no presente estudo, são apresentadas na tabela 8 e a figura 11 mostra o perfil dessas estatísticas ao longo das populações mundiais.

É possível observar que a população com maior índice de diversidade nas nossas genotipagens foi Tiriyo e a menos diversa Arara do Iriri (tabela 8).

Tabela 8. Estatísticas sumárias para os índices de diversidade nas populações nativas americanas genotipadas no presente estudo. He representa a heterozigose média, Θ H representa o índice de diversidade Theta H, K o número médio de alelos e R a amplitude média de variação do tamanho dos alelos dos microssatélites.

Populações	He	s.d.	Θ H	K	s.d.	R	s.d.
Tiriyó	0,70	0,10	1,86	4,90	1,19	4,30	1,42
Zo'é	0,57	0,15	1,55	3,54	1,13	3,30	1,62
Arara Laranjal	0,61	0,11	1,60	3,89	1,16	3,73	1,97
Arara do Iriri	0,43	0,22	1,54	2,80	1,20	2,78	1,59
Urubu-Kaapor	0,61	0,15	1,60	4,13	1,06	3,61	1,33
Awá-Guajá	0,56	0,17	1,53	3,51	0,91	3,16	1,30
Parakanã	0,61	0,17	1,59	3,82	1,30	3,48	1,74
Asurini	0,65	0,12	1,70	4,61	1,05	4,25	1,40
Araweté	0,57	0,16	1,54	3,56	1,06	3,25	1,45
Kayapó Krokraimoro	0,64	0,13	1,68	4,34	1,11	3,86	1,26
Kayapó Xikrin	0,66	0,11	1,72	4,64	0,98	4,11	1,24
Guarani Kaiowá	0,66	0,10	1,74	4,64	0,98	4,27	1,59
Guarani M'bya	0,65	0,12	1,68	4,33	1,03	3,84	1,50
Kaingang	0,64	0,12	1,67	4,54	1,22	3,97	1,46

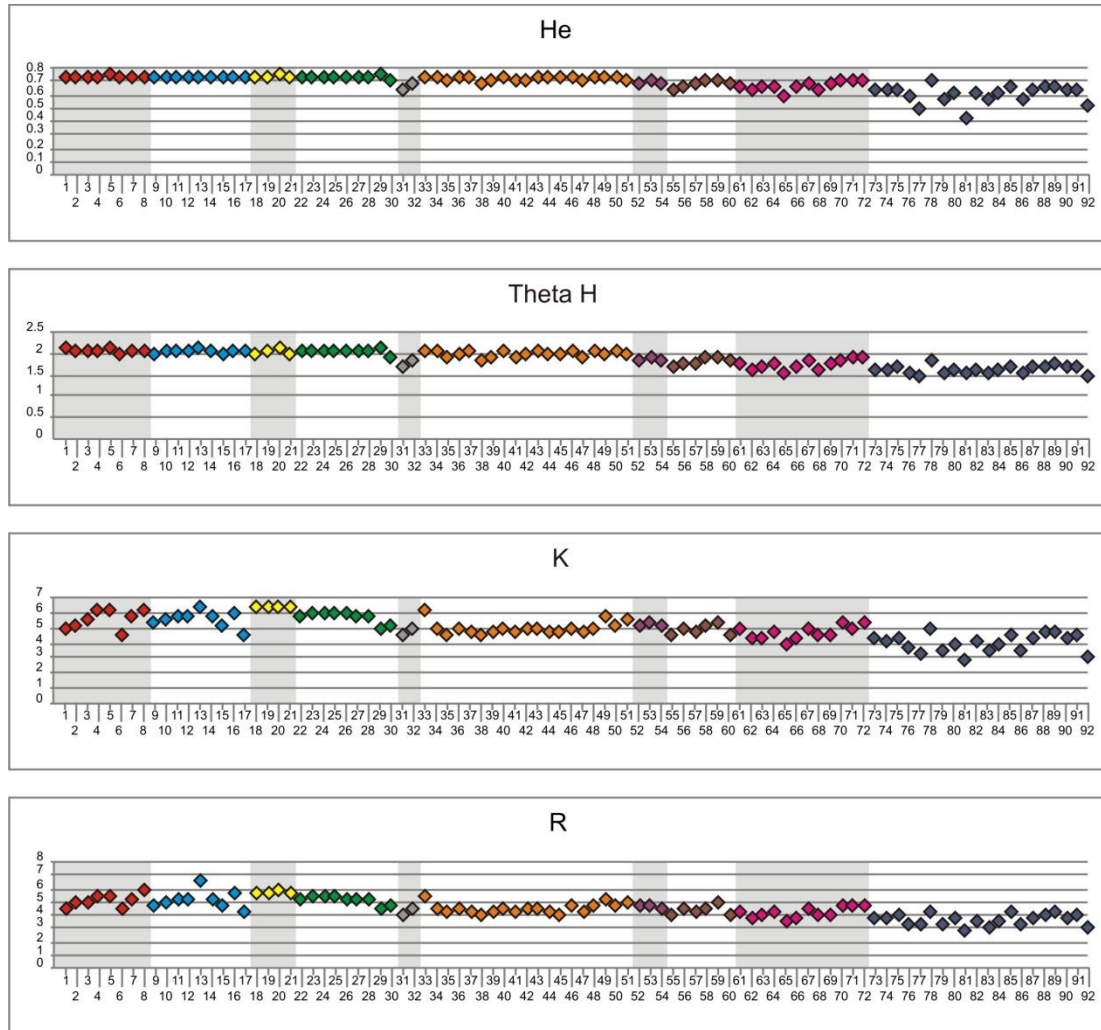


Figura 11. Estatísticas sumárias de diversidade genética ao longo das populações do mundo. Em vermelho as populações da África (1 Bantu da África do Sul, 2 Bantu do Kenia, 3 Ovimbundu, 4 Mandenka, 5 Yoruba, 6 San, 7 Mbuti Pigmeu, 8 Biaka Pigmeu), em azul populações da Europa (9 Orcadian, 10 Adygei, 11 Russa, 12 Basco, 13 Portuguesa, 14 Francesa, 15 Italiana, 16 Sardenha, 17 Toscana), em amarelo as populações do Oriente Médio (18 Mozabite, 19 Beduína, 20 Druza, 21 Palestina), em verde populações do Centro-Sul da Ásia (22 Balochi, 23 Brahui, 24 Makrani, 25 Sindhi, 26 Pathan, 27 Burusho, 28 Hazara, 29 Uygur, 30 Kalash), em cinza populações da Oceania (31 Melanésias e 32 Papua), em laranja populações do Leste Ásia (33 Han, 34 Han-NChina, 35 Dai, 36 Daur, 37 Hezhen, 38 Lahu, 39 Miao, 40 Oroqen, 41 She, 42 Tujia, 43 Tu, 44 Xibo, 45 Yi, 46 Mongólia, 47 Naxi, 48 Canbojano, 49 Japoneses, 50 Tundra Nentsi, 51 Yakut), em lilás as populações da América do Norte (52 Chipewyan, 53 Cree, 54 Ojibwa), em marrom as populações da Mesoamérica (55 Pima, 56 Mixtec, 57 Mixe, 58 Zapotec, 59 Maya, 60 Kaqchikel), em rosa as populações do Oeste da América do Sul (61 Cabecar, 62 Guaymi, 63 Embera, 64 Zenu, 65 Kogi, 66 Arhuaco, 67 Wayuu, 68 Waunana, 69 Inga, 70 Quechua, 71 Aymara, 72 Huilliche), em azul escuro as populações do Leste da América do Sul (73 TicunaA, 74 TicunaT, 75 Piapoco, 76 Karitiana, 77 Surui, 78 Tiriyo, 79 Zo'é, 80 Arara Laranjal, 81 Arara Iriri, 82 Urubu –Kapor, 83 Awá-Guajá, 84 Parakanã, 85 Assurini, 86 Araweté, 87 KayapóK, 88 KayapóX, 89 GuaraniK, 90 GuaraniM, 91 Kaingang, 92 Ache).

Descrição da diferenciação e estruturação populacional

Os índices de diferenciação populacional ao longo das regiões do mundo para os 61 microssatélites estão descritos na tabela 9 e a diferenciação par a par baseada no índice de diferenciação F_{ST} está representada graficamente na figura 12.

É possível observar que, de modo geral, as estimativas de F_{ST} (baseada no modelo de infinitos alelos) e R_{ST} (baseada no modelo *stepwise mutation*) se comportam do mesmo modo ao longo das regiões do mundo sendo que a comparação das matrizes de diferenciação par a par de F_{ST} com R_{ST} mostram que elas são positivamente correlacionadas ($r^2 = 0,734$).

Os maiores níveis de diferenciação encontram-se nas Américas, em especial entre populações do Leste da América do Sul ($F_{ST} = 0,092$ e $0,127$ $R_{ST} = 0,095$ e $0,125$ respectivamente). Além disso, a diferenciação no Leste da América do Sul é praticamente o dobro da descrita no Oeste da América do Sul ($F_{ST} = 0,127$ e $0,059$; $R_{ST} = 0,125$ e $0,598$ respectivamente) (tabela 9).

A figura 13 reproduz com 61 microssatélites o mesmo padrão descrito em outros estudos no qual ocorre a estruturação populacional correspondente às principais regiões do mundo. Contudo, é possível observar que populações no leste da Amazônia (Tiriyó, Zo'é, Arara do Laranjal, Arara do Iriri, Urubu-Kaapor, Awá-Guajá, Parakanã, Asurini, Araweté e Kayapó) e Centro Sul (Guarani, Kaingang e Ache) formam um grupo bem estruturado e distinto das demais populações da América.

Tabela 9. Índices de diferenciação populacional (F_{ST} e R_{ST}) em diferentes regiões do mundo.

Região	N	F_{ST}	I.C. 95%	valor -p	R_{ST}	I.C. 95%	valor -p
Mundo	92	0,068	0,063 0,073	0,000	0,096	0,083 0,111	0,000
África	8	0,026	0,019 0,031	0,000	0,029	0,021 0,037	0,000
Europa	9	0,005	0,002 0,009	1,000	0,010	0,003 0,016	0,911
Oriente Médio	4	0,011	0,008 0,013	0,000	0,012	0,008 0,016	0,003
Centro Sul Asia	9	0,011	0,009 0,013	0,730	0,018	0,012 0,025	0,001
Leste Asia	19	0,012	0,009 0,013	1,000	0,014	0,008 0,020	0,998
Oceania	2	0,082	0,059 0,108	0,000	0,226	0,129 0,314	0,000
América	41	0,092	0,087 0,098	0,000	0,095	0,087 0,104	0,000
América do Norte	3	0,036	0,021 0,047	0,000	0,050	0,017 0,091	0,000
América Central	6	0,041	0,033 0,049	0,000	0,044	0,034 0,053	0,000
Oeste América do Sul	12	0,059	0,054 0,064	0,000	0,059	0,049 0,071	0,000
Leste América do Sul	20	0,127	0,117 0,136	0,000	0,125	0,110 0,143	0,000

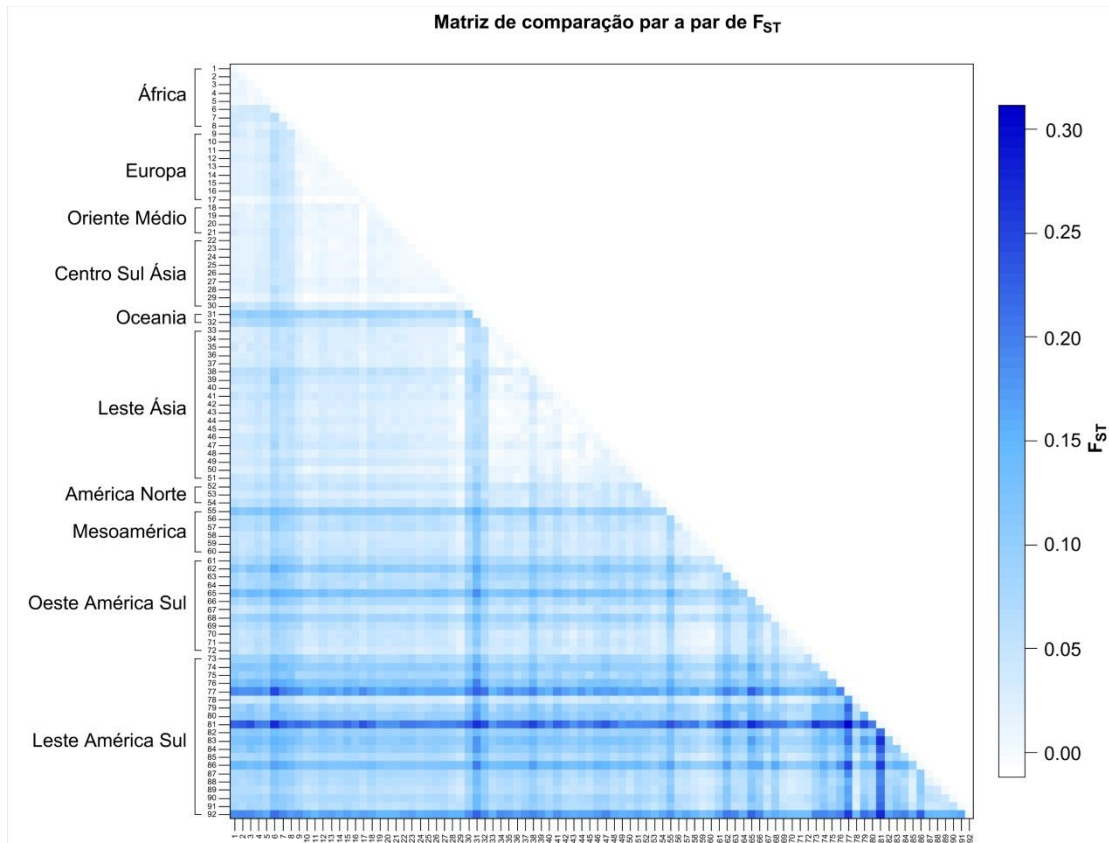


Figura 12. Índice de diferenciação populacional (F_{ST}) par a par das populações. Sendo as populações representadas: 1 Bantu Africa do Sul, 2 Bantu Kenia, 3 Ovibundum, 4 Mandenka, 5 Yoruba, 6 San, 7 Mbuti Pigmeu, 8 Biaka Pigmeu, 9 Orcadian, 10 Adygei, 11 Russa, 12 Basco, 13 Portuguesa, 14 Francesa, 15 Italiana, 16 Sardenha, 17 Toscana, 18 Mozabite, 19 Beduína, 20 Druza, 21 Palestina, 22 Balochi, 23 Brahui, 24 Makrani, 25 Sindhi, 26 Pathan, 27 Burusho, 28 Hazara, 29 Uygur, 30 Kalash, 31 Melanesios, 32 Papua, 33 Han, 34 Han-NChina, 35 Dai, 36 Daur, 37 Hezhen, 38 Lahu, 39 Miao, 40 Oroqen, 41 She, 42 Tujia, 43 Tu, 44 Xibo, 45 Yi, 46 Mongólia, 47 Naxi, 48 Canbojano, 49 Japoneses, 50 Tundra Nentsi, 51 Yakut, 52 Chipewyan, 53 Cree, 54 Ojibwa, 55 Pima, 56 Mixtec, 57 Mixe, 58 Zapotec, 59 Maya, 60 Kaqchikel, 61 Cabecar, 62 Guaymi, 63 Embera, 64 Zenu, 65 Kogi, 66 Arhuaco, 67 Wayuu, 68 Waunana, 69 Inga, 70 Quechua, 71 Aymara, 72 Huilliche, 73 Ticuna Arara, 74 TicunaTarapaca, 75 Piapoco, 76 Karitiana, 77 Surui, 78 Tiriyó, 79 Zo'é, 80 Arara Laranjal, 81 Arara Iri, 82 Urubu –Kaapor, 83 Awá-Guajá, 84 Parakanã, 85 Asurini, 86 Araweté, 87 Kayapó Krokraimoro, 88 Kayapó Xikrin, 89 Guarani Kaiowá, 90 Guarani M'bya, 91 Kaingang, 92 Ache.

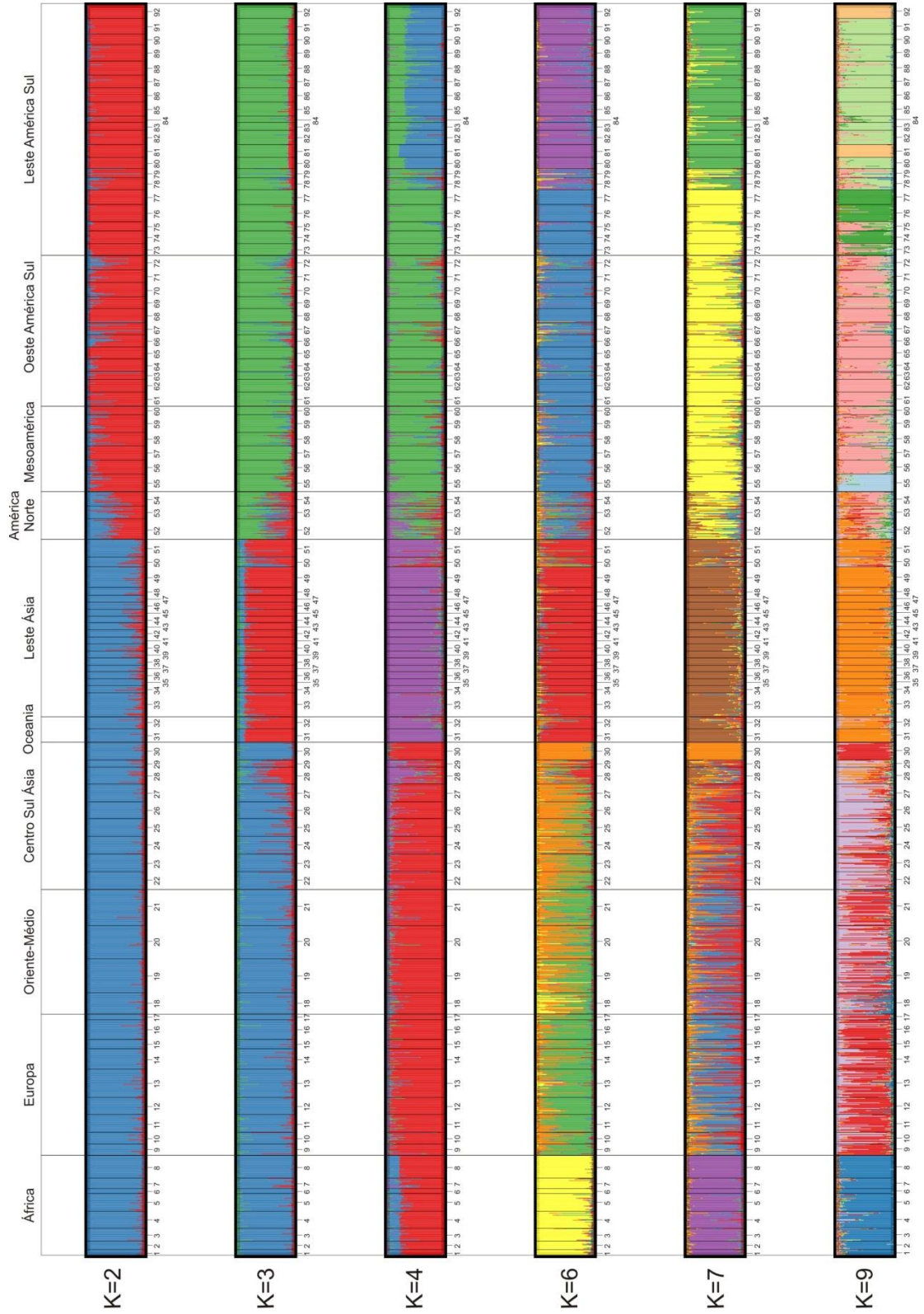


Figura 13. Estruturação populacional nas populações mundiais, mostrando a formação de grupos para as principais regiões: África, Europa e Sul da Ásia, Leste Ásia, América e Leste da América do Sul. Sendo as populações representadas: 1 Bantu Africa do Sul, 2 Bantu Kenia, 3 Ovibundum, 4 Mandenka, 5 Yoruba, 6 San, 7 Mbuti Pigmeu, 8 Biaka Pigmeu, 9 Orcadian, 10 Adygei, 11 Russa, 12 Basco, 13 Portuguesa, 14 Francesa, 15 Italiana, 16 Sardenha, 17 Toscana, 18 Mozabite, 19 Beduína, 20 Druza, 21 Palestina, 22 Balochi, 23 Brahui, 24 Makrani, 25 Sindhi, 26 Pathan, 27 Burusho, 28 Hazara, 29 Uygur, 30 Kalash, 31 Melanesios, 32 Papua, 33 Han, 34 Han-NChina, 35 Dai, 36 Daur, 37 Hezhen, 38 Lahu, 39 Miao, 40 Oroqen, 41 She, 42 Tujia, 43 Tu, 44 Xibo, 45 Yi, 46 Mongólia, 47 Naxi, 48 Canbojano, 49 Japoneses, 50 Tundra Nentsi, 51 Yakut, 52 Chipewyan, 53 Cree, 54 Ojibwa, 55 Pima, 56 Mixtec, 57 Mixe, 58 Zapotec, 59 Maya, 60 Kaqchikel, 61 Cabecar, 62 Guaymi, 63 Embera, 64 Zenu, 65 Kogi, 66 Arhuaco, 67 Wayuu, 68 Waunana, 69 Inga, 70 Quechua, 71 Aymara, 72 Huilliche, 73 Ticuna Arara, 74 Ticuna Tarapaca, 75 Piapoco, 76 Karitiana, 77 Surui, 78 Tiriyo, 79 Zo'é, 80 Arara Laranjal, 81 Arara Iriri, 82 Urubu –Kaapor, 83 Awá-Guajá, 84 Parakanã, 85 Asurini, 86 Araweté, 87 Kayapó Krokraimoro, 88 Kayapó Xikrin, 89 Guarani Kaiowá, 90 Guarani M'bya, 91 Kaingang, 92 Ache.

Analizamos também a estruturação populacional apenas nas populações nativas do continente americano (figura 14). Aqui descrevermos os cenários mais consistentes de estruturação populacional, ou seja, aqueles com algoritmo de Greedy $\geq 0,9$ ou maior probabilidade. Os cenários alternativos e a frequência em que apareceram estão escritos na tabela A.2 do ANEXO 1.

Observamos que primeiro ocorre a formação de um grupo composto pelas populações do leste da Amazônia e Centro Sul da América do Sul, distinto das demais populações ($K=2$). Em seguida, há a separação das populações do sudoeste da Amazônia (Karitiana, Surui e Ticuna) em relação às demais populações ($K=3$). A partir do $K=4$ até o $K=9$ observa-se o aparecimento de estruturação em populações individuais como Ache, Araweté, Arara do Iriri, Surui e Pima.

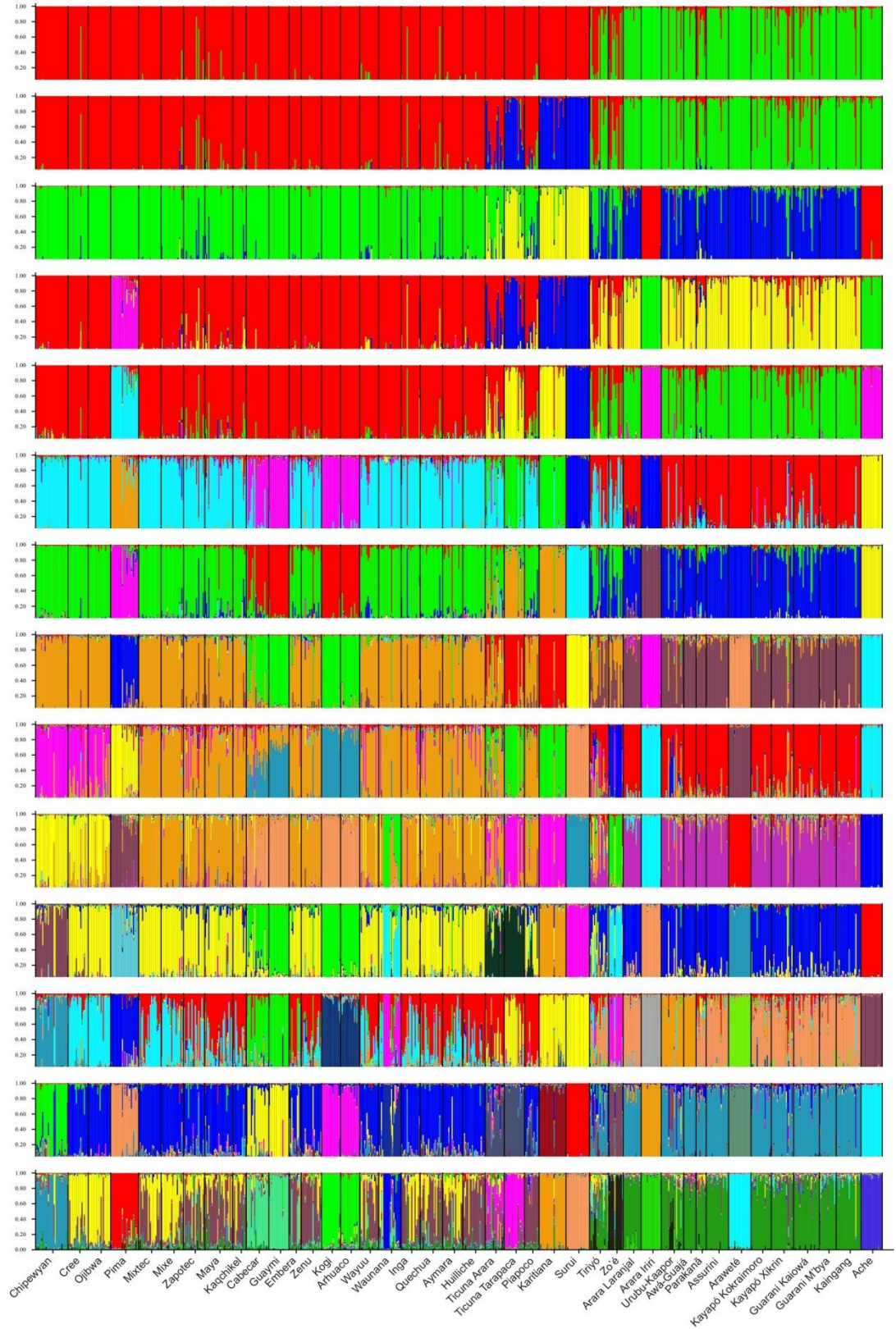


Figura 14. Padrão de estruturação populacional nas Américas.

A partir dessas análises é possível observar que as populações da América do Sul são as que apresentam maior grau de estrutura e diferenciação populacional. Para entendermos melhor como a alta diferenciação ocorre subdividimos a América do Sul nas regiões Oeste e Leste. O Oeste da América do Sul foi dividido em: Noroeste da América do Sul (formado pelas populações Cabecar¹, Guaymi¹, Embera, Zenu, Kogi, Arhuaco, Wayuu e Waunana) e Andes (Inga, Quechua, Aymara e Huilliche). O Leste da América do Sul foi dividido em Oeste da Amazônia (Ticuna Arara, Ticuna Tarapaca, Piapoco, Karitiana, Surui), Leste da Amazônia (Tiriyó, Zo'é, Arara do Laranjal, Arara do Iriri, Urubu-Kaapor, Awá-Guajá, Parakanã, Asurini, Araweté, Kayapó Krokraimoro e Kayapó Xikrin) e Centro Sul (Guarani Kaiowá, Guarani M'bya, Kaingang e Ache). Os resultados das análises de diferenciação nessas sub-regiões são apresentados na tabela 10.

Para a região Oeste da América do Sul é possível observar que as populações Andinas apresentam os menores índices de diferenciação ($F_{ST} = 0,018$ e $R_{ST} = 0,007$). Por outro lado, as populações do Noroeste apresentam maiores índices de diferenciação ($F_{ST} = 0,078$ e $R_{ST} = 0,079$). No Leste da América do Sul observamos que todos os grupos populacionais apresentam altos índices de F_{ST} (Centro-Sul: 0,107; Leste Amazônia: 0,113 e Sudoeste da Amazônia: 0,131) e R_{ST} (Centro Sul: 0,078; Leste Amazônia: 0,121; Sudoeste Amazônia: 0,136). Esse resultado corrobora os resultados prévios (TARAZONA-SANTOS *et al.*, 2001; WANG *et al.*, 2007).

Para testar a nossa hipótese de a elevada diferenciação populacional no Leste da América do Sul em outros estudos possa ser reflexo tanto da baixa amostragem de populações no leste como da inclusão de populações com histórias conhecidas de isolamento, utilizamos a seguinte estratégia:

- 1) Verificamos qual a contribuição de cada população individualmente para a diferenciação populacional da região à qual pertence;
- 2) Verificamos qual a contribuição dos pares de populações com maior índice de diferenciação na diferenciação populacional da região á que pertence.

Os resultados dessas análises são apresentados na tabela 11.

¹ As populações Cabecar e Guaymi estão localizadas em Costa Rica e Panamá respectivamente, mas devido ao relato de outro estudo (WANG *et al.*, 2007) de proximidade genética e linguística com o Noroeste da América do Sul estão sendo analisadas neste grupo no presente estudo.

Tabela 10. Índices de diferenciação populacional (F_{ST} e R_{ST}) ao longo das regiões da América da Sul.

Região	N	F_{ST}	I.C. 95%	valor -p	R_{ST}	I.C. 95%	valor -p
Oeste América do Sul	12	0,059	0,054 0,064	0,000	0,059	0,049 0,071	0,000
Noroeste da América do Sul	9	0,078	0,070 0,086	0,000	0,079	0,062 0,096	0,000
Andes	3	0,018	0,012 0,024	0,008	0,021	0,011 0,031	0,051
Leste América do Sul	20	0,127	0,117 0,135	0,000	0,125	0,110 0,143	0,000
Oeste da Amazônia	5	0,131	0,115 0,147	0,000	0,137	0,108 0,163	0,000
Leste da Amazônia	11	0,113	0,101 0,125	0,000	0,121	0,098 0,145	0,000
Centro-Sul	4	0,107	0,089 0,124	0,000	0,078	0,046 0,114	0,000

Esses resultados mostram que populações da América do Sul apresentam os maiores níveis de diferenciação em comparação com as demais populações do mundo. Contudo, essa diferenciação é maximizada pela presença de algumas populações muito diferenciadas. Quando essas populações *outliers* de diferenciação (Surui, Ticuna Tarapaca, Arara do Iriri, Araweté e Ache) são excluídas das análises, os níveis de diferenciação encontrados são similares entre o Noroeste e Leste da América do Sul, embora a diferenciação ainda seja mais elevada do que nas demais regiões do mundo.

Tabela 11. Comparação da contribuição de cada população na diferenciação populacional.

Região	F _{ST}	I.C. 95%	valor -p	R _{ST}	I.C. 95%	valor -p
Noroeste América do Sul	0,059	0,054 0,064	0,000	0,059	0,049 0,071	0,000
Sem Arhuaco	0,077	0,064 0,097	0,000	0,072	0,054 0,090	0,000
Sem Cabecar	0,083	0,075 0,091	0,000	0,081	0,061 0,101	0,000
Sem Embera	0,078	0,070 0,087	0,000	0,080	0,063 0,097	0,000
Sem Guaymi	0,074	0,066 0,084	0,000	0,079	0,061 0,096	0,000
Sem Kogi	0,070	0,062 0,079	0,000	0,075	0,058 0,091	0,000
Sem Waunana	0,077	0,068 0,086	0,000	0,081	0,063 0,098	0,000
Sem Wayuu	0,084	0,075 0,094	0,000	0,084	0,066 0,102	0,000
Sem Zenu	0,082	0,073 0,092	0,000	0,082	0,064 0,101	0,000
Sem Arhuaco e Kogi	0,066	0,054 0,075	0,000	0,062	0,045 0,080	0,000
Sem Kogi e Guaymi	0,065	0,056 0,075	0,000	0,073	0,055 0,090	0,000
Oeste da Amazônia	0,131	0,116 0,147	0,000	0,137	0,108 0,163	0,000
Sem Karitiana	0,129	0,104 0,147	0,000	0,148	0,111 0,183	0,000
Sem Piapoco	0,139	0,120 0,159	0,000	0,133	0,106 0,162	0,000
Sem Surui	0,096	0,082 0,111	0,000	0,091	0,066 0,116	0,000
Sem Ticuna Arara	0,145	0,126 0,165	0,000	0,153	0,121 0,186	0,000
Sem Ticuna Tarapaca	0,062	0,053 0,072	0,000	0,056	0,036 0,075	0,000
Sem Surui e Ticuna T	0,098	0,081 0,116	0,000	0,100	0,061 0,139	0,000

Continua

Tabela 11 (continuação): Comparação da contribuição de cada população na diferenciação populacional.

Região	F _{ST}	I.C. 95%	valor -p	R _{ST}	I.C. 95%	valor -p
Leste Amazônia	0,113	0,101 0,125	0,000	0,121	0,098 0,146	0,000
Sem Arara Iriri	0,086	0,073 0,096	0,000	0,086	0,071 0,102	0,000
Sem AraraLaranja	0,114	0,102 0,126	0,000	0,123	0,098 0,151	0,000
Sem Arawete	0,107	0,095 0,119	0,000	0,113	0,089 0,140	0,000
Sem Assurini	0,123	0,110 0,137	0,000	0,133	0,108 0,161	0,000
Sem Awa-Guajá	0,112	0,098 0,124	0,000	0,119	0,094 0,146	0,000
Sem Kayapó K	0,122	0,108 0,136	0,000	0,130	0,105 0,157	0,000
Sem Kayapó Xikrin	0,124	0,110 0,138	0,000	0,132	0,106 0,161	0,000
Sem Parakanã	0,111	0,099 0,123	0,000	0,117	0,094 0,142	0,000
Sem Tiryó	0,122	0,110 0,135	0,000	0,130	0,106 0,157	0,000
Sem Urubu-Kaapor	0,116	0,103 0,129	0,000	0,123	0,100 0,148	0,000
Sem Zoé	0,109	0,097 0,127	0,000	0,118	0,095 0,144	0,000
Sem Arara I e Araweté	0,076	0,063 0,085	0,000	0,074	0,060 0,090	0,000
Sem Arara I, Araweté e Zoe	0,066	0,058 0,075	0,000	0,067	0,054 0,081	0,000
Centro-Sul	0,107	0,089 0,125	0,000	0,078	0,046 0,114	0,000
Sem Ache	0,053	0,041 0,067	0,000	0,039	0,023 0,059	0,000
Sem Guarani Kaiowá	0,135	0,112 0,158	0,000	0,089	0,051 0,138	0,000
Sem Guarani M'bya	0,123	0,103 0,144	0,000	0,092	0,053 0,139	0,000
Sem Kaingang	0,118	0,095 0,142	0,000	0,087	0,047 0,133	0,000

Relação das populações do oeste e leste da América do Sul: possíveis rotas migratórias

Através de análise supervisionada no programa *STRUCTURE* constatamos que as populações do Leste da Amazônia e Centro-Sul apresentam maior componente de contribuição Andina em sua composição genética, enquanto que as populações do oeste da Amazônia apresentam maior componente do Noroeste da América do Sul (figura 15).

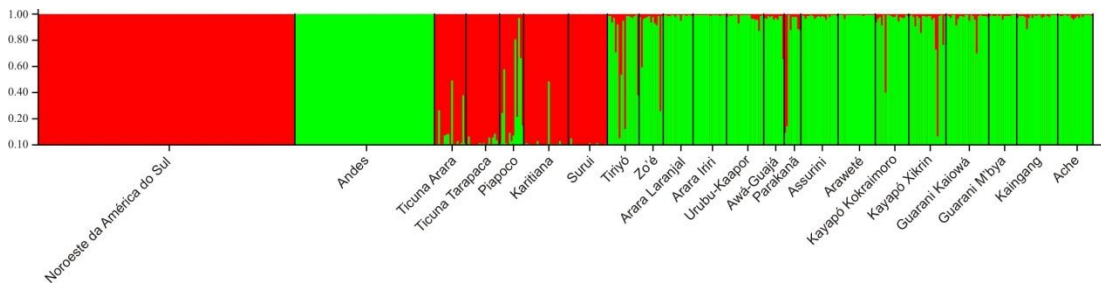


Figura 15. Influência das populações localizadas nas regiões Noroeste da América do Sul e Andinas na constituição das populações do Leste da América do Sul. Em vermelho está grupo formado pelas populações do Noroeste da América do Sul, em verde grupo formado pelas populações andinas. As demais populações representam o Leste da América do Sul, e tiveram seu componente de ancestralidade Andina e do Noroeste da América do Sul estimados.

Cenário Demográfico

A relação entre a estimativa Θ da variância do número de repetições dos alelos de microssatélites e a heterozigose podem ser informativas quanto ao passado demográfico das populações. Quando a taxa $\Theta (V)/ \Theta (He)$, chamado de índice de desequilíbrio β , é <1 indica que a população passou por expansão populacional, quando >1 indica que a população pode ter passado por um gargalo populacional.

Através da estimativa do índice de desequilíbrio β foi possível verificar que as populações da América, em especial do Leste da América do Sul são as que apresentam o maior número e intensidade de gargalos populacionais (figura 16).

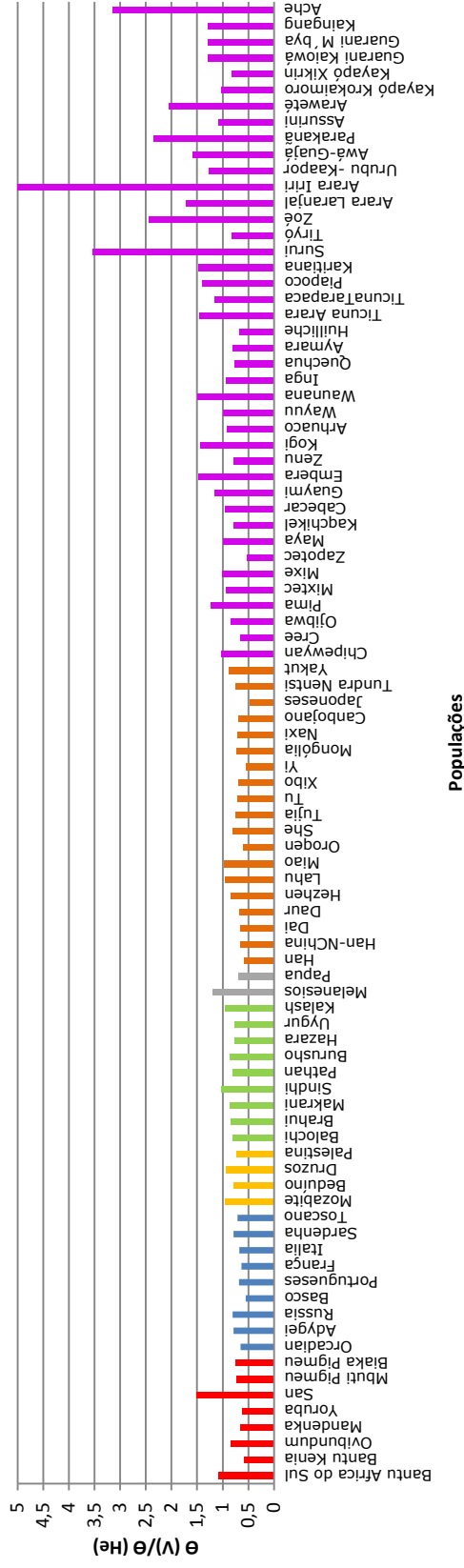


Figura 16. Estimativa de gargalo populacional através do Índice de Desequilíbrio β . Valores <1 indicam que a população passou por expansão populacional, quando >1 indicam que a população pode ter passado por um gargalo populacional. Em vermelho, populações da África, em azul populações da Europa, amarelo populações do Oriente Médio, verde populações do Centro-Sul Ásia, cinza populações da Oceania, em laranja populações do Leste da Ásia, em rosa populações nativas da América.

Discussão

Avaliação do banco de dados

Erros de genotipagem podem ocorrer devido a diferentes fatores que incluem mutações causadas pela PCR, erros de sequenciamento causados pela baixa qualidade e quantidade de DNA, artefatos bioquímicos (qualidade de reagentes, erros intrínsecos da *taq polimerase*, artefatos da eletroforese), e fatores humanos (troca de amostras, contaminação do DNA, erro de identificação e anotação dos resultados). Como consequência, esses erros podem gerar alelos nulos (não amplificação de um alelo por mutação na sequência de DNA onde ocorre a ligação do oligonucleotídeo iniciador), dominância de amplificação de um alelo, ou allelic dropout (que ocorre quando, em indivíduos heterozigotos, há a amplificação preferencial de um dos alelos) e falsos alelos (amplificação de produtos inespecíficos ou deslize da *taq polimerase* criando um alelo inexistente) (POMPANON *et al.*, 2005).

Esses erros podem ter sérias consequências, uma vez que a identificação errada de genótipos leva a estimativas incorretas das frequências alélicas, resultando em estimativas incorretas dos índices de diferenciação (F_{ST} e R_{ST}), falsas taxas de migração, falsa detecção de sinais de seleção e detecção de gargalos populacionais, por exemplo. Considerando que essas estimativas são a base analítica do presente estudo, buscamos validar a qualidade dos nossos dados em dois aspectos. Primeiro, investigamos se há erros de genotipagem acima de uma taxa aceitável. Segundo, tomamos medidas para nos assegurar que os nossos dados foram integrados de modo correto à base de dados pré-existente.

A maioria das amostras de DNA que utilizamos foi coletada na década de 90. Devido ao longo período de estocagem (quase 20 anos), grande parte das amostras apresentou degradação do material genético. Isso fez com que, em média 40%, das genotipagens tivessem que ser refeitas com ajustes no protocolo de PCR (alterações nas concentrações de cloreto de magnésio e na temperatura de hibridização do oligonucleotídeo iniciador na reação de PCR). Devido ao exposto, nós esperávamos que o padrão de qualidade das nossas genotipagens fosse inferior ao de outros estudos. O estabelecimento dos filtros de eliminação de indivíduos e *loci* com

excesso de falha de amplificação teve como objetivo minimizar os erros que eles porventura pudessem agregar ao nosso banco de dados.

As estimativas de falha de amplificação mostram que o nosso banco de dados apresenta maior percentagem (5,86%) do que a apresentada por WANG e colaboradores (3,7%). Como as falhas não estão concentradas em indivíduos ou *loci* específicos (figura 3), acreditamos que elas não refletem problemas com amostras ou marcadores específicos.

Comparando um subconjunto de sete *loci* feitos em duplicata para cerca de 300 amostras, nós estimamos a taxa de erro de genotipagem em 3,59% ($\pm 0,97\%$) (tabela 6). A percentagem de erros encontrada está acima dos padrões exigidos (ou esperados) para laboratórios prestadores de serviços de genotipagem (0,5% a 1% por *locus*), contudo está dentro da margem observada em outro estudos (2 a 15% por *locus*; revisado em POMPANON *et al.*, 2005). Os erros não aparentam ser sistemáticos sendo que o ganho e a perda de heterozigose ocorreram de modo simétrico entre as duas réplicas (Teste exato de Fisher, $p=0,07682$). A aleatoriedade com que os erros estão distribuídos sugere que o efeito que eles introduzirão na análise será de gerar ruído em relação ao sinal demográfico, de modo que inferências realizadas devem ser robustas ao seu efeito e não espúrias.

Um segundo controle para a qualidade de genotipagem foi realizado através da estimativa de *loci* que desviam do esperado sob panmixia (ex. desvios das proporções de Hardy-Weinberg). Nossos dados apresentam menos desvios da proporção de Hardy Weinberg (5,9%) do que o banco de dados de WANG e colaboradores (6,9%), bem como menor percentagem de *loci* que desviam por excesso de homozigose (66,7% no presente estudo e 82,7% em Wang e colaboradores). O valor de 5,9% de desvios significativos, para um valor crítico de 0,05, é muito próximo ao valor esperado caso os desvios estivessem ocorrendo ao acaso. Novamente, como os desvios de Hardy-Weinberg não estão concentrados em populações ou *loci* específicos, supomos que eles terão pouca influência sob os nossos resultados.

Uma das nossas maiores preocupações foi averiguar se os dados foram calibrados e integrados de forma correta aos bancos de dados pré-existentes. Essa verificação é importante principalmente para poder garantir que os resultados de

estruturação populacional na América do Sul são reais e não artefatos de erro de calibração. Por isso, o uso dos controles de genotipagem para calibração e o uso da comparação com outras populações foi fundamental. Os resultados obtidos, em particular o fato de que amostras com localização geográfica Europeia (Portugueses), Africana (Ovimbundu) e Centro-Americana (Guaymi) por nós genotipadas, quando analisadas através do programa STRUCTURE, agrupam-se com as populações do mesmo continente (ou sub-região) da qual foram amostradas (figura 7), demonstram que a calibração dos dados foi efetuada corretamente e que os resultados de estruturação são reais.

Estimativa de miscigenação

Desde as primeiras expedições de exploração do interior do território brasileiro, iniciadas em 1540, os portugueses começaram a entrar em contato com as populações nativas do interior do Brasil. Muitas missões religiosas foram feitas ao longo dos séculos, principalmente pela Igreja Católica (missões Jesuítas e Franciscanas), com intuito de catequizar e introduzir os princípios europeus de civilização aos nativos. Outras missões foram organizadas pelo governo brasileiro com o objetivo de explorar, conquistar e colonizar o território nacional, e escravizar o máximo de mão de obra indígena possível. Desse modo, não é de se estranhar que o contato entre nativos e europeus (ou euro-brasileiros) muitas vezes foi hostil o que levou muitas populações indígenas a abandonarem suas terras e buscar isolamento no interior das matas e do cerrado (CUNHA, 1992). No último século é recorrente o relato de que o governo brasileiro, através de órgãos como o SPI (Serviço de Proteção ao Índio) e posteriormente a FUNAI (Fundação Nacional do Índio), buscaram novamente estreitar as relações com os povos nativos. Na tabela 12 fazemos um breve resumo de quando ocorreu o contato com os colonizadores em cada uma das tribos genotipadas no presente estudo segundo o registro oficial por parte do governo brasileiro (Instituto Socio-Ambiental: <http://pib.socioambiental.org>). Porém, independente dos relatos de contato oficial, muitas populações indígenas convivem e interagem com populações não indígenas em sua vizinhança desde a chegada dos europeus e africanos no continente Americano.

Tabela 12. Resumo do perfil das populações investigada no presente estudo e seu contato com europeus e euro-brasileiros segundo os registros oficiais do governo.

População	Tamanho	Família Linguística	Início do contato	Informações Gerais
Tiriyó	1464	Karib	1906	Intensificação do contato a partir da década de 60 com as missões franciscanas. Hoje contam com o apoio da FUNAI e do governo do Amapá. Interação com populações urbanas na divisa do Amapá com as Guianas.
Zo'é	256	Tupi-Guarani	1989	O contato se deu a partir de missões protestantes e sertanistas da FUNAI. Contudo, há relatos de contatos ocasionais a pelo menos 50 anos com castanheiros e caçadores de pele.
Arara do Laranjal	453	Karib	1850	Tiveram convívio nem sempre pacífico com os seringueiros da região e chegaram a visitar cidades próximas com Altamira. A construção da Transamazônica dividiu suas terras e provocou mudanças socioeconômicas drásticas.
Arara do Iriri	56	Karib	1987	População fundada a partir de um casal de irmãos oriundos de Arara do Laranjal
Urubu-Kaapor	991	Tupi-Guarani	1600	Os relatos de contato com luso-brasileiros é bem antigo. Contudo, os conflitos com grupos indígenas e não indígenas locais marcam sua história. Falam português, e tem dependência de vários artefatos da cultura ocidental.
Awá-Guajá	60	Tupi-Guarani	Década de 70	Há pouca informação a respeito do contato. Há relatos de depopulação de 91 para 25 indivíduos entre 1976 a 1980 em decorrência do contato desordenado.
Parakanã	1266	Tupi-Guarani	1910	Entre 1910 e 1950 mais de 20 mulheres euro-descendentes foram raptadas. Dessas, dezessete tiveram filhos com Parakanãs
Asurini	154	Tupi-Guarani	1960	Contatos hostis iniciaram na década de 60. A partir da década de 70 os contatos se intensificaram e 50% da população morreu em decorrência de doenças. A recuperação populacional se deu a partir de casamentos inter étnicos com povos indígenas próximos.
Araweté	358	Tupi-Guarani	1971	Pouco se sabe a respeito do contato. Em 1976 a população sofre um decréscimo de 36% devido ao contato e a uma conjuntivite infecciosa.
Kayapó	8638	Gê	Início do séc. XX	A maioria dos grupos Kayapós está em contato com a nossa sociedade desde 1950-1960, quando missões para pacificá-los foram organizadas pelo governo brasileiro.
Guarani Kaiowá	4960	Tupi-Guarani	Séc. XIX	O Grupo Guarani, como um todo, sempre teve contato e conflitos com os euro-brasileiros. A partir de 1960 houve a intensificação da colonização e desapropriação das terras indígenas por parte dos colonizadores.
Guarani M'bya	500	Tupi-Guarani		
Kaingang	1000	Gê	Sec XVIII	Há relatos desde 1770 de expedições para tentar conquistar e ocupar as terras pertencentes aos Kaingang.

Fonte: <http://pib.socioambiental.org>

Devido à dificuldade em escravizar os nativos a mão de obra escrava africana foi introduzida no Brasil a partir da metade do século XVI. Até 1850, cerca de quatro milhões de africanos foram trazidos ao Brasil (IBGE, 2000). Muitos desses escravos africanos fugiram do trabalho forçado e adentraram e se escondiam na mata. Nas primeiras fugas se refugiavam em aldeias indígenas e mais tarde quando as fugas de escravos africanos eram mais frequentes se organizaram e formaram comunidades – quilombos. De modo geral, as comunidades quilombolas eram bastante fechadas e o casamento inter grupal dificilmente permitido, contudo ao longo dos anos isso mudou. Hoje é comum o relato da interação entre as comunidades indígenas e quilombolas (revisão em FOSTER, 2004).

Considerando que a miscigenação com grupos não ameríndios poderiam ter contribuído para a formação genética das populações do presente estudo, entendemos que é importante estimar o grau de miscigenação.

Através das nossas análises (tabela 7 e figuras 9 e 10) foi possível constatar que a miscigenação é baixa em nossas amostras. Como analisamos marcadores autossômicos, não conseguimos distinguir se o componente não-ameríndio ocorre de forma simétrica ou assimétrica (isso é, se a contribuição de homens e mulheres é igual ou distinta). Casos de assimetria nos componentes de miscigenação são comuns na literatura. Por exemplo, a análise de marcadores uniparentais (microsatélites no cromossomo Y e DNAMt) sugerem que há maior componente não ameríndio no cromossomo Y que no DNAMt nas as populações Embera, Ingano, Ticuna, Wayuu e Zenu (MESA *et al.*, 2000). O mesmo padrão foi descrito por RODRIGUES-DELFIN *et al.* (2001) em populações Andinas, Amazônicas e do Sul da América do Sul. A história relata que sempre vieram para o Brasil muito mais homens europeus e africanos do que mulheres (IBGE, 2000). Esse cenário cria um viés que acarreta num maior fluxo gênico entre homens de etnias não indígenas com mulheres indígenas.

Nas nossas genotipagens, as duas populações localizadas ao norte do rio Amazonas (Tiriyó e Zo'é) são as que apresentam o maior grau de miscigenação ($8,35\% \pm 0,35\%$ e $8,80\% \pm 0,28\%$ de componente africano e $2,95\% \pm 0,49\%$, $4,50\% \pm 0,42\%$ de componente europeu respectivamente). A população Tiriyó vive no estado do Amapá na divisa com as Guianas. A população mantém estreita relação de

intercambio econômico, inclusive com o deslocamento de pessoas entre os dois lados da fronteira. Além de sua própria língua Karib, também falam português, francês e compreende alemão, holandês e inglês, que são línguas faladas na região por grupos não indígenas. Essa intensa relação com outros povos pode ser responsável pelo maior grau de miscigenação observado em nosso estudo. SALZANO e CALLEGARI-JACQUES, (1988) estudaram a população Tiriyo e encontraram para o fator Rh negativo 3,3% de alelos africanos, introduzidos por miscigenação. Para outros conjuntos de marcadores de grupos sanguíneos os autores encontraram entre 3,4 e 3,1% de miscigenação africana e europeia respectivamente.

A população Zo'é, apesar de ser umas das últimas populações ameríndias a ser oficialmente contatada pela FUNAI (1989), apresentou um alto grau de miscigenação, em especial com africanos. É importante chamar a atenção que essa estimativa é influenciada basicamente por dois indivíduos. Um indivíduo apresentou 39% de ancestralidade africana, 53% de europeia e 8% indígena. Outro indivíduo apresentou 68% de ancestralidade africana. Sugerimos que esses indivíduos são imigrantes na população Zo'é, ou que houve troca de amostras. O presente estudo é o primeiro a estimar o grau de miscigenação nessa população.

SALZANO e CALLEGARI-JACQUES, 1988 também analisaram alelos do fator Rh negativo e 13 genes relacionados a grupos sanguíneos em outras populações nativas e constataram que a população Parakanã apresenta 1,2% de miscigenação com africanos e 0% com os europeus. No presente estudo encontramos $2,05\% \pm 0,21\%$ de componente africano e $1,00\% \pm 0,28\%$ de componente europeu. Esse resultado é condizente com os relatos de contato com quilombos e rapto de 20 mulheres euro-descendentes no início do século passado. Para a população Kayapó esses autores identificaram 1,2% de miscigenação com os africanos e 0,2% com os europeus, nós estimamos $1,2\% \pm 0,14\%$ de componente africano e $1,1\% \pm 0,28\%$ de componente europeu.

Gostaria de chamar atenção para o fato que os índices de miscigenação descritos no presente estudo para as populações Guarani Kaiowá ($2,05\% \pm 0,21\%$ - africana e $0,65\% \pm 0,07\%$ - europeia), Guarani M'bya ($0,50\% \pm 0,00\%$ - africana e $0,30\% \pm 0,00\%$ - europeia) e Kaingang ($0,40\% \pm 0,00\%$ - africana e $0,35\% \pm 0,07\%$ - europeia) podem ser subestimados nessas populações, uma vez que selecionamos

nossas amostras priorizando indivíduos sem sinais de miscigenação para os genótipos de *HLA*. Com base nos genótipos de *HLA*, PETZL-ERLER e SOTOMAIOR (1993) encontraram 2,5% de miscigenação em Guarani e 5,3% em Kaingang. TSUNETO *et al.*, 2003 analisando apenas genes *HLA* de classe II descreveram 7,2% de miscigenação em Kaingang, 3,7% em Guarani M'bya e não encontraram evidências de miscigenação em Guarani Kaiowá. Um estudo com *HLA-B* mostra que há 5% de alelos não indígenas nas populações Guarani (FRANCISCO, 2009).

Nosso objetivo principal com essa análise não era contar esse capítulo da história indígena, mas sim eliminar possíveis indivíduos miscigenados que possam introduzir viés para a análise da história pré-colombiana, que estamos tentando interpretar. Por isso, todos os indivíduos que apresentaram mais de 50% de componente não ameríndio foram eliminados das nossas análises. Entretanto, para análises voltadas para a história demográfica recente das populações Amazônicas, contemplando a miscigenação que ocorreu após a chegada dos europeus e africanos ao continente, a inclusão dessas amostras é apropriada.

Diversidade populacional e gargalos populacionais

É reconhecido que a região de origem de uma espécie é a que apresenta maior diversidade genética e as regiões por último povoadas são as com menor diversidade. As populações humanas africanas são as que apresentam o maior grau de diversidade no mundo. Há pouco tempo, a maioria dos pesquisadores acreditava que o local de origem do homem moderno era o leste da África. Contudo, estudos recentes baseados na genotipagem de 848 microssatélites e 476 indels em 2.432 africanos (TISHKOFF *et al.*, 2009) e 580.000 SNPs em 27 populações africanas (HENN *et al.*, 2011) revelam que a região sul do continente africano é a região que apresenta maior diversidade genética e menor desequilíbrio de ligação, sugerindo que esse poderia ser o local de origem do homem moderno.

Vários estudos relatam o padrão de decréscimo na diversidade genética à medida que as populações se afastam da África, sendo a América do Sul a região com menores índices de diversidade populacional para estudos envolvendo os mais diversos sistemas genéticos (INGMAN *et al.*, 2000; UNDERHILL *et al.*, 2000; PRUGNOLLE *et al.*, 2005; RAMACHANDRAN *et al.*, 2005; WANG *et al.*, 2007; TISHKOFF

et al., 2009). Esse mesmo padrão foi reproduzido no presente estudo a partir de 61 marcadores moleculares do tipo microssatélites (figura 11). A taxa de heterozigose média por nós estimada, para cada população, é praticamente a mesma que aquela estimada usando o conjunto total de marcadores (377 microssatélites no estudo de PRUGNOLLE *et al.*, 2005 e 678 microssatélites no estudo de WANG *et al.*, 2007). Isso mostra que mesmo com um número inferior de marcadores analisados, nós conseguimos captar o mesmo padrão de declínio de diversidade com distância percorrida desde o continente africano, escrito em por outros estudos.

Em 2009 AMOS e HOFFMAN realizaram análise com intuito de verificar a intensidade e quantificar os gargalos populacionais nas populações mundiais. Com base em 783 microssatélites de 53 populações do HGDP-CEPH os autores encontraram sinais de dois gargalos populacionais mais pronunciados. O primeiro referente à saída do homem moderno do continente africano e o segundo correspondente a entrada do homem moderno no continente americano.

Os microssatélites analisados no presente estudo são um subconjunto dos marcadores utilizados por AMOS e HOFFMAN (2009) e também percebemos que os sinais de gargalos populacionais são mais pronunciados no continente americano, em especial no leste da América do Sul (figura 16).

A baixa diversidade genética, observada em especial no leste da América do Sul, resulta de um efeito cumulativo da história do povoamento das Américas e da subsequente história de cada população. Como exposto acima, espera-se que a América do Sul apresente os menores índices de diversidade pelo simples fato de ser o último local que foi povoado pelo homem moderno, implicando que as populações que ali chegaram passaram por um maior número de "gargalos genéticos seriados" (RAMACHANDRAN *et al.* 2005).

A população que apresentou os menores índices de diversidade e maior sinal de gargalo populacional foi Arara do Iriri (tabela 8 e figura 16). RIBEIRO-SANTOS *et al.*, 1999 com base em marcadores moleculares (sete microssatélites autossômicos e dois no cromossomo Y), estimaram a heterozigose para essa população em 0,39, esse valor é similar ao descrito no presente estudo ($H_e = 0,43 \pm 0,22$). A população Arara do Iriri é um caso extremo de efeito fundador: ela foi fundada por um casal de irmãos expulsos da tribo Arara do Laranjal. Todos os indivíduos da população Arara

do Iriri descendem desse casal de irmãos e portanto a população apresenta no máximo 4 alelos por *locus* (eventos mutacionais não foram observados para esse conjunto de microssatélites).

As populações Awa-Guajá, Araweté e Zo'é também apresentaram baixos índices de diversidade (tabela 8). Awa-Guajá e Araweté são populações que apresentam em sua história recente gargalos populacionais resultantes do contato com grupos não indígenas e com doenças infecciosas (ver tabela 12 da discussão e figura 16). Em Zo'é, após eliminarmos os dois indivíduos com alto componente não ameríndio a população passou a exibir baixos níveis de diversidade ($H_e = 0,57 \pm 0,15$ e $K = 3,54 \pm 1,13$). A heterozigose e o baixo número de alelos em Zo'é também foram observados em outro estudo baseado em 11 microssatélites autossômicos ($H_e = 0,59 \pm 0,15$ e $K = 3,7 \pm 1,0$ - SANTOS *et al.*, 2009).

SANTOS *et al.* (2009) também estimaram a taxa de heterozigose e o número de alelos para outras populações do presente estudo (Tiriyó, Urubú-Kaapor, Awá-Guajá e Parakanã) e sempre encontraram resultados semelhantes, sendo a população Tiriyó a mais diversa, Urubu-Kaapor e Parakanã intermediárias, e Awa-Guajá a com menor diversidade.

A história de cada população e as particularidades da região leste da América do Sul será explorada em mais detalhes nas próximas seções.

Diferenciação e estrutura populacional

As análises de diferenciação populacional permitiram detectar o modo como a diversidade genética está distribuída entre as populações. Identificamos que a região que apresenta maior diferenciação inter-populacional é a América do Sul ($F_{ST} = 0,092$ e $R_{ST} = 0,095$ – ver detalhes na tabela 9 e figura 12) em especial a região Leste da América do Sul ($F_{ST} = 0,127$ e $R_{ST} = 0,125$) com praticamente o dobro da diferenciação encontrada no Oeste da América do Sul ($F_{ST} = 0,059$ e $R_{ST} = 0,060$). Esses resultados são semelhantes aos descritos por WANG *et al.*, 2007 que encontraram um valor de $F_{ST} = 0,081$ para a América, $F_{ST} = 0,147$ para o Leste da América do Sul e $F_{ST} = 0,057$ para o Oeste da América do Sul.

Através da análise de diferenciação par-a-par (figura 12) é possível observar que a América do Sul apresenta maior número de populações com alto grau de

diferenciação. Além disso, as análises realizadas no programa STRUCTURE mostram o Leste da América do Sul é a região que apresenta maior estruturação populacional (figura 14).

O grande número de populações, em especial no Leste da América do Sul, que formam grupos individuais cria uma multimodalidade nas análises do STRUCTURE, fazendo com que para valores de K entre 4 a 14 várias soluções de agrupamento sejam apresentadas (tabela A.2 – ANEXO 1). A partir de K=15 as populações ajustam-se nos grupos, e o particionamento dos grupos passa a ser mais regular, ou seja, ocorre a formação de um menor número de cenários alternativos. A mesma variação de modalidade também foi observada por WANG e colaboradores, contudo como o número de populações por ele usadas foi menor, a regularidade foi atingida com menor número de K (K=9).

Até pouco tempo atrás esses resultados corroborariam o estereótipo popular, cunhado na década de 50, da região amazônica como um ambiente homogêneo, de cultura uniforme, ocupada por pequenos povoamentos dispersos em tribos primitivas, isoladas ao longo da floresta tropical (revisão em HECKENBERGER e NEVES, 2009).

A Amazônia, até recentemente, era reconhecida como uma região sem uma civilização pré-colombiana clássica, ou seja, sem desenvolvimento de escrita, arquitetura, grãos para agricultura e domesticação de animais. Acreditava-se que a floresta não possuía capacidade de suporte ambiental para abrigar grandes densidades populacionais humanas e, por isso, as civilizações eram formadas por pequenos grupos de caçadores-coletores. Contudo, estudos antropológicos contemporâneos têm revelado uma nova e dinâmica história indígena, incluindo desde vias alternativas de domesticação de animais e agricultura, até estilos de vida que denotam uma elevada complexidade social (HECKENBERGER e NEVES, 2009).

Uma extensa rede de comércio foi identificada, que ocorria principalmente através dos principais tributários do rio Amazonas. O comércio era baseado na troca de alimentos agrícolas, animais e artefatos (cerâmica, pedras, etc.). Assim, os grupos que viviam próximos às margens dos rios possuíam maior controle sobre o comércio, o que estimulava alianças entre os grupos, como casamentos entre os

herdeiros dos líderes tribais, com o intuito de expandir o domínio das redes de comércio (HECKENBERGER, 2008).

Essa comunicação e integração nos sistemas regionais criou um pluralismo em lugar de uma homogeneização cultural. É possível observar que peças e artefatos de outras regiões eram incorporados em outras sociedades sem que houvesse a perda da identidade cultural local (HECKENBERGER e NEVES, 2009). A diversidade linguística é uma característica notável da Amazônia, e nenhuma família linguística dominou a região, como ocorreu, por exemplo, na Europa (Indo-Europeu), África Sub-Saariana (Niger-Congo) ou no Pacífico (Austronésia). Da mesma forma, nenhuma formação sócio-política foi forte o suficiente para expandir sua influência em larga escala, com ocorreu nos Andes, com o Império Inca. Na Amazônia nunca surgiu um império, mas isso não significa que não houve grandes cacicados e reinos, que podiam se unir nos tempos de guerra e se desestruturar nos tempos de paz (HECKENBERGER, 2008). Em termos gerais, a política amazônica era menos centralizada no que diz respeito aos aspectos tecno-econômicos, sócio-políticos e de recursos simbólicos (HECKENBERGER e NEVES, 2009).

Essa descentralização, ao contrário do que possa parecer, não é sinônimo de sociedade simples. Estudos recentes revelam que há pelo menos 2 mil anos, a Amazônia teve grandes povoamentos, estáveis e complexos. Em 2003, HECKENBERGER e colaboradores anunciaram a descoberta, na região do rio Xingu, de vestígios de 19 grandes povoados interligados por uma ampla rede de estradas. Havia ainda pontes, caminhos elevados, canais, tanques artificiais nos rios para a criação de peixes e outras estruturas elaboradas. Outra região densamente habitada, de acordo com a antropóloga Anna Roosevelt, foi a Ilha de Marajó (800 a 14.000 a.C). Os vestígios arqueológicos revelam uma sociedade sofisticada, que durou mais de mil anos e possivelmente possuía mais de 100 mil habitantes (revisado em MANN, 2007).

Frente ao exposto nós nos perguntamos por que é tão difícil visualizar esse cenário a partir dos dados genéticos? Por que nós observamos alta diferenciação entre as populações apesar desses relatos de interação entre as populações? Para compreender melhor como a diferenciação está ocorrendo nós subdividimos a América do Sul em regiões (Noroeste América do Sul, Andes, Oeste Amazônia, Leste

Amazônia e Centro-Sul). Essa análise revelou que a região Andina é a que apresenta menor diferenciação populacional ($F_{ST} = 0,0184$ e $R_{ST} = 0,02122$). Um ponto importante a ser esclarecido é que as populações do oeste do continente americano analisadas por TARAZONA-SANTOS *et al.* 2001, entram na categoria de Andinas e o $F_{ST} = 0,024$ descrito pelos autores é semelhante ao encontrado no presente estudo para essa região. Esse resultado possivelmente ilustra o efeito que o Império Inca teve na homogeneização das populações, como discutido por TARAZONA-SANTOS *et al.*, 2001. O Noroeste da América do Sul apresentou os maiores índices de diferenciação na região Oeste da América do Sul ($F_{ST}=0,071$ e $R_{ST}= 0,062$). Essa região nunca fez parte de nenhum grande império e, portanto pode apresentar uma dinâmica populacional diferente da região Andina.

Todas as subdivisões do Leste da América do Sul apresentaram altos níveis de diferenciação (Sudoeste Amazônia $F_{ST}=0,131$, $R_{ST}=0,137$; Leste Amazônia $F_{ST}=0,113$, $R_{ST}=0,121$; Centro-Sul $F_{ST}=0,107$, $R_{ST}=0,078$). Quando olhamos as análises de estruturação populacional (figura 14), a maneira com que o particionamento ocorre no leste da América do Sul, chama a atenção. As populações que primeiro se diferenciam das demais são as que apresentam relatos de apresentarem amostras aparentadas, ou gargalo populacionais recentes (Arara do Iriri, Ache, Surui, Araweté). Por isso, fizemos uma segunda análise eliminando uma a uma as populações para ver qual a contribuição individual de cada uma na diferenciação inter-populacional de cada região.

Observamos que cada uma das populações do Noroeste da América do Sul individualmente pouco influenciam na diferenciação inter-populacional da região (ver tabela 11). Contudo, o mesmo não acontece nas populações do Leste da América do Sul, sendo que, as populações que mais contribuem para o aumento da diferenciação inter-populacional possuem histórias evolutivas peculiares (Por exemplo, Arara do Iriri foi fundada por um casal de irmãos, resultando num efeito fundador extremo; Araweté sofreu um drástico declínio populacional há 30 anos em decorrência de doença; Ache e Zo'é, são populações que vivem isoladas e raramente interagem com outras populações indígenas; as amostras de Suruí foram coletadas como parte de estudos familiares com genealogias, portanto há grande número de indivíduos consanguíneos). Assim sendo, essas populações representam apenas uma

face da história das populações nativas do leste da América do Sul e o fato de que ao excluí-las os níveis de diferenciação no Noroeste e Leste da América do Sul são semelhantes desafia a ideia de que todas as populações no leste da América do Sul são altamente diferenciadas uma das outras.

O que percebemos com essas análises é que ao mantermos o banco de dados completo (sem excluir as populações *outliers*), observamos grandes índices de diferenciação inter-populacional nas populações do Leste em relação ao Oeste da América do Sul, corroborando os trabalhos anteriores de TARAZONA-SANTOS *et al.*, 2001 e WANG *et al.*, 2007. Por outro lado, ao excluirmos das análises populações individuais, os resultados obtidos corroboram a ideia de PUCCIARELLI *et al.*, 2006, que sugere que o Oeste e Leste da América do Sul são dois universos distintos, contudo com níveis de diferenciação inter-populacional similares .

A diferenciação entre Sudoeste e Leste da Amazônia

O resultado que mais nos chamou a atenção nas análises de estruturação populacional no continente americano foi que para $K=2$ o Leste da Amazônia e o Centro-Sul formam um grupo distinto das demais regiões da América. Para $K=3$ o Leste da Amazônia e Centro-Sul se mantêm como um grupo, e o Sudoeste da Amazônia forma outro grupo distinto das demais regiões do mundo (figuras 13 e 14). A princípio consideramos que esse resultado poderia ser erro de calibração ou genotipagem. Contudo, devido aos controles de genotipagens e os testes de calibração (figura 7 e 8) nos certificamos que não há erros.

O mesmo padrão de separação do sudoeste e leste da Amazônia já foi observado em estudos anteriores. SALZANO e CALLEGARI-JACQUES, 1988 mostram, através de análises antropométricas, que as populações do Noroeste da América do Sul e Sudoeste da Amazônia apresentam menor estatura que as populações do Leste da Amazônia e Centro Sul do continente. Os mesmos autores realizaram análise de componente principal com base em sete marcadores genéticos clássicos. O primeiro componente é responsável por 23% da variância observada e o padrão apresentado também aponta para uma diferença entre Sudoeste e Leste da região Amazônica. MENDES JUNIOR (2005), ao analisar os haplogrupos de DNAm, também encontrou diferenças na distribuição frequências dos haplogrupos ao longo da América do Sul.

Análise NM-MDS (análise de dispersão em escala multidimensional) com base em 96 populações da América do Sul, mostrou que as três dimensões são consistentes com a diferença entre sudoeste e leste da Amazônia.

O sudoeste da Amazônia apresenta os maiores níveis de diferenciação para a América do Sul ($F_{ST} = 0,131$). Essa região é representada por populações de diversos grupos linguísticos: Ticuna (pertence ao grupo linguístico Ticuna), Piapoco (Arawak), Karitiana (Tupi Guarani), e Suruí (Tupi Mondé). Além disso, a região é reconhecida por apresentar a maior riqueza cultural da Amazônia, além de apresentar as evidências mais antigas de ocupação contínua humana (iniciada a pelo menos 9.000 anos) e das terras pretas (solo antropomórfico repleto de matéria orgânica – 4.700 anos) (NEVES, 2006).

Alguns autores também sugerem que o sudoeste da região amazônica, mais especificamente a região de Rondônia, é o centro de dispersão do maior grupo linguístico indígena brasileiro, o Tupi. Os membros do grupo Tupi são os que mais expandiram no território brasileiro (Amazônia, litoral Atlântico até o Sul do Brasil), sendo um povo bélico e com política atrelada a assimilação dos povos conquistados. Contudo, diferentemente dos Incas, os Tupi não impõe sua cultura aos povos conquistados, preservando dessa forma a diversidade cultural (NOELLI, 2008). No presente estudo não testamos formalmente nenhuma hipótese envolvendo a diferenciação entre grupos linguísticos, por julgar que nossa amostragem apresenta uma sub-representação de grupos não-tupi. De qualquer modo, nossos resultados indicam que o contraste entre sudoeste e leste da Amazônia revela uma menor diferenciação entre populações de grupos linguisticamente distintos quando esses estão geograficamente mais próximos. Contudo essa questão merece ser explorada em mais detalhes futuramente.

Relação das populações do oeste e leste da América do Sul: possíveis rotas migratórias

Para testar o modelo proposto por SCHMITZ em 1983, realizamos uma análise supervisionada no programa STRUCTURE onde fixamos as regiões Noroeste da América do Sul e Andes para verificar a influência e proximidade genética dessas regiões nas populações do Leste da América do Sul.

Observamos que as populações do Sudoeste da Amazônia apresentam maior componente genético do Noroeste da América do Sul, a as populações do Leste da Amazônia apresentam maior componente Andino (figura 15). Em princípio, essa análise corrobora o modelo descrito por SCHMITZ em 1983 (revisão em SALZANO e CALLEGARI-JACQUES, 1988). Neste modelo o autor propõe que uma rota possível para o povoamento do Leste da América do Sul seria através de rios localizados na divisa entre a bacia Amazônica e o Rio da Prata. Seguindo essa rota os migrantes poderiam ter atingido a bacia do São Francisco, de onde dispersões secundárias teriam ocorrido tanto em direção à Amazônia como ao Sul do Brasil (figura 17). O Noroeste da Amazônia, por sua vez, poderia ter sido povoado através do noroeste da Venezuela e Guiana (as quais por sua vez têm influência das populações Colombianas).

Contudo chamamos a atenção para o fato que a análise supervisionada que realizamos é extremamente sensível à estruturação populacional e desta forma pode estar apenas refletindo o alto grau de estruturação do sudoeste e leste da Amazônia induzindo uma maior proximidade com regiões distintas do oeste da América do Sul. Por isso, novas análises utilizando diferentes abordagens (ex. ABC – Computação Bayseana Aproximada, teste de modelos migratórios com base nos programas IM e MIGRATE) deverão ser realizadas futuramente para averiguar a consistência desse resultado.

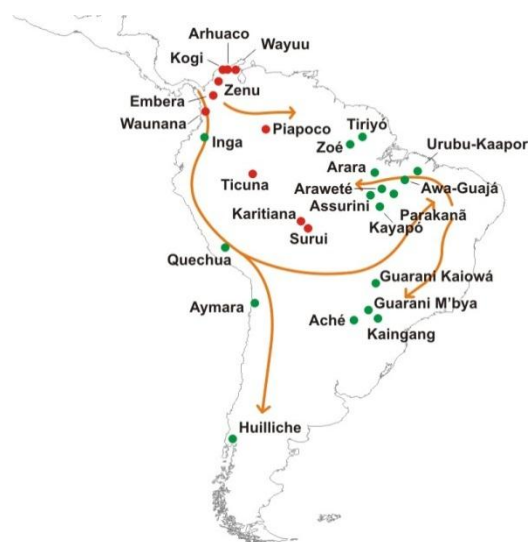


Figura 17. Rotas migratórias para o povoamento da América do Sul segundo o modelo de SCHMITZ, 1983.

Capítulo 2

Comparação entre efeitos demográficos e seletivos sobre a região do MHC com base em marcadores genéticos do tipo microssatélites

Introdução

A variação genética dentro, e a diferenciação entre populações humanas, é moldada tanto pela história demográfica, como por processos determinísticos, como a seleção natural, sendo que ambos alteram as frequências alélicas. Na última década, a disponibilidade de dados em escala genômica permitiu a investigação de maneira até então inédita da contribuição relativa desses processos microevolutivos para a variabilidade genética de nossa espécie. Duas principais abordagens vêm sendo empregadas. A primeira é a abordagem de "outliers", que busca genes com padrões de variação ou diferenciação extremos em relação à distribuição dos demais, sugerindo que estiveram sob seleção. Alternativamente, distribuições nulas dos padrões de variação genética podem ser geradas usando estimativas de parâmetros obtidas a partir da totalidade dos dados, para então se buscar genes com padrões extremos (revisão em RONALD e AKEY, 2005, HARRIS e MEYER, 2006; SABETI *et al.*, 2007; AKEY *et al.*, 2009).

Apesar de haver diferenças marcantes entre os diversos estudos voltados para detectar seleção natural recente, um resultado recorrente é a presença de genes de função imunológica dentre os que apresentam maior evidência de estarem sob seleção (BUSTAMANTE *et al.*, 2005; NIELSEN *et al.*, 2005; VOIGHT *et al.*, 2006; AKEY *et al.*, 2009). Entretanto, os genes *HLA*, que são provavelmente a classe de genes para a qual há maior evidência de seleção natural desde antes dos dados em escala genômica (HUGHES, 2002; MEYER *et al.*, 2006), não figuram nessa lista. A explicação para esse padrão é relativamente simples: a seleção que atua sobre esses genes é do tipo balanceadora, que aumenta a variação intra-populacional, e os testes de seleção positiva são sensíveis ao aumento das frequências de alelos específicos, um efeito esperado da seleção direcional. Estudos mais recentes, que definiram estatísticas sensíveis a seleção balanceadora (ANDRES *et al.*, 2009) mostram que os genes *HLA*, em especial o gene *HLA-B*, apresentam fortes sinais de seleção. Meta-

análises também reforçam o padrão de variação extremo desses genes (SOLBERG *et al.*, 2008).

Em populações nativas da América, o perfil da variação genética dos genes *HLA* (em especial *HLA-B* e *HLA-DRB1*) difere dos demais genes até então estudados. Essas populações apresentam, para esse conjunto de genes, uma alta diversidade e presença de alelos endêmicos em alta proporção, enquanto que, para os demais genes, é observada baixa diversidade genética e os alelos são, em geral, um subconjunto dos alelos encontrados na Ásia (revisado em SALZANO, 2002). Esse padrão de variação genética já foi interpretado como sendo o resultado de seleção natural favorecendo alelos de *HLA* que apresentam valor adaptativo elevado, que surgiram especificamente no continente americano, e respondem a patógenos encontrados naquela região (BELICH *et al.*, 1992; WATKINS *et al.*, 1992; CADAVID e WATKINS 1997; PARHAM *et al.*, 1997). Alternativamente, interpretações demográficas devem também ser consideradas: a ocupação das Américas resultou em gargalos populacionais extremos, levando as populações da América do Sul a apresentarem os menores níveis de diversidades do mundo. É, portanto, razoável supor que gargalos extremos tenham alterado as frequências alélicas levando as populações a apresentarem perfis alélicos "extremos", altamente diferenciados daqueles de outras regiões do mundo.

Para distinguir entre efeitos seletivos e demográficos, a estratégia ideal é a comparação de marcadores neutros com aqueles candidatos a estarem sob seleção. Estudos prévios (ex., MEYER *et al.* 2006, SANCHEZ-MAZAS, 2007) enfrentaram dificuldades na comparação da diversidade genética de genes *HLA* com a de outros marcadores que evoluem por mecanismos mutacionais diferentes (ex., comparando a evolução de sequências de DNA com a variação em microssatélites). Há dificuldades nesses casos, pois além de diferenças seletivas, a própria natureza dos marcadores e seu padrão mutacional podem contribuir para diferenças nos padrões de variação.

No presente estudo, adotamos como estratégia a comparação de níveis de variação de uma mesma classe de marcadores genéticos (os microssatélites) em diferentes regiões do genoma (na região MHC, que contém os genes *HLA*, em comparação com o restante do genoma). Essa abordagem permite investigar até

que ponto a seleção natural altera as frequências alélicas e a diferenciação populacional dos genes *HLA* de modo diferente do restante do genoma, dessa forma fornecendo subsídios para testes que separam as contribuições demográficas e seletivas para o perfil de genes *HLA* nas Américas. Portanto para separar esses dois processos, nós utilizamos 61 microssatélites espalhados pelo genoma (descritos no capítulo 1) para recriar a história demográfica, e 20 microssatélites restritos a região do MHC para identificar a história da região candidata à seleção. Se os dois conjuntos de microssatélites apresentarem o mesmo padrão, sugere-se que a variação na região do MHC das populações nativas americana é moldada principalmente pela demografia. Contudo se os dois conjuntos de microssatélites apresentarem padrões diferentes, isso significa que há sinais visíveis que a seleção atuou sob a região do MHC.

A seguir, faremos um breve resumo das principais características e evidências de seleção natural na região do MHC e dos genes *HLA*. (Para mais detalhes, revisamos esse tema em um manuscrito, apresentado no ANEXO 2. O manuscrito produzido em colaboração com colegas de doutorado foi publicado em 2009 como capítulo de abertura do livro “Estudos de Associação HLA x Doenças: Extratos do II Simpósio Brasileiro”). Em seguida apresentaremos como estudo de caso o gene *HLA-B*, discutiremos seu perfil genético nas populações nativas da América, e como no presente estudo pretendemos contribuir para as questões em aberto.

O MHC e genes HLA: características e evidências de seleção

A região do MHC, localizada no braço curto do cromossomo 6 (6p21.3) é uma das mais estudadas do genoma humano. Essa região apresenta a maior densidade de genes por kb (em média 1 gene a cada 16kb) e estima-se que cerca de 40% desses genes estão envolvidos na resposta imunológica (*MHC sequencing consortium*, 1999). Historicamente, o complexo gênico do MHC é dividido em três regiões: classe I, II e III. Nas regiões de classe I e II está localizada a família gênica HLA enquanto que na região de classe III encontram-se outros genes que codificam moléculas também relacionadas com a resposta imune adaptativa, como citocinas (TNF, $LT\alpha$ e $LT\beta$) e moléculas ligadas ao sistema complemento (C2, C4A, C4B) (ABBAS *et al.*, 2008).

As moléculas HLA são glicoproteínas de superfície celular que desempenham papel central na captura e apresentação de peptídeos na superfície celular. O complexo HLA/peptídeo antigênico é reconhecido na superfície celular pelos receptores de célula T durante a inspeção imunológica (figura 18).

O sistema HLA é dividido em dois grupos principais. Os genes de HLA de classe I, são subdivididos em genes clássicos (*HLA-A, B e C*) e não clássicos (*HLA-E, F e G*), e HLA de classe II (*HLA-DRA, DRB, DQA, DQB, DPA e DPB*). Via de regra, as moléculas de classe I ligam peptídeos antigênicos de origem endógena (por exemplo, patógenos que invadem a célula) e o complexo é reconhecido pelos linfócitos T CD8. Por outro lado, as moléculas de classe II se ligam a peptídeos antigênicos de origem exógena (patógenos que invadem o meio extra-celular e são fagocitados pela célula) e o complexo é reconhecido pelos linfócitos T CD4 (ABBAS *et al.*, 2008).

As moléculas de HLA em especial as moléculas de classe I também desempenham papel importante como ligantes dos receptores KIR (Killer cell Immunoglobulina-like Receptor). Esses receptores regulam a ativação e inibição das células NK (Natural Killer) responsáveis por reconhecer células infectadas e destruí-las. Dependendo do haplótipo HLA/KIR de cada indivíduo a interação entre essas moléculas pode ser mais ou menos forte resultando em perfis distintos de ativação e intensidade da resposta imune (ABBAS *et al.*, 2008).

Diversos estudos apontam que a região do MHC e, em especial os genes *HLA*, apresentam indícios de estarem evoluindo sob seleção do tipo balanceadora (HEDRICK e THOMSON, 1983; BUBB *et al.*, 2007; ver revisão em NUNES *et al.*, 2009 – ANEXO 2). Dentre as características compatíveis com a atuação desse tipo de seleção estão: 1) Níveis de diversidade alélica superior ao esperado sob neutralidade (HEDRICK e THOMSON, 1983; MEYER *et al.*, 2006); 2) Altos níveis de diferenciação entre alelos (10 vezes superior que a média genômica) (BUDD *et al.*, 2007); 3) Taxas de substituição não sinônima superior à sinônima na região da molécula responsável pela ligação do peptídeo (HUGHES e NEI, 1988); 4) Polimorfismos trans-específicos (isto é, compartilhamento de alelos entre humanos e chimpanzé) (KLEIN *et al.*, 1993) e 5) Altos níveis de desequilíbrio de ligação (HUTTLEY *et al.*, 1999). Tal desequilíbrio de ligação pode ter duas origens: a) efeito carona, onde a seleção num gene cria

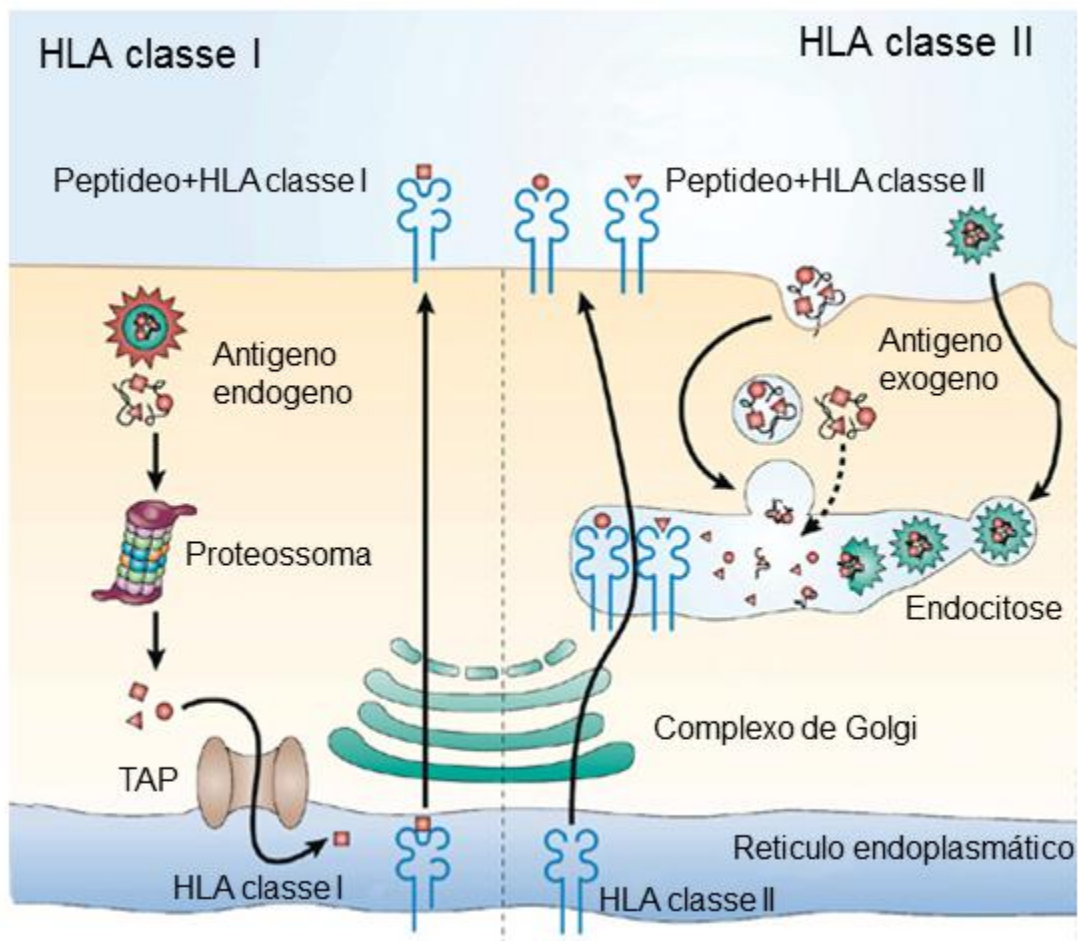


Figura 18. Representação gráfica do processamento e apresentação do peptídeo antigênicos pelas moléculas HLA. Na via de classe I, os peptídeos de origem endógena são processados pelo proteossoma, transportados pelas proteínas TAP até o retículo endoplasmático onde se encontra as moléculas de HLA de classe I, o peptídeo liga-se a fenda da molécula apresentadora e o complexo HLA + peptídeo é transportado através do complexo de Golgi para a superfície celular, onde serão reconhecidos pelos receptores dos linfócitos T CD8. Na via de classe II, antígenos exógenos são fagocitados pela célula e empacotados em vesícula fagocitária onde serão lisados. As moléculas HLA de classe II saem do retículo endoplasmático e vão até as vesículas fagocitárias onde ligam-se aos peptídeos. As vesículas contendo HLA+peptideo fundem-se a membrana plasmática e as moléculas de HLA de classe II apresentam os peptídeos que são reconhecidos por linfócitos T CD4.

Fonte: VILLADANGOS e SCHNORRER (2007) modificada.

desequilíbrio de ligação com outros, fisicamente próximos; b) a seleção atuando sobre combinação de alelos em mais de um gene, ou seja, o haplótipo é o alvo da seleção (ver detalhes em NUNES *et al.*, 2009 – ANEXO 2).

Partindo do princípio de que a variação genética observada nos genes *HLA* está correlacionada com o conjunto de peptídeos antigênicos que eles são capazes de apresentar, algumas hipóteses foram formuladas para explicar como a seleção atuou ou atua sobre esses genes. Uma das mais difundidas é a de vantagem do heterozigoto, segundo a qual indivíduos heterozigotos apresentam maior valor adaptativo em relação a indivíduos homozigotos, uma vez que possuiriam uma maior capacidade de apresentação de peptídeos diferentes quando comparado com indivíduos homozigotos (DOHERTY e ZINKERNAGEL, 1975). Outra hipótese é a de seleção dependente de frequência, segundo a qual uma mutação de escape que surge no genoma de agentes infecciosos exerceria seleção contra os alelos de *HLA* comuns, e favoreceriam os alelos raros, aos quais os patógenos ainda não estão adaptados (APANIUS *et al.*, 1997 e HEDRICK, 2002; revisão em NUNES *et al.*, 2009). Um terceiro modelo sugere que o regime de seleção pode alterar temporal e espacialmente dependendo da alteração do conjunto de patógenos e de sua distribuição geográfica (HEDRICK, 2002). Esse aspecto dinâmico do processo evolutivo resultaria na alta diversidade nesses genes, mas não pressupõe a vantagem do heterozigoto. Apesar de ainda haver debate sobre quais desses mecanismos melhor se adequa aos dados, a argumentação de que a seleção sobre genes *HLA* é movida por patógenos é altamente corroborada, inclusive por estudos que mostram uma correlação positiva entre a abundância de patógenos numa região geográfica e a diversidade de genes *HLA* (PRUGNOLLE *et al.*, 2005).

Portanto nossos objetivos ao caracterizar 20 microssatélites na região do MHC são a) estabelecer o potencial dos microssatélites como ferramenta para inferir seleção em genes *HLA* de populações ameríndias; b) inferir o padrão de variabilidade nas regiões adjacentes a genes aparentemente favorecidos pela seleção natural c) baseado nesses resultados, investigar a contribuição de eventos demográficos e seletivos para os níveis de variação e diferenciação genética dos genes *HLA* em populações nativas do continente americano.

O gene *HLA-B* e seu perfil nas Américas

Dentre os genes *HLA*, o gene *HLA-B* é conhecido por apresentar padrões de variação e diferenciação extremos, raros no restante do genoma. O gene *HLA-B* é o mais polialélico dentre os genes *HLA*, com mais de 2000 alelos descritos na população mundial (IMGT, 2011). Desde o início da década de 90 é reconhecido que as populações nativas, em especial as sul-americanas, apresentam um grande número de alelos endêmicos (isto é, estão presentes nas Américas e ausentes das demais regiões do mundo, BELICH *et al.*, 1992). Além disso, sua presença na América ocorre em altas frequências (CADAVID *et al.*, 1997) e cada população originalmente analisada (Guarani, Kaingang e Waorani) apresentava um conjunto distinto de alelos (BELICH *et al.*, 1992).

Os alelos de *HLA-B* estão diretamente envolvidos na proteção e/ou susceptibilidade a infecções virais (TRACHTENBERG *et al.*, 2003). Assim, é plausível imaginar que ao colonizar o continente americano as populações ameríndias ancestrais encontraram patógenos endêmicos que podem ter exercido pressão seletiva sobre genes do sistema imunológico, inclusive o *HLA-B*. Baseado nessa ideia foi elaborado um modelo, denominado de *turnover* de alelos, para explicar a “substituição” dos alelos do Velho Mundo por alelos endêmicos em alta proporção e frequência nas populações nativas americanas. De acordo com esse modelo, alelos exclusivos de populações nativas americanas são, de modo geral, produtos de conversões e recombinações gênicas de alelos presentes no Velho Mundo, e que teriam aumentado de frequência porque apresentavam com maior eficiência os antígenos patogênicos endêmicos do continente sul americano. Segundo os autores, o pequeno tamanho populacional e consequente intensa deriva genética associados com os baixos níveis de fluxo gênico entre as populações teriam contribuído para a efetivação desse processo (BELICH *et al.*, 1992; WATKINS *et al.*, 1992; CADAVID e WATKINS 1997; PARHAM *et al.*, 1997).

Esse modelo, apesar de amplamente citado na literatura, nunca havia sido detalhadamente testado, ou examinado num contexto regional mais amplo. Com esse intuito FRANCISCO em 2009 revisitou esse tema, analisando 26 populações nativo-americanas, e identificou um aumento na proporção e frequência de alelos endêmicos nas populações que habitam as regiões climáticas neotropicais.

Diferentemente de estudos anteriores, ele mostrou que as diferenças de perfis alélicos não ocorrem entre populações individuais, mas sim entre grupos de populações.

O autor também identificou através do teste de *D* de Tajima, sinais significativos de seleção balanceadora atuando ao longo do continente americano, porém não encontrou diferença entre as populações que apresentam maiores ou menores frequências de alelos endêmicos. Concluiu-se que o teste de neutralidade aplicado estava detectando um sinal seletivo anterior à entrada do homem moderno no continente americano.

Contudo, o autor nota que como a maioria dos alelos de *HLA-B* surgiu a partir da recombinação ou conversão gênica entre alelos pré-existentes, o teste de neutralidade de *D* de Tajima (que utiliza uma estatística baseado em dois parâmetros: o número médio de diferenças nucleotídicas entre pares de sequências e o número de sítios segregantes na amostra) seria “cego” para alelos que surgem a partir desses mecanismos genéticos (uma vez que a recombinação e conversão apenas “redistribuem” a variação previamente existente entre novos alelos). Desse modo, o teste estaria identificando principalmente seleção resultante da diferença entre as linhagens, e não entre os alelos de *HLA-B* nas Américas.

Os alelos dos genes *HLA* são primariamente classificados em linhagens. As linhagens consistem em conjuntos de alelos que são reconhecidos sorologicamente por apresentarem o mesmo conjunto de epítomos. Em nível molecular as linhagens apresentam haplótipos característicos que são compartilhados entre os alelos que pertencentes àquela linhagem.

A hipótese de pressão seletiva recente atuando nos alelos endêmicos e em alta frequência das populações ameríndias continua em aberto, e o presente estudo pretende contribuir para essa questão. Assim, genotipamos seis microssatélites adjacentes ao gene *HLA-B* em um subconjunto das populações e amostras utilizadas por FRANCISCO, (2009). Nosso objetivo é utilizar o desequilíbrio de ligação para verificar se microssatélites flaqueadores do gene *HLA-B* possuem um padrão de variação que evidencia uma mudança de frequência recente, compatível com efeito da seleção nesses alelos endêmicos e em alta frequência.

Material e métodos

Material

Amostras

Para os microssatélites da região do MHC, foram genotipados 590 indivíduos, 525 pertencentes a 28 populações nativo americanas, 25 indivíduos africanos da população Ovimbundu de Angola e 40 indivíduos europeus oriundos da população portuguesa (tabela 13, figura 19).

As amostras de DNA utilizadas no presente estudo foram obtidas através de doação de bancos de DNA, sob tutela dos Professores Dr. Eduardo José Melo dos Santos e João Guerreiro do Laboratório de Genética Humana e Médica da Universidade Federal do Pará – LGHM UFPA (amostras indígenas da Amazônia), do Professor Dr. Andrés Ruiz-Linares do Galton Institute da Universidade de Londres – GI UL (amostras das demais populações nativo-americanas) e do Professor Dr. Jorge Rocha da Universidade do Porto – UP (amostras de Ovimbundu e Portugueses).

Marcadores Moleculares

Microssatélites autossômicos da região do MHC

Foram genotipados 20 microssatélites autossômicos localizados na região do MHC, cuja identificação e informações gerais estão indicadas na tabela 14 e a localização na figura 20.

Microssatélites autossômicos espalhados pelo genoma

Utilizamos 61 microssatélites autossômicos (descritos no capítulo 1 da presente tese) como marcadores neutros para contrastar com os marcadores da região do MHC, candidatos à seleção.

Os conjuntos de microssatélites distribuídos ao longo do genoma e restritos a região do MHC foram genotipados exatamente nos mesmos indivíduos de cada população.



Figura 19. Mapa com a localização das populações analisadas no presente estudo. O ponto Ticuna representa as populações Ticuna Arara e Ticuna Tarapaca, Arara representa Arara do Laranjal e Arara do Iri e Kayapó representa Kayapó Krokraimoro e Kayapó Xikrin.

Tabela 13. Populações genotipadas pelo presente estudo, para 20 microssatélites localizados na região do MHC.

População	Localização	Tamanho amostral (N)	Procedência
Ovimbundu	Angola	25	CTM – UP
Portugueses	Portugal	40	CTM – UP
Chipewyan	Canadá	20	GI-UL
Cree	Canadá	18	GI-UL
Ojibwa	Canadá	20	GI-UL
Quiche	Guatemala	20	GI-UL
Cabecar	Costa Rica	20	GI-UL
Guaymi	Panamá	18	GI-UL
Kogi	Colômbia	17	GI-UL
Arhuaco	Colômbia	17	GI-UL
Wanana	Colômbia	20	GI-UL
Wayuu	Colômbia	17	GI-UL
Embera	Colômbia	17	GI-UL
Zenu	Colômbia	20	GI-UL
Ingá	Colômbia	17	GI-UL
Aymara	Chile	20	GI-UL
Huilliche	Chile	20	GI-UL
Ticuna Tarapaca	Amazonas	17	GI-UL
Ticuna Arara	Amazonas	19	GI-UL
Tiriyó	Pará	20	LGHM – UFPA
Zo'é	Pará	20	LGHM – UFPA
Arara do Laranjal	Pará	20	LGHM – UFPA
Arara do Iriri	Pará	18	LGHM – UFPA
Urubu-Kaapor	Maranhão	20	LGHM – UFPA
Awá-Guajá	Maranhão	14	LGHM – UFPA
Parakanã	Pará	11	LGHM – UFPA
Assurini	Pará	22	LGHM – UFPA
Awereté	Pará	22	LGHM – UFPA
Kayapó Krokraimoro	Pará	20	LGHM – UFPA
Kayapó Xikrin	Pará	21	LGHM – UFPA
Total		590	

Tabela 14. Microssatélites da região do MHC analisados no presente estudo.

<i>Locus</i>	Marcador	Tipo de repetição
D6S276	D6S276	CA (dinucleotídeo)
D6S105	D6S105	CA (dinucleotídeo)
D6S2972	MOG2	AC (dinucleotídeo)
D6S265	D6S265	CA (dinucleotídeo)
D6S2685	HLAC-CA	TA (dinucleotídeo)
D6S2674	HLABC-CA	CA (dinucleotídeo)
D6S2929	D6S2929	TTTG (tetranucleotídeo)
D6S2928	D6S2928	TTTG (tetranucleotídeo)
D6S2927	D6S2927	TTTG (tetranucleotídeo)
D6S2926	D6S2926	TTTC (tetranucleotídeo)
MICA	MICA-CA	GCT (trinucleotídeo)
D6S2789	TNF-CA	AG (dinucleotídeo)
D6S2671	BAT2-CA	GT (dinucleotídeo)
D6S273	DS273	GT (dinucleotídeo)
D6S2665	DRA-CA	AC (dinucleotídeo)
D6S2876	G51152	GT (dinucleotídeo)
D6S2874	TAP-CA	GA (dinucleotídeo)
D6S2749	RING3	GT (dinucleotídeo)
D6S439	D6S439	CA (dinucleotídeo)
D6S291	D6S291	CA (dinucleotídeo)

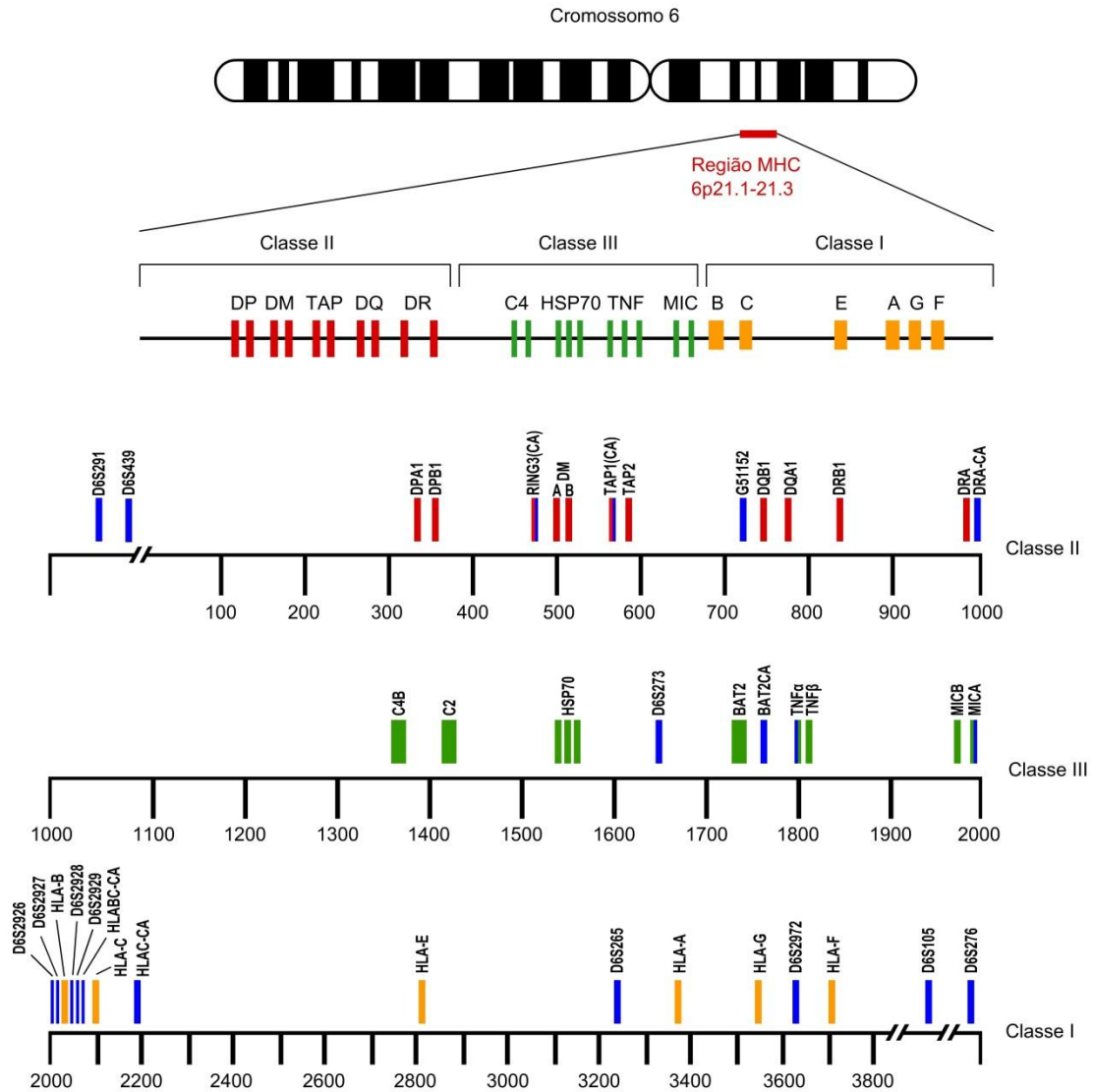


Figura 20. Localização dos microssatélites genotipados na região do MHC. As marcações em azul representam os microssatélites e em vermelho, genes da região de classe II, em verde, genes da região de classe III e em amarelo, genes da região de classe I do MHC.

Metodologia

Extração de DNA

As amostras analisadas tiveram seu DNA extraído pelos métodos de fenol-clorofórmio (SAMBROOK *et al.*, 1989) ou *salting out* (LAHIRI e NURENBERG, 1991) nos respectivos laboratórios que as doaram.

Amplificação do genoma inteiro

As amostras de DNA utilizadas passaram pelo processo de amplificação inteira do genoma realizado com o *kit GenomiPhi V2 DNA Amplification Kit (GE Healthcare)* seguindo instruções do fabricante.

As amostras de populações não brasileiras tiveram o processo de amplificação inteira do genoma realizada em seus laboratórios de origem.

Amplificação de fragmentos os microssatélites

As amplificações dos fragmentos de DNA contendo microssatélites foram realizadas através de reações de PCR individuais ou por multiplexagem dos oligonucleotídeos iniciadores na reação de PCR (tabela 15).

As reações de PCR foram realizadas com 0,2mM DNTPs, tampão 1x, 0,07U/uL de taq polimerase e cerca de 10ng de DNA (as concentrações de MgCl₂ e dos oligonucleotídeos iniciadores variaram como apresentado na tabela 15).

Os seguintes ciclos de amplificação foram realizados: 94°C por 11 minutos seguidos por 30 ciclos de 94°C por 1 minuto, TA (ver tabela 15) por 1 minuto, 72°C por 1 minuto e um ciclo de extensão final de 60°C por 1 hora.

A padronização dos protocolos de amplificação dos microssatélites foi realizada com auxílio das alunas Maria Helena T. Maia e Ana Paulo Castro da Universidade Federal do Pará.

Reação de genotipagem dos microssatélites

Agrupamento dos produtos de PCR na eletroforese

Para as eletroforeses, produtos de reações distintas de PCR com fluorescências distintas e com tamanhos de fragmentos que não se sobrepõem foram agrupados de acordo com o apresentado na tabela 15.

Tabela 15. Informações sobre microssatélites do MHC, reação de PCR (Temperatura de hibridação – TA; concentração dos oligonucleotídeos iniciadores – oligos e concentração de cloreto de magnésio – MgCl₂), os conjuntos de multiplexagem, e agrupamentos para a eletroforese.

Marcador	Tamanho (pb)	Fluorescência	TA (°C)	Oligos (pmol)	MgCl ₂ (mM)	Multiplex PCR	Agrupamento Eletroforese
<i>HLA-BC</i>	90-120	VIC	57	0,3	3,34	1	A
G51152	192-236	NED	57	0,3	3,34	1	A
D6S273	132-158	VIC	57	0,3	3,34	1	A
DRA-CA	240-270	NED	57	0,75	3,34	1	A
D6S2926	253-285	FAM	60	1,5	2,5	2	A
D6S2929	172-210	VIC	60	1,5	2,5	2	A
D6S2874	187-211	FAM	55	0,75	3,0	3	B
D6S291	166-186	NED	55	0,75	3,0	3	B
D6S276	63-151	NED	55	0,75	3,0	3	B
MICA	180-200	VIC	55	0,75	3,34	4	B
D6S2927	228-248	FAM	55	0,5	3,0	5	B
D6S2928	253-273	VIC	55	0,5	3,0	5	B
RING3	225-237	FAM	60	1,5	2,5	6	C
TNF	124-134	NED	60	0,8	2,5	6	C
BAT2CA	135-155	VIC	60	0,2	2,5	6	C
D6S439	270-300	NED	60	0,75	3,0	7	D
D6S265	176-218	VIC	55	0,75	3,0	7	D
D6S2885	100-120	VIC	55	0,75	3,0	7	D
D6S105	144-164	FAM	55	0,75	3,0	7	D
D6S2972	122-160	NED	55	1,5	2,5	8	D

Reação de desnaturação

Para a desnaturação foram misturados 0,5µL do produto de PCR acrescido de 0,5µL do marcador de tamanho molecular (*GeneScanTM – 500 ROXTM – Applied Biosystems*) e 7,5µL de formamida (*HI-DITM formamide – Applied Biosystems*). Essa

solução foi homogeneizada e aquecida a 95°C por 3 minutos e em seguida resfriada em gelo por 20 minutos.

Eletroforese

A eletroforese para separação dos diferentes tamanhos de fragmentos de DNA contendo os microssatélites foi realizada no aparelho *ABI 3130 Genetic Analyzer* do Centro de Estudos do Genoma Humano da Universidade de São Paulo.

Análise das genotipagens

Os microssatélites foram analisados com o auxílio do programa *Peak Scanner™ Software v.1.0* da *Applied Biosystem*.

Obtenção do banco de dados final utilizado nas análises

Com intuito de assegurar uma melhor acurácia nas análises usamos alguns critérios para excluir indivíduos e *loci* que apresentaram excesso de falha de amplificação:

- 1) Indivíduos com menos de 50% dos loci selecionados amplificados com sucesso (44 indivíduos) foram excluídos;
- 2) *Loci* com menos de 50% das amostras individuais com sucesso de amplificação (*DRA-CA* e *TNF*) foram excluídos;
- 3) *Loci* com falha de amplificação em todos os indivíduos de alguma população foram excluídos (*RING3* e *D6S2929*);
- 4) Indivíduos sem correspondência com o banco de dados de 61 microssatélites autossômicos espalhados pelo genoma (52 indivíduos excluídos – ver tabela 16).

Dessa maneira, o banco de dados final, utilizado nas análises, é composto por 494 indivíduos (tabela 16) e 16 marcadores moleculares do tipo microssatélites restritos a região do MHC, perfeitamente pareados com 61 microssatélites espalhados pelo genoma (Capítulo 1).

Para as análises restritas à questão de detecção de sinais de seleção em alelos do gene *HLA-B* nas Américas utilizamos os dados para genótipos de *HLA-B* gerados por FRANCISCO (2009), e os microssatélites flanqueadores do gene *HLA-B* aqui descritos (*MICA*, *D6S2926*, *D6S2927*, *D6S2928*, *HLA-BC* e *D6S2885*) (tabela 16). Esse banco é constituído por 415 amostras (tabela 16).

Tabela 16. Número de amostras utilizadas nas análises após a exclusão de indivíduos com excesso de falha de amplificação.

População	Tamanho amostral inicial	Amostras utilizadas nas análises de caracterização da região do MHC	Amostras utilizadas nas análises restritas ao gene <i>HLA-B</i>
Ovimbundu	25	16	0
Portugueses	40	23	0
Chipewyan	20	14	14
Cree	18	17	17
Ojibwa	20	17	15
Kaqchikel	20	12	19
Cabecar	20	20	19
Guaymi	18	18	18
Embera	17	11	14
Zenu	17	16	16
Kogi	20	12	12
Arhuaco	17	17	17
Wayuu	17	15	16
Waunana	20	20	20
Inga	17	17	14
Aymara	20	18	20
Huilliche	20	20	20
Ticuna Arara	17	17	17
Ticuna Tarapaca	19	16	14
Tiriyó	20	16	6
Zo'é	20	11	12
Arara Laranjal	20	16	11
Arara Iriri	18	17	9
Urubu Kaapor	20	20	18
Awá-Guajá	14	11	9
Parakanã	11	9	6
Assurini	22	20	19
Arawete	22	20	14
Kayapó Krokaimoro	20	18	15
Kayapó Xikrin	21	20	14
Total	590	494	415

Análise dos dados

Frequências alélicas

As frequências alélicas foram obtidas através da contagem direta do número de cromossomos e de indivíduos, respectivamente.

Equilíbrio de Hardy-Weinberg

As frequências observadas dos genótipos foram comparadas com aquelas esperadas de acordo com o princípio de equilíbrio de Hardy-Weinberg, através do teste exato de GUO e THOMPSON (1992), implementado no pacote de programas do Arlequin 3.5 (EXCOFFIER e LISCHER, 2010). Essa análise permite verificar a confiabilidade da metodologia de genotipagem dos microssatélites, uma vez que a existência de alelos nulos pode resultar em dados com excesso de homozigotos.

Descrição da variabilidade genética

Os índices de diversidade genética foram estimados pelo método do pacote computacional Arlequin 3.5 (EXCOFFIER e LISCHER, 2010).

A diversidade gênica corresponde à heterozigose esperada (probabilidade de amostrar ao acaso dois alelos em uma dada população e eles serem diferentes). A diversidade genética (\hat{H}) e a sua variância amostral ($V(\hat{H})$) foram calculadas da seguinte maneira:

$$\hat{H} = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$$

$$V(\hat{H}) = \frac{2}{n(n-1)} \left\{ 2(n-2) \left[\sum_{i=1}^k p_i^3 - \left(\sum_{i=1}^k p_i^2 \right)^2 \right] + \sum_{i=1}^k p_i^2 - \left(\sum_{i=1}^k p_i^2 \right)^2 \right\}$$

Onde n é o número de cópias gênicas na amostra, k o número de alelos e p_i a frequência dos i - alelos.

O theta-H é um estimador de theta ($4N\mu$), ele é específico para o caso de microssatélites, supondo o "stepwise mutation model". Essas qualificações são importantes pois a expressão abaixo não se aplica a outras classes de marcadores

$$\hat{\theta}_H = \frac{1}{(1 - \hat{H})^2} - 1,$$

Onde \hat{H} é a heterozigose esperada.

O número de alelos (K) foi estimado por contagem direta do número de alelos observados em cada *locus* e população

Diferenciação inter-populacional

Matrizes de diferenciação entre os pares de populações foram estimadas utilizando o programa *Arlequin v.3.5* (EXCOFFIER e LISCHER, 2010) para os índices de diferenciação populacional F_{ST} e R_{ST} . As significâncias foram obtidas através 10.000 permutações dos indivíduos entre as populações.

A diferenciação populacional para os índices F_{ST} e R_{ST} entre as regiões e/ou grupos de populações foi estimada através do particionamento hierárquico da variação genética ao longo dos *loci* (AMOVA). O intervalo de confiança foi obtido realizando reamostragens ao longo dos *loci* através de 10.000 *bootstraps*.

Detecção de loci sob seleção

Para detectar *loci* sob seleção utilizamos a abordagem baseada na relação entre o índice F_{ST} e a Heterozigose, implementada no programa *LOSITAN* (ANTAO *et al.*, 2008). O método cria uma distribuição nula a partir de simulações coalescentes condicionadas no valor de F_{ST} observado para os dados reais, assumindo um modelo de ilhas. A distribuição de valores de F_{ST} é sabidamente sensível à taxa de heterozigose, e dessa forma é essencial relacionar os valores de F_{ST} com os de H_e . O valor de F_{ST} estimado pelos 61 microssatélites distribuídos pelo genoma foram utilizados para gerar a distribuição. Em seguida, plotamos nessa distribuição os dados dos microssatélites do MHC e realizamos o teste para a hipótese deles desviarem dos valores esperados sob o modelo nulo. Os *loci* com valores de F_{ST} e H_e

que excedam o intervalo de confiança de 95% foram considerados como *outliers* da distribuição e, portanto, candidatos à seleção.

Determinação da fase gamética

A reconstrução dos haplótipos foi inferida a partir de análise Bayesiana implementada no programa *PHASE v.2.1* (STEPHENS *et al.*, 2004).

Desequilíbrio de ligação

O coeficiente de desequilíbrio de ligação D' foi obtido através da equação de Hedrick (1987) implementada no pacote do programa *Pypop* (LANCASTER *et al.*, 2007).

Heterozigose específica para os alelos de *HLA-B*

Criamos uma função em R baseada no índice de Heterozigose Haplótipo Específica – HSH (Haplotype Specific Heterozygosity) descrita por MALKKI *et al.* (2005) (Script no ANEXO 3). Esse índice estima a taxa de heterozigose para cada microssatélites condicional ao alelo de *HLA-B* em que ocorrem. Essa informação foi usada como ferramenta para detecção de alelos de *HLA-B* que possam ter sido selecionados.

O índice HSH é calculado através da estimativa das frequências alélicas de um determinado *locus* de microssatélite em um dado alelo de *HLA-B*. As frequências alélicas são então aplicadas na fórmula $HSH = 1 - \sum_{i=1}^k p_i^2$ onde k corresponde ao número de alelos de um determinado *locus* de microssatélite e p a frequência alélica (MALKKI *et al.*, 2005).

O objetivo dessa análise é observar se os alelos endêmicos, ou observados em alta frequência apenas no continente americano, apresentam a heterozigose específica associada a esses alelos (o índice HSH), reduzida em relação a alelos cosmopolitas e que apresentam alta frequência em outras regiões do mundo.

Diferenciação inter e intra linhagem de *HLA-B*

Usamos o índice de diferenciação populacional F_{ST} , implementado no pacote de programas Arlequin 3.5 (EXCOFFIER e LISCHER, 2010) para estimar o grau de diferenciação entre as linhagens e entre os alelos da mesma linhagem de *HLA-B*.

Associação de alelos de microssatélites e alelos e linhagens de *HLA-B*

Estimamos através de contagem direta a frequência que um determinado alelo de microssatélite estava associado a um determinado alelo ou linhagem de *HLA-B*.

Resultados

Avaliação dos dados

Para verificar se os dados gerados neste estudo apresentam o mesmo perfil de qualidade de genotipagem que os dados gerados no capítulo anterior, que serão usados como nossa referência de marcadores neutros, as análises descritas a seguir foram realizadas.

Quantificação das falhas de amplificação

Obtivemos 8,99% de falha de amplificação ao longo dos 16 *loci* de microssatélites das 30 populações genotipadas (figura 21e 22).

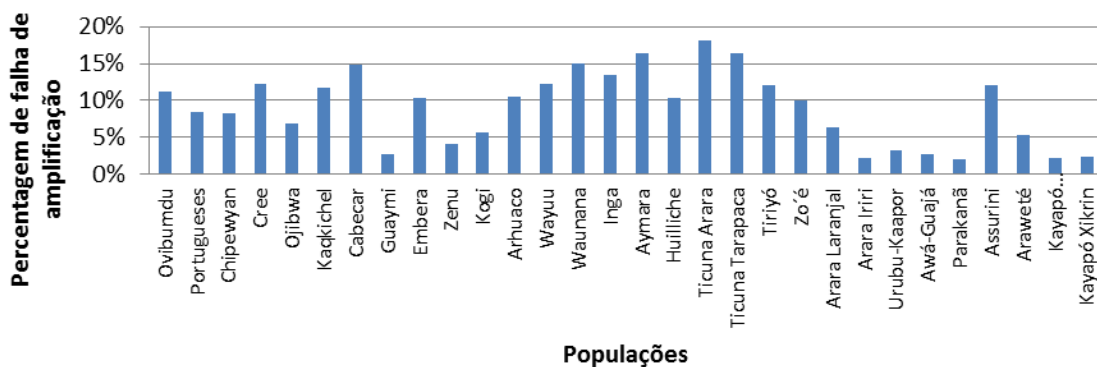


Figura 21. Percentagem de falha de amplificação ao longo das populações nos conjuntos de 16 *loci* de microssatélites da região do MHC.

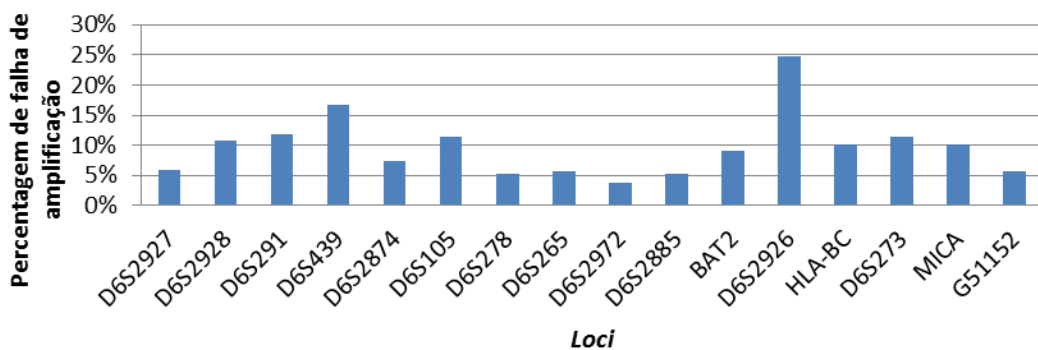


Figura 22. Percentagem de falha de amplificação ao longo dos 16 *loci* de microssatélites da região do MHC nas populações.

Estimativas de desvios de Hardy-Weinberg

Observamos 14,59% de desvios significativos de Hardy-Weinberg ($p < 0,05$). As figuras 23 e 24 mostram a distribuição dos desvios por população e por *loci* respectivamente. Após correção por Bonferroni, nenhum desvio permaneceu significativo ($p_{Bonferroni} = 0,00014$).

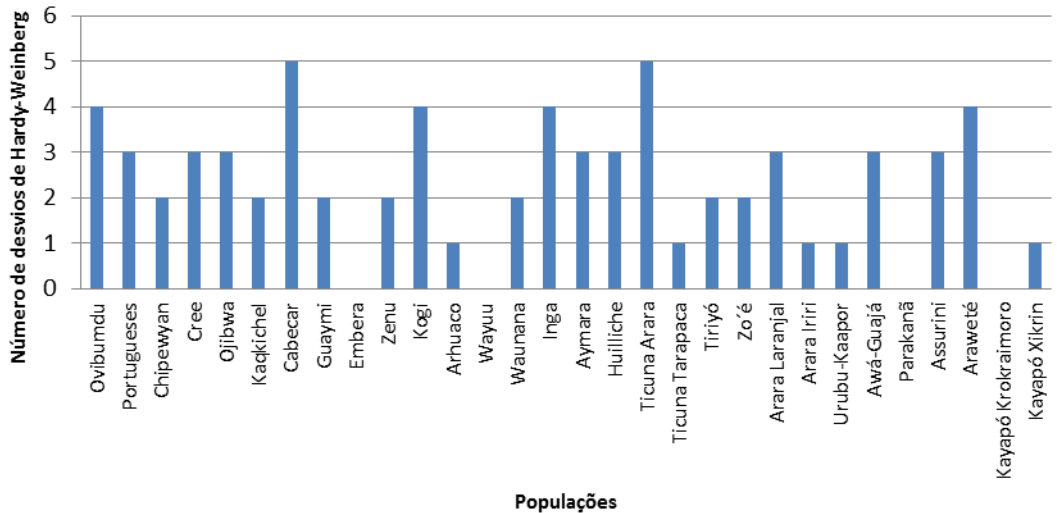


Figura 23. Distribuição dos desvios significativos de Hardy-Weinberg ($p < 0,05$) nas populações para o conjunto de 16 *loci* de microssatélites da região do MHC.

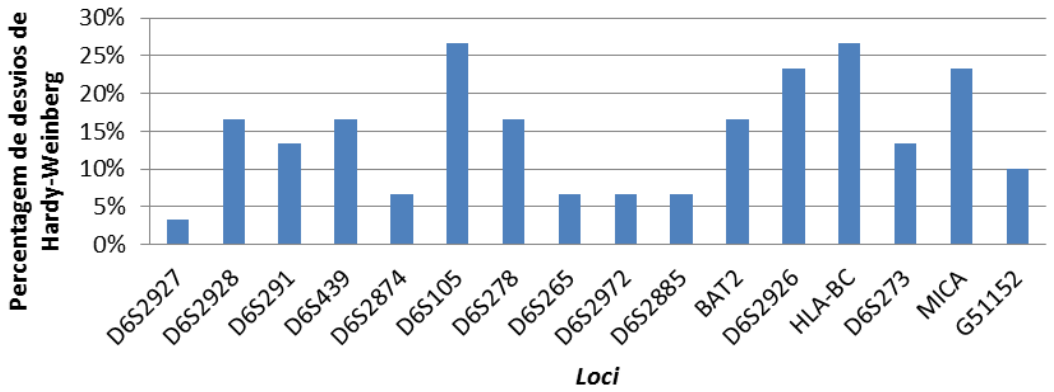


Figura 24. Distribuição dos desvios significativos de Hardy-Weinberg ($p < 0,05$) para cada um dos 16 *loci* da região do MHC nas populações.

Frequências alélicas

A distribuição das frequências alélicas para cada microssatélite ao longo das 30 populações analisadas é apresentada nas figuras 25 a 28. É possível observar que os alelos encontrados na América são um subconjunto dos alelos encontrados nas populações africana (Ovimbundu) e europeia (Portugueses). Contudo, a frequência desses alelos varia entre as populações. Através da comparação das frequências alélicas entre os pares de populações, realizada através do teste de Fisher (Tabelas A.3 a A.18 – ANEXO 1) é possível observar que a maioria da diferença da distribuição das frequências encontra-se entre africanos e europeus em relação aos nativo-americanos. Também observamos que as populações do leste (em especial as populações, Ticuna Tarapaca, Arara do Iriri e Araweté) são as que apresentam maior diferença na distribuição das frequências alélicas na América (Tabelas A.3 a A.18 ANEXO 1).

Perfil dos microssatélites da região do MHC

Verificamos que há heterogeneidade entre os perfis dos microssatélites. Na tabela 17 apresentamos um resumo do número médio de alelos e a heterozigose total média dos microssatélites entre as 30 populações. Os microssatélites *HLA-BC*, *D6S2972* e *D6S276* são os mais diversos ($H_e = 0,788; 0,709; 0,698$ e número de alelos ($k = 8,43; 5,90; 5,70$, respectivamente). *G51152*, *D6S2928* e *D6S291* são os menos diversos ($H_e = 0,428; 0,431; 0,432$ e $k = 2,90; 2,80; 3,67$ respectivamente).

Comparação de índices de diversidade

Na tabela 18 apresentamos a comparação dos índices de diversidade para os 61 microssatélites espalhados pelo genoma e os 16 microssatélites da região do MHC para cada uma das 30 populações analisadas.

É possível observar que, com exceção das populações da América do Norte e as não nativas, para todos os índices estimados (H_e , T_{θ} e K) os valores encontrados para os microssatélites do genoma são mais elevados que para os microssatélites do MHC (tabela 18 e figura 29).

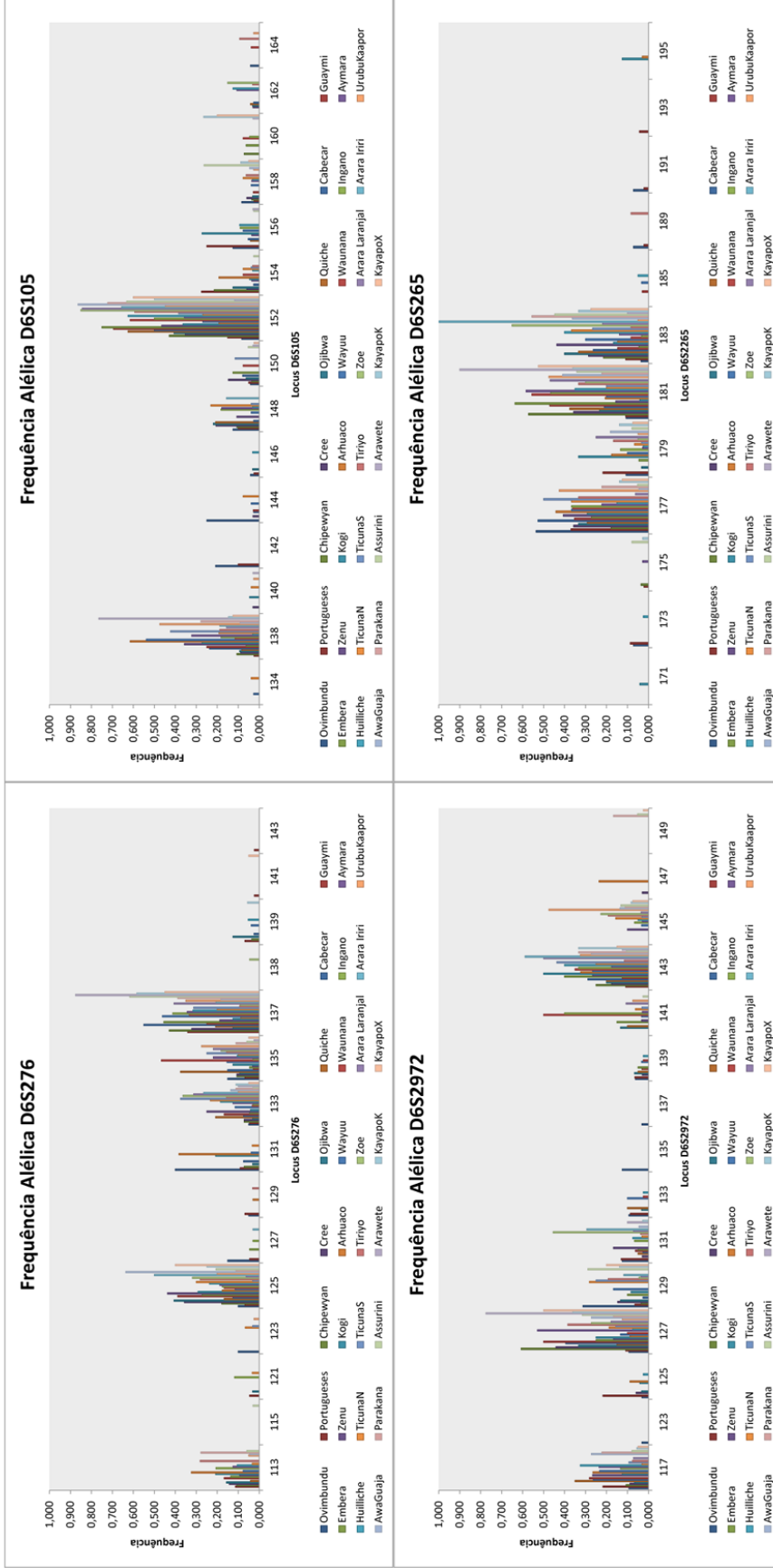


Figura 25. Distribuição das frequências alélicas dos microsatélites D6S276, D6S105, D6S2972, D6S265.

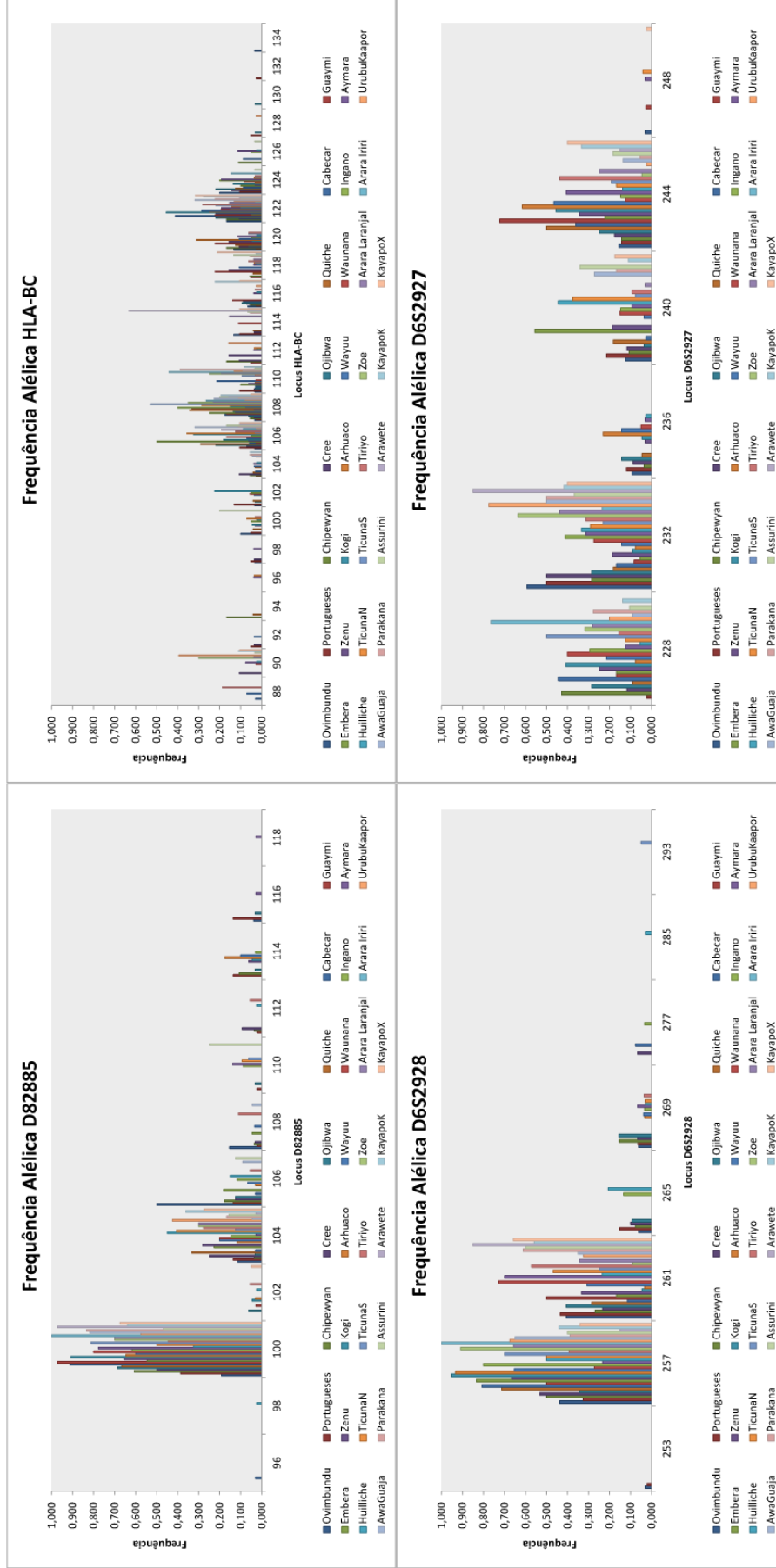


Figura 26. Distribuição das frequências alélicas dos microsossatélites D6S2885, HLA-BC, D6S2928, D6S2927.

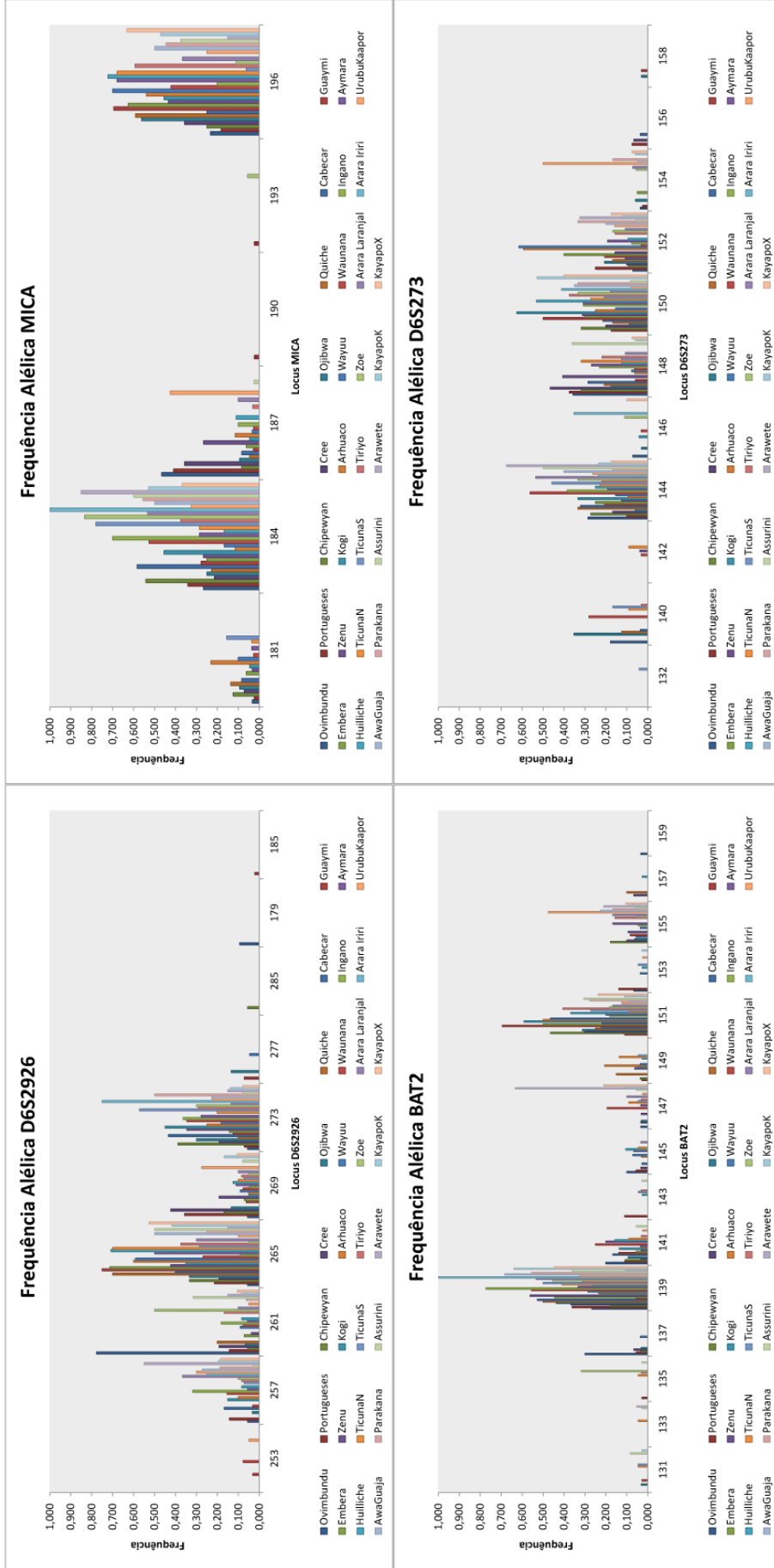


Figura 27. Distribuição das frequências alélicas dos microsossatélites *D6S2926*, *MICA*, *BAT2*, *D6S273*.

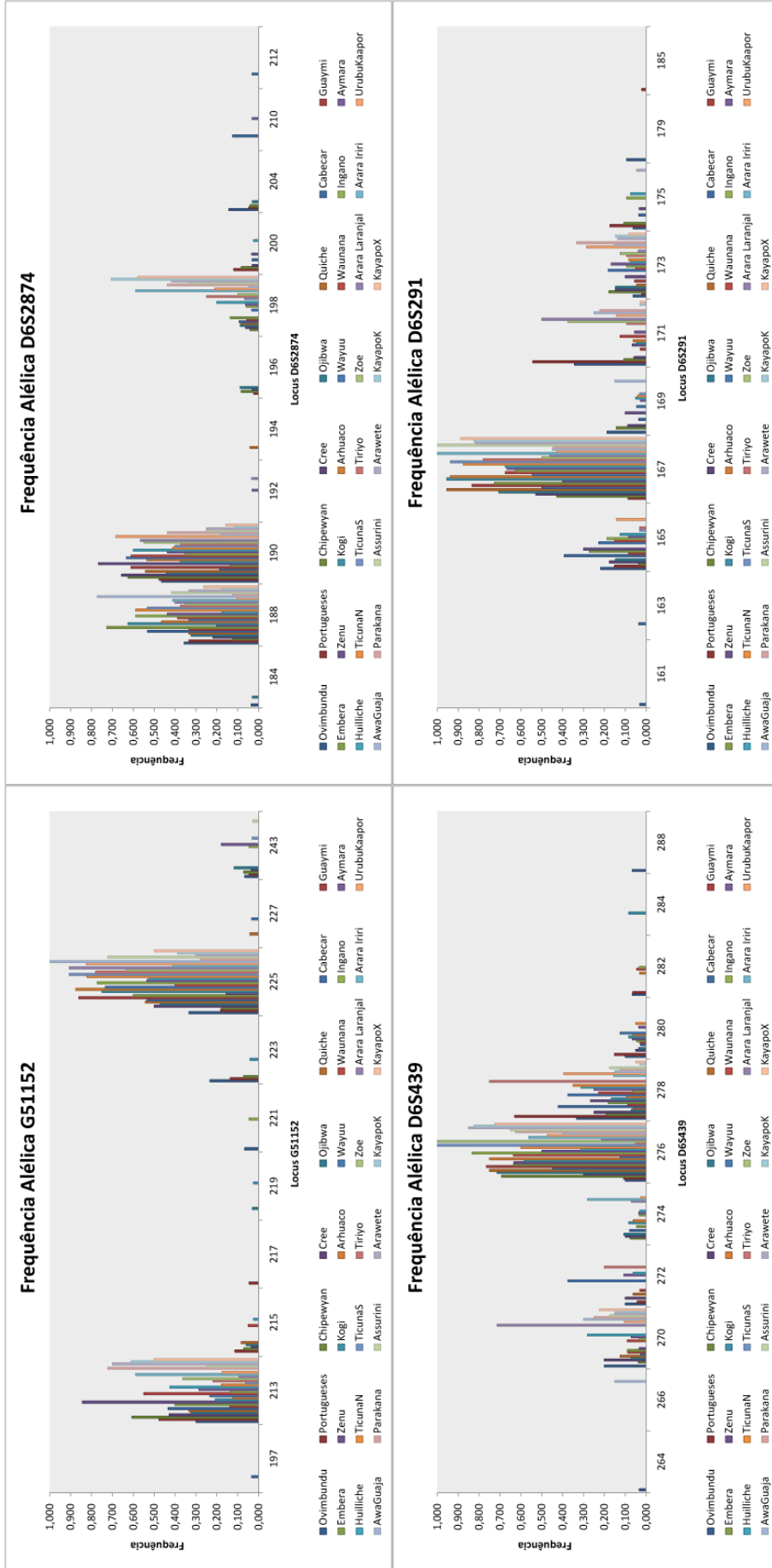


Figura 28. Distribuição das frequências alélicas dos microsossatélites G51152, D6S2874, D6S439, D6S291.

Tabela 17. Número médio de alelos e heterozigose média dos microssatélites restritos a região do MHC nas 30 populações analisadas.

<i>Loci</i>	Numero de alelos		Heterozigose total	
	Media	S.E.	Media	S.E.
<i>D6S276</i>	5,700	0,311	0,721	0,022
<i>D6S105</i>	5,433	0,394	0,609	0,031
<i>D6S2972</i>	5,900	0,301	0,733	0,019
<i>D6S265</i>	4,200	0,277	0,634	0,030
<i>D6S2885</i>	3,767	0,324	0,462	0,039
<i>HLABC</i>	8,433	0,509	0,817	0,015
<i>D6S2928</i>	2,800	0,206	0,447	0,033
<i>D6S2927</i>	4,233	0,164	0,646	0,023
<i>D6S2926</i>	4,467	0,184	0,650	0,022
<i>MICA</i>	3,267	0,179	0,526	0,027
<i>BAT2</i>	5,500	0,386	0,651	0,029
<i>D6S273</i>	4,667	0,211	0,703	0,016
<i>G51152</i>	2,900	0,222	0,443	0,032
<i>D6S2874</i>	3,433	0,218	0,562	0,016
<i>D6S439</i>	4,033	0,269	0,515	0,037
<i>D6S291</i>	3,667	0,273	0,450	0,046

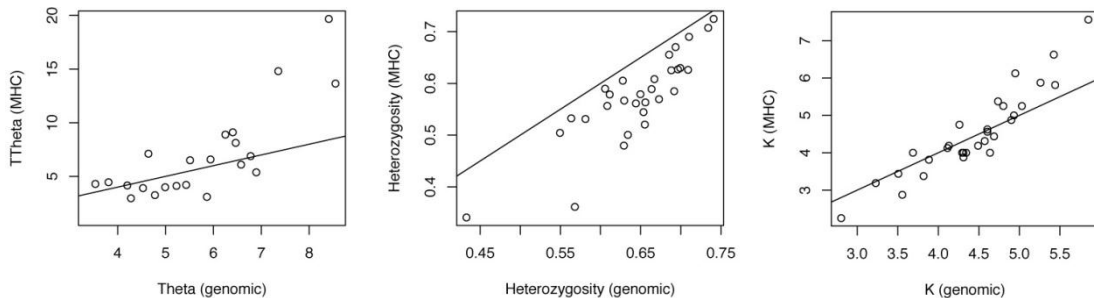


Figura 29. Gráfico de dispersão contrastando microssatélites restritos a região do MHC com microssatélites genômicos para os índices de diversidade Theta (h), Heterozigose (H_e) e o número de alelos (K).

Verificamos através do teste de Wilcoxon Mann Whitney que a diferença entre as taxas de heterozigose nas populações para o conjunto de microssatélites genômicos e do MHC diferem significativamente ($p=0,00001$), contudo o mesmo não é observado para a medida de diversidade Theta (H) ($p=0,7437$) e para o número de alelos ($p=0,7283$).

Tabela 18. Comparação dos índices de diversidade entre os microssatélites espalhados pelo genoma e os microssatélites da região do MHC. Onde He corresponde a Heterozigose, Theta (H) ao índice de diversidade e K ao número de alelos. Todos os índices correspondem à média para o conjunto dos *loci*.

População	Microssatélites Genômicos			Microssatélites MHC		
	He	Theta(H)	K	He	Theta (H)	K
Ovimbundu	0,73	2,06	5,42	0,74	2,11	6,62
Portugueses	0,74	2,10	5,85	0,76	2,25	7,56
Chipewyan	0,69	1,82	4,74	0,69	1,83	5,38
Cree	0,71	1,93	5,26	0,72	2,01	5,88
Ojibwa	0,69	1,85	4,95	0,70	1,89	6,13
Kaqchikel	0,69	1,85	4,57	0,61	1,60	4,31
Cabecar	0,66	1,74	4,93	0,61	1,61	5,00
Guaymi	0,63	1,64	4,30	0,50	1,50	4,00
Embera	0,65	1,71	4,31	0,57	1,54	3,88
Zenu	0,67	1,75	4,69	0,63	1,66	4,44
Kogi	0,58	1,55	3,69	0,55	1,52	4,00
Arhuaco	0,66	1,71	4,31	0,54	1,51	4,00
Wayuu	0,69	1,83	4,80	0,65	1,71	5,25
Waunana	0,63	1,64	4,49	0,59	1,57	4,19
Inga	0,67	1,77	4,61	0,59	1,57	4,56
Aymara	0,70	1,88	5,03	0,66	1,72	5,25
Huilliche	0,71	1,92	5,44	0,66	1,72	5,81
TicunaArara	0,63	1,64	4,26	0,63	1,64	4,75
TicunaTarapaca	0,63	1,66	4,12	0,52	1,50	4,13
Tiriyó	0,70	1,87	4,92	0,65	1,71	4,88
Zo'é	0,56	1,53	3,33	0,52	1,50	3,19
Arara Laranjal	0,61	1,60	3,89	0,58	1,55	3,81
Arara Iriri	0,45	1,52	2,89	0,34	1,72	2,25

continua

Tabela 18 (continuação). Comparação dos índices de diversidade entre os microssatélites espalhados pelo genoma e os microssatélites da região do MHC. Onde H_e corresponde a Heterozigose, Theta (H) ao índice de diversidade e K ao número de alelos. Todos os índices correspondem à média entre os *loci*.

População	Microssatélites Genômicos			Microssatélites MHC		
	H_e	Theta(H)	K	H_e	Theta (H)	K
Urubu –Kaapor	0,62	1,62	4,20	0,60	1,59	4,19
Awa-Guajá	0,57	1,54	3,59	0,55	1,52	3,44
Parakanã	0,61	1,60	3,90	0,62	1,61	3,38
Asurini	0,66	1,71	4,69	0,60	1,59	4,63
Arawete	0,58	1,55	3,62	0,37	1,64	2,88
Kayapó K	0,65	1,69	4,39	0,59	1,56	4,00
Kayapó X	0,66	1,74	4,69	0,59	1,56	4,00
Total	0,64	1,71	4,38	0,59	1,63	4,34

Diferenciação populacional: microssatélites restritos à região do MHC versus microssatélites espalhados pelo genoma

Na tabela 19 apresentamos a comparação do índice de diferenciação populacional (F_{ST}) entre os microssatélites espalhados pelo genoma e os restritos a região do MHC (essa mesma análise foi realizada com o índice R_{ST} , porém como o padrão apresentado foi similar ao F_{ST} , optamos por não apresentar esses resultados). De modo geral, o conjunto de microssatélites do MHC apresenta índices de diferenciação mais elevados que os espalhados pelo genoma, sendo que a maior diferença foi observada para a região Andina ($F_{ST \text{ Genômicos}} = 0,0217$ e $F_{ST \text{ MHC}} = 0,0993$).

Na figura 30 apresentamos a diferenciação populacional entre os pares de populações para os microssatélites genômicos e na figura 31 para os microssatélites da região do MHC.

Tabela 19. Comparação da diferenciação entre os microssatélites do genoma e restritos a região do MHC.

Região	Microssatélites do Genômicos			Microssatélites do MHC		
	F _{ST}	I.C 95%	valor-p	F _{ST}	I.C 95%	valor-p
Mundo	0,087	0,072 0,093	0,000	0,130	0,111 0,142	0,000
Africa e Europa	0,023	0,008 0,035	0,027	0,051	0,010 0,087	0,000
America	0,090	0,082 0,097	0,000	0,129	0,112 0,142	0,000
America do Norte	0,031	0,021 0,042	0,000	0,028	0,011 0,043	0,053
Noroeste América do Sul	0,077	0,068 0,085	0,000	0,108	0,082 0,131	0,000
Andes	0,022	0,014 0,030	0,001	0,099	0,050 0,136	0,000
Leste América do Sul	0,117	0,105 0,128	0,000	0,161	0,138 0,180	0,000

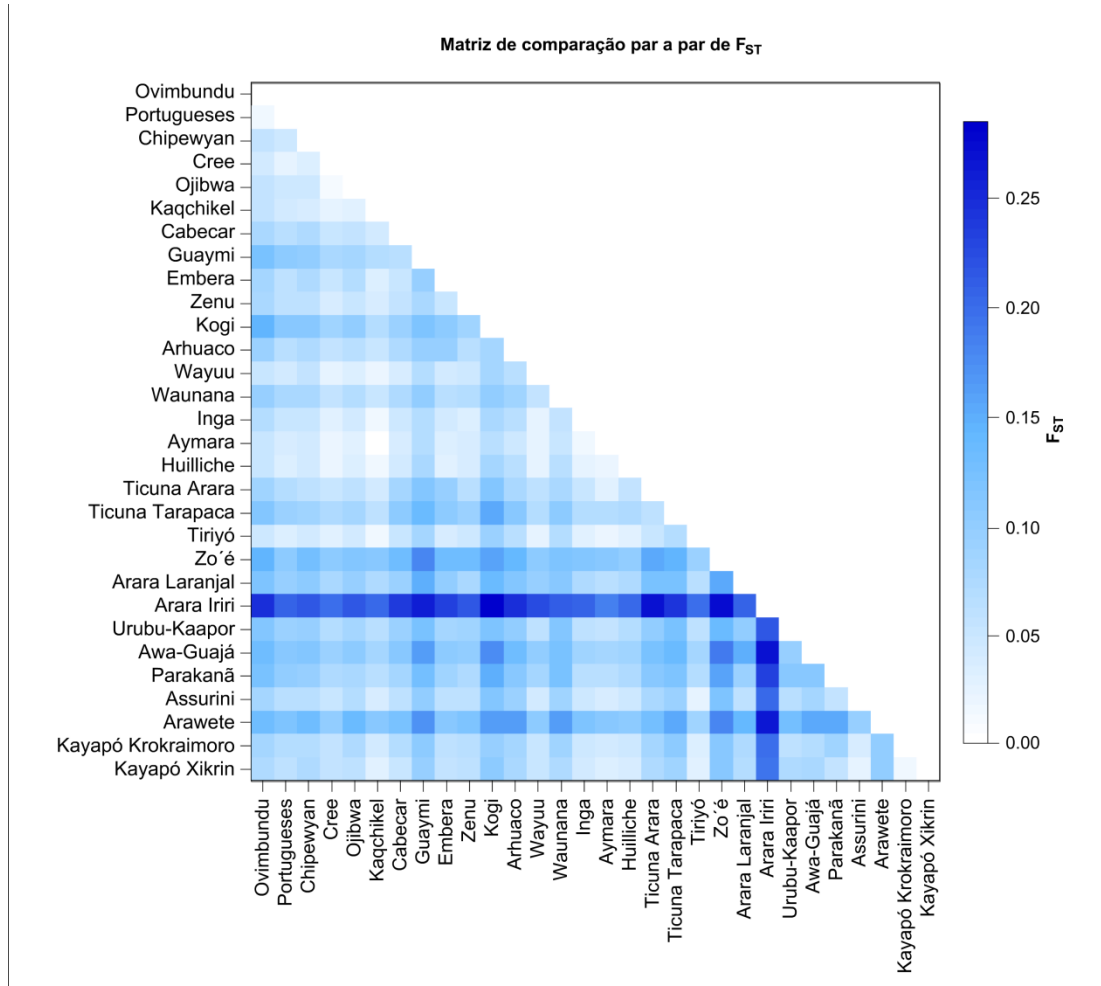


Figura 30. Diferenciação populacional par a par para os microsstatélites espalhados pelo genoma.

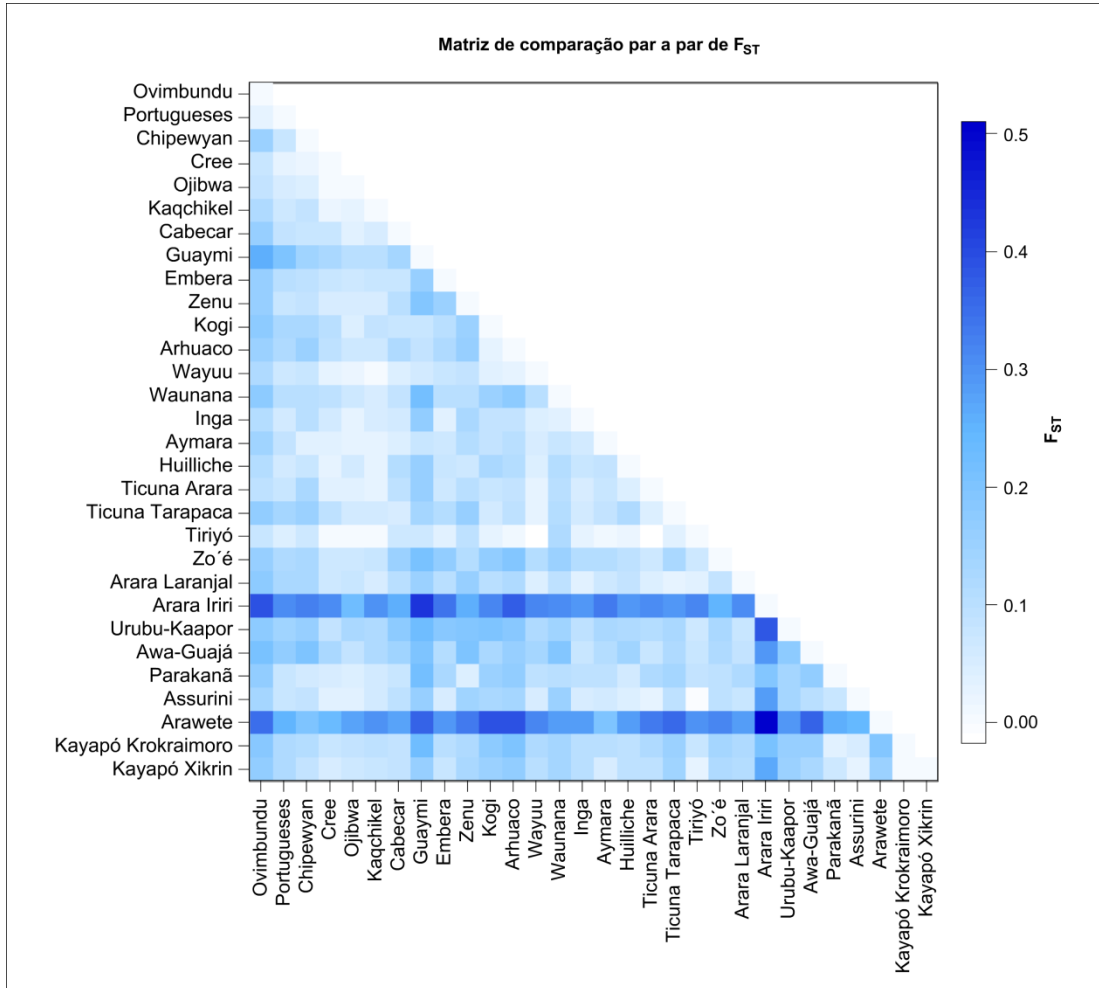


Figura 31. Diferenciação populacional par a par para os microsatélites do MHC.

Para verificar se o padrão de diferenciação populacional é maior na região do MHC realizamos uma regressão linear (figura 32). Observamos que há um deslocamento na direção dos valores mais elevados de F_{ST} para os microsatélites do MHC em relação à igualdade entre os dois subconjuntos de microsatélites (figura 32).

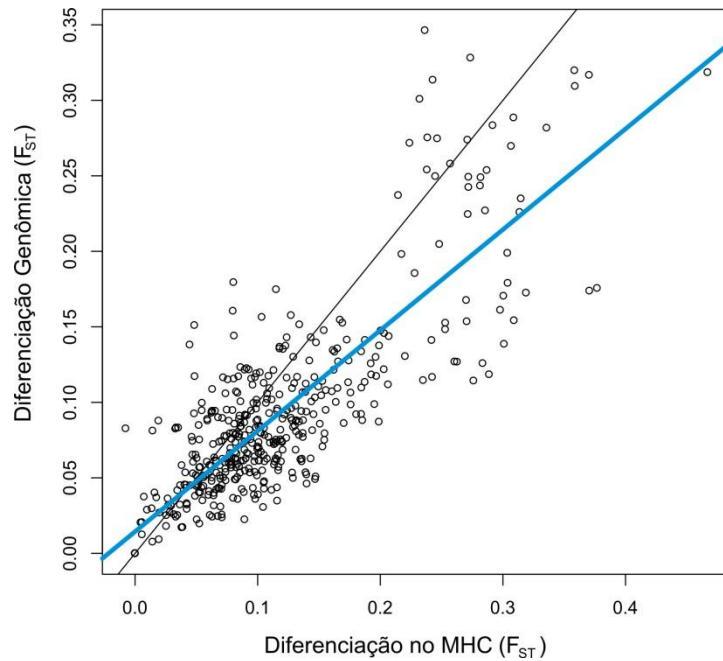


Figura 32. Regressão linear dos valores de F_{ST} par a par entre as 30 populações. Em azul a linha esperada sob a hipótese de que microssatélites do MHC e genômicos possuem os mesmos níveis de diferenciação, e em preto a linha observada para a relação entre eles (obtida usando um modelo de regressão linear).

Averiguamos se a maior diferenciação populacional observada na região do MHC estava atrelada a alguma população específica, como observada com os dados genômicos (Capítulo 1). Deste modo, verificamos qual a contribuição de cada população na diferenciação populacional da região geográfica a que pertence (tabela 20). Observamos que as populações Andinas e do Noroeste da América do Sul apresentam a mesma contribuição na diferenciação populacional. Na região leste da América do Sul, as populações Arara do Iriri e Araweté são as que individualmente mais contribuem para a diferenciação na região (sem elas o F_{ST} cai de 0,16 no leste da América do Sul para 0,1473 e 0,1346 respectivamente). Contudo, mesmo ao excluir essas populações da análise a diferenciação do MHC na região leste o continente sul americano é mais elevada que a diferenciação genômica ($F_{ST} = 0,1346$ e 0,1017 respectivamente).

Tabela 20. Comparação da contribuição de cada população na diferenciação populacional.

Região	F _{ST}	I.C 95%	valor-p
Andes	0,099	0,050 – 0,136	0,000
Sem Inga	0,084	0,038 – 0,114	0,000
Sem Aymara	0,125	0,077 – 0,177	0,000
Sem Huilliche	0,100	0,037 – 0,178	0,000
Noroeste América do Sul	0,108	0,081 – 0,131	0,000
Sem Arhuaco	0,103	0,083 – 0,125	0,000
Sem Cabecar	0,114	0,089 – 0,142	0,000
Sem Embera	0,109	0,085 – 0,136	0,000
Sem Guaymi	0,103	0,081 – 0,128	0,000
Sem Kogi	0,109	0,091 – 0,132	0,000
Sem Waunana	0,097	0,079 – 0,120	0,000
Sem Wayuu	0,117	0,093 – 0,144	0,000
Sem Zenu	0,110	0,091 – 0,131	0,000
Leste America do Sul	0,161	0,138 – 0,180	0,000
Sem Arara do Iriri	0,147	0,115 – 0,158	0,000
Sem Arara do Laranjal	0,165	0,150 – 0,181	0,000
Sem Araweté	0,134	0,114 – 0,158	0,000
Sem Assurini	0,170	0,149 – 0,193	0,000
Sem Awá-Guajá	0,164	0,147 – 0,184	0,000
Sem Kayapó K	0,170	0,153 – 0,190	0,000
Sem Kayapó Xikrin	0,170	0,152 – 0,191	0,000
Sem Parakanã	0,166	0,149 – 0,185	0,000
Sem Ticuna Arara	0,162	0,146 – 0,179	0,000
Sem Ticuna Tarapaca	0,167	0,149 – 0,187	0,000
Sem Tiriyó	0,166	0,151 – 0,184	0,000
Sem Urubu-Kaapor	0,159	0,141 – 0,181	0,000
Sem Zo'é	0,160	0,145 – 0,178	0,000

Em seguida, identificamos no gráfico da análise de regressão os pontos referentes às diferenciações par a par envolvendo as populações Arara do Iriri e Araweté (Figura 33). É possível observar que apesar delas terem as maiores diferenciações, as demais populações, mesmo menos diferenciadas, continuam apresentando um excesso de diferenciação para microssatélites do MHC, sugerindo que não eram as populações altamente diferenciadas que estavam alavancando a regressão.

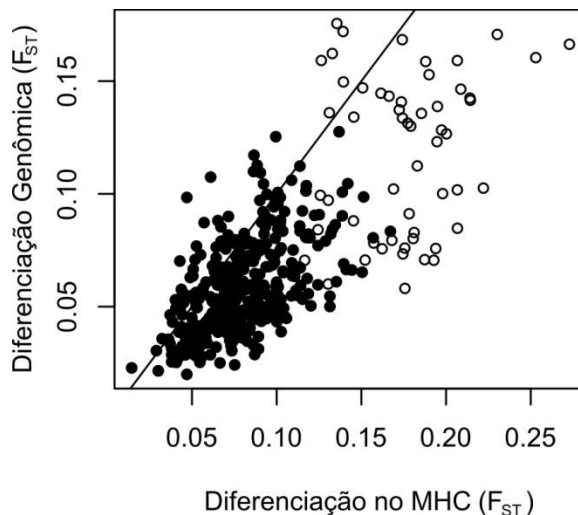


Figura 33. Regressão linear dos valores de F_{ST} par a par entre as 30 populações. Os círculos em branco representam as comparações par a par envolvendo as populações mais diferenciadas (Arara do Iriri e Araweté).

É difícil realizar um teste estatístico para a relação entre diferenciação nos microssatélites do MHC e genômicos, uma vez que os dados envolvem a comparação de pares de populações, e portanto são não-independentes, impedindo o uso de estatísticas paramétricas. Frente a esse desafio, nós usamos um método de re-amostragem para testar a hipótese de que seria esperado, simplesmente em função do tamanho finito da amostra, encontrar uma associação entre microssatélites do MHC e genômicos que apresentam diferenças tão extremas como a observada. Nós aleatorizamos os microssatélites genômicos e do MHC, mantendo a estrutura de dois grupos (um com 16 microssatélites e outro com 61 microssatélites, refletindo os mesmos tamanhos associados aos grupos de MHC e genômicos respectivamente). Repetimos essa aleatorização 1000 vezes e para cada uma calculamos a proporção de valores de F_{ST} do grupo de 16 microssatélites que eram superiores ao grupo de 61 microssatélites. O resultado está representado graficamente na figura 34 e evidencia que a associação de F_{ST} mais elevado para os microssatélites da região do MHC é mais extrema do que seria esperada ao acaso ($p < 0,001$).

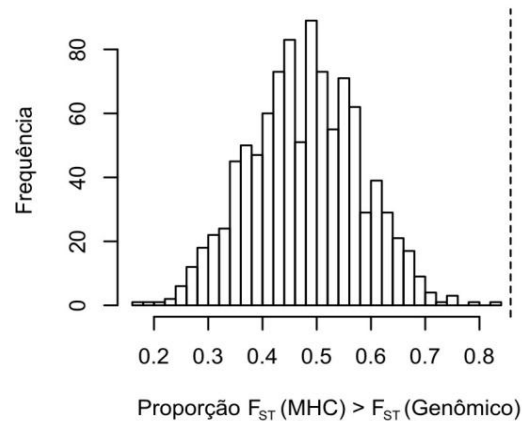


Figura 34. Teste de permutação com a proporção de valores de F_{ST} para o grupo de 16 microssatélites que são maiores do que para o grupo de 61 microssatélites.

Há relatos na literatura que a taxa de mutação pode diferir entre microssatélites que apresentam tipo de unidade de repetição distinta. Levando em consideração que os microssatélites genômicos analisados na presente tese são tetranucleotídeos e os do MHC, na sua maioria, são dinucleotídeos (ver tabela 14) utilizamos os microssatélites dinucleotídeos do banco de dados de Wang *et al.* (2007) para testar se a diferenças observadas para o índice F_{ST} poderia estar atrelada a diferenças no tipo de unidade de repetição.

Deste modo, primeiro comparamos os valores de médios de F_{ST} para di e tetranucleotídeos no banco de dados de Wang e colaboradores. O banco de dados contém 9 microssatélites do tipo dinucleotídicos e 512 tetranucleotídeos. Observamos que os valores médios de F_{ST} são similares entre eles ($0,0843 \pm 0,0164$ para o conjunto de microssatélites dinucleotídicos e $0,0805 \pm 0,0218$ para o conjunto de tetranucleotídeos).

Em seguidas realizamos a comparação das estimativas de F_{ST} par a par dos di com os tetranucleotídeos nas populações nativas da América genotipadas previamente por Wang *et al.*, (2007) (figura 35). Apesar de haver uma pequena tendência de maior diferenciação para dinucleotídeos, essa diferença é muito menos marcante do que aquela observada comparando microssatélites do MHC e genômicos: o número de pares de populações com F_{ST} maior para microssatélites do

MHC foi de 80%, em contraste com 59% para os do dinucleotídeos genômicos, quando comparados aos tetranucleotídeos genômicos. A baixa diferença no padrão de diferenciação de di e tetranucleotídeos espalhados pelo genoma indica que não é o tipo de repetição que explica o padrão observado nos contrastes entre microssatélites do MHC em relação aos genômicos.

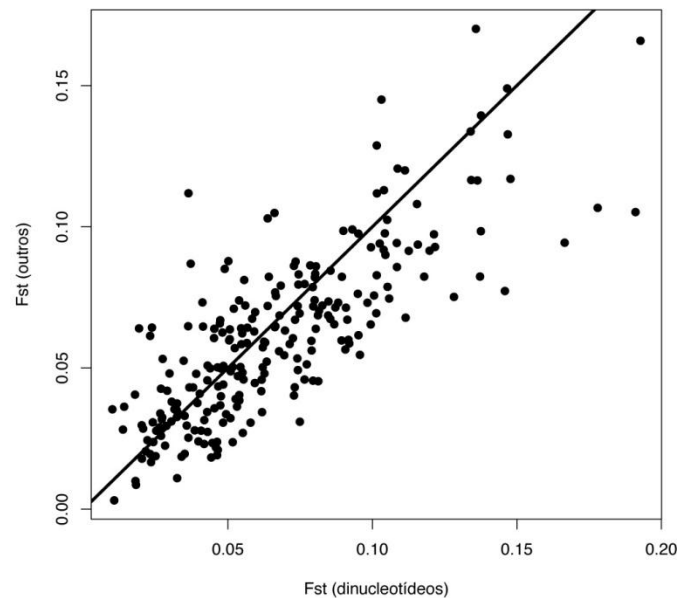


Figura 35. Regressão linear dos valores de F_{ST} par a par entre os microssatélites do tipo dinucleotídeos e tetranucleotídeos da base de dados de WANG *et al.*, (2007).

Teste de seleção na região do MHC por F_{ST} extremos

As análises de F_{ST} extremo, realizadas através do programa LOSITAN, mostram que 11 dos 16 microssatélites da região do MHC encontram-se fora da distribuição nula gerada a partir dos microssatélites genômicos para uma amostra incluindo 28 populações ameríndias, uma africana e uma europeia (figura 36). Se Africanos e Europeus são removidos, a distribuição nula se altera, e há diminuição do número de microssatélites do MHC que apresentam desvios significativos (cai de 11 para 7 microssatélites). Um efeito similar é observado em análises com conjuntos de populações menores (figuras 37 a 41).

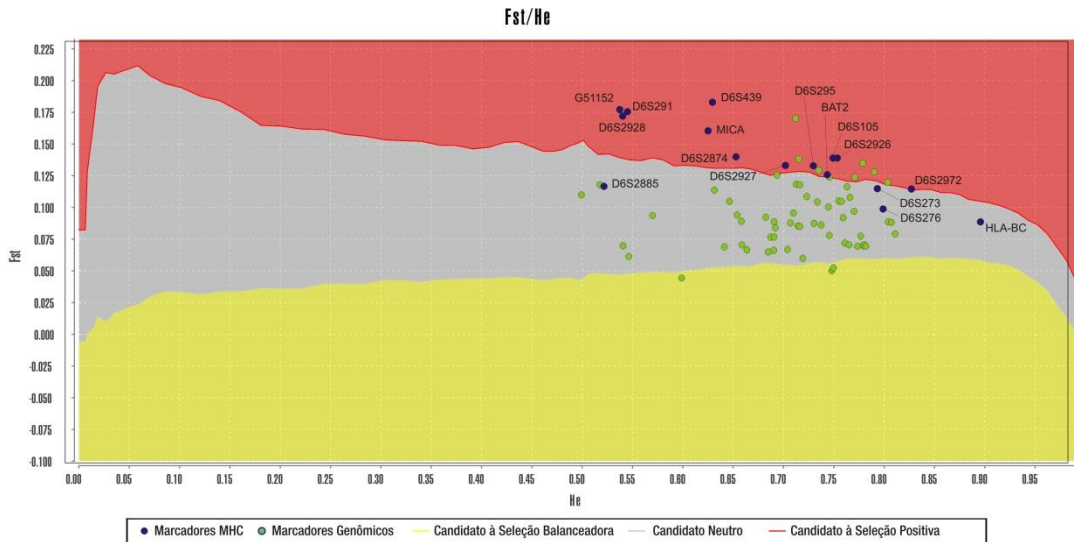


Figura 36. Comparação do F_{ST} e heterozigose (H_e) para identificação de loci candidatos à seleção, nas 30 populações analisadas no presente estudo. A região em amarelo representa candidatos à seleção balanceadora, a região cinza é a região neutra e a região vermelha candidata à seleção positiva.

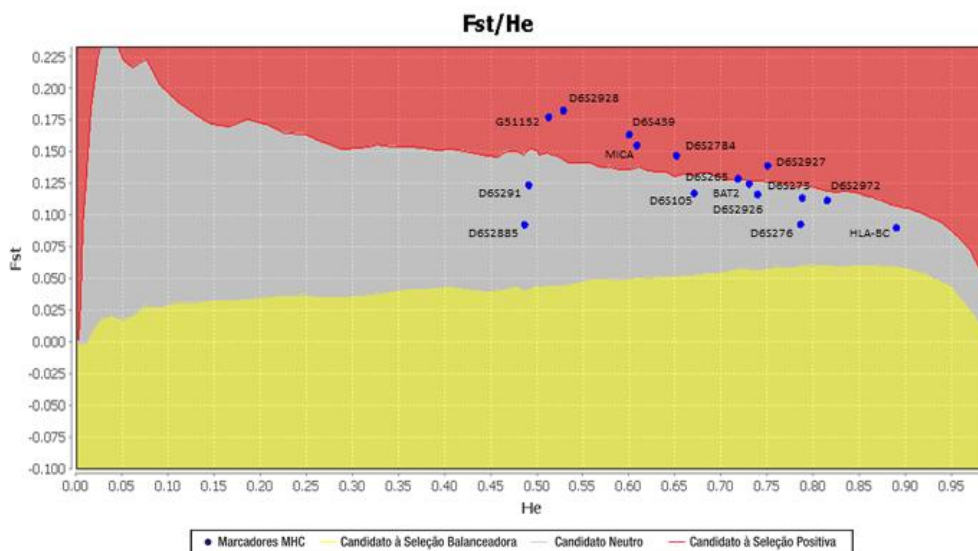


Figura 37. Comparação do F_{ST} e heterozigose (H_e) para identificação de loci candidatos à seleção, nas 28 populações nativas da América analisadas no presente estudo. A região em amarelo representa candidatos à seleção balanceadora, a região cinza é a região neutra e a região vermelha a candidata à seleção positiva.

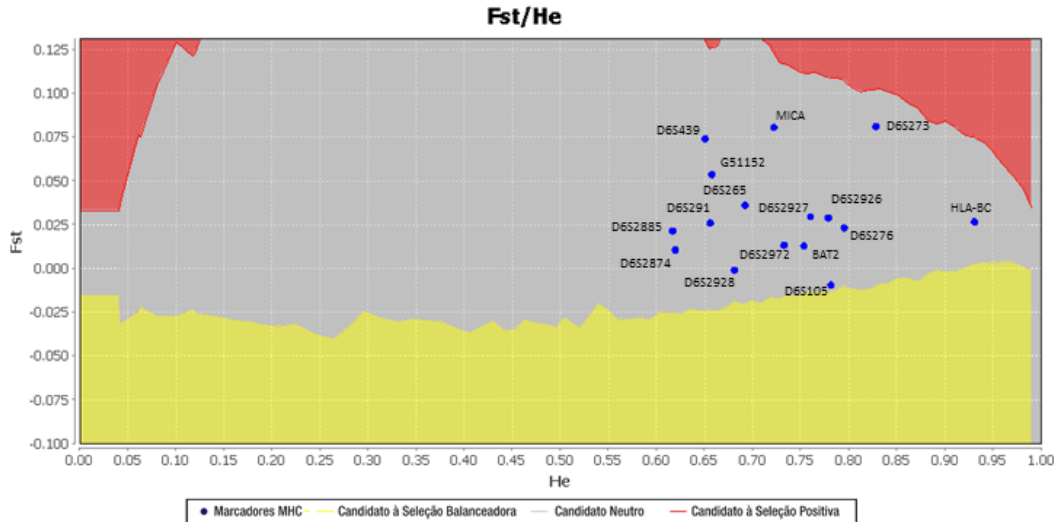


Figura 38. Comparação do FST e heterozigose (He) para identificação de loci candidatos à seleção, nas três populações da América do Norte analisadas no presente estudo. A região em amarelo representa candidatos à seleção balanceadora, a região cinza é a região neutra e a região vermelha candidata à seleção positiva.

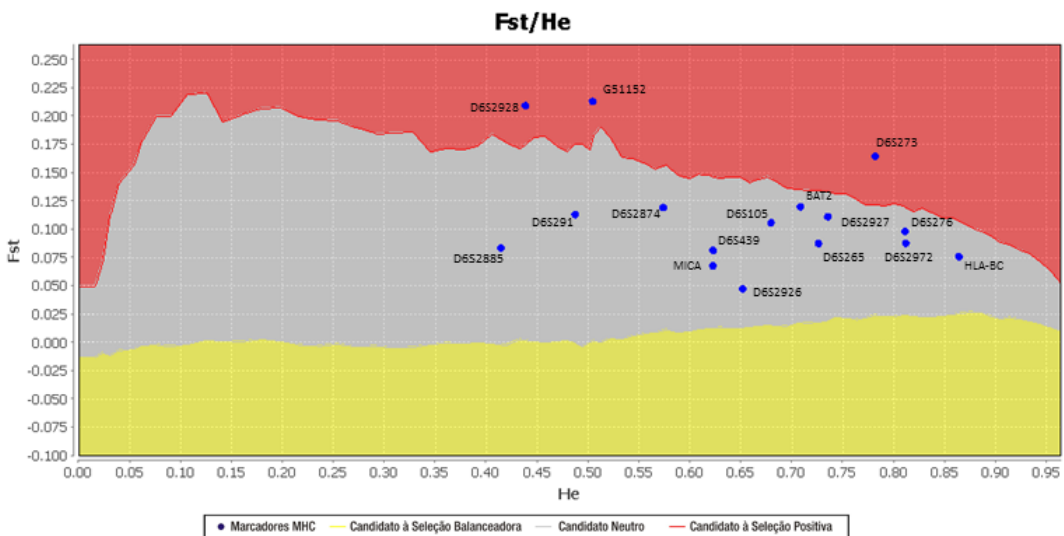


Figura 39. Comparação do F_{ST} e heterozigose (He) para identificação de *loci* candidatos à seleção, nas seis populações do Noroeste da América do Sul analisadas no presente estudo. A região em amarelo representa candidatos à seleção balanceadora, a região cinza é a região neutra e a região vermelha candidata à seleção positiva.

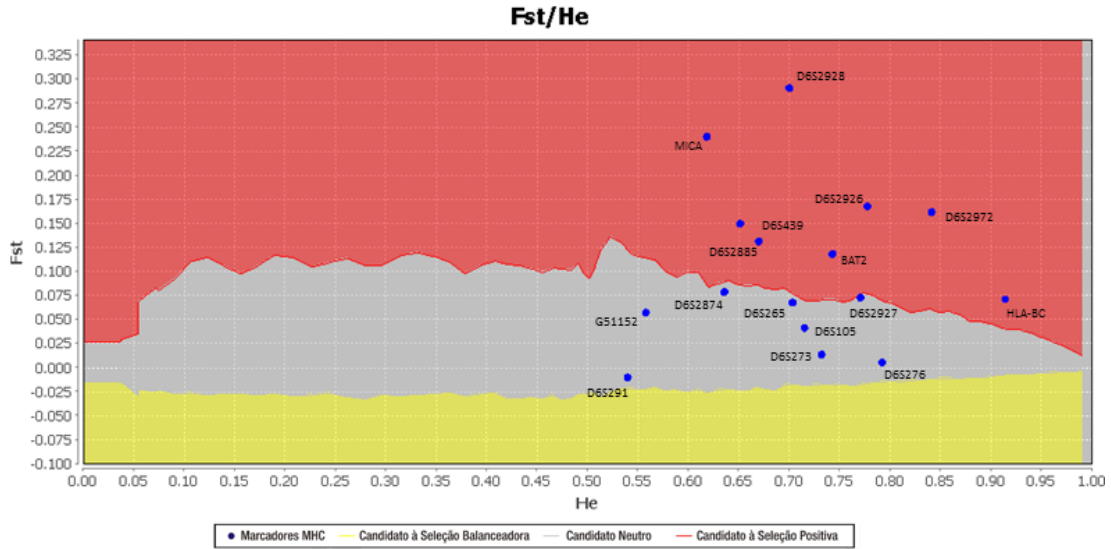


Figura 40. Comparação do F_{ST} e heterozigose (He) para identificação de *loci* candidatos à seleção, nas três populações Andinas analisadas no presente estudo. A região em amarelo representa candidatos à seleção balanceadora, a região cinza é a região neutra e a região vermelha candidata à seleção positiva.

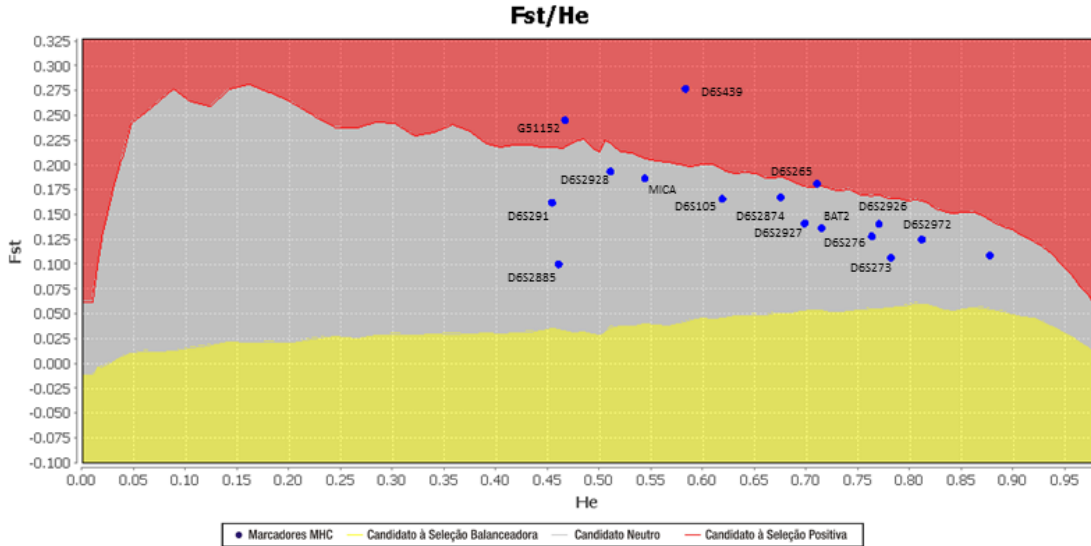
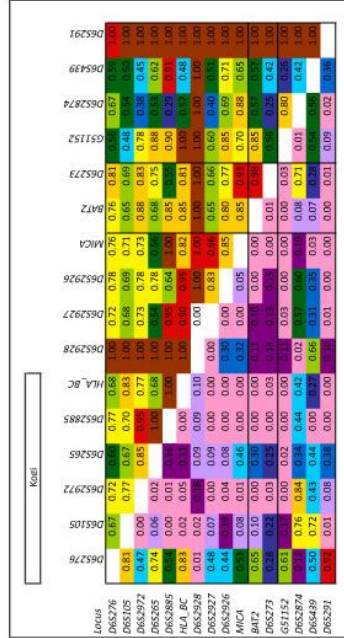
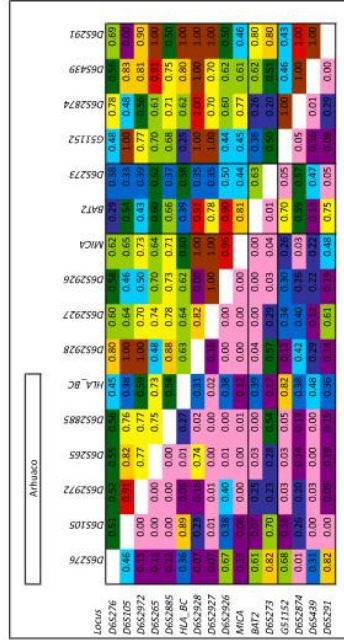
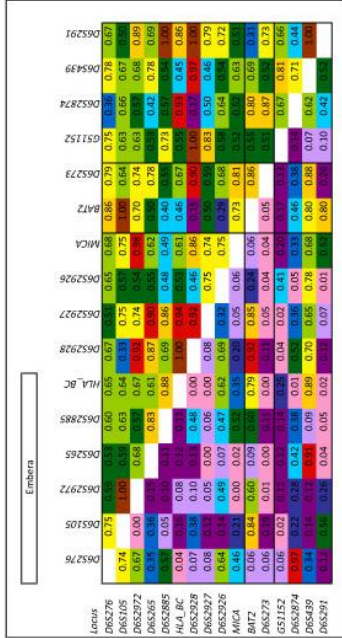
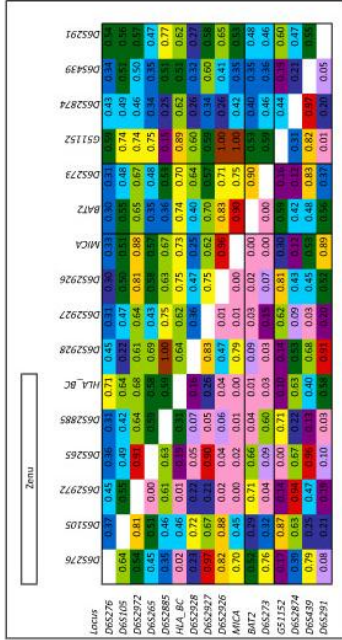
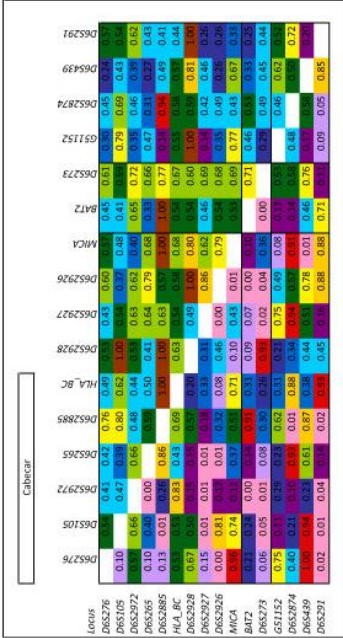
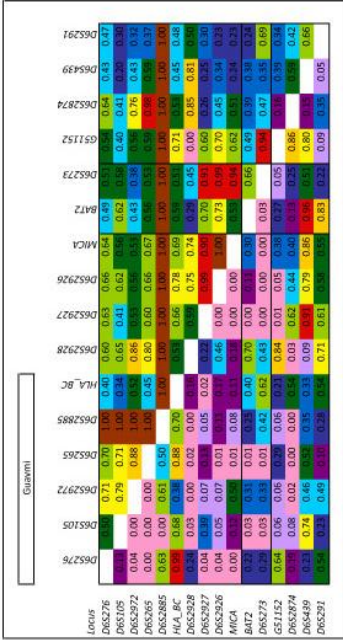
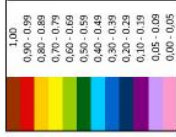


Figura 41. Comparação do F_{ST} e heterozigose (He) para identificação de *loci* candidatos à seleção, nas 13 populações do Leste da América do Sul analisadas no presente estudo. A região em amarelo representa candidatos à seleção balanceadora, a região cinza é a região neutra e a região vermelha candidata à seleção positiva.

Desequilíbrio de ligação

Na figura 42 mostramos que os microssatélites na região do MHC não estão associados aleatoriamente e que cada população apresenta um padrão próprio de desequilíbrio de ligação. Também observamos que há um aumento no desequilíbrio de ligação entre *loci* que estão localizados fisicamente mais próximos e esse diminui à medida que os *loci* distanciam (figura 43). Além disso, o desequilíbrio de ligação é maior nas populações do leste e noroeste da América do Sul e reduz gradativamente na direção dos Andes, América do Norte, Portugueses e Ovimbundu (figura 43).

B.



C.

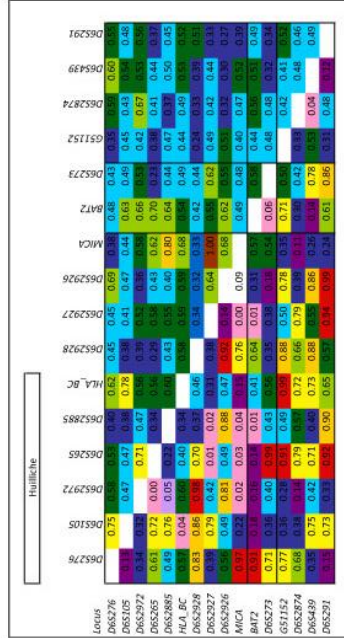
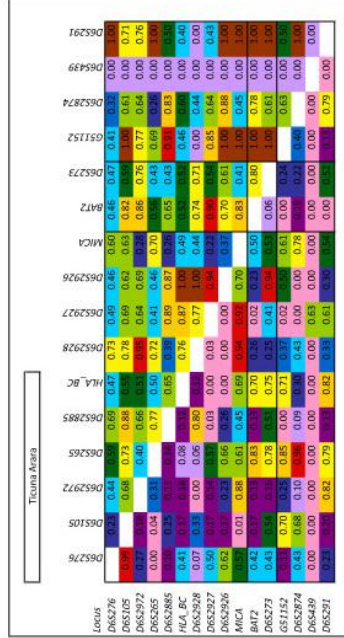
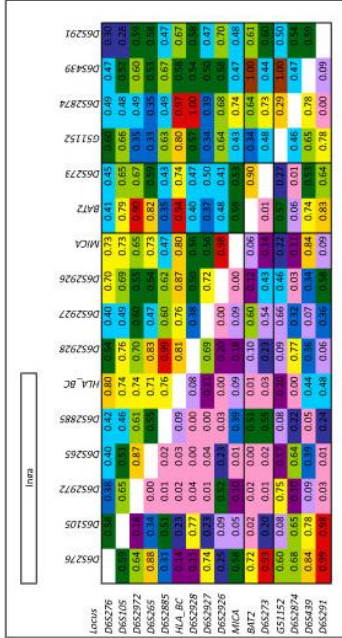
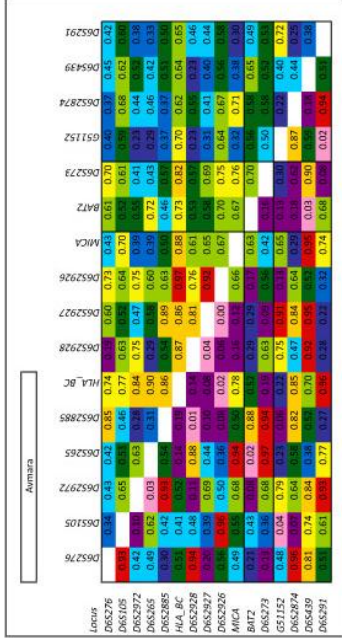
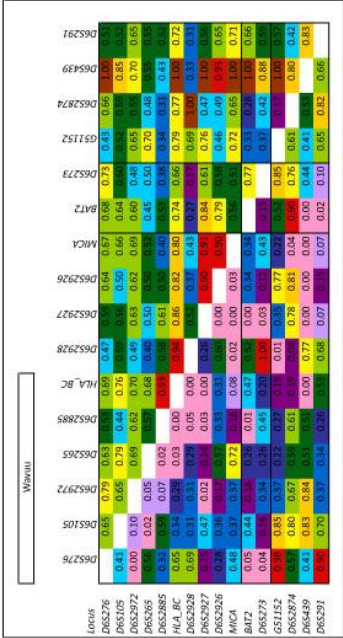
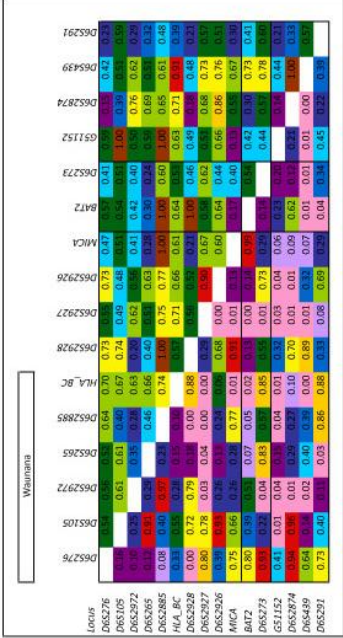
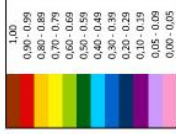


Figura 42. Padrão de desequilíbrio de ligação para os microssatélites restritos à região do MHC em cada uma das populações analisadas no presente estudo. A diagonal superior refere-se ao valor de D' , a diagonal inferior se refere ao valor de p . As linhas verticais e horizontais separam os *loci* das classes I, III e II do MHC.

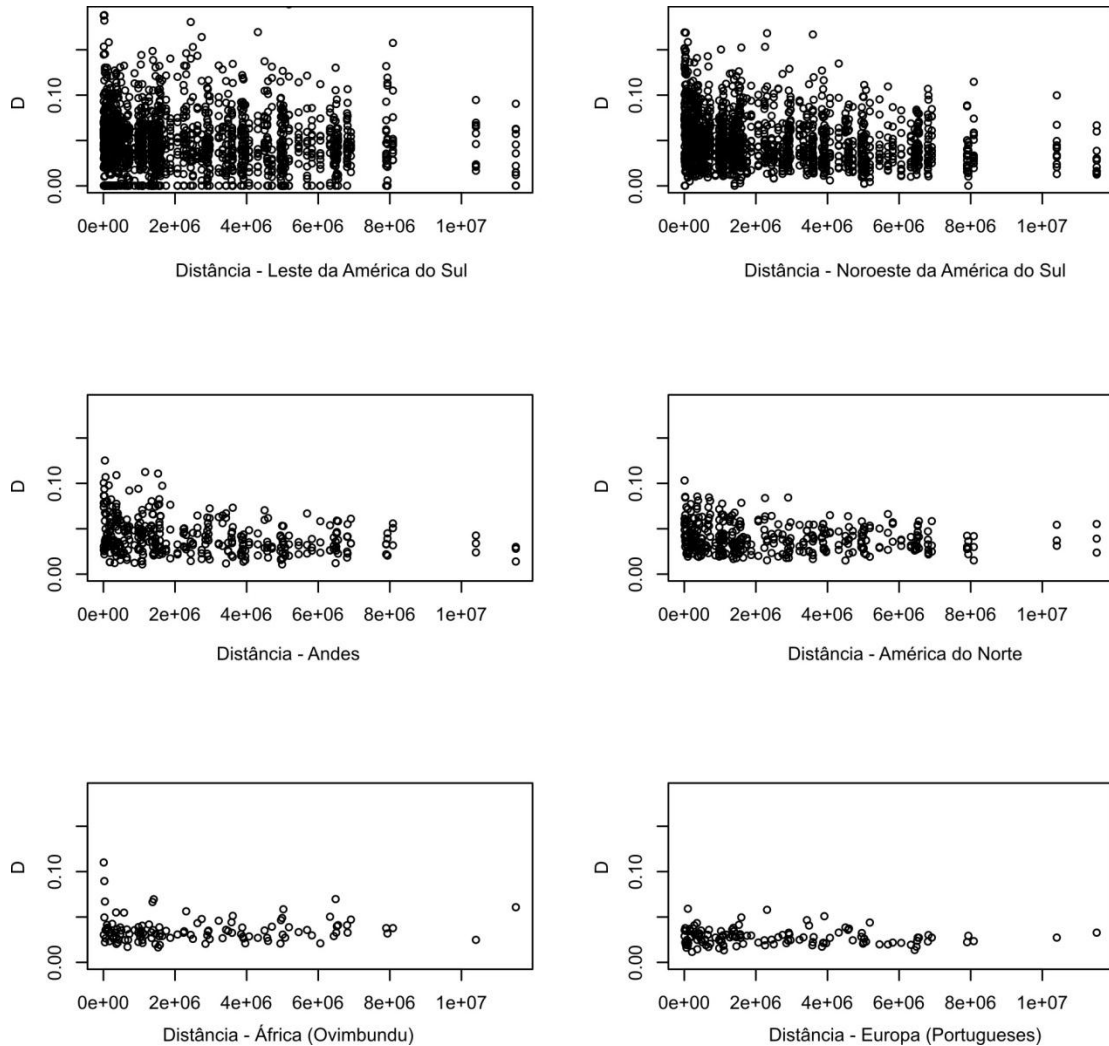


Figura 43. Desequilíbrio de ligação por distância física (em pares de bases) dos microssatélites em cada região geográfica. No eixo x o valor do desequilíbrio no eixo y da distancia entre os *loci*.

Desequilíbrio de ligação entre HLA-B e microssatélites adjacentes

Como verificamos no item anterior, os microssatélites do MHC estão em desequilíbrio de ligação. Devido ao fato que as análises incluindo os genótipos para o gene *HLA-B* são constituídas por um subconjunto dos dados analisados até o momento (ver item tabela 16), refizemos a análise de desequilíbrio de ligação.

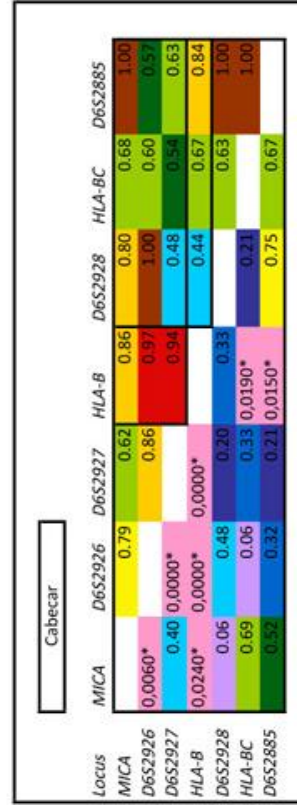
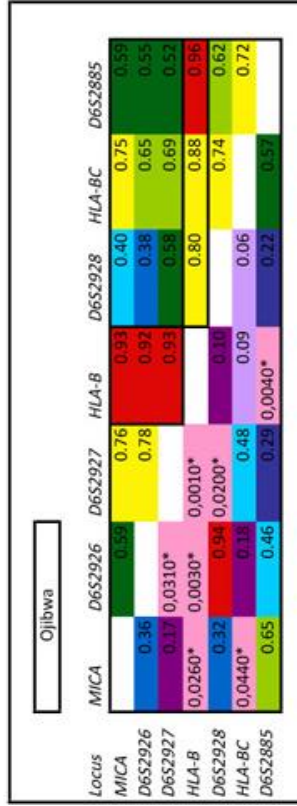
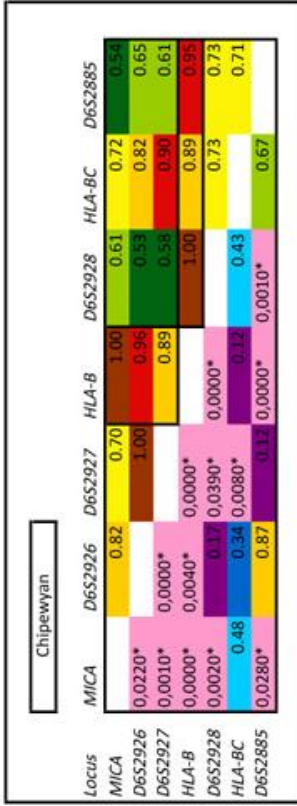
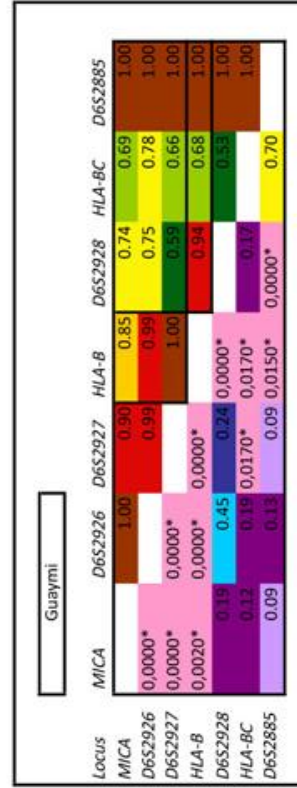
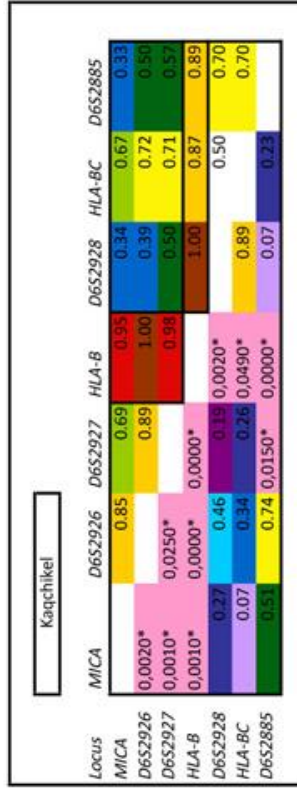
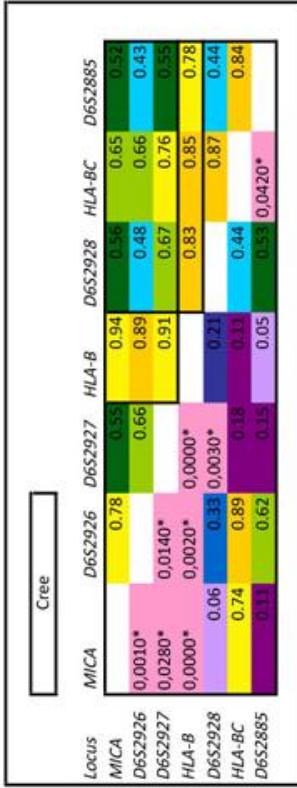
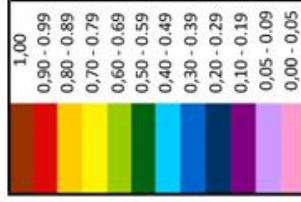
O desequilíbrio de ligação entre o gene *HLA-B* e os microssatélites adjacentes (*MICA*, *D6S2926*, *D6S2927*, *D6S2928*, *HLA-BC* e *D6S2885*) em cada uma das 28 populações ameríndias pode ser observado na figura 44. Apesar das peculiaridades de cada população, o gene *HLA-B* apresenta alto nível de desequilíbrio de ligação (D'entre 1,00 e 0,39) com os demais *loci* de microssatélites adjacentes.

Heterozigose específica dos alelos de microssatélites em função do alelo *HLA-B*

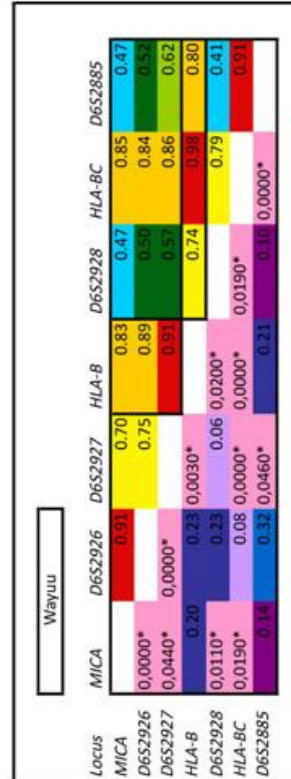
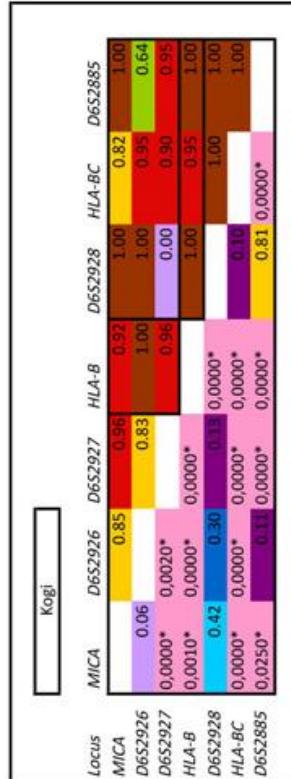
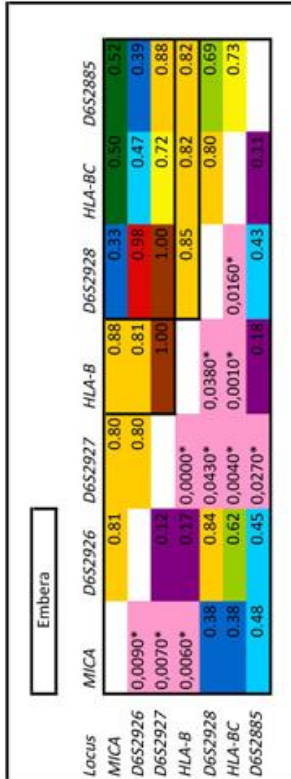
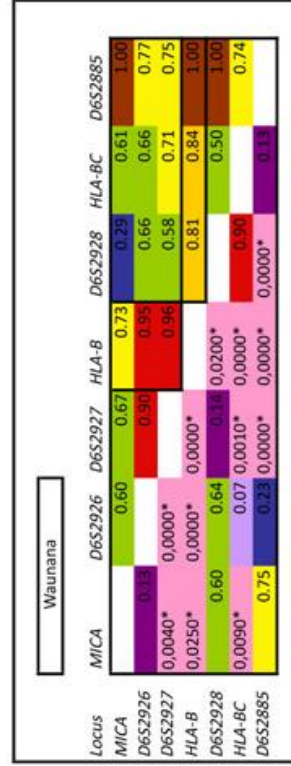
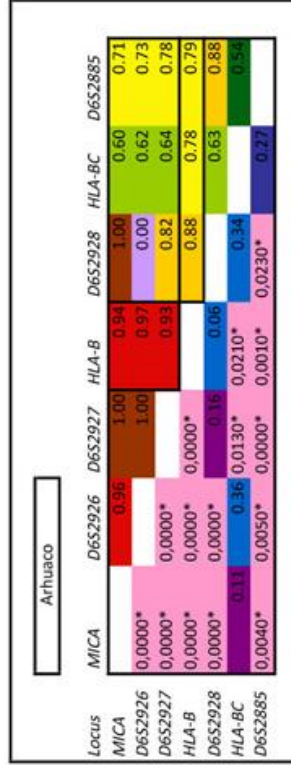
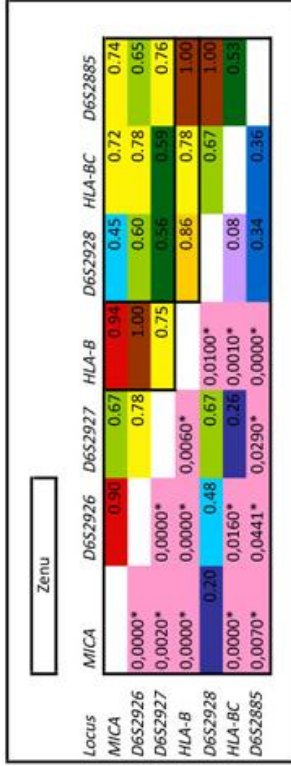
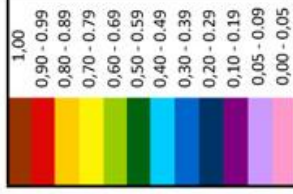
Usando a classificação estabelecida por FRANCISCO (2009), separamos os alelos de *HLA-B* mais comuns da amostra (acima de 5 cópias) em dois grupos: a) endêmicos/alta frequência nas Américas e baixa frequência nas demais regiões do Mundo, e b) Cosmopolitas/Alta frequência em outras regiões do mundo.

Em seguida estimamos a heterozigose alelo específica (HSH) dos microssatélites associados aos alelos de *HLA-B* em toda a amostragem, sem levar em consideração em que população eles ocorrem. De modo geral, é possível observar que a heterozigose observada nos microssatélites que flanqueiam alelos específicos de *HLA-B* é menor que a heterozigose total de cada *locus* (tabela 21). Contudo não observamos diferenças entre a HSH dos alelos endêmicos e cosmopolitas (tabela 21).

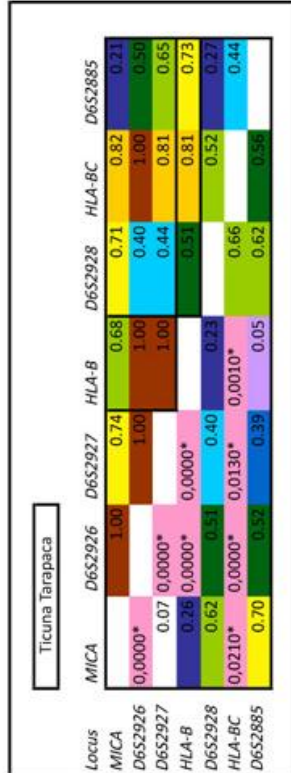
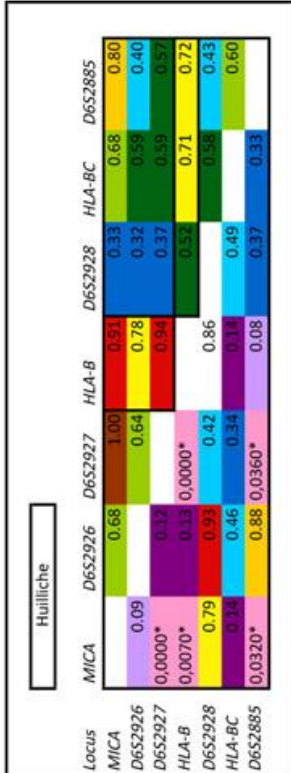
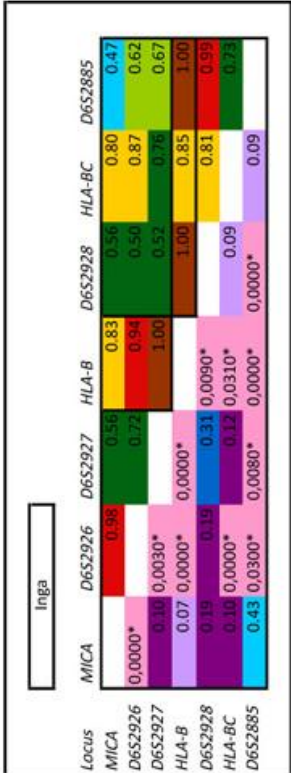
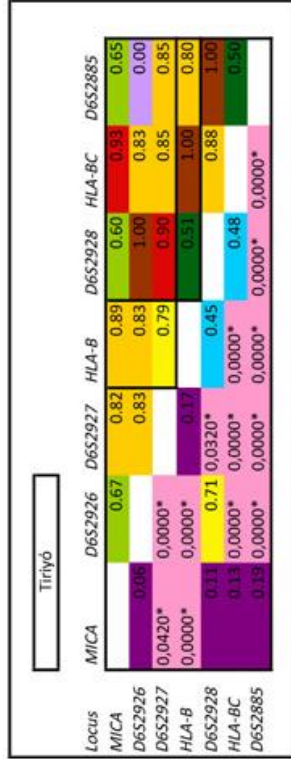
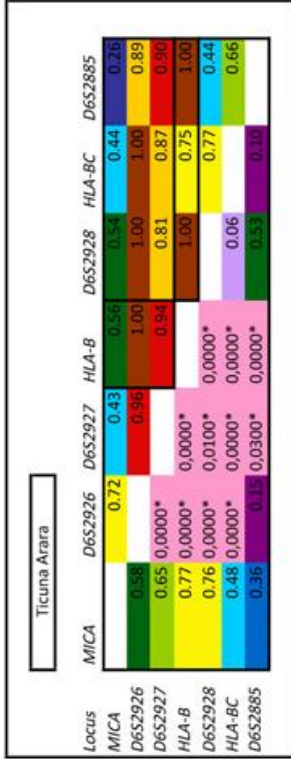
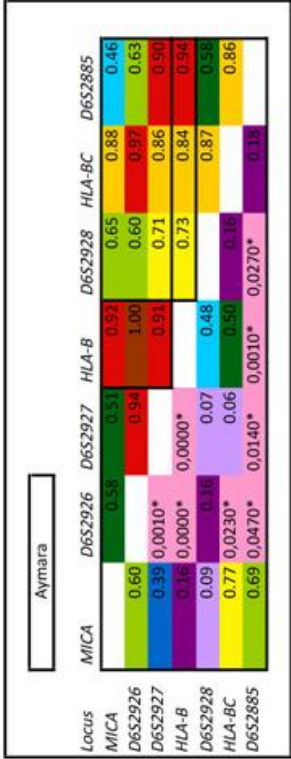
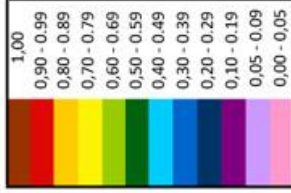
A.

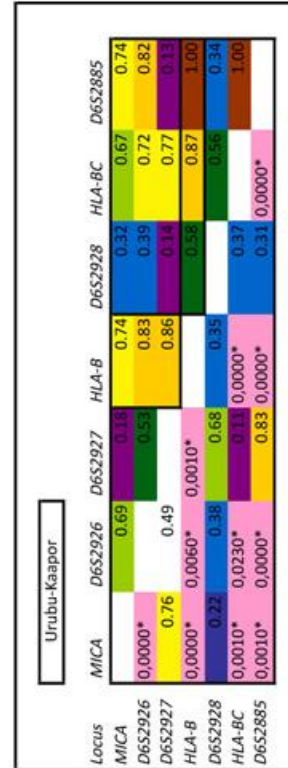
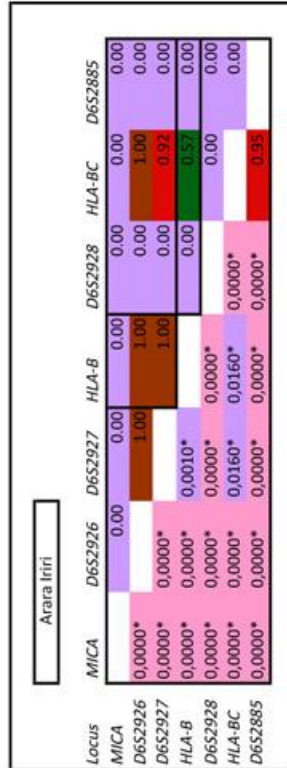
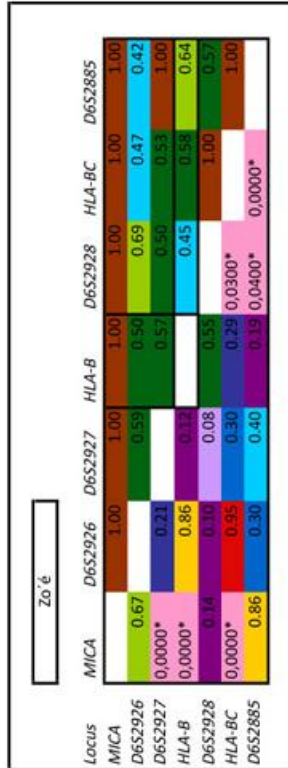
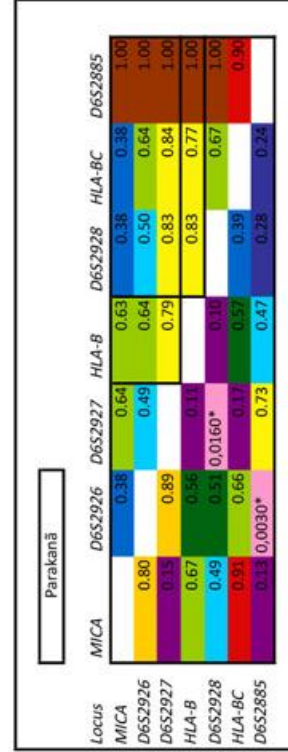
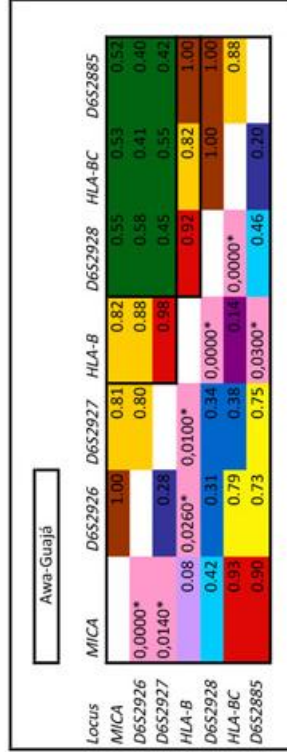
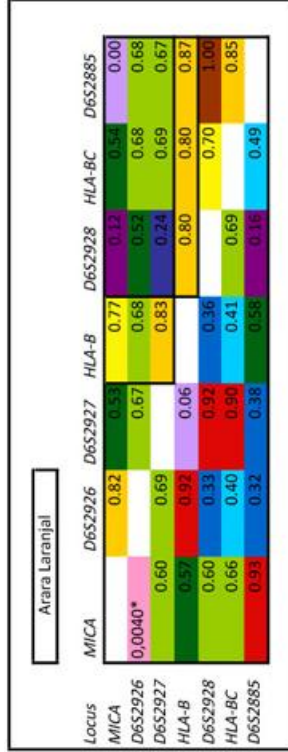
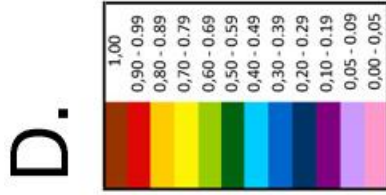


B.



C.





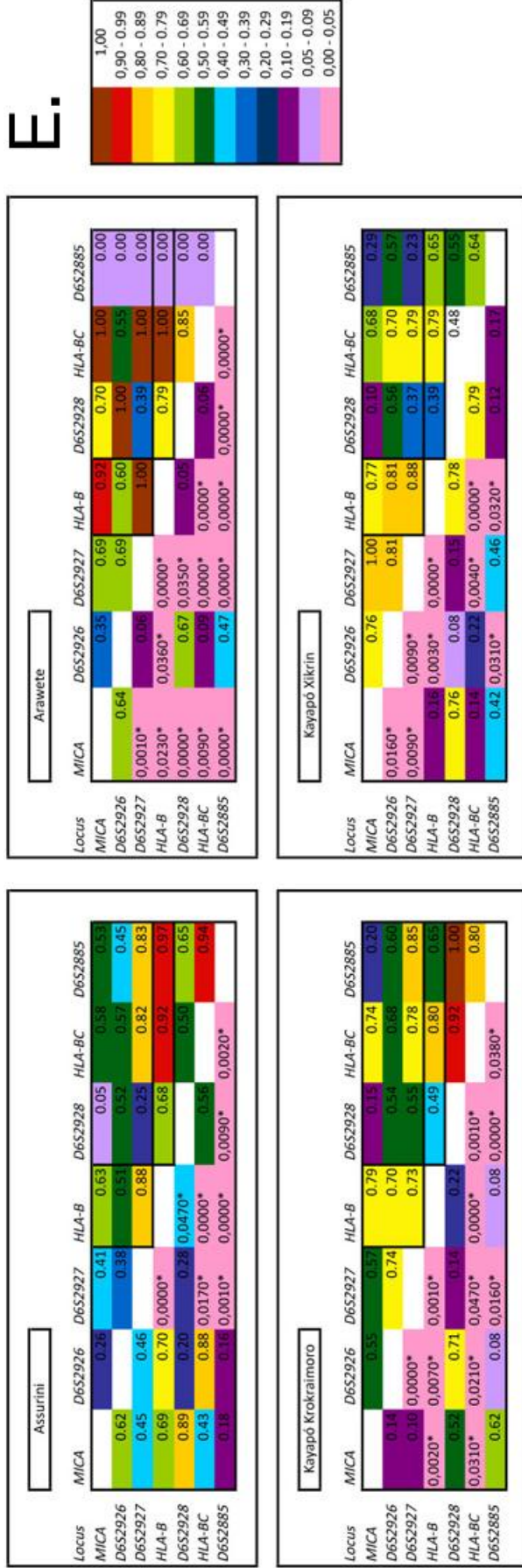


Figura 44. Desequilíbrio de ligação entre os microsatélites *MICA*, *D6S2926*, *D6S2927*, *D6S2928*, *HLA-BC* e *D6S2885* com o gene *HLA-B*. em cada uma das 30 populações analisadas no presente estudo. A diagonal superior corresponde aos valores de *D'* e a diagonal inferior refere-se a valores de *p*.

Tabela 21. Heterozigose específica para os alelos de microssatélites em função dos alelos do gene *HLA-B*. Onde N é o número de cópias dos alelos na amostra.

Alelos <i>HLA-B</i>	N	<i>MICA</i> (N alelos)	<i>D6S2926</i> (N alelos)	<i>D6S2927</i> (N alelos)	<i>D6S2928</i> (N alelos)	<i>HLA-BC</i> (N alelos)	<i>D6S2885</i> (N alelos)
Endêmicos/ Alta frequência nas Américas e baixa frequência nas demais regiões do Mundo							
*1504	50	0,29 (2)	0,54 (6)	0,08 (3)	0,27(2)	0,50(7)	0,07(2)
*350401	30	0,45 (4)	0,34(3)	0,18(4)	0,35(2)	0,74(7)	0,12(3)
*350501	37	0,19(2)	0,24(3)	0,20(4)	0,39(2)	0,57(8)	0,10(2)
*3543	57	0,03(2)	0 (1)	0,03 (2)	0,07 (2)	0,66 (7)	0,16 (3)
*3549	09	0,37 (2)	0,20 (2)	0,20 (2)	0 (1)	0,49 (3)	0 (1)
*3903	10	0,54(3)	0,32(2)	0,42(2)	0,48(2)	0,46 (3)	0,48 (2)
*390501	37	0,42 (3)	0,35 (3)	0,15 (3)	0,27 (2)	0,66 (8)	0,62 (5)
*3909	22	0 (1)	0,08 (2)	0 (1)	0,61 (4)	0,65 (5)	0,47 (3)
*4004	22	0,16 (2)	0,71 (4)	0,31 (3)	0,45 (3)	0,75 (8)	0,17 (3)
*4802	13	0,50 (2)	0,56 (4)	0 (1)	0,35 (2)	0,39 (5)	0,14 (2)
*520102	27	0,32 (3)	0,65 (5)	0,37(4)	0,42 (2)	0,75 (8)	0,43 (3)
Cosmopolitas/Alta frequência em outras regiões do mundo							
*150101	25	0,29 (3)	0,34 (4)	0,34 (4)	0,32 (2)	0,77 (6)	0,27 (2)
*350101	52	0,18 (4)	0,24 (4)	0,14 (4)	0,38 (3)	0,77 (10)	0,21 (4)
*390101	7	0 (1)	0,24 (2)	0,24 (2)	0 (1)	0,61 (3)	0,69 (4)
*400201	11	0,33 (4)	0,22 (6)	0,14 (3)	0,31 (5)	0,72 (13)	0,03 (3)
*440201	8	0,37 (2)	0,71 (4)	0,37 (2)	0,37 (2)	0,81 (6)	0 (1)
*480101	9	0,49 (2)	0,34 (2)	0 (1)	0,72 (4)	0,83 (7)	0,44 (2)
*510101	61	0,58 (3)	0,57 (6)	0,21 (5)	0,53 (4)	0,82 (12)	0,55 (6)
Heteroz total		0,61	0,72	0,73	0,51	0,88	0,43

Diferenciação inter e intra-linhagem

Na análise de diferenciação, com base nos microssatélites flanqueadores de *HLA-B* é possível constatar que existe grande diferença entre as linhagens e uma diminuição da diferenciação entre alelos de microssatélites dentro de uma mesma linhagem (tabela 22). Isso poderia ser um indicativo de que, um conjunto específico de alelos de microssatélites poderia estar associado a cada linhagem. Também é possível observar que a diferenciação varia ao longo dos *loci* dentro de cada linhagem. Assim, os *loci* de microssatélites com alto grau de diferenciação dentro das linhagens podem indicar que o conjunto de alelos daquele *locus* é diferente entre os alelos de *HLA-B*.

Tabela 22. Índice de diferenciação (F_{ST}) por locus de microssatélite intra e inter-linhagens de *HLA-B*.

Linhagem <i>Loci</i>	*15	*35	*39	*40	*44	*51	Linhagens
<i>MICA</i>	0,0545	0,0689	0,1434	0,0038	0,4397	0,1771	0,5213
<i>D6S2926</i>	0,1621	0,0300	0,1299	0,0527	0,4149	0,0680	0,5614
<i>D6S2927</i>	0,3964	0,0269	0,0868	0,0454	0,0204	0,1835	0,7403
<i>D6S2928</i>	0,4407	0,4936	0,2392	0,0375	0,0667	0,1091	0,1834
<i>HLA-BC</i>	0,2747	0,1130	0,0726	0,0173	0,0597	0,0666	0,1338
<i>D6S2885</i>	0,0702	0,0150	0,2043	0,0509	0,1864	0,0932	0,4364

Associação de alelos de microssatélites com linhagens e alelos de *HLA-B*

Investigamos a frequência com que um determinado alelo de microssatélite (o mais frequente dentro de cada linhagem) é encontrado em uma determinada linhagem de *HLA-B* (tabela 23). Apesar de linhagens distintas compartilharem alelos mais frequentes de determinados *loci*, de modo geral, cada linhagem está associada a um conjunto característico de alelos dos microssatélites adjacentes (haplótipo formado pelos microssatélites adjacente), corroborando a hipótese apresentada no item anterior. Também é possível observar que os alelos dos microssatélites localizados entre *HLA-B* e o gene *MICA* caracterizam melhor as linhagens que os alelos dos microssatélites situados entre *HLA-B* e *HLA-C* (tabela 23).

Em seguida, verificamos como os alelos de microssatélites estão associados aos alelos de *HLA-B*. Observamos uma alta associação entre os alelos de microssatélite e os alelos de *HLA-B* (42 a 100%) (tabela 24).

Tabela 23. Associação do alelo mais comum do microssatélite com as linhagens de *HLA-B*.

Linhagem	N	Microssatélite	Alelo	N	Associação
*15	139	MICA	184	119	85,61%
		D6S2926	257	87	62,89%
		D6S2927	232	113	81,29%
		MICA:D6S2926:D6S2927	184:257:232	75	53,95%
		D6S2928	257	80	57,55%
		D6S2885	100	126	90,65%
		D6S2928: D6S2885	257:100	69	49,64%
*35	234	MICA	196	206	88,03%
		D6S2926	265	212	90,60%
		D6S2927	244	213	91,02%
		MICA:D6S2926:D6S2927	196:265:244	179	76,49%
		D6S2928	261	118	50,43%
		D6S2885	100	210	89,74%
		D6S2928: D6S2885	261:100	112	47,86%
*39	101	MICA	196	76	75,25%
		D6S2926	265	81	80,20%
		D6S2927	240	90	89,11%
		MICA:D6S2926:D6S2927	196:265:240	65	64,35%
		D6S2928	257	70	69,31%
		D6S2885	104	51	50,49%
		D6S2928: D6S2885	257:104	39	38,61%
*40	144	MICA	184	116	80,65%
		D6S2926	273	122	84,72%
		D6S2927	228	130	90,03%
		MICA:D6S2926:D6S2927	184:273:228	95	65,97%
		D6S2928	257	114	78,08%
		D6S2885	100	140	95,89%
		D6S2928: D6S2885	257:100	111	76,03%
*44	8	MICA	184	6	75%
		D6S2926	261	3	37,5%
		D6S2927	232	6	75%
		MICA:D6S2926:D6S2927	184:261:232	3	37,5%
		D6S2928	261	6	75%
		D6S2885	100	8	100%
		D6S2928: D6S2885	257:100	6	75%

Continua

Tabela 23 (continuação). Associação do alelo mais comum do microssatélite com as linhagens de *HLA-B*.

Linhagem	N	Microssatélite	Alelo	N	Associação
*48	28	MICA	184	15	53,57%
		D6S2926	273	20	71,43%
		D6S2927	232	27	96,43%
		MICA:D6S2926:D6S2927	184:273:232	9	32,14%
		D6S2928	261	17	60,71%
		D6S2885	100	19	67,85%
		D6S2928: D6S2885	261:100	16	57,14%
*51	72	MICA	187	42	58,33%
		D6S2926	269	46	63,89%
		D6S2927	232	63	87,5%
		MICA:D6S2926:D6S2927	187:269:232	29	40,28%
		D6S2928	257	48	66,67%
		D6S2885	104	44	61,11%
		D6S2928:D6S2885	257:104	35	48,61%
*52	27	MICA	184	22	81,48%
		D6S2926	269	15	55,56%
		D6S2927	232	21	77,78%
		MICA:D6S2926:D6S2927	184:269:232	10	37,04%
		D6S2928	257	19	70,37%
		D6S2885	104	19	70,37%
		D6S2928: D6S2885	257:104	16	59,26%

Tabela 24. Associação dos alelos mais frequentes dos microssatélites com os alelos de *HLA-B*.

Alelo <i>HLA-B</i>	N	Microssatélite	Alelo	N	Associação
B*1504	50	MICA	184	41	82,00%
		D6S2926	257	32	64,00%
		D6S2927	232	48	96,00%
		MICA:D6S2926:D6S2927	184:257:232	28	56,00%
		D6S2928	261	34	82,00%
		D6S2885	100	47	94,00%
		D6S2928: D6S2885	261:100	32	64,00%
B*150501	9	MICA	184	9	100%
		D6S2926	273	8	88,89%
		D6S2927	228	9	100%
		MICA:D6S2926:D6S2927	184:273:288	8	88,89%
		D6S2928	257	9	100%
		D6S2885	100	9	100%
		D6S2928: D6S2885	257:100	9	100%
B*1508	19	MICA	184	14	73,68%
		D6S2926	257	10	52,63%
		D6S2927	232	16	84,21%
		MICA:D6S2926:D6S2927	184:257:232	8	42,10%
		D6S2928	257	13	68,42%
		D6S2885	100	14	73,68%
		D6S2928: D6S2885	257:100	8	42,10%
B*1520	28	MICA	184	28	100%
		D6S2926	257	18	64,28%
		D6S2927	232	22	78,57%
		MICA:D6S2926:D6S2927	184:257:232	18	64,28%
		D6S2928	257	23	82,14%
		D6S2885	100	28	100%
		D6S2928: D6S2885	257:100	23	82,14%
B*150101	25	MICA	184	21	84,00%
		D6S2926	257	20	80,00%
		D6S2927	232	20	80,00%
		MICA:D6S2926:D6S2927	184:257:232	15	60,00%
		D6S2928	257	20	80,00%
		D6S2885	100	21	84,00%
		D6S2928: D6S2885	257:100	17	68,00%
B*3504	30	MICA	196	21	70%
		D6S2926	265	24	80%
		D6S2927	244	27	90%
		MICA:D6S2926:D6S2927	196:265:244	15	50%
		D6S2928	261	23	76,67%
		D6S2885	100	28	93,34%
		D6S2928: D6S2885	261:100	22	73,34%

continua

Tabela 24 (continuação). Associação dos alelos mais frequentes dos microsstatélites com os alelos de *HLA-B*.

B*350501	35	MICA	196	31	88,57%
		D6S2926	265	30	85,71%
		D6S2927	244	31	88,57%
		MICA:D6S2926:D6S2927	196:265:244	24	68,57%
		D6S2928	261	27	77,14%
		D6S2885	100	34	97,14%
		D6S2928: D6S2885	261:100	27	77,14%
B*3510	10	MICA	196	10	100%
		D6S2926	265	10	100%
		D6S2927	244	10	100%
		MICA:D6S2926:D6S2927	196:265:244	10	100%
		D6S2928	261	9	90%%
		D6S2885	100	10	100%
		D6S2928: D6S2885	261:100	9	90%
B*350101G	52	MICA	196	47	90,38%
		D6S2926	265	45	86,54%
		D6S2927	244	48	92,31%
		MICA:D6S2926:D6S2927	196:265:244	41	78,84%
		D6S2928	261	39	75%
		D6S2885	100	46	88,46%
		D6S2928: D6S2885	261:100	37	71,15%
B*354301G	57	MICA	196	56	98,24%
		D6S2926	265	57	100%
		D6S2927	244	56	98,24%
		MICA:D6S2926:D6S2927	196:265:244	55	96,49%
		D6S2928	257	55	96,49
		D6S2885	100	52	91,23%
		D6S2928: D6S2885	257:100	50	87,72%
B*390101	7	MICA	196	7	100%
		D6S2926	265	6	85,71%
		D6S2927	240	6	85,71%
		MICA:D6S2926:D6S2927	196:265:240	5	71,43%
		D6S2928	257	7	100%
		D6S2885	114	3	42,86%
		D6S2928: D6S2885	257:114	3	42,86%
3903	10	MICA	196	6	60%
		D6S2926	265	8	80%
		D6S2927	240	7	70%
		MICA:D6S2926:D6S2927	196:265:240	6	60%
		D6S2928	261	6	60%
		D6S2885	104	6	60%
		D6S2928: D6S2885	261:104	2	20%

continua

Tabela 24 (continuação). Associação dos alelos mais frequentes dos microssatélites com os alelos de *HLA-B*.

B*390501	37	MICA	196	27	72,97%
		D6S2926	265	29	78,38%
		D6S2927	240	34	91,89%
		MICA:D6S2926:D6S2927	196:265:240	22	59,46%
		D6S2928	257	31	83,78%
		D6S2885	104	20	54,05%
		D6S2928: D6S2885	257:104	20	54,05%
B*390900	22	MICA	196	22	100%
		D6S2926	265	21	94,59%
		D6S2927	240	22	100%
		MICA:D6S2926:D6S2927	196:265:240	21	94,59%
		D6S2928	257	10	45,45%
		D6S2885	104	15	68,18%
		D6S2928: D6S2885	257:104	6	27,27%
B*4004	22	MICA	184	20	90,91%
		D6S2926	273	16	72,73%
		D6S2927	228	18	81,82%
		MICA:D6S2926:D6S2927	184:273:228	14	63,63%
		D6S2928	257	15	68,18%
		D6S2885	100	20	90,91%
		D6S2928: D6S2885	257:100	14	63,63%
B*400201G	118	MICA	184	95	80,51%
		D6S2926	273	104	88,13%
		D6S2927	228	109	92,37%
		MICA:D6S2926:D6S2927	184:273:228	80	67,79%
		D6S2928	257	97	82,20%
		D6S2885	100	116	98,30
		D6S2928: D6S2885	257:100	95	80,51%
B*480201G	13	MICA	196	7	53,84%
		D6S2926	273	8	61,54%
		D6S2927	232	13	100%
		MICA:D6S2926:D6S2927	196:273:232	5	38,46%
		D6S2928	261	10	76,92%
		D6S2885	100	12	92,31%
		D6S2928: D6S2885	257:100	95	80,51%
B*510101G	61	MICA	187	33	54,09%
		D6S2926	269	38	62,29%
		D6S2927	232	54	88,52%
		MICA:D6S2926:D6S2927	187:269:232	23	37,70%
		D6S2928	257	38	62,29%
		D6S2885	104	38	62,29%
		D6S2928: D6S2885	257:104	29	47,54%
B*520102	27	MICA	184	22	81,48%
		D6S2926	269	15	55,56%
		D6S2927	232	21	77,78%
		MICA:D6S2926:D6S2927	184:269:232	10	37,04%
		D6S2928	257	19	70,37%
		D6S2885	104	19	70,37%
		D6S2928: D6S2885	257:104	16	59,26%

Discussão

Falha de amplificação

O nível de falha de amplificação dos produtos de PCR foi mais elevado para os microssatélites na região do MHC do que no resto do genoma (8,99% e 5,86% respectivamente). A maior falha de amplificação para o conjunto de microssatélites do MHC pode ser atribuída a dois fatores. a) Degradação da fluorescência dos oligonucleotídeos iniciadores. Os oligonucleotídeos iniciadores utilizados na amplificação dos microssatélites do MHC foram gentilmente cedidos pelo prof. Dr. Eduardo Santos da Universidade Federal do Pará. Esses oligos foram adquiridos no ano de 2003 e apresentaram menor intensidade de fluorescência que os iniciadores utilizados para a genotipagem dos microssatélites genômicos, adquiridos em 2009. A degradação da fluorescência ao longo do tempo é algo reconhecido pelo fabricante dos oligos (Life Technology LTDA). Como nota de esclarecimento, vale a pena descrever que uma característica da genotipagem de microssatélites é que, de modo geral, os picos de fluorescência dos alelos em heterozigose não apresentam a mesma intensidade, sendo que o alelo com maior número de repetições apresenta menor intensidade de fluorescência (ou seja, um menor número de cópias desse alelo é amplificado em relação ao alelo de menor tamanho). Para nos assegurarmos de que erros de genotipagens não foram introduzidos devido à baixa intensidade de fluorescência e que indivíduos heterozigotos não deixaram de ser detectados, consideramos apenas as genotipagens que apresentaram picos de fluorescência superior a 100 UF (unidades de fluorescência). Com esse critério, alguns genótipos homozigotos foram excluídos e contribuíram para o aumento da percentagem de falha de amplificação. b) Degradação do material genético. Como discutido no capítulo anterior, as amostras utilizadas no presente estudo foram coletadas no início da década de 90 e apresentam o DNA degradado. Além disso, pudemos observar que as amostras doadas pelo prof. Dr. Andrés Ruiz-Linares são as que apresentam maior discrepância de falha de amplificação entre o conjunto de microssatélites do MHC e genômicos (figuras 21 e 22). Resaltando que as genotipagens dos microssatélites genômicos para as amostras do professor Andrés foi realizada pela empresa de genotipagem Marshfield no trabalho de Wang *et al.*,

2007. Devido à dificuldade de amplificação dessas amostras, não apenas no presente estudo, mas também em outros realizados em nosso grupo de pesquisa (comunicação pessoal com Rodrigo S. Francisco e Márcia R. Pincerati) supomos que essa amostra também possa apresentar material genético degradado. Assim, como relatado para os microssatélites genotipados no capítulo 1, também realizamos alterações nos protocolos de PCR para possibilitar a genotipagem de amostras com problemas de amplificação, e em alguns casos separamos os multiplex em PCR individuais.

Desequilíbrio de Hardy-Weinberg e detecção de alelos nulos

Os microssatélites da região do MHC também apresentaram maior proporção de desvios de Hardy-Weinberg quando comparados com os microssatélites espalhados pelo genoma (14,9% e 5,9% respectivamente).

É importante ressaltar que os microssatélites da região do MHC violam alguns pré-supostos para a análise de Hardy-Weinberg: 1) *loci* não segregam independentemente. Os microssatélites do MHC analisados no presente estudo estão fisicamente próximos (entre 16 até 79412 kb de distância) e como visto nas figuras 23 e 24 apresentam diferentes níveis de desequilíbrio de ligação. Portanto, a frequência dos alelos de diferentes *loci* não é totalmente independente. 2) *loci* podem estar sob influência de seleção natural. Partimos do pressuposto que os *loci* de microssatélites estão em desequilíbrio de ligação com os genes HLA que estão sob regime seletivo. A seleção natural causa mudanças não casuais nas frequências alélicas, o que leva a desvios nas estimativas de Hardy-Weinberg. Esses fatores podem contribuir para a discrepância entre os resultados obtidos com os microssatélites espalhados pelo genoma (neutros e segregam independentemente).

Com essas questões em mente, utilizamos o teste de Hardy-Weinberg apenas com intuito de detectar *loci* que apresentam um excesso de desvios em relação aos demais *loci* analisados. Esses excessos poderiam ser indicativos de que há algum alelo nulo na amostragem. Como não observamos *outliers* na distribuição, supomos que a nossa amostragem não apresenta alelos nulos.

Perfil dos microssatélites da região do MHC

Os microssatélites do MHC não apresentam perfis homogêneos quanto à diversidade de cada *locus*. Sendo os microssatélites *HLA-BC*, *D6S2972* e *D6S276* os que apresentam maiores níveis de diversidade e *G51152*, *D6S2928* e *D6S291* são os com menores níveis de diversidade (tabela 17). OLIVEIRA, 2007 analisou as populações indígenas Guarani e Kaingang e também observou que os microssatélites *HLA-BC*, *D6S2972* e *D6S276* são os mais diversos ($H_e=0,778$; $0,788$; $0,708$) e *G51152* e *D6S291* os menos diversos ($H_e=0,43$ e $0,418$). Estudos com populações miscigenadas também relatam o mesmo perfil (MALKKI *et al.*, 2005; ABUAZAR, 2008), portanto podemos concluir que essas características são intrínsecas desses microssatélites e não das populações nativas americanas.

MALKKI *et al.*, 2005 observaram que há um aumento das estimativas de heterozigose para microssatélites fisicamente próximos a genes HLA de classe I e sugere que isso poderia ser um efeito da seleção balanceadora atuando nas regiões adjacentes. No nosso estudo também verificamos que os maiores índices de diversidade correspondem a microssatélites localizados na região de classe I, contudo observamos que o microssatélite *D6S2928*, flanqueador do gene *HLA-B*, apresenta baixa diversidade (tabela 17). Esse marcador não foi analisado por outros autores, porém o padrão observado é o mesmo tanto nas populações Ameríndias como em Ovimbundu e Portugueses. Apesar da baixa diversidade apresentada por esse marcador é possível observar nas análises de associação dos alelos de microssatélites com linhagens e alelos de *HLA-B* (Tabelas 22 e 23) que a associação deste *locus* é baixa quando comparada com os demais microssatélites. Basicamente esse *locus* tem dois alelos mais frequentes (alelo 257 e 261) e um mesmo alelo de *HLA-B* pode estar associado com os dois alelos de *D6S2928*. Para explicar esse padrão sugerimos que esse *locus* apresenta um modelo mutacional com recorrência de homoplasia na formação desses dois alelos (257 e 261).

Averiguação de indícios de seleção na região do MHC

Para o conjunto de microssatélites do MHC, as populações Ovimbundu (Angola) e Portugueses apresentam a maior diversidade de alelos, sendo que os alelos observados na América são um subconjunto dos encontrados nessas

populações (figuras 25 a 28, tabela 18). As populações da América do Sul, por sua vez, são as que possuem a menor diversidade alélica (tabela 18) e seus alelos são um subconjunto daqueles presentes nas demais populações. Esse padrão é o mesmo observado para os dados de microssatélites genômicos e reflete bem a história das migrações sequencias do homem moderno após deixar o continente africano discutidas no capítulo anterior (INGMAN *et al.*, 2000; THOMPSON *et al.*, 2000; UNDERHILL *et al.*, 2000; PRUGNOLLE *et al.*, 2005; RAMACHANDRAN *et al.*, 2005; WANG *et al.*, 2007; TISHKOFF *et al.*, 2009).

Apesar de haver um grande compartilhamento de alelos entre as populações do presente estudo, foi possível constatar através do teste exato de Fisher que a distribuição das frequências alélicas difere entre as populações. Os Ovimbundu e Portugueses apresentam as frequências mais distintas em relação às populações da América. Na comparação entre as populações da América percebemos que há muita diferença na distribuição das frequências alélicas, e que mesmo populações geograficamente próximas apresentam diferenças significativas. Observamos que algumas populações do leste da América do Sul, como Ticuna Tarapaca, Arara do Iriri e Araweté são as que com maior recorrência apresentam suas frequências alélicas significativamente distintas das demais populações (Tabelas A.3 a A.18 - ANEXO 1).

As análises de diferenciação populacional baseadas no índice F_{ST} quantificam a variância na frequência dos alelos entre as populações. Valores de F_{ST} baixos indicam que a frequência dos alelos nas populações é semelhante, enquanto valores altos significam que as frequências alélicas são diferentes entre as populações. Se um *locus* for favorecido pela seleção natural espera-se que esse *locus* apresente uma estimativa de F_{ST} mais elevada do que *loci* que diferem em suas frequências alélicas apenas como resultado da deriva genética. Por outro lado, *locus* sob regime de seleção balanceadora apresentam frequências alélicas parecidas e esperasse que apresentem estimativas de F_{ST} mais baixa que *loci* que diferem por deriva.

Como mencionado anteriormente os genes HLA apresentam evidencias de estarem evoluindo sob seleção balanceadora. Desta forma esperávamos que os microssatélites da região do MHC apresentassem baixos valores de F_{ST} . Contudo, nossas análises mostram que os valores de F_{ST} para os microssatélites localizados na região do MHC são mais elevados do que as estimativas para os microssatélites

espalhados pelo genoma (tabela 19), sendo esse efeito observado tanto entre populações do continente americano quanto em comparações envolvendo populações fora da América (Ovimbundu e Portugueses).

Averiguamos se a maior diferenciação poderia estar relacionada a alguma população específica (como observado para o conjunto de microssatélites genômicos no capítulo 1). Verificamos que as populações Andinas e do Noroeste da América do Sul apresentam a mesma contribuição na diferenciação de suas respectivas regiões (tabela 20). No Leste da América do Sul as populações que individualmente mais contribuem para o aumento da diferenciação na região são Arara do Iriri e Araweté (F_{ST} sem as populações=0,1473 e 0,1346 respectivamente). Essas mesmas populações já haviam sido identificadas como as que mais contribuem para a diferenciação no leste da região amazônica, quando analisamos os microssatélites genômicos (ver capítulo 1). Como discutimos no capítulo 1, essas populações apresentam histórias evolutivas muito peculiares. Arara do Iriri é uma população formada a partir de um casal de irmãos que foram expulsos de sua tribo de origem (Arara do Laranjal; ver detalhes no capítulo 1). Por outro lado, Araweté apresenta uma história de declínio populacional em 1976 em decorrência de doença adquirida através do contato com populações não ameríndias (exploraremos a história de Araweté em mais detalhes no próximo item). Contudo, verificamos que independente dessas duas populações, os níveis de diferenciação na região do MHC ainda são mais elevados que para as demais regiões do genoma (figuras 32 e 33 e tabela 20).

Em seguida, através de testes de permutação, verificamos que a chance de ao acaso reunirmos os valores mais elevados de F_{ST} , em um conjunto de 16 microssatélites, é significativamente baixo ($p < 0,001$) (figura 34).

Por fim, examinamos se os nossos resultados poderiam estar sendo influenciados pela comparação de microssatélites com diferenças no tipo de unidade de repetição (a maioria dos microssatélites da região do MHC são dinucleotídeos - com exceção de *MICA*, *D6S2926*, *D6S2927* e *D6S2928* e os microssatélites espalhados pelo são tetranucleotídeos), uma vez que alguns autores defendem que a taxa de mutação pode variar entre os tipos de microssatélites. Ao compararmos os dados de dinucleotídeos e tetranucleotídeos genômicos

disponibilizados por WANG *et al.* (2007) verificamos que o valor de F_{ST} médio entre os *loci* é praticamente o mesmo. Também realizamos uma comparação par a par entre os valores F_{ST} das populações genotipadas por WANG *et al.* (2007) e observamos uma pequena tendência de maior diferenciação entre os dinucleotídeos genômicos em relação aos tetra (figura 35). Entretanto, essa diferenciação não é tão pronunciada como a observada para os microssatélites do MHC. Portanto a diferenciação não é totalmente explicada pelo tipo de unidade de repetição do microssatélite.

Frente ao exposto, concluímos que a região do MHC apresenta de fato um sinal de diferenciação mais pronunciada quando comparado com o resto do genoma. Recentemente, BHATIA *et al.* (2011) analisaram 500.000 SNPs em 6209 afro-americanos, 756 nigerianos e 2946 gambianos e detectaram que a região do MHC apresenta a maior diferenciação em relação ao resto do genoma.

A aparente discrepância entre os nossos resultados e os baixos valores de F_{ST} observados para os genes *HLA* entre populações de diversas regiões do mundo (MEYER *et al.*, 2006) reflete uma limitação da medida de F_{ST} frente ao complexo padrão de diversidade dos genes *HLA*, onde grupos de populações ou regiões geográficas podem apresentar conjuntos de alelos distintos mas com padrões de frequências alélicas similares (SOLBERG *et al.*, 2008). No caso, específico de *HLA-B*, é nítido a distinção entre o conjunto de alelos observados nas populações do leste da América do Sul e o observado nas populações nativas do México (FRANCISCO, 2009), mas mesmo assim, o valor estimado de F_{ST} é baixo entre essas regiões (Rodrigo dos Santos Francisco, comunicação pessoal). A explicação para essa observação é que a medida F_{ST} (e outras medidas correlacionadas) estima a fração da variância total da variação que está distribuída entre populações, e é fortemente afetado pela variância intra-populacional. Nos casos de populações diferenciadas, porém internamente muito variáveis (como ocorre nos genes *HLA*), valores baixos de F_{ST} podem ser obtidos (até mesmo na condição extrema, mas não incomum para alelos de *HLA-B*, por exemplo, em que populações não partilham nenhum alelo). Esse problema pode ser minimizado através da implementação do índice de diferenciação D descrito por JOST (2008), que é uma medida de "verdadeira diferenciação". Análises preliminares com os genes *HLA* clássicos nas populações

nativas da América, utilizando essa medida mostram que há alta diferenciação entre as populações (Diogo Meyer, comunicação pessoal), mesmo nos casos em que o F_{ST} é baixo. Nós aplicamos esse índice aos dados de microssatélites e observamos que o padrão de diferenciação para esse conjunto de marcadores é semelhante aos obtidos com as medidas de F_{ST} e R_{ST} tanto para os microssatélites da região do MHC como para os microssatélite espalhados pelo genoma.

Perfil dos desvios de neutralidade para a região do MHC

Através da distribuição nula criada a partir da relação entre F_{ST} e H_e dos microssatélites espalhados pelo genoma, percebemos que parte dos microssatélites restritos a região do MHC apresentam níveis de diferenciação significativamente elevados (figuras 36 a 41). Na análise incluindo as populações africana (Ovimbundu), europeia (Portugueses) e nativo-americanas verificamos que 11 dos 16 microssatélites da região do MHC desviam da distribuição nula na direção de excesso de diferenciação (figura 36), e nenhum apresenta redução de diferenciação, ou seja, baixo índice de F_{ST} . Esse padrão é muito diferente do que seria esperado se os microssatélites do MHC se comportassem como uma amostra aleatória dos genômicos, ou se estivessem sob seleção balanceadora. Ao restringirmos as análises apenas às populações do continente americano, verificamos que o número de microssatélites da região do MHC fora da distribuição nula cai de onze para sete. Análises regionais mostram que o número e a identidade dos microssatélites do MHC fora da distribuição nula variam entre as regiões do continente (figuras 38 a 41) sendo que a América do Norte não apresenta diferença na distribuição dos microssatélites genômicos e do MHC. O Noroeste e o Leste da América do Sul apresentam três microssatélites distintos do MHC que desviam da distribuição (*D6S2928*, *D6S273* e *G51152* para o noroeste e *D6S265*, *G51152* e *D6S439* para o leste) (figuras 39 e 41). As populações Andinas apresentaram o maior número de microssatélites fora da distribuição nula, oito dos 16 (*D6S2972*, *HLA-BC*, *D6S2885*, *D6S2928*, *D6S2926*, *MICA*, *BAT2* e *D6S439*). Essas diferenças podem indicar que pressões seletivas regionais podem ter contribuído para a homogeneização local e diferenciação inter-regional dos genes adjacentes a esses microssatélites. Esse

resultado é semelhante ao sugerido por FRANCISCO (2009) com base no perfil do gene *HLA-B*.

Uma região geográfica que nos chama a atenção é a Andina. As populações andinas apresentam as maiores diferenças nos valores de F_{ST} , com baixos índices para os microssatélites genômicos ($F_{ST} = 0,0217$) e alta diferenciação para os microssatélites do MHC ($F_{ST} = 0.0993$) (tabela 19). As populações que representam a região andina nas nossas análises são Inga (ou Ingano), Aymara e Huilliche. Essas populações estão distantes geograficamente (Inga localiza-se no sul da Colômbia, Aymara no norte do Chile e Huilliche no Sul do Chile). Inga e Aymara fizeram parte do Império Inca, e Huilliche viveu na vizinhança (MANN, 2007). É reconhecido que o Império Inca estreitou relações e promoveu a interação entre os povos da região andina, o que indiretamente levou a um maior fluxo gênico na região (MANN, 2007), o que explica os baixos valores de F_{ST} para os microssatélites espalhados pelo genoma na região (TARAZONA-SANTOS *et al.*, 2001; WANG *et al.*, 2007). Por outro lado, a análise do gene *HLA-B* nessas populações revelam que a composição alélica difere entre as populações (FRANCISCO, 2009). Em Inga os alelos mais frequentes são: *HLA-B*400201* (26,7%), *B*150101* (16,7%), e *B*3543* (13,3%), em Aymara - *B*3505* (37,5%) e *B*480101* (15,0%) e em Huilliche *B*3909* (42,1%) e *B*510101* (13,2%), sendo que em Huilliche há sinais de seleção positiva no gene *HLA-B*, detectado através do teste de neutralidade de Ewens-Watterson (FRANCISCO, 2009). Portanto, por mais homogenizador que tenha sido o papel do Império Inca na região Andina, não vemos esse sinal na região do MHC.

Outra população que chama a atenção quanto à discrepância entre os valores de F_{ST} para os microssatélites genômicos é Araweté. Essa população, como citado no item anterior, apresenta altos valores de diferenciação tanto para os microssatélites genômicos como para os do MHC. Contudo, a diferenciação do MHC é mais elevada que em Arara do Iriri (a população fundada por um casal de irmãos). Os relatos na literatura revelam que em julho de 1976 funcionários da FUNAI encontraram a população Araweté doente em decorrência de uma conjuntivite infecciosa e decidiram levá-los até o posto de saúde mais próximo, a 100 km de distância. A caminhada durou 17 dias e por não enxergarem o caminho muitos se perderam na mata e morreram de fome. Estima-se que em torno de 36% da

população morreu durante esse episódio (Ver mais detalhes em: <http://pib.socioambiental.org/pt/povo/arawete>). Araweté é a população que mais contribui na diferenciação da região leste para os dados de microssatélites do MHC (para os dados de microssatélites espalhados pelo genoma, a diferenciação ao excluir Arara do Iriri e Araweté é $F_{ST} = 0,0865$ e $0,1071$ respectivamente– tabela 20). Indagamos até que ponto o episódio descrito acima pode ter exercido alguma pressão seletiva sob o MHC. Analisando os dados para o gene *HLA-B* nessa população, observamos que o alelo *HLA-B*1504* ocorre com frequência de 67,86%, sendo que de 14 indivíduos genotipados, 12 apresentam pelo menos uma cópia desse alelo. Os indivíduos que não apresentam o alelo *HLA-B*1504*, apresentam o alelo da mesma linhagem *HLA-B*1520* (Rodrigo dos Santos Francisco, comunicação pessoal). Portanto, é possível que tenha ocorrido um favorecimento desse alelo *HLA-B*1504* ou da linhagem *15, nessa população. Contudo como não dispomos de amostras dessa população coletadas anterior a 1976 (data do declínio populacional pontual), não sabemos como era a composição genética da população naquela época e, portanto não podemos descartar a hipótese que o alelo *B*1504* já ocorria em alta frequência nessa população.

Desequilíbrio de ligação

Através das nossas análises é possível observar que os microssatélites da região do MHC apresentam alto nível de desequilíbrio de ligação (figuras 42 a 46). Diversos estudos relatam que a região do MHC está entre as regiões do genoma com maior nível de desequilíbrio de ligação (HEDRICK e THOMSON, 1986; HUTTLEY *et al.* 1999), sendo que esse níveis excedem o esperado para populações que evoluem de modo neutro (HEDRICK e THOMSON, 1986).

Também observamos que cada população apresenta suas peculiaridades quanto ao nível de desequilíbrio (figuras 42 a 46). A associação não aleatória de alelos de dois ou mais *loci* reflete muito da história evolutiva das populações. Fatores como sistemas de casamentos, subdivisões, gargalos populacionais e o fluxo gênico, associados à ação da seleção natural, recombinação e conversão gênica, além do modelo mutacional do *locus* em questão, moldam os padrões de desequilíbrio de ligação em cada população (revisão em SLATKIN, 2008).

É possível observar que devido ao excesso de *loci* em homozigose na população Arara do Iriri, esses *loci* estão totalmente ligados com os demais (e por isso o desequilíbrio é calculado como zero). Outras populações com alto níveis de desequilíbrio de ligação são Kogi e Araweté. Kogi é uma população que vive isolada na Sierra Nevada de Santa Marta na Colômbia e, de acordo com as análises realizadas por WANG *et al.* (2007), as amostras analisadas no presente estudo apresentam indivíduos com alto grau de parentesco (pais, irmãos e primos em primeiro grau), o que explica o elevado grau de desequilíbrio de ligação. Araweté, como discutido no tópico anterior é uma população que sofreu um forte gargalo populacional recentemente o que pode levar a elevação dos níveis de desequilíbrio de ligação.

A heterozigose associada aos alelos de *HLA-B*

Determinamos a heterozigose associada a cada um dos alelos de *HLA-B* mais comuns nas populações nativas do continente americano. Essa análise teve como objetivo verificar como a diversidade dos microssatélites adjacentes está associada aos alelos de *HLA-B*. Alelos de *HLA-B* cujos microssatélites apresentam baixa heterozigose são candidatos à seleção direcional. Já alelos com excesso de variação seriam aqueles associados à seleção balanceadora, que teria mantido alelos por longos períodos de tempo, aumentando a diversidade intra-alélica.

Observamos que de modo geral, a heterozigose alelo específica é menor que a heterozigose média ao longo de todos os *loci* de microssatélites e alelos de *HLA-B* (tabela 21). Isso é um indicativo de que há uma associação de alelos específicos dos microssatélites com os alelos de *HLA-B*, o que não é inesperado, dado a ligação física. Também observamos que, os alelos dos microssatélites localizados entre os genes *MICA* e *HLA-B* (*MICA*, *D6S2926* e *D6S2927*) apresentam menor heterozigose que os alelos localizados entre os genes *HLA-B* e *-C* (*D6S2928*, *HLA-BC* e *D6S2885*) (tabela 21). O mestrando Fabio Mendes, do nosso grupo de pesquisa, realizou uma análise com SNPs na região MHC, a partir de dados preliminares do projeto “1000 Genomas” e verificou que o maior pico de diversidade da região encontra-se entre os genes *HLA-B* e *C*. Ainda não está claro qual seria o fenômeno que estaria mantendo a diversidade dessa região.

Na tentativa de compreender se há diferenças na heterozigose alelos específica dos alelos endêmicos e cosmopolitas de *HLA-B* estimamos a HSH para o alelo endêmico e cosmopolita mais frequente nas populações. Observamos que alguns alelos endêmicos de *HLA-B*, como *B*3909* e *B*3543*, apresentam baixa heterozigose associada aos alelos de microssatélites (tabela 21). Das 22 cópias do alelo *B*3909*, 16 ocorrem na população Huilliche, sendo que todos apresentam o mesmo haplótipo para os microssatélites *MICA*, *D6S2926* e *D6S2927* (196:265:240:*3909). Como discutido anteriormente, essa população apresenta sinais de seleção positiva para os dados de *HLA-B* (FRANCISCO, 2009). O alelo *B*3543* apresenta distribuição predominante nas populações da América Central (Kaqkichel, Cabecar e Guaymi) e do Noroeste da América do Sul (Zenu, Kogi e Arhuaco) e também apresenta conservação do haplótipo *MICA*, *D6S2926* e *D6S2927* (196:265:244:*3543).

O alelo *B*1504* é o único alelo realmente endêmico do continente americano e ocorre apenas em populações da região Amazônica (Waunana, Tiriyo, Arara do Laranjal, Urubu-Kaapor, Assurini, Araweté e Kayapó), Nós verificamos que a heterozigose alelo específica do *B*1504* é similar aos demais alelos de *HLA-B* nas Américas. Contudo, como mencionado acima, na população Araweté praticamente todos os indivíduos (12 de 14) apresentam pelo menos uma cópia do alelo *B*1504* (FRANCISCO, 2009) e ao analisarmos o haplótipo estendido do alelo *B*1504* nessa população observamos que todos os indivíduos apresentam o haplótipo 184:261:232:*1504:261:114:100 para os loci *MICA*, *D6S2926*, *D6S2927*, *HLA-B*, *D6S2928*, *HLA-BC*, *D6S2885*. Porém como discutido anteriormente, no caso específico desse alelo em Araweté é extremamente difícil distinguir qual o efeito da demografia e da seleção nesse processo de homozigose haplótipo estendida.

Devido à restrição de tempo para finalizar a presente tese não foi possível realizar a análise de heterozigose alelo específica para cada população individualmente. Considerando que cada população possui sua história evolutiva julgamos essa análise essencial para compreender se há diferenças nos perfis dos alelos endêmicos e cosmopolitas nas populações nativas do continente americano.

Diferenciação inter e intra-linhagem

Os alelos de *HLA* são primariamente classificados em linhagens. Cada linhagem é formada por um conjunto de alelos, que compartilham sítios polimórficos em suas sequências de DNA. Em uma escala evolutiva, as linhagens existem mesmo antes da divergência entre humanos e chimpanzé.

Nossos resultados de diferenciação evidenciam que há grande diferenciação entre as principais linhagens de *HLA-B* nas Américas (*15, *35, *39, *40,*48 e *51) e menor diferenciação dentro das linhagens (tabela 22). Além disso, mostramos que haplótipos específicos de microssatélites apresentam forte associação com determinadas linhagens (tabela 23). Por outro lado, *loci* de microssatélites com alto grau de diferenciação dentro das linhagens podem indicar que o conjunto de alelos daquele *locus* é diferente entre os alelos dentro da linhagem de *HLA-B* (tabela 22 e 24). Se levarmos em consideração a longa existência das linhagens, dentro da escala evolutiva, esperaríamos que houvesse um grande número de alelos diferentes de microssatélites associado as linhagens e que o nível de diferenciação entre as linhagens fosse baixo.

Esses resultados podem ser explicados de duas maneiras: Na primeira, sugerimos devido ao gargalo populacional ocorrido na entrada do homem moderno na América, ocorreu perda da diversidade dos alelos de microssatélites adjacentes ao gene *HLA-B*, e como consequência haveria uma maior homozigose dos alelos de microssatélites associados as linhagens. Uma segunda explicação seria que esses resultados podem ser um indicativo de que não apenas os alelos de *HLA-B*, mas também as linhagens podem ser alvo da seleção.

O questionamento se linhagens ou alelos de *HLA* são alvos de seleção foi tema de um estudo recente por parte de nosso grupo (BITARELLO, 2011). A autora analisou nos genes *HLA* clássicos os códons responsáveis na molécula pela ligação ao peptídeo antigênico e comparou a relação entre a taxa de substituição sinônima e não sinônima nessa região da molécula entre os alelos e as linhagens de *HLA*.

As análises revelaram que diferenças significativamente maiores entre pares de alelos de linhagens distintas que para pares de alelos da mesma linhagem, porém os padrões de variação dentro das linhagens de *HLA* e entre as linhagens são incompatíveis com o cenário neutro. Esse resultado é coerente com a hipótese de

que tanto alelos como linhagens de *HLA* podem ser alvos de seleção natural. Para respondermos se é a demografia ou a seleção que está moldando a variabilidade nos microssatélites adjacentes às linhagens de *HLA-B*, o doutorando Rodrigo dos Santos Francisco está realizando a genotipagem dos genes *HLA-B* nas amostras Ovimbundu e Portuguesas, genotipadas para os microssatélites do MHC no presente estudo.

Discussão Geral e Conclusões

Neste estudo buscamos contribuir para o esclarecimento de questões em aberto a respeito dos processos demográficos e seletivos da história das populações nativas do continente americano.

Mostramos que, assim como observado por outros autores (ex., TARAZONA-SANTOS *et al.*, 2001; WANG *et al.*, 2007; PUCCIARELLI *et al.*, 2006), existe uma clara diferenciação entre as populações que vivem nas regiões leste e oeste da América do Sul. Essa diferenciação foi inferida através da análise baseada em genótipos multi-*locus* usando o método bayesiano de atribuição de grupos a genótipos, implementado no programa STRUCTURE, e aponta que há diferenças nos perfis genotípicos de cada uma dessas regiões. Ao subdividirmos a região oeste em Andina e Noroeste, mostramos que elas também são diferenciadas, e que as populações da região andina apresentam índices de diferenciação extremamente baixos ($F_{ST} = 0,018$), achado coerente com efeito homogenizador associado à dinâmica populacional dessa região, que inclui os descendentes do Império Inca, o qual estreitou relações e promoveu a interação entre os povos da região andina, o que indiretamente levou a um maior fluxo gênico na região como sugerido por TARAZONA-SANTOS *et al.* (2001). Já a região noroeste apresenta um F_{ST} mais elevado (0,0783), possível reflexo do modo de vida semelhante ao observado no leste da América do Sul.

Com uma amostragem de 20 populações na porção leste da América do Sul, (5 da região oeste da Amazônia, 11 do leste da Amazônia e 4 do centro-sul do leste da América do Sul) pudemos confirmar que essa região apresenta os menores índices de diversidade e maior diferenciação populacional do mundo. À luz da atual visão da região amazônica pré-colombiana como dispendo de uma ampla rede de comércio e maior interação entre as populações, buscamos esclarecer porque os dados genéticos sugerem uma história demográfica aparentemente mais fácil de conciliar com uma história de populações relativamente pequenas e isoladas. Identificamos que algumas populações específicas (como, Arara do Iriri, Araweté, Surui e Ticuna Tarapaca) contribuem desproporcionalmente para a elevação da diferenciação populacional na região. Essas populações apresentam histórias evolutivas marcadas por gargalos populacionais, efeito fundador extremo e amostras incluindo indivíduos aparentados, o que contribui para menor diferenciação intra e maior diferenciação inter-populacional (como verificado previamente – WANG *et al.*, 2007). Por outro lado, mostramos que as demais populações da região leste apresentam índices de diferenciação inter-populacional

semelhante aos observados para as populações da região Noroeste da América do Sul corroborando os estudos morfológicos de PUCCIARELLI et al. (2006) . Como mencionado anteriormente, a região Noroeste da América do Sul e a Amazônia brasileira (onde estão localizadas a maioria das nossas amostras) apresentam bioma e clima semelhante, o que pode ter favorecido as duas regiões a apresentarem modos de vida semelhante.

Nosso estudo documentou uma grande diferenciação entre a região leste e sudoeste da Amazônia. Esse resultado já havia sido verificado através de análises com marcador uniparental (DNAmT – MENDES JUNIOR, 2005) e com marcadores genéticos clássicos e dados antropométricos (SALZANO e CALLEGARI-JACQUES, 1988). Contudo, esta é a primeira vez que esse resultado é apresentado com base em marcadores moleculares autossômicos. Também verificamos, através de análises de ancestralidade, que a região sudoeste da Amazônia apresenta um predomínio de componente genético das populações do Noroeste da América do Sul, enquanto as populações do leste apresentam maior componente genético das populações Andinas. Apesar desses resultados corroboram a hipótese de rotas migratórias para o leste da América do Sul descritas por SCHMITZ em 1983 julgamos a nossa análise sensível a estruturação populacional por isso sugerimos que novas análises baseadas em ABC, por exemplo, sejam realizadas futuramente.

O desafio de oferecer um panorama da história e demografia das populações nativas sul-americanas tinha dois objetivos: (a) contribuir para a compreensão da história das populações nativas, e (b) criar uma base de dados neutra, que serviu de referência para o estudo de como a seleção natural moldou a variação em genes específicos de populações nativas da América do Sul. Esse segundo aspecto será discutido a seguir.

Estudos recentes mostram que a história demográfica (geografia) tem um papel central na determinação dos padrões de variação genômicos. Por exemplo, COOP *et al.* (2009), mostraram que até mesmo genes com forte evidência de estarem sob seleção positiva apresentam padrões de variação global que são em sua maior parte explicáveis pela história demográfica das populações, reveladas pelos demais *loci*. No presente estudo buscamos investigar de que modo a história demográfica (notadamente caracterizada por gargalos populacionais e concomitante perda de diversidade, no caso das populações indígenas da América do Sul), interage com os processos seletivos, que são muito bem documentados para genes da região do MHC, levando a um aumento de diversidade em relação às expectativas neutras.

Assim, seguindo a estratégia de COPP *et al.*, (2009), analisamos os padrões de variação e diferenciação genética em regiões candidatas a seleção natural, comparando-as com medidas desses mesmos parâmetros, porém vindo de regiões

aleatórias do genoma, dessa forma refletindo a história demográfica. Nossa abordagem consistiu na comparação de microssatélites da região do MHC e fora dele, uma estratégia que contorna as dificuldades metodológicas associadas à comparação de diferentes classes de marcadores moleculares (ex. comparação de genes HLA com microssatélites ou RFLPs – MEYER *et al.*, 2006; SANCHEZ-MAZAS *et al.*, 2007).

Nosso resultado mais surpreendente foi que os microssatélites da região do MHC apresentam diferenciação populacional superior ao observado para os microssatélites espalhados pelo genoma (nosso controle demográfico), tanto nas análises de regressão como nas de F_{ST} *outliers*, em que simulações são usadas para inferir um intervalo de confiança associado à amplitude de valores de F_{ST} esperada sob neutralidade. Testes de aleatorização mostraram que a diferenciação maior para microssatélites do MHC não é explicável por efeitos amostrais (isto é, o tamanho pequeno da amostra de microssatélites do MHC sendo empregados), e uma comparação entre di e tetranucleotídeos sugere que uma diferenciação maior para dinucleotídeos, mesmo que exista, tem um efeito sensivelmente menor do que aquele observado na comparação de microssatélites do MHC e do restante do genoma.

Nossa expectativa original era baseada na premissa de que a seleção balanceadora, por reduzir a intensidade de deriva genética, resultaria em níveis de diferenciação reduzidos em relação à média (presumivelmente neutra) genômica (ex., AKEY *et al.*, 2002, MEYER *et al.*, 2006). Como os genes *HLA* estão sob intensa seleção balanceadora, e os microssatélites do MHC, em muitos casos, estão em desequilíbrio de ligação com os genes clássicos, nossa expectativa era de que os microssatélites refletiriam essa consequência do processo seletivo, resultando em valores mais baixos de F_{ST} para os microssatélites do MHC em relação aos genômicos.

Nossa interpretação para o achado de maior diferenciação nos microssatélites do MHC leva em conta a forma como a seleção natural atua sobre os genes *HLA*. Ao invés de supor que a seleção balanceadora favorece heterozigotos de modo geral, independentemente do genótipo específico (e, portanto da especificidade da resposta que montam frente a patógenos), é razoável supor que os alelos favorecidos em cada população sejam determinados por um processo co-evolutivo, determinado pela interação com os patógenos (ex. BORGHANS *et al.*, 2004). Tal regime pode, simultaneamente, resultar em aumento de diversidade intra-populacional (uma das assinaturas de seleção balanceadora), e aumentar a diferenciação populacional. Em outras palavras, por mais que o regime de seleção balanceadora favoreça a diversidade, não necessariamente os mesmos alelos serão

favorecidos nas diferentes populações ou grupos. Nosso resultado é consistente com uma série de achados recentes, que sugerem que o regime de seleção que atua sobre genes *HLA* pode resultar em maior diferenciação populacional, ao contrário do que se supunha anteriormente (ex. BHATIA *et al.*, 2011). A razão que esse padrão não era visto na análise de genes *HLA*, mas tornou-se visível no estudo de microssatélites que os flanqueiam, resulta da limitação de medidas de F_{ST} no contexto de marcadores altamente polimórficos: a alta variabilidade intra-populacional reduz os valores de F_{ST} , mesmo diante da alta diferenciação no conjunto de alelos.

Também verificamos que existe uma associação não aleatória de alelos de microssatélites com os alelos e as linhagens do gene *HLA-B*. Esse resultado também é surpreendente, se levarmos em consideração que as linhagens de *HLA-B* são antigas, muitas existindo desde antes da especiação humana. Diante dessa antiguidade, esperaríamos uma maior diversidade entre os alelos de microssatélites associados a cada uma delas. Sugerimos que o padrão observado pode ser decorrente de dois fatores, não mutuamente exclusivos: a) o gargalo populacional na ocupação do continente americano pode ter diminuído consideravelmente a diversidade associada às linhagens; b) a seleção natural estaria atuando não apenas no nível alélico, mas também ao nível de organização biológica que corresponde às linhagens, resultando numa diminuição da variação dentro destas.

No presente estudo mostramos que a história demográfica teve um papel decisivo para moldar a variação genética nas populações nativas do continente americano. Contudo, baseado em nosso controle demográfico (perfil dos microssatélites espalhados pelo genoma) percebemos que ela não explica todo o padrão de diferenciação observado nos microssatélites do MHC, e que a seleção natural resultou em um aumento da diferenciação populacional para esses genes.

Referências Bibliográficas

Abbas AK, LICHTMAN AH, PILLAI S. *Imunologia celular e molecular*. Rio de Janeiro: Elsevier, 2008.

Abuazar CS, (2008) Análise de microssatélites da região MHC em uma amostra de doadores voluntários de medula óssea. Universidade Federal do Paraná. Dissertação de mestrado

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Niskerson A, Kruglyak L. "Population history and natural selection shape patterns of genetic variation in 132 genes". *PloS Biol*. 2004, 2:1591-99.

Akey JM. "Constructing genomic maps of positive selection in humans: where do we go from here?" *Genome research*. 2009, 19(5), 711-22.

Amos W, Hoffman JI. "Evidence that two main bottleneck events shaped modern human genetic diversity". *Proc Biol Sci*. 2010, 7;277(1678):131-7.

Andres AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR, Gutenkunst RN, *et al*. "Targets of Balancing Selection in the Human Genome." *Mol Biol Evol* . 2009.26 (12) (December): 2755-2764.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. "LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method". *BMC Bioinformatics*. 2008, 28;9:323.

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK. "The nature of selection on the major histocompatibility complex". *Crit. Rev. Immunol.*, 1997, 17(2), pp. 179-224.

Araújo A, Reinhard KJ, Ferreira LF, Gardner SL. "Parasites as a probes for prehistoric human migrations". *Trends Parasit*. 2008, 24(3):112-115.

Balaresque PL, Ballereau SJ, Jobling MA. "Challenges in human genetic diversity: demographic history and adaptation". *Hum Mol Genet*. 2007, Oct 15;16.

Bamshad M e Wooding SP. "Signatures of natural selection in the human genome". *Nature Rev. Genet*. 2003, 4: 99-111.

Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL. An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA*. 1997, 94:4516-4519.

Belich MP, Madrigal JA, Hildebrand WH, Zemmour J, Williams RC, Luz R, Petzl-Erler ML, Parham P. "Unusual *HLA-B* alleles in two tribes of Brazilian Indians". *Nature*. 1992 May 28;357(6376):326-9.

Bhatia G, Patterson N, Pasaniuc B, Zaitlen N, Genovese G, Pollack S, Mallick S, Myers S, Tandon A, Spencer C, Palmer CD, Adeyemo AA, Akylbekova EL, Cupples LA, Divers J, Fornage M, Kao WH, Lange L, Li M, Musani S, Mychaleckyj JC, Ogunniyi A, Papanicolaou G, Rotimi CN, Rotter JJ, Ruczinski I, Salako B, Siscovick DS, Tayo BO, Yang Q, McCarroll S, Sabeti P, Lettre G, De Jager P, Hirschhorn J, Zhu X, Cooper R, Reich D, Wilson JG, Price AL. "Genome-wide Comparison of African-Ancestry Populations from CARE and Other Cohorts Reveals Signals of Natural Selection". *Am J Hum Genet.* 2011, 89(3):368-81.

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi M, Videl-Rioja LB, Herrera RJ, Lopez-Camelo JS. "Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations". *Am J Hum Genet.* 1998, 63:1862-71.

Bitarello BD, (2011) Seleção natural em genes HLA: uma investigação da localização molecular e temporal dos eventos de seleção. Universidade de São Paulo. Dissertação de Mestrado.

Black FL. "Why did they die?" *Science.* 1992, 5089(258):1739-40.

Bonato SL, Salzano FM. "A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data". *Proc Natl Acad Sci USA.* 1997, 94:1866-71.

Bubb K.L.; Bovee D.; Buckley D.; Haugen E.; Kibukawa M.; Paddock M.; Palmieri A.; Subramanian S.; Zhou Y.; Kaul R.; Green P.; Olson M.V. "Scan of human genome reveals no new *Loci* under ancient balancing selection". *Genetics*, 2006, 173(4):2165-77.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG. "Natural selection on protein-coding genes in the human genome". *Nature.* 2005, 437(7062):1153-7.

Cadauid LF, Watkins DI. "The duplicative nature of the MHC class I genes: an evolutionary perspective". *Eur J Immunogenet.* 1997, 24(4):313-22.

Cavalli-Sforza LL, Menozzi P e Piazza A. *The history and geography of human genes.* Princeton: Princeton University Press, 1994.

Cavalli-Sforza LL, Feldman MW. "The Application of Molecular Genetic Approaches to the Study of Human Evolution". *Nature Genetics Supplement.* 2003, 33: 266-275.

Confalonieri U, Ferreira, LF, Araújo A. "Intestinal helminths in Lowland South America Indians: some evolutionary interpretations". *Hum. Bio.* 1991, 63:863-73.

Confaloneiri UEC. "Environmental change and human health in the Brazilian Amazon". *Global Change & Human Health*. 2000, 1(2):174-183

Confalonieri, UEC. "Saúde na Amazônia: um modelo conceitual para a análise de paisagens e doenças". *Est. Av.* 2005, 19(53):221-35.

Cornuet JM and Luikart G. "Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data". *Genetics*. 1997, 144:2001-2014.

Cruxent JM. Theses for meditation on the origin and distribution of man in South America. In: *Biomedical challenges presented by the Amerindian Indian Pan American*. Health Organization Scientific Publication. 1968. Vol 165 Washington, DC 11-16.

Cunha MC. (1992). *Historia dos índios no Brasil*. São Paulo: Companhia das Letras. 611p.

De Boer RJ; Borghans JA; van Boven M; Keşmir C; Weissing FJ. "Heterozygote advantage fails to explain the high degree of polymorphism of the MHC". *Immunogenetics*, 2004, 55(11), pp. 725-31.

Diamond J. *Armas Germes e Aço: os destinos das sociedades humanas*. Tradução: Silvia de Souza Costa. Rio de Janeiro: Record. 2006. 8ed.

Dillehay TD. *Monte Verde, a late pleistocene settlement in Chile vol 2 The archaeological context and interpretation*. Smithsonian Institution Press: Washington, DC. 1997.

Dillehay TD, Ramirez C, Pino M, Collins MB, Rossen J, Pino-Navarro JD. "Monte Verde: seaweed, food, medicine, and the peopling of South America". *Science*. 2008, 320:784-6.

Dixon EJ. "Human colonization of the Americas: timing, technology and process. Quaternary ". *Science Reviews*. 2001, v.20, p.277-299.

Doherty, PC. e Zinkernagel, RM. "Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex". *Nature*, 1975, 256, pp. 50-52.

Excoffier L, Lischer HEL. "Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows". *Molecular Ecology Resources*. 2010, 10:564-567.

Fiedel SJ. "The peopling of the New World: present evidence, new theories, and future directions". *J. Archaeol Res*. 2000, 8:39-103.

Fladmark KR. "Routes – alternate migration corridors for early man in North-

America". *Am. Antiq.* 1979, 44:55-69.

Foster P, Harding R, Torroni A, Bandelt HJ. "Origin and evolution of Native American mtDNA variation: a reappraisal". *Am J Hum Genet.* 1996, 1:151-57.

Foster ELS. (2004) *Racismo e movimentos instituintes na escola.* Universidade Federal Fluminense. Tese de doutorado.

Francisco R. (2009). *Polimorfismo e Evolução do gene HLA-B nas Américas.* Universidade de São Paulo. Dissertação de mestrado.

Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. "Mitochondrial DNA diversity in South América and the Genetic History of Andean Highlanders". *Mol Biol Evol.* 2003, 20:1682-91.

González-José R, González-Martín A, Hernández M, Pucciarelli HM, Sardi M, Rosales A, Van der Molen S. "Cranio-metric evidence for Paleoamerican survival in Baja California". *Nature.* 2003, 425:62-5.

González-José R, Bortolini MC, Santos FR, Bonatto SL. "The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view". *Am. J. Phys Anth.* 2008, 137:175-87.

Greenberg JH, Turner CG, Zegura SL. "The settlement of the America – a comparison of linguistic, dental, and genetic-evidence". *Curr Anthropol.* 1986, 27:477-97.

Guo SW, Thompson EA. "Performing the exact test of Hardy-Weinberg proportion for multiplex alleles". *Biometrics.* 1992, 48:361-72.

Gurgel C. *Doenças e Curas: o Brasil nos primeiros séculos.* São Paulo: Contexto. 2010.

Harris, EE e Meyer D. "The Molecular Signature of Selection Underlying Human adaptations." *Am. J. of Phys. Anth.*, 2006 doi:10.1002/ajpa.

Heckenberger M. 2006. "Ecologia e Poder: a base simbólica da economia na Amazônia". In: Forline L., Murrieta R.S.S., Viera, I.C.G. org. *Amazônia além dos 500 anos.* Belém: Museu Paraense Emílio Goeldi. pp.39:69

Heckenberger M. 2008. "History, ecology and alterity – visualizing polity in ancient Amazonia". In: Silverman H, Isbell W.H. editors. *Handbook of South American Archaeology.* New York: Springer. pp311:340.

Heckenberger M., Neves EG. "Amazonian archaeology". *Annu Rev Antropol.* 2009, 38:251-66

Hedrick PW e Thomson G. "Evidence for balancing selection at HLA". *Genetics,* 1983,

104(3):449-56.

Hedrick PW, 1987 "Gametic disequilibrium measures: proceed with caution". *Genetics* 117: 331–41.

Hedrick PW. "Balancing selection and MHC". *Genetica*. 1999, 104:207–214.

Hedrick PW. "Pathogen resistance and genetic variation at MHC *loci*". *Evolution*, 2002, 56(10), pp. 1902-8.

Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodrigues-Botigué L, Ramachandran S, Hon L, Brisbin A, Lin AA, Underhill PA, Comas D, Kidd KK, Norman PJ, Parham P, Bustamante CD, Mountain JL, Feldman MW. "Hunter-gatherer genomic diversity suggests a southern African origin for modern humans". *Proc Natl Acad Sci USA*. 2011, 108:5154-62.

Hughes AL, Nei M. "Pattern of nucleotide substitution at major histocompatibility complex class I *loci* reveals overdominant selection". *Nature*.1988 Sep 8;335(6186):167-70.

Hughes AL. "Natural selection and the diversification of vertebrate immune effectors." *Immunological reviews*. 2002, 190: 161-8.

Huttley GA; Smith MW; Carrington M; O'Brien SJ "A scan for linkage disequilibrium across the human genome". *Genetics*, 1999, 152(4):1711-22.

IBGE [On line]. Disponível: <http://www.ibge.gov.br/brasil500/index2.html> [capturado em 10 de setembro de 2011].

Ingman M, Kaessmann H, Paabo S, Gyllensten U. "Mitochondrial genome variation and the origin of modern humans". *Nature*. 2000, 408:708–13.

Instituto Socio Ambiental [On line]. Disponível: www.socioambiental.org [capturado em 10 de setembro de 2011].

Jablonski NG. "The evolution of human skin and skin color". *Annu.Rev. Anthropol.* 2004, 33:585–623.

Jakobsson M, Rosenberg NA. "Clumpp: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure". *Bioinformatics*. 2007, 23(14): 1801-06.

Jobling MA, Hurles M, Tyler-Smith C (2004) *Human evolutionary genetics: origins, people e disease*. Garland Science, New York.

Jost L. " G_{ST} and its relatives do not measure differentiation". *Molecular Ecology*. 2008, 17:4015–26.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR and Hammer MF. "Ancestral Asian source(s) of New World Y-chromosome founder haplotypes". *Am J Hum Genet.* 1999, 64:817-31.

Keyeux G, Rodas C, Gevez N, Carter D. "Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations". *Hum. Biol.* 2002, 74:211-33.

Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB. "Signatures of population expansion in microsatellite repeat data". *Genetics.* 1998, 148:1921-30.

King JP, Kimmel M, Chakraborty R. "A power analysis of microsatellite based statistics for inferring past population growth". *Mol. Biol. Evol.* 2000, 17(12):1859-2000.

Klein J, Satta Y, Takahata N, O'hUigin C. "Trans-specific Mhc polymorphism and the origin of species in primates". *J Med Primatol.* 1993, 22(1):57-64.

Kwiatkowski DP. "How malaria has affected the human genome and what human genetics can teach us about malaria". *Am. J. Hum. Genet.* 2005, 77: 171–92.

Lahiri DK e Nurenberg Jr JL. "A rapid non-enzymatic method for the preparation of HMW DANN from blood for RFLP studies". *Nucl Acid Res.* 1991, 19:5444.

Lancaster AK, Single RM, Solberg OD, Nelson MP, Thomson G. "PyPop update - a software pipeline for large-scale multi/*locus* population genomics" *Tissue Antigens*, 2007 69 (s1), 192-97.

Lell JT, Schurr TG, Sukernik RI, Starikovskaya YB, Torroni A, Moore LG, Troup GM, Wallace DC. "Y chromosome polymorphisms in Native American and Siberian populations: Identification of Native American Y chromosome haplotypes". *Hum Genet.* 1997,100:536-43.

Lell JT, Sukernik RI, Starikovskaya YB, Jin L, Su B, Schurr TG, Underhill P, Wallace DC. "The dual origins and Siberian affinities of Native American Y chromosomes". *Am J Hum Genet.* 2002, 70:192–206.

Lewontin RC.(1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Malkki, M.; Single, R.; Carrington, M.; Thomson, G.; Petersdorf, E. "MHC microsatellite diversity and linkage disequilibrium among common HLA-A, HLA-B, DRB1 haplotypes: implications for unrelated donor hematopoietic transplantation and disease association studies". *Tissue Antigens*, 2005, 66:114-124.

Mann CC (2007) 1491 Novas revelações das Américas antes de Colombo. Tradução:

Renato Aguiar. Rio de Janeiro: Objetiva. 471p.

Mendes Junior CT (2005) DNA mitocondrial na Amazônia Brasileira: Estrutura genética regional e inferências continentais. Universidade de São Paulo. Tese de Doutorado

Merriwether DA, Rothhammer F, Ferrell RE. "Distribution of the four founding lineage haplotype in Native Americans suggests a single wave of migration for the New World". *Am J Phys Anthropol.* 1995, 98:411-30.

Mesa NR, Mondragon MC, Soto ID, Parra MV, Duque D, Ortiz-Barrientos D, *et al.* "Autosomal, mtDna, and y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America". *Am j hum genet* 2000; 67: 1277-86.

Meyer, D, RM Single, SJ Mack, HA Erlich, and G Thomson. 2006. "Signatures of demographic history and natural selection in the human major histocompatibility complex *loci*." *Genetics* 173 (4): 2121-42.

Neves WA, Prous A, González-José R, Kipnis R, Powell J."Early Human skeletal Remains from Santana do Riacho, Brazil: implications for the settlement of the New World", *Journal of Human Evolution.* 2003, vol. 45: 19-42.

Neves EG, 2006. *Arqueologia da Amazônia.* Rio de Janeiro: Jorge Zahar. 86p

Neves EG, Petersen JB. 2008. "Political economy and pre-Colombian landscape transformations in Central Amazonia". In: Silverman H, Isbell W.H. editors. *Handbook of South American Archaeology.* New York: Springer. pp279:309

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. "Genomic scans for selective sweeps using SNP data". *Genome Res.* 2005, 15(11):1566-75.

Noelli FS. (2008) "The Tupi expansion". In: Silverman H, Isbell W.H. editors. *Handbook of South American Archaeology.* New York: Springer. pp. 659:670.

Oliveira, LA. (2007). Diversidade populacional de microssatélites do complexo principal de histocompatibilidade (MHC) em populações indígenas do Paraná e do Mato Grosso do Sul. Universidade Federal do Paraná. Dissertação de Mestrado.

Parham P, Arnett KL, Adams EJ, Little AM, Tees K, Barber LD, Marsh SG, Ohta T, Markow T, Petzl-Erler ML. "Episodic evolution and turnover of *HLA-B* in the indigenous human populations of the Americas". *Tissue Antigens.* 1997, 50(3):219-32.

Pena SDJ, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, Munkhtuja B, Oyunsuren T. "A major founder Y-chromosome haplotype in Amerindians". *Nat Genet.*1995, 11:15–16.

Petzl-Erler ML, Sotomaior LVS. "The HLA polymorphism of two distinctive South-American Indian tribes: the Kaingang and the Guarani". *Tiss. Antig.* 1993, 41:227-37.

Pompanon F, Bonin A, Taberlet P. "Genotyping errors: causes, consequences and Solutions". *Nature.* 2005, 6;847-59.

Pritchard JK, Wen X, Falush D. Structure software: version 2.3. 2010. Disponível em: <http://pritch.bsd.uchicago.edu/structure.html>

Prugnolle F.; Manica A.; Balloux F. "Geography predicts neutral genetic diversity of human populations". *Current Bio.* 2005. 15:59-60.

Pucciarelli HM, Neves WA, González-José R, Saedi ML, Rozzi FR, Sruck A, Bonilla MY. "East-West cranial differentiation in pre-Columbian human populations of South America". *J Hom Comp Hum Biol.* 2006. 57:133-50.

Ramachandran S., Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. "Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa". *PNA.* 2005, 102:15942-47.

Ribeiro-Santos AKC, Guerreiro JF, Santos SEB, Zago MA. "The Split of the Arara Population: Comparison of Genetic Drift and Founder Effect". *Hum Hered.* 1999, 51:79-84.

Rodrigues-Delfin LA, Rubin de Celis VE, Zago MA. "Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: data from Y chromosome and mitochondrial DNA". *Hum Hered.* 2001, 51:97-106.

Rogers RA; Rogers LA; Martin LD. "How the door opened: the peopling of the new world". *Human Molecular Genetics.* 1995, 4:1869-74.

Ronald J, Akey JM. "Genome-wide scans for *loci* under selection in humans". *Hum Genomics* 2005 Jun;2(2):113-25.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA e Feldman MW. "Genetic structure of human populations". *Science.* 2002, 298:2381-5.

Rosenberg NA. "Distruct: a program for the graphical display of population structure". *Molecular Ecology Notes.* 2004, 4:137-8.

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, *et al.* "Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure". *PLoS Genet.* 2005,1(6): e70.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, *et al.* "Genome-wide detection and characterization of positive selection in human populations". *Nature*. 2007, 18;449(7164):913-8

Salzano FM, Callegari-Jacques SM (1988). *South American Indians: a case study in evolution*. Oxford University Press, New York.

Salzano FM. "Molecular variability in Amerindians: widespread but uneven information". *An Acad Bras Cienc*. 2002, 74(2):223-63.

Salzano FM, Callegari-Jacques SM. "Amerindian and Amerindian autosome molecular variability – a test analysis". *Genetica*. 2006, 126:237-42.

Salzano FM. *The prehistoric Colonization of the Americas*. In: *Antropological Genetic*. Cambridge: Cambridge University Press, 2007.p. 433-55.

Sambrook J, Fritsch EF, Maniatis T. (1989). *Molecular cloning – a laboratory manual*. 2ed. Cold Sping Harbor Laboratory Press.

Sanchez-Mazas, A. "An apportionment of human HLA diversity." *Tissue Antigens*. 2007, 1 (8): 198-202.

Santos SEB, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AKC, Hutz MH, Tovo-Rodrigues L, Salzano FM, Callegari-Jacques SM. "Autosomal STR Analyses in Native Amazonian Tribes Suggest a Population Structure Driven by Isolation by Distance". *Hum Bio*. 2009: 81(1):77-88.

Schroeder KB, Schurr TG, Long JC, Rosemberg NA, Crawford MH, Tarskaia LA, Osipova LP, Zhadanov SI, Smith DG. "A private allele ubiquitous in the Americas". *Biol Lett*. 2007, 3(2): 218–223.

Schroeder KB, Jakobsson M, Crawford MH, Schurr TG, Boca SM, Conrad DF. "Haplotypic background of a private allele at high frequency in the Americas". *Mol. Biol. Evol*. 2009, 26:995-1016.

Schurr TG, Sherry ST. "Mitochondrial DNA and Y chromosome diversity and the peopling of the American evolutionary and demographic evidence". *Am J Hum Biol*. 2004,16:420-39.

Slatkin M. "Linkage disequilibrium--understanding the evolutionary past and mapping the medical future". *Nat Rev Genet*. 2008,9(6):477-85.

Solberg OD, Mack SJ, Lancaster AK, Single RM, Sanchez-Mazas A, Tsai Y, and Thomson G *Balancing selection and heterogeneity across the classical human leukocyte antigen loci: A meta-analytic review of 497 population studies*. *Human Immunology*, 2008; 69(7): 443-464.

Stephens M, Donnelly P. "A comparison of bayesian methods for haplotype reconstruction from population genotype data". *Am J Hum Genet.* 2003, 73(5):1162-9.

Stone AC, Stoneking M. "Ancient DNA from a pre-Colombian Amerindian population". *Am J Phys Anthropol.* 1993, 92:463-71.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ *et al.* "Beringian standstill and spread of Native American founders". *PLoS ONE.* 2007: 9e829.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, Stefano GFD, Labarga CM, Rickards O, Tyler-Smith C, Pena SDJ, Santos FR. "Genetic differentiation in south amerindians is related to environmental and cultural diversity: evidence from the Y chromosome". *Am. J. Hum. Genet.* 2001, 68:1485-96.

Tishkoff SA, Kidd KK. "Implications of biogeography of human populations for 'race' and medicine". *Nat Genet.* 2004 Nov;36(11 Suppl):S21-7.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman, JS, Powell K., Mortensen HM, Hirbo JB, Osman M. *et al.* "Convergent adaptation of human lactase persistence in Africa and Europe". *Nat. Genet.* 2007, 39:31-40.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM. "The genetic structure and history of Africans and African Americans". *Science.* 2009, 324 (5930):1035-44.

Torrini A, Nell JV, Barrantes R, Schurr TG, Wallace DC. "Mitochondrial DNA 'clock' for Amerindian and its implications for timing their entry into North American". *Proc. Nat. Ac. Sci. USA.* 1994. 91:1158-62.

Trachtenberg, E; Korber, B; Sollars, C; Kepler, TB; Hraber, PT; Hayers E.; *et al.* "Advantage of rare HLA supertype in HIV disease progression". *Nat. Med.* 2003, 9(7):928-35.

Tsuneto LT, Probst CM, Hutz MH, Salzano FM, Rodrigues-Delfin LA, Zago MA, Hill K, Hurtado AM, Riberio-dos Santos AKC, Petzl-Erler ML. "HLA class II diversity in seven Amerindian populations. Clues about the origins of the Ache". *Tiss Antig.* 2003, 62:512-26.

Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL. "A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history". *Proc Natl Acad Sci USA.* 1996, 93:196-200.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir

B, Bertranpetit, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ. "Y chromosome sequence variation and the history of human populations". *Nat Genet.* 2000, 26:358–61.

Villadangos JA, Schnorrer P. "Intrinsic and cooperative antigen-presenting functions of dendritic-cell subsets in vivo". *Nat Rev Immunol.* 2007,7(7):543-55.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. "A map of recent positive selection in the human genome". *PLoS Biol.* 2006 Mar;4(3):e72. Epub 2006 Mar 7. Erratum in: *PLoS Biol.* 2006 Apr;4(4):e154. *PLoS Biol.* 2007 Jun;5(6):e147.

Vuilleumier SB. Pleistocene changes in flora and fauna of South America. *Science.* 1971, 173:771-80.

Wallace DC. "Mitochondrial DNA variation in human evolution, degenerative disease, and aging". *Am J Hum Genet.* 1995, 57:201-23.

Walters MR, Stafford Jr TW. "Redefining the age of Clovis: implications from peopling of the Americas". *Science.* 2007, 315:1122-26.

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. "Genetic variation and population structure in native Americans". *PLoS Genet.* 2007,3(11).

Watkins DI, McAdam SN, Liu X, Strang CR, Milford EL, Levine CG, Garber TL, Dogon AL, Lord CI, Ghim SH, *et al.* "New recombinant *HLA-B* alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I *loci*". *Nature.* 1992 May 28;357(6376):329-33

Zegura SL, Karafet TM, Zhivotovskiy LA, Hammer MF. "High-Resolution SNPs and Microsatellites Haplotype Point to a Single, Recent Entry of Native American Y Chromosomes into the Americas". *Mol Biol Evol.* 2004, 21:164-75.

Zhivotovskiy LA, Rosenberg N, Feldman M. "Features of evolution and expansion of modern humans, inferred from genomewide microsatellites markers". *Am J Hum Gen.* 2003, 72:1171-86.

ANEXO 1

Resultados complementares

Tabela A.1. Indivíduos excluídos das análises de miscigenação.

ID	POP	Probabilidade	ID	POP	Probabilidade
2385	Chipewyan	0.005	2440	Ojbwua	0.107
2387	Chipewyan	0.360	861	Maya	0.496
2391	Chipewyan	0.381	2021	Cabecar	0.459
2393	Chipewyan	0.017	2026	Cabecar	0.087
2394	Chipewyan	0.489	2028	Cabecar	0.242
2395	Chipewyan	0.019	2571	Embera	0.000
2397	Chipewyan	0.062	2460	Zenu	0.143
2455	Chipewyan	0.291	2575	Arhuaco	0.022
2520	Chipewyan	0.080	2732	Arhuaco	0.490
2800	Chipewyan	0.027	2525	Wayuu	0.406
2401	Cree	0.225	2527	Wayuu	0.000
2403	Cree	0.012	2535	Wayuu	0.468
2405	Cree	0.164	2538	Wayuu	0.018
2406	Cree	0.208	2582	Waunana	0.000
2407	Cree	0.102	2771	Quechua	0.358
2408	Cree	0.000	2774	Quechua	0.401
2413	Cree	0.281	2126	Huilliche	0.355
2414	Cree	0.001	2128	Huilliche	0.226
2418	Cree	0.421	2130	Huilliche	0.027
2424	Ojbwua	0.004	2132	Huilliche	0.474
2425	Ojbwua	0.179	2136	Huilliche	0.348
2428	Ojbwua	0.264	2428	Ojbwua	0.264
2435	Ojbwua	0.382	2435	Ojbwua	0.382

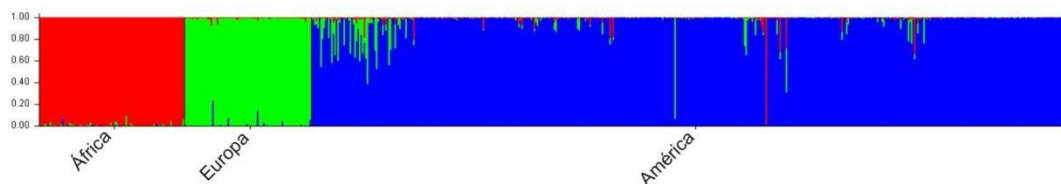


Figura A.1: Grau de miscigenação nas populações parentais

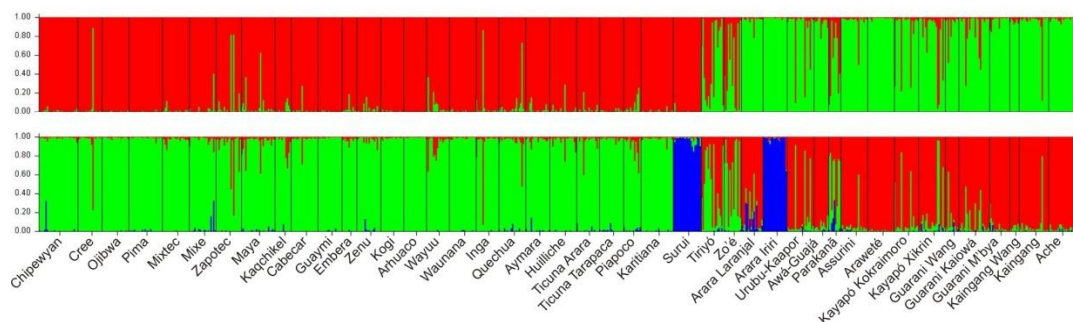


Figura A.2. Estruturação populacional com K=2 e K=3 com as populações Guarani e Kaingang duplamente genotipadas pelo presente estudo e por Wang e colaboradores. Sendo as populações 89 Guarani genotipadas por Wang, 90 e 91 Guarani genotipadas no presente estudo; 92 Kaingang genotipada por Wang e 93 Kaingang genotipada no presente estudo. Em K=3 as populações em azul representam Surui e Arara do Iriri.

Tabela A.2. Cenários alternativos de estruturação populacional e a frequência que ocorreram (mostrando apenas frequências acima de 8%)

K	Estruturação	%
2	AM/LAMZ	92
2	AM/SUR, ARI, ACH	2
3	AM/TIT, KAR, SUR/LAMZ	32
3	AM/LAMZ/ARI, ACH	12
3	AM/ TIT, KAR, SUR, ACH/LAMZ	12
3	AM/SUR, ARI, ACH/LAMZ	20
4	AM/TIR, KAR, SUR/LAMZ/ARI, ACH	24
4	AM/PIM, ACH/ TIT, KAR, SUR/LAMZ	16
4	AM/TIT, KAR, SUR/LAMZ/ARI	16
4	AM/TIT, KAR, SUR, ACH/LAMZ/ARI	12
5	AM/PIM/TIT, KAR, SUR/LAMZ/ARI, ACH	16
5	AM/TIT, KAR/ SUR, ACH/ ARI/LAMZ	8
5	AM/PIM, SUR/ TIT, KAR, ACH/ARI/LAMZ	8
6	AM/CAB,GUY,KOG,ARH/ TIC, KAR, SUR/ ARI/ACH/LAMZ	12
6	AM/PIM/TIC, KAR/SUR/ARI/ACH/LAMZ	8
6	AM/PIM/CAB,GUY,KOG,ARH/TIT,KAR,SUR,ACH/ARI/LAMZ	8
7	AM/PIM/CAB, GUY,KOG,ARH/TIT, KAR/SUR/ARI,ACH/LAMZ	12
8	AM/PIM/CAB,GUY,KOG,ARH/TIT,KAR/SUR/ARI/ACH/LAMZ	16

continua

Tabela A.2 (continuação). Cenários alternativos de estruturação populacional e a frequência que ocorreram (mostrando apenas frequências acima de 8%).

K	Estruturação	%
9	AM/PIM/CAB,GUY,KOG,ARH/TIT, KAR/SUR/ARI/AWE/ACH/LAMZ	10
9	AM/AN/PIM/CAB,GUY, KOG,ARH/TIT, KAR/SUR/ARI/AWE/ACH/LAMZ	16
9	AM/AN/PIM/CAB,GUY,KOG,ARH,WAY/TIT,KAR/SUR/ARI/AWE/ACH/LAMZ	8
9	AM/CHI/PIM/CAB,GUY,KOG,ARH/TIT/KAR/SUR/ARI/ACH/LAMZ	8
9	AM/PIM/CAB,GUY,KOG,ARH,WAY/TIT,KAR/SUR/ARI/AWE/ACH/LAMZ	8
10	AM/AN/PIM/CAB,GUY/KOG,ARH,WAY/TIT,KAR/SUR/ARI/ACH/LAMZ	20
10	AM/PIM/CAB,GUY/KOG,ARH, WAY/TIT/KAR/SUR/ARI/ACH/LAMZ	12
10	AM/AN/PIM/CAB,GUY,KOG,ARH/TIT,KAR/SUR/ARI/AWE,GUA,KAI/ACH/LAMZ	8
10	AN/PIM/CAB,GUY,KOG,ARH/TIT,KAR/SUR/ZOE/ARI,ACH/AWE/LAMZ	8
11	AM/CHI/PIM/CAB,GUY/KOG,ARH,WAY/TIT,KAR/SUR/ZOE/ARI/ACHE	8
11	AM/AN/PIM/CAB,GUY,KOG,ARH/TIC,KAR/SUR/WAY,ZOE/ARI/AWE/ACH/LAMZ	20
11	AM/AN/PIM/CAB,GUY/KOG,ARH/ZOE,WAY/TIT,KAR/SUR/ARI/AWE/ACH/LAMZ	8
11	AM/PI/CAB,GUY,KOG,ARH/ZOE,WAY/TIT/KAR/SUR,ARI/KAA,AWA/AWE/ACH/LAMZ	8
12	AM/CHI/PIM/CAB,GUY,KOG,ARH/WAY,ZOE/TIT/KAR/SUR/ARI/AWE/ACHE/LAMZ	20
12	AM/AN/PIM/CAB,GUY/KOG, ARH/TIT,KAR,SUR/ZOE/ARI/AWA,KAA/AWE/ACH/LAMZ	12
12	AM/AN/PIM/CAB,GUY/KOG,ARH/TIT/KAR/SUR/ARI/WAY,ZOE/AWE/ACH/LAMZ	16
13	AM/AN/PIM/CAB,GUY/KOG/ARH/TIT/KAR/SUR/ARI/AWA,KAA/AWE,GUA,KAI/ACH/LAMZ	12
13	AM/AN/PIM/CAB,GUY/KOG/ARH/ZOE,WAY/TIC,KAR/SUR/ARI/AWA,KAA/AWE/ACH/LAMZ	12
13	AM/AN/PIM/CAB,GUY/KOG,ARH/WAY/TIC,KAR/SUR/ARI/ZOE/AWE/ACH/LAMZ	16
14	AM/CHI/CRE,OJB/PIM/CAB,GUY,KOG,ARH/WAY/TIT/KAR/SUR/ARI/AWA,KAA/AWE/ACH/LAMZ	16
14	AM/AN/PIM/CAB,GUY/KOG,ARH/WAY/TIC,KAR/SUR/ZOE/ARI/AWE/ACH/LAMZ	8
14	AM/CHI/PIM/CAB,GUY/KOG,ARH/WAY,ZOE/TIT/KAR/SUR/ARI/KAA,AWA/AWE/ACH/LAMZ	8
14	AM/CHI/CRE,OJB/PIM/KOG,ARH/CAB,GUY/WAY/TIT/KAR/SUR/ZOE/ARI/AWE/ACH/LAMZ	8

Legenda de abreviaturas: AM – América, AN – América do Norte, LAMZ – Leste da Amazônia, ACH – Ache, ARI – Arara do Iriri, ARH – Arhuaco, AWA – Awá-Gujá, AWE Araweté, CAB – Cabecar, CHI – Chipewyan, CRE – Cree, GUA –Guarani, GUY – Guaymi, KAA – Urubu-Kaapor, KAI – Kaingang, KAR – Karitiana, KOG – Kogi, OJB – Ojibwa, PIM – Pima, SUR – Surui, TIC – Ticuna, TIT – Ticuna Tarapaca, WAY – Wayuu, ZOE – Zo’é.

Tabela A.3. Teste Exato de Fisher entre as populações para o locus D6S276.

POP	OVI	POR	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WAW	ING	AYM	HUI	TIA	TIT	TIR	ZOE	ARL	ARI	KAA	AWA	PRK	ASS	AWE	KAK	KAX		
POR	0,00																															
CHI	0,00	0,60																														
CRE	0,00	0,03	0,55																													
OJI	0,00	0,03	0,17	0,23																												
KAQ	0,00	0,00	0,08	0,07	0,01																											
CAB	0,00	0,03	0,51	0,01	0,00	0,02																										
GUA	0,00	0,00	0,09	0,56	0,05	0,01	0,00																									
BEM	0,00	0,28	0,74	0,81	0,10	0,02	0,05	0,38																								
ZEN	0,00	0,00	0,03	0,26	0,03	0,01	0,00	0,84	0,10																							
KOG	0,01	0,06	0,23	0,10	0,24	0,01	0,00	0,02	0,06	0,01																						
ARH	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,11																						
WAY	0,00	0,53	0,98	0,29	0,04	0,30	0,74	0,04	0,47	0,02	0,05	0,00																				
WAW	0,00	0,00	0,01	0,01	0,00	0,19	0,03	0,00	0,00	0,00	0,00	0,06	0,06																			
ING	0,00	0,14	0,16	0,16	0,02	0,00	0,00	0,03	0,70	0,00	0,01	0,00	0,08	0,00																		
AYM	0,00	0,02	0,51	0,55	0,02	0,46	0,04	0,07	0,31	0,04	0,02	0,00	0,74	0,07	0,03																	
HUI	0,00	0,02	0,54	0,55	0,09	0,17	0,02	0,33	0,36	0,24	0,01	0,00	0,68	0,00	0,02	0,58																
TIA	0,00	0,01	0,20	0,21	0,06	0,27	0,01	0,17	0,12	0,43	0,05	0,00	0,23	0,00	0,01	0,15	0,29															
TIT	0,00	0,00	0,00	0,00	0,00	0,17	0,00	0,00	0,00	0,00	0,00	0,05	0,00	0,00	0,01	0,00	0,02															
TIR	0,00	0,00	0,02	0,18	0,08	0,06	0,00	0,13	0,03	0,09	0,10	0,00	0,01	0,00	0,00	0,12	0,09	0,05	0,00													
ZOE	0,00	0,00	0,00	0,01	0,00	0,25	0,00	0,01	0,00	0,11	0,00	0,00	0,01	0,00	0,00	0,02	0,05	0,51	0,01	0,03												
ARL	0,00	0,00	0,01	0,00	0,00	0,19	0,04	0,00	0,00	0,00	0,00	0,18	0,01	0,00	0,05	0,02	0,02	0,85	0,00	0,02												
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,04	0,01	0,32	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,08	0,00	0,03	0,00											
KAA	0,00	0,00	0,28	0,29	0,00	0,71	0,08	0,01	0,16	0,01	0,00	0,00	0,59	0,39	0,00	0,86	0,21	0,21	0,03	0,01	0,02	0,07	0,00									
AWA	0,00	0,00	0,01	0,14	0,07	0,00	0,00	0,35	0,08	0,62	0,01	0,00	0,00	0,00	0,00	0,00	0,04	0,21	0,00	0,01	0,02	0,00	0,45	0,00								
PRK	0,00	0,40	0,68	0,48	0,06	0,09	0,03	0,20	0,60	0,04	0,09	0,73	0,00	0,43	0,66	0,47	0,10	0,01	0,23	0,00	0,02	0,00	0,22	0,00								
ASS	0,00	0,02	0,49	0,27	0,00	0,00	0,14	0,01	0,82	0,00	0,00	0,40	0,00	0,12	0,11	0,07	0,01	0,00	0,00	0,00	0,00	0,00	0,07	0,00	0,21							
AWE	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00						
KAK	0,00	0,00	0,05	0,01	0,00	0,00	0,03	0,00	0,06	0,00	0,00	0,05	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,11	0,00					
KAX	0,00	0,00	0,04	0,10	0,00	0,00	0,01	0,01	0,17	0,01	0,00	0,02	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,05	0,00	0,19	0,00	0,08				

Tabela A.5. Teste Exato de Fisher entre as populações para o locus D6S2972

POP	PO	VI	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WOU	ING	AYM	HUI	TIA	TIT	ZOE	ARL	ARI	AAA	AWA	PRK	ASS	AWE	KAK	KAX
POR	0,55																												
CHI	0,00	0,00																											
CRE	0,00	0,00	0,20																										
OJI	0,00	0,00	0,04	0,56																									
KAQ	0,00	0,00	0,37	0,88	0,67																								
CAB	0,00	0,00	0,00	0,34	0,28	0,18																							
GUA	0,00	0,00	0,30	0,46	0,04	0,36	0,05																						
BEM	0,00	0,00	0,23	0,03	0,00	0,03	0,00	0,29																					
ZEN	0,00	0,00	0,00	0,19	0,71	0,23	0,32	0,01	0,00																				
KOG	0,01	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00																			
ARH	0,06	0,22	0,00	0,03	0,06	0,02	0,02	0,01	0,02	0,01	0,14																		
WAY	0,01	0,06	0,01	0,24	0,74	0,24	0,15	0,06	0,01	0,21	0,02	0,49																	
WOU	0,00	0,00	0,14	0,07	0,00	0,06	0,01	0,62	0,55	0,00	0,00	0,00	0,00																
ING	0,01	0,01	0,01	0,00	0,01	0,00	0,07	0,67	0,00	0,13	0,02	0,11																	
AYM	0,00	0,00	0,95	0,14	0,02	0,21	0,00	0,52	0,84	0,00	0,00	0,01	0,02	0,46	0,14														
HUI	0,00	0,00	0,19	0,23	0,41	0,54	0,00	0,06	0,01	0,03	0,00	0,00	0,22	0,00	0,00	0,05													
TIA	0,00	0,02	0,01	0,32	1,00	0,35	0,25	0,02	0,00	0,51	0,01	0,21	0,96	0,00	0,01	0,18													
TIT	0,02	0,02	0,01	0,42	0,72	0,25	0,87	0,07	0,01	0,42	0,00	0,24	0,67	0,01	0,01	0,68													
TIR	0,57	0,65	0,08	0,07	0,10	0,07	0,02	0,12	0,46	0,01	0,12	0,61	0,41	0,05	0,66	0,24	0,06	0,15	0,16										
ZOE	0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,14	0,01	0,00	0,00										
ARL	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,01	0,00	0,01	0,01	0,00	0,00	0,12	0,01									
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00								
AAA	0,00	0,00	0,02	0,03	0,00	0,02	0,00	0,25	0,75	0,00	0,00	0,05	0,01	0,50	0,68	0,27	0,00	0,00	0,38	0,00	0,00	0,00							
AWA	0,00	0,00	0,03	0,01	0,03	0,02	0,00	0,00	0,01	0,00	0,00	0,01	0,04	0,00	0,00	0,03	0,18	0,03	0,00	0,08	0,17	0,76	0,00	0,00					
PRK	0,00	0,00	0,04	0,20	0,74	0,38	0,06	0,02	0,00	0,46	0,00	0,01	0,23	0,00	0,00	0,01	0,63	0,41	0,14	0,03	0,09	0,00	0,00	0,04					
ASS	0,00	0,00	0,05	0,00	0,02	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,13	0,01	0,00	0,01	0,37	0,05	0,00	0,00	0,46	0,12			
AWE	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
KAK	0,00	0,00	0,12	0,05	0,14	0,12	0,00	0,01	0,02	0,00	0,00	0,02	0,17	0,00	0,01	0,10	0,24	0,14	0,02	0,25	0,07	0,37	0,00	0,00	0,79	0,14	0,49	0,00	0,00
KAX	0,00	0,00	0,31	0,06	0,04	0,17	0,00	0,03	0,10	0,00	0,00	0,00	0,03	0,01	0,01	0,32	0,26	0,03	0,00	0,12	0,02	0,22	0,00	0,00	0,39	0,06	0,10	0,00	0,56

Tabela A.6. Teste Exato de Fisher entre as populações para o locus D6S265

POP	PO	VI	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WOU	ING	AYM	HUI	TIA	TIT	ZOE	ARL	ARI	CAA	AWA	PRK	ASS	AWE	KAK	KAX
POR	0,55																												
CHI	0,00	0,00																											
CRE	0,00	0,00	0,20																										
OJI	0,00	0,00	0,04	0,56																									
KAQ	0,00	0,00	0,37	0,88	0,67																								
CAB	0,00	0,00	0,00	0,34	0,28	0,18																							
GUA	0,00	0,00	0,30	0,46	0,04	0,36	0,05																						
BEM	0,00	0,00	0,23	0,03	0,00	0,03	0,00	0,29																					
ZEN	0,00	0,00	0,00	0,19	0,71	0,23	0,32	0,01	0,00																				
KOG	0,01	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00																			
ARH	0,06	0,22	0,00	0,03	0,06	0,02	0,02	0,01	0,02	0,01	0,14																		
WAY	0,01	0,06	0,01	0,24	0,74	0,24	0,15	0,06	0,01	0,21	0,02	0,49																	
WOU	0,00	0,00	0,14	0,07	0,00	0,06	0,01	0,62	0,55	0,00	0,00	0,00																	
ING	0,01	0,01	0,01	0,00	0,01	0,00	0,07	0,67	0,00	0,00	0,13	0,02	0,11																
AYM	0,00	0,00	0,95	0,14	0,02	0,21	0,00	0,52	0,84	0,00	0,00	0,01	0,02	0,46	0,14														
HUI	0,00	0,00	0,19	0,23	0,41	0,54	0,00	0,06	0,01	0,03	0,00	0,00	0,22	0,00	0,00	0,05													
TIA	0,00	0,02	0,01	0,32	1,00	0,35	0,25	0,02	0,00	0,51	0,01	0,21	0,96	0,00	0,01	0,18													
TIT	0,02	0,02	0,01	0,42	0,72	0,25	0,87	0,07	0,01	0,42	0,00	0,24	0,67	0,01	0,01	0,01	0,02	0,68											
TIR	0,57	0,65	0,08	0,07	0,10	0,07	0,02	0,12	0,46	0,01	0,12	0,61	0,41	0,05	0,66	0,24	0,06	0,15	0,16										
ZOE	0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,14	0,01	0,00	0,00									
ARL	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,01	0,00	0,01	0,01	0,00	0,00	0,12	0,01									
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00								
CAA	0,00	0,00	0,02	0,03	0,00	0,02	0,00	0,25	0,75	0,00	0,00	0,05	0,01	0,50	0,68	0,27	0,00	0,00	0,02	0,38	0,00	0,00							
AWA	0,00	0,00	0,03	0,01	0,03	0,02	0,00	0,00	0,01	0,00	0,00	0,01	0,04	0,00	0,00	0,03	0,18	0,03	0,00	0,08	0,17	0,76	0,00	0,00					
PRK	0,00	0,00	0,04	0,20	0,74	0,38	0,06	0,02	0,00	0,46	0,00	0,01	0,23	0,00	0,00	0,01	0,63	0,41	0,14	0,03	0,09	0,00	0,00	0,00	0,04				
ASS	0,00	0,00	0,05	0,00	0,02	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,13	0,01	0,00	0,01	0,37	0,05	0,00	0,00	0,46	0,12			
AWE	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		
KAK	0,00	0,00	0,12	0,05	0,14	0,12	0,00	0,01	0,02	0,00	0,00	0,02	0,17	0,00	0,01	0,10	0,24	0,14	0,02	0,25	0,07	0,37	0,00	0,00	0,79	0,14	0,49	0,00	
KAX	0,00	0,00	0,31	0,06	0,04	0,17	0,00	0,03	0,10	0,00	0,00	0,00	0,03	0,01	0,01	0,32	0,26	0,03	0,00	0,12	0,02	0,22	0,00	0,00	0,39	0,06	0,10	0,00	0,56

Tabela A.7. Teste Exato de Fisher entre as populações para o locus D6S2885

POP	OVI	POR	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WAU	ING	AYM	HUI	TIA	TIT	ZOE	ARL	AAA	AWA	PRK	ASS	AWE	KAK	KAX	
		0,00																											
CHI	0,00	0,20																											
CRE	0,00	0,02	0,08																										
OJI	0,00	0,06	0,57	0,01																									
KAQ	0,00	0,00	0,00	0,06	0,00																								
CAB	0,00	0,00	0,02	0,00	0,15	0,01																							
GUA	0,00	0,00	0,00	0,00	0,03	0,00	0,28																						
BEM	0,03	0,10	0,19	0,63	0,11	0,07	0,01	0,00																					
ZEN	0,00	0,00	0,00	0,02	0,00	0,57	0,00	0,00	0,02																				
KOG	0,00	0,00	0,04	0,00	0,40	0,00	0,66	0,70	0,00	0,01																			
ARH	0,00	0,03	0,16	0,00	0,16	0,02	0,02	0,00	0,02	0,16	0,13																		
WAY	0,00	0,17	0,35	0,17	0,12	0,15	0,01	0,00	0,43	0,41	0,02	0,63																	
WAU	0,00	0,00	0,00	0,00	0,00	0,37	0,03	0,01	0,01	0,14	0,01	0,01	0,03																
ING	0,00	0,04	0,15	0,09	0,41	0,05	0,03	0,00	0,32	0,04	0,01	0,06	0,31	0,01															
AYM	0,00	0,00	0,00	0,00	0,07	0,00	0,05	0,01	0,00	0,00	0,07	0,00	0,00	0,00	0,04														
HUI	0,00	0,00	0,00	0,25	0,00	0,02	0,00	0,00	0,26	0,00	0,00	0,00	0,01	0,00	0,00	0,00													
TIA	0,00	0,00	0,00	0,01	0,00	0,19	0,00	0,00	0,01	0,06	0,00	0,00	0,01	0,01	0,03	0,00	0,01												
TIT	0,00	0,00	0,00	0,00	0,06	0,08	0,10	0,01	0,01	0,06	0,05	0,02	0,03	0,19	0,17	0,26	0,00	0,04											
TIR	0,01	0,07	0,08	0,69	0,05	0,10	0,00	0,00	0,48	0,04	0,00	0,04	0,37	0,01	0,07	0,00	0,32	0,03	0,01										
ZOE	0,00	0,01	0,01	0,11	0,01	1,00	0,03	0,00	0,13	0,62	0,01	0,05	0,24	0,52	0,09	0,01	0,03	0,20	0,28	0,14									
ARL	0,00	0,00	0,00	0,03	0,00	1,00	0,00	0,00	0,03	0,54	0,00	0,01	0,09	0,40	0,02	0,00	0,00	0,10	0,10	0,03	1,00								
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,24	1,00	0,00	0,00	0,15	0,00	0,00	0,01	0,00	0,02	0,00	0,00	0,01	0,00	0,00	0,00							
KAA	0,00	0,00	0,00	0,00	0,00	0,60	0,00	0,00	0,01	0,15	0,00	0,00	0,01	0,05	0,00	0,00	0,01	0,14	0,00	0,01	0,41	0,33	0,00						
AWA	0,00	0,02	0,52	0,09	0,74	0,01	0,76	0,04	0,25	0,01	0,36	0,06	0,27	0,03	0,23	0,05	0,00	0,00	0,13	0,15	0,04	0,01	0,02	0,00					
PRK	0,00	0,01	0,04	0,07	0,14	0,30	0,37	0,03	0,09	0,30	0,08	0,15	0,26	1,00	0,17	0,08	0,00	0,04	0,65	0,08	0,45	0,33	0,04	0,07	0,31				
ASS	0,00	0,00	0,00	0,01	0,02	0,00	0,00	0,00	0,04	0,00	0,00	0,00	0,01	0,00	0,37	0,01	0,00	0,01	0,01	0,01	0,00	0,00	0,00	0,01	0,01				
AWE	0,00	0,00	0,00	0,00	0,01	0,00	0,61	1,00	0,00	0,00	0,32	0,00	0,00	0,03	0,00	0,03	0,00	0,08	0,00	0,01	0,00	1,00	0,00	0,18	0,29	0,00			
KAK	0,00	0,00	0,00	0,01	0,00	1,00	0,00	0,00	0,02	0,29	0,00	0,00	0,03	0,13	0,00	0,00	0,11	0,02	0,02	0,77	0,61	0,00	0,64	0,00	0,21	0,00	0,00		
KAX	0,00	0,00	0,00	0,00	0,00	0,51	0,00	0,00	0,01	0,24	0,01	0,01	0,02	0,23	0,01	0,00	0,04	0,05	0,07	0,65	0,56	0,00	0,14	0,01	0,34	0,00	0,00	0,44	

Tabela A.8. Teste Exato de Fisher entre as populações para o locus *HLA-BC*

POP	POI	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WOU	ING	AYM	HUI	TIA	TIT	TIR	ZOE	ARL	ARI	CAA	AWA	PRK	ASS	AWE	KAK	KAX	
POR	0,06																													
CHI	0,00	0,00																												
CRE	0,21	0,08	0,12																											
OJI	0,14	0,00	0,00	0,08																										
KAQ	0,23	0,03	0,01	0,13	0,39																									
CAB	0,00	0,00	0,00	0,00	0,05	0,00																								
GUA	0,00	0,00	0,00	0,00	0,00	0,00	0,00																							
BEM	0,00	0,00	0,00	0,00	0,05	0,01	0,00	0,00																						
ZEN	0,00	0,00	0,09	0,02	0,00	0,01	0,00	0,02	0,00																					
KOG	0,07	0,00	0,00	0,02	0,45	0,31	0,24	0,00	0,01	0,02																				
ARH	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00																			
WAY	0,00	0,00	0,01	0,01	0,00	0,03	0,00	0,00	0,00	0,08	0,01	0,00																		
WOU	0,00	0,00	0,00	0,00	0,01	0,00	0,01	0,00	0,07	0,00	0,07	0,00	0,00																	
ING	0,09	0,00	0,00	0,02	0,07	0,02	0,13	0,00	0,02	0,01	0,24	0,02	0,00	0,24																
AYM	0,12	0,00	0,36	0,63	0,03	0,03	0,07	0,00	0,00	0,02	0,04	0,00	0,00	0,00	0,23															
HUI	0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,01	0,02	0,00														
TIA	0,00	0,00	0,00	0,00	0,01	0,02	0,00	0,00	0,36	0,01	0,13	0,00	0,00	0,06	0,08	0,00	0,13													
TIT	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,01	0,14	0,00	0,03	0,21	0,00	0,00	0,00												
TIR	0,01	0,00	0,00	0,00	0,03	0,03	0,10	0,00	0,00	0,05	0,24	0,00	0,21	0,00	0,05	0,00	0,04	0,02												
ZOE	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00											
ARL	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,00	0,00	0,00	0,01	0,22	0,00	0,00	0,02	0,01	0,01	0,01										
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00									
CAA	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00								
AWA	0,00	0,00	0,00	0,00	0,02	0,01	0,06	0,00	0,13	0,00	0,11	0,00	0,02	0,10	0,01	0,00	0,04	0,25	0,02	0,07	0,00	0,14	0,00	0,00						
PRK	0,00	0,02	0,00	0,02	0,00	0,01	0,01	0,00	0,01	0,02	0,00	0,59	0,01	0,01	0,00	0,00	0,00	0,02	0,02	0,01	0,63	0,03	0,00	0,13						
ASS	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,02	0,01	0,02	0,00	0,00	0,03	0,07	0,00	0,01	0,29	0,00	0,02	0,00	0,11	0,00	0,00	0,01	0,03				
AWE	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
KAK	0,00	0,00	0,00	0,00	0,00	0,00	0,04	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,01	0,02	0,00	0,00	0,00	
KAX	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Tabela A.9. Teste Exato de Fisher entre as populações para o locus D6S2928

POP	OWI	POR	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WOU	ING	AYM	HUI	TIA	TIT	TIR	ZOE	ARL	ARI	AAA	AWA	PRK	ASS	AWE	KAK	KAX			
POR	0,79																																
CHI	0,65	0,26																															
CRE	0,42	0,14	0,59																														
OJI	0,68	0,65	0,64	0,20																													
KAQ	0,38	0,04	0,17	0,32	0,04																												
CAB	0,00	0,00	0,01	0,12	0,00	0,31																											
GUA	0,16	0,01	0,01	0,01	0,01	0,21	0,00																										
BEM	0,05	0,00	0,06	0,14	0,00	0,67	0,54	0,02																									
ZEN	0,09	0,00	0,02	0,05	0,00	1,00	0,05	0,21	0,32																								
KOG	0,00	0,00	0,00	0,01	0,00	0,13	0,36	0,00	0,31	0,02																							
ARH	0,00	0,00	0,00	0,00	0,00	0,03	0,16	0,00	0,28	0,01	1,00																						
WAY	0,41	0,02	0,22	0,28	0,04	1,00	0,08	0,19	0,38	0,77	0,02	0,02																					
WOU	0,12	0,04	0,00	0,00	0,02	0,02	0,00	0,18	0,00	0,01	0,00	0,00	0,01																				
ING	0,00	0,00	0,00	0,03	0,00	0,01	0,05	0,00	0,03	0,00	0,11	0,09	0,00	0,00																			
AYM	0,11	0,04	0,01	0,00	0,06	0,01	0,00	0,03	0,00	0,00	0,00	0,00	0,01	0,48	0,00																		
HUI	0,28	0,26	0,26	0,48	0,09	0,18	0,00	0,00	0,04	0,02	0,00	0,00	0,05	0,00	0,01	0,00																	
TIA	0,55	0,05	0,08	0,06	0,07	0,54	0,00	0,90	0,04	0,25	0,00	0,00	0,64	0,21	0,00	0,16	0,01																
TIT	0,14	0,00	0,09	0,16	0,01	1,00	0,27	0,08	0,70	0,51	0,05	0,06	0,66	0,01	0,00	0,00	0,06	0,12															
TIR	0,52	0,10	0,06	0,03	0,10	0,13	0,00	0,51	0,01	0,07	0,00	0,00	0,23	0,65	0,00	0,43	0,01	0,72	0,03														
ZOE	0,00	0,00	0,01	0,02	0,00	0,18	0,57	0,00	0,64	0,05	1,00	0,75	0,05	0,00	0,05	0,00	0,01	0,00	0,15	0,00													
ARL	0,07	0,00	0,02	0,03	0,00	0,75	0,04	0,22	0,21	1,00	0,01	0,00	0,77	0,01	0,00	0,00	0,01	0,31	0,43	0,07	0,05												
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,04	0,00	0,39	0,22	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,15	0,00												
AAA	0,02	0,00	0,01	0,01	0,00	1,00	0,02	0,16	0,23	1,00	0,01	0,00	0,56	0,01	0,00	0,00	0,18	0,34	0,03	0,06	1,00	0,00											
AWA	0,35	0,02	0,11	0,16	0,03	0,73	0,09	0,40	0,28	1,00	0,02	0,02	1,00	0,03	0,00	0,01	0,06	0,63	0,73	0,14	0,06	1,00	0,00	1,00									
PRK	0,41	0,11	0,04	0,04	0,07	0,09	0,00	0,56	0,02	0,08	0,00	0,00	0,07	0,74	0,00	0,32	0,01	0,71	0,05	1,00	0,00	0,08	0,00	0,05	0,19								
ASS	0,04	0,00	0,00	0,00	0,00	0,06	0,00	0,48	0,00	0,03	0,00	0,00	0,03	0,59	0,00	0,08	0,00	0,29	0,01	0,68	0,00	0,04	0,00	0,02	0,10	1,00							
AWE	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,20	0,00	0,14	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,08	0,02						
KAK	0,12	0,01	0,01	0,00	0,01	0,12	0,00	0,81	0,01	0,14	0,00	0,00	0,09	0,28	0,00	0,05	0,00	0,71	0,05	0,60	0,00	0,14	0,00	0,07	0,26	0,77	0,65	0,01					
KAX	0,03	0,01	0,00	0,00	0,00	0,03	0,00	0,23	0,00	0,01	0,00	0,00	0,01	0,79	0,00	0,20	0,00	0,18	0,01	0,44	0,00	0,02	0,00	0,01	0,03	0,77	0,65	0,07	0,35				

Tabela A.10. Teste Exato de Fisher entre as populações para o locus D6S2927

POP	POV	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WAU	ING	AYM	HUI	TIA	TIT	TIR	ZOE	ARL	ARI	CAA	AWA	PRK	ASS	AWE	KAK	KAX
	POR	0,76																											
CHI	0,00	0,00																											
CRE	0,29	0,44	0,11																										
OJI	0,00	0,00	0,36	0,23																									
KAQ	0,01	0,02	0,03	0,09	0,09																								
CAB	0,00	0,00	0,16	0,00	0,08	0,02																							
GUA	0,00	0,00	0,00	0,00	0,00	0,03	0,01																						
BEM	0,00	0,00	0,01	0,00	0,00	0,09	0,00	0,00																					
ZEN	0,00	0,00	0,35	0,07	0,18	0,70	0,10	0,00	0,11																				
KOG	0,00	0,00	0,05	0,00	0,17	0,04	0,60	0,14	0,00	0,10																			
ARH	0,00	0,00	0,00	0,00	0,01	0,06	0,00	0,02	0,00	0,00	0,04																		
WAY	0,00	0,00	0,04	0,02	0,64	0,31	0,07	0,05	0,00	0,21	0,48	0,37																	
WAU	0,00	0,00	0,98	0,08	0,23	0,01	0,03	0,00	0,01	0,27	0,01	0,00	0,01																
ING	0,00	0,00	0,61	0,16	0,06	0,01	0,01	0,00	0,00	0,14	0,00	0,00	0,50																
AYM	0,04	0,02	0,08	0,30	0,26	0,83	0,03	0,02	0,01	0,60	0,05	0,01	0,28	0,03	0,07														
HUI	0,02	0,17	0,00	0,06	0,00	0,04	0,00	0,00	0,12	0,01	0,00	0,00	0,00	0,00	0,01	0,01													
TIA	0,01	0,05	0,04	0,08	0,01	0,13	0,00	0,00	0,30	0,18	0,00	0,00	0,00	0,05	0,14	0,14	0,79												
TIT	0,00	0,00	0,95	0,01	0,16	0,01	0,53	0,00	0,00	0,20	0,10	0,00	0,02	0,56	0,30	0,04	0,00	0,01											
TIR	0,00	0,00	0,05	0,07	0,08	0,51	0,07	0,01	0,00	0,44	0,03	0,00	0,09	0,01	0,07	1,00	0,00	0,04	0,05										
ZOE	0,00	0,00	0,07	0,03	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,08	0,00	0,00	0,03	0,00										
ARL	0,00	0,00	0,28	0,11	0,19	0,02	0,12	0,00	0,00	0,06	0,02	0,00	0,02	0,09	0,33	0,24	0,00	0,01	0,30	0,25	0,13								
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
CAA	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
AWA	0,13	0,30	0,05	0,38	0,01	0,06	0,00	0,00	0,03	0,05	0,00	0,00	0,00	0,05	0,27	0,11	0,63	0,67	0,01	0,06	0,01	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PRK	0,01	0,03	0,48	0,28	0,07	0,01	0,00	0,00	0,01	0,06	0,00	0,00	0,00	0,44	0,78	0,06	0,22	0,18	0,04	0,26	0,17	0,00	0,03	0,39	0,80	0,16	0,00	0,00	0,00
ASS	0,01	0,05	0,01	0,07	0,00	0,06	0,00	0,00	0,08	0,06	0,00	0,00	0,00	0,01	0,10	0,04	0,72	0,84	0,00	0,03	0,00	0,00	0,00	0,00	0,80	0,16	0,00	0,00	0,00
AWE	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
KAK	0,01	0,01	0,05	0,24	0,05	0,21	0,01	0,00	0,00	0,20	0,00	0,00	0,02	0,02	0,21	0,84	0,01	0,08	0,03	0,81	0,01	0,32	0,00	0,00	0,22	0,13	0,11	0,00	0,00
KAX	0,05	0,01	0,00	0,01	0,00	0,09	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,14	0,00	0,02	0,00	0,04	0,00	0,00	0,00	0,00	0,05	0,00	0,02	0,00	0,07

Tabela A.11. Teste Exato de Fisher entre as populações para o locus D6S2926

POP	POI	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WOU	ING	AYM	HUI	TIA	TIT	ZOE	ARL	ARI	CAA	AWA	PRK	ASS	AWE	KAK	KAX	
POR	0,00																												
CHI	0,00	0,02																											
CRE	0,00	0,13	0,12																										
OJI	0,00	0,07	0,48	0,02																									
KAQ	0,00	0,02	0,12	0,01	0,10																								
CAB	0,00	0,00	0,01	0,00	0,01	0,01																							
GUA	0,00	0,00	0,02	0,00	0,01	0,19	0,00																						
BEM	0,00	0,02	0,31	0,01	0,23	0,89	0,01	0,87																					
ZEN	0,00	0,00	0,88	0,05	0,33	0,06	0,00	0,03	0,36																				
KOG	0,00	0,00	0,28	0,00	0,14	0,02	0,81	0,02	0,05	0,10																			
ARH	0,00	0,00	0,15	0,00	0,14	0,17	0,24	0,68	0,57	0,14	0,64																		
WAY	0,00	0,02	0,41	0,01	0,49	0,69	0,00	0,37	1,00	0,44	0,05	0,43																	
WOU	0,00	0,02	0,29	0,01	0,16	0,07	0,11	0,01	0,11	0,10	0,57	0,25	0,07																
ING	0,00	0,01	0,01	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,07	0,00	0,00	0,31															
AYM	0,00	0,06	0,81	0,04	0,60	0,40	0,10	0,51	0,86	0,80	0,48	0,91	0,93	0,57	0,03														
HUI	0,00	0,01	0,00	0,00	0,00	0,24	0,00	0,14	0,38	0,00	0,00	0,05	0,14	0,00	0,00	0,11													
TIA	0,00	0,03	0,51	0,02	0,36	0,58	0,04	0,90	1,00	0,58	0,20	0,82	0,87	0,19	0,00	0,94	0,19												
TIT	0,00	0,00	0,75	0,01	0,35	0,01	0,35	0,02	0,05	0,29	0,84	0,39	0,08	0,58	0,09	0,68	0,00	0,21											
TIR	0,00	0,04	0,45	0,03	0,33	0,35	0,04	0,02	0,37	0,23	0,30	0,30	0,40	0,63	0,12	0,81	0,02	0,30	0,33										
ZOE	0,48	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,30	0,00	0,00	0,01	0,07										
ARL	0,00	0,06	0,01	0,00	0,01	0,04	0,00	0,00	0,03	0,00	0,06	0,05	0,01	0,19	0,18	0,13	0,01	0,04	0,02	0,16	0,00								
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00							
CAA	0,00	0,05	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,16	0,12	0,01	0,00	0,00	0,03	0,01	0,00	0,10	0,00						
AWA	0,00	0,00	0,01	0,00	0,01	0,02	0,32	0,03	0,05	0,00	0,28	0,42	0,01	0,11	0,01	0,13	0,00	0,10	0,06	0,05	0,00	0,11	0,00	0,00					
PRK	0,00	0,00	0,15	0,00	0,09	0,03	0,51	0,00	0,02	0,03	0,80	0,12	0,02	0,66	0,39	0,33	0,00	0,04	0,73	0,38	0,03	0,09	0,01	0,03	0,15				
ASS	0,01	0,00	0,01	0,00	0,00	0,74	0,00	0,00	0,42	0,00	0,00	0,01	0,21	0,00	0,00	0,14	0,09	0,12	0,00	0,21	0,00	0,00	0,00	0,00	0,00	0,00			
AWE	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,15	0,00	0,00	0,00	0,00	0,00	0,10	0,00	0,00	0,00	0,00	0,02	0,00		
KAK	0,00	0,15	0,07	0,01	0,04	0,11	0,00	0,02	0,20	0,04	0,08	0,21	0,10	0,19	0,02	0,56	0,10	0,30	0,05	0,37	0,00	0,60	0,00	0,03	0,07	0,05	0,02	0,00	
KAX	0,00	0,04	0,01	0,00	0,01	0,29	0,00	0,04	0,34	0,01	0,02	0,18	0,13	0,05	0,00	0,36	0,37	0,24	0,01	0,22	0,00	0,36	0,00	0,00	0,06	0,01	0,08	0,00	0,81

Tabela A.12. Teste Exato de Fisher entre as populações para o locus MICA

POP	POI	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WOU	ING	AYM	HUI	TIA	TIT	TIR	ZOE	ARL	ARI	CAA	AWA	PRK	ASS	AWE	KAK	KAX	
POR	0,96																													
CHI	0,01	0,02																												
CRE	0,64	0,42	0,04																											
OJI	0,00	0,00	0,12	0,12																										
KAQ	0,00	0,00	0,13	0,06	0,91																									
CAB	0,01	0,03	1,00	0,03	0,10	0,10																								
GUA	0,00	0,00	0,00	0,00	0,16	0,00	0,00																							
BEM	0,01	0,01	0,18	0,15	1,00	0,89	0,19	0,59																						
ZEN	0,44	0,26	0,06	0,79	0,29	0,11	0,07	0,01	0,37																					
KOG	0,01	0,01	0,47	0,03	0,52	0,55	0,57	0,29	0,82	0,16																				
ARH	0,00	0,00	0,02	0,06	0,38	0,54	0,01	0,00	0,40	0,05	0,04																			
WAY	0,00	0,00	0,01	0,01	0,62	0,87	0,01	0,23	0,90	0,03	0,20	0,34																		
WOU	0,00	0,00	0,31	0,00	0,13	0,12	0,52	0,06	0,32	0,01	0,95	0,00	0,02																	
ING	0,01	0,04	0,37	0,01	0,01	0,00	0,60	0,00	0,02	0,02	0,31	0,00	0,00	0,23																
AYM	0,00	0,00	0,01	0,00	0,32	0,41	0,01	0,81	0,79	0,01	0,54	0,02	0,46	0,17	0,00															
HUI	0,00	0,00	0,00	0,01	0,16	0,10	0,00	0,27	0,44	0,08	0,04	0,02	0,19	0,00	0,00	0,08														
TIA	0,00	0,00	0,01	0,00	0,32	0,41	0,01	0,81	0,79	0,01	0,55	0,02	0,45	0,16	0,00	1,00	0,08													
TIT	0,00	0,00	0,07	0,00	0,00	0,00	0,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
TIR	0,00	0,00	0,02	0,00	0,21	0,14	0,04	0,72	0,57	0,02	0,66	0,00	0,12	0,61	0,02	0,58	0,11	0,58	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
ZOE	0,00	0,00	0,06	0,00	0,00	0,00	0,15	0,00	0,00	0,00	0,01	0,00	0,00	0,03	0,40	0,00	0,00	0,00	0,10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
ARL	0,00	0,01	0,23	0,01	0,06	0,03	0,41	0,04	0,16	0,10	0,59	0,00	0,00	0,50	0,44	0,01	0,01	0,01	0,00	0,19	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
CAA	0,80	0,94	0,00	0,21	0,00	0,00	0,00	0,00	0,00	0,19	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
AWA	0,00	0,00	0,06	0,00	0,07	0,06	0,09	0,20	0,17	0,01	1,00	0,00	0,02	0,92	0,05	0,24	0,01	0,24	0,00	0,75	0,02	0,26	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PRK	0,00	0,00	0,15	0,00	0,07	0,05	0,26	0,12	0,14	0,01	0,93	0,00	0,01	1,00	0,13	0,12	0,01	0,12	0,00	0,51	0,06	0,39	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ASS	0,00	0,00	0,08	0,00	0,01	0,00	0,18	0,01	0,06	0,00	0,55	0,00	0,00	0,78	0,28	0,01	0,00	0,01	0,00	0,27	0,04	0,49	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
AWE	0,00	0,00	0,01	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,15	0,00	0,00	0,00	0,00	0,02	0,00	0,49	0,00	0,03	0,00	0,01	0,04	0,02	0,00	0,00
KAK	0,00	0,00	0,01	0,00	0,01	0,01	0,03	0,05	0,04	0,00	0,41	0,00	0,00	1,00	0,03	0,07	0,00	0,07	0,00	0,27	0,01	0,14	0,00	0,00	1,00	1,00	0,63	0,00	0,00	0,00
KAX	0,00	0,00	0,00	0,00	0,02	0,03	0,00	0,46	0,13	0,00	0,16	0,00	0,02	0,13	0,00	0,42	0,01	0,43	0,00	0,70	0,00	0,01	0,00	0,00	0,42	0,25	0,04	0,00	0,00	0,24

Tabela A.13. Teste Exato de Fisher entre as populações para o locus BAT2

POP	POI	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WOU	ING	AYM	HUI	TIA	TIT	TIR	ZOE	ARL	ARI	CAA	AWA	PRK	ASS	AWE	KAK	KAX	
POR	0,01																													
CHI	0,00	0,00																												
CRE	0,00	0,01	0,29																											
OJI	0,00	0,01	0,28	0,76																										
KAQ	0,00	0,01	0,00	0,14	0,03																									
CAB	0,00	0,01	0,02	0,14	0,68	0,01																								
GUA	0,00	0,00	0,18	0,00	0,00	0,00	0,00																							
BEM	0,00	0,02	0,10	0,37	0,75	0,06	0,22	0,01																						
ZEN	0,00	0,00	0,04	0,36	0,58	0,23	0,20	0,00	0,21																					
KOG	0,00	0,00	0,26	0,05	0,17	0,02	0,02	0,14	0,39	0,02																				
ARH	0,00	0,00	0,01	0,00	0,00	0,06	0,00	0,00	0,09	0,00	0,24																			
WAY	0,00	0,06	0,48	0,67	0,51	0,10	0,09	0,02	0,83	0,17	0,62	0,37																		
WOU	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00																	
ING	0,00	0,00	0,00	0,12	0,20	0,05	0,14	0,00	0,03	0,56	0,00	0,01	0,00	0,00																
AYM	0,00	0,00	0,12	0,27	0,38	0,00	0,47	0,00	0,04	0,09	0,01	0,00	0,08	0,00	0,01															
HUI	0,00	0,07	0,03	0,07	0,06	0,03	0,09	0,00	0,37	0,00	0,51	0,04	0,78	0,00	0,00	0,06														
TIA	0,00	0,08	0,00	0,03	0,10	0,11	0,03	0,00	0,03	0,13	0,01	0,01	0,03	0,01	0,01	0,06	0,04													
TIT	0,00	0,13	0,01	0,23	0,45	0,48	0,03	0,00	0,55	0,53	0,04	0,02	0,53	0,00	0,28	0,01	0,07	0,33												
TIR	0,00	0,00	0,52	0,57	0,64	0,01	0,05	0,01	0,38	0,34	0,08	0,00	0,41	0,00	0,03	0,14	0,01	0,16												
ZOE	0,00	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,00	0,01	0,00	0,01	0,02												
ARL	0,00	0,00	0,00	0,29	0,12	0,01	0,02	0,00	0,01	0,44	0,00	0,00	0,01	0,00	0,18	0,08	0,00	0,01	0,09	0,17	0,00									
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00								
CAA	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,02	0,00	0,02	0,00	0,05	0,00						
AWA	0,00	0,00	0,00	0,05	0,04	0,00	0,02	0,00	0,00	0,27	0,00	0,00	0,00	0,00	0,23	0,02	0,00	0,00	0,01	0,02	0,01	0,32	0,00	0,11						
PRK	0,00	0,00	0,08	0,61	0,51	0,05	0,17	0,00	0,08	0,77	0,01	0,00	0,22	0,00	0,30	0,11	0,01	0,01	0,30	0,74	0,02	0,50	0,00	0,11	0,32					
ASS	0,00	0,00	0,08	0,34	0,96	0,02	0,10	0,00	0,55	0,31	0,11	0,00	0,22	0,00	0,11	0,11	0,02	0,24	0,72	0,69	0,03	0,11	0,00	0,00	0,02	0,38				
AWE	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
KAK	0,00	0,00	0,00	0,04	0,10	0,01	0,01	0,00	0,01	0,29	0,00	0,00	0,00	0,00	0,39	0,00	0,00	0,02	0,32	0,01	0,00	0,72	0,00	0,00	0,24	0,27	0,10	0,00		
KAX	0,00	0,00	0,00	0,05	0,04	0,00	0,00	0,00	0,01	0,08	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,04	0,05	0,00	0,46	0,00	0,00	0,01	0,12	0,06	0,00	0,27		

Tabela A.14. Teste Exato de Fisher entre as populações para o locus D6S273

POP	OVI	POR	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WOU	ING	AYM	HUI	TIA	TIT	TIR	ZOE	ARL	ARI	CAA	AWA	PRK	ASS	AWE	KAK	KAX
POR	0,00																													
CHI	0,00	0,12																												
CRE	0,00	0,60	0,49																											
OJI	0,04	0,00	0,00	0,00																										
KAQ	0,09	0,02	0,26	0,06	0,05																									
CAB	0,03	0,17	0,99	0,60	0,00	0,77																								
GUA	0,00	0,00	0,06	0,00	0,00	0,02	0,02																							
BEM	0,00	0,00	0,01	0,00	0,00	0,02	0,01	0,24																						
ZEN	0,30	0,54	0,47	0,00	0,04	0,26	0,01	0,00	0,00																					
KOG	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00																				
ARH	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,19	0,00	0,00																			
WAY	0,00	0,01	0,00	0,00	0,00	0,01	0,00	0,01	0,22	0,00	0,00	0,79																		
WOU	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00																	
ING	0,00	0,01	0,88	0,13	0,00	0,23	0,89	0,06	0,01	0,12	0,01	0,00	0,00	0,00																
AYM	0,00	0,26	0,77	0,30	0,00	0,22	0,65	0,14	0,03	0,70	0,00	0,02	0,02	0,00	0,40															
HUI	0,00	0,00	0,32	0,01	0,00	0,03	0,11	0,53	0,02	0,04	0,06	0,00	0,00	0,00	0,37	0,22														
TIA	0,01	0,00	0,25	0,06	0,00	0,14	0,30	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,12	0,13	0,01													
TIT	0,01	0,00	0,03	0,00	0,00	0,21	0,12	0,00	0,00	0,00	0,00	0,00	0,00	0,04	0,07	0,00	0,16													
TIR	0,00	0,05	0,83	0,12	0,00	0,39	0,74	0,29	0,03	0,45	0,00	0,01	0,01	0,00	0,58	0,92	0,56	0,10	0,02											
ZOE	0,00	0,00	0,03	0,00	0,00	0,03	0,03	0,15	0,44	0,00	0,09	0,00	0,01	0,00	0,04	0,03	0,06	0,00	0,04											
ARL	0,00	0,00	0,12	0,00	0,00	0,14	0,21	0,01	0,04	0,00	0,00	0,00	0,00	0,00	0,32	0,04	0,03	0,01	0,05	0,04	0,19									
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,00								
KAA	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00							
AWA	0,00	0,00	0,03	0,00	0,00	0,04	0,03	0,30	0,71	0,00	0,03	0,01	0,01	0,00	0,06	0,03	0,10	0,00	0,01	0,06	0,75	0,33	0,00	0,00						
PRK	0,00	0,00	0,01	0,00	0,00	0,01	0,00	0,09	0,69	0,00	0,00	0,05	0,05	0,00	0,00	0,02	0,01	0,00	0,00	0,02	0,30	0,03	0,00	0,02	0,33					
ASS	0,01	0,00	0,13	0,03	0,00	0,06	0,24	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,17	0,01	0,00	0,01	0,01	0,01	0,08	0,00	0,00	0,00	0,00	0,00				
AWE	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00			
KAK	0,00	0,00	0,09	0,00	0,00	0,01	0,02	0,60	0,15	0,00	0,08	0,00	0,00	0,00	0,09	0,07	0,62	0,00	0,00	0,17	0,28	0,07	0,00	0,00	0,47	0,20	0,00	0,00		
KAX	0,00	0,00	0,04	0,00	0,00	0,01	0,01	0,15	0,16	0,00	0,01	0,00	0,00	0,00	0,03	0,06	0,09	0,00	0,00	0,07	0,70	0,03	0,00	0,00	0,20	0,26	0,00	0,00	0,33	

Tabela A.15. Teste Exato de Fisher entre as populações para o locus G51152

POP	OVI	POR	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WAU	ING	AYM	HUI	TIA	TIT	ZOE	ARL	ARI	AAA	AWA	PRK	ASS	AWE	KAK	KAX			
	0,03																															
CHI	0,04	0,70																														
CRE	0,01	0,02	0,11																													
OJI	0,01	0,01	0,06	0,66																												
KAQ	0,00	0,01	0,02	0,73	0,32																											
CAB	0,00	0,00	0,00	0,89	0,08	0,25																										
GUA	0,00	0,00	0,00	0,00	0,01	0,01																										
BEM	0,01	0,00	0,01	0,92	0,20	0,55	0,87	0,04																								
ZEN	0,00	0,00	0,04	0,00	0,00	0,00	0,00	0,00																								
KOG	0,01	0,00	0,00	0,10	0,04	0,17	0,11	0,37	0,32	0,00																						
ARH	0,00	0,00	0,00	0,00	0,01	0,01	1,00	0,04	0,00	0,35																						
WAY	0,00	0,00	0,00	0,09	0,03	0,37	0,13	0,26	0,42	0,00	0,85	0,25																				
WAU	0,00	0,00	0,06	0,58	0,05	0,30	0,24	0,00	0,27	0,01	0,01	0,00	0,00																			
ING	0,01	0,00	0,00	0,08	0,11	0,07	0,03	0,35	0,10	0,00	0,73	0,44	0,38	0,00																		
AYM	0,01	0,00	0,01	0,19	0,59	0,04	0,04	0,00	0,08	0,00	0,04	0,00	0,03	0,00	0,13																	
HUI	0,00	0,00	0,01	0,91	0,21	0,57	0,93	0,00	0,90	0,00	0,11	0,00	0,12	0,47	0,02	0,02																
TIA	0,00	0,00	0,00	0,03	0,01	0,06	0,03	0,74	0,11	0,00	0,72	0,72	0,75	0,00	0,54	0,02	0,03															
TIT	0,00	0,00	0,00	0,00	0,00	0,00	0,34	0,01	0,00	0,18	0,67	0,06	0,00	0,63	0,01	0,00	0,23															
TIR	0,00	0,00	0,00	0,06	0,01	0,06	0,07	0,53	0,21	0,00	0,73	0,51	0,88	0,00	0,31	0,02	0,06	0,76	0,15													
ZOE	0,01	0,00	0,00	0,75	0,15	0,55	0,76	0,10	1,00	0,00	0,33	0,05	0,53	0,18	0,16	0,09	0,83	0,20	0,01	0,35												
ARL	0,00	0,00	0,00	0,00	0,01	0,00	0,71	0,01	0,00	0,18	1,00	0,12	0,00	0,37	0,00	0,00	0,46	1,00	0,30	0,04												
ARI	0,00	0,00	0,01	0,30	0,01	0,04	0,26	0,00	0,26	0,03	0,01	0,00	0,01	0,54	0,00	0,00	0,26	0,00	0,00	0,17	0,00											
KAA	0,00	0,00	0,00	0,01	0,00	0,02	0,02	0,76	0,11	0,00	0,40	0,75	0,44	0,00	0,25	0,00	0,01	1,00	0,16	0,77	0,12	0,50	0,00									
AWA	0,00	0,00	0,00	0,00	0,00	0,00	0,10	0,00	0,00	0,03	0,15	0,02	0,00	0,13	0,00	0,00	0,07	0,52	0,04	0,00	0,28	0,00	0,05									
PRK	0,01	0,05	0,30	0,20	0,03	0,03	0,10	0,00	0,06	0,46	0,00	0,00	0,00	0,34	0,00	0,01	0,15	0,00	0,00	0,03	0,00	0,38	0,00	0,00								
ASS	0,00	0,00	0,00	0,34	0,09	0,13	0,11	0,25	0,59	0,00	0,66	0,18	0,87	0,00	0,60	0,10	0,12	0,64	0,15	0,88	0,72	0,08	0,01	0,33	0,02	0,00						
AWE	0,00	0,00	0,01	0,04	0,00	0,01	0,03	0,00	0,05	0,17	0,00	0,00	0,00	0,19	0,00	0,00	0,03	0,00	0,00	0,02	0,00	0,34	0,00	0,00	1,00	0,00						
KAK	0,00	0,00	0,01	0,19	0,01	0,03	0,17	0,00	0,17	0,06	0,00	0,00	0,00	0,49	0,00	0,00	0,18	0,00	0,00	0,10	0,00	1,00	0,00	0,00	0,55	0,00	0,47					
KAX	0,00	0,00	0,00	0,44	0,01	0,08	0,54	0,00	0,59	0,00	0,02	0,00	0,03	0,31	0,00	0,01	0,66	0,01	0,00	0,02	0,43	0,00	0,49	0,00	0,00	0,15	0,03	0,11	0,37			

Tabela A.16. Teste Exato de Fisher entre as populações para o locus D6S2874

POP	OVI	POR	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WAU	ING	AYM	HUI	TIA	TIT	TIR	ZOE	ARL	ARI	KAA	AWA	PRK	ASS	AWE	KAK	KAX		
POR	0,11																															
CHI	0,04	0,32																														
CRE	0,02	0,15	0,71																													
OJI	0,15	0,05	0,36	0,44																												
KAQ	0,08	0,04	0,14	0,71	0,47																											
CAB	0,00	0,00	0,00	0,00	0,01	0,03																										
GUA	0,03	0,02	0,04	0,64	0,21	0,74	0,00																									
BEM	0,00	0,00	0,00	0,00	0,01	0,01	0,27	0,00																								
ZEN	0,01	0,08	0,30	0,70	0,01	0,07	0,00	0,15	0,00																							
KOG	0,06	0,04	0,00	0,01	0,05	0,06	0,05	0,04	0,04	0,00																						
ARH	0,10	0,08	0,01	0,08	0,09	0,20	0,00	0,36	0,00	0,05	0,28																					
WAY	0,06	0,06	0,09	0,64	0,26	0,62	0,00	1,00	0,00	0,31	0,06	0,42																				
WAU	0,04	0,04	0,00	0,09	0,03	0,16	0,00	0,44	0,00	0,14	0,11	0,62	0,60																			
ING	0,01	0,01	0,00	0,01	0,10	0,21	0,09	0,16	0,22	0,00	0,59	0,21	0,13	0,04																		
AYM	0,06	0,02	0,02	0,20	0,29	0,81	0,17	0,44	0,07	0,02	0,43	0,55	0,47	0,14	0,62																	
HUI	0,00	0,00	0,09	0,52	0,03	0,22	0,00	0,09	0,00	0,03	0,00	0,06	0,00	0,00	0,02																	
TIA	0,02	0,02	0,00	0,00	0,01	0,04	0,00	0,05	0,01	0,00	0,79	0,45	0,06	0,15	0,44	0,20	0,00															
TIT	0,03	0,02	0,00	0,10	0,23	0,38	0,10	0,25	0,14	0,00	0,47	0,32	0,32	0,08	0,92	0,90	0,01	0,42														
TIR	0,00	0,00	0,00	0,03	0,11	0,18	0,02	0,04	0,05	0,00	0,01	0,01	0,02	0,00	0,06	0,15	0,11	0,00	0,13													
ZOE	0,08	0,07	0,18	0,82	0,60	1,00	0,05	0,91	0,02	0,09	0,08	0,24	0,72	0,22	0,31	0,88	0,40	0,06	0,50	0,32												
ARL	0,10	0,07	0,01	0,13	0,08	0,30	0,00	0,45	0,00	0,12	0,17	0,89	0,69	0,70	0,07	0,61	0,00	0,17	0,16	0,01	0,29											
ARI	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
KAA	0,00	0,00	0,03	0,11	0,01	0,05	0,00	0,02	0,00	0,01	0,00	0,00	0,02	0,00	0,00	0,00	0,70	0,00	0,00	0,01	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
AWA	0,00	0,00	0,00	0,00	0,02	0,27	0,00	0,77	0,00	0,20	0,02	0,00	0,00	0,00	0,47	0,14	0,00	0,08	0,35	0,01	0,02	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PRK	0,00	0,00	0,03	0,04	0,04	0,04	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,36	0,00	0,00	0,18	0,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ASS	0,00	0,00	0,00	0,00	0,04	0,04	0,06	0,01	0,12	0,00	0,01	0,00	0,00	0,00	0,03	0,04	0,02	0,00	0,06	0,62	0,09	0,00	0,00	0,00	0,00	0,02	0,14					
AWE	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,33	0,02	0,00	0,00	0,00	0,00	0,00	0,21	0,80				
KAK	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
KAX	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

0,00 0,11 0,16 0,38 0,58

Tabela A.17. Teste Exato de Fisher entre as populações para o locus D6S439

POP	POI	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WOU	ING	AYM	HUI	TIA	TIT	TIR	ZOE	ARL	ARI	CAA	AWA	PRK	ASS	AWE	KAK	KAX	
POR	0,01																													
CHI	0,00	0,00																												
CRE	0,28	0,00	0,07																											
OJI	0,00	0,00	0,68	0,07																										
KAQ	0,00	0,00	0,30	0,14	0,57																									
CAB	0,00	0,00	0,36	0,09	0,03	0,01																								
GUA	0,00	0,00	0,30	0,03	0,45	1,00	0,00																							
BEM	0,00	0,00	0,88	0,37	0,84	0,61	0,42	0,82																						
ZEN	0,00	0,00	0,42	0,03	0,17	0,13	0,33	0,34	0,70																					
KOG	0,00	0,00	0,35	0,05	0,34	0,07	0,14	0,07	0,52	0,24																				
ARH	0,00	0,00	0,50	0,00	0,64	0,16	0,02	0,28	0,53	0,20	0,50																			
WAY	0,50	0,12	0,00	0,31	0,00	0,01	0,01	0,00	0,01	0,01	0,00	0,00																		
WOU	0,00	0,00	0,59	0,08	0,20	0,45	0,14	0,47	1,00	0,54	0,11	0,18	0,00																	
ING	0,00	0,00	0,69	0,00	0,65	0,78	0,00	0,75	0,59	0,07	0,12	0,81	0,00	0,46																
AYM	0,02	0,00	0,39	0,74	0,20	0,52	0,29	0,30	0,71	0,39	0,22	0,09	0,25	0,42	0,06															
HUI	0,05	0,00	0,01	0,74	0,00	0,05	0,01	0,00	0,08	0,00	0,00	0,00	0,03	0,05	0,00	0,27														
TIA	0,00	0,00	0,26	0,03	0,05	0,03	0,53	0,10	0,50	0,93	0,32	0,17	0,01	0,40	0,02	0,40	0,01													
TIT	0,00	0,00	0,21	0,00	0,34	0,57	0,01	0,55	0,21	0,09	0,12	0,51	0,00	0,11	0,84	0,05	0,00	0,05												
TIR	0,01	0,04	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,19	0,00	0,00	0,00	0,00	0,00	0,00											
ZOE	0,00	0,00	0,04	0,00	0,09	0,04	0,00	0,16	0,02	0,01	0,01	0,17	0,00	0,01	0,67	0,00	0,00	0,01	0,00	0,00										
ARL	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00									
ARI	0,00	0,00	0,14	0,00	0,09	0,00	0,02	0,00	0,04	0,00	0,06	0,04	0,00	0,01	0,02	0,01	0,00	0,01	0,02	0,00	0,00	0,00								
CAA	0,00	0,00	0,24	0,09	0,02	0,03	0,44	0,01	0,39	0,16	0,01	0,00	0,00	0,37	0,00	0,23	0,15	0,18	0,01	0,00	0,00	0,00	0,00							
AWA	0,00	0,00	0,01	0,10	0,01	0,10	0,00	0,02	0,07	0,00	0,00	0,00	0,00	0,06	0,00	0,02	0,09	0,00	0,01	0,00	0,00	0,00	0,00	0,01						
PRK	0,01	0,00	0,19	0,18	0,27	0,62	0,01	0,55	0,67	0,11	0,03	0,02	0,00	0,48	0,21	0,23	0,21	0,03	0,08	0,00	0,01	0,00	0,00	0,14	0,29					
ASS	0,00	0,00	0,11	0,01	0,05	0,30	0,00	0,29	0,43	0,07	0,00	0,00	0,00	0,45	0,05	0,05	0,02	0,02	0,04	0,00	0,00	0,00	0,00	0,09	0,05	0,83				
AWE	0,00	0,00	0,00	0,00	0,02	0,20	0,00	0,08	0,01	0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,33	0,00	0,10	0,00	0,00	0,00	0,00	0,07	0,01			
KAK	0,00	0,00	0,05	0,00	0,14	0,65	0,00	0,51	0,11	0,01	0,00	0,01	0,00	0,09	0,33	0,00	0,00	0,00	0,38	0,00	0,15	0,00	0,00	0,00	0,00	0,39	0,12	0,86		
KAX	0,00	0,00	0,02	0,00	0,07	0,52	0,00	0,23	0,10	0,00	0,00	0,00	0,00	0,08	0,07	0,00	0,00	0,00	0,11	0,00	0,02	0,00	0,00	0,00	0,00	0,70	0,23	0,24	0,60	

Tabela A.18. Teste Exato de Fisher entre as populações para o locus D6S291

POP	POI	CHI	CRE	OJI	KAQ	CAB	GUA	BEM	ZEN	KOG	ARH	WAY	WAU	ING	AYM	HUI	TIA	TIT	TIR	ZOE	ARL	ARI	KAA	AWA	PRK	ASS	AWE	KAK	KAX
POR	0,00																												
CHI	0,00	0,00																											
CRE	0,00	0,00	0,17																										
OJI	0,00	0,00	0,00	0,16																									
KAQ	0,00	0,00	0,00	0,01	0,04																								
CAB	0,00	0,00	0,00	0,03	0,00	0,00																							
GUA	0,00	0,00	0,00	0,07	0,31	0,51	0,00																						
BEM	0,00	0,00	0,00	0,04	0,08	0,02	0,42	0,14																					
ZEN	0,00	0,00	0,15	0,77	0,02	0,00	0,27	0,01	0,04																				
KOG	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,35	0,02	0,00																			
ARH	0,00	0,00	0,00	0,00	0,00	0,31	0,00	0,24	0,00	0,00	1,00																		
WAY	0,00	0,00	0,06	0,72	0,52	0,01	0,11	0,10	0,12	0,57	0,00	0,00																	
WAU	0,00	0,00	0,00	0,13	0,07	0,02	0,00	0,33	0,09	0,05	0,04	0,02	0,05																
ING	0,00	0,00	0,01	0,07	0,25	0,02	0,09	0,11	0,13	0,11	0,00	0,00	0,28	0,04															
AYM	0,00	0,00	0,11	0,61	0,53	0,08	0,00	0,50	0,02	0,11	0,04	0,01	0,54	0,25	0,10														
HUI	0,00	0,00	0,06	0,18	0,21	0,06	0,08	0,18	0,11	0,14	0,01	0,00	0,38	0,03	0,67	0,21													
TIA	0,00	0,00	0,02	0,03	0,06	1,00	0,00	0,36	0,01	0,00	0,49	0,08	0,04	0,01	0,02	0,27	0,15												
TIT	0,00	0,00	0,00	0,00	0,01	0,80	0,00	0,38	0,03	0,00	0,80	0,49	0,00	0,01	0,00	0,02	0,06	0,42											
TIR	0,00	0,00	0,02	0,07	0,09	0,34	0,00	0,54	0,01	0,00	0,34	0,16	0,03	0,32	0,03	0,63	0,04	0,31	0,04										
ZOE	0,00	0,02	0,41	0,14	0,01	0,00	0,00	0,03	0,00	0,09	0,01	0,01	0,02	0,25	0,01	0,24	0,01	0,02	0,00	0,34									
ARL	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00									
ARI	0,00	0,00	0,00	0,00	0,00	0,41	0,00	0,10	0,00	0,00	0,41	0,49	0,00	0,00	0,00	0,00	0,07	0,49	0,01	0,00	0,00								
KAA	0,00	0,00	0,06	0,22	0,03	0,00	0,00	0,01	0,00	0,09	0,00	0,00	0,15	0,05	0,01	0,26	0,00	0,00	0,00	0,05	0,33	0,00	0,00						
AWA	0,00	0,00	0,46	0,08	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,01	0,01	0,01	0,00	0,07	0,00	0,02	0,00	0,04	0,71	0,07	0,00	0,04					
PRK	0,00	0,00	0,13	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,01	0,01	0,00	0,10	0,00	0,00	0,00	0,08	0,65	0,02	0,00	0,28	0,26				
ASS	0,00	0,00	0,00	0,00	0,01	1,00	0,00	0,18	0,02	0,00	1,00	0,51	0,00	0,01	0,00	0,01	0,02	0,49	1,00	0,08	0,00	0,00	0,00	0,00	0,00	0,00			
AWE	0,00	0,00	0,04	0,02	0,10	0,35	0,00	0,27	0,01	0,00	0,17	0,02	0,04	0,01	0,11	0,23	0,19	0,65	0,05	0,28	0,02	0,00	0,02	0,00	0,01	0,01	0,11		
KAK	0,00	0,00	0,00	0,01	0,07	0,39	0,00	0,21	0,00	0,00	0,11	0,05	0,01	0,02	0,01	0,35	0,01	0,58	0,05	0,47	0,03	0,00	0,04	0,00	0,00	0,02	0,08	0,80	
KAX	0,00	0,00	0,00	0,00	0,03	0,78	0,00	0,39	0,00	0,00	0,38	0,31	0,00	0,02	0,00	0,14	0,01	0,87	0,19	0,43	0,03	0,00	0,24	0,00	0,00	0,00	0,28	0,60	0,73

ANEXO 2

Manuscrito capítulo de livro

Estudos de Associação HLA x Doenças: Extratos do II Simpósio Brasileiro. Organizadores: Luís Cristóvão de Moraes Sobrino Porto e Luciane Faria de Souza Pontes. – Rio de Janeiro. EdUERJ, 244p 2009.

Capítulo 1: Uma visão evolutiva sobre a variabilidade em genes MHC

Kelly Nunes, Márcia Regina Pincerati, Rodrigo dos Santos Francisco e Diogo Meyer

Acaso e Seleção

Desde seu estabelecimento como disciplina, o “cálice sagrado” da genética de populações é compreender os processos que explicam a origem e manutenção da variabilidade genética nas populações [1]. Dentre os processos evolutivos que moldam a variabilidade genética estão a deriva genética (mudança aleatória de frequência alélica entre gerações) e a seleção natural (sobrevivência diferencial de indivíduos portadores de diferentes genótipos).

Grande parte da pesquisa em genética evolutiva consiste em compreender quais diferenças, entre populações ou espécies, resultam do acúmulo passivo de mutações sem efeitos seletivos (evolução neutra, que se dá por deriva genética), e quais resultam das vantagens seletivas associadas a certas mutações (a evolução adaptativa, que se dá por seleção natural).

Ao longo da última década, utilizando análises populacionais, foi possível vasculhar o genoma humano em busca de genes sob seleção natural. Foi encontrado um grande número de genes que aparentam estar evoluindo sob o efeito da seleção (revisado em [2]). Por exemplo, há fortes evidências de que a seleção natural favoreceu determinados alelos dos genes G6PD e Duffy [3,4] em populações africanas, em função de seu efeito protetor frente à malária.

O estudo de genes envolvidos na pigmentação humana também revelou casos de seleção natural, incluindo os genes SLC24A5 e MATP [5]. Há evidências de que certos alelos desses genes resultam num “clareamento” da pele, e que teriam aumentado de frequência pela ação da seleção natural em populações europeias. Essas populações localizam-se em altas latitudes, onde a incidência de luz ultravioleta (UV) é menor. Como a luz UV é

necessária para a síntese de vitamina D, a pele escura seria desvantajosa em relação a pele clara, pois dificultaria o acesso da luz aos capilares.

Há também vários exemplos de genes que estiveram sob seleção em escalas de tempo maiores. Por exemplo, comparando os genomas de humanos e chimpanzés, há evidências de que diversos genes envolvidos na formação de gametas evoluíram sob seleção [6].

Entretanto as maiorias dos genes humanos não apresentam evidências de ter sido selecionado: a maioria deles evoluem de modo neutro. Quando comparamos suas frequências nas populações, esses genes apresentam níveis de diferenciação compatíveis com o acúmulo passivo de mutações.

De modo semelhante, a maior parte das diferenças entre humanos e chimpanzés também parece ser resultado do acúmulo passivo de mutações, sendo que menos de 5% das diferenças entre essas espécies resultam de seleção natural [7].

Atualmente, não há dúvidas de que muitos genes evoluem de modo neutro, e outros são selecionados. O desafio é definir quais genes se diferenciaram de modo neutro, e quais genes se diferenciaram sob seleção, e entender porque a seleção atuou sobre determinados genes.

A seguir, definiremos diferentes tipos de seleção natural e revisaremos brevemente algumas ferramentas que geneticistas utilizam para, através de análises de populações atuais, inferir se houve seleção natural no passado. Depois, aplicaremos essas ferramentas ao estudo de genes MHC.

Tipos de seleção natural

Dizemos que há seleção natural sobre um gene se houver variação nas probabilidades de sobrevivência e reprodução dos indivíduos portadores de diferentes genótipos desses genes. Se todos os genótipos conferem a mesma probabilidade de sobrevivência e reprodução, não há seleção natural. Nesse caso, dizemos que o gene evolui de modo neutro.

As diferenças nas probabilidades de sobrevivência e reprodução conferidas pelos genótipos definem qual regime de seleção está atuando. Há três tipos básicos de regimes seletivos: 1. Seleção positiva – ocorre quando há um alelo vantajoso que aumenta de frequência na população. 2. Seleção negativa ou purificadora – ocorre quando há um alelo que diminui a aptidão do seu portador e, portanto, tende a ser removido da população; 3.

Seleção balanceadora – ocorre quando a seleção natural resulta na manutenção de variação na população, ao invés de favorecer ou remover alelos específicos.

Como saber se um gene esteve sob seleção natural?

A ação da seleção natural pode causar mudanças na composição genética de uma população. Por exemplo, a seleção positiva pode aumentar a frequência de um alelo vantajoso, diminuindo a variabilidade populacional para o *locus* selecionado. Por outro lado, a seleção balanceadora mantém muitos alelos numa população, resultando numa variabilidade aumentada para o *locus* selecionado.

A tarefa de um geneticista de populações é medir a variabilidade genética numa amostra populacional e usar essa informação para responder à seguinte pergunta: a população amostrada esteve (ou está) sob seleção natural para o *locus* cuja variabilidade foi quantificada?

Há diversas medidas que servem para descrever a variação genética numa amostra populacional, e para cada uma dessas medidas podemos prever de que modo a seleção resultaria numa alteração da variabilidade genética. Nossa discussão enfatizará os efeitos da seleção balanceadora, a mais relevante para explicar a variabilidade em genes MHC [tabela 1].

Para testar se a variação genética revela a ação da seleção, nos baseamos na Teoria Neutra da Evolução Molecular [8]. Essa teoria permite que façamos previsões sobre a variabilidade esperada numa amostra populacional, supondo que ele esteja evoluindo sem a ação da seleção natural.

Comparando a variação observada com aquela esperada na ausência de seleção, podemos realizar testes de neutralidade. Dizemos que rejeitamos a hipótese nula de neutralidade quando a variação observada em dados reais é incompatível com aquela esperada para uma população evoluindo de modo neutro (isto é, sem seleção) [6].

Na prática, a predição de variação esperada na ausência de seleção é obtida utilizando abordagens analíticas (isto é, baseando-se nas expressões matemáticas que descrevem a evolução neutra) e simulações de computador (cf. [2] para uma revisão dos métodos usados em testes de neutralidade).

Um dos maiores desafios no estudo da seleção natural é que outras forças evolutivas também moldam a variação genética. A história demográfica das populações – incluindo os padrões de migração e mudanças nos tamanhos populacional ao longo do tempo – podem influenciar a variação genética de modo semelhante à seleção natural [9].

Tabela 1. Medidas de variabilidade genética em amostras populacionais e sua expectativa sob seleção balanceadora.

Medidas de variação	Definição do cálculo	Expectativa sob seleção balanceadora
Taxa de Heterozigose	$1 - \sum p_i^2$ onde p_i é a frequência do alelo i . É uma medida de diversidade da amostra	Maior do que esperado com base na taxa de mutação e tamanho populacional.
Divergência entre alelos	Medida do número de diferenças nucleotídicas entre sequências de DNA de diferentes alelos	Maior que esperado para alelos de genes neutros.
Desequilíbrio de ligação entre genes	Medida do grau de associação entre os alelos de diferentes <i>loci</i> . Geralmente não há desequilíbrio de ligação entre <i>loci</i> fisicamente distantes.	No caso de seleção favorecendo combinações de alelos em diferentes <i>loci</i> , espera-se um desequilíbrio de ligação mais alto do que o observado para genes neutros.
Taxa de substituição não-sinônima e sinônima	O número de mudanças sinônimas e não sinônimas em relação ao número de sítios sinônimos e não sinônimos	No caso da seleção favorecendo a diversificação funcional dos alelos, a taxa não-sinônima pode exceder a sinônima.

Por exemplo, uma população cujo crescimento é recente possui padrões de variação genética similares aos observados num *locus* favorecido por seleção natural. Isso faz sentido, se considerarmos que um gene vantajoso se tornou comum na população de modo relativamente rápido, e tal aumento de frequência é justamente aquilo que acontece quando há uma expansão populacional.

Outro caso em que a história demográfica e seletiva de uma população é difícil de distinguir ocorre quando há subdivisões populacionais (isso é, quando um grupo que consideramos uma população é na verdade um conjunto de subpopulações parcialmente diferenciadas). Nesse caso, a diferenciação entre as populações pode resultar num aumento de variação genética na população “geral” (formada pelas várias subpopulações), gerando um efeito semelhante ao da seleção balanceadora.

Como separar os aspectos da variação genética que resultam dos efeitos seletivos daqueles que resultam da história demográfica? Felizmente, há uma solução para esse desafio. A seleção natural molda as características de um ou poucos *loci*, enquanto os efeitos demográficos afetam o genoma como um todo.

Conseqüentemente, se estamos interessados em testar uma hipótese sobre como a seleção atuou sobre os genes de função imunológica, podemos comparar seus padrões de variação com aqueles de *loci* espalhados pelo genoma. Se, por exemplo, o aumento da variabilidade observado para genes de função imunológica não for observado para regiões não-codificadoras, temos indícios de que o aumento da variação é consequência da seleção, e não da história demográfica das populações [10].

Hipóteses seletivas para genes MHC

A variação no repertório de genes MHC de um indivíduo está correlacionada com o conjunto de peptídeos que ele é capaz de apresentar e, conseqüentemente, com o conjunto de patógenos aos quais ele poderá iniciar uma resposta adaptativa. Esse raciocínio levou Doherty e Zinkernagel (1975) a proporem que camundongos heterozigotos teriam “vigilância imunológica elevada” para a região do MHC [11].

Conseqüentemente, a seleção natural favoreceria camundongos heterozigotos para genes MHC, uma vez que eles seriam capazes de responder a uma gama mais ampla de patógenos. A seleção natural que favorece heterozigotos é chamada de sobredominância ou vantagem do heterozigoto.

Outra teoria para explicar a alta diversidade dos genes MHC é a seleção dependente de frequência. Neste caso, os alelos de um gene MHC são seletivamente vantajosos quando raros e menos favorecidos quando comuns [12]. Há uma lógica evolutiva para isso: os patógenos estariam menos aptos para “escapar” das moléculas MHC raras, pois teriam sido menos expostos a elas. Isso geraria uma vantagem seletiva aos portadores das moléculas de MHC raras, que se tornariam progressivamente mais comuns.

Tornando-se comuns, essas moléculas MHC entrariam em contato com os patógenos mais frequentes, aumentando a probabilidade de a seleção favorecer um patógeno com uma mutação de escape. Isso reduziria a vantagem adaptativa dos portadores dessas moléculas, fazendo com que elas abaissem de frequência, dando início a um novo ciclo [12].

Entretanto, a função de genes HLA na apresentação de peptídeos também é compatível com outros processos seletivos. Hedrick (2002) argumentou que mudanças nos

regimes de seleção ao longo do tempo, causadas pelas alterações do conjunto de patógenos aos quais uma população está exposta, resultariam numa constante variação da pressão seletiva sobre genes MHC [13]. Esse aspecto dinâmico do processo evolutivo resultaria na alta diversidade nesses genes, mas não pressupõe a vantagem de heterozigoto.

Outro cenário seletivo envolve a distribuição geográfica dos patógenos: se os patógenos em diferentes regiões do mundo são distintos, as pressões seletivas sobre os genes MHC refletiria a heterogeneidade espacial dos regimes seletivos.

Evidências populacionais de que genes MHC estão sob seleção natural

Motivados pela hipótese de Doherty e Zinkernagel (1975) [11], e pelos avanços no conhecimento funcional dos genes MHC, diversos estudos populacionais testaram a hipótese de que esses genes estão sob seleção natural (resumos na tabela 2).

- Genes HLA são mais diversos do que o esperado sob neutralidade. Hedrick e Thomson (1983), usando dados sorológicos, mostraram que a maior parte das populações humanas possui uma variabilidade maior para os genes HLA-A, B e C do que o esperado sob neutralidade [14]. Posteriormente, esse resultado foi estendido para os genes HLA de classe II apenas para DPB1 não há evidências de seleção balanceadora [10];
- Alelos de genes HLA são altamente diferenciados. Os alelos de genes HLA são muito diferentes uns dos outros. Por exemplo, a diferença entre dois alelos de *HLA-B* podem chegar a 4%, mais que uma ordem de grandeza acima da média genômica (~0,08%). Um estudo recente, comparando a divergência entre genes HLA com os demais *loci* do genoma humano, não encontrou nenhuma região cuja diferenciação fosse similarmente elevada [15]. A seleção balanceadora pode explicar a existência de alelos HLA altamente diferenciados nas populações humanas: os alelos que sofrem mutação, por apresentem “novos” antígenos, teriam uma maior chance de persistir na população e não seriam perdidos por deriva;
- Taxas de substituição não sinônimas são elevadas na fenda apresentadora de peptídeos. Substituições não sinônimas alteram o aminoácido especificado por um códon e, por isso, são normalmente removidas pela ação da seleção natural. Já as sinônimas não alteram a aptidão dos seus portadores (pois não mudam o aminoácido especificado por um códon) e não são eliminadas pela seleção. Portanto, as taxas sinônimas são geralmente mais elevadas do que as não sinônimas. Entretanto, nos códons responsáveis pela fenda apresentadora de

peptídeos dos genes HLA, a situação se inverte: as taxas de substituição não sinônimas são mais elevadas [16];

Tabela 2. Resultados esperados para medidas de variabilidade genética em amostras vindas de populações evoluindo sob seleção balanceadora e resultados obtidos em estudos de genes HLA

Medida de variação	Expectativa sob seleção balanceadora	Resultados para HLA
Taxa de heterozigose	Maior do que o esperado com base na taxa de mutação e tamanho populacional	Muitas populações humanas apresentam maior diversidade do que esperado para os genes HLA-DQB1, DRB1, A, B e C. Não há evidência de seleção balanceadora em DPB1
Grau de diferenciação entre alelos	Maior que esperado para alelos de genes neutros	Não há nenhum <i>locus</i> no genoma humano com alelos tão diferentes quanto os de genes HLA
Desequilíbrio de ligação entre genes	No caso de seleção favorecendo combinações de alelos em diferentes <i>loci</i> espera-se um desequilíbrio de ligação mais alto do que o observado para genes neutros	Há desequilíbrio de ligação significativo entre genes clássicos fisicamente distantes na região do MHC de humanos.
Taxas de substituição não-sinônima e sinônima	No caso da seleção favorecendo a diversificação funcional dos alelos, a taxa não sinônima pode exceder a sinônima	A taxa não-sinônima é mais elevada do que a sinônima nos códonos envolvidos na apresentação de peptídeos

- Desequilíbrio de ligação elevado: Alelos de diferentes genes HLA estão associados de maneira não aleatória, originando haplótipos, muito mais comuns do que o esperado pelas frequências alélicas dos genes individuais. Huttley *et al.* (1999) demonstraram que o nível de desequilíbrio de ligação na região do MHC de humanos é excepcionalmente alto, quando comparado com outras regiões do genoma [17]. Mais recentemente, estudos com SNPs na região do MHC mostram que há extensos

haplótipos com baixa variação ao redor de genes clássicos (revisado em [18]), indicando que alelos vantajosos devem ter aumentado de frequência recentemente.

Em conjunto, esses resultados indicam que os genes MHC foram alvos de seleção natural. No entanto, Garrigan e Hedrick (2003) recomendam cautela na interpretação de resultados de análises evolutivas [19]. Em primeiro lugar, em muitos casos não há evidências de seleção natural (por exemplo, mais da metade das populações estudadas para DQB1I não apresentam desvios da neutralidade). Uma explicação para esse resultado é o baixo poder estatístico dos testes de neutralidade: mesmo quando há seleção, os testes podem deixar de detectá-la.

Por outro lado, para certos testes há também o risco de falsos positivos: a seleção natural não está atuando, porém há evidências de seleção. Esse cenário é bastante plausível quando houve intensa seleção no passado, mas esta não atua mais. Nesses casos, o vestígio da seleção nos dados, tal como a alta taxa não-sinônima, demora milhares de gerações para desaparecer, mesmo que a seleção natural não esteja mais atuando [19].

Diante desses desafios, há duas estratégias complementares que podem ser adotadas. Primeiro, recomenda-se o uso concomitante de vários testes de neutralidade, pois eles diferem em seu poder estatístico e tendência de falso-positivos. Segundo, o papel da seleção natural pode ser investigado por meio de estudos que complementam as inferências evolutivas. Por exemplo, quando há uma correlação entre a presença de uma variante genética e o fenótipo (por exemplo, suscetibilidade ou resistência a patógenos), temos uma evidência de que aquela variação genética tem o potencial de ser selecionada, pois ela possui uma manifestação fenotípica.

Evidências populacionais de que genes HLA estão sob seleção mediada por patógenos

Diversos estudos tem mostrado que a suscetibilidade a doenças infecciosas e à variabilidade nas características da resposta imune possuem componentes genéticos (revisado em [20,21]), com um importante envolvimento dos genes HLA.

Há, atualmente, uma convincente lista de alelos HLA que mostram clara associação com doenças infecciosas. Por exemplo, alelos comuns do oeste africano mostram associação com proteção à malária, fornecendo evidências de seleção em genes HLA [22].

Associação de alelos HLA com HIV mostram que algumas linhagens HLA de classe I, como B*27 e B*57, estão associados com menor progressão da doença [23]. Já os alelos

B*35 e Cw*04 mostram forte associação com progressão mais rápida da doença [24]. Recentemente, Gao *et al.* (2001) demonstraram que uma única substituição de aminoácido entre alelos HLA pode ter efeito substancial na taxa de progressão da Aids [25].

Associações entre genes HLA e doenças infecciosas são consistentes com a ideia de que patógenos exercem uma pressão seletiva sobre esses genes. Nas seções a seguir discutiremos o que os estudos de associação revelam sobre a coevolução de patógenos e genes HLA. Começaremos discutindo exemplos da seleção favorecendo heterozigotos e seleção dependente de frequência. Depois, apresentaremos alguns exemplos da relação entre polimorfismos de genes HLA e suscetibilidade e/ou resistência. Finalmente discutimos a relação entre estudos de associação e a diferenciação entre populações.

Vantagem de heterozigotos para HLA em doenças infecciosas

Demonstrar que indivíduos heterozigotos para genes HLA apresentam uma vantagem frente a doenças infecciosas representaria um forte apoio à hipótese de que a seleção balanceadora contribui para a diversidade desses genes. Carrington *et al.* (1999) demonstraram uma associação altamente significativa entre a taxa de homozigose dos genes HLA de classe I e uma progressão mais rápida para Aids em indivíduos caucasóides e afro-americanos [24]. Os autores verificaram que a progressão da Aids era mais pronunciada em indivíduos homozigotos para mais de um *locus*, sendo menor quando apenas um dos *loci* HLA era homozigoto e maior quando dois ou três *loci* eram homozigotos [24].

No caso de resistência ao gene da hepatite B (HBV) uma maior taxa de homozigose em *loci* de classe II aumenta o risco de persistência desse vírus. Em HCV-infectados, Hraber *et al.* (2007) mostraram que entre os heterozigotos para o gene HLA-DRB1 havia menos infectados do que entre os homozigotos, indicando um papel protetor da condição heterozigota [26].

Eles também demonstraram que as diferenças foram mais pronunciadas com alelos representados como supertipos (definidos com base na especificidade da região âncora principal) do que com genótipos de baixa resolução, indicando que portadores de diferentes supertipos para alelos de DRB1 teriam vantagem contra a progressão da infecção por HCV.

Esses resultados mostram que há doenças para as quais encontramos associações entre heterozigotos e proteção de modo consistente com um modelo de vantagem de heterozigotos.

Seleção dependente de frequência em doenças infecciosas

O modelo de seleção dependente de frequência é particularmente plausível no caso de doenças com agentes infecciosos com alta taxa de mutação como vírus. Nesses casos, há grande probabilidade de surgir uma “mutação de escape” no genoma do agente infeccioso, tornando-o resistente à apresentação pelas moléculas MHC mais comuns na população.

O trabalho de Trachtenberg *et al.* (2003) ilustra esse cenário [27]. Esses autores encontraram uma correlação entre a frequência dos supertipos HLA e a carga viral do HIV em indivíduos norte-americanos. Esse resultado sugere que ocorreu adaptação do HIV aos alelos mais frequentes da população, proporcionando uma vantagem seletiva para aqueles indivíduos que expressam alelos HLA raros.

Apesar de instigante, esse resultado apresenta limitações. A principal delas é a correlação negativa entre a frequência dos alelos HLA e a carga viral dos pacientes era quase inteiramente explicada pela baixa carga viral numa única classe de alelos, aqueles pertencentes ao supertipo B*58. Além disso, não havia evidencia de carga viral aumentada nos supertipos mais comuns, como seria esperado sob a hipótese de seleção dependente de frequência.

Coevolução entre patógenos e hospedeiros

Patógenos com mutações que evitam a sua apresentação por moléculas MHC são favorecidos pela seleção natural; de modo simétrico, são favorecidos os hospedeiros com moléculas MHC capazes de apresentar os patógenos resistentes. Esse cenário define um processo coevolutivo [28].

Mutações de escape em vírus são favorecidas pela pressão seletiva exercida pelo sistema imune. Elas podem fazer com que um peptídeo viral deixe de ser apresentado por moléculas MHC e, conseqüentemente, que as células infectadas não sejam reconhecidas pelos linfócitos T citotóxicos. As mutações de escape tendem a ser mantidas após a transmissão para novos hospedeiros que compartilham a mesma molécula MHC.

Todavia, a capacidade de uma mutação de escape persistir numa linhagem viral é limitada pelo valor funcional ou estrutural do resíduo alterado para a perpetuação do vírus [28]. Como exemplo, alguns trabalhos têm revelado que indivíduos portadores dos alelos B*57 ou B*5801 apresentaram associação com menor progressão de infecção por HIV. Essa associação parece estar relacionada com mutação de escape viral.

As mutações T242N no epítopo viral GagTW10 do HIV e A163X (X=G, N, D ou S) no epítopo viral KF11 não são reconhecidas por moléculas codificadas por esses alelos HLA e,

consequentemente, não são apresentadas na superfície celular [29,30]. Contudo, essas mutações ocorrem em domínios importantes das proteínas viral e resultam numa menor carga viral e numa menor progressão da infecção por HIV [31].

Chooopera *et al.* (2008) encontraram as mutações T242N e A146X (X= P ou S) dos epitopos virais em indivíduos B*57 ou B*5801 negativos [32]. Os autores sugerem que a transmissão pode ter ocorrido a partir de indivíduos B*57 ou B*5801 positivos. A observação de que alguns indivíduos apresentam a mutação viral H219Q, uma mutação compensatória que recupera o valor adaptativo do vírus com a mutação T242N, sugere que possa estar ocorrendo nestes indivíduos B*57 ou B*5801 negativos uma reversão para o fenótipo viral ancestral.

Outro exemplo é visto para o vírus Epstein-Barr. Esse vírus contém sequências de proteínas que, quando apresentadas para células T por moléculas HLA-A11, desencadeiam a resposta imune. Campos-Lima *et al.*, (1997) descobriram que essas sequências são comuns em regiões geográficas nas quais o alelo A11 é raro [33].

Já a forma alterada da proteína viral (que apresenta a mutação de escape) é encontrada nas regiões onde o alelo A11 é comum. Essas mutações de escape estão associadas a perda do controle imune e menor responsividade de células CD8+. Estudos recentes tem mostrado o papel dos polimorfismos HLA na transmissão e no acúmulo dessas variantes de escape em doenças como hepatite C e HIV, e sua relação com resposta imune em nível populacional e como limitador da evolução viral.

Variações de sequência em parasitas de malária também podem ser consequência da pressão seletiva exercida pelo sistema imune. A principal proteína de superfície de esporozoítos (forma parasitária inoculada pelos mosquitos), a proteína CS, apresenta polimorfismos nas duas regiões da proteína envolvidas no reconhecimento de epítomos por células CD4+ e CD8+, e há um excesso de substituições não sinônimas. Esses resultados indicam que essas regiões da proteína de superfície do esporozoíto estão sob pressão seletiva [22].

Variação genética nas populações de hospedeiros e parasitas

Variação genética e geográfica dos parasitas pode fazer com que o mesmo alelo HLA confira diferentes respostas frente à infecção em populações de áreas geográficas distintas. Por exemplo, o alelo HLA-A2 apresenta associação com a alta severidade e persistência viral nas infecções por HBV em japoneses. Contudo, esse mesmo alelo confere maior eliminação do HBV na população russa. O alelo Cw*04 foi associado à persistência a infecção por HCV

em euro-americanos, mas foi associado à proteção nas populações japonesas e russas (revisado em [34]).

Thio *et al* (2001) estudaram o gene DQB1 em afro-americanos e em euro-americanos todos infectados por HCV [35]. O alelo DQB1*0301 foi associado à eliminação viral em afro-americanos enquanto o mesmo alelo foi a persistência viral em euro-americanos. Os alelos DQB1*0101 e *0501, por outro lado foram associados a eliminação viral em euro-americanos.

Essas incongruências nos padrões de associação podem estar relacionadas com a variação genética dos agentes infecciosos nas diferentes regiões geográficas, como evidenciadas por estudos feitos com outros sistemas genéticos.

Por exemplo, Lively e Dybdahl (2000) infectaram populações de caramujo *Potamopyrgus antipodarum* com o trematódea *Microphallus* sp. E observaram que hospedeiros com genótipos mais frequentes são mais comumente infectados por parasitas simpátricos, isto é, parasitas que apresentam a mesma origem geográfica que os hospedeiros [36]. Os caramujos com genótipos mais raros eram menos susceptíveis à infecção pelos parasitas simpátricos.

Esses resultados sugerem que os parasitas tendem a se adaptar ao genótipos hospedeiros mais comuns. De modo semelhante, as diferenças nos padrões de associação dos vírus de hepatite B e C, em populações humanas distintas, poderiam ser reflexo da adaptação desses patógenos aos genótipos HLA mais comuns nessas populações [21].

Abordagens experimentais no estudo da seleção natural sobre os genes da região do MHC

Uma forma de investigar as consequências evolutivas de variação em genes da região do MHC é comparar como diferenças nessa região influenciam a resposta a infecções impostas em condições controladas.

Utilizando caramujo, Penn *et al.* (2002) infectaram linhagens heterozigotas apenas para a região do MHC com as bactérias *Salmomella* entérica ou *Listeria monocytogenes* [37]. Eles observaram que os heterozigotos, em média, respondiam à infecção de maneira mais eficiente que os homozigotos, eliminando patógenos e apresentando sintomas menos intensos. Porém, os heterozigotos não apresentavam uma sobrevivência maior que o genótipo homozigoto mais eficiente. Isso sugere que um efeito de dominância, em que o alelo mais vantajoso confere aos indivíduos que o possuem uma melhor resposta imune tanto em homozigose quanto em heterozigose.

Uma vez que a resistência conferida aos patógenos pelos genótipos MHC se revelou dominante nas infecções por um único patógeno, seria possível que numa infecção dupla os heterozigotos apresentassem um maior valor adaptativo. Para testar essa hipótese, camundongos foram infectados simultaneamente com *S. enterica* e o vírus murinho encefalomielite de Theiler (TMEV) [38].

Durante as coinfeções, os heterozigotos mostraram-se mais eficientes na resposta imunológica do que os homozigotos. Essa situação se manifestou exclusivamente nos indivíduos que apresentavam alelos cujos padrões de suscetibilidade fossem opostos, isto é, quando um alelo conferia resistência a uma infecção e susceptibilidade à outra e vice-versa.

Esse resultado poderia explicar a vantagem de heterozigoto em humanos frente a infecções causadas por HIV, por exemplo. O HIV apresenta uma alta taxa de mutação, gerando um cenário parecido com o que ocorre quando há coinfeções.

Ilmonen *et al* (2007) investigaram de que modo a variabilidade de genes dentro e fora do MHC interage para produzir fenótipos de resistência e suscetibilidade [39]. Esses autores confirmaram a influencia dos genótipos MHC nos padrões de resistência à *S. enterica*, mesmo em animais com maior variabilidade em outras regiões além do MHC.

Entretanto, em contraste com os trabalhos anteriores, o padrão de resistência revelou-se recessivo com relação aos genótipos MHC, ao invés de dominante ou sobredominante. Além disso, uma análise de indivíduos em ambientes seminaturais (nos quais a exposição aos patógenos era aumentada) mostrou que os heterozigotos não eram superiores aos homozigotos.

Dessa forma, apesar de confirmar a associação do MHC com resistência à patógenos, o experimento mostrou que a vantagem de heterozigotos vista em ensaios com linhagens congênicas não ocorria em estudos que utilizavam animais com maior variabilidade no restante do genoma.

Existem inúmeros genes que estão envolvidos na resposta imunológica, seja ela inata ou adaptativa. Apesar disso, são detectadas diferenças nas capacidades de resposta imune quando a única região variável do genoma é o MHC. Porém os trabalhos experimentais que verificaram essa influencia apresentam como maior limitação o fato de terem utilizado linhagens congênicas de camundongos.

O estudo com organismos menos endocruzados, mais parecidos com as populações naturais, mostrou um resultado conflitante quanto ao padrão de suscetibilidade dos genótipos. Os resultados, como afirmam os próprios autores, colocam em dúvida todos os experimentos feitos com linhagens congênicas. Além disso, a maioria dos estudos foi

realizada utilizando poucas espécies de patógenos, o que também pode ter influenciado esses resultados.

Portanto, apesar de revelarem que genes MHC influenciam as respostas imunológicas nos animais testados, as abordagens experimentais não permitem distinguir entre diferentes padrões de herança da resistência e/ou suscetibilidade a doenças (dominante, recessivo ou sobredominante) e carecem de realismo biológico em diversos aspectos (por exemplo, por tipicamente usarem linhagens congênicas e infectarem os animais com uma ou poucas linhagens de patógenos).

Abordagens teóricas no estudo de genes MHC

A diversidade de abordagens descritas neste capítulo documenta de modo convincente que genes MHC foram alvos da seleção natural. Entretanto, o mecanismo seletivo que atua sobre esses genes permanece uma questão em aberto. Conforme detalhado no item “Hipóteses seletivas para genes MHC”, modelos de seleção favorecendo heterozigotos, dependentes de frequência, ou de seleção variável ao longo do tempo e no espaço, são todos capazes de explicar os padrões de variação observados.

Distinguir entre esses modelos é biologicamente importante, pois, por mais que suas consequências evolutivas sejam similares (o aumento da variabilidade nos genes MHC, entre outros fatores), os processos biológicos podem ser substancialmente diferentes. Por exemplo, num modelo simples de vantagem de heterozigoto [40] o que é favorecido seletivamente é a própria condição de ser um heterozigoto para o MHC.

Por outro lado, segundo um modelo de seleção favorecendo alelos raros, a seleção favoreceria alelos MHC raros, uma vez que os patógenos seriam selecionados para evadir a apresentação pelas moléculas mais comuns na população, dessa forma favorecendo diversas variantes mais raras.

A possibilidade de distinguir entre esses dois modelos foi alvo de estudos teóricos [41,42] que utilizam simulações em computador para prever a diversidade esperada no MHC sob os dois modelos (vantagem de heterozigotos e coevolução entre patógenos e hospedeiros).

As simulações revelaram que as condições necessárias para que a vantagem de heterozigotos explique a grande variabilidade dos genes MHC são muito restritivas: todos os heterozigotos teriam que possuir genótipos com aptidão semelhante entre si e uniformemente superior aos homozigotos. Já o modelo de coevolução de patógenos e hospedeiros revelou-se muito mais verossímil pois uma simulação que incorporava a

capacidade do patógeno adaptar-se ao hospedeiro resultava na manutenção de uma grande variabilidade alélica no MHC, ainda que não tenha sido a condição heterozigota em si que tenha sido selecionada, mas sim a combinação de alelos do genótipo.

Esses estudos teóricos têm importantes implicações sobre como enxergamos a variabilidade em genes MHC. Diferentemente de uma visão bastante difundida, não é a variabilidade em si que é favorecida por seleção. Ao invés disso, seriam alelos específicos, cuja vantagem seletiva resulta da pressão exercida pelos patógenos, que seriam favorecidos.

Uma vez que os patógenos em si são variáveis e estão constantemente evoluindo, a pressão resulta na manutenção não de um, mas de vários alelos diferentes nos genes MHC. Desta forma, os estudos de Borghans *et al.* (2004) e Boer *et al.* (2004) sugerem que a seleção natural, sem dúvida, aumenta a variabilidade dos genes MHC, mas isso é subproduto da coevolução de patógenos e hospedeiros, e não uma consequência direta da vantagem de ser heterozigoto [41,42].

Considerações finais

Os resultados de estudos populacionais deixam claro que a seleção natural atuou sobre os genes do MHC. Todavia, a perspectiva evolutiva não documenta a atuação da seleção em populações contemporâneas; ao invés disso, os resultados permitem inferir que os genes MHC foram selecionados no passado, resultando em níveis de variação nas populações atuais que não são compatíveis com a neutralidade. Além disso, esses resultados não identificam o tipo de seleção que atua ou atuou (dependente de frequência, sobredominante, ou variável ao longo do tempo e espaço).

Os estudos de associação, por outro lado, examinam populações contemporâneas e testam hipóteses de haver uma correlação entre determinadas variantes genéticas e fenótipos. A rica documentação de tais associações não deixa dúvida de que a variação em gene MHC tem consequências fenotípicas, que devem ser alvo da seleção natural.

Há também exemplos convincentes de que as mudanças evolutivas em patógenos resultam de seleção exercida pelas características imunológicas do hospedeiro. Entretanto, dentre os estudos de associação que revisamos, encontramos poucos que documentam a vantagem de heterozigotos e um único que é consistente com seleção dependente de frequência.

A relativa escassez de estudos apoiando esses dois regimes de seleção não é surpreendente: amostras populacionais precisam ser imensas para detectar seleção em

populações, com exceção dos casos de seleção extremamente intensa. Além disso, a seleção dependente de frequência é um processo dinâmico, e, mesmo que ela esteja ocorrendo, o “retrato” instantâneo das frequências alélicas e fenotípicas não seria necessariamente capaz de evidenciá-la.

As abordagens experimentais, nas quais genótipos e pressões seletivas são manipuladas, podem complementar os estudos populacionais e de associação. Os trabalhos que revisamos mostram que infecções com múltiplos patógenos revelam a vantagem do heterozigoto, mas este efeito é perdido em estudos com animais de background genético exocruzado. Esse resultado deixa claro que a vantagem de heterozigoto não é um padrão geral, e que em situações biologicamente plausíveis o regime de seleção pode favorecer alelos individuais, e não genótipos heterozigotos.

Referencias bibliográficas

1. Lewontin, R.C. *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press, 1974.
2. Harris E. E. e Meyer D. “The molecular signature of selection underlying human adaptations”. *Am. J. Phys. Anthropol.*, 2006, Suppl. 43, pp. 89-130.
3. Tishkoff S.A.; Varkonyi R.; Cahinhinan N.; Abbes S.; Argyropoulos G.; Destro-Bisol G.; Drousiotou A.; Dangerfield B.; Lefranc G.; Loiselet J.; Piro A.; Stoneking M.; Tagarelli A.; Tagarelli G.; Touma E.H.; Williams S.M.; Clark A.G. “Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance”. *Science*, 2001, 293, pp. 455-462.
4. Hamblin M.T. e Di Rienzo A. “Detection of the signature of natural selection in humans: Evidence from the Duffy blood group *locus*”. *Am. J. Hum. Genet.*, 2000, 66, pp. 1669–1679.
5. McEvoy B.; Beleza S.; Shriver M.D. “The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model”. *Hum. Mol. Genet.*, 2006, 15, pp.15.
6. Nielsen R. “Molecular signatures of natural selection”. *Annu. Rev. Genet.*, 2005, 39, pp. 197-218.
7. Eyre-Walker A. “The genomic rate of adaptive evolution”. *Trends Ecol. Evol.*, 2006, 21(10), pp. 569-75.
8. Kimura M. *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press, 1983.
9. Bamshad M. e Wooding S.P. “Signatures of natural selection in the human genome”. *Nat. Rev. Genet.*, 2003, 4(2), pp. 99-111.

10. Meyer D.; Single R.M.; Mack S.J.; Erlich H.A.; Thomson G. "Signatures of demographic history and natural selection in the human major histocompatibility complex *Loci*". *Genetics*, 2006, 173(4), pp. 2121-42.
11. Doherty, P. C. e Zinkernagel, R. M. "Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex". *Nature*, 1975, 256, pp. 50-52.
12. Apanius V.; Penn D.; Slev P.R.; Ruff L.R.; Potts W.K. "The nature of selection on the major histocompatibility complex". *Crit. Rev. Immunol.*, 1997, 17(2), pp. 179-224.
13. Hedrick P.W. "Pathogen resistance and genetic variation at MHC *loci*". *Evolution*, 2002, 56(10), pp. 1902-8.
14. Hedrick P.W. e Thomson G. "Evidence for balancing selection at HLA". *Genetics*, 1983, 104(3):449-56.
15. Bubb K.L.; Bovee D.; Buckley D.; Haugen E.; Kibukawa M.; Paddock M.; Palmieri A.; Subramanian S.; Zhou Y.; Kaul R.; Green P.; Olson M.V. "Scan of human genome reveals no new *Loci* under ancient balancing selection". *Genetics*, 2006, 173(4), pp. 2165-77.
16. Hughes A.L. e Nei M. "Pattern of nucleotide substitution at major histocompatibility complex class I *loci* reveals overdominant selection". *Nature*, 1988, 8;335(6186), pp. 167-70.
17. Huttley G.A.; Smith M.W.; Carrington M.; O'Brien S.J. "A scan for linkage disequilibrium across the human genome". *Genetics*, 1999, 152(4), pp. 1711-22.
18. Traherne J.A. "Human MHC architecture and evolution: implications for disease association studies". *Int. J. Immunogenet.*, 2008, 35(3), pp. 179-92.
19. Garrigan D. e Hedrick P.W. "Perspective: detecting adaptive molecular polymorphism: lessons from the MHC". *Evolution*, 2003, 57(8), pp. 1707-22.
20. Cooke G.S. e Hill A.V.S. "Genetics of susceptibility to human infectious disease". *Nat. Rev. Genet.*, 2001, 2 (12), pp.967-77.
21. Frank S. "Immunology and Evolution of Infectious Disease". Princeton (NJ): Princeton University Press., 2002, pp.1-348.
22. Hill A.V.; Jepson A.; Plebanski M.; Gilbert S.C. "Genetic analysis of host-parasite coevolution in human malaria". *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 1997, 29;352(1359), pp. 1317-25.
23. Kaslow R.A.; Carrington M.; Apple R.; Park L.; Muñoz A.; Saah A.J.; Goedert J.J.; Winkler C.; O'Brien S.J.; Rinaldo C.; Detels R.; Blattner W.; Phair J.; Erlich H.; Mann D.L. "Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection". *Nat. Med.*, 1996, 2(4), pp. 405-11.
24. Carrington M.; Nelson G.W.; Martin M.P.; Kissner T.; Vlahov D.; Goedert J.J.; Kaslow R.; Buchbinder S.; Hoots K.; O'Brien S.J. "HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage". *Science*, 1999, 12;283(5408), pp. 1748-52.

25. Gao X.; Nelson G.W.; Karacki P.; Martin M.P.; Phair J.; Kaslow R.; Goedert J.J.; Buchbinder S.; Hoots K.; Vlahov D.; O'Brien S.J.; Carrington M. "Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS". *N. Engl. J. Med.*, 2001, 344(22), pp. 1668-75.
26. Hraber P.; Kuiken C.; Yusim K. "Evidence for human leukocyte antigen heterozygote advantage against hepatitis C virus infection". *Hepatology*, 2007, 46(6), pp. 1713-21.
27. Trachtenberg E.; Korber B.; Sollars C.; Kepler T.B.; Hraber P.T.; Hayes E.; Funkhouser R.; Fugate M.; Theiler J.; Hsu Y.S.; Kunstman K.; Wu S.; Phair J.; Erlich H.; Wolinsky S. "Advantage of rare HLA supertype in HIV disease progression". *Nat. Med.*, 2003, 9(7), pp. 928-35.
28. Moore C.; John M.; James I.R.; Christiansen F.T.; Witt C.S.; Mallal S.A. "Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level". *Science*, 2002, 296(5572), pp. 1439-43.
29. Martinez-Picado J.; Prado J.G.; Fry E.E.; Pfafferott K.; Leslie A.; Chetty S.; Thobakgale C.; Honeyborne I.; Crawford H.; Matthews P.; Pillay T.; Rousseau C.; Mullins J.I.; Brander C.; Walker B.D.; Stuart D.I.; Kiepiela P.; Goulder P. "Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1". *J. Virol.*, 2006, 80(7), pp. 3617-23.
30. Brockman M.A.; Schneidewind A.; Lahaie M.; Schmidt A.; Miura T.; Desouza I.; Ryvkin F.; Derdeyn C.A.; Allen S.; Hunter E.; Mulenga J.; Goepfert P.A.; Walker B.D.; Allen T.M. "Escape and compensation from early *HLA-B57*-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A". *J. Virol.*, 2007, 81(22), pp. 12608-18.
31. Schneidewind A.; Brockman M.A.; Yang R.; Adam R.I.; Li B.; Le Gall S.; Rinaldo C.R.; Craggs S.L.; Allgaier R.L.; Power K.A.; Kuntzen T.; Tung C.S.; LaBute M.X.; Mueller S.M.; Harrer T.; McMichael A.J.; Goulder P.J.; Aiken C.; Brander C.; Kelleher A.D.; Allen T.M. "Escape from the dominant *HLA-B27*-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication". *J. Virol.*, 2007, 81(22), pp. 12382-93.
32. Chopera D.R.; Woodman Z.; Mlisana K.; Mlotshwa M.; Martin D.P.; Seoighe C.; Treurnicht F.; de Rosa D.A.; Hide W.; Karim S.A.; Gray C.M.; Williamson C.; CAPRISA 002 Study Team. "Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage". *PLoS Pathog.*, 2008, 21;4(3),
33. Campos-Lima P.O.; Levitsky V.; Imreh M.P.; Gavioli R.; Masucci M.G. "Epitope-dependent selection of highly restricted or diverse T cell receptor repertoires in response to persistent infection by Epstein-Barr virus". *J. Exp. Med.*, 1997, 7;186(1), pp. 83-9.
34. Singh R.; Kaul R.; Kaul A.; Khan K. "A comparative review of HLA associations with hepatitis B and C viral infections across global populations". *World J. Gastroenterol.*, 2007, 28;13(12), pp. 1770-87.
35. Thio C.L.; Thomas D.L.; Goedert J.J.; Vlahov D.; Nelson K.E.; Hilgartner M.W.; O'Brien S.J.; Karacki P.; Marti D.; Astemborski J.; Carrington M. "Racial differences in HLA class II

- associations with hepatitis C virus outcomes". *J. Infect. Dis.*, 2001, 1;184(1), pp. 16-21.
36. Lively C.M. e Dybdahl M.F. "Parasite adaptation to locally common host genotypes". *Nature*, 2000, 8;405(6787), pp. 679-81.
37. Penn D.J.; Damjanovich K.; Potts W.K. "MHC heterozygosity confers a selective advantage against multiple-strain infections". *Proc. Natl. Acad. Sci. U S A*, 2002, 20;99(17), pp. 11260-4.
38. McClelland E.E.; Penn D.J.; Potts W.K. "Major histocompatibility complex heterozygote superiority during coinfection". *Infect. Immun.*, 2003, 71(4), pp. 2079-86.
39. Ilmonen P.; Penn D.J.; Damjanovich K.; Morrison L.; Ghotbi L.; Potts W.K. "Major histocompatibility complex heterozygosity reduces fitness in experimentally infected mice". *Genetics*, 2007, 176(4), pp. 2501-8.
40. Takahata N.; Nei M. "Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex *loci*". *Genetics*, 1990, 124(4), pp. 967-78.
41. Borghans J.A.; Beltman J.B.; De Boer R.J. "MHC polymorphism under host-pathogen coevolution". *Immunogenetics*, 2004, 55(11):732-9.
42. De Boer R.J.; Borghans J.A.; van Boven M.; Keşmir C.; Weissing F.J. "Heterozygote advantage fails to explain the high degree of polymorphism of the MHC". *Immunogenetics*, 2004, 55(11), pp. 725-31.

ANEXO 3

Script em linguagem R para a função HSH

```

hsh<- function(x,y,simular=FALSE)
{
n=length(x)
freq.norm<- rep(0,n)
hsh.real<- rep(0,n)
for (i in 1:n) # Normalização das frequências alélicas
  {
    freq.norm <-(x/sum(x))
  }
cat("\n\n", "Frequencias alélicas normalizadas")
print(freq.norm) # retorna o resultado das frequências normalizadas
for (i in 1:n) # Estimativa de hsh
  {
    pi2<- freq.norm^2
    hsh.real<- 1-(sum(pi2))
  }
cat("\n\n", "HSH")
print (hsh.real) #retorna o hsh calculado para seus dados

if (simular=="TRUE") # se desejar simular os dados
{
y # informar um conjunto de dados ou distribuição de frequências
sample(y,n, replace=T)-> resample
n.resample <- length(resample)
freq.norm.resample<- rep(0,n.resample)
hsh.resample.vector<- rep(0, 10000)
hsh.resample <- 0
for (j in 1:10000) # repetir 10000 vezes a estimativa de HSH a partir dos dados
embaralhados.
  {

```

```

sample(y,n, replace=T)-> resample # embaralha e reamostra n valores 10000 vz
  for (i in 1:n.resample)
  {
    freq.norm.resample<-(resample/sum(resample)) #normaliza a frequência
    pi2.resample<- freq.norm.resample^2
    hsh.resample<- 1-(sum(pi2.resample)) #calcula hsh das reamostragens
    hsh.resample.vector[j]<- hsh.resample
  }
# cálculo do p-value empirico:
table(hsh.resample.vector>=hsh.real)-> p # Essa tabela mostra a proporção de re-
amostragens com hsh maior ou igual ao hsh observado
pvalue <- p[[1]]/(p[[1]]+p[[2]]) # o p-value vem direto dessa tabela
cat ("\n\n", "pvalue")
print(pvalue)
hist(hsh.resample.vector, xlim=0:1, nclass=20, main= "HSH Distribution", xlab="HSH", ylab=
"Frequency") # versão gráfica do resultado, com o valor empírico na distribuição
abline(v=hsh.real, lwd=3, col= "red") # linha mostrando onde cai o hsh real/observado
}

```