

Márcia Regina Pincerati

Genética de populações em genes de função
imunológica: um estudo em populações Ameríndias

Population genetics of immune function genes: a study
of Amerindian populations

São Paulo

2011

Márcia Regina Pincerati

Genética de populações em genes de função
imunológica: um estudo em populações Ameríndias

Population genetics of immune function genes: a study
of Amerindian populations

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

Orientador: Dr. Diogo Meyer

São Paulo

2011

Pincerati, Márcia Regina
Genética de populações em genes de
função imunológica: um estudo em
populações Ameríndias

Número de páginas: 180

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

1. Seleção natural 2. Populações
Ameríndias I. Universidade de São
Paulo. Instituto de Biotecnologia.
Departamento de Genética e Biologia
Evolutiva.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).
Orientador(a)

À minha família

“E assim, depois de muito esperar, num dia como outro qualquer, decidi triunfar...
Decidi não esperar as oportunidades e sim, eu mesmo buscá-las.
Decidi ver cada problema como uma oportunidade de encontrar uma solução.
Decidi ver cada deserto como uma possibilidade de encontrar um oásis.
Decidi ver cada noite como um mistério a resolver.
Decidi ver cada dia como uma nova oportunidade de ser feliz.
Naquele dia descobri que meu único rival não era mais que minhas próprias limitações e que enfrentá-las era a única e melhor forma de as superar.
Naquele dia, descobri que eu não era o melhor e que talvez eu nunca tivesse sido.
Deixei de me importar com quem ganha ou perde.
Agora me importa simplesmente saber melhor o que fazer.
Aprendi que o difícil não é chegar lá em cima, e sim deixar de subir.
Aprendi que o melhor triunfo é poder chamar alguém de "amigo".
Descobri que o amor é mais que um simples estado de enamoramento, "o amor é uma filosofia de vida".
Naquele dia, deixei de ser um reflexo dos meus escassos triunfos passados e passei a ser uma tênue luz no presente.
Aprendi que de nada serve ser luz se não iluminar o caminho dos demais.
Naquele dia, decidi trocar tantas coisas...
Naquele dia, aprendi que os sonhos existem para tornar-se realidade.
E desde aquele dia já não durmo para descansar... simplesmente durmo para sonhar.”

Walt Disney

AGRADECIMENTOS

Às agências de fomento FAPESP, CNPq e CAPES pelos auxílios concedidos.

Ao meu orientador Prof. Dr. Diogo Meyer pela oportunidade, pela confiança, pelos ensinamentos passados. Ficam aqui o meu agradecimento, respeito e admiração.

Ao Prof. Eduardo José Melo dos Santos por ceder as amostras para que esse trabalho fosse realizado, pelas sugestões e pela orientação fundamental para que esse trabalho fosse concluído. Agradeço também o pessoal do LGHM/UFGA por terem me acolhido tão bem na minha estadia em Belém.

Ao Prof. João Morgante por ter permitido utilização do laboratório fazendo com que esse doutorado fosse possível e pelas gostosas conversas no cafézinho. Além das palavras de incentivo quando estava finalizando a tese. Afinal, reforço positivo é tudo nessa hora!!

À Dr. Mary Carrington por ter aceitado me orientador durante meu estágio no exterior e por ter sido uma fofa querida comigo durante minha permanência no seu laboratório. À Pat Martin por ter supervisionado meu trabalho. Agradeço também a todos do Laboratório de Imunologia Experimental do *National Cancer Institute-Frederick* pela acolhida, ajuda e ensinamentos. Em especial à Fuh-Mei por estar sempre disposta a ajudar, à Smita Kulkarni pela ajuda nos desenhos experimentais e pela companhia nas madrugadas trabalhando, à Rhasmi Thomas pelas conversas e passeios e à Richard Apps por compartilhar idéias e valiosas discussões, além, claro, de ter cedido as amostras para parte análises deste trabalho.

Aos amigos Danilo Augusto e Jan Ševčík por darem suporte durante minha estadia no exterior e serem responsáveis por momentos marcantes da minha vida. Afinal, como esquecer o primeiro boneco de neve ou as tentativas desastradas de chegar no laboratório a pé na neve?. Como esquecer as mensagens nas horas difíceis e os “morning emails”??

À Danilo do Couto Rosa pelo amor e amizade (sem esquecer o dinheiro emprestado e ainda não devolvido) que foram fundamentais nesses quatro anos de doutorado.

Aos amigos antigos e novos de laboratório Rodrigo Ramalho, Kelly Nunes, Fábio Mendes, Gustavo França, Bárbara Bitarello, Maria Helena Maia, Rodrigo Francisco, Nádia Moraes, Billy Requena, Fernanda Almeida, Gisele Dantas, Sofia Marques, Camila Clozato, Ana Carolina Pavan, Juliana Machado Ferreira, Caroline Aires e Felipe Martins, por serem excelentes companheiros de trabalho, fazendo os dias no laboratório sempre uma diversão (não tão sempre assim..hehe). Sem esquecer das estagiárias bonitinhas: Anna Carolina, Vanessa e Gabi. Mas, aqui fica a bronca: cadê minha tese pronta??? Brincadeira...

Agradecimento também às gatas garotas: Ana Carolina Pavan, Beatriz Kohlbach, Daniela Prioli, Juliana Machado Ferreira e Gisele Dantas pelos momentos históricos vividos nas noites paulistanas.

Às amigas e companheiras de apartamento Katia Ktreutz e Cecília Ikegami pelos ótimos momentos vividos, pela amizade e pela paciência de me agüentar na reta final do doutorado.

Aos amigos Gisele Dantas, Maria Helena H. T. Maia e Rodrigo dos Santos Francisco pela incrível amizade e apoio nos momentos difíceis (que não foram poucos). Em especial a Gigi pelos momentos de ombro amigo, por ter cuidado de mim e ajudado na formatação da tese. A “Mano” Maria por ser sempre prestativa nos momentos de apuro, pelas discussões e leituras atenciosas dos textos e por sempre ver o lado positivo. E a mente brilhante e incansável de Rodrigo (popularmente conhecido como Latino) sempre disposto a horas de discussão, fossem elas durante simples conversas ou ligações no meio da noite. Sem esquecer as longas conversas por skype nas madrugadas que espantaram os momentos de solidão e o desespero durante a minha estadia no exterior.

À todos os meus amigos, de perto ou de longe, recentes ou de longa data, por me fazerem a minha vida tão especial.]

Ao meu amigo e amorzinho Júlio Freire pelo apoio e por não ter desistido de mim, mesmo passando 5 meses só me vendo interagir com o computador.

Aos meus tios Maurício, Eliana, Márcia e João pelo amor e patrocínio no início do doutorado. Afinal a espera por uma bolsa nunca é fácil!!

Aos meus irmãos Walkíria, Douglas e Kiko por fazerem parte da minha vida, pelo amor e pelas ajudas nos momentos de apuros. Agradeço também a mais nova integrante da família Berg-Pincerati, a minha sobrinha Hannah, que tem a magia de me fazer feliz só por saber que ela existe

Aos meus pais, pelo amor e incentivo em todos os momentos da minha vida. Eu poderia escrever uma tese de doutorado só descrevendo meu amor por vocês e não seria suficiente para expressar o que sinto por vocês. Vocês são tudo para mim!!!

À Deus.

ÍNDICE

1. INTRODUÇÃO	1
1.1. ESTUDOS DE SELEÇÃO NATURAL	2
1.2. POPULAÇÕES AMERÍNDIAS E DIVERSIDADE GENÉTICA.....	5
1.3. SELEÇÃO E DEMOGRAFIA	7
1.4. GENES CANDIDATOS.....	8
1.4.1. Os genes <i>KIR</i>	9
1.4.1.1. Os receptores de células NK: Função e diversidade entre populações	9
1.4.1.2. Evidências de ação de seleção natural.....	12
1.4.2. O gene <i>HLA-G</i>	15
1.4.2.1. Evidência de seleção em <i>HLA-G</i>	18
2. JUSTIFICATIVA.....	21
3. OBJETIVOS	22
3.1. OBJETIVO GERAL.....	22
3.2. OBJETIVOS ESPECÍFICOS.....	22
4. MATERIAIS E MÉTODOS	23
4.1. AMOSTRA	23
4.1.1. Amostras indígenas.....	23
4.1.2. Amostras de células de trofoblasto.....	23
4.2. EXTRAÇÃO DE DNA.....	25
4.3. TÉCNICA DE AMPLIFICAÇÃO DE GENOMA INTEIRO.....	25
4.4. METODOLOGIA DE ANÁLISE DOS GENES CANDIDATOS	25
4.4.1. <i>KIR2DL4</i>	25
4.4.2. <i>KIR3DL1/S1</i>	28
4.4.3. <i>HLA-G</i>	30
4.4.3.1. Região promotora.....	30
4.4.3.2. O polimorfismo de 14pb	32
4.4.3.3. Região 3'UTR.....	34
4.4.4. Metodologia de análise das regiões presumidamente neutras.....	35
4.5. ANÁLISE DOS RESULTADOS	36
4.5.1. Frequências de portadores, alélicas e genotípicas.....	36
4.5.2. Estimativa da frequência de alelos nulos para os genes <i>KIR</i>	36
4.5.3. Frequências haplotípicas.....	37
4.5.4. Equilíbrio de Hardy-Weinberg.....	37
4.5.5. Desequilíbrio de ligação.....	37
4.5.6. Testes de neutralidade.....	37
4.5.7. Diferenciação inter-populacional.....	39
4.5.8. Inferências demográficas a partir da análise de marcadores neutros	39
5. RESULTADOS.....	40
5.1. GENE <i>KIR2DL4</i>	40
5.1.1. Caracterização da variabilidade alélica.....	40
5.1.2. Determinação das frequências alélicas.....	43
5.1.3. O teste do equilíbrio de Hardy-Weinberg para avaliação das metodologias aplicadas.....	48
5.1.4. Distribuição das frequências alélicas.....	52
5.1.5. Diferenciação inter-populacional (<i>Fst</i>).....	54

5.1.6. Testes de neutralidade.....	61
5.2. GENE KIR3DL1/S1	65
5.2.1. Caracterização da variabilidade alélica.....	65
5.2.2. O teste do equilíbrio de Hardy-Weinberg para avaliação das metodologias aplicadas.....	74
5.2.3. Distribuição das frequências alélicas.....	77
5.2.4. Diferenciação inter-populacional (<i>F_{st}</i>).....	80
5.2.5. Testes de neutralidade.....	84
5.3. O PAPEL DA DEMOGRAFIA	90
5.4. HLA-G.....	94
5.4.1. Região promotora	94
5.4.1.1. Caracterização da variação de sequência e determinação dos haplótipos	94
5.4.1.2. Teste de neutralidade	99
5.4.1.3. Polimorfismo 14pb.....	100
5.4.1.4. Região 3'UTR.....	104
5.4.1.5. Análise das regiões promotora e 3'UTR do gene HLA-G.....	106
5.4.1.5. Investigação da relação entre polimorfismos do gene HLA-G e níveis de expressão da molécula.....	110
6. DISCUSSÃO	111
6.1. CONSIDERAÇÕES SOBRE A ANÁLISE DO GENE <i>KIR2DL4</i>	111
6.2. CONSIDERAÇÕES SOBRE A ANÁLISE DO GENE <i>KIR3DL1/S1</i>	115
6.3. CONSIDERAÇÕES SOBRE AS ANÁLISES DO GENE HLA-G.....	119
7. CONSIDERAÇÕES FINAIS.....	123
8. RESUMO	129
9. ABSTRACT	130
10. REFERÊNCIAS BIBLIOGRÁFICAS	131
ANEXO	149
PROJETO ADICIONAL: Análise populacional do GENE <i>MC1R</i>.....	149

1. INTRODUÇÃO

A identificação dos processos que moldam a variação genética nas populações naturais permanece um problema chave na genética de populações. Mutação, migração, deriva, recombinação, seleção em locos individuais, os efeitos da seleção em sítios ligados e a história demográfica, conjuntamente, participam deste processo, embora a importância relativa desses fatores ainda não seja clara (NACHMAN & CROWELL 2000).

Recentemente, o estudo dos efeitos da seleção natural sobre genes humanos tem assumido maior proeminência, e um grande número de genes sob seleção na história da nossa espécie foi descoberto (ex. VALLENDER & LAHN, 2004). Em paralelo, houve importantes avanços nos métodos empregados para detectar a seleção natural a partir de dados genéticos (NIELSEN, 2005a). Atualmente, há grande esforço voltado para caracterizar regimes de seleção que atuam sobre genes específicos, assim como em distinguir os efeitos da seleção natural e da história demográfica de populações (BAMSHAD & WOODING, 2003).

Dentre as categorias de genes estudadas, os genes do sistema imune parecem ser um importante alvo de seleção natural, pois vem consistentemente aparecendo em estudos evolutivos como uma classe intensamente selecionada (ex. NILSEN *et al.*, 2005b). Evidência de seleção natural positiva é relativamente comum em genes de função imunológica quando comparada com outras classes funcionais de genes, sugerindo que as respostas à patógenos têm exercido uma importante pressão seletiva durante a história evolutiva humana (FERRER-ADMETLLA *et al.*, 2008).

Estudos populacionais constituem uma das formas de se testar a hipótese de que genes específicos ou regiões do genoma estão sob ação de seleção natural. Nesse contexto, o estudo da variação genética em populações Ameríndias apresenta algumas peculiaridades, uma vez que essas populações (1) ocuparam recentemente uma nova

região com novos desafios ambientais; (2) apresentam história demográfica distinta das demais populações do mundo, caracterizada por gargalos populacionais na fundação e pequenos tamanhos efetivos das populações. Essas características tornam a análise de populações indígenas particularmente informativa para estudos de seleção natural e sobre as interações entre processos demográficos e seletivos.

Dessa forma, no presente estudo foram analisados três genes de função imunológica em populações Ameríndias. Foram escolhidos os genes *KIR2DL4*, *KIR3DL1/S1* e *HLA-G* devido aos seus papéis funcionais e fortes evidências de ação de seleção natural em populações anteriormente estudadas. Estudos prévios analisaram os efeitos de seleção sobre populações Africanas, Européias e Asiáticas (TAN *et al.*, 2005; YAWATA *et al.*, 2006; NORMAN *et al.*, 2007). O presente trabalho propõe o estudo dos efeitos da seleção natural sobre populações Ameríndias a fim de investigar se a seleção natural, que parece ser relevante numa escala global deixa uma marca quando populações de escalas geográficas menores e história evolutiva recente, como o caso das populações Ameríndias, são analisadas. Além disso, o presente trabalho pretende examinar a interação dos efeitos de seleção e demografia nos padrões de variação encontrados nessas populações.

1.1. ESTUDOS DE SELEÇÃO NATURAL

De um modo geral, duas classes de fatores explicam a variação genética. Primeiro, a história demográfica das populações, incluindo fatores tais como os tempos de divergência entre populações, seus tamanhos populacionais efetivos, a ocorrência de gargalos e expansões populacionais e a troca de genes entre populações (fluxo gênico). Segundo, a seleção natural também afeta os padrões de variação intra- e inter-populacional. Conseqüências desse processo incluem uma maior diferenciação genética (no gene selecionado) quando os regimes de seleção diferem entre as populações, para o gene sob consideração. Alternativamente, regimes de seleção balanceadora (que favorecem a

manutenção da diversidade em populações) podem resultar numa diminuição da diferenciação entre populações, com relação ao esperado se não houvesse seleção (revisado em BAMSHAD & WOODING, 2003).

A detecção da seleção natural a partir de dados genéticos se transformou profundamente nas últimas décadas, com avanços nos métodos analíticos e do volume de dados genéticos gerados, permitindo estudos mais efetivos e estatisticamente mais poderosos (NIELSEN, 2005a). Duas classes principais de abordagem vêm sendo utilizadas. A primeira consiste em definir a priori genes candidatos a seleção natural, com base no conhecimento funcional ou estudos populacionais prévios. Esses genes são então estudados numa escala populacional e padrões de variação intra e inter-populacionais (e às vezes também os inter-específicos) são analisados utilizando os testes para detectar seleção (SABETI *et al.*, 2002). Tais estudos permitem testar a hipótese de que os genes candidatos estão sob seleção, fazer inferências sobre o regime seletivo ao qual o gene está submetido, e comparar os regimes de seleção entre diferentes populações (HARRIS & MEYER, 2006).

A segunda abordagem consiste em realizar varreduras genômicas. Essa segunda abordagem difere da anterior na medida em que tem como objetivo encontrar genes selecionados sem a definição prévia de candidatos e requer a análise molecular muito mais extensa, e em geral é empreendida utilizando grupos de populações relativamente reduzidos, enfatizando os efeitos de seleção que diferenciam humanos de chimpanzés ou “grandes grupos” geográficos, tais como os africanos, europeus e asiáticos (ex. NIELSEN, 2005b; VOIGHT *et al.*, 2006).

Com a rápida expansão dos dados genômicos e o aumento das ferramentas analíticas, recentes estudos têm demonstrado que a seleção positiva tem um papel importante na evolução do genoma humana (ex. VOIGHT *et al.*, 2006; WILLIAMSON *et al.*, 2007). Estudos de regiões genômicas que têm sido alvo de seleção positiva são importantes para o entendimento da história evolutiva humana e de como as populações se adaptaram

aos diferentes ambientes. Além disso, a descrição dessas regiões é uma poderosa ferramenta para identificar regiões que são funcionalmente relevantes (RONALD & AKEY, 2005).

Sinais de pressão seletiva têm sido descritos em vários genes (AKEY *et al* 2002; VOIGHT *et al*, 2006; MYLES *et al*, 2008). Diferentes abordagens genômicas têm mostrado que a imunidade é uma das classes funcionais mais enriquecidas com genes selecionados positivamente (ANDRÉS *et al.*, 2009). Seleção balanceadora tem sido invocada para explicar o padrão de variação encontrado em alguns sistemas gênicos relacionados com a resposta imune para manter altos níveis de variabilidade genética (HUGHES *et al*, 2005). Exemplos clássicos de seleção balanceadora atuando em genes do sistema imune são os genes *HLA*, cuja variabilidade está entre as maiores descritas na espécie humana. A variação no repertório de genes *HLA* de um indivíduo está correlacionada com o papel dessas moléculas na apresentação de antígenos na resposta imune adaptativa, ou seja, 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 imune (SOLBERG *et al*, 2008). FERRER-ADMETLLA *et al.* (2008) também demonstrou assinaturas de seleção positiva, especialmente de seleção balanceadora, em genes da resposta imune inata, confirmando a importância da plasticidade do sistema imune nos processos evolutivos adaptativos.

Mudanças na regulação gênica também contribuem para adaptação de populações naturais e diferenças na regulação da expressão têm uma importante função na variação fenotípica entre e dentro de espécies (HAHN 2007). Há fortes evidências de que a seleção natural também atua sobre regiões *cis*-reguladoras (WRAY, 2007). Evidências de seleção balanceadora têm sido identificadas em regiões *cis*-reguladoras de genes de função imunológica, tais como *HLA-G* (TAN *et al.*, 2005) e *HLA-DPA1* e *HLA-DPB1* (LIU *et al.*, 2006).

1.2. POPULAÇÕES AMERÍNDIAS E DIVERSIDADE GENÉTICA

As populações Ameríndias apresentam um padrão de variação genética característico, provavelmente resultado das peculiaridades de sua história demográfica. A migração do homem moderno para as Américas representou um dos últimos grandes eventos de povoamento. Evidências vindas de várias ciências foram utilizadas para a elaboração de teorias para explicar a colonização das Américas (KLEIN, 1999). Uma das teorias mais aceitas é de que no final do Paleolítico, marcado por um intenso período glacial, houve a formação de espessas calotas de gelo, fazendo com que os níveis dos oceanos baixassem 140m. Formou-se uma ponte continental onde hoje é o Estreito de Bering, fazendo com que o Alaska e regiões livres de gelo do Canadá se tornassem uma extensão da Sibéria, onde não havia glaciares devido ao clima seco na época. Entre 22.000 a 12.000 anos atrás (sendo a data mais antiga sugerida a partir de estudos de mtDNA e cromossomo Y, e mais recente por evidências arqueológicas e paleoclimáticas) especula-se que chegaram os primeiros habitantes da América, que conseguiram alcançar regiões meridionais do continente a partir de um corredor livre de gelo, ou por uma rota pela costa (KLEIN, 1999).

Os ameríndios formam o grupo populacional com os menores níveis de diversidade genética na espécie humana. Estudos com *mtDNA* (DNA mitocondrial) mostraram que os nativos americanos foram fundados por um número limitado de linhagens maternas que deixaram a Ásia (SCHURR *et al.*, 1990). Outro estudo com *mtDNA* revelou que os ameríndios apresentam níveis de variação três vezes menores que os encontrados em africanos (SALZANO, 2002). WANG *et al.* (2007), utilizando dados de 678 *loci* microssatélites, verificaram que as populações Ameríndias apresentam o menor número de alelos por locus e o menor número de alelos privados, quando comparadas ao resto do mundo. Esses resultados são condizentes com o efeito da colonização serial a partir da África. Como citado acima, a América foi a última das grandes massas de terra a ser

ocupada e, portanto, está no final da série de colonização, sendo precedida por um grande número de efeitos de fundador, que levaram a uma significativa perda de diversidade.

Comparações entre as populações Ameríndias e Asiáticas sugerem que dois haplótipos do cromossomo Y, originários da região central da Sibéria, tenham contribuído para a população fundadora das Américas (revisado por SALZANO, 2002). WANG *et al.* (2007) também mostraram que dentre as populações do velho mundo, as Siberianas são as que apresentam os maiores níveis de similaridade com os ameríndios, indicando que a colonização das Américas provavelmente se deu a partir do nordeste asiático.

Na América do Sul existem claramente dois padrões de distribuição de diversidade, evidenciado por vários marcadores genéticos. Estudos do cromossomo Y em Ameríndios mostraram que diferentes populações da América do Sul apresentam dois padrões distintos de diversidade: as populações andinas são mais homogêneas e com uma maior diversidade intra-populacional quando comparadas às populações da porção oriental do continente (onde ocorre uma grande diversidade inter-populacional, porém, com baixos índices de diversidade dentro das populações individuais). Nos Andes as condições geográficas e culturais favoreceram o fluxo gênico e o estabelecimento de uma cultura relativamente uniforme, que teve seu ápice com o estabelecimento do império Inca. Nas populações da Bacia Amazônica, Planalto Brasileiro e Chaco, o efeito de deriva foi importante, devido aos pequenos tamanhos efetivos e isolamento (TARAZONA-SANTOS *et al.* 2001). Dados de microsatélites autossômicos reforçam essas considerações ao mostrarem que populações da região andina apresentam valores intermediários de F_{st} , enquanto as populações do leste apresentam os valores mais altos (WANG *et al.*, 2007). Dessa forma, esses resultados revelam que as populações da porção ocidental da América do Sul experimentaram níveis mais altos de fluxo gênico e/ou apresentam um maior tamanho efetivo que as populações do leste do continente, onde o efeito de deriva genética foi mais importante devido aos pequenos tamanhos efetivos e isolamento.

Análises de 72 microssatélites em 14 populações do mundo sugerem que os Ameríndios tiveram uma expansão muito mais recente, quando comparados com as populações dos demais continentes (ZHIVOTOVSKY, 2000). Estudos que utilizaram *mtDNA* também mostraram que a população ancestral dos Ameríndios provavelmente sofreu uma rápida expansão em direção ao sul do continente (FAGUNDES *et al.*, 2008; KITCHEN *et al.*, 2008).

A combinação de fatores que moldaram a história de populações Ameríndias torna-as particularmente interessantes para estudos de seleção natural. Primeiro, elas ocuparam novos ambientes num período relativamente recente (fator esse que parece estar associado à presença de assinaturas de seleção em populações de outras regiões do mundo, ex. AKEY *et al.* 2004). Segundo, elas apresentam uma estrutura demográfica extrema entre as populações humanas, com alta diferenciação inter-populacional (ZAGO *et al.*, 1996; SALZANO e CALLEGARI-JACQUES, 1988; WANG *et al.*, 2007), oferecendo um estudo de caso para os efeitos da história demográfica sobre os fatores seletivos.

Além disso, de um modo geral, poucos marcadores autossômicos foram analisados nessas populações. Há estudos para STRs e marcadores clássicos, mas poucos para haplótipos definidos ao nível de DNA. Diante disso, a análise de mais marcadores poderá auxiliar no entendimento do padrão de variação genética e contribuir para a compreensão da história evolutiva e demográfica das populações Ameríndias.

1.3. SELEÇÃO E DEMOGRAFIA

Como mencionado acima, a seleção natural apresenta um papel crucial na modulação dos padrões de variação intra e inter-populacional. Entretanto, a história demográfica também pode deixar padrões similares fazendo com que, muitas vezes, seja

difícil distinguir entre os dois processos. Um ponto chave para estudos populacionais é justamente a distinção entre a ação da seleção natural da história demográfica.

Sabe-se que as diferentes formas de seleção deixam suas assinaturas em locos específicos, enquanto a história demográfica tende a deixar sinais ao longo de todo o genoma dentro de uma população. Os fatores demográficos são os responsáveis pela alteração das frequências alélicas nos genes neutros, exceto nos casos em que esses genes estão ligados a um gene selecionado, tornando-os influenciados indiretamente pela seleção (revisado em BAMSHAD & WOODING, 2003).

Há duas principais formas de lidar com o desafio de distinguir entre os efeitos da seleção e da história demográfica. Uma forma é a realização de testes para detecção de seleção natural, levando-se em conta os possíveis efeitos demográficos, como o que foi realizado no trabalho de AKEY *et al.* (2004). Outra forma é comparar diferentes conjuntos de marcadores, onde alguns estão sob um regime de seleção neutra, e outros possivelmente sob ação de seleção natural. Nessa segunda abordagem são os próprios marcadores neutros que são usados para determinar o papel da demografia no padrão de variação encontrado.

1.4. GENES CANDIDATOS

Como mostrado acima, evidências de seleção natural atuando em várias regiões do genoma humana vêm crescendo e muitos estudos apontam os genes de sistema imune como uma categoria de genes que tem sido alvo de seleção natural (ex. ANDRÉS *et al.*, 2009).

No presente trabalho, foram escolhidos três genes de função imunológica para estudo de seleção natural em populações Ameríndias. Nas seções a seguir serão apresentados os genes candidatos à seleção natural que foram escolhidos para o presente estudo e a motivação para o estudo dos mesmos.

1.4.1. Os genes *KIR*

1.4.1.1. Os receptores de células NK: Função e diversidade entre populações

As células *Natural Killer* (NK) humanas são linfócitos do sistema imune inato que são cruciais na defesa do hospedeiro e na regulação imune. As células NK são derivadas da medula óssea que compartilham o mesmo progenitor com as células T, não expressam receptores de superfície celular antígeno-específico e compreendem cerca de 10-15% dos linfócitos circulantes (revisado em MIDDLETON & GONZALEZ, 2009). Devido à sua produção de citocina e quimiocinas e sua habilidade de lisar células alvo sem requerer uma sensibilização *a priori* (por isso o termo “assassinas naturais”), as células NK são importantes componentes do sistema imune na primeira linha de defesa contra agentes infecciosos virais e células cancerígenas, além de apresentarem uma relevante função na reprodução (ORANGE & BALLAS, 2006).

Apesar de serem classificadas como componentes do sistema imune inato, as células NK não são “armas” independentes do sistema imune, nem são células nas quais faltam especificidade. As células NK têm receptores que reconhecem moléculas HLA de classe I e sua função é altamente integrada com outras células do sistema imune inato e adaptativo (ORANGE & BALLAS, 2006).

Embora as células NK possam ser ativadas de várias maneiras, muitos detalhes do mecanismo de ativação permanecem pouco elucidados. O mecanismo de ativação de células NK melhor caracterizado é a reduzida expressão de moléculas HLA de classe I. A reduzida expressão dessas moléculas na superfície celular é conseqüência comum de infecção intracelular, já que muitos patógenos desenvolveram mecanismos que sabotam a expressão de moléculas HLA de classe I e apresentação de antígeno às células T CD8. Tais mecanismos permitem ao vírus evitar que as células infectadas sejam detectadas por

linfócitos T. Mecanismos similares também são usados por células tumorais, que apresentam mutações que reduzem a expressão de moléculas HLA de classe I (revisado em VILCHES & PARHAM, 2002).

Nas duas últimas décadas, avanços significativos também têm sido feitos no entendimento da regulação das células NK. Vários receptores ativadores e inibidores foram descobertos e acredita-se que o balanço entre os sinais desses receptores possibilita as células NK de fornecerem efetiva função de matar as células alvo enquanto mantém a tolerância ao próprio (CHEENT & KHAKOO, 2009).

Com base na estrutura de proteínas, os receptores de células NK podem ser classificados em dois principais grupos: a superfamília de receptores imunoglobulina-símiles (IgSF) e a superfamília de receptores de lectina tipo C cálcio independente. Pertencente a superfamília de receptores de lectina tipo C está o heterodímero CD94:NKG2A, mapeado no cromossomo 12. No cromossomo 19 está localizado o complexo LCR (*Leukocyte Receptor Complex*) que contém genes que codificam moléculas relacionadas à superfamília das imunoglobulinas, como os receptores KIR. Os receptores KIR são uma família diversa de glicoproteínas de superfície celular que são expressas em células NK e em uma subpopulação de linfócitos T CD8+ e que se ligam às moléculas HLA de classe I (VILCHES & PARHAM 2002).

A família gênica *KIR* é composta atualmente por 15 genes (*KIR2DL1*, *KIR2DL2*, *KIR2DL3*, *KIR2DL4*, *KIR2DL5A*, *KIR2DL5B*, *KIR2DS1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS4*, *KIR2DS5*, *KIR3DL1*, *KIR3DL2*, *KIR3DL3* e *KIR3DS1*) e dois pseudogenes (*KIR2DP1* e *KIR3DP1*) (<http://www.ebi.ac.uk/idp/kir>). A classificação de genes *KIR* é determinada de acordo com o número de domínios extracelulares, dois domínios (2D) ou três domínios (3D) e se possuem a cauda citoplasmática longa (L) ou curta (S). Os receptores com longa cauda citoplasmática contêm imunoreceptores com motivos inibitórios baseados em tirosina (ITIM), enquanto os receptores com cauda citoplasmática curta têm uma função potencialmente ativadora. Similaridades estruturais e de sequência entre as diferentes moléculas KIR definem os três tipos de domínios na proteína, denominados domínios D0,

D1 e D2 (BASHIROVA *et al.*, 2006).

Os genes que codificam os receptores KIR mostram extensivo polimorfismo. Eles estão agrupados em uma das regiões mais variáveis do genoma humano em termos de conteúdo gênico e polimorfismo de seqüência (MIDDLETON & GONZALEZ &, 2009). Haplótipos *KIR* exibem variabilidade em número e tipo de genes apresentados e em polimorfismo alélico de genes individuais (PARHAM *et al.*, 2005). Dois grupos de haplótipos *KIR* humanos são definidos. O haplogrupo *A* tem uma organização fixada de sete genes, a maioria dos quais com potencial inibidor. O haplogrupo *B* tem um número variável de genes *KIR*, muitos dos quais, de função ativadora (UHRBERG *et al.*, 1997; YAWATA *et al.*, 2002). Comparações entre populações revelaram diferenças nas frequências de haplogrupos *A* e *B*. Em populações do leste Asiático, o haplogrupo *A* é dominante. Em populações européias e africanas os haplogrupos *A* e *B* são igualmente abundantes. Já nas populações nativas da Austrália e América, o haplogrupo *B* predomina.

Outra característica marcante dos genes *KIR* é a sua rápida evolução. Uma das hipóteses para explicar a rápida evolução dos genes *KIR* está relacionada ao papel central atribuído à interação KIR:HLA na modulação da resposta imune. Recentes estudos sugerem que as diferenças encontradas entre humanos e outras espécies são conseqüências da rápida evolução dos genes *HLA* de classe I, um processo conduzido por pressões idiossincráticas de patógenos sobre a imunidade do hospedeiro. Nesse modelo, os genes codificando receptores de células NK com especificidade para moléculas *HLA* de classe I clássicas evoluíram rapidamente para seguir a evolução de seus ligantes, mantendo, dessa forma, as interações funcionais apropriadas (KHAKOO *et al.*, 2000; KHAKOO *et al.*, 2002). A idéia de co-evolução entre genes *HLA* e *KIR* também é sustentada por estudos que revelam que associações a doenças dependem das combinações de alelos *HLA* e *KIR* presentes, assim como por estudos populacionais que indicam correlações entre as frequências de genes *HLA* e *KIR* em diferentes populações humanas (NORMAN *et al.* 2007; SINGLE *et al.* 2007; GUINAN *et al.*, 2010).

1.4.1.2. Evidências de ação de seleção natural

Dado que a ativação das células NK é fundamental nas fases iniciais da resposta imune, uma questão relevante é se a grande diversidade haplotípica e o polimorfismo alélico dos genes *KIR* são conseqüências de seleção natural ou de mudanças neutras.

YAWATA *et al.* (2006), através de estudos funcionais e populacionais, mostraram evidência de ação de seleção natural no aumento não usual de um haplótipo A em populações japonesas e a ação de seleção positiva na parte centromérica dos genes *KIR* (contendo os genes *KIR2DL1* e *KIR2DL3*). Em contraste com a região centromérica, a região telomérica (contendo os genes *KIR2DL4*, *KIR3DL1/S1*, *KIR2DS4* e *KIR3DL2*) mostra forte evidência de ação de seleção balanceadora (YAWATA *et al.*, 2006).

NORMAN *et al.* (2007) também mostraram a ação de seleção natural atuando no gene *KIR3DL1/S1* em diversas populações. Os autores também mostraram que a seleção balanceadora vem mantendo antigas três linhagens nesse locus, sendo duas linhagens do receptor *KIR3DL1* (*3DL1*015* e *3DL1*005*) e uma linhagem do alótipo *KIR3DS1*.

Evidência de seleção balanceadora mantendo a diversidade haplotípica de genes *KIR* também foi encontrada na tribo Ameríndia Yucpa, localizada na Venezuela (GENDZEKHADZE *et al.*, 2009). Nesse trabalho, os autores descrevem um possível modelo para explicar a diversidade dos haplótipos *KIR* e a manutenção da mesma pela ação de seleção natural. Segundo eles, haplótipos A e B seriam mantidos nas populações devidos aos seus papéis na imunidade e reprodução, respectivamente. Os haplótipos A são protetores contra alguns tipos de infecções, mas, são fatores de risco para distúrbios de gravidez, como pré-eclampsia e aborto recorrente. Já certas combinações de haplótipos B seriam protetores de doenças da gravidez. Dessa forma, haplótipos A são principalmente selecionados por seu papel na defesa imune, enquanto os haplótipos B são selecionados pelo seu papel na reprodução. Neste modelo, a seleção balanceadora que tem mantido os

haplótipos A e B em todas as populações seria decorrente de distintas pressões sobre o sistema imune e reprodutivo.

GUINAN *et al.* (2010), analisando uma população irlandesa também encontrou diversidade mais alta do que o esperado em vários genes *KIR*, consistente com história de seleção balanceadora atuando para manter várias variantes alélicas em alta frequência nessa população. Os autores também observaram que a diversidade em genes *KIR* aumenta quando haplótipos contendo polimorfismos funcionais dos genes *KIR2DL3*, *KIR3DL1/S1* e *KIR2DS4* são analisados.

Como citado nos exemplos acima, os genes *KIR* apresentam evidências de seleção natural moldando o padrão de variabilidade desses genes entre populações. Dentre os genes *KIR*, os genes *KIR3DL1/S1* e *KIR2DL4* estão entre aqueles que apresentam características marcantes em termos de diversidade e as maiores e mais significativas evidências de ação de seleção balanceadora (YAWATA *et al.*, 2006; GUINAN *et al.*, 2010).

O gene *KIR3DL1/S1* é o único locus *KIR* que codifica um alótipo ativador (*KIR3DS1*) e/ou um alótipo inibidor (*KIR3DL1*). O alótipo *KIR3DL1* pode estar presente tanto em haplótipos A como em haplótipos B. O alótipo *3DS1* está presente somente nos haplótipos B. O alótipo *KIR3DL1* tem especificidade para moléculas HLA-B (e algumas moléculas HLA-A) com o motivo sorológico Bw4. Embora o ligante para o alótipo *KIR3DS1* não tenha sido definido por ensaios de ligação, estudos populacionais e de associação a doenças sugerem que eles interagem com moléculas HLA-B com isoleucina na posição 80 (THOMAS *et al.*, 2008).

O alótipo *KIR3DL1* possui grande importância na modulação da ativação das células NK e sua frequência é elevada em todas as populações estudadas (SINGLE *et al.* 2007). Polimorfismos nesse locus apresentam consequências fenotípicas nas quais diferentes alelos são expressos em diferentes níveis (alto, baixo e nulo) na superfície de células NK (GARDINER *et al.*, 2001; YAWATA *et al.*, 2006)

O gene *KIR3DL1/S1* tem se destacado em termos de sua importância médica. A associação desse locus com o seu ligante (a molécula HLA-B) mostra-se importante na

modulação da progressão da AIDS e hepatite C, bem como na proteção contra condições auto-imunes, como endometriose e colangite esclerosante primária, além da associação de do alótipo *KIR3DS1* com a proteção contra linfoma de Hodgkin e predisposição ao câncer cervical (KULKARNI *et al.*, 2008).

O gene *KIR2DL4* codifica um receptor para a molécula HLA-G e é expresso em praticamente todas as células NK e, em significativa proporção, em células NK uterinas durante o primeiro trimestre da gravidez. Esse padrão de expressão, aliados a estudos envolvendo associações deste gene com complicações gestacionais, indicam uma importante função deste gene na manutenção do sistema imune, principalmente na interface materno-fetal (SCHELLEKENS *et al.*, 2008).

O gene *KIR2DL4* apresenta algumas particularidades com relação aos demais genes *KIR*: 1) Ele é considerado um gene moldura (*framework*), estando presente em quase todos os genomas; 2) A estrutura genômica é diferente quando comparado aos demais genes *KIR2DL*: a maioria dos genes *KIR2DL* consistem em 9 exons, incluindo o pseudoexon 3. *KIR2DL4* consiste de somente 8 exons, faltando o exon 4. Como os exons 3-5 codificam domínios extra-celulares, isso resulta em uma diferente combinação de domínios; 3) A molécula *KIR2DL4* apresenta uma combinação única do domínio ITIM intracelular e uma argenina carregada positivamente na região transmembrana, sugerindo que *KIR2DL4* exerce tanto as funções inibitória (pela presença do motivo ITIM) como a ativadora (pela presença do aminoácido argenina carregado positivamente) e que pode mediar diferentes funções sob diferentes circunstâncias (GOODRIDGE *et al.*, 2003).

A grande diversidade, as fortes evidências de ação de seleção natural e as características funcionais distintas desses genes fazem deles excelentes candidatos a estudos populacionais. Além disso, apesar de muito já se conhecer sobre a diversidade dos genes *KIR*, as estimativas de diversidade desse complexo provavelmente são subestimadas, o que limita as inferências sobre os possíveis mecanismos evolutivos atuando nesses genes e a real magnitude da diversidade dentro do complexo *KIR*. Essa subestimativa é decorrente do fato que a maioria dos estudos de diversidade de genes *KIR*

em populações investigou apenas a diversidade desses genes com relação ao conteúdo gênico desse complexo, o que permite apenas afirmar se certo locus está presente ou não (SANTOS EJM, comunicação pessoal).

Nesse contexto, o presente estudo propôs a caracterização alélica dos genes *KIR3DL1/S1* e *KIR2DL4*. A caracterização alélica desses genes trará um panorama mais completo de sua diversidade e contribuirá para um melhor entendimento de como as forças evolutivas atuam sobre esses polimorfismos.

1.4.2. O gene *HLA-G*

O MHC (*Major Histocompatibility Complex*) ou, CPH (Complexo Principal de Histocompatibilidade) é uma região cromossômica contendo mais de 200 genes. Localiza-se no braço curto do cromossomo 6, na posição 6p21.3. Encontra-se dividido em três regiões: classe I (mais telomérica), classe II (mais centromérica) e classe III (ROBINSON *et al.*, 2003).

As moléculas HLA de classe I subdividem-se em clássicas e não-clássicas. As clássicas (Ia) são altamente polimórficas e compreendem as moléculas HLA-A, -B e -C (ROBINSON *et al.*, 2003). As não clássicas compreendem as moléculas HLA-E, HLA-F e HLA-G. Os genes *HLA* clássicos e não clássicos diferem em relação ao nível de polimorfismos, distribuição tecidual e função. A função primária de genes *HLA* de classe I clássicos é a apresentação de antígenos, enquanto os genes não clássicos exibem novas funções, todas as quais ainda não tão bem esclarecidas (O'CALLAGHAN *et al.*, 1998).

O gene *HLA-G* é um gene *HLA* de classe I não clássico expresso predominantemente em tecidos placentários que apresenta propriedades imunoreguladoras características, atuando na modulação do sistema imune materno durante a gravidez (OBER *et al.*, 2003).

Durante a gravidez, o sistema imune materno está em contato próximo com células e tecidos do feto semi-alógeno. Portanto, mecanismos específicos devem existir para

modular e moderar o sistema imune materno, de modo que a mãe não rejeite o seu próprio feto (KOVATS *et al.*, 1990). As células trofoblásticas da placenta, as quais se originam do feto, não expressam moléculas HLA de classe I clássicas, exceto pela baixa expressão de HLA-C (KING *et al.*, 1996; 2000). Dessa maneira, as células fetais semi-alogênicas não entram em contato direto com o sistema imune materno. Entretanto, sabe-se que células que não expressam moléculas HLA em sua superfície sofrem lise celular através de células NK. A forte expressão de moléculas HLA-G, juntamente com a expressão de HLA-E e -F na placenta, previne tal evento (ISHITANI *et al.*, 2003). Aberrações nos mecanismos de modulação do sistema imune materno podem conduzir a complicações na gravidez. A presença de certos alelos de *HLA-G* e níveis de expressão de proteínas e RNAm já foram associados com problemas na gravidez como pré-eclampsia, aborto e taxas de sucesso na gravidez após fertilização *in vitro* (HUNT *et al.*, 2005; MENIER *et al.*, 2010).

A descoberta em 1990 de que o gene *HLA-G* era um gene *HLA* não clássico expresso em células da placenta na interface materno-fetal, chamou a atenção de biólogos e imunologistas da reprodução e por vários anos os estudos envolvendo esse gene ficaram focados na capacidade desta molécula em proteger o feto do sistema imune materno. Novos estudos têm demonstrado que a expressão de HLA-G não é restrita somente a células fetais. Em condições não patológicas, HLA-G também é expresso por células epiteliais tímicas adultas, córnea, matriz de unha, entre outros, os quais são considerados sítios imunologicamente privilegiados. Dessa forma, esses estudos mostram a relevância dessa molécula em novos contextos e sua associação em outras condições patológicas, que vão além do sucesso ou insucesso reprodutivo.

Estudos têm demonstrado que HLA-G inibe a resposta imune por diferentes mecanismos. A expressão de HLA-G por células alvo confere resistência contra células *Natural Killer* (NK) ou citotoxicidade mediada por células T citotóxicas CAROSELLA *et al.* (2007a). Esse grupo também demonstrou que HLA-G é capaz de inibir resposta aloproliferativas de células CD4⁺ de três diferentes maneiras: a) como moléculas solúveis; b) como moléculas de membrana; c) como molécula de membrana expressa por células

apresentadoras de antígenos estimuladas. Em todos esses casos, HLA-G induz a diferenciação de células T reguladoras CD4⁺ que são capazes de inibir a reatividade de outras células T. Além disso, experimentos *in vivo* demonstraram que HLA-G interrompe a maturação de células dendríticas, conduzindo à inibição da resposta imune celular, diferenciação de células reguladoras e sobrevivência prolongada de alo-enxerto (FAVIER *et al.*, 2007).

A análise da relação entre HLA-G e tumores é relativamente recente, mas expandiu rapidamente no campo de pesquisa sobre o câncer. A expressão de HLA-G foi primeiramente descrita em tumores sólidos, mas com o passar dos anos o número de artigos publicados de associações de HLA-G e câncer tem aumentado. Atualmente, a expressão de HLA-G tem sido identificada na maioria dos tumores, incluindo câncer de colo, melanoma, câncer renal, câncer de pulmão e mama, glioblastoma e linfomas cutâneos (revisado de CAROSSELA *et al.*, 2007b).

De maneira similar aos tumores, vírus desenvolvem estratégias para escapar da reação imune do hospedeiro. Dessa forma, a indução da expressão de HLA-G por células infectadas por vírus pode ser um mecanismo adicional que ajudam os vírus a subverter a defesa do hospedeiro. Diante disso, estudos foram realizados para verificar a função de HLA-G após infecção com citomegalovírus (CMV), vírus da imunodeficiência humana (HIV) e viroses neurotrópicas. Em pacientes com HIV, a expressão de HLA-G foi regulada positivamente em células T CD8⁺ e monócitos. Enquanto pacientes progressores expressam altos níveis plasmáticos de HLA-G, pacientes não-progressores não expressam. Este fato sugere um novo mecanismo através do qual o HIV escapa da citotoxicidade das células imunes. O mesmo foi verificado em monócitos infectados por CMV, na qual a expressão de formas solúveis de HLA-G é induzida. Estudos em infecções virais neurotrópicas mostram que os vírus da herpes simplex tipo I e vírus rábico também regulam positivamente a expressão neuronal de isoformas de HLA-G em células infectadas e células vizinhas (revisado de CAROSELLA *et al.*, 2008).

1.4.1.1. Evidência de seleção em *HLA-G*

A região codificadora do gene *HLA-G* é praticamente monomórfica, o que é notavelmente contrastante com os genes altamente polimórficos *HLA* de classe Ia e II. Entretanto, polimorfismos nas regiões não codificadoras foram descritos, a maioria dos quais na região promotora (TAN *et al.* 2005).

O fato de altos níveis de expressão de *HLA-G* serem protetores na gravidez e esses mesmos altos níveis contribuírem para doenças imuno-mediadas sugere que deve haver um balanço ótimo nos níveis de expressão do feto para a vida adulta. Essas considerações motivaram estudos para a caracterização das regiões reguladoras do gene *HLA-G*.

A região promotora de *HLA-G* é muito diferente daquela dos demais genes *HLA*, com vários dos elementos *cis*-reguladores conservados em todos os outros genes *HLA* sendo interrompidos ou deletados em *HLA-G*. Por exemplo, o acentuador A, o módulo SXY e o elemento regulador estimulado por interferon (ISRE) adquiriram mutações que os tornaram não-funcionais. Como consequência, o gene *HLA-G* não responde a NF- κ B e trans-ativadores de classe II e é menos responsivo a interferon- γ (IFN- γ) quando comparado com outros genes *HLA*. Além disso, enquanto todos os elementos reguladores conhecidos estão contidos dentro de aproximadamente 500pb a montante do exon 1, os elementos reguladores de *HLA-G* estendem-se a 1500pb do exon 1 (SOLIER *et al.*, 2001).

TAN *et al.* (2005) analisaram a região promotora do gene *HLA-G* em populações Afro-americana, Euro-americana e Chinesa. Em todas as três populações analisadas foram observados altos níveis de diversidade nucleotídica, chegando a ser até 8 vezes mais alta que a média do genoma humano. Na análise da frequência de polimorfismos, empregando-se teste D de Tajima, foi encontrado um valor de D positivo nas três populações, sugerindo que a seleção balanceadora esteja influenciando o padrão de variação na região promotora

de *HLA-G*.

Comparações de *HLA-G* com *HLA-A* e o pseudogene *HLA-J* mostraram que o padrão de variação da região promotora de *HLA-G* não é devido à seleção em outros genes *HLA* (TAN *et al.*, 2005). Uma comparação com *Patr-G* de chimpanzés (gene homólogo ao *HLA-G* humano) evidenciou que nenhum dos 27 sítios polimórficos encontrados em humanos são polimórficos em chimpanzés, sugerindo que esses polimorfismos são específicos em humanos e se originaram depois da diferenciação entre humanos e chimpanzés.

O estudo de TAN *et al.* (2005) também sugeriu que a seleção manteve duas linhagens alélicas no locus *HLA-G* e que elas estariam relacionadas com atividades promotoras diferentes para manter balanço entre haplótipos de alta e baixa expressão. Essas atividades promotoras diferentes podem resultar em padrões de expressão espaço-temporais variáveis dependendo das diferentes necessidades imunológicas encontradas. Por exemplo, durante a gravidez, haplótipos conferindo alta expressão podem ser favorecidos na ausência de infecção, enquanto haplótipos de baixa expressão podem ser favorecidos na presença de infecção.

A região 3'UTR do gene *HLA-G* é caracterizada pela presença de um polimorfismo de inserção/deleção de 14 pb no exon 8 (HARRISON *et al.*, 1993). Tal polimorfismo pode ter conseqüências funcionais, uma vez que já foi verificada associações com aborto recorrente e pré-eclampsia (HVIID *et al.*, 2002; HVIID *et al.*, 2004; HYLENIUS *et al.*, 2004). As conseqüências funcionais desse polimorfismo podem estar associadas com o fato de que a inserção de 14pb é conhecida por gerar um sítio de splicing adicional, na qual 92 pares de bases são removidos do início do exon 8. ROUSSEAU *et al.* (2003), verificaram que mRNAs de *HLA-G* que apresentam essa deleção de 92pb são mais estáveis que as outras formas de mRNA, sugerindo que esta região pode estar envolvida em mecanismos controlando a regulação pós-transcricional. Entretanto, a presença da inserção foi associada a baixos níveis de RNAm (HVIID *et al.*, 2003)

MENDES-JUNIOR *et al.* (2007), sugerem a possibilidade de haver seleção natural

contra alelos com a inserção, uma vez que o mesmo encontra-se relacionado com complicações na gravidez. Nesse estudo, foram analisadas populações indígenas amazônicas com o objetivo de verificar o efeito de seleção em populações isoladas. Como resultado, esses autores não encontraram desvio de neutralidade nas populações em direção a seleção contra alelos com a inserção, mas acharam uma tendência de seleção balanceadora e sugerem, portanto, que a influência de seleção natural não pode ser descartada até que novas populações sejam estudadas.

A região *3'UTR* apresenta, além do polimorfismo de inserção e deleção de 14pb mencionado acima, vários sítios polimórficos que podem influenciar a transcrição e/ou tradução de HLA-G. Entre esses estão o SNP na posição +3142, o qual está localizado em uma região alvo de micro-RNAs que degradam o RNAm de HLA-G e o SNP na posição +3187, que está relacionado com estabilidade do RNAm (revisado de DONADI *et al.*, 2011).

Dados de análise de seqüenciamento nessa região em populações européias revelam uma alta diversidade nucleotídica (ALVAREZ *et al.*, 2009), mas, até o momento evidências de ação de seleção nessa região não foram documentadas.

Diante dessas informações, as regiões promotora e *3'UTR* do gene *HLA-G* mostram-se boas candidatas para estudos de seleção natural em populações. Dessa forma, o presente trabalho propôs o estudo das regiões reguladoras do gene *HLA-G* em populações Ameríndias e em uma população não-Ameríndia.

2. JUSTIFICATIVA

Estudos populacionais geram informações sobre a variação intra- e inter-populacional. Através da análise baseada na teoria da genética de populações é possível fazer inferências sobre os processos demográficos (a história das mudanças de tamanhos e estrutura de populações, assim como seus deslocamentos), genéticos (os processos de recombinação e mutação) e seletivos (a sobrevivência diferencial de indivíduos em decorrência da presença de variantes genéticas específicas) que moldaram os níveis de variação observados.

A maior parte dos estudos dos genes *KIR2DL4*, *KIR3DL1/S1* e *HLA-G* e dos estudos de seleção natural em geral, tratam a variação genética humana numa escala global comparando amostras das principais regiões do mundo. Há comparativamente poucos estudos feitos em escalas geográficas e históricas menores, com amostragens mais densas de várias populações de uma mesma região. Dessa forma, populações Ameríndias oferecem uma ótima oportunidade para estudos de seleção natural, pois as populações sul-americanas são as mais diferenciadas de qualquer região do mundo e estão distribuídas por um extenso território geográfico, criando potencial para heterogeneidade ambiental e, portanto, seletiva.

Além disso, a análise da variação de genes específicos em populações Ameríndias e a comparação com marcadores neutros podem auxiliar na compreensão de como a demografia atua na variação intra- e inter-populacional, permitindo examinar a importância relativa desse processo na diferenciação entre as populações.

3. OBJETIVOS

3.1. OBJETIVO GERAL

Este trabalho tem como objetivo geral o estudo dos polimorfismos alélicos e haplotípicos dos genes *KIR2DL4*, *KIR3DL1/S1* e *HLA-G*, em populações indígenas das Américas, para uma análise micro-evolutiva a fim de contribuir para o entendimento de como os fatores evolutivos, tais como a seleção natural e deriva, e os fatores demográficos moldam a variação genética nessas populações.

3.2. OBJETIVOS ESPECÍFICOS

- Caracterizar o polimorfismo dos genes *KIR2DL4* e *KIR3DL1/S1* e das regiões não-codificadoras do gene *HLA-G* em populações Ameríndias, utilizando a técnica de sequenciamento;
- Estimar as frequências alélicas e genotípicas e verificar a conformidade das proporções genotípicas com respeito às expectativas em equilíbrio de Hardy-Weinberg;
- Descrever os alelos e haplótipos mais freqüentes e compará-los entre as populações estudadas e com outros grupos populacionais;
- Testar a hipótese nula de que esse marcador evolui de modo consistente com um modelo de neutralidade e equilíbrio populacional;
- Analisar o polimorfismo de regiões presumidamente neutras para verificar os efeitos de processos demográficos.

4. MATERIAIS E MÉTODOS

4.1. AMOSTRA

4.1.1. Amostras indígenas

O presente trabalho analisou amostras de populações indígenas Americanas distribuídas nas regiões Norte, Centro e Sul do continente Americano, sendo a maioria proveniente de populações indígenas da Amazônia (figura 1).

As amostras de DNA que utilizamos neste trabalho foram obtidas através de doação de bancos de amostras pertencentes ao Laboratório de Genética Humana e Médica da Universidade Federal do Pará (LGHM-UFGPA), do *Galton Laboratory da University College London* (GL-UCL) e do *Laboratory of Experimental Immunology do National Cancer Institute* (NCI-Frederick).

O número de amostras e populações difere entre os genes estudados devido às questões específicas associadas a cada um dos estudos e serão detalhados posteriormente. O número pequeno de amostras para algumas das populações estudadas não compromete o projeto, pois o objetivo é obter uma amostra da diversidade da região como um todo, fazendo um estudo do impacto da seleção dentro do continente sul-americano e não apenas nas populações individuais.

4.1.2. Amostras de células de trofoblasto

O presente trabalho também analisou amostras de DNA de 49 células de trofoblasto extraviloso provenientes da população inglesa. Essas amostras foram obtidas a partir de tecidos placentários e deciduais de gravidez normais entre 6-12 semanas de gestação. As amostras de DNA tiveram a expressão da proteína HLA-G na superfície medidas através de

citometria de fluxo (APPS, manuscrito em preparação). As amostras de DNA que utilizamos neste trabalho são pertencentes ao *Laboratory of Experimental Immunology* do *National Cancer Institute* (NCI-Frederick).

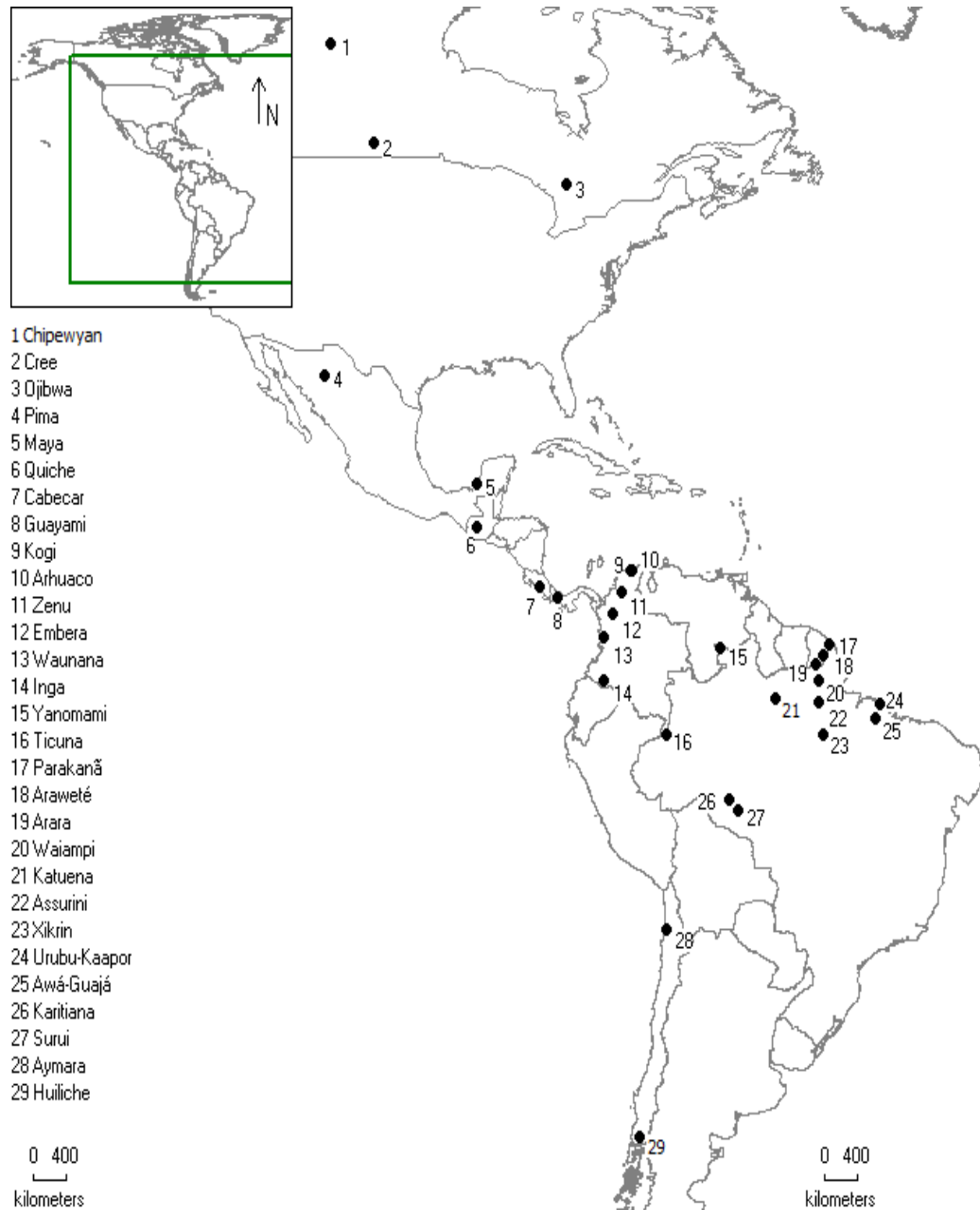


Figura 1. Mapa das Américas mostrando a localização geográfica das populações estudadas neste trabalho.

4.2. EXTRAÇÃO DE DNA

Os métodos para a extração de DNA dessas amostras foram: fenol-clorofórmio de acordo com o descrito por SAMBROOK *et al.* (1989) e *salting out* com base na técnica de LAHIRI & NURNBERG (1991), modificada.

4.3. TÉCNICA DE AMPLIFICAÇÃO DE GENOMA INTEIRO

Foi utilizada a técnica de amplificação de genoma inteiro através da metodologia de *multiple displacement amplification* (MDA), implementada pelo kit *GenomiPhi* (GE Healthcare) para obtenção de maior quantidade de DNA para cada amostra. Na amplificação de genoma inteiro por MDA, a DNA polimerase do bacteriófago Phi29 amplifica exponencialmente DNA molde, utilizando oligonucleotídeos aleatórios de seis pares de base como iniciadores.

4.4. METODOLOGIA DE ANÁLISE DOS GENES CANDIDATOS

4.4.1. *KIR2DL4*

A caracterização de presença e ausência desse locus foi realizada pelo Professor Eduardo José Melo dos Santos no Laboratório de Imunologia Experimental no *National Cancer Institute* em Frederick através de PCR-SSP (primer sequência específico), como descrito por KULKARNI *et al.* (2010).

Para a caracterização dos alelos do gene *KIR2DL4* os exons 3, 5, 7 e 9 foram amplificados e seqüenciados. A amplificação dessas regiões constituem em duas reações de PCR, uma para a amplificação dos exons 3 e 5 e outra para os exons 7 a 9. Dessa

forma, para a genotipagem desse locus, foram realizadas duas reações de PCR para cada indivíduo.

Os iniciadores utilizados encontram-se na tabela 1. As condições de amplificação desse gene encontram-se na tabela 2.

Tabela 1. Oligonucleotídeos iniciadores utilizados para a amplificação dos exons do gene *KIR2DL4*.

Oligonucleotídeos iniciadores		Região amplificada
<i>2DL4/1 – Forward</i>	5' GGT CAG GAC AAG CCC TTC T 3'	Exons 3-5
<i>2DL4/4 – Reverse</i>	5' CCT CAC CTG TGA CAG AAA CAG 3'	
<i>2DL4/5 – Forward</i>	5' GCC AGA CAC CTG CAT GCT G 3'	Exons 7-9
<i>2DL4/8 – Reverse</i>	5' CTG GCC CTT CTC AGA GGA G 3'	

Tabela 2. Protocolo para a amplificação do gene *KIR2DL4*.

Reagentes	Concentrações
Tampão	1X
Dntp	0,2mM
MgCl ₂	1,5mM
Iniciador_Forward	0,4 mM
Iniciador_Reverse	0,4 Mm
<i>Taq Platinum</i>	0,25 U/ml
<i>DNA</i>	50 ng

As condições de ciclagem foram as mesmas para todos os exons: 95°C por 2 minutos, 35 ciclos de 95 °C por 15 segundos, 60 °C por 15 segundos, 72 °C por 45 segundos e um ciclo final de extensão de 72 °C por 7 minutos.

Após a amplificação, foi feita a verificação do tamanho e da concentração dos produtos da PCR através de corrida eletroforética em gel de agarose 2,0 % juntamente com o marcador *Low DNA Mass Ladder (Invitrogen)*.

Os produtos de PCR foram submetidos à purificação com método automatizado utilizando *beads* magnéticos (*AMPURE XP, Beckman Coulter Genomics*).

Os produtos purificados foram submetidos a quatro reações de seqüenciamento, cada uma referente a um exon específico. Para as reações de seqüenciamento foram utilizados iniciadores *2DL4/1, 2DL4/4, 2DL4/5 e 2DL4/7* (5' CTG GCC CTT CTC AGA GGA G 3'). O protocolo de seqüenciamento é mostrado na tabela 3. O protocolo de ciclagem utilizado para o seqüenciamento das amostras amplificadas consiste uma etapa inicial de 94°C por 1 minuto e 25 ciclos de 94°C por 5 segundos, 50 °C por 5 segundos e 60 °C por 4 minutos.

Tabela 3. Protocolo da reação de seqüenciamento para o gene *KIR2DL4*.

Reagentes	Volumes
Produto de PCR purificado	3.0 µl
Iniciador	0.2 µl
<i>Dye Terminator v.3.1 (Applied Biosystems)</i>	0.25 µl
Tampão 5X	1.875 µl
Água	4.75 MI

As reações de seqüenciamento foram submetidas à corrida eletroforética em seqüenciador no Laboratório de Imunologia Experimental no *National Cancer Institute*.

Para análise da qualidade das sequências e identificação dos polimorfismos no gene *KIR2DL4* o programa Mutation Surveyor® (Softgenetics) foi usado.

A identificação dos alelos foi feita com base na combinação de sítios polimórficos nesses quatro éxons, de acordo com o *Immuno Polymorphism Database* (IPD, <http://www.ebi.ac.uk/ipd/kir>).

4.4.2. *KIR3DL1/S1*

A caracterização de presença e ausência desse locus também foi realizada pelo Professor Eduardo José Melo dos Santos através da metodologia de PCR-SSP (primer sequência específico), como descrito por KULKARNI *et al.* (2010).

Para a genotipagem do alótipo *KIR3DL1* os exons 3, 4, 5 e 9 foram amplificados e seqüenciados. Cada um dos exons exigiu uma amplificação separada, de modo que para a genotipagem de *KIR3DL1* foram necessárias quatro reações de PCR diferentes para cada indivíduo analisado. Os iniciadores utilizados encontram-se na tabela 4. As condições de amplificação desse gene encontram-se na tabela 5.

Tabela 4. Oligonucleotídeos iniciadores utilizados para a genotipagem do alótipo *KIR3DL1*.

Oligonucleotídeos iniciadores		Região amplificada
<i>3DL1/2.5 – Forward</i>	5' TTC TTG GTC CAG AGG GCC GGT 3'	Exon 3
<i>3DL1/3.3- Reverse</i>	5' CTG TGA CCA TGA TCA CCA C 3'	
<i>3DL1/4.5 – Forward</i>	5' GAA ACC ACA GAA AAC CTT CCC 3'	Exon 4
<i>3DL1/5.3A – Reverse</i>	5' AGA GAG AAG GTT TCT CAT ATG 3'	
<i>3DL1/5.5 – Forward</i>	5' GCC TCT TCT CCT TCC AGG TCC 3'	Exon 5
<i>3DL1/5.3B – Reverse</i>	5' CAC CTG TGA CAG AAA CAA G 3'	
<i>3DL1/7.5 – Forward</i>	5' AGT GGT CAT CAT CCT CTT CAT C 3'	Exon 9
<i>3DL1/9.3 – Reverse</i>	5' GTG TAC AAG ATG GTA TCT GTA 3'	

Tabela 5. Protocolo para a amplificação do alótipo *KIR3DL1*.

Reagentes	Concentrações
Tampão	1X
dNTP	0,2mM
MgCl ₂	1,5mM
Iniciador_Forward	0,4 mM
Iniciador_Reverse	0,4 mM
<i>Taq Platinum</i>	0,5 U/ml
<i>DNA</i>	50 ng

Para a amplificação dos exon 3, as condições de ciclagem foram: 95°C por 2 minutos, 5 ciclos de 95 °C por 15 segundos, 70 °C por 15 segundos, 72 °C por 90 segundos, 26 ciclos de 95 °C por 15 segundos, 65 °C por 15 segundos, 72 °C por 90 segundos, 4 ciclos de 95 °C por 15 segundos, 55 °C por 1 minuto, 72 °C por 2 minutos e um ciclo final de extensão de 72 °C por 7 minutos.

Para a amplificação dos exons 4, 5 e 9, as condições de ciclagem foram: 95°C por 2 minutos, 5 ciclos de 95 °C por 15 segundos, 65 °C por 15 segundos, 72 °C por 90 segundos, 26 ciclos de 95 °C por 15 segundos, 60 °C por 15 segundos, 72 °C por 90 segundos, 4 ciclos de 95 °C por 15 segundos, 55 °C por 1 minuto, 72 °C por 2 minutos e um ciclo final de extensão de 72 °C por 7 minutos.

Os produtos foram submetidos à purificação com método automatizado utilizando *beads* magnéticos (*AMPURE XP, Beckman Coulter Genomics*).

Os produtos purificados foram submetidos a reações de seqüenciamento cujo protocolo é mostrado na tabela 6. Para as reações de seqüenciamento foram utilizados os iniciadores *3DL1/3.3*, *3DL1/4.5*, *3DL1/5.3B* e o iniciador *3DL1/1172* (5' CTG CAG GGA ACA

GAA CAG CC 3'). O protocolo de ciclagem utilizado para o seqüenciamento das amostras amplificadas consiste uma etapa inicial de 94°C por 1 minuto e 25 ciclos de 94°C por 5 segundos, 50 °C por 5 segundos e 60 °C por 4 minutos.

Tabela 6. Protocolo da reação de seqüenciamento para alótipo *KIR3DL1*.

Reagentes	Volumes
Produto de PCR purificado	3.0 µl
Iniciador	0.2 µl
<i>Dye Terminator v.3.1</i> (Applied Biosystems)	0.25 µl
Tampão 5X	1.875 µl
Água	5.0 µl

As reações de seqüenciamento foram submetidas à corrida eletroforética em seqüenciador no Laboratório de Imunologia Experimental no *National Cancer Institute*.

O programa Mutation Surveyor® (Softgenetics) foi usado para análise da qualidade das sequências e identificação dos polimorfismos no alótipo *KIR3DL1/S1*.

A identificação dos alelos foi feita com base na combinação de sítios polimórficos nesses quatro éxons, de acordo com o *Immuno Polymorphism Database* (IPD, <http://www.ebi.ac.uk/ipd/kir>).

4.4.3. HLA-G

4.4.3.1. Região promotora

A região promotora do gene *HLA-G* foi analisada em 61 amostras Ameríndias e em 49 amostras de DNA de células. Para a amplificação do segmento que contém a região promotora do gene *HLA-G* foi utilizada a reação de PCR (*Polimerase Chain Reaction*). Os primers utilizados, descritos por OBER *et al.* (2003), e o respectivo tamanho do fragmento amplificado encontram-se na tabela 7.

Tabela 7. Oligonucleotídeos iniciadores utilizados para a amplificação da região promotora do gene *HLA-G*.

Primers	Tamanho do fragmento
<i>Forward</i> 5' ACA TTC TAG AAG CTT CAC AAG AAT G 3'	
<i>Reverse</i> 5' TGG GCC TTG GTG TTC CGT G3'	1742 pb

Após a amplificação, foi feita a verificação do tamanho e da concentração dos produtos da PCR através de corrida eletroforética em gel de agarose 2,0 % juntamente com o marcador *Low DNA Mass Ladder (Invitrogen)*.

Os produtos da PCR foram purificados, utilizando-se as enzimas exonuclease I (*GE Healthcare*) e fosfatase alcalina de camarão (*GE Healthcare*). As concentrações dos reagentes utilizados no protocolo de purificação foram as seguintes: 1µ de cada solução de enzima para cada 10 µl de produto de PCR.

Após a purificação, os produtos de PCR foram seqüenciados em ambas as direções utilizando-se o kit *Big Dye Terminator v.3 (Applied Biosystems)*. A reação foi feita utilizando-se os primers da reação de PCR descritos acima, assim como outros 6 primers internos (G-908: 5' TTC ACC TCA CAG TTG TAA GTG TTC 3'; G-830F: 5' CAC ACG GAA ACT TAG GGC TAC G 3'; G-1123F: 5' GCC TCG CTG GGT GTT CTT TGC 3'; G-304R: 5' GCC AAG CGT TCT GTC TCA GTG T 3'; GPR-247: 5' CTC AAG CGT GGC TCT CAG GGT C 3'; GIN1: 5' GTT TCC CTC CTG ACC CCG CAC T 3').

Após a reação de seqüenciamento, as amostras foram precipitadas e os produtos desta reação foram submetidos à corrida eletroforética no seqüenciador ABI PRISM™ 3100 (*Applied Biosystems*) do laboratório Biologia Molecular de Plantas do Departamento de Botânica do Instituto de Biociências da USP-SP e no seqüenciador ABI PRISM™ 3100 (*Applied Biosystems*) do Departamento de Bioquímica-IQ USP-SP.

Para análise da qualidade, montagem e edição das leituras (*reads*) geradas através do seqüenciamento automático, foi usado o programa Phred/Phrap/Consed. O programa Phred realiza a identificação e avalia a qualidade da leitura dos nucleotídeos presentes em seqüência nos cromatogramas gerados pelo seqüenciador automático. O Phrap utiliza os arquivos gerados pelo Phred e realiza a montagem das seqüências de modo a formar uma única seqüência. O Consed é o programa de visualização e edição dos contigs montados pelo Phrap (EWING & GREEN, 1998; EWING *et al.*, 1998; GORDON *et al.*, 1998).

4.4.3.2. O polimorfismo de 14pb

Os iniciadores utilizados, descritos por HVIID *et al.* (2002), e o respectivo tamanho do fragmento amplificado encontram-se na tabela 8. As condições de amplificação desse segmento encontram-se na tabela 9.

Tabela 8. Oligonucleotídeos iniciadores utilizados para a amplificação do polimorfismo de 14pb do gene *HLA-G*.

Iniciadores		Tamanho do fragmento
Forward	5'-GTGATGGGCTGTTTAAAGTGTCACC-3'	210-224 pb
Reverse	5'-GGAAGGAATGCAGTTCAGCATGA-3'	

Tabela 9. Protocolo para amplificação do polimorfismo de 14pb.

Reagentes	Concentrações
Tampão	1X
Dntp	0,2mM
MgCl ₂	1,5mM
HLA-G_14For	0,4 mM
HLA-G_14Rev	0,4 mM
<i>Taq Platinum</i>	0,75 U/ml
<i>DNA</i>	50 ng

As siglas HLA-G_14For e HLA-G_14Rev referem-se aos iniciadores 5' e iniciador 3', respectivamente. O volume final de reação é de 25µl.

As condições de ciclagem foram: 94°C por 2 minutos, 30 ciclos de 94°C por 30 segundos, 64°C por um minuto, 72°C por dois minutos e um ciclo final de extensão de 72°C por 10 minutos.

Os produtos da reação de PCR foram analisados através da separação eletroforética em gel de poliacrilamida 8% não-desnaturante, preparado através da mistura dos seguintes componentes:

- 7,5 ml de acrilamida + bis-acrilamida (29 g de acrilamida recristalizada + 1 g de bisacrilamida recristalizada, diluídas em 100 ml de água deionizada);
- 1.25 ml de glicerol;
- 16,25 ml de tampão TBE 0.5x (TBE 10x: pH 8,0 – 108 g de Tris + 53 g de Ácido Bórico + 7,44 g de EDTA, diluídos em 1 litro de água deionizada);
- 25 µl de TEMED (N, N, N', N', tetrametiletilenodiamina);
- 250 µl de persulfato de potássio [100 mg/ml].

Após a polimerização do gel, este é colocado em uma cuba de eletroforese vertical, contendo tampão TBE 0.5x nas porções superior (anódica) e inferior (catódica). Em cada poço do gel, aplica-se 5 µl de produto da PCR, juntamente com 1 µl de tampão de carregamento (*loading buffer* – TBE 1x; Xilenocianol 0,25%; Azul de Bromofenol 0,25%; EDTA 0,1M pH 8,0; Ficoll 15%) e com o corante gel-red para visualização das bandas em luz ultra-violeta. A cuba de eletroforese foi conectada a uma fonte de eletroforese de alta voltagem e a separação eletroforética dos fragmentos foi realizada durante 4 horas a 200 volts, com amperagem liberada.

4.4.3.3. Região 3'UTR

As 49 amostras de DNA de células foram amplificadas e seqüenciadas para a região 3'UTR. Todos os iniciadores utilizados foram desenhados para o presente projeto baseados no conhecimento da seqüência da região. Para a amplificação de toda a região 3'UTR do gene *HLA-G* foram utilizados os iniciadores: CG3_F2 5' TCT CGA GGG TCC AAG ACT AGG AGG 3' (forward) e CG3_R2 5' ACT GGG CGC AGC CCC ATC TA 3' (reverse). O protocolo para amplificação encontra-se na tabela 10.

Tabela 10. Protocolo para primeira amplificação da região 3'UTR do gene *HLA-G*.

Reagentes	Concentrações
Tampão	1X
Dntp	0,2Mm
MgCl ₂	1,5Mm
Iniciador_Forward	0,5 Mm
Iniciador_Reverse	0,5 mM
<i>Taq Platinum</i>	0,25 U/ml
<i>DNA</i>	50 ng

As condições de ciclagem foram: 95°C por 2 minutos, 35 ciclos de 95 °C por 15 segundos, 65 °C por 15 segundos, 72 °C por 45 segundos e um ciclo final de extensão de 72 °C por 7 minutos.

Os produtos de PCR foram submetidos à purificação com método automatizado utilizando *beads* magnéticos (*AMPURE XP, Beckman Coulter Genomics*). Para as reações de seqüenciamento foram utilizados os dois iniciadores descritos acima e mais dois iniciadores internos: 5' GTG ATG GGC TGT TTA AAG TGT CAC C 3' (forward) e 5' GGA AGG AAT GCA GTT CAG CAT GA 3'. Para os indivíduos heterozigotos para o polimorfismo de inserção/deleção de *14pb*, foram utilizados iniciadores adicionais para confirmação dos polimorfismos adjacentes ao mesmo, foram eles: *CG3_F4* 5'AGT GGC AAG TCC CTT TGT GAC 3' (forward), *CG3_IF3* 5' TGT GAC TTC AAG AAC CCT GAC 3' (forward), *CG3_IF5* 5'GTC TTC CAT TTA TTT TGT CTC T (reverse) e *CG3_R4* 5' CAA GAT TTG TTC ATG CCT TCC C 3' (reverse). O protocolo de sequenciamento encontra-se na tabela 11.

Tabela 11. Protocolo da reação de seqüenciamento para a região *3'UTR* do gene *HLA-G*.

Reagentes	Volumes
Produto de PCR purificado	2.5 µl
Iniciador	0.2 µl
<i>Dye Terminator v.3.1</i> (Applied Biosystems)	0.25 µl
Tampão 5X	1.875 µl
Água	4.675 µl

4.4.4. Metodologia de análise das regiões presumidamente neutras

A análise de marcadores neutros como controles genômicos para efeitos demográficos foi feita através da análise de polimorfismos em microssatélites pela aluna de doutorado Kelly Nunes. Foram analisados 23 microssatélites espalhados pelo genoma (*D5S816*, *D1S551*, *D7S3061*, *D4S3248*, *D16S539*, *D9S922*, *D10S1426*, *D7S821*, *D13S317*, *D8S2324*, *D19S559*, *D6S1040*, *D20S482*, *D9S301*, *D18S535*, *D12S395*, *D17S1299*, *D1S1660*, *D3S3039*, *D10S2470*, *D4S2366*, *D3S3038*, *New*). A genotipagem dos microssatélites foi baseada na amplificação por reação de PCR (*Polimerase Chain Reaction*). Os microssatélites foram agrupados em reações multiplex e submetidos à corrida eletroforética no seqüenciador automático MegaBACE 1000 (*GE Healthcare*).

4.5. ANÁLISE DOS RESULTADOS

4.5.1. Frequências de portadores, alélicas e genotípicas

As frequências de portadores, alélicas e genotípicas foram obtidas por contagem direta.

4.5.2. Estimativa da frequência de alelos nulos para os genes *KIR*

As estimativas de frequências de alelos nulos foram feitas através de três métodos 1) Considerando apenas a frequência dos homozigotos para a ausência do gene; 2) Inferindo o genótipo dos indivíduos heterozigotos (para a presença do alelo nulo) através da análise dos haplótipos contendo a combinação de presença e ausência dos outros genes do complexo *KIR*. Os 14 genes do complexo foram analisados para presença e ausência e o haplótipo dos indivíduos homozigotos para o alelo nulo foi determinado. Todos os indivíduos que não se enquadravam nesse perfil foram considerados verdadeiros homozigotos. 3)

Estimativa das frequências alélicas por máxima verossimilhança, usando o algoritmo de expectativa e maximização (EM) de DEMPSTER *et al.* (1977), implementado no programa GENEPOP 3.4 (RAYMOND & ROUSSET, 1995)

4.5.3. Frequências haplotípicas

Para a determinação dos haplótipos e das frequências haplotípicas foi utilizado uma abordagem Bayesiana implementada no programa PHASE (STEPHENS, *et al.*, 2001) para inferência de fase dos dados gerados.

4.5.4. Equilíbrio de Hardy-Weinberg

As frequências observadas dos genótipos foram comparadas com aquelas esperadas de acordo com o equilíbrio de Hardy-Weinberg, através do método de GUO & THOMPSON (1992), realizado através do uso do pacote de programas Arlequin 3.5 (EXCOFFIER & LISHER, 2010).

4.5.5. Desequilíbrio de ligação

A determinação do grau de desequilíbrio de ligação entre os polimorfismos das regiões promotora e 3'UTR do gene *HLA-G* foi realizado utilizando-se o programa HAPLOVIEW (BARRETT *et al.*, 2005)

4.5.6. Testes de neutralidade

Foi utilizado o teste de Ewens-Watterson para testar desvios da neutralidade seletiva para os dados alélicos (WATTERSON, 1978; EWENS, 1972). Nesse teste, a taxa de homozigose observada (F_{obs}) é comparada com a esperada (F_{esp}). O esperado é derivado de amostras simuladas com um mesmo tamanho e número de alelos da amostra observada (a simulação é baseada na teoria de amostragem de alelos sob neutralismo de EWENS (1972). Amostras com $F_{obs} > F_{esp}$, apresentam poucos alelos em alta frequência e vários em baixa, situação compatível com seleção direcional ou expansão populacional. Amostras com $F_{obs} < F_{esp}$, apresentam alelos em frequências intermediárias, situação compatível com seleção balanceadora, gargalo populacional recente ou estruturação populacional (Watterson, 1978). A significância desse teste é dada pela proporção de F_{esp} que são inferiores ao observado. Esse teste foi realizado utilizando a implementação do pacote de programas Arlequin v.3.5.1.2 (EXCOFFIER & LISHER, 2010).

Análises ao nível molecular foram feitas através do teste D de Tajima (TAJIMA, 1989). O teste *D de Tajima* compara dois estimadores do parâmetro θ , também conhecido como $4N\mu$ para os *loci* nucleares, onde N representa o tamanho da população e μ é a taxa de mutação. Um dos estimadores é o π ou diversidade nucleotídica, o número médio de diferenças par a par entre as seqüências da amostra ($\theta\pi$). O outro estimador é o número de sítios segregantes na amostra (θS). Tajima (1989) demonstrou que sob evolução neutra não há diferenças entre esses dois estimadores de θ . Valores positivos de D refletem a presença de polimorfismos em frequências intermediárias na amostra, fenômeno causado pela seleção balanceadora ou gargalo populacional recente. Valores negativos são um reflexo da presença de vários polimorfismos de baixa frequência nas seqüências, fenômeno causado pela ação da seleção direcional, purificadora ou expansão populacional. A significância do teste *D de Tajima* é dada pela proporção de valores de *D* de amostras obtidas por simulação coalescente (com o mesmo número de sítios segregantes que a amostra observada) menor que o *D* observado. O teste D de Tajima foi realizado utilizando o pacote de programas Arlequin v.3.5.1.2 (EXCOFFIER & LISHER, 2010).

4.5.7. Diferenciação inter-populacional

A estatística *Fst* foi utilizada como medida de similaridade genética entre cada par de populações e realizada através do pacote de programas Arlequin v.3.5.1.2 (EXCOFFIER & LISHER, 2010).

4.5.8. Inferências demográficas a partir da análise de marcadores neutros

Os dados gerados a partir da genotipagem de 23 marcadores microssatélites autossômicos foram utilizados para fazer inferências sobre a história demográfica das populações através de correlações entre a diversidade genética dos microssatélites (através da medida theta H, que é um estimador de $4N\mu$ obtido a partir da taxa de heterozigose dos microssatélites) e a diversidade haplípica dos genes *KIR*. A possibilidade de terem ocorrido recentes gargalos ou expansões foi testada utilizando métodos baseados na diversidade alélica através do índice de GARZA & WILLIAMSON (2001) e realizada através do pacote de programas Arlequin v.3.5.1.2 (EXCOFFIER & LISHER, 2010).

5. RESULTADOS

5.1. GENE *KIR2DL4*

5.1.1. Caracterização da variabilidade alélica

Para o gene *KIR2DL4* foram analisados 219 indivíduos pertencentes a 10 tribos localizadas na região Amazônica, sendo 39 indivíduos pertencentes à tribo Arara, 13 à tribo Assurini, 18 à tribo Awá-Guajá, 25 à tribo Araweté, 56 à tribo Xikrin, 12 à tribo Urubu-Kaapor, 8 à tribo Katuena, 23 à tribo Parakanã, 3 à tribo Waiampi e 1 indivíduo pertencente à tribo Yanomami.

Apesar de *KIR2DL4* ser considerado um gene moldura (*framework*), isto é, presente em todos os haplótipos *KIR*, alguns estudos relatam a existência de haplótipos nos quais esse gene está ausente. Dessa forma, a primeira parte das análises com o gene *KIR2DL4* consistiu na caracterização de presença e ausência desse locus. Essa parte do trabalho foi realizada pelo Professor Eduardo José Melo dos Santos no Laboratório de Imunologia Experimental no *National Cancer Institute* em Frederick, Estados Unidos. Dos 219 indivíduos analisados, 2.3% apresentaram o gene *KIR2DL4* ausente em ambos os cromossomos.

Para os indivíduos contendo pelo menos uma cópia do gene *KIR2DL4*, a genotipagem foi feita através do seqüenciamento dos éxons 3, 5, 7 e 9. Os alelos foram identificados com base na combinação de sítios polimórficos nesses quatro éxons, de acordo com o *Immuno Polymorphism Database* (IPD, <http://www.ebi.ac.uk/ipd/kir>). Apesar de terem sido seqüenciados os quatro éxons mais polimórficos e que definem o maior número de alelos para o gene *KIR2DL4*, um caso de ambigüidade foi encontrado: os alelos *KIR2DL4*00102* e *KIR2DL4*00105* apresentam uma única diferença no éxon 8 desse gene

(que não foi seqüenciado no presente trabalho). Dessa maneira, para nomear esse alelo, foi seguida a nomenclatura na qual o alelo é definido pelos dígitos separados por barra (*KIR2DL4*00102/05*).

O seqüenciamento dos quatro éxons do gene *KIR2DL4* nas populações Ameríndias estudadas revelou a presença de 7 alelos, dos quais 6 já foram previamente descritos em outras populações: *KIR2DL4*00501*, *KIR2DL4*00102/05*, *KIR2DL4*011*, *KIR2DL4*00801*, *KIR2DL4*00802*, *KIR2DL4*006*. O alelo novo foi denominado de *KIR2DL4*New*. Esse alelo é idêntico ao *KIR2DL4*00801*, exceto pela mutação A → G na posição 568 no éxon 5.

Os sítios polimórficos que discriminam os alelos encontrados são mostrados na tabela 12. As frequências de portadores dos alelos foram calculadas para o grupo Ameríndio e são apresentadas na tabela 13.

Tabela 12. Sítios polimórficos dos alelos do gene *KIR2DL4* encontrados no presente trabalho.

	Exon 3	Exon 5						Exon 7		Exon 9
Proteína	Domínio DO	Domínio D2						Domínio transmembrana		Região citoplasmática
Aminoácido¹	Y30C	G109P	T115A		G167R		A186P			
Nucleotídeo	158	395	412	480	568	615	625	762	811	1023
<i>2DL4*00101</i>	A	C	A	A	G	A	G	T	A	G
<i>2DL4*00102/05</i>	-	-	-	G	-	G	-	C	-	A
<i>2DL4*00501</i>	G	-	G	G	-	G	C	C	-	-
<i>2DL4*006</i>	-	C	G	G	-	G	C	C	-	-
<i>2DL4*00801</i>	-	-	G	G	-	G	C	-	*	G
<i>2DL4*00802</i>	-	-	G	-	-	G	C	-	*	G
<i>2DL4*011</i>	G	-	G	G	-	G	C	-	*	G
<i>2DL4*New</i>	-	-	G	G	A	G	C	-	*	G

O alelo *KIR2DL4*00101* foi usado como referência

Traços indicam identidade com a referência

* representa a deleção de uma adenina no final do exon 7, introduzindo um códon de parada prematuro na proteína

¹ Posições que apresentam mudanças de aminoácidos não-sinônimas

Tabela 13. Frequências de portadores de alelos do gene *KIR2DL4* no conjunto de populações Ameríndias.

Alelos	%
<i>2DL4*00501</i>	79.0
<i>2DL4*00102/05</i>	51.1
<i>2DL4*011</i>	5.9
<i>2DL4*New</i>	3.2
<i>2DL4*00802</i>	0.46
<i>2DL4*00801</i>	0.9
<i>2DL4*006</i>	0.46

5.1.2. Determinação das frequências alélicas

A metodologia de genotipagem dos genes *KIR* baseia-se na detecção de ausência e presença de genes específicos. Como mencionado na introdução, os haplótipos *KIR* são compostos por diferentes números de genes, sendo alguns ausentes e outros presentes. Dessa forma, a presença ou ausência do gene é tratada como estado alélico. Um desafio na análise das frequências alélicas dos genes desse complexo, quando baseada somente em genotipagem de presença e ausência e sequenciamento, é a distinção entre os indivíduos que são realmente homocigotos daqueles que são heterocigotos para a ausência do gene. A dificuldade resulta do fato desses dois genótipos serem, no referido desenho experimental, indistinguíveis.

Como mencionado no item anterior, o *KIR2DL4* é considerado um gene moldura (*framework*). Entretanto, a ausência do gene, embora rara, foi documentada em 8 indivíduos de um painel de 12.741 indivíduos analisados (www.allelefreqencies.net). No presente trabalho, a ausência desse gene no conjunto de populações Ameríndias analisadas tem uma frequência significativamente maior, superior a 4% em 6 tribos estudadas. Dessa forma, fez-se necessário investigar a contribuição da ausência do gene em heterocigose na determinação das frequências alélicas e verificar qual é o peso da mesma nos testes a

serem realizados. Para tentar solucionar esse problema, três diferentes metodologias foram empregadas. A primeira consistiu em assumir que todos os indivíduos que apresentam apenas um alelo por sequenciamento são homozigotos para aquele alelo (denominado aqui de método 1). A segunda metodologia consistiu em inferir quais indivíduos são heterozigotos (isto é, carregam um alelo nulo) através da análise dos haplótipos contendo a combinação de presença e ausência dos outros genes do complexo *KIR*. Os 14 genes do complexo foram analisados para presença e ausência e o haplótipo dos indivíduos homozigotos para o alelo nulo foi determinado. Todos os indivíduos que não se enquadravam nesse perfil foram considerados verdadeiros homozigotos (método 2). A terceira metodologia consiste em estimar as frequências alélicas por máxima verossimilhança, usando o algoritmo de expectativa e maximização (EM) de DEMPSTER *et al.* (1977) (método 3). Nas tabelas 14, 15 e 16 são apresentadas as frequências alélicas do gene *KIR2DL4* para cada uma das metodologias citadas.

Tabela 14. Frequências alélicas do gene *KIR2DL4* para a método 1.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã	Waiampi	Yanomami
Alelos\2N	78	30	36	50	112	24	16	84	6	2
2DL4*00501	56.4	50.0	61.1	92.0	54.5	45.9	56.2	46.4	66.7	0.5
2DL4*00102/05	43.6	43.3	5.6	4.0	28.6	37.5	25.0	52.4	33.3	0.5
2DL4*011	0	6.7	22.2	0	6.3	0	6.3	0	0	0
2LD4*New	0	0	0	4.0	3.5	8.3	12.5	1.2	0	0
2DL4*00802	0	0	0	0	0.9	0	0	0	0	0
2DL4*00801	0	0	0	0	1.8	0	0	0	0	0
2DL4*006	0	0	0	0	0.9	0	0	0	0	0
Nulo	0	0	11.1	0	3.5	8.3	0	0	0	0

Tabela 15. Frequências alélicas do gene *KIR2DL4* para a método 2.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã	Waiampi	Yanomami
Alelos\2N	78	30	36	50	112	24	16	84	6	2
2DL4*00501	56.4	50.0	41.7	88.0	48.2	41.7	43.8	35.7	66.7	0.5
2DL4*00102/05	43.6	43.3	5.6	4.0	24.1	33.3	25.0	58.3	33.3	0.5
2DL4*11	0	6.7	13.9	0	5.3	0	6.2	0	0	0
2LD4*New	0	0	0	4.0	2.7	8.3	12.5	1.2	0	0
2DL4*00802	0	0	0	0	0.9	0	0	0	0	0
2DL4*00801	0	0	0	0	1.8	0	0	0	0	0
2DL4*006	0	0	0	0	0.9	0	0	0	0	0
Nulo	0	0	38.8	4.0	16.1	16.7	12.5	4.8	0	0

Tabela 16. Frequências alélicas do gene *KIR2DL4* para a método 3.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã	Waiampi	Yanomami
Alelos\2N	78	30	36	50	112	24	16	84	6	2
2DL4*00501	56.4	50.0	46.6	92.0	46.7	38.9	54.1	45.5	66.7	0.5
2DL4*00102/05	43.6	43.3	5.5	4.0	25.0	32.6	25.0	51.4	33.3	0.5
2DL4*11	0	6.7	15.5	0	5.5	0	6.3	1.2	0	0
2LD4*New	0	0	0	4.0	2.7	4.5	9.0	0	0	0
2DL4*00802	0	0	0	0	0.9	0	0	0	0	0
2DL4*00801	0	0	0	0	1.8	0	0	0	0	0
2DL4*006	0	0	0	0	0.9	0	0	0	0	0
Nulo	0	0	32.4	0	16.5	24.0	5.6	1.9	0	0

5.1.3. O teste do equilíbrio de Hardy-Weinberg para avaliação das metodologias aplicadas

As frequências genóticas obtidas pelos métodos 1 e 2 foram testadas para o equilíbrio de Hardy-Weinberg. Os resultados dos testes são mostrados nas tabelas 17 e 18.

Para o método 1 foram encontrados desvios significativos nas tribos Arara, Awá-Guajá, Xikrin e Urubu-Kaapor, sendo que na primeira há um excesso de heterozigotos e nas demais há um excesso de homozigotos. Os desvios significativos do equilíbrio tendendo a um excesso de homozigotos podem indicar problemas de origem metodológica, como a presença de alelo nulo. O resultado nessas três tribos não é surpreendente, uma vez que elas apresentam indivíduos com a ausência do gene em homozigose, sugerindo que a presença do alelo nulo pode ser relativamente alta. O resultado do teste de equilíbrio de Hardy-Weinberg indica que a metodologia 1 subestima a frequência dos alelos nulos, levando a desvios de equilíbrio.

Quando a segunda metodologia foi empregada, foram encontrados desvios significativos para as tribos Arara, Xikrin e Parakanã, sendo todas as três com um excesso de heterozigotos. O resultado do teste de equilíbrio de Hardy-Weinberg indica que, ao contrário da metodologia anterior, essa abordagem superestima a frequência de alelos nulos.

O terceiro método para estimar as frequências alélicas, a partir da presença de um alelo nulo, assume que as populações estão em equilíbrio de Hardy-Weinberg. Sendo assim, populações com desvios das condições de equilíbrio podem ter as estimativas das frequências alélicas enviesadas. As populações Ameríndias são conhecidas por serem endogâmicas, fato esse que pode causar desvios de equilíbrio, na direção de um excesso de homozigose. Portanto, assumir que as populações estão em equilíbrio poderia não ser adequado nesse cenário.

Para testar o pressuposto de que as populações estão em equilíbrio, as taxas de heterozigose observada e esperada foram calculadas para marcadores microssatélites nessas populações. Em média, considerando 23 microssatélites autossômicos, os valores observados não desviaram do equilíbrio. A frequência de desvios em direção de excesso de homozigotos é de 4.96%, que é aproximadamente o próprio valor nominal usado nos testes. Portanto, não há evidências de um desvio significativo causado por endogamia, indicando que os desvios, quando encontrados, são provavelmente decorrência de erros metodológicos mais do que características históricas das populações (Kelly Nunes, comunicação pessoal).

Diante da ausência de evidência de desvio de Hardy-Weinberg para marcadores espalhados pelo genoma, a metodologia 3 foi considerada a mais adequada.

Tabela 17. Resultados dos testes de equilíbrio de Hardy-Weinberg para o método 1.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã	Waiampi
Het. Obs.	0.7195	0.5333	0.2222	0.1600	0.4286	0.4167	0.6250	0.4762	0.6667
Het. Esp.	0.4982	0.5770	0.5778	0.1535	0.6203	0.6630	0.6417	0.5161	0.5333
<i>P</i>	0.008	1.000	0.000	1.000	0.000	0.008	0.120	0.870	1.000

Het. Obs.: Taxa de Heterozigose observada

Het. Esp.: Taxa de Heterozigose esperada

p: Valor de *p*

Tabela 18. Resultados dos testes de equilíbrio de Hardy-Weinberg para o método 2.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã	Waiampi
Het. Obs.	0.8750	0.5333	0.7778	0.3157	0.8409	0.6364	0.8750	0.8276	1.0000
Het. Esp.	0.5059	0.5770	0.6714	0.2902	0.7487	0.7273	0.7583	0.5862	0.5862
<i>P</i>	0.000	1.000	0.760	1.000	0.005	0.079	0.074	0.002	1.000

Het. Obs.: Taxa de Heterozigose observada

Het. Esp.: Taxa de Heterozigose esperada

p: Valor de *p*

5.1.4. Distribuição das frequências alélicas

Uma vez que o método 3 foi avaliado como sendo o mais apropriado para a determinação das frequências alélicas, a discussão da distribuição das frequências dos alelos encontrados no presente trabalho foi baseada nos resultados obtidos através dessa metodologia.

Os alelos mais frequentes encontrados nas tribos indígenas são *KIR2DL4*00501* e *KIR2DL4*00102/05*, respectivamente. Esses alelos também são os mais frequentes em populações não-Ameríndias (tabela 19). Entretanto, há grande variação na frequência de vários outros alelos, sendo mais marcante nos alelos *KIR2DL4*00501* (cuja frequência varia de 51.3% a 4.5%), *KIR2DL4*00102/03/05* (57% a 18.4%), *KIR2DL4*00801* (23% a 0.5%) e *KIR2DL4*00802* (20.7% a 0.3%).

Tabela 19. Frequências alélicas do gene *KIR2DL4* em várias populações.

Alelos/N	Populações					
	Francesa ¹	Irlandesa ²	República do Congo ¹	Japonesa ³	Chinesa ⁴	Ameríndios
	49	230	49	132	78	219
*00102/03/05	18.4	25.9	47.2	51.3	57.0	30.2
*00201	1.0	0	2.5	0	0	0
*00202	0	0	1.0	0	1.9	0
*00203	0	0	3.6	0	0	0
*005	15.0	21.9	4.5	17.1	15.4	51.3
*011	16.4	13.0	4.5	12.8	14.1	3.3
*006	2.0	2.6	2.0	12.8	7.1	0.2
*00801	23.0	20.4	1.0	6.0	5.1	0.5
*00802	20.7	16.1	11.4	0	0	0.3
*12	0	0	6.1	0	0	0
*New^a	0	0	0	0		1.8
*New 1^b	0	0	1.0	0	0	0
*New 2^b	0	0	1.0	0	0	0
*New 3^b	0	0	1.0	0	0	0
*New 4^b	0	0	1.0	0	0	0
*Nulo/blank^c	3.5	0	3.4	0	0	12.4

¹ Dados de frequências alélicas obtidos em BUHLER *et al.* (2009). ² Dados de frequências alélicas obtidos em GUINAN *et al.* (2010). ³ Dados de frequências alélicas obtidos em YAWATA *et al.* (2006). ⁴ Dados de frequências alélicas obtidos em ZHU *et al.* (2006). ^a Alelo novo descrito no presente trabalho. ^b Alelos novos descritos em BUHLER *et al.* (2009). *Blank^c*: Alelo não definido (ou indetectável) encontrado em BUHLER *et al.* (2009) através do algoritmo EM para múltiplos genótipos. Para fim de comparação foram agrupados na mesma categoria dos alelos nulos.

5.1.5. Diferenciação inter-populacional (F_{ST})

Foi calculado o F_{ST} entre os pares de populações não-Ameríndias e o grupo Ameríndio como um todo (figura 2, tabela 20). A maior parte das comparações, com exceção daquela entre duas populações asiáticas e as duas populações europeias, revelou diferenciação significativa. As comparações entre o grupo Ameríndio e as demais populações analisadas também mostram valores de F_{ST} significativos, inclusive quando comparados com populações Asiáticas, ancestrais das populações Ameríndias.

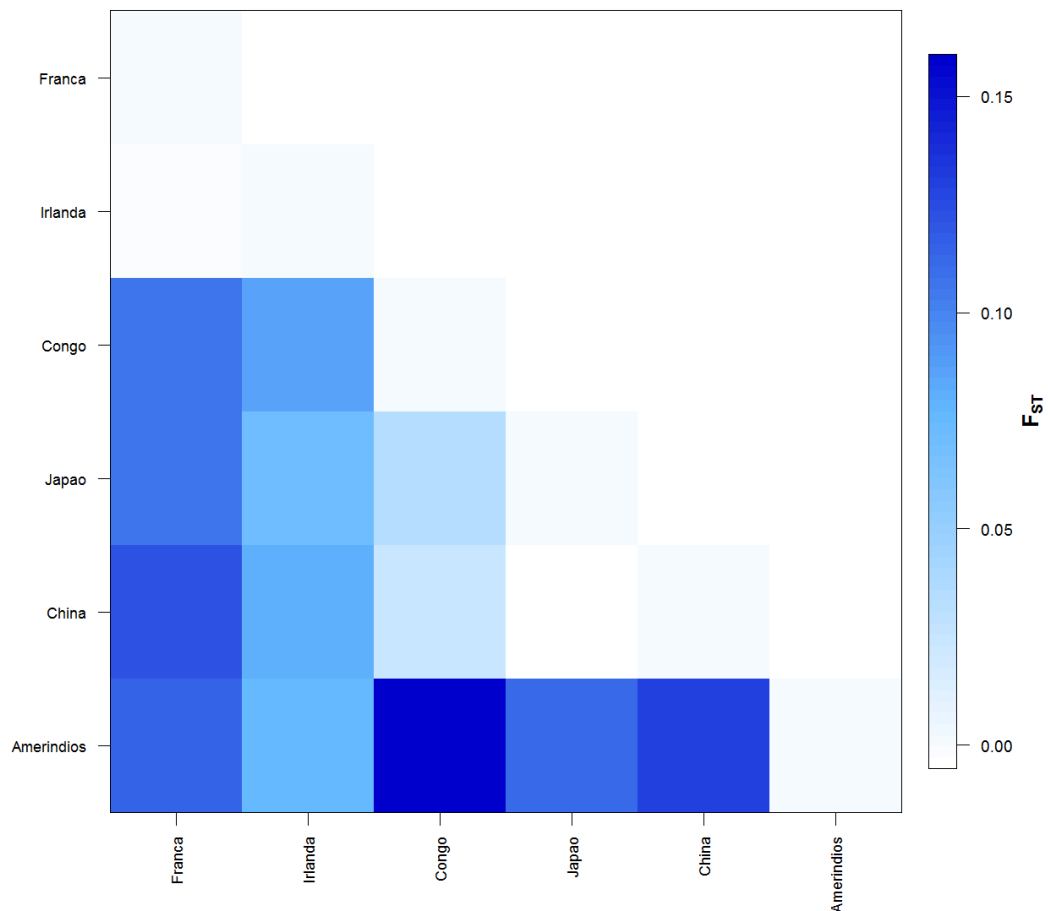


Figura 2. Diferenciação inter-populacional entre as populações não-Ameríndias e o grupo Ameríndio.

Tabela 20. Valores de *Fst* obtidos nas comparações entre o grupo Ameríndio e populações não-Ameríndias.

	França	Irlanda	República do Congo	Japão	China	Ameríndios
França	0.00000					
Irlanda	-0.00242	0.00000				
República do Congo	0.10770*	0.08628*	0.00000			
Japão	0.10676*	0.06930*	0.03367*	0.00000		
China	0.12235*	0.08149*	0.02437*	-0.00555	0.00000	
Ameríndios	0.11529*	0.07677*	0.15962*	0.11146*	0.12978*	0.00000

* Valores de *Fst* significativos

Comparações par-a-par entre as tribos estudadas no presente trabalho também foram realizadas. Apesar de termos definido que o método 3 é o método mais adequado para a estimativa das frequências alélicas, foram calculados os valores de *Fst* entre todas as tribos (com exceção de Yanomami, representada por apenas um indivíduo) para os três métodos de estimativa de frequências alélicas. O objetivo foi verificar se os valores de *Fst* são sensíveis ao método de estimativa de frequência alélica escolhido.

Os valores de *Fst* obtidos e suas respectivas significâncias são mostrados nas tabelas 21, 22 e 23 e representados na figura 3.

Para as três metodologias de estimativa de frequências alélicas empregadas, as tribos que apresentaram maior diferenciação com as demais, nas comparações par-a-par, foram Awá-Guajá e Araweté.

Para verificar a possibilidade da diferenciação encontrada para as tribos Awá-Guajá e Araweté resultarem de gargalos populacionais recentes, estimamos o índice de Garza-Williamson (GW, utilizado para quantificar a ocorrência de gargalo populacional através de seu efeito sobre a distribuição de frequências alélicas em microssatélites) foi aplicado para o conjunto de 23 marcadores microssatélites (tabela 24). Os dados mostram que as tribos Awá-Guajá e Araweté são justamente aquelas que apresentam os menores valores de GW obtido entre as populações analisadas, sugerindo a ocorrência de efeito de gargalo populacional mais pronunciado nessas tribos.

Tabela 21. Valores de *Fst* obtidos nas comparações entre as populações Ameríndias pelo método 1.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã	Waiampi
Arara	0.00000								
Assurini	-0.01507	0.00000							
Awá-Guajá	0.15190*	0.11922*	0.00000						
Araweté	0.27270*	0.34015*	0.17817*	0.00000					
Xikrin	0.01508	0.00081	0.05854*	0.16826*	0.00000				
Urubu-Kaapor	-0.00040	-0.01876	0.11474*	0.33223*	-0.00594	0.00000			
Katuena	0.01400	-0.00356	0.03223	0.23099*	-0.02736	-0.02275	0.00000		
Parakanã	0.00518	-0.00950	0.20847*	0.36065*	0.04747*	0.00443	0.05431	0.00000	
Waiampi	-0.07379	-0.06724	0.02613	0.23610	-0.07326	-0.06162	-0.09196	-0.01467	0.00000

* Valores de *Fst* significativos

Tabela 22. Valores de *Fst* obtidos nas comparações entre as populações Ameríndias pelo método 2.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã	Waiampi
Arara	0.00000								
Assurini	-0.01507	0.00000							
Awá-Guajá	0.22161*	0.17554*	0.00000						
Araweté	0.27270*	0.34015*	0.35588*	0.00000					
Xikrin	0.04795*	0.02675	0.05157*	0.19074*	0.00000				
Urubu-Kaapor	0.03287	0.00658	0.07107	0.34475*	-0.01138	0.00000			
Katuena	0.03998	0.00589	0.04542	0.32112*	-0.02665	-0.04380	0.00000		
Parakanã	0.04971*	0.02238	0.25530*	0.43324*	0.09871*	0.04539	0.07901	0.00000	
Waiampi	-0.01946	-0.01030	0.15452*	0.18765*	0.00589	0.01926	0.00085	0.11753	0.00000

* Valores de *Fst* significativos

Tabela 23. Valores de *Fst* obtidos nas comparações entre as populações Ameríndias pelo método 3.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã	Waiampi
Arara	0.00000								
Assurini	-0.01507	0.00000							
Awá-Guajá	0.19376*	0.14967*	0.00000						
Araweté	0.27270*	0.34015*	0.30508*	0.00000					
Xikrin	0.04664*	0.02419	0.03763*	0.20016*	0.00000				
Urubu-Kaapor	0.06859*	0.03518	0.04927	0.38998*	-0.00624	0.00000			
Katuena	0.00655	-0.01035	0.05113	0.22642*	-0.02122	0.00711	0.00000		
Parakanã	0.00586	-0.01034	0.20773*	0.35199*	0.05813*	0.04866	0.04081	0.00000	
Waiampi	-0.01946	-0.01030	0.11840*	0.18765*	0.00929	0.05825	-0.05569	0.04265	0.00000

* Valores de *Fst* significativos

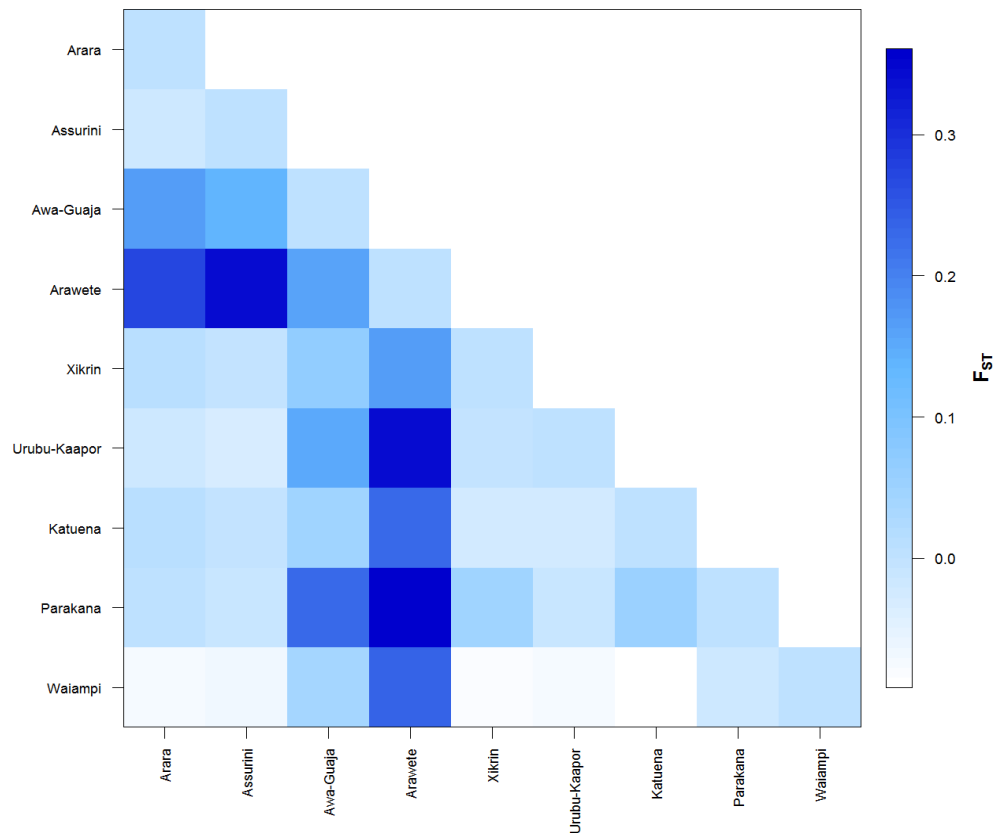


Figura 3. Diferenciação inter-populacional entre as populações Ameríndias.

Tabela 24. Resultados do teste de Garza-Williamson.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Parakana
Índice de Garza-Williamson	0.1777	0.1890	0.1493	0.1446	0.2392	0.1899
s.d	0.0473	0.0511	0.0465	0.0871	0.0397	0.0497

s.d: desvio padrão

Para quantificar o grau de semelhança entre as estimativas de F_{ST} obtidas a partir das frequências estimadas pelos diferentes métodos, foi estimada a correlação entre os valores de F_{ST} entre pares de populações obtidos por cada par de métodos. Os resultados mostraram que, para as estimativas de F_{ST} , as correlações entre todos os métodos são altas

($R > 0.95$ em todos os casos), indicando que a análise de diferenciação populacional é robusta a efeitos associados ao método de estimativa de frequências alélicas escolhido.

5.1.6. Testes de neutralidade

A hipótese de haver seleção natural atuando no gene *KIR2DL4* foi estudada através dos testes de Ewens-Watterson e D de Tajima. Assim como as análises de comparação inter-populacional (F_{st}), os testes de neutralidade foram aplicados aos dados obtidos através dos três métodos de estimativa de frequências alélicas. Novamente, o objetivo foi verificar se os resultados obtidos por esses testes são sensíveis ao método de estimativa de frequência alélica escolhido (tabelas 25, 26 e 27).

Tabela 25. Resultados obtidos nos testes de neutralidade pelo método 1.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã
Teste de Ewens-Watterson								
F observado	0.50822	0.44222	0.43827	0.84960	0.38520	0.36458	0.39844	0.49008
F esperado	0.80879	0.59499	0.48665	0.63079	0.34348	0.46101	0.41423	0.65350
<i>P</i>	0.0500	0.1790	0.44500	0.84200	0.70030	0.24900	0.58900	0.22800
D de Tajima								
D	3.01661	1.94290	-0.80109	-1.51568	0.86287	1.40021	0.85212	1.29883
<i>P</i>	1.00000	0.97900	0.24300	0.04600	0.83200	0.91300	0.82800	0.89400

p: Valor de *p*

Tabela 26. Resultados obtidos nos testes de neutralidade pelo método 2.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã
Teste de Ewens-Watterson								
F observado	0.50822	0.44222	0.34722	0.77920	0.32047	0.31944	0.28906	0.47024
F esperado	0.80649	0.59961	0.49470	0.52738	0.34350	0.45360	0.33045	0.56209
<i>P</i>	0.0540	0.16600	0.15500	0.90600	0.52600	0.10000	0.37500	0.36400
D de Tajima								
D	3.05287	1.94290	-0.72534	-1.42634	0.93242	1.41642	1.11460	1.59581
<i>P</i>	1.00000	0.98100	0.26400	0.07500	0.84700	0.91700	0.87700	0.93500
<i>p</i> : Valor de <i>p</i>								

Tabela 27. Resultados obtidos nos testes de neutralidade pelo método 3.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Katuena	Parakanã
Teste de Ewens-Watterson								
F observado	0.50822	0.44222	0.34551	0.84960	0.31353	0.31597	0.39062	0.46740
F esperado	0.78921	0.59057	0.49847	0.62902	0.34371	0.45349	0.33404	0.54781
<i>P</i>	0.05500	0.18000	0.14700	0.83200	0.49000	0.11900	0.83800	0.37600
D de Tajima								
D	3.01661	1.94290	-0.63648	-1.51568	0.80901	1.08689	0.51038	1.66035
<i>P</i>	0.99800	0.9700	0.29300	0.0390	0.82000	0.87400	0.73400	0.93400

p: Valor de *p*

A aplicação do teste D de Tajima revelou desvios de neutralidade significativos nas tribos Arara e Assurini, apresentando um valor de D positivo. Embora não significativos, de maneira geral há mais valores de D positivos (6) do que negativos (2) nas tribos.

Já o teste de Ewens-Waterson não mostrou desvios com relação ao esperado sob neutralidade, embora para a tribo Arara o valor de p está perto do limiar de significância.

Assim como para os dados de F_{st} , os resultados obtidos pelos testes de neutralidade nos três métodos analisados apresentaram resultados muito semelhantes, com correlações de $R > 0.99$, indicando que a análise dos testes de neutralidade também é robusta a efeitos associados ao método escolhido.

5.2. GENE KIR3DL1/S1

5.2.1. Caracterização da variabilidade alélica

Para o gene *KIR3DL1/S1* foram analisadas 569 amostras pertencentes a 14 tribos Ameríndias (tabela 28).

Na primeira parte deste trabalho as amostras foram analisadas para presença e ausência dos alótipos *KIR3DL1* e *KIR3DS1*, respectivamente. As frequências de portadores dos dois alótipos encontram-se na tabela 28. Nas figuras 4 e 5 foram apresentadas as distribuições de frequências desses alótipos em populações mundiais previamente estudadas, cujos resultados foram sumarizados no site www.allelefreqencies.net. Essas figuras são apresentadas para permitir uma comparação com os dados das tribos Ameríndias obtidas no presente trabalho.

Tabela 28. Frequência de portadores dos alótipos *KIR3DL1* e *KIR3DS1* em populações Nativo-Americanas.

População	Frequência de portadores	
	<i>KIR3DL1</i>	<i>KIR3DS1</i>
Arara (n=31)	74.2	90.3
Assurini (n=20)	80.0	75.0
Awá-Guajá (n=13)	42.9	71.4
Araweté (n=38)	15.8	97.4
Xikrin (n=54)	59.3	66.7
Urubu-Kaapor (n=50)	66.0	76.0
Karitiana (n=55)	81.8	80.0
Katuena (n=15)	80.0	80.0
Maya (n=50)	94.0	56.0
Pima (n=95)	77.9	70.5
Parakanã (n=30)	70.0	80.0
Surui (n=46)	91.3	30.4
Ticuna (n=63)	93.7	46.0
Waiampi (n=9)	88.9	62.5

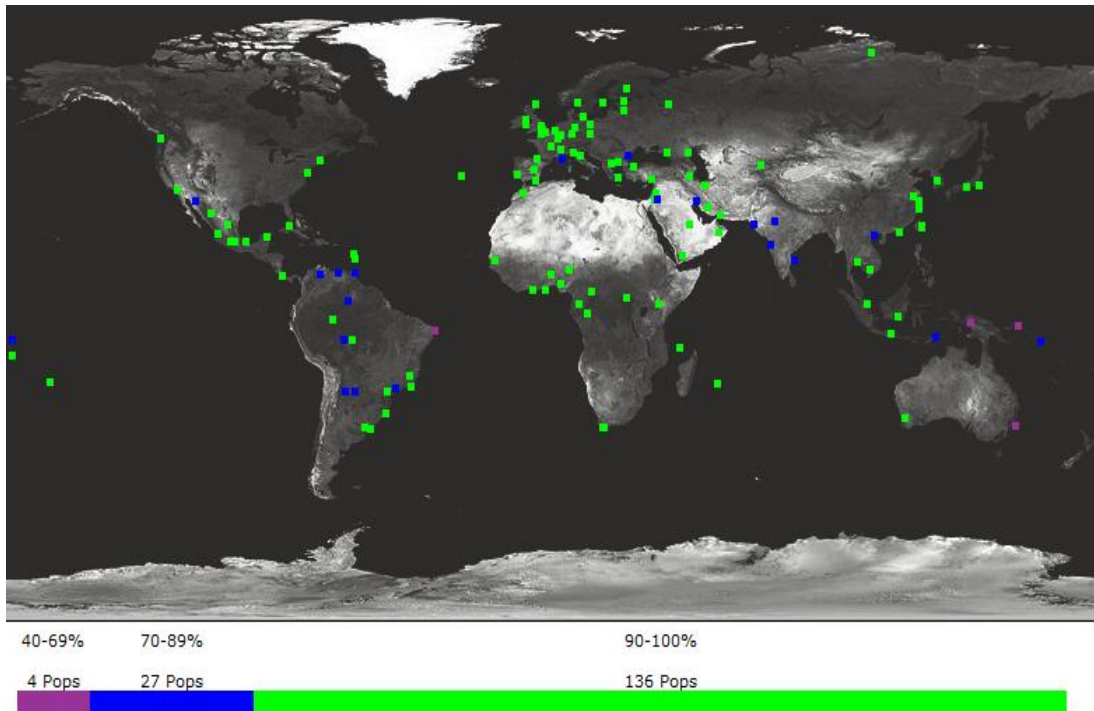


Figura 4. Distribuição da frequência do alótipo *KIR3DL1* em populações mundiais previamente estudadas (fonte: www.allelefrequencies.net).

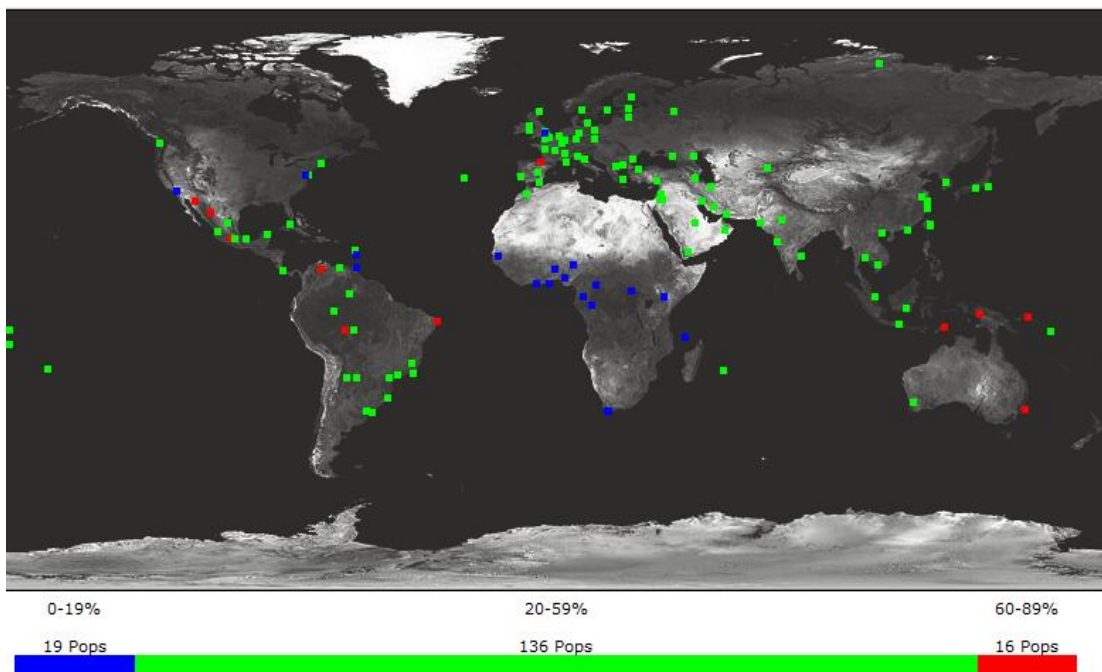


Figura 5. Distribuição da frequência do alótipo *KIR3DS1* em populações mundiais previamente estudadas (fonte: www.allelefrequencies.net).

É interessante notar que das 14 tribos Ameríndias analisadas neste trabalho, 11 apresentaram frequências de portadores do alótipo *KIR3DS1* maior que 60%, sendo que em algumas tribos esse valor passa de 90%. Das 171 populações mundiais analisadas, apenas 16 apresentam frequência desse alótipo superior a 60%. Esse resultado mostra que as tribos Ameríndias, especialmente nas Amazônicas, o alótipo *KIR3DS1* é mantido em frequências excepcionalmente altas.

Uma vez que os dados de presença e ausência não inferem a variabilidade alélica do locus, o alótipo *KIR3DL1* (o mais polimórfico dentre os genes *KIR*) foi seqüenciado para os éxons 3, 4, 5 e 9. Assim como para o locus *KIR2DL4*, os alelos foram identificados com base na combinação de sítios polimórficos nesses quatro éxons, de acordo com o *Immuno Polymorphism Database* (IPD, <http://www.ebi.ac.uk/ipd/kir>). O alótipo *KIR3DS1*, por ser pouco variável, não foi seqüenciado e será tratado nas análises como um alelo do locus *KIR3DL1/S1*.

O seqüenciamento dos quatro éxons revelou a presença de 7 alelos para o alótipo *KIR3DL1* nas tribos Ameríndias (*KIR3DL1*001*, *KIR3DL1*01502*, *KIR3DL1*029*, *KIR3DL1*00501*, *KIR3DL1*00401*, *KIR3DL1*007* e *KIR3DL1*008*), sendo que todos já foram previamente descritos. Os sítios polimórficos que permitem a discriminação dos alelos são mostrados na tabela 29.

Tabela 29. Posições dos nucleotídeos polimórficos nos alelos do alótipo *KIR3DL1*.

Exon	3						4	5				9	
Proteína	Domínio DO						Domínio D1	Domínio D2				Região Citoplasmática	
Aminoácido*	R31H	R44G	I47V	I54L	S86L		P182S	Q256K		W283L		E373Q	
Nucleotídeo	84	155	193	202	223	320	607	732	829	882	911	1180	1194
<i>3DL1*001</i>	A	G	A	A	A	C	C	T	C	C	G	G	C
<i>3DL1*01502</i>	G	-	-	G	T	-	-	C	-	-	-	-	-
<i>3DL1*029</i>	G	-	-	G	T	-	-	C	A	-	-	-	-
<i>3DL1*00501</i>	-	-	-	-	-	-	T	-	-	T	T	-	-
<i>3DL1*00401</i>	-	A	G	-	-	T	T	-	C	T	T	C	-
<i>3DL1*007</i>	G	-	-	G	T	-	C	-	-	-	-	C	T
<i>3DL1*008</i>	G	A	-	G	T	-	-	-	-	-	-	-	-

Traços indicam identidade com a referência. *Posições que apresentam mudanças de aminoácidos não-sinônimas.

As frequências de portadores dos alelos do locus *KIR3DL1/S1* foram calculadas para as tribos Ameríndias agrupadas e são apresentadas na tabela 30. Os números amostrais diferem dos mostrados na tabela 15, pois nem todos os indivíduos que tiveram a genotipagem para presença e ausência dos alótipos puderam ser sequenciados.

Tabela 30. Frequências de portadores do gene *KIR3DL1/S1* no grupo de populações Nativo-Americanas.

Alelo	Frequência(%)
	N=459
3DS1	71.5
3DL1*01502	43.1
3DL1*029	19.6
3DL1*00501	10.9
3DL1*001	4.1
3DL1*008	1.5
3DL1*00401	0.9
3DL1*007	0.2

Assim como para o gene *KIR2DL4*, a metodologia de genotipagem do locus *KIR3DL1/S1* baseia-se na detecção de ausência e presença de genes *KIR* específicos. Aqui, novamente, o desafio na análise das frequências alélicas é a distinção entre os indivíduos que são realmente homozigotos daqueles que são heterozigotos para a ausência do gene. As mesmas três metodologias descritas para as análises do gene *KIR2DL4* também foram aplicadas para as análises com o gene *KIR3DL1/S1*. A seguir serão apresentadas as frequências alélicas do gene *KIR3DL1/S1* nas tribos indígenas aqui estudadas para cada uma das metodologias (tabelas 31, 32 e 33).

Tabela 31. Frequências alélicas do gene *KIR3DL1/S1* para a método 1.

	Arara	Assurini	Awa-Guajá	Araweté	Xikrin	Urubu-Kaapor	Karitiana	Katuena	Maya	Pima	Parakanã	Surui	Ticuna	Waiampi
Alelos\2N	62	32	40	72	98	76	94	18	80	136	52	70	92	10
3DS1	58.1	50.0	60.0	93.1	54.1	61.8	50.0	61.1	30.0	52.2	57.7	22.9	26.1	40.0
3DL1*01502	32.3	25.0	2.5	0	13.3	28.9	43.6	22.2	36.2	16.2	13.5	52.9	38.0	30.0
3DL1*029	9.6	18.8	2.5	2.8	16.3	0	6.4	0	10.0	25.0	28.8	1.4	6.5	20.0
3DL1*00501	0	6.2	25.0	0	5.1	2.7	0	5.6	11.3	2.2	0	18.5	20.7	10.0
3DL1*001	0	0	0	4.1	6.1	4.0	0	11.1	6.3	3.7	0	1.4	1.1	0
3DL1*007	0	0	0	0	1.02	0	0	0	0	0	0	0	0	0
3DL1*00401	0	0	0	0	0	0	0	0	3.7	0.7	0	0	0	0
3DL1*008	0	0	0	0	0	0	0	0	0	0	0	0	7.6	0
NULO	0	0	10.0	0	4.08	2.6	0	0	2.5	0	0	2.9	0	0

Tabela 32. Frequências alélicas do gene *KIR3DL1/S1* para a método 2.

	Arara	Assurini	Awa-Guajá	Araweté	Xikrin	Urubu-Kaapor	Karitiana	Katuena	Maya	Pima	Parakanã	Surui	Ticuna	Waiampi
Alelos\2N	62	32	40	72	98	76	94	18	80	136	52	70	92	10
3DS1	58.1	50.0	40.0	80.6	44.9	55.3	43.6	50.0	27.5	42.7	48.1	18.6	30.4	30.0
3DL1*01502	32.3	25.0	2.5	0	10.2	26.3	43.6	22.2	33.7	15.4	13.5	45.7	33.7	30.0
3DL1*029	9.6	18.8	2.5	2.8	15.3	0	6.4	0	10.0	22.8	28.8	1.4	6.5	20.0
3DL1*00501	0	6.2	15.0	0	4.1	1.3	0	5.6	11.3	2.2	0	18.6	19.6	10.0
3DL1*001	0	0	0	4.2	5.1	4.0	0	11.1	6.3	3.7	0	1.4	1.1	0
3DL1*007	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0
3DL1*00401	0	0	0	0	0	0	0	0	3.7	0.7	0	0	0	0
3DL1*008	0	0	0	0	0	0	0	0	0	0	0	0	7.6	0
NULO	0	0	40.0	12.4	19.4	13.1	6.4	11.1	7.5	12.5	9.6	14.3	1.1	10.0

Tabela 33. Frequências alélicas do gene *KIR3DL1/S1* para a método 3.

	Arara	Assurini	Awa-Guajá	Araweté	Xikrin	Urubu-Kaapor	Karitiana	Katuena	Maya	Pima	Parakanã	Surui	Ticuna	Waiampi
Alelos\2N	62	32	40	72	98	76	94	18	80	136	52	70	92	10
3DS1	56.4	49.3	45.2	93.0	45.1	53.2	50.0	50.0	29.39	49.9	57.69	20.87	25.01	38.35
3DL1*01502	33.9	24.7	2.5	0	10.0	26.0	43.6	16.7	34.65	15.7	13.46	49.43	36.65	27.86
3DL1*029	9.7	18.3	2.5	2.8	14.8	0	6.4	0	10.0	23.4	28.85	1.43	6.52	20.0
3DL1*00501	0	6.3	17.1	0	4.2	1.4	0	5.6	11.25	2.3	0	17.85	19.01	10.0
3DL1*001	0	0	0	4.2	5.2	2.7	0	15.4	6.25	3.8	0	1.43	1.09	0
3DL1*007	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0
3DL1*00401	0	0	0	0	0	0	0	0	3.75	0.1	0	0	0	0
3DL1*008	0	0	0	0	0	0	0	0	0	0	0	0	7.61	0
NULO	0	1.4	32.7	0	19.7	16.7	0	12.3	4.71	4.8	0	8.98	4.11	3.78

5.2.2. O teste do equilíbrio de Hardy-Weinberg para avaliação das metodologias aplicadas

Os resultados dos testes do equilíbrio de Hardy-Weinberg são mostrados nas tabelas 34 e 35.

Assim como para o gene *KIR2DL4*, os resultados dos testes do equilíbrio de Hardy-Weinberg para a primeira metodologia empregada mostraram desvios significativos para as populações Awá-Guajá, Xikrin e Urubu- Kaapor, sendo que em todas as três há um excesso de homozigotos. Novamente, o resultado do teste indica que essa metodologia subestima a frequência de alelos nulos (pois há uma subestimativa de indivíduos heterozigotos para o alelo nulo), sendo que para tribos que apresentam a ausência do gene em homozigose esse efeito é maior, levando a desvios em relação ao esperado sob equilíbrio.

Para a segunda metodologia empregada, foram encontrados desvios significativos para as populações Arara, Karitiana e Pima, todas as três tendendo a um excesso de heterozigotos. Os resultados do teste de equilíbrio de Hardy-Weinberg indicam um excesso de heterozigotos, assim como os encontrados para o gene *KIR2DL4*, e reforçam a idéia de que essa abordagem superestima a frequência de alelos nulos em algumas populações.

O terceiro método de estimativa assume que as populações estão em equilíbrio de Hardy-Weinberg. Como apresentado anteriormente, dados de microssatélites mostraram que, em média, não há desvio de equilíbrio nessas populações (Kelly Nunes, comunicação pessoal). Dessa forma, a metodologia 3 pode ser aplicada para os dados deste estudo.

Tabela 34. Resultados dos testes de equilíbrio de Hardy-Weinberg pelo método 1.

	Arara	Assurini	Awá- Guajá	Araweté	Xikrin	Urubu- Kaapor	Karitiana	Katuena	Maya	Pima	Parakanã	Surui	Ticuna	Waiampi
Het. Obs.	0.6774	0.6250	0.2000	0.1389	0.4082	0.3684	0.6170	0.5556	0.7750	0.5441	0.5769	0.6000	0.6522	0.6000
Het. Esp.	0.5584	0.6694	0.5808	0.1334	0.6619	0.5379	0.5617	0.5948	0.7595	0.6416	0.5769	0.6418	0.7424	0.7778
p	0.52	0.73	0.00	1.00	0.00	0.00	0.72	0.12	0.51	0.23	0.92	0.05	0.29	0.19

Het. Obs.: Taxa de Heterozigose observada

Het. Esp.: Taxa de Heterozigose esperada

p : Valor de p

Tabela 35. Resultados dos testes de equilíbrio de Hardy-Weinberg pelo método 2.

	Arara	Assurini	Awá- Guajá	Araweté	Xikrin	Urubu- Kaapor	Karitiana	Katuena	Maya	Pima	Parakanã	Surui	Ticuna	Waiampi
Het. Obs.	0.9130	0.6250	0.8000	0.3889	0.8333	0.6286	0.8140	0.7778	0.8750	0.8709	0.8333	0.8286	0.7610	0.8000
Het. Esp.	0.6184	0.6694	0.6731	0.3376	0.7785	0.6406	0.6230	0.7124	0.7867	0.7562	0.6932	0.7114	0.7648	0.8444
<i>P</i>	0.00	0.73	0.73	1.00	0.15	0.07	0.00	0.19	0.63	0.03	0.26	0.15	0.54	0.36

Het. Obs.: Taxa de Heterozigose observada

Het. Esp.: Taxa de Heterozigose esperada

p: Valor de *p*

5.2.3. *Distribuição das frequências alélicas*

A discussão da distribuição das frequências dos alelos do gene *KIR3DL1/S1* encontrados no presente trabalho foi baseada nos resultados obtidos através da metodologia 3.

Comparada com a diversidade alélica em 28 populações analisadas por NORMAN et al. (2007) (tabela 36), percebe-se que as tribos indígenas apresentam diversidade reduzida em relação ao número de alelos do alótipo *KIR3DL1*. Por exemplo, as populações africanas apresentam em média 32 alelos, as populações européias apresentam 27 alelos, as populações do sul Asiático 17 alelos, as populações do sudeste e leste Asiático com 13 alelos e as tribos Ameríndias aqui estudadas com 7 alelos. Comparações com as populações Asiáticas estudadas por esses autores mostraram que os alelos encontrados nas tribos indígenas são um sub-conjunto dos alelos descritos para populações do leste e sudeste Asiático, resultado que se encaixa com o modelo de colonização da América a partir da Ásia.

Os alelos do alótipo *KIR3DL1* encontrados nas tribos estudadas no presente trabalho também são os mesmos encontrados nas tribos indígenas Venezuelanas analisadas por NORMAN et al. (2007), com exceção do alelo *KIR3DL1*001* que é compartilhado com todas as outras 25 populações mundiais analisadas por esses autores, mas não é representado nas tribos indígenas Venezuelanas.

5.2.4. Diferenciação inter-populacional (*Fst*)

Foram calculados os valores de *Fst* entre todas as tribos estudadas neste trabalho, com exceção da tribo Waiampi, representada por apenas três indivíduos. Os valores de *Fst* obtidos e suas significâncias são mostrados nas tabelas 37, 38 e 39 e representados na figura 6.

Aqui, assim como nas análises para o gene *KIR2DL4*, houve uma concordância nos resultados de comparações inter-populacionais. As tribos que apresentaram maior diferenciação nas comparações par-a-par foram Awá-Guajá e Araweté. As tribos Surui e Ticuna também mostraram valores de *Fst* significativos mais altos que as demais.

Tabela 37. Valores de *Fst* obtidos nas comparações entre as populações Ameríndias pelo método 1.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Karitiana	Katuena	Maya	Pima	Parakanã	Surui	Ticuna	Waiamp i
Arara	0.00000													
Assurini	-0.00364	0.00000												
Awá-Guajá	0.11213*	0.07277*	0.00000											
Araweté	0.25273*	0.30953*	0.22830*	0.00000										
Xikrin	0.02647*	-0.00442	0.04392*	0.17188*	0.00000									
Urubu-Kaapor	-0.00109	0.02430	0.08816*	0.20332*	0.03529*	0.00000								
Karitiana	0.00488	0.02421	0.17028*	0.32928*	0.07475*	0.02602*	0.00000							
Katuena	-0.00426	0.00376	0.04308	0.24529*	0.00079	-0.02469	0.03417	0.00000						
Maya	0.05601*	0.02434*	0.13204*	0.36703*	0.06884*	0.07743*	0.03706*	0.05463	0.00000					
Pima	0.03162*	-0.00759	0.08618*	0.19990*	0.00044	0.05948*	0.07684*	0.03103	0.07361*	0.00000				
Parakanã	0.04389*	0.00199	0.10021*	0.24086*	0.00510	0.07818*	0.10310*	0.05217	0.10471*	-0.00764	0.00000			
Surui	0.13639*	0.11431*	0.23309*	0.50946*	0.17696*	0.15722*	0.08121*	0.15006*	0.02102	0.18459*	0.22985*	0.00000		
Ticuna	0.09300*	0.05564*	0.14526*	0.39931*	0.10752*	0.11233*	0.06789*	0.09031*	0.00322	0.11557*	0.14846*	0.01065	0.00000	
Waiampi	-0.01236	-0.05875	0.07595*	0.48604*	-0.01276	0.02737	-0.00621	0.00194	-0.03793	-0.02025	0.00337	0.03617	-0.0145	0.00000

* Valores de *Fst* significativos

Tabela 38. Valores de *Fst* obtidos nas comparações entre as populações Ameríndias pelo método 2.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Karitiana	Katuena	Maya	Pima	Parakanã	Surui	Ticuna	Waiamp i
Arara	0.00000													
Assurini	-0.02117	0.00000												
Awa-Guajá	0.18302*	0.13907*	0.00000											
Araweté	0.22295*	0.17254*	0.21305*	0.00000										
Xikrin	0.06596*	0.03433*	0.03787*	0.16059	0.00000									
Urubu- Kaapor	0.01185	0.02291	0.10123*	0.13311*	0.04413*	0.00000								
Karitiana	-0.00001	0.03455	0.19077*	0.27362*	0.09456*	0.03448*	0.00000							
Katuena	-0.00334	0.00090	0.06743	0.13056*	0.01136	-0.02883	0.03101	0.00000						
Maya	0.01545*	0.02987*	0.11940*	0.25952*	0.04587*	0.04734*	0.01899*	0.01876	0.00000					
Pima	0.03674*	0.01108	0.08613*	0.16982*	0.00103	0.05010*	0.07350*	0.02316	0.03842*	0.00000				
Parakanã	0.04871*	0.00458	0.11452*	0.18139*	0.01340	0.06860*	0.09711*	0.04385	0.06363*	-0.00556	0.00000			
Surui	0.08991*	0.11173*	0.16365*	0.37048*	0.12071*	0.10855*	0.05232*	0.08476*	0.01520	0.11918*	0.15992*	0.00000		
Ticuna	0.05244*	0.05421*	0.15801*	0.30588*	0.08916*	0.08532*	0.04883*	0.05348*	0.00189	0.07785*	0.10752*	0.01714*	0.00000	
Waiampi	-0.02094	-0.02826	0.07877	0.30271*	-0.01228	0.02147	-0.00878	-0.01163	-0.04692	-0.02918	-0.01148	-0.00098	-0.0257	0.00000

* Valores de *Fst* significativos

Tabela 39. Valores de *Fst* obtidos nas comparações entre as populações Ameríndias pelo método 3.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Karitiana	Katuena	Maya	Pima	Parakanã	Surui	Ticuna	Waiampi
Arara	0.00000													
Assurini	-0.00360	0.00000												
Awá-Guajá	0.15749*	0.10612*	0.00000											
Araweté	0.26880*	0.30953*	0.35365*	0.00000										
Xikrin	0.06775*	0.02507	0.02395	0.23028*	0.00000									
Urubu-Kaapor	0.02351	0.03007	0.06436*	0.24484*	0.02918*	0.00000								
Karitiana	-0.00030	0.02421	0.19472*	0.32928*	0.10420*	0.04100*	0.00000							
Katuena	0.03489	0.01621	0.03863	0.31607*	0.00195	-0.00733	0.06474*	0.00000						
Maya	0.05038*	0.02369	0.11332*	0.35999*	0.06029*	0.05279*	0.04055*	0.03219	0.00000					
Pima	0.03485*	-0.00713	0.09748*	0.20086*	0.01627*	0.05105*	0.07478*	0.02551	0.06103	0.00000				
Parakanã	0.04860*	0.00199	0.14262*	0.24086*	0.04092*	0.08459*	0.10310*	0.06672	0.10146*	-0.00326	0.00000			
Surui	0.12185*	0.10773*	0.18743*	0.49596*	0.14783*	0.11405*	0.08255*	0.11750*	0.01812*	0.16176*	0.21908*	0.00000		
Ticuna	0.08331*	0.05338*	0.12941*	0.39343*	0.09356*	0.08262*	0.06680*	0.06726*	0.00080*	0.10019*	0.14330*	0.00711	0.00000	
Waiampi	-0.01688	-0.05875	0.08724	0.48604*	0.00453	0.01635	-0.00621	-0.00348	-0.03759	-0.02336	0.00337	0.03027	-0.0161	0.00000

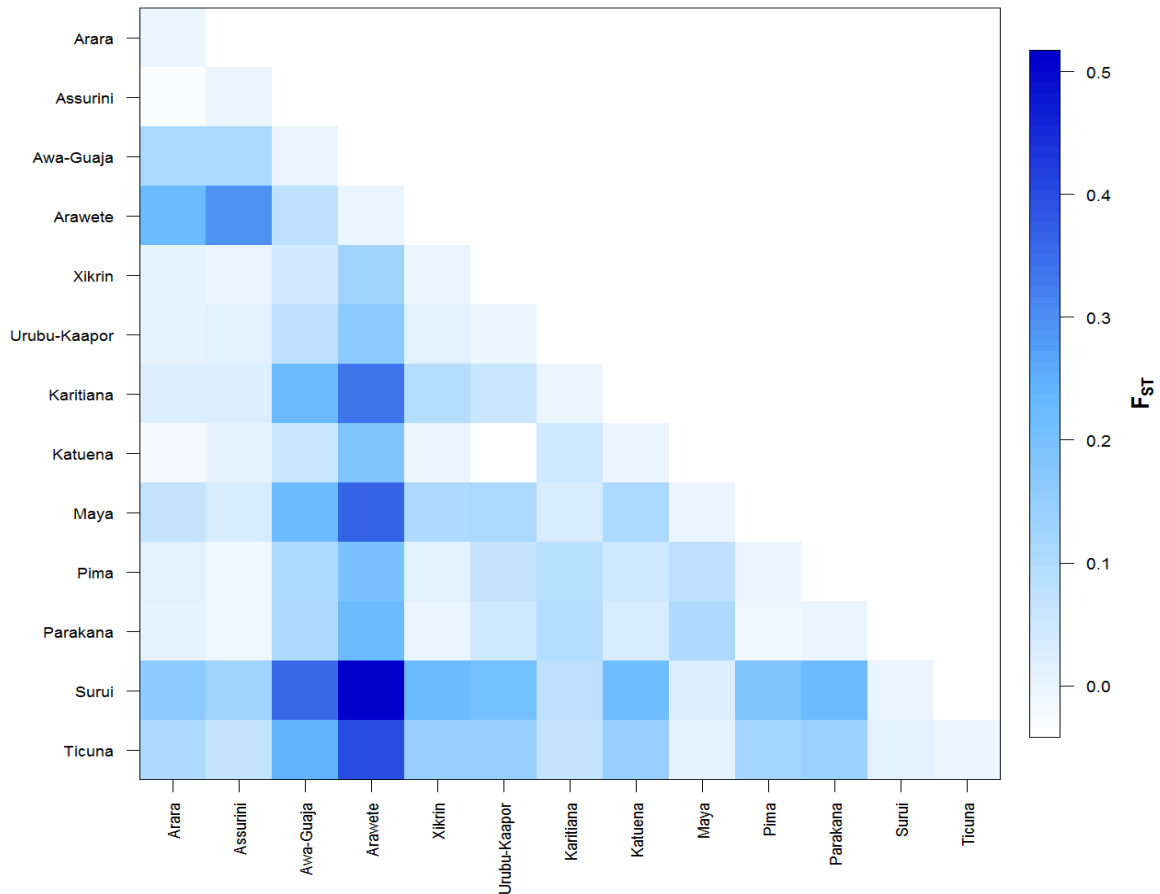


Figura 6. Diferenciação inter-populacional entre as populações Ameríndias.

5.2.5. Testes de neutralidade

Para investigar a ação de seleção natural no gene *KIR3DL1/S1* nas tribos aqui estudadas, os testes de neutralidade de Ewens-Watterson e D de Tajima foram empregados. A primeira parte das análises foi feita considerando a frequência de presença e ausência dos alótipos *KIR3DL1* e *KIR3DS1*, através do teste de Ewens-Waterson para o método 3. Para o teste D de Tajima, apenas o alótipo *KIR3DL1* foi analisado, nas tribos com

número amostral maior que 5. O teste foi aplicado aos resultados obtidos pelos três métodos de estimativa de frequências alélicas e os resultados encontram-se nas tabelas 40, 41, 42 e 43.

Tabela 40. Resultados obtidos no teste de neutralidade de Ewens-Waterson para presença e ausência dos alótipos *KIR3DL1* e *KIR3DS1*.

	Arara	Assurini	Awá-Guajá	Araweté	Xikrin	Urubu-Kaapor	Karitiana	Katuena	Maya	Pima	Parakana	Surui	Ticuna
2N	62	32	40	72	98	76	94	18	81	136	52	70	92
F observado	0.50832	0.50000	0.35875	0.87076	0.36672	0.39785	0.50000	0.41358	0.51837	0.45978	0.51183	0.54327	0.56356
F esperado	0.78901	0.75531	0.60862	0.80085	0.67127	0.65926	0.80683	0.55985	0.65448	0.68639	0.77813	0.65694	0.66899
<i>p</i>	0.06800	0.02100	0.03400	0.54000	0.01800	0.06000	0.00300	0.14200	0.31000	0.12000	0.08200	0.34400	0.35300

p: valor de *p*

Tabela 41. Resultados obtidos no teste de neutralidade de D de Tajima pelo metodo 1.

	Arara	Assurini	Xikrin	Urubu-Kaapor	Karitiana	Maya	Pima	Parakanã	Surui	Ticuna
N	26	16	41	27	47	54	65	22	52	69
D de Tajima	0.6689	-0.4151	0.7394	-0.2257	-0.1939	1.0276	-0.6476	1.1422	1.3667	2.0445
p	0.8500	0.3840	0.8040	0.4440	0.2970	0.8610	0.2950	0.9060	0.9140	0.9740

p : valor de p

Tabela 42. Resultados obtidos no teste de neutralidade de D de Tajima pelo metodo 2.

	Arara	Assurini	Xikrin	Urubu-Kaapor	Karitiana	Maya	Pima	Parakanã	Surui	Ticuna
N	26	16	36	24	47	52	63	22	47	64
D de Tajima	0.6689	-0.4151	0.5469	-0.7755	-0.1939	1.1010	-0.6198	1.1422	1.5851	2.0927
ρ	0.8470	0.3720	0.7510	0.2380	0.3130	0.8830	0.3190	0.8850	0.9430	0.9860

ρ : valor de p

Tabela 43. Resultados obtidos no teste de neutralidade de D de Tajima pelo metodo 3.

	Arara	Assurini	Xikrin	Urubu-Kaapor	Karitiana	Maya	Pima	Parakanã	Surui	Ticuna
N	26	16	35	24	47	53	62	22	56	67
D de Tajima	0.6690	-0.4151	0.5921	-0.7755	-0.1939	1.0641	-0.6045	1.1422	1.4519	1.9900
<i>P</i>	0.8390	0.3580	0.7730	0.2490	0.2930	0.8780	0.3150	0.8960	0.9310	0.9720

p: valor de *p*

As análises de desvio de neutralidade para a ausência e presença dos alótipos *KIR3DL1* e *KIR3DS1* através do teste de Ewens-Watterson mostraram desvios significativos de neutralidade para as tribos Assurini, Xikrin e Karitiana, indicando ação de seleção natural mantendo os dois alótipos em frequências intermediárias. Apesar de não terem sido encontrados desvios significativos nas outras tribos estudadas, com exceção de Araweté, as populações estudadas mostraram um excesso na taxa de heterozigose, reforçando a evidência de seleção balanceadora atuando nessas tribos.

A aplicação do teste D de Tajima revelou um desvio de neutralidade positivo e significativo com relação ao esperado para a tribo Ticuna. Esse resultado pode indicar que além da manutenção dos alótipos desse gene, a seleção natural também está atuando para manter diferentes linhagens do alótipo *KIR3DL1* nessa tribo.

5.3. O PAPEL DA DEMOGRAFIA

Para poder analisar o papel da demografia na diversidade dos genes *KIR* nas populações estudadas foram realizadas comparações com os dados dos 23 marcadores microssatélites autossômicos. A figura 7 mostra a comparação entre a diversidade nucleotídica do gene *KIR2DL4* e a diversidade dos microssatélites analisados (através do parâmetro theta, que é $4N\mu$, estimado a partir da taxa de heterozigose desses loci), para populações genotipadas para ambos marcadores. Os resultados mostraram que a diversidade de ambos está correlacionada ($p=0.02$), o que reitera a importância da história demográfica moldando a diversidade das populações, mesmo sobre genes selecionados.

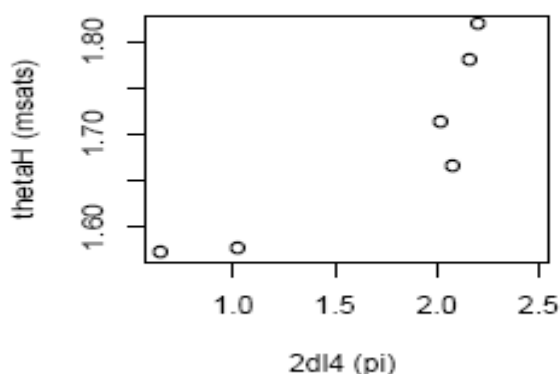


Figura 7. Correlação entre os índices de diversidade para os dados do gene *KIR2DL4* e os marcadores microssatélites, para as populações Arara, Assurini, Awá-Guajá, Araweté, Xikrin e Parakanã.

É importante notar que o fato da história demográfica (inferida pelo padrão de variação dos marcadores microssatélites) influenciar a diversidade do gene *KIR2DL4* não significa necessariamente ausência de seleção nesse locus. Como exemplo, sabe-se que genes *HLA* possuem taxa de heterozigose fortemente associada com a diversidade de microssatélites, apesar de estarem sob seleção (MEYER D, dados não publicados).

Para testar se a demografia pode explicar os resultados obtidos pelo teste D de Tajima para o gene *KIR2DL4*, a relação entre os valores de D de Tajima e a diversidade dos marcadores microssatélites foi analisada. Os resultados indicam que os valores de D de Tajima encontrados para o gene *KIR2DL4* não são totalmente explicáveis pela demografia (uma vez que a relação não é linear).

Esses resultados indicam que a demografia tem um importante papel na modulação das frequências alélicas do gene *KIR2DL4*, entretanto, para algumas populações, esse efeito não é suficiente para explicar todos os padrões encontrados, ressaltando que o efeito de seleção balanceadora encontrado nesse trabalho não é reflexo da demografia.

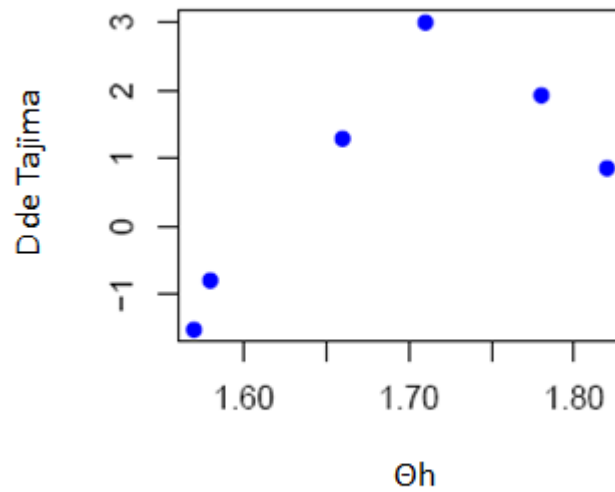


Figura 8. Correlação entre os índices de diversidade dos marcadores microssatélites e os valores de D de Tajima obtidos para o gene *KIR2DL4*.

Já as análises com gene *KIR3DL1/S1* apresentaram uma correlação negativa e não significativa. Esse resultado difere dos achados para *KIR2DL4* e para genes *HLA* clássicos. A correlação negativa encontrada pode ser decorrência do fato de que as análises feitas continham apenas a informação de diversidade do alótipo *KIR3DL1* (obtida através do seqüenciamento desse alótipo), não sendo representativa da diversidade total do locus. Dessa forma, não sendo totalmente informativa para as análises em questão.

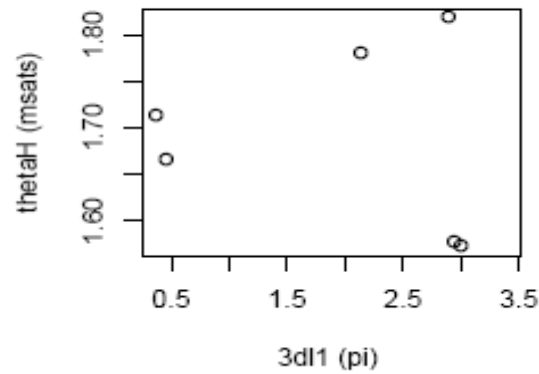


Figura 8. Correlação entre os índices de diversidade para os dados do gene *KIR3DL1/S1* e os marcadores microsatélites, para as populações Arara, Assurini, Awá-Guajá, Araweté, Xikrin e Parakanã.

No presente trabalho também foi analisada a relação entre a diversidade encontrada para o gene *KIR2DL4* e para o gene *KIR3DL1/S1* através da comparação entre as taxas de heterozigose desses dois marcadores (figura 9).

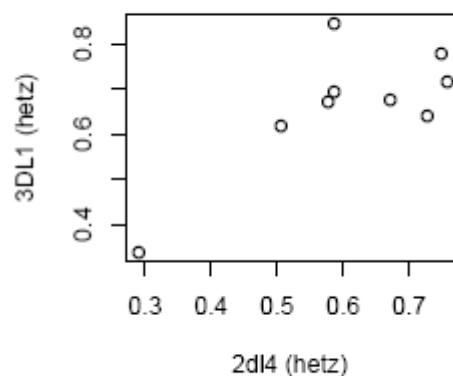


Figura 9. Correlação entre os índices de diversidade para os dados do gene *KIR2DL4* e os marcadores microsatélites.

Os resultados mostram que há correlação entre os marcadores quando comparadas as taxas de heterozigose dos dois locus. Essa correlação é essencialmente resultado da baixa variação da população Araweté, ou seja, a população menos variável apresenta esse padrão para os dois marcadores, como seria de se esperar em se tratando de um padrão de origem demográfica que afetou tanto o gene *KIR2DL4* quanto *KIR3DL1/S1*. Entretanto, a associação física desses marcadores pode implicar que ambos partilham a mesma história, seja ela demográfica ou seletiva.

5.4. HLA-G

5.4.1. Região promotora

5.4.1.1. Caracterização da variação de sequência e determinação dos haplótipos

Para a análise da região promotora do gene *HLA-G* em populações indígenas americanas foram analisados 61 indivíduos pertencentes às populações Ojibwa (n=7), Chipewyan (n=4), Guaymi (n=11), Arhuaco (n=6), Embera (n=1), Zenu (n=2), Inga (n=2), Ticuna (n=3), Xikrin (n=3), Arara (n=14), Urubu-Kaapor (n=2), Assurini (n=3), Aymara (n=1) e Huiliche (n=2).

Foram encontrados 25 polimorfismos de nucleotídeo único (SNPs), sendo 24 deles previamente descritos em populações Afro-Americanos, Euro-Americanos e Chineses, nas posições: -1306 (A/G), -1179 (G/A), -1155 (G/A), -1140 (T/A), -1138 (A/G), -964 (A/G), -810 (C,T), -762 (T/C), -725 (G/T), -716 (G/T), -689 (G/A), -666 (T/G), -633 (A/G), -540insG, -533delA, -509 (C/G), -486 (C/A), -477 (G/C), -443 (G/A), -400 (G/A), -391 (G/A), -369 (A/C), -201 (A/G), -56 (C/T), e um polimorfismo novo na posição -1098 (G/A). As posições dos polimorfismos encontrados no presente trabalho são mostradas na figura 10.

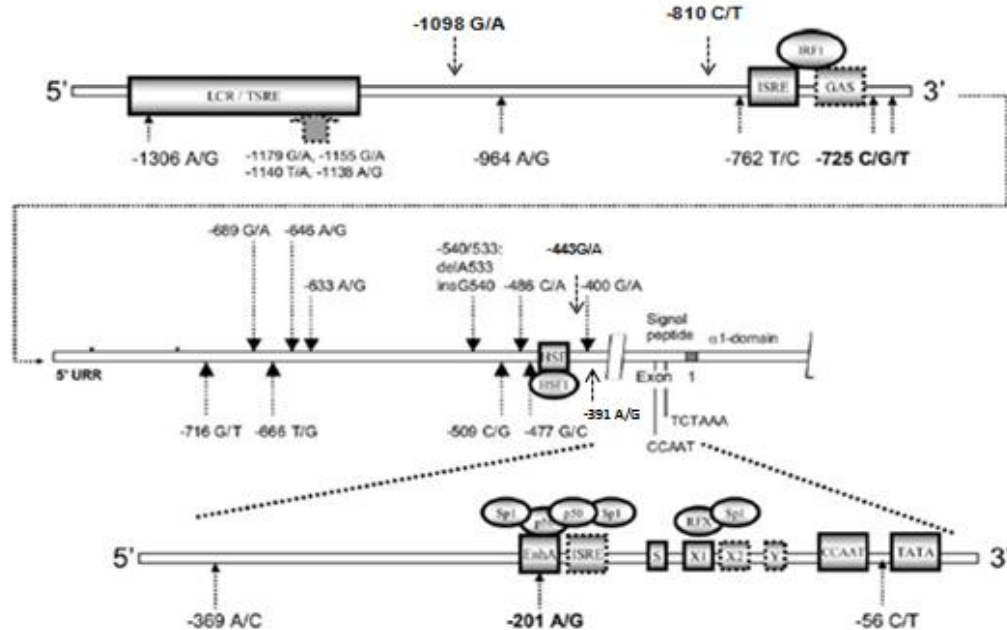


Figura 10. Desenho esquemático da região 5'UTR do gene *HLA-G* mostrando a localização dos polimorfismos encontrados no presente trabalho. Em destaque, os fatores de transcrição e a localização de sítios de ligação dos sítios de transcrição conhecidos nessa região (Figura modificada de LARSEN *et al.*, 2009).

Como já destacado por TAN *et al.* (2005), a maioria dos polimorfismos para a região promotora do gene *HLA-G*, que incluem os polimorfismos encontrados no presente trabalho, está localizada perto ou dentro de elementos reguladores, sugerindo que as mesmas afetam a ligação de fatores nucleares específicos e, dessa forma, podem influenciar os níveis de transcrito.

Com base nos polimorfismos encontrados foram identificados 35 diferentes haplótipos nas populações indígenas americanas (Tabela 44). Os haplótipos foram nomeados de 5'UTR_1 a 5'UTR_35, com base em suas frequências.

Os três haplótipos mais comuns encontrados nas populações indígenas americanas também são encontrados em populações Euro-Americana, Afro-Americana e Chinesa (TAN *et al.*, 2005). O haplótipo 5'UTR_1, cuja frequência é de 16.5%, também é o mais freqüente

em populações Euro-Americanas (34%), Afro-Americanas (22.7%) e Chineses (37%). Já o haplótipo *5'UTR_2*, o segundo haplótipo mais frequente em tribos indígenas americanas (13.9%), foi encontrado somente na população Chinesa (9%). O haplótipo *5'UTR_3*, que apresenta a frequência de 13.1%, também foi encontrado em altas frequências em populações Euro-Americanas (43.6%), Afro-Americanas (30.7%) e em populações Chinesas (31%).

Os demais haplótipos também apresentam semelhanças aos já descritos (TAN *et al.*, 2005; HVIID *et al.*, 2005), variando em apenas poucas posições entre eles, sendo que a maioria dos haplótipos menos frequentes aqui descritos são provavelmente exclusivos de tribos indígenas americanas.

Os nossos resultados mostram que os haplótipos mais frequentes encontrados nas Américas são compartilhados com outras populações mundiais, o que sugere que esses haplótipos são mantidos através da história evolutiva das populações, indicando uma possível importância na manutenção desses haplótipos na linhagem humana.

O trabalho de TAN *et al.* (2005) sugere que a seleção natural vem mantendo duas linhagens alélicas no locus *HLA-G*, representados pelas posições -1306 e +15, sendo uma linhagem definida pela combinação G-G (denominada de linhagem "G-G") e a outra pela combinação A-A (denominada de linhagem "A-A"). Os resultados das análises dos haplótipos em populações Ameríndias mostram que eles apresentam essas duas linhagens.

Tabela 44. Haplótipos encontrados para a região promotora do gene *HLA-G* para populações Ameríndias e suas respectivas frequências.

Haplótipo	-1305	-1179	-1155	-1140	-1138	-1098	-964	-810	-762	-725	-716	-689	-666	-633	-540	-533	-509	-486	-477	-443	-400	-391	-369	-201	-56	+15	Frequência	
5ÚTR_1	G	A	G	A	A	G	G	C	C	C	T	A	G	G	-	A	C	A	C	G	G	G	C	G	C	G	16.5	
5ÚTR_2	A	G	A	A	A	G	A	C	T	C	G	G	T	A	-	A	C	C	G	A	G	G	A	A	C	A	13.9	
5ÚTR_3	A	G	G	T	A	G	A	C	T	C	G	G	T	A	-	A	C	C	G	G	G	G	A	A	C	A	13.1	
5ÚTR_4	G	G	G	A	G	G	G	C	C	T	T	A	G	G	G	A	G	A	G	G	A	A	A	A	G	T	G	8.2
5ÚTR_5	G	A	G	A	A	G	G	C	C	C	T	G	G	G	-	A	C	A	C	G	G	G	C	G	C	G	7.0	
5ÚTR_6	A	G	A	A	A	G	A	C	T	C	G	A	T	A	-	A	C	C	G	A	G	G	A	A	C	A	5.6	
5ÚTR_7	G	G	G	A	G	G	G	C	C	T	T	A	G	G	G	A	C	A	G	G	A	A	A	A	G	T	G	4.7
5ÚTR_8	G	A	G	A	A	G	G	C	C	C	T	A	G	G	-	A	C	A	G	G	G	G	C	G	C	G	2.4	
5ÚTR_9	A	G	G	T	A	G	A	C	T	C	G	A	T	A	-	A	C	C	G	G	G	G	A	A	C	A	2.4	
5ÚTR_10	A	G	A	A	A	G	A	C	T	C	G	A	T	A	-	A	C	C	G	G	G	G	A	A	C	A	1.8	
5ÚTR_11	G	A	G	A	A	G	G	C	T	C	T	A	G	G	-	A	C	A	C	G	G	G	C	G	C	G	1.7	
5ÚTR_12	A	G	A	A	A	G	A	C	C	C	G	A	T	A	-	A	C	A	G	A	G	G	A	A	C	A	1.6	
5ÚTR_13	A	G	A	A	A	G	A	C	T	C	G	A	T	A	-	A	C	A	G	A	G	G	A	A	C	A	1.5	
5ÚTR_14	G	A	G	A	A	G	G	C	C	C	T	A	G	G	-	-	C	A	C	G	G	G	C	G	C	G	1.5	
5ÚTR_15	G	G	G	A	G	G	G	C	C	T	T	A	G	G	G	A	C	A	G	A	A	A	A	A	G	T	G	1.0
5ÚTR_16	A	G	A	A	A	G	A	C	T	C	G	A	T	A	-	A	C	A	G	G	G	G	A	A	C	A	1.0	
5ÚTR_17	A	A	A	A	A	G	A	C	T	C	G	G	T	A	-	A	C	C	G	A	G	G	A	A	C	A	1.0	
5ÚTR_18	A	G	G	A	A	G	A	C	T	C	G	A	T	A	-	A	C	C	G	A	G	G	A	A	C	A	1.0	
5ÚTR_19	A	G	G	A	A	G	A	C	T	C	G	G	T	A	-	A	C	C	G	A	G	G	A	A	C	A	1.0	
5ÚTR_20	G	A	G	A	A	G	G	C	C	C	T	G	G	G	G	A	C	A	C	G	G	G	C	G	C	G	1.0	
5ÚTR_21	A	G	A	A	A	G	A	C	T	C	G	G	T	A	-	A	C	C	G	G	G	G	A	A	C	A	1.0	

Haplótipo	-1305	-1179	-1155	-1140	-1138	-1098	-964	-810	-762	-725	-716	-689	-666	-633	-540	-533	-509	-486	-477	-443	-400	-391	-369	-201	-56	+15	Frequência	
5'UTR_22	A	G	A	A	A	G	A	C	T	C	G	G	T	A	G	A	C	C	G	A	G	G	A	A	C	A	1.0	
5'UTR_23	G	G	G	A	A	G	G	C	C	T	T	A	G	G	G	A	C	C	G	G	G	G	A	G	T	G	1.0	
5'UTR_24	G	G	G	A	G	G	G	C	C	T	T	A	G	G	-	A	G	A	G	G	A	A	A	A	G	T	G	1.0
5'UTR_25	G	G	G	A	G	G	G	C	C	T	T	G	G	G	G	A	G	A	G	G	A	A	A	A	G	C	G	1.0
5'UTR_26	G	A	G	A	A	G	G	C	C	G	T	A	G	G	-	A	C	A	C	G	G	G	G	C	G	C	G	1.0
5'UTR_27	A	G	A	T	A	G	A	C	T	C	G	G	T	A	-	A	C	C	G	A	G	G	A	A	C	A	1.0	
5'UTR_28	G	G	G	A	G	G	G	C	C	C	T	A	G	G	G	A	G	A	G	G	A	A	A	A	G	T	G	0.9
5'UTR_29	G	G	G	A	A	G	G	C	C	T	T	A	G	G	G	A	C	A	G	G	G	G	G	C	G	T	G	0.8
5'UTR_30	A	A	G	A	A	G	G	C	C	C	T	A	G	G	-	A	C	A	C	G	G	G	G	C	G	C	G	0.8
5'UTR_31	A	G	G	T	A	G	A	C	C	C	G	A	T	A	-	A	C	A	G	A	G	G	A	A	C	A	0.7	
5'UTR_32	G	G	G	A	G	A	G	C	C	T	T	A	G	G	G	A	C	C	G	G	G	G	G	A	G	T	G	0.6
5'UTR_33	G	G	G	A	G	A	G	C	C	T	T	G	G	G	G	A	G	A	G	G	A	A	A	A	G	T	G	0.5
5'UTR_34	A	G	A	A	A	G	A	T	T	C	G	G	T	A	-	A	C	C	G	A	G	G	A	A	C	A	0.4	
5'UTR_35	G	A	G	A	A	G	G	C	C	C	T	G	G	G	-	-	C	A	C	G	G	G	C	G	C	G	0.4	

Os haplótipos foram nomeados de 5'UTR_1 a 5'UTR_35, com base em suas frequências.

5.4.1.2. Teste de neutralidade

Para verificar se a seleção natural está atuando na região promotora do gene *HLA-G* em populações Nativo-Americanas o teste D de Tajima foi aplicado. Na análise com as amostras foram reunidas em um único, foi observado um desvio significativo da expectativa de neutralidade e equilíbrio populacional, com valor de D positivo ($D=3,01$; $p<0,01$). Além disso, observamos altos níveis de variação nucleotídica. A diversidade nucleotídica foi de 0.00701, o que é uma ordem de grandeza mais elevada que a média para o genoma humano (0.00075) (SACHIDANANDAM *et al.*, 2001).

Resultados de D de Tajima positivos podem indicar a ação de seleção balanceadora como também subdivisão populacional. Para tentar minimizar esse efeito, uma vez que o grupo Ameríndio analisado no presente trabalho é composto por populações Nativo-Americanas de várias regiões do continente, a análise de D de Tajima foi feita para as populações divididas conforme localização geográfica de acordo com o proposto por YANG *et al.* (2010). As populações foram separadas por regiões: América do Norte, Noroeste da América do Sul, Andes e Amazônia. Os valores de D encontrados e o respectivo valor de significância estão na tabela 45.

Tabela 45. Valores de D de Tajima para os grupos Ameríndios.

	Noroeste	Amazônia	Norte	Andes
N	40	50	22	10
D de Tajima	1.71731	2.33456	1.39307	0.18199
P	0.97250	0.99500	0.94000	0.56500

p: valor de p

Os resultados de D de Tajima para os grupos separados mostram que os grupos Noroeste e Amazônia apresentam sinal de seleção balanceadora quando analisadas individualmente. Apesar do teste em populações da América do Norte não ser significativo,

o valor de p do mesmo encontra-se perto do limiar de significância. O grupo dos Andes foi o que apresentou um menor valor de D (próximo a 0, não significativo), entretanto, esse padrão encontrado pode ser devido ao baixo número amostral analisado nesse grupo.

Os nossos resultados sugerem que a seleção natural atuou na região promotora do gene *HLA-G*, gerando um aumento da variabilidade em relação ao esperado sob neutralidade. Apesar desse resultado já ser conhecido para populações Euro-Americanas, Afro-Americanas e Asiáticas (TAN *et al.*, 2005), tal resultado não havia sido documentado em populações Ameríndias.

. O fato das populações Ameríndias terem experimentado um intenso gargalo populacional associado à colonização do continente e, no entanto, apresentarem evidências de seleção balanceadora, pode ser interpretado de duas formas. Uma possibilidade é que a variação acumulada antes da ocupação do continente gerou variação que não foi perdida no processo de colonização do continente, e a assinatura seletiva que vemos é a persistência de um sinal de seleção antiga. A segunda interpretação é a de que a seleção balanceadora atuou sobre o gene em populações Ameríndias, contribuindo para a geração do padrão observado. As análises presentes não permitem distinguir entre esses dois cenários, ambos envolvendo seleção natural, uma vez que atribuem a variação observada a eventos que operaram em diferentes escalas de tempo.

5.4.1.3. Polimorfismo 14pb

Para a análise desse polimorfismo, 200 indivíduos, pertencentes a 14 populações Nativo-Americanas distribuídas por todo o continente Americano, foram analisadas. Para todas as populações estudadas foram encontrados ambos os alelos (de inserção e deleção). Não foram encontrados desvios significativos do equilíbrio de Hardy-Weinberg em todas as populações analisadas. As frequências dos alelos nas populações encontram-se na tabela 46.

Tabela 46. Frequência do polimorfismo de *14pb* do gene *HLA-G* nas populações Nativas Americanas estudadas.

População	Tamanho amostral	-14 bp (%)	+14 pb (%)
Chipewyan	17	82.3	17.7
Cree	11	81.8	18.2
Ojibwa	16	56.2	43.8
Quiche	14	42.9	57.1
Cabecar	14	67.9	32.1
Arhuaco	16	65.6	34.4
Zenu	16	59.4	40.6
Embera	14	60.7	39.3
Kogi	19	39.5	60.5
Inga	17	38.2	61.8
Aymara	11	59.1	40.9
Huiliche	18	22.2	77.8
Waunana	15	46.7	53.3
Ticuna	2	50.0	50.0
Total	200	54.3	45.7

Foi aplicado o teste de Ewens-Watterson para verificar se a seleção natural está influenciando esse padrão de variação encontrado para o polimorfismo de *14pb* do gene *HLA-G*. Para a amostra agrupada das populações Ameríndias, o teste indicou um desvio em direção a um excesso na taxa de heterozigose, com um valor de p perto do limiar de significância (tabela 47), sugerindo uma possível assinatura de seleção balanceadora nesse polimorfismo. Quando as populações foram analisadas separadamente, não foi encontrado significância no teste em nenhuma das populações estudadas. Entretanto, para todas as

populações há um excesso na taxa de heterozigose (tabela 48), compatível com a ação de seleção balanceadora. O fato de não termos encontrado significância para o teste de *Ewens-Watterson* pode ser devido ao baixo poder estatístico do teste, em particular em situações envolvendo sistemas com poucos alelos.

O fato de dispormos de diversas amostras populacionais permitiu realizar um teste adicional. Sob um cenário neutro haveriam valores de F maiores e menores do que o esperado em iguais proporções. Nós observamos 14 casos de F menor que o observado em 14 populações, que sob a hipótese nula tem $p < 0.001$ (teste binomial). Portanto, o fato de termos encontrado um desvio em direção a um excesso na taxa de heterozigose em todas as populações estudadas, é compatível com a ação de seleção atuando no polimorfismo de *14pb* em populações Ameríndias.

Tabela 46. Resultados obtidos no teste Ewens-Watterson para a amostra total.

	População Ameríndia
2N	400
F observado	0.5036
F esperado	0.8475
P	0.0290

Tabela 47. Resultados obtidos no teste Ewens-Watterson para as populações Ameríndias analisadas individualmente.

	Arhuaco	Embera	Ingano	Kogi	Zenu	Aymara	Cabecar	Chipewyan	Cree	Huiliche	Ojibwa	Quiche	Ticuna	Waunana
2N	32	28	34	38	32	22	28	34	22	36	32	28	4	30
F obs	0.5488	0.5223	0.5277	0.5222	0.5176	0.5165	0.5638	0.7093	0.7025	0.6543	0.5078	0.5102	0.5000	0.5022
F esp	0.7677	0.7502	0.7583	0.7644	0.7622	0.7545	0.7534	0.7646	0.7388	0.7635	0.7639	0.7507	0.5921	0.7552
<i>P</i>	0.1560	0.1230	0.1270	0.1040	0.1120	0.1190	0.2160	0.4020	0.4430	0.3220	0.0780	0.0880	0.2630	0.0570

F obs: F observado

F esp: F esperado

p: valor de *p*

5.4.1.4. Região 3'UTR

A ação de seleção natural sobre regiões regulatórias de genes envolvidos na resposta imune tem aparecido recorrentemente em diversos estudos (TAN *et al.*, 2005; LIU *et al.*, 2006; HAYGOOD *et al.*, 2010). Portanto, além de investigarmos o papel da seleção natural na região promotora, o presente trabalho também investigou o papel da seleção natural na região 3'UTR do gene *HLA-G*. A região 3'UTR do gene *HLA-G* foi seqüenciada em 49 amostras de uma população europeia. As análises revelaram a presença de 10 polimorfismos nas posições +2798 (G/A), +2804 (G/T), +2838 (C/G), 14pb inserção/deleção, +3003 (T/C), +3010 (G/C), +3035 (C/T), +3142 (C/G), +3187 (G/A), +3196 (C/G) (figura 11). Todos os polimorfismos aqui encontrados já foram previamente descritos.

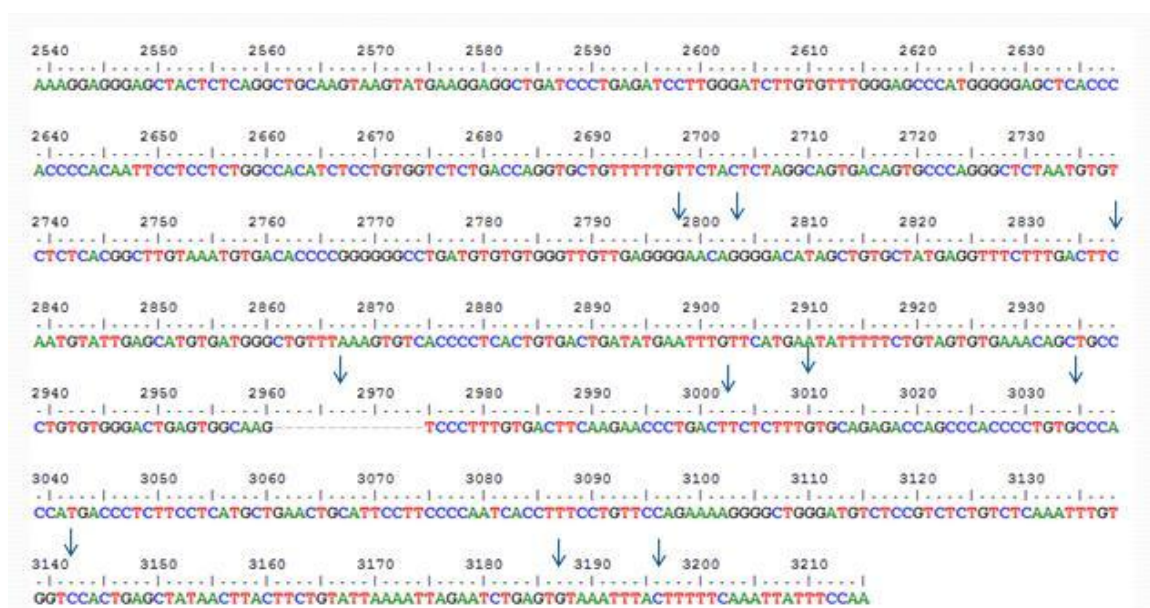


Figura 11. Estrutura genética do gene *HLA-G*, contendo a região 3'UTR. As flechas indicam as posições dos polimorfismos encontrados nesse trabalho (Figura modificada de CASTELLI *et al.*, 2010).

Com base nos polimorfismos encontrados foram identificados 12 diferentes haplótipos (tabela 48). Os haplótipos mais freqüentes encontrados na população européia aqui estudada também foram os mais freqüentes encontrados em uma população urbana brasileira (CASTELLI *et al.*, 2010).

Tabela 48. Haplótipos encontrados para a região *3'UTR* do gene *HLA-G* para populações Nativo-Americanas e suas respectivas frequências.

Haplótipos	+2798	+2804	+2838	+14pb	+3003	+3010	+3035	+3142	+3187	+3196	Frequência (%)
<i>3'UTR_1</i>	G	G	C	Del	T	G	C	C	G	C	32.6
<i>3'UTR_2</i>	A	T	G	Ins	T	C	C	G	A	G	26.3
<i>3'UTR_3</i>	G	G	C	Del	C	G	C	C	A	C	11.2
<i>3'UTR_4</i>	G	T	G	Del	T	C	C	G	A	C	11.2
<i>3'UTR_5</i>	G	G	C	Del	T	G	C	C	A	C	4.1
<i>3'UTR_6</i>	G	T	G	Ins	T	C	T	G	A	C	4.1
<i>3'UTR_7</i>	A	T	G	Ins	T	C	C	G	A	C	4.4
<i>3'UTR_8</i>	G	T	G	Ins	T	C	C	G	A	C	1.9
<i>3'UTR_9</i>	G	T	G	Ins	T	C	C	G	A	G	1.2
<i>3'UTR_10</i>	G	G	G	Ins	T	G	C	G	A	C	1.0
<i>3'UTR_11</i>	G	G	G	Ins	T	C	C	G	A	C	1.0
<i>3'UTR_12</i>	G	T	G	Ins	T	G	C	G	A	C	1.0

Para verificar se a seleção natural está atuando também na região *3'UTR* do gene *HLA-G*, o teste D de Tajima foi aplicado. Foi observado um desvio significativo da expectativa de neutralidade e equilíbrio populacional, com valor de D positivo ($D=2.81$; $p=0.9990$), corroborando a idéia da influência de seleção balanceadora nessa região, reforçando os resultados obtidos nas análises do polimorfismo de *14pb*.

5.4.1.5. Análise das regiões promotora e 3'UTR do gene HLA-G

Evidências de seleção balanceadora foram previamente descritas para a região promotora do gene *HLA-G* (TAN *et al.*, 2005). No presente trabalho, nós mostramos que a região 3'UTR também apresenta sinais de seleção balanceadora. A existência de duas regiões em um mesmo gene com assinaturas de seleção balanceadora significa que o padrão observado pode ser resultado de seleção em apenas uma das regiões, levando a um padrão de seleção na outra, mesmo que a segunda não seja selecionada. Alternativamente, ambas as regiões podem ser selecionadas. Para examinar essa questão, foram determinados os haplótipos para a combinação dessas duas regiões e o grau de associação entre os polimorfismos nas regiões promotora (5'UTR) e 3'UTR do gene *HLA-G*, quantificando os níveis de desequilíbrio de ligação.

Para esse fim, a amostra europeia caracterizada para a região 3'UTR analisada no tópico anterior foi também seqüenciada para a região promotora. Foram analisadas 42 amostras e encontrados 25 polimorfismos, sendo esses nas posições: -1306 (A/G), -1179 (G/A), -1155 (G/A), -1140 (T/A), -1138 (A/G), -1121 (C,T), -964 (A/G), -762 (T/C), -725 (G/T), -716 (G/T), -689 (G/A), -666 (T/G), -646 (A,G), -633 (A/G), -540insG, -533delA, -509 (C/G), -486 (C/A), -483 (A,G), -477 (G/C), -400 (G/A), -391 (G/A), -369 (A/C), -201 (A/G), -56 (C/T).

A análise conjunta das regiões promotora e 3'UTR nessas amostras revelou a presença de 22 diferentes haplótipos (tabela 49). Os dois haplótipos mais freqüentes encontrados apresentam frequências similares (28.4 e 22.3%) e somam mais de 50% de todos os haplótipos encontrados. Comparando esses dois haplótipos, percebe-se que 22 das 36 posições polimórficas analisadas diferem entre os mesmos, mostrando uma grande divergência entre ambos, o que pode estar relacionado com atividades reguladoras distintas. Dessas 22 posições, 13 estão localizadas na região promotora, uma no éxon 1 (na posição +15, que define as diferentes linhagens da região promotora do gene juntamente com a posição -1306), e 8 na região 3'UTR do gene. Das 13 posições divergentes localizadas na região promotora desses dois haplótipos, 8 estão localizadas dentro ou perto de sítios de

ligação a fatores de transcrição, indicando que possivelmente elas são importantes para a regulação do gene e apresentam funções antagônicas, o que vai de acordo com o estudo de TAN *et al.* (2005).

Dos 8 polimorfismos divergentes encontrados na região 3'UTR, três já foram descritos como tendo influência na variação de expressão da molécula. São eles: o polimorfismo de 14pb, a posição +3142 e a posição +3187. Estudos prévios têm reportado que o alelo de inserção de 14pb, o alelo G na posição +3142 e o alelo A na posição +3187 estão associados com baixa expressão da molécula HLA-G. CASTELLI *et al.* (2010) analisando uma população urbana brasileira encontraram haplótipos na qual esses três polimorfismos são encontrados associados e sugerem que, *in vivo*, possivelmente haja um efeito aditivo entre esses polimorfismos na regulação da expressão da molécula. No presente trabalho, nós verificamos que a inserção no polimorfismo de 14pb está sempre associada ao alelo G na posição +3142 e ao alelo A na posição +3187 em todos os haplótipos contendo a inserção de 14pb, condizentes com os achados por CASTELLI *et al.* (2010).

No presente trabalho, nós também verificamos que esses três polimorfismos estão ligados em um dos haplótipos mais freqüentes e a combinação oposta é encontrada no outro haplótipo mais freqüente, corroborando a idéia de que esses dois principais haplótipos encontrados pelas análises conjuntas das regiões promotora e 3'UTR devem possuir atividades antagônicas no que diz respeito a expressão da molécula HLA-G. Além disso, é interessante notar que esses dois haplótipos mais comuns são formados pelas duas linhagens da região promotora propostas por TAN *et al.* (2005) em combinação com os dois haplótipos divergentes da região 3'UTR. Os resultados das análises com as duas regiões mostram alto desequilíbrio de ligação entre os polimorfismos, o que mostra uma forte associação das regiões promotora e 3'UTR do gene *HLA-G*. É interessante notar que as posições +3142 e +3187 (funcionalmente implicados com diferenças na expressão de HLA-G) estão em desequilíbrio de ligação com todas as demais posições analisadas. Já o polimorfismo de 14pb, apesar de mostrar um desequilíbrio de ligação alto com as demais, possui um menor valor do que o encontrado para as posições +3142 e +3187.

Tabela 49. Haplótipos encontrados para as regiões promotora e 3'UTR do gene *HLA-G*

Haplótipo	-1306	-1179	-1155	-1140	-1138	-1121	-964	-762	-725	-716	-689	-666	-646	-633	-540	-533	-509	-486	-483	-477	-400	-391	-369	-201	-56	+15	+2798	+2804	+2838	14pb	+3003	+3010	+3035	+3142	+3187	+3196	Freq.
1	G	A	G	A	A	C	G	C	C	T	A	G	A	G	-	A	C	A	A	C	G	G	C	G	C	G	G	G	C	-	T	G	C	C	G	C	28.4
2	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	A	T	G	+	T	C	C	G	A	G	22.3
3	G	A	G	A	A	T	G	C	T	T	A	G	A	G	-	A	C	A	A	C	G	G	C	G	C	G	G	G	C	-	C	G	C	C	A	C	8.1
4	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	A	T	G	+	T	C	C	G	A	C	5.4
5	G	A	G	A	A	C	G	C	C	T	A	G	A	G	-		C	A	A	C	G	G	C	G	C	G	G	G	C	-	T	G	C	C	A	C	4.1
6	G	A	G	A	A	C	G	C	T	T	A	G	A	G	-	A	C	A	A	C	G	G	C	G	C	G	G	G	C	-	C	G	C	C	A	C	4.1
7	G	A	G	A	A	C	G	C	C	T	A	G	A	G	-	A	C	A	A	C	G	G	C	G	C	G	G	G	C	-	C	G	C	C	A	C	1.4
8	G	A	G	A	A	C	G	C	C	T	A	G	A	G	-	A	C	A	G	C	G	G	C	G	C	G	G	G	C	-	T	G	C	C	G	C	1.4
9	G	G	G	A	G	C	G	C	G	T	A	G	A	G	-	A	G	A	A	G	A	A	A	G	T	G	G	T	G	+	T	C	T	G	A	C	1.4
10	A	G	G	A	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	A	T	G	+	T	C	C	G	A	G	1.4
11	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	C	A	C	A	A	T	G	+	T	C	C	G	A	G	1.4
12	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	G	G	G	+	T	G	C	G	A	C	1.4
13	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	G	G	G	+	T	C	C	G	A	C	1.4
14	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	G	T	G	+	T	G	C	G	A	C	1.4
15	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	G	T	G	+	T	C	C	G	A	C	1.4
16	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	G	T	G	+	T	C	C	G	A	G	1.4
17	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	A	T	C	+	T	C	C	G	A	G	1.4
18	A	G	G	T	A	C	A	T	C	G	G	T	A	A	G	A	C	C	A	G	G	G	A	A	C	A	G	T	G	+	T	C	T	G	A	C	0.82
19	G	G	G	A	G	C	G	C	G	T	A	G	G	G	-	A	G	A	A	G	A	A	A	G	T	G	G	T	G	+	T	C	T	G	A	C	0.72
20	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	A	T	G	+	T	C	C	G	A	G	0.62
21	G	G	G	A	G	C	G	C	G	T	A	G	A	G	G	A	G	A	A	G	A	A	A	G	T	G	G	T	G	+	T	C	T	G	A	C	0.54
22	A	G	G	T	A	C	A	T	C	G	G	T	A	A	-	A	C	C	A	G	G	G	A	A	C	A	G	T	G	+	T	C	T	G	A	C	0.54

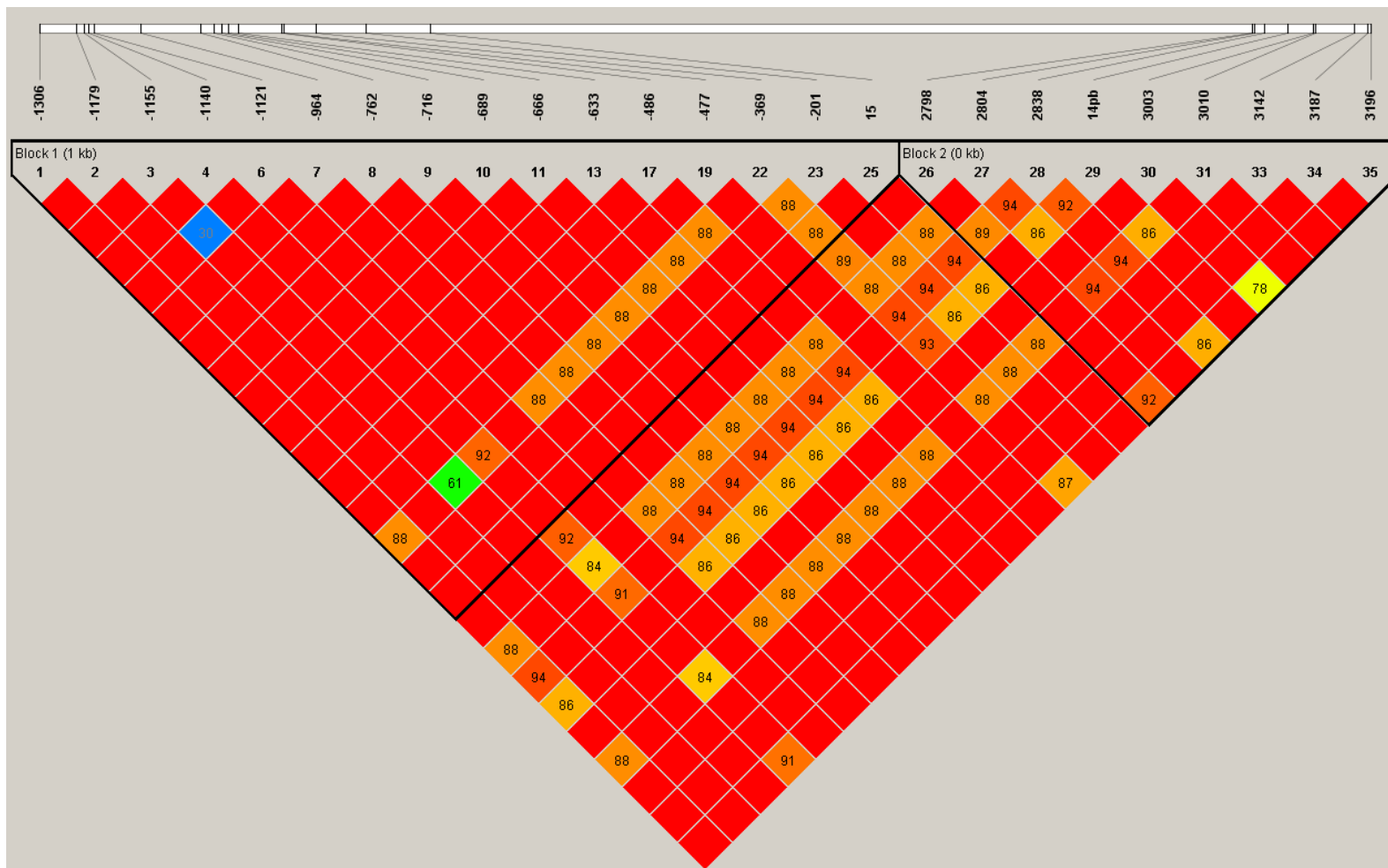


Figura 4. Desequilíbrio de ligação entre os polimorfismos encontrados para a região promotora e 3'UTR do gene *HLA-G* em uma população europeia. Os blocos representam os polimorfismos presentes nas regiões promotora de 3'UTR, respectivamente. O esquema de cores é GOLD HEATMAP, que indicam valores de D' aumentando do azul ao vermelho. Os valores de D são indicados na figura. Foi adotado um filtro para os alelos de menor frequência ($MAF=0.042$)

5.4.1.5. *Investigação da relação entre polimorfismos do gene HLA-G e níveis de expressão da molécula*

As amostras analisadas para a região promotora e *3'UTR* da população europeia citadas acima são amostras de DNA de células de trofoblasto, cujo os níveis de expressão da molécula HLA-G foram medidos por citometria de fluxo (APPS, manuscrito em preparação). O resultado do perfil de expressão da molécula HLA-G nessas células mostram diferenças entre os diferentes doadores.

Para investigar se os polimorfismos encontrados nas regiões promotora e *3'UTR* do gene *HLA-G* estão associados com as diferenças nos padrões de expressão da molécula HLA-G encontrados, análises de correlação entre os genótipos individuais determinados para as posições polimórficas encontradas nessas duas regiões e os níveis de expressão foram realizadas.

Todos os 36 polimorfismos encontrados foram analisados e não foi encontrada nenhuma correlação entre os polimorfismos individuais e variação na expressão da molécula HLA-G nas amostras analisadas.

6. DISCUSSÃO

6.1. CONSIDERAÇÕES SOBRE A ANÁLISE DO GENE *KIR2DL4*

Muito já se conhece sobre a variabilidade haplotípica dos genes *KIR* em populações no que diz respeito à presença e ausência de diferentes loci *KIR*. Entretanto, a caracterização alélica dos mesmos ainda é pouco explorada. Para tentar suprir a falta de informação alélica populacional do gene *KIR2DL4*, populações Ameríndias foram analisadas. Além de caracterizar os alelos para esse locus, o presente trabalho buscou estimar as frequências alélicas através de três metodologias diferentes e testar se há sinal de seleção natural atuando nesse gene em populações Ameríndias.

O seqüenciamento do gene *KIR2DL4* nas populações Ameríndias estudadas no presente trabalho revelou a presença de sete alelos, sendo seis destes compartilhados com populações não-Ameríndias e um alelo novo.

Um importante aspecto levantado no presente trabalho é o desafio de se trabalhar com genes que podem estar ausentes no haplótipo, dificultando as inferências sobre os genótipos dos indivíduos estudados. Três diferentes metodologias foram empregadas para tentar inferir as frequências alélicas a partir dos dados de sequenciamento e verificar se as três metodologias abordadas causam impacto nos testes realizados. Através do teste de Hardy-Weinberg pode-se verificar que a primeira metodologia subestima a frequência de possíveis indivíduos heterozigotos para a ausência do gene em um dos cromossomos e, esse efeito, como já era de esperar, é mais pronunciado em populações pequenas e que apresentam a ausência do gene em homozigose.

O teste de Hardy-Weinberg na segunda metodologia empregada revela o oposto: uma possível superestimativa do alelo nulo em heterozigose, mostrando uma limitação dessa abordagem na inferência dos genótipos.

Já a terceira metodologia permite a estimativa de frequências alélicas quando um alelo nulo está presente assumindo que as populações estão em equilíbrio de Hardy-Weinberg. Embora populações Ameríndias sejam conhecidamente endogâmicas, não se encaixando perfeitamente nos pressupostos do teste, dados obtidos com marcadores microssatélites estudados pela aluna de Doutorado Kelly Nunes foram usados para testar se esses marcadores apontam para desvio de equilíbrio nessas populações. Os dados com esses marcadores mostraram que em média, não há desvios significativos de equilíbrio para os marcadores analisados. Sendo assim, essa metodologia foi considerada a mais adequada para determinação das frequências alélicas quando apenas dados de presença e ausência do gene e seqüenciamento dos mesmos são utilizados. Dessa forma, os dados de frequências alélicas obtidas por essa metodologia foram utilizados para as comparações com dados de frequência em populações não-Ameríndias.

Apesar de haver 37 alelos descritos para esse locus (www.ebi.ac.uk/ipd/kir), as populações Ameríndias apresentaram 7 alelos, sendo apenas 1 não descrito previamente. Os dados de frequências dos alelos mostram que as populações Ameríndias compartilham, em sua maioria, os alelos freqüentes encontrados em populações mundiais (ZHU *et al.*, 2006; YAWATA *et al.*, 2006, BUHLER *et al.*, 2009; GUINAN *et al.*, 2010). O fato de que maior parte das populações investigadas, incluindo as populações Ameríndias aqui estudadas, apresentar um pool pequeno de alelos, pode ser um indicativo que mais importante do que manter a diversidade através de um maior número de alelos é manter alelos específicos em frequências intermediárias em todas as populações.

Comparações inter-populacionais através da estatística *Fst* mostraram que os grupos populacionais de diferentes regiões geográficas apresentam diferenças significativas nas comparações par-a-par, incluindo o grupo Ameríndio que apresentou diferenças significativas inclusive com populações Asiáticas. Esses resultados mostram que essas populações, apesar de compartilharem os mesmos alelos, apresentam diferenças significativas entre elas, provavelmente devido a diferenças nas frequências

desses alelos. Esse resultado pode sugerir que esses alelos estão sendo selecionados em diferentes regiões geográficas, devido a pressões seletivas encontradas nessas regiões.

Comparações inter-populacionais através da estatística *Fst* também foram realizadas para comparação entre as diferentes tribos estudadas no presente trabalho. Os resultados mostraram que as tribos Ameríndias apresentam diferenças significativas entre elas, sendo que as populações Awá-Guajá e Araweté se mostram as mais diferentes entre todos os pares de comparações.

Dados do Instituto Sócio-Ambiental (<http://pib.socioambiental.org/p>) mostram que as populações Awá-Guajá e Araweté sofreram grandes declínios populacionais devido ao contato recente com a sociedade atual. De modo geral, para a população Awá-Guajá, sítios arqueológicos demonstram que a transição para a agricultura trouxe consigo complicações na medida em que caçadores-coletores se assentam de modo mais permanente, ora por mudança de dieta, ora por doenças contagiosas decorrentes da falta de saneamento e da maior concentração de população.

A população Araweté, imediatamente anterior ao contato com a FUNAI, era de pelo menos 200 pessoas. Devido às condições em que o contato se realizou, a mortalidade causada por epidemias e desnutrição levou o grupo ao mínimo de 120 pessoas, em 1977. Em setembro de 1992 a população chegou a 206, alcançando assim o efetivo da época pré-contato. Em 2000 a população foi acometida por um surto de varicela (popularmente conhecida como catapora), mas a queda populacional não foi tão intensa quanto na década de 70. Esses declínios populacionais e posterior expansão podem justificar porque essas duas populações são mais diferenciadas que as demais no que diz respeito a frequência dos alelos encontrados.

Para testar se a diferenciação mais acentuada encontrada para as tribos Awá-Guajá e Araweté é decorrência de gargalos populacionais mais pronunciados nessas tribos, foi aplicado o teste de Garza-Williamson em dados de marcadores microssatélites analisados nessas mesmas populações. Os dados mostram que as populações Awá-Guajá e Araweté apresentam os menores valores dentre as populações analisadas,

demonstrando um efeito de gargalo populacional mais pronunciado nessas populações, possivelmente em decorrência dos fatores históricos de contato com a sociedade atual, mencionados acima.

Apesar de termos julgado o método 3 de estimativa de frequências alélicas mais adequado para as análises a serem realizadas, uma preocupação no nosso trabalho foi verificar se os resultados obtidos para os três métodos de estimativa de frequências alélicas poderiam ser são sensíveis ao método escolhido (uma vez que todos os métodos tratam-se de estimativas). Para tanto foi analisada a correlação entre os valores de *Fst* entre pares de populações obtidos por cada par de métodos. Os resultados mostraram que as correlações entre todos os métodos são altas, indicando que os resultados obtidos para a estimativa de *Fst* são robustos a efeitos associados ao método escolhido, assegurando a confiança nos resultados obtidos.

O gene *KIR2DL4* é um dos genes *KIR* que mostra uma assinatura forte de seleção natural, freqüentemente apresentando maior diversidade do que o esperado por neutralidade (YAWATA *et al.*, 2006; BUHLER *et al.*, 2009; GUINAN *et al.*, 2010) . No presente trabalho, foi investigada se há evidência de seleção natural atuando no gene *KIR2DL4* através das análises dos testes de neutralidade de Ewens-Watterson e *D* de Tajima, em populações indígenas da Amazônia. As populações Arara e Assurini apresentaram um valor de *D* positivo significativo, o que aponta para um possível efeito de seleção balanceadora. Entretanto, não encontramos significância no teste de Ewens-Watterson. O teste de *Ewens-Watterson* é mais adequado que o teste *D* de *Tajima* na captação de sinais seletivos ou demográficos mais recentes (GARRIGAN & HEDRICK, 2003). O fato de termos encontrado significância no teste *D* de *Tajima* e não no teste de *Ewens-Watterson* pode indicar que o sinal seletivo encontrado nas populações aqui estudadas são antigos e, possivelmente, anteriores à entrada no continente americano.

O fato de que nem todas as populações mostraram o mesmo perfil nos testes pode ser um indicativo de que a demografia, em alguns casos, foi mais forte do que a seleção, apagando os sinais da mesma. Para verificar o papel da demografia na diversidade do

gene *KIR2DL4* em populações Ameríndias, comparações com dados de marcadores microssatélites foram realizadas. Os resultados mostram que a diversidade de ambos está correlacionada, mostrando a importância da história demográfica moldando a diversidade dessas populações. Entretanto, correlações não significativas entre os resultados do teste D de Tajima e as medidas de diversidade genômica mostram que a demografia não é suficiente para explicar todos os padrões encontrados, corroborando a ideia de que os sinais seletivos encontrados nas populações Ameríndias são de fato evidência de ação de seleção natural atuando nesse locus. Além disso, estudos de resequenciamento em outros marcadores em populações Ameríndias mostram que os valores de D encontrados no presente trabalho são maiores do que os previamente descritos, sendo que grande parte desses trabalhos mostraram valores de D negativos e não significativos (revisado de RAMALHO *et al.*, 2010), o que fortalece os nossos resultados.

6.2. CONSIDERAÇÕES SOBRE A ANÁLISE DO GENE *KIR3DL1/S1*

O gene *KIR3DL1/S1* é conhecido entre os genes *KIR* como tendo alto polimorfismo alélico e estudos prévios mostram evidência de seleção natural nesse locus em populações humanas (YAWATA *et al.*, 2006; NORMAN *et al.*, 2007; GUINAN *et al.*, 2010).

Assim como para o locus *KIR2DL4*, o presente trabalho propôs a caracterização alélica para o gene *KIR3DL1/S1*, a comparação de metodologias para estimativa das frequências alélicas através de três diferentes abordagens e testar a hipótese de que há sinal de seleção natural atuando nesse gene em populações Ameríndias.

A primeira parte do trabalho consistiu da análise da frequência dos portadores dos alótipos *KIR3DL1* e *KIR3DS1* nas populações indígenas. Foi verificado que em populações Ameríndias, especialmente, para as populações indígenas Amazônicas, o alótipo *KIR3DS1* é mantido em altas frequências, sugerindo que pressões seletivas podem estar atuando nesse locus para favorecer o alótipo ativador.

A segunda parte do trabalho consistiu na caracterização alélica do alótipo *KIR3DL1* nas populações Ameríndias. Assim como para o locus *KIR2DL4*, as populações Ameríndias apresentaram 7 alelos. Os dados de frequências dos alelos mostram que as populações Ameríndias compartilham os mesmos alelos mais freqüentes em diferentes populações mundiais descritos em NORMAN et al. (2007). Além disso, os alelos encontrados aqui são um subconjunto dos alelos encontrados em populações Asiáticas, condizente com a história evolutiva das populações Nativo-Americanas.

Como parte deste estudo, comparamos metodologias para determinação das frequências alélicas desse locus. As três diferentes metodologias empregadas mostraram resultados semelhantes aos encontrados nas análises do gene *KIR2DL4*: 1) A primeira metodologia subestima a frequência de possíveis indivíduos heterozigotos para a ausência do gene em um dos cromossomos, levando a um excesso de homozigotos em algumas populações; 2) A segunda metodologia empregada superestima a presença do alelo nulo em heterozigose; 3) A terceira metodologia assume que as populações estão em equilíbrio de Hardy-Weinberg para a estimativa das frequências alélicas. Viés nas estimativas pode ocorrer caso as populações estudadas não estejam em equilíbrio.

Utilizando o mesmo raciocínio mencionado na discussão do gene *KIR2DL4*, através da comparação com dados de marcadores microssatélites, a terceira metodologia empregada foi julgada a mais adequada para determinação das frequências alélicas.

As comparações inter-populacionais através da estatística *Fst* mostraram que as populações Awá-Guajá e Araweté apresentam uma diferenciação inter-populacional mais acentuada do que demais. Como mencionado para o locus *KIR2DL4*, a diferenciação populacional encontrada para as tribos Awá-Guajá e Araweté podem indicar que os recentes declínios populacionais sofridos por essas duas tribos tiveram conseqüências mais pronunciadas na distribuição das frequências alélicas do que se observa nas outras populações indígenas. As populações Surui e Ticuna também são um pouco mais diferenciadas das demais e esses resultados apontam para uma diferenciação entre populações Amazônicas mais a oeste do continente quando comparadas com as

populações mais a leste. Esses resultados corroboram estudos prévios que mostram que populações do oeste da América do Sul apresentam maior diversidade do que as populações do leste do continente (TARAZONA-SANTOS *et al.*, 2001). Comparações com marcadores microssatélites nessas populações, como foi realizado para os dados das análises do gene *KIR2DL4*, poderão fortalecer essa hipótese.

Para as análises de diferenciação inter-populacional, assim como para os resultados obtidos para o gene *KIR2DL4*, as três metodologias de estimativa de frequências alélicas trazem resultados bastante similares, mostrando que os dados de comparação inter-individual discutidos aqui são bastante confiáveis independentemente do método utilizado.

A última parte do trabalho investigou se há evidências de atuação de seleção natural no gene *KIR3DL1/S1*. As análises de desvio de neutralidade para a ausência e presença dos alótipos *KIR3DL1* e *KIR3DS1* mostraram desvios significativos de neutralidade para as tribos Assurini, Xikrin e Karitiana, indicando que a seleção natural vem mantendo os dois alótipos em frequências intermediárias nessas populações. O fato de ter se encontrado desvio em direção a um excesso na taxa de heterozigose em quase todas as tribos analisadas (exceto para a tribo Araweté), embora não significativo, reforça a idéia de seleção natural atuando nessas populações.

A aplicação do teste D de Tajima revelou desvio de neutralidade significativo para os Ticuna, apresentando um valor de D positivo. Esse resultado pode indicar que além da manutenção dos alótipos desse gene, a seleção natural também está atuando para manter diferentes linhagens do alótipo *KIR3DL1* nessa população.

NORMAN *et al.* (2007) mostraram que três diferentes linhagens, duas linhagens do alótipo inibidor *KIR3DL1* (*015 e *005) e uma linhagem do alótipo *KIR3DS1*, evoluíram sob ação de seleção natural, levando as três linhagens a adquirirem especificidades de ligação diferentes. Estudos funcionais corroboram as conclusões desses autores mostrando inibição ou citotoxicidade diferencial quando diferentes alótipos de *KIR3DL1* são desafiados com seu respectivo ligante YAWATA *et al.*, 2006; THANANCHAI *et al.*, 2007).

No presente trabalho, nós verificamos que a maioria das populações tendem a manter um balanço entre duas principais linhagens descritas por NORMAN *et al.* (2007) representadas aqui pelos alótipos de *KIR3DL1* e *KIR3DS1*. Já o valor de D de Tajima positivo na tribo Ticuna mostra o papel da seleção natural também mantendo diferentes linhagens para o alótipo *KIR3DL1*.

O fato de no presente trabalho ter se encontrado assinatura de seleção natural em populações Ameríndias corrobora a idéia de que a manutenção desse padrão seja ancestral e a importância da manutenção do balanço entre os perfis ativadores e inibidores, além da manutenção de diferentes alelos do alótipo *KIR3DL1*. Entretanto, o fato de não encontrarmos o mesmo padrão em todas as populações pode evidenciar o papel da demografia no padrão de variação encontrado.

Outra possível explicação para esse cenário é que para populações indígenas, principalmente as da região Amazônica, há um desvio de regime seletivo nesse gene em direção a seleção direcional, verificado pela alta frequência de indivíduos portadores do alótipo *KIR3DS1* em populações. Nesse segundo caso, o perfil mais ativador estaria sendo favorecido. Perfis KIR ativadores são importantes em contextos de infecções (BASHIROVA *et al.*, 2006). É sabido que populações Ameríndias sofreram com epidemias trazidas por populações européias no período de colonização e algumas também sofreram com o contato mais recente. Dessa forma, perfis ativadores seriam importantes e selecionados positivamente.

Devido a essas características, uma hipótese provável é que as epidemias que atingiram o continente americano tenham deixado impacto seletivo, causando um aumento de frequência de perfis KIR ativadores. Entretanto, a medida que os efeitos dessa pressão seletiva foram sendo minimizados, um balanço entre os perfis ativadores e inibidores vem sendo restaurado.

6.3. CONSIDERAÇÕES SOBRE AS ANÁLISES DO GENE HLA-G

Apesar de ter sido primeiramente descrito em células de placenta e por muitos anos ter se acreditado ter como única função a proteção do feto na gravidez, os últimos anos de pesquisa com o gene *HLA-G* tem revelado que ele apresenta uma importante função na regulação do sistema imune (CAROSELLA *et al.*, 2008).

O fato de *HLA-G* apresentar uma região codificadora bem menos polimórfica do que os genes *HLA* clássicos e evidências de que sua expressão deve estar sob intenso controle e regulação, fez com que grande atenção fosse dada a estudos com as regiões reguladoras desse gene.

A região promotora do gene *HLA-G* é única entre os genes *HLA*, na qual vários elementos cis-reguladores são deletados ou modificados (SOLIER *et al.*, 2001) Análises dessa região em populações mostraram que muitos dos polimorfismos descritos coincidem com ou estão perto de sítios de ligação a fatores de transcrição, indicando que esses podem influenciar a expressão de transcritos de *HLA-G*. Essa região também tem alta diversidade nucleotídica e evidência de ação de seleção balaceadora, mantendo na espécie humana duas linhagens de promotores divergentes que podem ter atividades promotoras diferentes (TAN *et al.*, 2005). A região *3'UTR* do gene *HLA-G* também apresenta polimorfismos associados com diferenças na expressão da molécula [ROUSSEAU *et al.* (2003); VEIT *et al.*, (2009)], alta diversidade nucleotídica (sendo esta mais alta do que o estimado para os exons 2, 3 e 4 juntos) (ALVAREZ *et al.*, 2009) e hipóteses sobre a ação de seleção balaceadora atuando também em um polimorfismo contido nessa região (MENDES-JUNIOR *et al.*, 2007).

Diante das características marcantes das regiões reguladoras do gene *HLA-G*, o presente trabalho teve o intuito de fazer uma investigação dessas regiões em populações Ameríndias a fim de tentar contribuir para um melhor entendimento da história evolutiva dessas regiões. Os resultados aqui encontrados nos levaram a um estudo mais amplo dessas regiões em uma população européia, culminando com um estudo de associação

entre os polimorfismos descritos para essas duas regiões e diferenças nos níveis de expressão da molécula.

A primeira parte do estudo caracterizou a variabilidade da região promotora desse gene em populações Ameríndias. Os resultados de seqüenciamento mostraram a presença de 25 polimorfismos nessa região, muitos dos quais localizados perto ou dentro de sítios de ligação de fatores de transcrição, em concordância ao já encontrado para outras populações mundiais analisadas.

Os resultados das análises haplotípicas mostraram que os haplótipos mais freqüentes encontrados nas populações Ameríndias também são compartilhados com outras populações mundiais estudadas. Além disso, as análises dos haplótipos em populações Ameríndias mostram que eles estão enquadrados nas duas linhagens previamente descritas por TAN *et al.* (2005), reiterando a importância dessas duas linhagens alélicas para o locus *HLA-G*.

Para verificar se em populações Ameríndias também são encontrados sinais de seleção balanceadora na região promotora, nós aplicamos o teste D de Tajima. Os resultados apontaram para um D positivo, indicando a influência de seleção balanceadora no padrão de variação encontrado também em populações Nativo-Americanas. Além disso, observamos altos níveis de variação nucleotídica, cerca de 10 vezes mais elevado que a média para o genoma humano.

O fato de termos encontrado retenção de um montante substancial de diversidade alélica, mesmo em populações que sofreram perda de heterozigose, como é o caso das populações Nativo-Americanas, é compatível com a hipótese de que a seleção balanceadora vem mantendo a variabilidade dessa região gênica também nas populações Nativo-Americanas. Entretanto, não podemos descartar que os resultados aqui encontrados sejam um sinal antigo de seleção que ainda não foi apagado nessas populações.

No presente estudo também investigamos a variabilidade populacional do polimorfismo de 14pb em populações Nativo-Americanas. Um estudo prévio feito em

populações indígenas da Amazônia sugere uma leve tendência de ação de seleção balanceadora atuando no polimorfismo de *14pb* localizado na região 3'UTR do gene *HLA-G* nessas populações (MENDES-JUNIOR *et al.*, 2007). Entretanto, os autores sugerem a necessidade de que mais populações fossem analisadas para tentar elucidar os mecanismos envolvidos na manutenção da variabilidade desse polimorfismo.

Os resultados obtidos neste trabalho (que contempla populações indígenas de todo o continente Americano) mostram que todas as populações Ameríndias apresentam ambos os alelos de inserção e deleção. Para testar se há um desvio de neutralidade, o teste de Ewens-Watterson foi aplicado. Os resultados, apesar de não significativos, apontam para um excesso de desvios na direção de taxas de heterozigose maiores do que esperada sob o modelo de neutralidade e equilíbrio populacional

O conjunto de populações apresenta um significativo viés na direção de excesso de heterozigose, indicando uma possível ação de seleção balanceadora nesses polimorfismos em populações indígenas de todo o continente Americano.

Uma vez que variações em regiões regulatórias são alvo de seleção natural em vários genes do sistema imune, investigamos a variação na região 3'UTR do gene *HLA-G* e se a mesma está sob ação de seleção natural através do seqüenciamento de indivíduos de uma população urbana européia. Para tanto, o teste D de Tajima foi aplicado e mostrou um valor de D positivo significativo.

Dados de análise de seqüenciamento nessa região em populações européias apontam para uma alta diversidade nucleotídica (ALVAREZ *et al.*, 2009), embora nesse estudo, testes de neutralidade não tenham sido aplicados, deixando em aberto a questão de uma possível ação de seleção natural mantendo essa alta diversidade nucleotídica. Análises de desvio de neutralidade dessa região em uma população urbana brasileira não revelaram resultados significativos nos testes de Ewens-Watterson e D de Tajima. Entretanto, os autores chamam a atenção para o fato de que esses resultados podem estar mascarados pela história da população (CASTELLI *et al.*, 2010). No presente estudo, nós demonstramos evidência de seleção natural atuando na região 3'UTR do gene *HLA-G*,

indicando que assim como a região promotora, a região *3'UTR* de HLA-G também mostra evidência de seleção balanceadora influenciando o padrão de variação encontrado.

Além da caracterização da região *3'UTR*, a região promotora também foi seqüenciada nessas amostras com o intuito de ver a relação entre a variação encontrada na região promotora e *3'UTR*. Os haplótipos conjuntos foram inferidos e mostraram a presença de dois haplótipos mais freqüentes nessa população com uma combinação oposta de alelos na maioria das posições analisadas, mostrando uma grande divergência entre ambos, que pode estar relacionado com atividades reguladoras distintas.

O fato de termos encontrados dois haplótipos mais freqüentes tão diferentes aponta para o fato de que não só a região promotora, mas, toda a região reguladora do gene *HLA-G*, mantém duas linhagens divergentes, possivelmente relacionadas à manutenção entre alta e baixa expressão da molécula. Esses resultados também apontam para a idéia de que mais do que os SNPs individuais, a associação entre eles é determinante na regulação da expressão. Os dados de análise de desequilíbrio de ligação entre os polimorfismos, que mostram um alto desequilíbrio de ligação entre os polimorfismos, ajudam a reforçar essa idéia.

No presente trabalho nós também analisamos possíveis associações entre os polimorfismos individuais encontrados e variações nos níveis de expressão da molécula HLA-G. Análises dos SNPs individuais não mostraram associação dos mesmos com os diferentes níveis de expressão. Esse resultado também reforça a idéia de que muito mais do que SNPs individuais, a combinação desses SNPs em diferentes haplótipos é responsável pela variação dos níveis de expressão encontrados.

7. CONSIDERAÇÕES FINAIS

Estudos recentes têm mostrado que muitos genes na espécie humana foram alvo de seleção natural (AKEY *et al* 2002; VOIGHT *et al*, 2006; MYLES *et al*, 2007). A origem deste grande interesse é aprender mais sobre os processos evolutivos que moldaram a variabilidade genética em populações humanas e em outros organismos, além de fornecer informações relevantes a respeito da importância funcional das regiões analisadas (NIELSEN, 2005a).

Dentre as categorias de genes que mostram grande influência da ação de seleção natural, os genes do sistema imune aparecem em destaque (ANDRÉS *et al.*, 2009). Grande interesse também tem sido voltado para estudos de seleção natural em regiões reguladoras, as quais têm demonstrado apresentar um importante papel na evolução adaptativa (WRAY *et al.*, 2007).

Apesar do crescente conhecimento envolvendo estudos de seleção natural em populações humanas, os mesmos tratam a variação genética humana numa escala global contemplando, na maioria das vezes, populações em escalas geográficas maiores, o que permite poucas inferências sobre a história evolutiva mais recente, como é o caso da diferenciação das populações Ameríndias. Dessa forma, estudos com populações Ameríndias podem contribuir significativamente para ampliar o conhecimento de como a seleção natural vêm moldando os padrões de variação em populações humanas. Além disso, estudos de seleção natural nessas populações são interessantes devido à sua história demográfica e a ocupação de um novo ambiente que pode ter trazido novas pressões seletivas, tais como contato com novos patógenos, motivando o estudo de genes de sistema imune nessas populações.

Dessa forma, o presente trabalho investigou questões relativas ao papel da seleção natural na história evolutiva de genes de função imunológica em populações Ameríndias, tais como: 1) Se genes que apresentam evidências de seleção natural globalmente também apresentam o mesmo padrão em Ameríndios; 2) Se o padrão de variação encontrado traz informações sobre a forma como a seleção e processos demográficos interagem; 3) Se genes do sistema imune aqui estudados apresentam um padrão de variação distinto da média genômica.

A primeira parte do trabalho consistiu do estudo de dois genes *KIR*. Para o gene *KIR2DL4* foram encontrados sete alelos em populações Ameríndias, sendo um deles descrito pela primeira vez no presente trabalho. Os alelos encontrados em populações Ameríndias também são compartilhados por populações não-Ameríndias. Apesar de compartilharem os mesmos alelos, comparações de diferenciação inter-populacional mostraram que os grupos populacionais de diferentes regiões geográficas apresentam diferenças significativas nas frequências dos alelos encontrados, sugerindo que diferentes alelos podem estar sendo favorecidos em diferentes regiões geográficas.

Evidência de seleção natural atuando no gene *KIR2DL4* foi de fato observada em populações indígenas da Amazônia, nas quais as populações Arara e Assurini apresentaram um valor de D positivo significativo, compatível com efeito de seleção balanceadora.

Para o gene *KIR3DL1/S1*, as análises de frequência dos portadores dos alótipos *KIR3DL1* e *KIR3DS1* mostraram que para populações Ameríndias, especialmente, para as populações indígenas Amazônicas, o alótipo *KIR3DS1* é mantido em altas frequências, sugerindo um efeito seletivo favorecendo o alótipo ativador nessas populações. Entretanto, resultados prévios com marcadores microssatélites mostram que populações Ameríndias apresentam alta diferenciação em relação as demais populações do mundo, resultado coerente com o gargalo populacional associado à ocupação do continente (WANG *et al.*, 2007). Dessa forma, o padrão observado também é compatível com essa recente história demográfica.

A caracterização alélica do alótipo *KIR3DL1* nas populações Ameríndias revelou a presença de 7 alelos. Os dados de frequências dos alelos mostram que os alelos encontrados aqui são um subconjunto dos alelos encontrados em populações Asiáticas, condizente com a história evolutiva das populações Ameríndias.

Evidências de atuação de seleção natural no gene *KIR3DL1/S1* foram encontradas em populações e mostraram que a seleção atua em dois níveis diferentes de variação: (1) as análises para a ausência e presença dos alótipos *KIR3DL1* e *KIR3DS1* mostraram desvios significativos de neutralidade para as tribos Assurini, Xikrin e Karitiana, consistente com a seleção natural mantendo os dois alótipos em frequências intermediárias nessas populações; (2) através da manutenção de diferentes linhagens do alótipo *KIR3DL1*, verificada na tribo Ticuna na qual encontramos desvio de neutralidade significativo com um valor de D positivo.

NORMAN *et al.* (2007) mostraram que três diferentes linhagens, duas linhagens do alótipo inibidor *KIR3DL1* (*015 e *005) e uma linhagem do alótipo *KIR3DS1*, evoluíram sob ação de seleção natural, mostrando que a seleção balanceadora no locus *KIR3DL1/S1* é antiga. Dessa forma, as assinaturas de seleção natural encontradas em populações Ameríndias refletem o padrão global encontrado para esse gene. Entretanto, a ausência de evidência de seleção natural em algumas populações estudadas pode ser consequência de dois processos: 1) A história demográfica recente das populações resultou em deriva genética suficientemente intensa para causar grande diferenciação genética, dessa forma, apagando o sinal de seleção papel da demografia no padrão de variação encontrado; alternativamente, o próprio regime de seleção pode ter sido alterado, com um desvio de regime seletivo nesse gene em direção a seleção direcional, favorecendo o o alótipo *KIR3DL1*.

Uma forma de compreender melhor os padrões de variação encontrados para os genes *KIR2DL4* e *KIR3DL1/S1* é analisar esses dois genes conjuntamente e contextualizá-los em haplótipos através das análises de presença e ausência dos demais genes do complexo. Apesar de termos analisados esses dois genes individualmente, eles estão

localizados próximos fisicamente e análises dos haplótipos constituídos pelos alelos desses dois genes podem fornecer informações adicionais sobre os efeitos seletivos e demográficos atuando na variação desses genes. Análises desses genes com os seus respectivos ligantes também podem contribuir para um melhor entendimento de como os processos evolutivos estão atuando no padrões de variação encontrados para o gene *KIR3DL1/3DS1*. Evidências de atuação de seleção favorecendo determinadas combinações de alelos dos genes *HLA-B* e *KIR3DL1/3DS1*, mostrando o papel da co-evolução entre receptor-ligante tem sido mostradas (SINGLE et al., 2007; NORMAN et al., 2007).

Regiões reguladoras sob ação de seleção natural também foram analisadas no presente trabalho. As regiões promotora e *3'UTR* do gene *HLA-G* foram investigadas para ação de seleção natural. Evidência de seleção natural atuando na região promotora desse gene já havia sido documentada em populações Euro-Americanas, Afro-Americanas e Chinesa e no presente trabalho foi mostrada evidência de ação de seleção natural atuando na região promotora do gene *HLA-G* em populações Ameríndias. Entretanto, não é possível diferenciar se esse sinal ocorreu no passado e persiste nas populações atuais, mesmo na ausência de seleção, ou se a seleção natural atua na região promotora em populações Ameríndias atualmente.

A região *3'UTR* do gene *HLA-G* também foi investigada através da análise de um polimorfismo específico (polimorfismo de *14pb*) em populações Ameríndias e também através do sequenciamento da região *3'UTR* em uma população européia. Em ambos os casos, evidência de seleção natural foi encontrada para a região *3'UTR*.

Além das análises das regiões promotora e *3'UTR* individualmente, haplótipos formados pela combinação dos polimorfismos nas duas regiões foram inferidos. Os resultados das análises conjuntas mostraram a presença de dois haplótipos mais frequentes que apresentam grande divergência na combinação de alelos na maioria das posições analisadas. O fato de termos encontrados dois haplótipos tão divergentes e que correspondem aos dois haplotipos mais frequentes aponta para o fato de que não só a

região promotora, mas, também, a região *3'UTR* do gene *HLA-G* mantém duas linhagens divergentes, possivelmente relacionadas com a manutenção entre alta e baixa expressão da molécula. A análise de desequilíbrio de ligação entre os polimorfismos mostrou uma alta associação entre os sítios dessas duas regiões do gene, reforçando essa idéia.

Por último, associações entre os polimorfismos individuais encontrados e variações nos níveis de expressão da molécula *HLA-G* foram analisadas. Análises dos SNPs individuais não mostraram associação desses com as diferenças nos níveis de expressão, o que pode indicar que a regulação entre alta e baixa expressão dessa molécula está relacionada com o efeito combinado dos polimorfismos ao invés de polimorfismos individuais.

Juntos esses resultados mostram que as regiões promotora e *3'UTR* apresentam um padrão de variação indicativo de seleção natural balanceadora e que as duas regiões estão altamente associadas, possivelmente para manter um balanço entre atividades reguladoras diferentes.

O presente trabalho teve como importante desafio o fato de trabalhar com genes cuja metodologia de genotipagem não permite a inferência de possíveis haplótipos com ausência do gene, fato muito comum nos genes *KIR* para os quais há variação de presença e ausência de genes específicos. Três metodologias para estimar as frequências alélicas desses genes foram testadas. A metodologia baseada em máxima-verossimilhança foi julgada a mais adequada, por não introduzir os vieses extremos de excesso e deficiência de homozigose, observados nos métodos 1 e 2. A plausibilidade biológica de que não há tais desvios foi apoiada na observação de dados de microssatélites genotipados para as mesmas populações. É interessante notar que as análises de comparação inter-populacional e testes de neutralidade não foram sensíveis ao método escolhido, mostrando que apesar da limitação metodológica, as inferências feitas no presente trabalho são robustas.

No presente trabalho nós também tentamos lidar com o desafio de se trabalhar com populações cuja história demográfica é complexa e inclui eventos recentes de gargalos

populacionais. A abordagem usada foi comparar os dados para os genes aqui analisados com dados para marcadores neutros (que são informativos sobre como a demografia moldou a variação nas populações). Essas comparações permitiram algumas conclusões, tais como: (1) Em populações Ameríndias, desvios de equilíbrio de Hardy-Weinberg são, em média, pouco frequentes, permitindo a confiança na abordagem de estimativa de frequências alélicas que assume que populações estão em equilíbrio, (2) Efeitos de gargalo recente foram mais pronunciados nas populações Awá-Guajá e Araweté, mostrando que as altas diferenças entre essas e as demais, nas análises dos genes *KIR*, são provavelmente consequência de efeitos demográficos, (3) A diversidade do gene *KIR2DL4* está correlacionada com a diversidade dos marcadores neutros, mostrando que a demografia tem um papel importantíssimo na modulação da diversidade dessas populações; (4) Não há uma correlação clara entre diversidade dos marcadores neutros com os resultados dos testes de neutralidade para o gene *KIR2DL4*, mostrando que a demografia não é suficiente para explicar todos os padrões encontrados, corroborando a idéia de que os sinais seletivos encontrados nas populações Ameríndias são de fato evidencia de ação de seleção natural atuando nesse locus.

8. RESUMO

Sinais de pressão seletiva têm sido descritos em vários genes. Dentre esses, os genes do sistema imune parecem ser um importante alvo de seleção natural. Diversos estudos têm demonstrado a atuação de pressões seletivas também modulando o padrão de diversidade de regiões reguladoras. Populações Ameríndias possuem uma história evolutiva relativamente recente, com processos demográficos marcantes e novos desafios ambientais trazendo diferentes pressões seletivas. Dessa forma, elas oferecem ótimas oportunidades para estudos de seleção natural. Com o intuito de ampliar os conhecimentos sobre como a seleção natural vem atuando nas populações humanas, o presente estudo propôs uma análise populacional de três *loci* de função imunológica, caracterizados previamente como alvos de seleção, em populações Ameríndias: dois pertencentes ao complexo *KIR* (*KIR2DL4* e *KIR3DL1/S1*) e regiões regulatórias do gene *HLA-G*. Evidências de ação de seleção balanceadora em populações Ameríndias foram encontradas para os genes *KIR2DL4* e *KIR3DL1/S1*. Para *KIR3DL1/S1* foram observadas evidências de seleção natural atuando em dois níveis diferentes de variação: (1) para a ausência e presença dos alótipos *KIR3DL1* e *KIR3DS1* e (2) para a manutenção de diferentes linhagens do alótipo *KIR3DL1*. Comparações da diversidade do gene *KIR2DL4* com marcadores autossômicos mostram que os processos demográficos têm um importante papel na modulação da variabilidade genética observada nessas populações, mas não explicam todos os padrões de variação encontrados. Os resultados de análises das regiões promotora e 3'UTR do gene *HLA-G* mostraram evidências de seleção balanceadora atuando na região promotora em populações Ameríndias e na região 3'UTR em uma população europeia. Além disso, análises dessas duas regiões conjuntas mostraram que essas estão altamente associadas e dois haplótipos divergentes em alta frequência foram encontrados, os quais podem estar relacionados com diferentes atividades reguladoras. Análises dos SNPs encontrados nas regiões reguladoras do *HLA-G* não mostraram associação com diferenças nos níveis de expressão da molécula. Esse resultado sugere que mais do que SNPs individuais, a combinação desses SNPs em diferentes haplótipos é responsável pela variação dos níveis de expressão da molécula *HLA-G*.

9. ABSTRACT

The diversity of several genes has been explained by the action of natural selection. Among them, the immune system genes are highlighted. In addition, several studies have demonstrated the action of selective pressure modulating the diversity of regulatory regions. Amerindian populations have a relatively recent evolutionary history, with remarkable demographic processes and new environmental challenges with different selection pressures. Therefore, those populations are excellent candidates for studies of natural selection. In order to broaden the knowledge of how natural selection has been acting in human populations, this study proposed a population analysis of three loci, previously characterized as targets of selection in Amerindian populations: two from the *KIR* complex (*KIR2DL4* and *KIR3DL1 / S1*) and regulatory regions of *HLA-G*. Evidence of action of balancing selection in Amerindian populations was found for *KIR2DL4* and *KIR3DL1/S1* genes. For *KIR3DL1/S1* gene were observed evidences of natural selection acting on two different levels of variation: (1) absence and presence of *KIR3DL1* and *KIR3DS1* allotypes and (2) maintenance of different strains of *KIR3DL1* allotype. Correlations between the diversity of *KIR2DL4* and autosomal markers show that the demographic processes have an important role in modulating the genetic variation observed in these populations, but do not explain all the variation patterns found. The results of analysis of promoter regions and 3'UTR of the *HLA-G* gene revealed evidence of balancing selection acting at the promoter region in Amerindian populations and at the 3'UTR region in a European population. Furthermore, the analysis of these two regions together showed that t both are highly related and divergent haplotypes at high frequency were found, which may be related to different regulatory activities to maintain a balance of high and low expression of the gene. Analyses of SNPs found in the regulatory regions of *HLA-G* were not associated with differences in expression levels of the molecule. Together, these results suggest that the effects of individual SNPs in the regulatory region of the *HLA-G* gene are not independent and balancing selection has maintained these lineages.

10. REFERÊNCIAS BIBLIOGRÁFICAS

AKEY JM, ZHANG G, ZHANG K, JIN L & SHRIVER MD. Interrogating a high-density SNP map for signatures of natural selection. **Genome Research** 12: 1805-1814, 2002.

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 Biology** 2: 1591-1599, 2004.

Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. GONZALEZ-GALARZA FF, CHRISTMAS S, MIDDLETON D & JONES AR. **Nucleic Acid Research** 39: 913-919, 2011. <http://www.allelefreqencies.net>

ALVAREZ M, PIEDADE J, BALSEIRO S, RIBAS G & REGATEIRO F. HLA-G 3'-UTR SNP and 14bp deletion polymorphisms in Portuguese and Guinea-Bissau populations. **International Journal of Immunogenetics** 36: 361-366, 2009.

ANDRES AM, HUBISZ MJ, INDAP A, TORGERSON DG, DEGENHARDT JD, BOYKO AR, GUTENKUNST RN, WHITE TJ, GREEN ED, BUSTAMANTE CD, CLARK AG & NIELSEN R. Targets of balancing selection in the human genome. **Molecular Biology Evolution** 26: 2755-2764, 2009.

BAMSHAD M & WOODING SP. Signatures of natural selection in the human genome. **Nature Review Genetics** 4: 99 –111, 2003.

BARRETT JC, FRY B, MALLER J & DALY MJ. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**. 15;21(2):263-5, 2005.

BASHIROVA AA, MARTIN MP, McVICAR DW & CARRINGTON M. The killer immunoglobulin-like receptor gene cluster: Tuning the genome defense. **Annual Reviews of Genomics and Human Genetics**. 7: 277-300, 2006.

BELICH MP, MADRIGAL JA, HILDEBRAND WH, ZEMMOUR J, WILLIAMS RC, LUZ R & PETZL-ERLER. Unusual HLA-B alleles in two tribes of Brazilian Indians. **Nature** 28;357(6376):326-329, 1992.

BUHLER S, CRISTOFARO JD, FRASSATI C, BASIRE A, GALICHER V, CHIARONI J & PICARD C. High levels of molecular polymorphism at the KIR2DL4 locus in French and Congolese populations: Impact for antropology and clinical studies. **Human Immunology** 70: 953-959, 2009.

CAROSELLA E D, FAVIER B, ROUS-FREISS N, MOREAU P & LEMAOULT. Beyond the increasing complexity of the immunomodulatory HLA-G molecule. **Blood Journal** 19: 2008.

CAROSELLA ED, MOREAU P, LEMAOULT J & ROUAS-FREISS N. HLA-G: from biology to clinical benefits. **Trends in Immunology** 29(3):125-32, 2008a.

CAROSELLA ED & HORUZSKO A . HLA-G and cancer. **Seminars in Cancer Biology** 17: 411-412, 2007b.

CASTELLI EC, MENDES-JUNIOR CT, DEGHAIDE NHS, ALBUQUERQUE RS, MUNIZ YCN, SIMÕES RT, CAROSELLA ED, MOREAU P & DONADI EA. The genetic structure of

3, untranslated region of the HLA-G gene: polymorphisms and haplotypes. **Genes and Immunity** 11: 134-141, 2010.

CHEENT K & KHAKOO SI. Natural killer cells: integrating diversity with function. **Immunology** 126(4):449-57, 2009.

DEMPSTER, A. P., LAIRD, N. M., & ROBIN, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm (with discussion). **Journal of Royal Statistical Society B** 39: 1–38, 1977.

DONADI EA, CASTELLI EC, ARNAIZ-VILLENA A, ROGER M, REY D, MOREAU P. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. **Cell Mol Life Sci.** 68(3):369-95, 2011.

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.** 10: 564-567, 2010.

EXCOFFIER L, SMOUSE PE & QUATTRO JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. **Genetics** 131:.479-491,1992.

FAGUNDES NJ, KANITZ R, ECKERT R, VALLS AC, BOGO MR, SALZANO FM, SMITH DG, SILVA WA JR, ZAGO MA, RIBEIRO-DOS-SANTOS AK, SANTOS SE, PETZL-ERLER ML & BONATTO SL. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. **American Journal Human Genetics** 82: 583-92, 2008.

EWENS, W. The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3:87-112, 1972.

EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:86-94, 1998.

EWING B, HILLIER L, WENDL M, GREEN P: Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* 8:175-185, 1998.

FAVIER B, LEMAOULT J & CAROSELLA ED. Functions of HLA-G in the immune system. *Tissue Antigens* 69: 150-152, 2007.

FERRER-ADMETLLA A, BOSCH E, SIKORA M, MARQUÈS-BONET T, RAMÍREZ-SORIANO A, MUNTASELL A, NAVARRO A, LAZARUS R, CALAFELL F, BERTRANPETIT J & CASALS F. Balancing selection is the main force shaping the evolution of innate immunity genes. *Journal of Immunology* 181: 1315-1322, 2008.

GARDINER CM, GUETHLEIN LA, SHILLING HG, PANDO M, CARR WH, RAJALINGAM R, VILCHES C, PARHAM P. Different NK cell surface phenotypes defined by the DX9 antibody are due to KIR3DL1 gene polymorphism. *Journal of Immunology* 1;166(5):2992-3001, 2001.

GARRIGAN D & HEDRICK P. Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* 57 (8): 1707-22, 2003.

GARZA JC & WILLIAMSON EG. Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* 10:305-318, 2001.

GENDZEKHADZE K, NORMAN PJ, ABI-RACHED L, GRAEF T, MOESTA AK, LAYRISSE Z, PARHAM P. Co-evolution of KIR2DL3 with HLA-C in a human population retaining minimal essential diversity of KIR and HLA class I ligands. **Proceedings of the National Academy of Sciences**. 23;106(44):18692-7, 2009.

GOODRIDGE JD, WITT CS, CHRISTIANSEN FT & WARREN HS. KIR2DL4 (CD158d) genotype influences expression and function in NK cells. **Journal of Immunology**, 15: 171(4) 1768-1774, 2003.

GOODRIDGE JP, LATHBURY LJ, STEINER NK, SHULSE CN, PULLIKOTIL P, SEIDAH NG, HURLEY CK, CHRISTIANSEN FT & WITT CS. Three common alleles of KIR2DL4 (CD158d) encode constitutively expressed, inducible and secreted receptors in NK cells. **European Journal Immunology** 37:199-211, 2007.

GORDON D, ABAJIAN C, GREEN P: Consed: a graphical tool for sequence finishing. **Genome Research** 8: 195-202, 1998.

GUINAN KJ, CUNNINGHAM RT, MEENAGH A, GONZALEZ A, DRING MM, MCGUINNESS BW, MIDDLETON D & GARDINER CM. Signatures of natural selection and coevolution between killer cell immunoglobulin-like receptors (KIR) and HLA class I genes. **Genes and Immunity** 11: 467-478, 2010.

GUO S & THOMPSON E. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. **Biometrics**; 48:361-372, 1992.

HAHN MW. Detecting natural selection on cis-regulatory DNA **Genetica** 129: 7-18, 2006.

HARRIS EE & MEYER D. The molecular signature of selection underlying human adaptations. **American Journal Phys Anthropology** 131:89-130, 2006.

HARRISON GA, HUMPHREY KE, JAKOBSEN IB, COOPER DW. A 14 bp deletion polymorphism in the HLA-G gene. **Hum Mol Genet.** 2(12):2200, 1993.

HAYGOOD R, BABBITT CC, FEDRIGO O, WRAY GA. Contrasts between adaptive coding and noncoding changes during human evolution. **Proc Natl Acad Sci U S A** 107(17):7853-7, 2010.

HUGHES AL, PACKER B, WELCH R, CHANOCK SJ & YEAGER M. High level of functional polymorphism indicates a unique role of natural selection at human immune system loci. **Immunogenetics**: 57:821-827, 2005.

HUNT JS, PETROFF MG, MCINTIRE RH & OBER C. HLA-G and immune tolerance in pregnancy. **FASEB J.** 19(7):681-693, 2005.

HVIID TV, HYLENIUS S, HOEGH AM, KRUSE C & CHRISTIANSEN OB . HLA-G polymorphisms in couples with recurrent spontaneous abortions. **Tissue Antigens** 60:122-132, 2002.

HVIID TVF, HYLENIUS S, LINDHARD A & CHRISTIANSEN OB. Association between human leukocyte antigen-G genotype and success of in vitro fertilization and pregnancy outcome. **Tissue Antigens** 64: 66, 2004

HVIID TV, RIZZO R, MELCHIORRI L, STIGNANI M, BARICORDI OR. Polymorphism in the 5' upstream regulatory and 3' untranslated regions of the HLA-G gene in relation to soluble HLA-G and IL-10 expression. **Hum Immunol.** 67(1-2):53-62, 2006.

HYLENIUS S, ANDERSEN AM, MELBYE M, HVIID TV. Association between HLA-G genotype and risk of pre-eclampsia: a case-control study using family triads. **Mol Hum Reprod.**;10(4):237-46, 2004.

ISHITANI A, SAGESHIMA N, LEE N, DOROFEEVA N, HATAKE K, MARQUARDT H & GERAGHTY DE. Protein expression and peptide binding suggest unique and interacting functional roles for HLA-E, F, and G in maternal-placental immune recognition. **Journal of Immunology** 1: 171(3):1376-1384, 2003.

KHAKOO SI, RAJALINGAM R, SHUM BP, WEIDENBACH K, FLODIN L, MUIR DG, CANAVEZ F, COOPER SL, VALIANTE NM, LANIER LL & PARHAM P. Rapid evolution of NK cell receptor systems demonstrated by comparison of chimpanzees and humans. **Immunity.** 12(6):687-98, 2000.

KHAKOO SI, GELLER R, SHIN S, JENKINS JA & PARHAM P. The D0 domain of KIR3D has a major histocompatibility complex class I binding enhancer. **Journal of Experimental Medicine** 196: 911-921, 2002.

KIDD JR, PAKSTIS AJ, ZHAO H, LU R-B, OKONOFUA FE, ODUNSI A, GRICORENKO E, BONNE-TAMIR B, FRIED-LAENDER J, SCHULZ LO, PARNAS J & KIDD KK. Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, *PAH*, in a global representation of populations. **American Journal Human Genetics** 66:1882-1899, 2000.

KING A, BOOCOOCK C, SHARKEY AM, GARDNER L, BERETTA A, SICCARDI AG & LOKE YW. Evidence for the expression of HLAA-C class I mRNA and protein by human first trimester trophoblast. **Journal of Immunology** 15;156(6):2068-2076, 1996.

KING A, BURROWS TD, HIBY SE, BOWEN JM, JOSEPH S, VERMA S, LIM PB, GARDNER L, LE BOUTEILLER P, ZIEGLER A, UCHANSKA-ZIEGLER B & LOKE YW. Surface expression of HLA-C antigen by human extravillous trophoblast. **Placenta** 376-387, 2000.

KLEIN RG. **The human career. Chicago University Press, 1999.**

KITCHEN, A, MIYAMOTO MM, MULLIGAN CJ. A three-stage colonization model for the peopling of the Americas. **PLoS ONE** 3: 1596, 2008.

KOVATS S, MAIN EK, LIBRACH C, STUBBEBINE M, FISHER SJ & DeMARS R. A class I antigen, HLA-G, expressed in the human trophoblasts. **Science** 248: 200-223, 1990

KRONER A, GRIMM A, JOHANNSEN K, MAURER M & WIENDL H. The genetic influence of the nonclassical MHC molecule HLA-G on Multiple Sclerosis. **Human Immunology** 68: 422-425, 2007.

KUHNER MK, YAMATO J & FELSENSTEIN J. Maximum likelihood estimation of population growth rates based on the coalescent. **Genetics** 149:429-34, 1998.

KULKARNI S, MARTIN MP & CARRINGTON M. The Yin and Yang of HLA and KIR in human disease. **Seminars Immunology** 20:343-52, 2008.

KULKARNI S, MARTIN MP, CARRINGTON M. KIR genotyping by multiplex PCR-SSP. **Methods Mol Biol.** 612:365-75, 2010.

LAHIRI DK & NURENBERG JR JI. A rapid non-enzymatic method for the preparation of HMW DANN from blood for RFLP studies. **Nucleic Acid Research** 19: 5444, 1991.

LARSEN MH & HVIID TV. Human leukocyte antigen-G polymorphism in relation to expression, function, and disease. **Human Immunology** 70(12):1026-34, 2009.

LIU X, FU Y, LIU Z, LIN B, XIE Y, LIU Y, XU Y, LIN J, FAN X, DONG M, ZENG K, WU CI & XU A. An ancient balanced polymorphism in a regulatory region of human major histocompatibility complex is retained in Chinese minorities but lost worldwide. **American Journal of Human Genetics.** 78(3):393-400, 2006.

MARTIN M P, QI Y, GAO X, YAMADA E, MARTIN JN, PEREYRA F, COLOMBO S, BROWN EE, LESLEY SW, PHAIR J, GOEDERT JJ, BUCHBINDER S, KIRK DG, TELENTI A, CONNORS M, O'BRIEN SJ, WALKER BD, PARHAM P, DEEKS SG, MCVICAR DW & CARRINGTON M. Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. **Nature Genetics** 39:733-740, 2007.

MENDES-JUNIOR CT, CASTELLI EC, SIMÕES RT, SIMÕES AL & DONADI EA. HLA-G 14-bp polymorphism at exon 8 in Amerindian populations from the Brazilian Amazon. **Tissue Antigens.** 70:255-260, 2007.

MENIER C, ROUAS-FREISS N, FAVIER B, LEMAULT J, MOREAU P & CAROSELLA ED. Recent advances on the non-classical major histocompatibility complex class I HLA-G molecule. **Tissue Antigens.** vol:20120-20126, 2010.

MIDDLETON D & GONZALEZ F. The extensive polymorphism of KIR genes. *Immunology*, 129:8-19, 2009

MORETTA L & MORETTA A. Killer immunoglobulin-like receptors. **Current Opinion Immunology** 16:626-33, 2004.

MYLES S, TANG K, SOMEL M, GREEN RE, KELSO J & STONEKING M. Identification and analysis of genomic regions with large between-population differentiation in humans. **Annual Human Genetic** 72:99-110, 2008.

NACHMAN MW & CROWELL SL. Contrasting evolutionary histories of two introns of the Duchene Muscular Dystrophy, *dmd*, in humans. **Genetics** 155: 1855-1864, 2000.

NIELSEN R. Molecular signatures of natural selection. **Annual Review Genetic** 39:197-218, 2005a.

NIELSEN R, BUSTAMANTE C, CLARK AG, GLANOWKI S & SACKTON TB. A scan for positively selected genes in the genomes of humans and chimpanzees. **PLoS Biology** 3:e170, 2005b.

NORMAN PJ, ABI-RACHED L, GENDZEKHADZE K, KORBEL D, GLEIMER M, ROWLEY D, BRUNO D, CARRINGTON CV, CHANDANAYINGYONG D, CHANG YH, CRESPI C, SARUHAN-DIRESKENELI G, FRASER PA, HAMEED K, KAMKAMIDZE G, KORAM KA, LAYRISSE Z, MATAMOROS N, MILÀ J, PARK MH, PITCHAPPAN RM, RAMDATH DD, SHIAU MY, STEPHENS HA, STRUIK S, VERITY DH, VAUGHAN RW, TYAN D, DAVIS RW, RILEY EM, RONAGHI M & PARHAM P. Unusual selection on the KIR3DL1/S1 natural killer cell receptor in Africans. **Nature Genetic** 39:1092-1099, 2007.

O'CALLAGHAN CA, BELL JI. Structure and function of the human MHC class Ib molecules HLA-E, HLA-F and HLA-G. **Immunol Rev.** 163:129-38, 1998.

OBER C, ALDRICH CL, CHERNONEVA I, BILLSTRAND C, RAHIMOV F, GRAY HL & HYSLOP T. Variation in the HLA_G promoter region influences miscarriage rates. **American Journal of Human Genetics** 72:1425-1435, 2003.

ORANGE JS & BALLAS ZK. Natural killer cells in human health and disease. **Clinical Immunology** 118(1):1-10, 2006

PARHAM P, ARNETT KL, ADAMS EJ, LITTLE AM, TEES K, BARBER LD, MARSH SGE, OHTA T, MARKOW T & PETZL-ERLER ML. Episodic evolution and turnover of *HLA-B* in the indigenous human populations of the americas. **Tissue Antigens** 50:219-232, 1997.

PARHAM, P. MHC class I molecules and KIRs in human history, health and survival. **Nature Reviews Immunology.** 5(3): 201-14, 2005.

Povos indígenas do Brasil [on line]. Disponível: www.socioambiental.org. [capturado em 25 de Abril de 2006].

RAMALHO RF, SANTOS EJ, GUERREIRO JF & MEYER D. Balanced polymorphism in bottlenecked populations: the case of the CCR5 5' cis-regulatory region in Amazonian Amerindians. **Human Immunology** 71:922-8,2010.

RAYMOND & ROUSSET. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. **Journal of Heredity**, 86:248-249, 1995

RONALD J & AKEY JM. Genome-wide scans for loci under selection in humans. **Human Genomics** 2: 113-125, 2005.

ROBINSON J, WALLER MJ, STOEHR P & MARSH SGE. IPD - the Immuno Polymorphism Database. *Nucleic Acids Research*,331:523-526, 2005. < <http://www.ebi.ac.uk/ipd/kir/>>
Acessado em: 10 de maio de 2011.

ROUSSEAU P, LE DISCORDE M, MOUILLOT G, MARCOU C, CAROSELLA ED & MOREAU P. The 14 bp deletion-insertion polymorphism in the 3' UT region of the HLA-G gene influences HLA-G mRNA stability. **Human Immunology** 64(11):1005-10, 2003.

SABETI PC, REICH DE, HIGGINS JM, LEVINE HZ, RICHTER DJ, SCHAFFNER SF, GABRIEL SB, PLATKO JV, PATTERSON NJ, MCDONALD GJ, ACKERMAN HC, CAMPBELL SJ, ALTSHULER D, COOPER R, KWIATKOWSKI D, WARD R & LANDER ES. Detecting recent positive selection in the human genome from haplotype structure. **Nature** 419:832-837, 2002.

SACHIDANANDAM R, WEISSMAN D, SCHMIDT SC, KAKOL JM, STEIN LD, MARTH G, SHERRY S, MULLIKIN JC, MORTIMORE BJ, WILLEY DL, HUNT SE, COLE CG, COGGILL PC, RICE CM, NING Z, ROGERS J, BENTLEY DR, KWOK PY, MARDIS ER, YEH RT, SCHULTZ B, COOK L, DAVENPORT R, DANTE M, FULTON L, HILLIER L, WATERSTON RH, MCPHERSON JD, GILMAN B, SCHAFFNER S, VAN ETEN WJ, REICH D, HIGGINS J, DALY MJ, BLUMENSTIEL B, BALDWIN J, STANGE-THOMANN N, ZODY MC, LINTON L, LANDER ES, ALTSHULER D; INTERNATIONAL SNP MAP WORKING GROUP. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. **Nature**. 409(6822):928-33, 2001.

SALZANO F & CALLEGARI-JAQCUES S. **South American Indians: a case study in evolution.** Clarendon Press, Oxford, 1988.

SALZANO FM. Molecular variability in amerindians: widespread but uneven information. **Anual Academica Brasileira de Ciência** 74:..223-263, 2002.

SAMBROOK J, FRITSCH EF & MANIATIS T. **Molecular cloning- a laboratory manual.** 2a. ed., Cold Spring Harbor laboratory Press, Plainview.

SCHELLEKENS J, TILANUS MG & ROZEMULLER EH. The elucidation of KIR2DL4 gene polymorphism. **Molecular Immunology** 45:1900-1906, 2008.

SCHURR TG, BALLINGER SW, GAN YY, HODGE JA, MERRIWETHER DA, LAWRENCE DN, KNOWLER WC, WEISS KM & WALLACE DC. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies suggesting they derived from four primary maternal lineages. **American Journal Human Genetics** 46: 613-623, 1990.

SINGLE RM, MARTIN MP, GAO X, MEYER D, YEAGER M, KIDD JR, KIDD KK & CARRINGTON M. Global diversity and evidence for coevolution of KIR and HLA. **Nature Genetics** 39:1114-1119, 2007.

SOLBERG OD, MACK SJ, LANCASTER AK, SINGLE RM, TSAI Y, SANCHEZ-MAZAS A & THOMSON G. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. **Human Immunology** 69: 443-64, 2008.

SOLIER C, MALLET V, LENFANT F, BERTRAND A, HUCHENQ A, LE & BOUTEILLER P. HLA-G unique promoter region: functional implications. **Immunogenetics** 53: 617-625, 2001.

STEPHENS M, SMITH N & DONNELLY P. A new statistical method for haplotype reconstruction from population data. **American Journal of Human Genetics**, 68:978-989, 2001.

TAJIMA F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics** 123:95, 1989.

TAN Z, SHON MS & OBER C. Evidence of balancing selection at the *HLA-G* promoter region. **Human Molecular Genetics** 14:3619-3628, 2005.

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. **American Journal Human Genetics** 68:1485–1496, 2001.

THANANCHAI H, GILLESPIE G, MARTIN MP, BASHIROVA A, YAWATA N, YAWATA M, EASTERBROOK P, MCVICAR D W, MAENAKA K, PARHAM P, CARRINGTON M, DONG T & ROWLAND-JONES S. Cutting Edge: Allele-Specific and Peptide-Dependent Interactions between KIR3DL1 and HLA-A and HLA-B. **Journal of Immunology** 178: 33–37, 2007.

THOMAS R, YAMADA E, ALTER G, MARTIN MP, BASHIROVA AA, NORMAM PJ, ALTFELD M, PARHAM P, ANDERSON SK, MCVIAR DW e CARRINGTON M. Novel KIR3DL1 alleles and their expression. Levels on NKcells: convergent evolution of KIR3DL1 phenotype variation? **Journal of Immunology** 180: 6743-6750, 2008.

UHRBERG M, VALIANTE NM, SHUM BP, SHILLING HG, LIENERT-WEIDENBACH K, CORLISS B, TYAN D, LANIER LL & PARHAM P. Human diversity in killer cell inhibitory receptor genes. **Immunity**;7(6):753-63, 1997.

VALLENDER EJ & LAHN BT. Positive selection on the human genome. **Human Molecular Genetics**. 13 Spec No 2:R245-54, 2004.

VEIT TD & CHIES JAB. Tolerance versus immune response-MicroRNAs as important elements in the regulation of the HLA-G gene expression. **Transplant Immunology** 20: 229-231, 2009.

VILCHES C & PARHAM P. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. **Annual Review Immunology** 20:217-251, 2002.

VOIGHT BF, KUDARAVALLI S, WEN X & PRITCHARD JK. A map of recent positive selection in the human genome. **PLoS Biology** 4:72-76, 2006.

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 Genetics** 3:e185, 2007.

WATTERSON, G. 1978. The homozygosity test of neutrality. **Genetics** 88:405-417, 1978.

WATIKINS DI, MCADAM SN, LIU X, STRANG CR, MILFORD EL, LEVINE CG, GARBER TL, DOGON AL, LORD CI & GHIM SH. New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. **Nature** 28;357(6376):329-33, 1992.

WIENDL H, MITSDOERFFER M, HOFMEISTER V, WISCHHUSEN J, BORNEMANN A, MEYERMANN R, WEISS EH, MELMS A & WELLER M. A functional role of HLA-G expression in human gliomas: an alternative strategy of immune escape. **Journal of Immunology** 68:4772-4780, 2002.

WILLIAMSON SH, HUBISZ MJ, CLARK AG, PAYSEUR BA, BUSTAMANTE CD & NIELSEN R. Localizing recent adaptive evolution in the human genome. **PLoS Genetic** 3:e90, 2007.

WITT CS, MARTIN A, CHRISTIANSEN FT. Detection of KIR2DL4 alleles by sequencing and SSCP reveals a common allele with a shortened cytoplasmic tail. **Tissue Antigens**. 56(3):248-57, 2000.

WRAY GA. The evolutionary significance of cis-regulatory mutations. **Nature Review Genetics** 8: 206-216, 2007.

YAMATA M, YAMATA N, MCQUEEN KL, CHENG NW, GUETHLEIN LA e RAJALINGAM R, SHILLING HG & PARHAM P. Predominance of group a KIR haplotypes in Japanese associated with diverse Nk cell repertoires of KIR expression. **Immunogenetics** 54: 543-550, 2002.

YANG NN, MAZIÈRES S, BRAVI C, RAY N, WANG S, BURLEY MW, BEDOYA G, ROJAS W, PARRA MV, MOLINA JA, GALLO C, POLETTI G, HILL K, HURTADO AM, PETZLERLER ML, TSUNETO LT, KLITZ W, BARRANTES R, LLOP E, ROTHHAMMER F, LABUDA D, SALZANO FM, BORTOLINI MC, EXCOFFIER L, DUGOUJON JM, RUIZ-LINARES A. Contrasting patterns of nuclear and mtDNA diversity in Native American populations. **Ann Hum Genet** 74(6):525-38, 2010.

YAWATA MN, YAWATA N, DRAGHI M, LITTLE A-M, PARTHENIOU F & PARHAM, P. Roles for HLA and KIR polymorphisms in natural killer cell repertoire selection and modulation of effector function. **The Journal of Experimental Medicine** 203:633–645, 2006.

ZAGO MA, SILVA JUNIOR W, TAVELLA MH, SANTOS SBE, GUERREIRO JF & FIGUEIREDO MS. Interpopulational and intrapopulational genetic diversity of amerindians as revealed by six variable number of tandem repeats. **Human Heredity** 46: 274-289, 1996.

ZHIVOTOVSKY LA ,BENNETT L, BOWCOCK AM & FELDMAN MW. Human population expansion and microsatellite variation. **Molecular Biology Evolution** 17: 757-767, 2000.

ZHU FM, JIANG K, LV QF, JE J & YAN LX. Investigation of Killer cell immunoglobulin-like receptor KIR2DL4 diversity by sequence-based typing in Chinese population. *Tissue Antigens*, 67: 214-221, 2006.

ANEXO

PROJETO ADICIONAL: Análise populacional do GENE *MC1R*

Além do estudo dos genes *KIR* e *HLA-G*, o presente projeto de doutorado também propôs estudar a variação da região codificadora do gene *MC1R* em populações Ameríndias. Os dados para a análise desse gene foram gerados e analisados durante o período de doutoramento. Entretanto, como esse gene está relacionado com pigmentação, ficando fora do contexto adotado na presente tese (seleção natural em genes do sistema imune), optamos por transformar esse capítulo em um manuscrito para submissão e incluí-lo na presente tese como um projeto adicional. Assim, apresentamos uma tese dentro de um grande tema central sem desmerecer a importância e o tempo dedicado ao estudo desse gene durante os quatro anos de doutoramento.

A seguir, apresentamos uma cópia do manuscrito referente ao estudo do gene *MC1R* em populações indígenas americanas. Além do gene *MC1R*, o manuscrito analisa um polimorfismo no gene *MATP*, cujos dados foram obtidos no laboratório da Professora Doutora Regina Célia Mingroni Netto.

**THE GENETICS OF PIGMENTATION IN AMERINDIANS: *MC1R* AND *MATP*
POLYMORPHISMS.**

**PINCERATI MR¹, AURICCHIO MTBM², GUERREIRO JF³, SANTOS EJM³,
MINGRONI-NETTO R², PETZL-ERLER ML⁴, TARZONA-SANTOS E⁵, HURTADO
AM⁶, RUIZ-LINARES A⁷, MEYER D¹**

1. Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, 05508-090, São Paulo, São Paulo, Brazil.
2. Centro de Estudos do Genoma Humano, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, 05508-090, São Paulo, São Paulo, Brazil.
3. Laboratório de Genética Humana e Médica, Instituto de Ciências Biológicas, Universidade Federal do Pará, 66075-110, Belém, Pará, Brazil.
4. Laboratório de Genética Molecular Humana, Departamento de Genética, Universidade Federal do Paraná, 81531-980, Curitiba, Paraná, Brazil.
5. Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, 31270-910, Belo Horizonte, Minas Gerais, Brazil.
6. Department of Anthropology, University of New Mexico, 87131, Albuquerque, USA.
7. The Galton Laboratory, Department of Biology, University College London, WC1E 6BT, London, United Kingdom.

ABBREVIATED TITLE: Polymorphisms in *MC1R* and *MATP* gene.

KEY WORDS: pigmentation, Amerindians, *MC1R*, *MATP*.

CORRESPONDENCE TO: Diogo Meyer

Rua do Matão, 277.

Instituto de Biociências

CEP 05508-090

Tel. (+55 11) 3091 8092

Fax (+55 11)

diogo@ib.usp.br

FINANCIAL SUPPORT: This work was supported by the Brazilian National Research Council (CNPq) (grant 474510/2004-5) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), grants 2003/01583-9 and 2006/

ABSTRACT

Skin pigmentation is one of most diverse human phenotypes and varies across human populations. The correlation between the skin pigmentation and sunlight suggests that the variation in pigmentation is due adaptation to different ultraviolet radiation. Some *MC1R* and *MATP* variants have been associated with normal skin variation and, therefore, good candidates to studies of evolution of skin pigmentation as well as the study of Amerindian populations that show darker skin pigmentation when compared with their ancestors. In this study, we resequence the entire coding region of the *MC1R* gene and the *C394G* polymorphism in the *MATP* gene in Amerindian populations. We found seven variants in the *MC1R* gene, five of them are non-synonymous sites (*Leu44His*, *Arg163Gln*, *Ala181Thr*, *Pro268Leu* and *Ala299Ser*) and two are synonymous sites (*Asp42* and *Thr314*). Five of the seven variants found (*Asp42*, *Leu44His*, *Ala181Thr*, *Pro268Leu* and *Ala299Ser*) have been not reported previously. We also find eight *MC1R* haplotypes. When compared the frequencies of these alleles between Amerindian and Asian populations. We found that they are very similar, mainly because of the sharing of the *Arg163Gln* allele in high frequencies. For the *MATP* polymorphism, the allele *G* is almost fixed in Amerindian populations. Asian populations also present this allele in high frequency. Though has been postulate that the *MC1R* and *MATP* gene are correlated with variation in skin pigmentation, our results demonstrated that variants in *MC1R* gene and the *C394G* *MATP* polymorphism cannot explain the morphological difference between Amerindians and Asian populations.

INTRODUCTION

The correlation between skin pigmentation and incident ultraviolet radiation (UV) suggests that sunlight is one of the most important environmental variables shaping normal pigmentation variation, leading to adaptive hypotheses for variation in pigmentation (Jablonski and Chaplin, 2000).

Over the last decade remarkable progress has been made in understanding the genetic basis of pigmentation and currently several loci that contribute to normal variation in pigmentation have been detected (reviewed in McEvoy *et al.*, 2006). Recent studies have shown evidence for positive selection associated to the transition of dark to light skin, leading to large haplotypes of low variation or large F_{st} values surrounding loci involved in pigmentation. Interestingly, this transition appears to have involved different loci in Europeans (e.g. *TYRP1*, *SLC24A5*, *MATP*, *MYO5A*) and Asians (e.g., *ADAM17*, *ADAMTS20*). The functional argument underlying this scenario is that darker skin is favored in regions of high ultraviolet radiation (UVR) due to its protection against sunburn and skin cancer, but that at higher latitudes individuals with lighter pigmentation were favored by experiencing increasing penetration of UVR, allowing sufficient vitamin D photosynthesis (reviewed in McEvoy *et al.*, 2006).

The first gene discovered to be associated to normal skin pigmentation variation was the Melanocortin 1 Receptor gene (*MC1R*) (Harding *et al.*, 2000). The MC1R protein is a cell-surface receptor of melanocytes and a major player in the melanin production pathway, regulating the switch between the production of eumelanin (red/yellow melanins) and pheomelanin (brown/black melanins) (Makova and Norton, 2005). The *MC1R* gene is highly polymorphic in humans, shows high association to the red hair color (RHC) phenotype (Box *et al.*, 1997; Sturm *et al.*, 2003), and has increased polymorphism in European as compared

to African populations (Rana *et al.*, 1999; Harding *et al.*, 2000). The high diversity of *MC1R* in European populations, along with the high number of non-synonymous polymorphisms found in these populations, has been interpreted as an evidence for diversifying selection acting in Europeans (Rana *et al.*, 1999), which would favor lighter skin in regions of low UV incidence due to loss-of-function mutations in the *MC1R* gene. However, the same finding has also been interpreted to be a consequence of relaxed functional constraints in regions of low UV incidence, where deviation from normal eumelanin production is tolerated (Harding *et al.*, 2000).

A second locus associated with variation in normal pigmentation is the Membrane Associated Transporter Protein locus (*MATP*), whose product plays a role in the trafficking and intracellular processing of the tyrosinase (TYR), which catalyzes the first two steps in the melanin synthesis pathway (Costin *et al.*, 2003). *MATP* gene polymorphisms are associated to variation of normal human skin, hair and eye color, with two alleles, *C374G* and *G272A* are significantly associated with dark hair, skin and eye color in Europeans (Graf *et al.*, 2005).

While much of the investigation on the role of selection on genes involved in skin color has focused on the transition from dark to light skin, Amerindians offer an opportunity to investigate change towards darker pigmentation. Jablonski and Chaplin (2000) showed that four Native America populations (the Caingang and Guarani of Brazil, and the Maranon and Nunoa of Peru) had lower reflectance values than Asian populations included in the survey (Cambodian, Southern Chinese, Tibetan, Japanese, Nepalese, and Vietnamese). In the present study, we investigate the genetic variability of two well documented pigmentation genes (*MC1R* and *MATP*) in Amerindians, by resequencing of the entire coding region of the *MC1R* gene and by genotyping of the *C374G* polymorphism of the *MATP* locus. Our goal is to characterize the diversity of the *MC1R* and *MATP* genes and to investigate if the transition

from lighter to darker skin, observed in the contrast between Amerindians and North-East Asians, is correlated to variability in these genes. Our samples include populations from Eastern Amazonia and the Andes, which are regions with high UV incidence.

MATERIALS AND METHODS

Population samples and DNA extraction

We analyzed DNA samples from 82 individuals for *MC1R* and 60 for *MATP* genes (Table 1). Samples are curated by the Laboratory of Human and Medical Genetics of Universidade Federal do Pará, Brazil (Arara do Laranjal, Arara do Iriri, Kayapo, Krokaimoro, Parakana, Katuena Zoe, Waiampi and Awa-Guaja); Laboratory of Human Molecular Genetics of Universidade Federal do Paraná, Brazil (Guarani-Kaiowa, Guarani-Nandeva, Guarani-Mbia, Ache and Kaingang); Departament of General Biology of Universidade Federal de Minas Gerais, Brazil (Quechua population), Department of Genetics of Universidade Federal do Rio Grande do Sul, Brazil (Guaymi, Zoro, Lengua, Xavante, Bribri, Guatuso, Kuben-Kran-Keng), and The Galton Laboratory, University College London, United Kingdom (Arhuaco). Appropriate consent was obtained from tribes to use the DNA in genetic and anthropological studies.

Amplification and sequencing of MC1R gene

The nucleotide sequence of *MC1R* gene (a single 954 bp exon) was obtained using a nested polymerase chain reaction (PCR) and direct sequencing method. The primers for amplification and sequencing were based on Box *et al.* (1997).

All nucleotide polymorphisms were confirmed by using sequencing primers in both directions. Sequences were analyzed using the program package Phred/Phrap/Consed (Ewing and Green, 1998; Ewing *et al.*, 1998; Gordon *et al.*, 1998).

For novel mutations found in the present study, a second round of amplification and sequencing was performed to validate the findings.

Amplification and genotyping of MATP gene

For the identification of the *C374G* polymorphism in the *MATP* gene (1122G>C), the MegaBace SNUPe genotyping kit (Amersham Biosciences) was employed. Products of the SNUPe reactions were purified with AutoSeq96 Dye Terminator clean-up kit (Amersham Biosciences) and were analysed with the MegaBACE 1000 DNA analysis system (Amersham Biosciences).

Data Analysis

Allele frequencies were obtained by direct counting and Hardy-Weinberg equilibrium was assessed using the Guo & Thompson's method (1992), implemented in ARLEQUIN v. 3.1 (Excoffier *et al.*, 2005). The same software was used to perform the Tajima's D and Fu's F_s tests. Comparisons of the allelic frequencies between Asian and Amerindians populations to test for population differentiation were performed via contingency tables.

RESULTS AND DISCUSSION

The entire coding sequence of the *MC1R* gene was sequenced in 82 individuals. Compared to the human consensus sequence (Rana *et al.*, 1999), our samples differed at five non-synonymous sites (*Leu44His*, *Arg163Gln*, *Ala181Thr*, *Pro268Leu* and *Ala299Ser*) and two synonymous sites (*Asp42* and *Thr314*). The *Arg163Gln* and *Thr314* variants have been previously reported in several human populations, but the other five mutations were found for the first time in our study. Only the *Arg163Gln* mutation was found at high frequencies (with the A allele found at an overall frequency of 74%). All others were restricted to single individuals or to very closely related family members (Table 2).

Arg163Gln mutation is also the most frequent variant observed in Asian populations (65.5%), a frequency which is substantially higher than that of Europeans (average 3.5%) and Africans (0.7%) (Gerstenblith *et al.*, 2007). The synonymous variant *Thr314* was found in only one Kaingang individual. This variant is found at high frequencies in Africa (44.44%) and intermediate frequency in Asia (13.27%) (compiled by Savage *et al.*, 2008).

Among the five new variants in the *MC1R* gene described in this study, only *Asp42* is synonymous. This variant was only found in individuals of the Arara do Iriri population, with the derived allele reaching the frequency of 79% in this population. According to Ribeiro-dos-Santos *et al.* (1999), oral reports indicate that the individuals who constitute this village today are descendants from a single couple of siblings who were expelled from the largest Arara village. Therefore, the high frequency of this allele is a result of founder effect.

The non-synonymous *Ala299Ser* variant, found in a single individual from the Mapuera population, coincides with a polymorphic position already described (*Ala299Thr*). The variant *Pro268Leu* was found in homozygous state in the Kaingang population in a single individual and the non-synonymous variants *Leu44His* and *Ala181Thr* were found in

individuals in heterozygous state in the Aché and Guarani-Kaiowá populations, respectively.

Overall, we observed significant deviation from the Hardy-Weinberg equilibrium ($p < 0.001$), with an excess of homozygous genotypes. This deviation from Hardy-Weinberg equilibrium is likely a consequence of the fact that our overall sample consists of a pool of individuals from highly differentiated populations, resulting in an excess of homozygotes (the Wahlund effect).

We compiled a list of the variants found in East Asian populations (Table 2) and compared the frequencies of the most frequent variants found in Asian and Amerindian populations: *Val92Met* (13.41% and 0%, respectively), *Arg163Gln* (75.51% and 74%), and *Thr314* (13.27% and 1.2%). The frequencies for *Arg163Gln* did not differ significantly between Asian and Amerindians populations ($p = 0.83$). The frequencies of *Val92Met* and *Thr314* differ significantly between the two populations ($p < 0.01$ for both). However, Nakayama *et al.* (2006) show that the *Val92Met* has increased allele frequency in the Southeast Asian populations, and its frequency decreases in East Asian populations. They also show that the *Val92Met* allele arose on a haplotype carrying the *Thr314* variant. We postulate that differences between the allele frequency of these two variants are due to demographic events leading to a recent increase of this haplotype in Asian populations, most probably after the split of between Amerindians and Asians. Overall, the *MC1R* allele frequencies in Amerindian populations are very similar to that of Asian population, mainly due to the sharing the high frequency variant *Arg163Gln*. The main differences observed among these regions concern variants at substantially lower frequencies.

We also determined the haplotypes at the *MC1R* locus. We found eight different haplotypes in the samples studied (Figure 1), with the most common *MC1R* haplotype found

in Amerindian populations is also common in East Asian populations, (Nakayama *et al.*, 2006).

We also compared the frequencies between Amerindian populations living in different UV incidence areas. No difference between allele frequencies were found (data not shown).

In the present study we found five new variants that show very low frequencies (except for the *Asp42* variant, which shows a high frequency, but is found in a single highly inbred population, the Arara do Iriri). The new alleles found in this study, due to low frequencies and the fact that they were found only in the present samples, may be considered private alleles of Amerindians populations. Private *MC1R* alleles have also been found in Asian populations (Nakayama *et al.*, 2006). The presence of private alleles in these populations may be an indicative of a continued relaxation of functional constraint on the *MC1R* gene in non-African populations, as postulated by Harding *et al* (2000).

For the *MATP* gene, a total of 60 Amerindians from different populations were analyzed for the *C374G* SNP. The *G* allele was found at a frequency of 98%. This allele shows high frequencies in non-Caucasian population, and has been found at a frequency of 89% in Asians (Graft *et al.*, 2005). The *C374G* SNP at the *MATP* locus was reported to be associated with normal variation in pigmentation in a population of European ancestry (Graf *et al.*, 2005), and valuable as an ancestry informative marker for Europeans (Yuasa *et al.*, 2004). Our results for the *C374G* polymorphism are consistent with the idea that the *374*C* allele originated outside of Africa after the divergence of modern Asians and Europeans. Our results show that for *C374G* polymorphism there is no significant difference between Amerindian and Asian populations in their allele frequency.

Although variants in the *MC1R* and *MATP* genes show a strong correlation with normal skin pigmentation variation, we found a lack of differentiation in their variants in

contrasts between Asians and Amerindians. Recent studies involving characterization of positive selection in human populations have shown new regions of the genome that have experienced recent evidence of selection, many of which show strong evidence of population-specific selective sweeps (Williamson *et al.*, 2007). Population-specific selective sweeps have been found in several genes related to skin pigmentation, suggesting the importance of multiple loci in the evolution of skin pigmentation. Therefore, our results confirm the idea that, despite the influence of *MC1R* and *MATP* genes in the variation in pigmentation of the skin and hair, other genes are involved in the control of pigmentation in the modifications in pigmentation seen in Amerindian populations .

In summary, in the present study we characterized the variation in the entire coding region of the *MC1R* gene and we genotyped the *C374G* position of *MATP* in Amerindians populations, which until this work had been poorly studied. Our goal was to test whether populations that show a transition to darker pigmentation display an underlying genetic change at these loci. Comparisons between our results and data from Asian populations show that there is no marked difference in the allele frequencies in these genes that can explain the morphological difference in their skin pigmentation. However, low frequency non-synonymous variants are fairly common in the Americas, and found restricted to individual populations. The identification and analysis of additional genes involved in human skin pigmentation will help clarify how the evolution of skin pigmentation in Amerindian populations took place.

ACKNOWLEDGMENTS

We thank the individuals of the population samples for their collaboration. We also thank to our laboratory colleagues for the technical assistance, critical comments, helpful discussion and friendship.

REFERENCES

BOX NF, WYETH JR, O'GORMAN LE, MARTIN NG, STURM RA. Characterization of melanocyte stimulating hormone receptor variant alleles in twins with red hair. **Hum Mol Genet.** 1997;6:1891-7.

COSTIN GE, VALENCIA JC, VIEIRA WD, LAMOREUX ML, HEARING VJ. Tyrosinase processing and intracellular trafficking is disrupted in mouse primary melanocytes carrying the underwhite (uw) mutation. A model for oculocutaneous albinism (OCA) type 4. **J Cell Sci.** 2003;116:3203-12.

EWING B, HILLIER L, WENDL M, GREEN P. Basecalling of automated sequencer traces using phred I. Accuracy assessment. **Genome Research** 1998;8:175-185.

EWING, B. AND GREEN, P. Basecalling of automated sequencer traces using Phred II. Error probabilities. **Genome Research** 1998;8:186-194.

EXCOFFIER, L. G. LAVAL, AND S. SCHNEIDER. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. **Evolutionary Bioinformatics Online** 2005;1:47-50.

GERSTENBLITH MR, GOLDSTEIN AM, FARGNOLI MC, PERIS K, LANDI MT. Comprehensive evaluation of allele frequency differences of MC1R variants across populations. **Hum Mutat.** 2007;28:495-505.

GORDON, D., C. ABAJIAN, AND P. GREEN. Consed: A Graphical Tool for Sequence Finishing. **Genome Research** 1998;8:195-202

GRAF J, HODGSON R, VAN DAAL A. Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. **Hum Mutat.** 2005;25:278-84.

HARDING RM, HEALY E, RAY AJ, ELLIS NS, FLANAGAN N, TODD C, DIXON C, SAJANTILA A, JACKSON IJ, BIRCH-MACHIN MA, REES JL. Evidence for variable selective pressures at *MC1R*. **Am J Hum Genet.** 2000;66:1351-61.

JABLONSKI NG and CHAPLIN G. The evolution of human skin coloration. **J Hum Evol.** 2000;39:57-106.

MAKOVA K & NORTON H. Worldwide polymorphism at the *MC1R* locus and normal pigmentation variation in humans. **Peptides** 2005;26:1901-8.

MCEVOY B, BELEZA S, SHRIVER MD. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. **Hum Mol Genet.** 2006;15:176-81.

NAKAYAMA K, SOEMANTRI A, JIN F, DASHNYAM B, OHTSUKA R, DUANCHANG P, ISA MN,SETTHEETHAM-ISHIDA W, HARIHARA S, ISHIDA T. Identification of novel functional variants of the melanocortin 1 receptor gene originated from Asians. **Hum Genet.** 2006;119:322-330.

RANA BK, HEWETT-EMMETT D, JIN L, CHANG BH, SAMBUUGHIN N, LIN M, WATKINS S, BAMSHAD M, JORDE LB, RAMSAY M, JENKINS T, LI WH. High polymorphism at the human melanocortin 1 receptor locus. **Genetics** 1999;151:1547-57.

RIBEIRO-DOS-SANTOS AK, GUERREIRO JF, SANTOS SE, ZAGO MA. The split of the Arara population: comparison of genetic drift and founder effect. **Hum Hered.** 2001;51:79-84.

SAVAGE SA, GERSTENBLITH MR, GOLDSTEIN AM, MIRABELLO L, FARGNOLI MC, PERIS K, LANDI MT. Nucleotide diversity and population differentiation of the melanocortin 1 receptor gene, MC1R. **BMC Genet.** 2008;10:9:31

STURM RA, DUFFY DL, BOX NF, NEWTON RA, SHEPHERD AG, CHEN W, MARKS LH, LEONARD JH, MARTIN NG. Genetic association and cellular function of MC1R variant alleles in human pigmentation. **Ann N Y Acad Sci.** 2003;994:348-58.

VALVERDE P, HEALY E, JACKSON I, REES J, THODY A. A variants of the melanocyte stimulating hormone receptor gene are associated with red hair and fair skin. **Nature genetics** 1995;11:328-330.

WILLIAMSON SH, HUBISZ MJ, CLARK AG, PAYSEUR BA, BUSTAMANTE CD, NIELSEN R. Localizing recent adaptive evolution in the human genome. **PLoS Genet.** 2007;3:90.

YUASA I, UMETSU K, WATANABE G, NAKAMURA H, ENDOH M, IRIZAWA Y.
MATP polymorphisms in Germans and Japanese: the L374F mutation as a population
marker for Caucasoids. **Int J Legal Med.** 2004;118:364-6.

TABLE 1. Populations considered in this study for the *MC1R* and *MATP* genes.

Population	Localization	Languages group	Sample size <i>MC1R</i>	Sample size <i>MATP</i>
Arara do Laranjal	Brazil/Pará	Karíb	10	8
Arara do Iriri	Brazil/Pará	Karíb	12	4
Kayapó Krokaimoro	Brazil/Pará	Ge	9	12
Kayapó Xikrim	Brazil/Pará	Ge	3	14
Kayapó Kuben-Kran-Kegn	Brazil/Pará	Caiapó	1	0
Parakanã	Brazil/Pará	Tupi	1	9
Katuena	Brazil/Amazônia	Karíb	4	4
Zoe	Brazil/Pará	Tupi	1	1
Waiampi	Brazil/Amapá	Tupi	6	5
Awá-Guajá	Brazil/Maranhão	Tupi	2	3
Guarani-Kaiowá	Brazil/Mato Grosso do Sul	Guarani	3	0
Guarani-Ñandeva	Brazil/Mato Grosso do Sul	Guarani	3	0
Guarani-Mbiá	Brazil/Paraná	Guarani	2	0
Kaingang	Brazil/Paraná	Ge	4	0
Aché	Paraguai	Guarani	4	0
Guaymi	Panamá	Guaymi	1	0
Zoró	Brazil/Mato Grosso	Guarani	1	0
Lengua	Paraguai	Lengua	1	0
Xavante	Brazil/Mato Grosso	Ge	1	0
Bribri	Costa Rica	Chibchan	1	0
Guatuso	Costa Rica	Rama	1	0
Quechua	Peru	Quechua	7	0
Arhuaco	Colombia	Chibchan	4	0
Total			82	60

TABLE 2. Allele frequencies (%) of *MC1R* polymorphisms in Asian and Native-American populations.

Protein	Nucleotide	Amerindians	Asia
Non-synonymous			
<i>Leu44His^c</i>	131 T→A	0.6	0
<i>Arg67Gln^a</i>	200 G→A	0	1.17
<i>Val92Met^a</i>	274 G→A	0	13.41
<i>Ile120Thr^a</i>	359 T→C	0	1.75
<i>Phe174Δ^b</i>	439-441del	0	0.004
<i>Arg151Cys^b</i>	451 C→T	0	0.002
<i>Thr157Ile^b</i>	470 C→T	0	0.004
<i>Pro159Thr^b</i>	475 C→A	0	0.002
<i>Arg163Gln^a</i>	488 G→A	74.0	75.51
<i>Ala166Gln^b</i>	497 C→G	0	0.004
<i>Ala181Thr^c</i>	541 G→A	0.6	0
<i>Fen196Leu^a</i>	586 T→C	0	0.15
<i>Pro268Leu^c</i>	803 T→C	1.2	0
<i>Ala299Ser^c</i>	895 G→T	0.6	0
Synonymous			
<i>Asp42^c</i>	126 C→T	12.0	0
<i>Thr314^a</i>	942 A→G	1.2	13.27

^a Frequencies from East and Southeast Asian populations compiled by Savage et al., 2008.

^b Frequencies from East Asian populations obtained by Nakayama et al., 2006.

^c New variants found for the first time in this study.

FIGURE 1. Network constructed from haplotypes of the *MC1R* gene. HCS= Human Consensus Sequence

