Rodrigo dos Santos Francisco

Polimorfismo e Evolução do Gene HLA-B nas Américas

São Paulo 2009

Polimorfismo e Evolução do Gene HLA-B nas Américas

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Ciências, na Área de Genética e Biologia Evolutiva

Orientador: Prof. Dr. Diogo Meyer

São Paulo 2009 Francisco, Rodrigo dos Santos Polimorfismo e evolução do gene *HLA-B* nas Américas 68 pp.

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

1. *HLA-B* 2. Ameríndios 3. Seleção Natural I. Universidade de São Paulo. Instituto de Biociências. 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. Dr. Diogo Meyer (Orientador)

Dedico essa dissertação aos meus pais, Ricardo e Iara e à minha irmã Juliana

Agradecimentos

A DEUS PELO TÉRMINO DE MAIS ESSA ETAPA.

AOS MEUS PAIS E MINHA IRMÃ PELO AMOR E COMPREESSÃO NESSES ANOS DE DISTÂNCIA.

AO PROFESSOR DIOGO MEYER PELA ORIENTAÇÃO E CONVIVÊNCIA.

AO PROFESSOR JOÃO STENGHEL MORGANTE PELA ACOLHIDA NO LABEC.

AOS PROFESSORES MARIA LUIZA PETZL-ERLER DA SAUDOSA UNIVERSIDADE FEDERAL DO PARANÁ E EDUARDO JOSÉ MELO DOS SANTOS DA UNIVERSIDADE FEDERAL DO PARÁ PELAS VÁRIAS CONTRIBUIÇÕES A ESSE TRABALHO.

AO GUTO PELO COMPANHERISMO E HORAS DE CONVERSAS.

À PINCERATI PELO MEU SEGUNDO NOME: LATINO!!!!!

AOS AMIGOS LABEQUIANOS: NANA, FELIPE, KELIZICHA, JUJÚ, GIGÍ, CAROL, ANINHA, RAMALHO, FEZINHA E SOFIA POR ESSES ANOS DE BOA CONVIVÊNCIA E AMISADE

À MINHA SEGUNDA FAMÍLIA DE SÃO MATEUS: TONI, RODES, APOLO, POLYANA, SEU OTACÍLIO, DONA JOSEFA E POP PELA ADOÇÃO NESSES ÚLTIMOS TRÊS ANOS.

Aos meus amigos de Curitiba: Cidoca, Rodolfinho, Loreley Jóice, Robinho, Ju Celma, Zé, Cris, Leti e Pati pela amisade a mais de 400 km.

AOS COMPANHEIROS DE REPÚBLICA: BRENO, JOAQUIM, JURANDIR, FERNANDO E ELAIR PELA CONVIVÊNCIA DIÁRIA.

À FAPESP PELA BOLSA CONCEDIDA

SUMÁRIO

LISTA DE TABELAS	vi
LISTA DE FIGURAS	vii
RESUMO	viii
1. INTRODUÇÃO	1
1.1. O CONTEXTO DA RESPOSTA IMUNE E ESTRUTURA DOS GENES	
HLA	2
1.2. EVIDÊNCIAS DE SELEÇÃO NATURAL NOS GENES HLA DE CLASSE I	
CLÁSSICOS	6
1.2.1. Evidências Populacionais de Seleção Natural nos Genes HLA	6
1.2.2. Modelos de Seleção Natural em Genes HLA	10
1.3. POPULAÇÕES AMERÍNDIAS	11
1.3.1. Variação dos genes HLA em Ameríndios	13
2. OBJETIVOS	16
2.1. OBJETIVOS ESPECÍFICOS	16
3. MATERIAIS E MÉTODOS	17
3.1. AMOSTRAS POPULACIONAIS	17
3.2. AMPLIFICAÇÃO, PURIFICAÇÃO E SEQÜENCIAMENTO DOS ÉXONS 2	
E 3 E DO ÍNTRON 2 DO GENE <i>HLA-B</i>	20
3.3. METODOS ANALÍTICOS	23
3.4. NOMENCLATURA DOS ALELOS DO GENE HLA-B	25
4. RESULTADOS	28
4.1. COMPARAÇÃO DOS RESULTADOS OBTIDOS PELAS TÉCNICAS PCR-	
SBT E PCR-SSOP	28
4.2. VERIFICAÇÃO DO EQUILÍBRIO DE HARDY-WEINBERG (HW)	31
4.3. DESCRIÇÃO DA VARIAÇÃO DO GENE <i>HLA-B</i>	32
4.3.1. Distribuição e Freqüências Alélicas	33
4.4. TESTES DE DESVIO DE NEUTRALIDADE	48
5. DISCUSSÃO	52
6. CONCLUSÕES	62
7. REFERÊNCIAS BIBLIOGRÁFICAS	64

LISTA DE TABELAS

Tabela 1.	Dados lingüísticos, localização geográfica e tamanho amostral das	
	populações estudadas	18
Tabela 2.	Seqüências dos oligonucleotídeos iniciadores	22
Tabela 3.	Concentração dos reagentes utilizados no mix de amplificação	22
Tabela 4.	Volumes dos reagentes utilizados na reação de seqüenciamento	22
Tabela 5.	Protocolo de ciclagem	23
Tabela 6.	Lista de alelos com éxons 2 e 3 idênticos que encontrados neste	
	trabalho	27
Tabela 7.	Classificação dos alelos de acordo com a sua distribuição	
	geográfica	27
Tabela 8.	Número de indivíduos tipados em cada técnica e discrepâncias de	
	resultados	30
Tabela 9.	Indivíduos com discrepância entre os resultados de PCR-SBT e PCR-	
	SSOP	30
Tabela 10.	Resultados do teste de equilíbrio de Hardy-Weinberg	32
Tabela 11.	Distribuição alélica na população Tundra Nentsi	34
Tabela 12.	Tamanho das amostras totais e sem indivíduos miscigenados	50
Tabela 13.	Resultados dos testes de neutralidade	51

LISTA DE FIGURAS

Figura 1.	Processamento e apresentação de peptideos via moleculas HLA de classe I	4	
Figura 2.	Esquema da molécula HLA de classe I obtida por cristalografia de raio X		
Figura 3.	Mapa do MHC humano		
Figura 4.	Distribuição da diversidade nos genes HLA de classe I clássicos	9	
Figura 5.	Mapa das Américas mostrando a localização geográfica das populações		
	amostradas nesse trabalho	19	
Figura 6.	Esquema do posicionamento dos iniciadores utilizados na amplificação e		
	seqüenciamento	23	
Figura 7.	Alinhamento dos alelos cosmopolitas e endêmicos	35	
Figura 8.	Freqüências alélicas das populações Chipewyan e Cree	37	
Figura 9.	Freqüências alélicas das populações Ojibwá e Mixe		
Figura 10.	Freqüências alélicas das populações Mixstec e Zapotec	39	
Figura 11.	Freqüências alélicas das populações Quiche, Cabecar e Guaymi	39	
Figura 12.	Freqüências alélicas das populações Zenu e Embera	40	
Figura 13.	Freqüências alélicas das populações Kogi, Arhuaco e Wayuu	41	
Figura 14.	Freqüências alélicas das populações Waunana e Ingano (Inga)	42	
Figura 15.	Freqüências alélicas das populações Aymara e Huilliche	42	
Figura 16.	Freqüências alélicas das populações <i>Ticuna N</i> e <i>Ticuna S</i>	44	
Figura 17.	Freqüências alélicas das populações Zoé e Katuena	45	
Figura 18.	Freqüências alélicas das populações Arara do Iriri, Arara do Laranjal e		
	Kaiapó	45	
Figura 19.	Freqüências alélicas da população Guarani	46	
Figura 20.	Freqüência dos alelos cosmopolitas, endêmicos e não-amerindios	48	
Figura 21.	Regressão linear entre a distância em km da Beríngia e a freqüência de		
	alelos endêmicos	48	

RESUMO

Existem evidências de que o gene *HLA-B*, cujo produto tem como função a apresentação de peptídeos a linfócitos CD8⁺ e ligação a receptores KIR, esteja sob a ação da seleção natural. As populações nativas da América do Sul possuem uma grande quantidade de alelos de *HLA-B* exclusivos (endêmicos) em freqüências elevadas, enquanto que as populações nativas norte-americanas compartilham seus alelos com a Ásia. A hipótese elaborada para explicar a diferença de perfil alélico entre as populações norte e sul-americanas é conhecida como *turnover* de alelos. Segundo essa hipótese, os alelos exclusivos de ameríndios teriam aumentado de freqüência e substituído os ancestrais por apresentarem com eficiência os antígenos dos patógenos sul-americanos e/ou devido ao efeito de deriva genética, que provavelmente foi importante como conseqüência dos pequenos tamanhos efetivos populacionais. Esse processo teria ocorrido várias vezes, explicando as diferenças nos perfis alélicos entre as populações sul-americanas, cada uma das quais com um determinado conjunto de alelos endêmicos.

Para verificar se as populações que possuem maiores freqüências de alelos endêmicos apresentam sinais da atuação da seleção natural, nós analisamos o polimorfismo do gene *HLA-B* em 474 amostras oriundas de 26 populações nativo-americanas e uma siberiana, quantificamos a freqüência de alelos endêmicos por população e região, e aplicamos os testes de neutralidade *Ewens-Watterson* e *D de Tajima*.

Nós encontramos uma correlação positiva entre aumento da distância do Estreito de Bering e maior freqüência de alelos endêmicos ($r^2 = 0,351$, p<0,01)). Entretanto, essa correlação desapareceu quando excluímos as populações norte-americanas ($r^2 = 0,109$, p>0,10), mostrando que a correlação é uma conseqüência das diferenças entre as populações localizadas na região neártica, que praticamente não possuem alelos endêmicos, e as populações neotropicais, que apresentam uma maior freqüência de alelos endêmicos. Esse resultado está de acordo com a hipótese de que há uma correlação entre o ambiente tropical e alterações do perfil alélico das populações ameríndias. O teste de *D de Tajima* mostrou desvios significativos de neutralidade, na direção de seleção balanceadora, ao longo de todo o continente (p<0,05), não havendo diferenças entre as populações com maiores ou menores freqüências de alelos endêmicos. Esse resultado é, provavelmente, conseqüência da captação de um sinal seletivo anterior à entrada no continente americano.

Nossos resultados mostram que a composição de alelos endêmicos difere entre os grupos populações, e não entre as populações individuais. Portanto, é plausível que a seleção natural tenha promovido a substituição dos alelos asiáticos antes do estabelecimento das populações ameríndias individuais que conhecemos hoje. Essa hipótese está de acordo com a quase total ausência de populações desviando significativamente do esperado segundo o modelo neutro, quando aplicamos o teste de *Ewens-Watterson*, que ao contrário do *D de Tajima*, é menos afetado por eventos seletivos antigos.

1. INTRODUÇÃO

Há um crescente interesse na detecção de genes e/ou regiões genômicas que foram alvos da seleção natural, com intuito de desvendar os processos evolutivos que atuaram na nossa e em outras espécies. A inferência de quais genes são selecionados indica quais são as regiões funcionalmente importantes do genoma (Nielsen, 2005) e contribui para o debate sobre a importância dos processos de deriva e seleção natural na evolução molecular. Vários estudos demonstraram a existência de regiões cujos padrões de variação são incompatíveis com regimes neutros de evolução e, paralelamente, houve importantes avanços nos métodos empregados para detectar a seleção natural a partir de dados genéticos (Bustamante, Fledel-Alon *et al.*, 2005; Nielsen, 2005; Sabeti, Schaffner *et al.*, 2006).

Muitos genes que participam da resposta imune apresentam evidências de seleção natural (Vallender e Lahn, 2004). Dentre esses, os genes *HLA (Human Leukocyte Antigen)*, devido à sua importância funcional, apresentam fortes evidências de estarem evoluindo sob seleção natural (Meyer e Thomson, 2001). Estudos populacionais mostraram que o gene *HLA-B*, um dos genes que apresenta os maiores níveis de diversidade na nossa espécie (Robinson, Waller *et al.*, 2003), apresenta grande diferenciação entre os ameríndios da América do Sul e outras populações, inclusive de ameríndios norte-americanos. Esse padrão pode ser interpretado seletivamente.

No presente estudo nós exploramos essa questão caracterizando a variação do gene *HLA-B* em 26 populações nativas de diversas regiões do continente americano para verificar a existência de sinais da atuação da seleção natural. Até o momento, nenhum estudo analisou a variabilidade genética em genes *HLA-B* para um conjunto de populações ameríndias com ampla distribuição geográfica e diversidade lingüística. Nas próximas seções revisaremos o papel dos genes *HLA* no sistema imune e apresentaremos as evidências de atuação da seleção

1.1.1. O CONTEXTO DA RESPOSTA IMUNE E ESTRUTURA DOS GENES HLA

O sistema imune é constituído por uma coleção de células e macromoléculas que nos protegem da uma infinidade de microorganismos patogênicos. Grosseiramente, podemos classificar o sistema imune em inato e adaptativo (Abbas e Lichtman, 2005). O sistema imune inato constitui a primeira linha de defesa do organismo e seus componentes são encontrados ao longo de toda a árvore filogenética dos animais (Parham, 2005). Os componentes do sistema imune inato incluem barreiras químicas e físicas (os epitélios em geral e as substâncias antimicrobianas produzidas por eles), células fagocíticas (macrófagos e neutrófilos), células NK (*natural killers*) e proteínas solúveis (sistema complemento e mediadores inflamatórios) (Abbas e Lichtman, 2005). Esses componentes já estão prontos mesmo antes de uma infecção e respondem rapidamente aos agentes invasores. Com exceção das células NK, os componentes do sistema imune inato reconhecem estruturas genéricas e essenciais à sobrevivência dos microrganismos, impedindo o estabelecimento da maioria das infecções.

Mecanismos de evasão do sistema imune inato surgiram em vários microrganismos. Como resposta, outro conjunto de células e moléculas permite a eliminação de infecções que escapam do controle da imunidade inata, montando uma resposta específica contra o agressor. Essa é a função do sistema imune adaptativo (Janeway, Travers *et al.*, 2001).

A resposta adaptativa é extremamente específica, normalmente sendo desencadeada por, e dirigida a, um determinado antígeno. Antígenos são os alvos da resposta imune, geralmente macromoléculas presentes nos patógenos. Toda a resposta adaptativa é dependente da função de linfócitos B e T. Os linfócitos B reconhecem antígenos solúveis (desde pequenas moléculas a macroestruturas encontradas nas superfícies de patógenos) e produzem os anticorpos, proteínas solúveis que se ligam aos antígenos, neutralizando sua atividade. Os anticorpos marcam os alvos e recrutam moléculas do sistema complemento e células fagocíticas, que por sua vez, eliminam os agentes infecciosos (Janeway, Travers *et al.*, 2001). A resposta mediada por anticorpos é o principal mecanismo de defesa contra patógenos extracelulares e suas toxinas.

Contra antígenos intracelulares existem os mecanismos da resposta adaptativa celular. Quando há uma infecção viral ou alteração de proteínas próprias como ocorre no câncer, peptídeos das proteínas estranhas são clivados no citoplasma da célula e bombeados para o retículo endoplasmático. No retículo esses peptídeos associam-se às moléculas HLA de classe I e o complexo HLA + peptídeo é levado à superfície celular. Linfócitos T CD8⁺ reconhecem esse complexo e desencadeiam uma resposta que culmina na eliminação da célula apresentadora (figura 1) (Doherty e Zinkernagel, 1975b). As células NK também interagem com as moléculas HLA de classe I, entretanto, elas são sensíveis às variações nos níveis de expressão dessas moléculas (Parham, 2005). Alguns patógenos e células cancerosas induzem a diminuição da expressão das moléculas HLA, um possível mecanismo de evasão do reconhecimento pelos linfócitos T. As células NK expressam receptores inibidores, os receptores KIR (*Killer cell immunoglobulin-like receptors*), que interagem com as moléculas HLA de classe I. Na ausência do sinal inibidor das moléculas HLA, as células NK desencadeiam uma reação citotóxica, eliminando a célula infectada.



Figura 1. Processamento e apresentação de peptídeos via moléculas HLA de classe I. Fonte: Male e colaboradores (2006), modificado.

As moléculas HLA-A, -B e -C, também conhecidas como moléculas HLA de classe I clássicas, são as principais responsáveis pela apresentação de peptídeos citoplasmáticos. Essas moléculas são constituídas por duas sub-unidades: a *cadeia* α e a β 2-*microglobulina*. A *cadeia* α é codificada pelos genes *HLA* de classe I clássicos, *HLA-A*, -*B* e -*C*, e apresenta três domínios extracelulares: α_1 , α_2 e α_3 . Os domínios α_1 e α_2 formam a fenda apresentadora de peptídeos, um sulco flanqueado por duas α -hélices dispostas sobre um assoalho de folhas β -pregueadas (figura 2) (Bjorkman, Saper *et al.*, 1987). Pequenos "bolsos" (*pockets*) na fenda da molécula HLA acomodam seis dos nove resíduos de aminoácido dos peptídeos (Saper, Bjorkman *et al.*, 1991).



Figura 2. Esquema e estrutura da molécula de HLA de classe I obtida por cristalografia de raio X. A *cadeia* α é mostrada em azul e a cadeia β_2 -*microglobulina* em vermelho. A região trans-membrana e a cauda citoplasmática são mostradas apenas no esquema. A molécula de HLA na sua forma expressa sempre apresenta um peptídeo inserido na fenda formada pelos domínios $\alpha_1 e \alpha_2$, nessa representação no entanto, o peptídeo não é mostrado. Fonte: Abbas e colaboradores (2005).

Os genes *HLA* estão localizados no Complexo Principal de Histocompatibilidade Humano, o MHC (*Major Histocompatibility Complex*) (figura 3). O MHC localiza-se no braço curto do cromossomo 6 (6p21.3), é uma das regiões com os maiores níveis de polimorfismo do genoma e apresenta uma alta concentração gênica. Cerca de 40% dos genes localizados no MHC estão envolvidos em algum processo imunológico (consórcio de seqüenciamento do MHC, 1999).

O MHC apresenta três regiões: classe I, II e III. A região de classe II, mais centromérica do complexo, estende-se por aproximadamente um megabase (Mb) e contém os genes *HLA* de classe II. A região de classe III também se estende por um Mb e apresenta genes envolvidos na resposta imune, porém não possui genes *HLA*. A região de classe I é a

mais telomérica, estende-se por dois Mb e contém os genes *HLA-A*, *-B* e *-C*, cada um dos quais com 649, 1029 e 350 alelos descritos respectivamente (Robinson, Waller *et al.*, 2003).



Figura 3. Mapa do MHC humano. O eixo horizontal marca a distância em Mb com relação ao inicio do complexo que está mais próximo do centrômero. Os genes *HLA* de classe II são mostrados em amarelo. Alguns dos genes não-*HLA* são representados em azul. Os genes *HLA* de classe I são mostrados em vermelho. As caixas mostram a localização dos genes *HLA* de classe I clássicos: *HLA-A*, *-B* e *-C*. Fonte: Janeway e colaboradores (2001), modificada.

1.2. EVIDÊNCIAS DE SELEÇÃO NATURAL NOS GENES *HLA* DE CLASSE I CLÁSSICOS

1.2.1. Evidências Populacionais de Seleção Natural nos Genes HLA

As regiões codificantes dos genes HLA-A, -B e -C apresentam características compatíveis com a atuação de seleção natural, tais como:

- Níveis de diversidade superiores ao esperado segundo o neutralismo: para HLA-B, por exemplo, existem em média 35,5 alelos por população (Meyer, Singe et al., 2007). Esses alelos estão distribuídos em freqüências intermediárias, levando os índices de diversidade a valores superiores ao esperado segundo simulações que levam em consideração o tamanho das amostras e o número de alelos (Solberg, Mack et al., 2008);
- Altos níveis de diferenças entre os alelos: a divergência média entre dois alelos dos genes HLA é de 4%, valor dez vezes superior ao observado em outros loci do genoma (Bubb, Bovee et al., 2006). Essa diferenciação é fruto dos grandes tempos de coalescências das linhagens dos genes HLA que parecem ser mais antigas que os tempos de especiação de humanos, chimpanzés e gorilas, fenômeno conhecido como polimorfismo trans-específico (Ayala, Escalante et al., 1994);
- 3. Taxas de substituição não-sinônimas são elevadas na região de ligação dos peptídeos: de maneira geral, substituições não-sinônimas alteram o aminoácido codificado por um códon, e por isso são normalmente removidas pela seleção natural. Já as sinônimas não alteram o aminoácido e não são eliminadas pela seleção. Portanto, as taxas de substituição sinônimas são geralmente mais elevadas do que as não-sinônimas. Entretanto, a situação se inverte nos códons que codificam as regiões responsáveis pela ligação dos peptídeos dos genes *HLA*: a taxa de substituição não-sinônima é mais elevada (Hughes e Nei, 1988);
- 4. Desequilíbrio de ligação elevado: os níveis de desequilíbrio de ligação na região MHC são extremamente altos (Huttley, Smith *et al.*, 1999). A seleção aumenta o desequilíbrio de ligação em torno de variantes vantajosas, elevando a freqüência de todas as mutações fisicamente ligadas a essas variantes (Gordo e Charlesworth, 2001). O resultado no MHC é que os alelos de diferentes genes *HLA* estão associados de maneira não aleatória, originando haplótipos que são muito mais comuns do que seria

esperado com base na freqüência alélica dos genes sem desequilíbrio de ligação.

Os níveis de diversidade observados nos genes *HLA* são incompatíveis com o modelo de evolução neutra, implicando que o modelo alternativo de seleção balanceadora seja mais plausível. Vários dos polimorfismos moleculares atribuídos a esse tipo de seleção ocorrem em proteínas celulares que interagem com microorganismos (Garrigan e Hedrick, 2003). Nas moléculas HLA, praticamente toda a variação localiza-se nos domínios α_1 e α_2 , nos sítios constituintes da fenda apresentadora de peptídeos (figura 4). Portanto, as regiões com maior evidência da atuação da seleção balanceadora são as que interagem diretamente com os antígenos, indicando que os microrganismos patogênicos sejam a fonte da seleção atuante nos genes *HLA*. Essa hipótese evolutiva é conhecida como seleção dirigida por patógenos.

Estudos de associação com doenças infecciosas corroboram a hipótese de seleção dirigida por patógenos. Por exemplo, foram descritas associações entre alelos *HLA* e resistência à malária em populações do oeste africano (Hill, Jepson *et al.*, 1997). Alelos de *HLA-B* foram associados a uma maior ou menor progressão à infecção por HIV (Kaslow, Carrington *et al.*, 1996; Carrington, Nelson *et al.*, 1999).



Figura 4. Distribuição da diversidade nos genes *HLA* de classe I clássicos. Acima à esquerda é mostrado um esquema de uma molécula HLA de classe I. A cadeia α é mostrada em amarelo. Acima à direita é mostrada a fenda apresentadora.de peptídeos. Abaixo é mostrada a distribuição da variação ao longo dos domínios α_1 , $\alpha_2 e \alpha_3$. Nas três representações, os resíduos com os maiores níveis de diversidade medidos pela taxa de heterozigose são marcados em vermelho (taxa de heterozigose ou *H*, é a probabilidade de se amostrar dois alelos distintos na população dada por $H=1-\sum_{i}^{n} p_i^2$, onde p_i representa a freqüência de cada i alelo). Fonte: Janeway e colaboradores (2001).

Se realmente foi a seleção dirigida por patógenos que moldou o polimorfismo dos genes *HLA*, é plausível supor que regiões geográficas com alta diversidade patogênica apresentem um maior sinal seletivo. Prugnolle e colaboradores (2005) mostraram que há uma correlação positiva entre a carga de patógenos de uma determinada região geográfica e os níveis de diversidade nos *loci HLA*, especialmente em *HLA-B*.

Todos esses resultados sustentam a hipótese de seleção dirigida por patógenos, mas ainda há debate a respeito dos mecanismos específicos de como a interação com os microorganismos moldou a variação dos genes *HLA*. Na próxima seção nós mostraremos/discutiremos alguns dos modelos propostos para explicar os níveis de polimorfismo observado nesses genes e porque o estudo de populações ameríndias é informativo para o entendimento dos processos seletivos que atuam nos genes *HLA*.

1.2.2. Modelos de Seleção Natural em Genes HLA

Dentre os vários modelos evolutivos propostos, o mais aceito é o de vantagem de heterozigotos ou sobredominância (Doherty e Zinkernagel, 1975a). Nesse modelo indivíduos heterozigotos apresentam um valor adaptativo superior aos homozigotos, pois possuem uma maior capacidade de apresentação peptídica. Takahata e Nei (1990), utilizando simulações coalescentes, demonstraram que a sobredominância é capaz de explicar os altos níveis de diversidade observados nos genes *HLA* se o valor adaptativo for igual entre os heterozigotos e superior ao de todos os homozigotos.

O apoio ao modelo de sobredominância vem sendo desafiado por estudos recentes. Trabalhos utilizando simulações computacionais re-visitaram os estudos de Takahata e Nei (1990), e alteraram o pressuposto de que todos os heterozigotos possuíam aptidão superior ao dos homozigotos e idênticas entre si. Ao invés, foi feita uma modelagem assumindo uma hipótese biologicamente mais plausível, na qual a aptidão dos heterozigotos era determinada pela soma dos valores adaptativos de cada alelo individualmente. Tal análise indicou que a sobredominância não é capaz de explicar o polimorfismo observado nos genes *HLA* (De Boer, Borghans *et al.*, 2004).

Borghans e colaboradores (2004) realizaram simulações matemáticas do modelo de co-evolução entre patógenos e sistema imunológico. De acordo com esse modelo, as mudanças em genes dos patógenos e genes imunológicos ocorreriam paralelamente: os patógenos tenderiam a acumular mutações de escape, isto é, mutações que evitassem a apresentação pelas moléculas HLA mais comuns na população de hospedeiros. Por outro lado, novas moléculas HLA que apresentassem de forma eficiente essas mutações de escape,

seriam adaptativas e aumentariam de freqüência na população de hospedeiros, recomeçando o ciclo. Esse modelo de co-evolução entre patógenos e hospedeiros se mostrou mais adequado que o de sobredominância simples, explicando de maneira satisfatória várias das características dos dados reais tal como o grande número de alelos, altas taxas de heterozigose e grande persistência dos alelos nas populações ao longo das gerações.

De acordo com o modelo de co-evolução, espera-se que populações localizadas em regiões geográficas distintas possuam alelos *HLA* que garantam a geração de respostas imunes contra os microorganismos endêmicos de cada região. O interesse em estudar os genes *HLA* em populações ameríndias está atrelado a essa idéia, uma vez que a América possui uma grande heterogeneidade de ambientes, e foi colonizada a partir de populações vindas do clima ártico, no qual o perfil de patógenos difere significativamente daquele presente nas regiões tropicais desse continente.

No presente trabalho pretendemos investigar se os regimes seletivos impostos pelo ambientes americanos foram intensos o suficiente para marcar a variação dos genes *HLA* nas populações ameríndias. Para responder essa questão é preciso entender os padrões de colonização das Américas. Nas próximas seções faremos um breve resumo dos modelos mais aceitos de ocupação desse continente, baseados em dados de diferentes disciplinas.

1.2. POPULAÇÕES AMERÍNDIAS

As populações ameríndias constituem os povos nativos do continente americano que falam línguas pertencentes ao tronco lingüístico *Ameríndio* (Mithun, 1999). Segundo essa classificação tradicional, dois outros grupos lingüísticos são encontrados em populações nativas da América do Norte: *Na-Dene* e o *Esquimó-Aleut* (Mithun, 1999).

Os ameríndios formam o grupo populacional com os menores níveis de variabilidade na espécie humana. Por exemplo, estudos com *DNA* mitocondrial (*DNAmt*) revelaram que os

essas populações apresentam níveis de variação três vezes menores que os encontrados em africanos (Salzano, 2002). Wang e colaboradores (2007), utilizando dados de 678 *loci* microssatélites, verificaram que as populações ameríndias possuem o menor número de alelos por lócus e o menor número de alelos privativos, quando comparadas ao resto do mundo. Parte desses resultados é explicada pelo efeito da colonização serial a partir da África (Salzano, 2002; Wang, Lewis *et al.*, 2007). 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.

Dentre as populações do Velho Mundo, as siberianas são as que apresentam os maiores níveis de similaridade genética com os ameríndios (Wang, Lewis *et al.*, 2007), indicando que a colonização das Américas provavelmente se deu a partir do nordeste asiático. Os cenários mais aceitos de colonização do continente americano apontam a *Beríngia* como provável região utilizada na entrada das Américas. A *Beríngia*, ponte continental que uniu a Eurásia e a América do Norte por aproximadamente 60.000 anos durante o Pleistoceno, ficou emersa até aproximadamente 10.000 anos atrás, quando os níveis oceânicos subiram mais de 120 metros devido ao fim do período glacial (Hopkins, 1982). Atualmente, o estreito de Bering apresenta mais de 100 quilômetros de extensão e separa as Américas da Sibéria.

Trabalhos que utilizaram *DNAmt* mostraram que a população ancestral dos ameríndios provavelmente entrou no continente americano numa data recente (não mais que 20.000 anos) e sofreu uma rápida expansão em direção ao sul do continente (Fagundes, Kanitz *et al.*, 2008; Kitchen, Miyamoto *et al.*, 2008). Wang e colaboradores (2007) mostraram que o alelo 275 do *locus* D9S1120 está presente em todas as 29 populações nativo-americanas analisadas (numa freqüência de 36,4% considerando a amostra total). Esse mesmo alelo está ausente nas demais populações mundiais, exceto na Sibéria. A alta freqüência com que esse alelo foi encontrado e o seu compartilhamento entre as populações ameríndias estão de acordo com o modelo de colonização a partir da Beríngia e rápida expansão.

Na América do Sul existem claramente dois padrões de distribuição de diversidade, evidenciados por vários marcadores genéticos. Dados de microssatélites do cromossomo Y (Tarazona-Santos, Carvalho-Silva *et al.*, 2001) e microssatélites autossômicos (Wang, Lewis *et al.*, 2007) mostraram que populações da região andina, apresentam valores intermediários de F_{st} , enquanto as populações do leste apresentam os valores mais altos. Esses resultados foram interpretados como evidência de que as populações da porção ocidental da América do Sul experimentaram níveis mais altos de fluxo gênico e/ou apresentaram 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.

Qual a implicação desses resultados nos estudos de seleção natural no genoma humano? As populações ameríndias enfrentaram drásticas mudança ambientais, quando partiram da região ártica a 20.000 anos atrás, e seguiram para o sul do continente. É plausível imaginar que as mudanças ambientais tenham criado pressões seletivas, inclusive sobre os genes do sistema imune. Essa hipótese explica as diferenças de composição alélica do gene *HLA-B* entre os ameríndios sul- e norte-americanos e na próxima seção nós descreveremos os trabalhos que a propuseram.

1.2.2. Variação dos genes HLA em Ameríndios

As primeiras análises da diversidade dos gene *HLA* em populações ameríndias foram baseadas em dados sorológicos. Elas demonstraram o mesmo padrão descrito para outros marcadores genéticos: as populações ameríndias apresentavam um subconjunto da diversidade encontrada no continente asiático (Kostyu e Amos, 1981). O nível de resolução dos dados sorológicos é de linhagens alélicas, portanto, podemos afirmar que as linhagens alélicas dos genes *HLA* encontradas nas Américas são um subconjunto das presentes em populações do Velho Mundo.

No início da década de 90, Belich e colaboradores (1992) e Watkins e colaboradores (1992), utilizando técnicas moleculares, descreveram uma série de alelos do gene *HLA-B* exclusivos de populações nativas da América do Sul (populações *Guarani, Kaingang* e *Waorani*). Esse padrão é contrastante com o obtido a partir dos marcadores genéticos estudados até então, inclusive com relação aos demais genes *HLA-A* e *-C*. Os novos alelos de *HLA-B* eram, em sua maioria, frutos de conversão gênica entre os alelos encontrados no Velho Mundo. Watikins e colaboradores (1992) analisaram paralelamente a população norteamericana *Zuni* e demonstraram que os alelos nela presentes são os mesmos encontrados na Eurásia. Apesar da presença de novos alelos, as populações sul-americanas possuíam o mesmo número de alelos das norte-americanas, indicando que os alelos antigos foram substituídos.

Para explicar essas observações foi proposta a hipótese do *turnover* alélico, segundo a qual a população ancestral dos ameríndios teria passado de um ambiente glacial ao norte da América do Norte para uma região tropical na América do Sul. Os trópicos apresentam uma maior biodiversidade, inclusive de microorganismos patogênicos, que por sua vez, criaram as pressões seletivas que favoreceram os novos alelos. Os novos alelos acabaram substituindo os antigos e o pequeno tamanho efetivo associado a baixos níveis de fluxo gênico entre as populações provavelmente contribuíram nesse processo (Belich, Madrigal *et al.*, 1992; Watkins, Mcadam *et al.*, 1992; Cadavid e Watkins, 1997; Parham, Arnett *et al.*, 1997).

Outros trabalhos de descrição das freqüências de alelos de genes *HLA* foram publicados na última década, porém, a hipótese do *tunover* alélico nunca foi revisitada. Nenhum trabalho testou a hipótese de haver uma correlação entre a presença de alelos exclusivos de ameríndios em freqüências elevadas e desvios com relação ao modelo de evolução neutra, algo esperado sob o pressuposto de que a presença dos alelos endêmicos seria resultado de seleção natural. A substituição dos alelos presentes em populações ancestrais por um novo conjunto, restrito aos ameríndios, é sem dúvida um resultado

consistente com o papel da seleção natural na diferenciação entre populações. Porém, cabe perguntar se, nas freqüências alélicas de populações atuais, haveria evidências genéticas de tal processo seletivo? Nosso principal objetivo foi testar se as populações nativo-americanas que apresentam repertórios alélicos muito distintos daqueles das populações asiáticas possuem desvios com relação às previsões do modelo de evolução neutro e equilíbrio populacional.

2. **OBJETIVOS**

Nesse trabalho investigamos a variação do gene *HLA-B* numa extensa amostra de populações ameríndias com o objetivo de responder às seguintes questões: qual é a distribuição dos alelos exclusivos de *HLA-B* dentro do continente americano? Onde esses alelos são encontrados em maior freqüência? Existe uma correlação entre sinais de atuação de seleção natural e a presença em freqüências elevadas de alelos exclusivos?

2.1. OBJETIVOS ESPECÍFICOS

- Seqüenciar os éxons 2 e 3 do gene *HLA-B* em populações ameríndias das Américas do Norte, Central e do Sul;
- Comparar os resultados de seqüenciamento com o de tipagem por PCR-SSOP (tipagem de produtos de PCR utilizando oligonucleotídeos sonda seqüênciaespecíficos), realizada no mesmo conjunto de amostras que seqüenciamos;
- Havendo correspondência entre os resultados obtidos pelas técnicas de PCR-SBT (tipagem baseada em seqüenciamento de produtos de reação de PCR, para detalhes da técnica ver seção 3.2) e PCR-SSOP, mesclar os dados obtidos por ambas as técnicas e descrever o polimorfísmo do gene *HLA-B* no continente americano;
- Testar a hipótese de que populações com altas freqüências de alelos endêmicos estão evoluindo de modo consistente com um modelo de neutralidade e equilíbrio populacional;

3. MATERIAIS E MÉTODOS

3.1. AMOSTRAS POPULACIONAIS

As amostras de DNA que utilizamos nesse trabalho foram obtidas através de doação de bancos de amostras pertencentes ao Laboratório de Genética Humana e Médica da Universidade Federal do Pará (LGHM-UFPA), Laboratório de Genética Molecular Humana da Universidade Federal do Paraná (LGMH-UFPR) e do *Galton Laboratory* da *University College London* (GL-UCL) (tabela 1).

Recebemos as amostras vindas dos laboratórios da UFPA e da UFPR na forma de DNA genômico extraído pelos métodos de fenol-clorofórmio, como descrito por Sambrook e Russell (2001) e *salting out* com base na técnica de Lahiri e Nurnberger (1991), modificada.

Utilizamos a técnica de amplificação de genoma inteiro através da metodologia de *multiple displacement amplification* (MDA), implementada pelo kit *GenomiPhi*, comercializado pela *GE Healthcare*. 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. As amostras vindas do GL-UCL sofreram o processo de amplificação de genoma inteiro no laboratório de origem.

Um total de 474 amostras oriundas 26 populações nativo-americanas e uma siberiana foram utilizadas nesse trabalho (figura 5 e tabela 1). Dessas amostras, 249 foram tipadas por mim durante meu mestrado utilizando a técnica de PCR-SBT e 370 foram tipadas pela pesquisadora Dra. Carolyn Hurley do Department of Oncology and Microbiology & Immunology da Georgetown University (DOMI-GU) que utilizou a técnica de PCR-SSOP, de modo que 145 amostras foram tipadas utilizando-se as duas técnicas (tabela 8).

Como resultado dessa colaboração, nós obtivemos um conjunto de dados que foi gerado em dois distintos laboratórios, utilizando duas diferentes técnicas de tipagem. Isso nos

permitiu testar a confiabilidade dos métodos de genotipagem e diminuir o risco de eventuais erros. A comparação dos resultados obtidos pelas duas técnicas será descrita na seção 4.1 (tabela 8).

 Tabela 1. Dados lingüísticos, localização geográfica e tamanho amostral das populações estudadas.

População	Localização	Língua – Grupo Lingüístico	Amostra (N)	Laboratório
Nentsi da Tundra	Sibéria/Rússia	Samoyédico - Urálico	20	GL-UCL
Chipewyan	Canadá	Dene Suline - Na Dene	25	GL-UCL
Cree	Canadá	Ojibwemowin - Algoquiano	18	GL-UCL
Ojibwa	Canadá	Almosan-Keresiouan	16	GL-UCL
Mixtec	México	Otomangue	20	GL-UCL
Zapotec	México	Otomangue	16	GL-UCL
Mixe	México	Penutian	20	GL-UCL
Quiche	Guatemala	Penutian	19	GL-UCL
Cabecar	Costa Rica	Chibcha	20	GL-UCL
Guaymi	Panamá	Chibcha	17	GL-UCL
Embera	Colômbia	Paezan	14	GL-UCL
Waunana	Colômbia	Paezan	20	GL-UCL
Zenu*	Colômbia	-	20	GL-UCL
Kogi	Colômbia	Chibcha	14	GL-UCL
Arhuaco	Colômbia	Chibcha	17	GL-UCL
Wayuu	Colômbia	Arawaque	19	GL-UCL
Ingano	Colômbia	Quéchua	15	GL-UCL
Aymara	Chile	Aymara	20	GL-UCL
Huilliche	Chile	Araucano	20	GL-UCL
Ticuna N	Brasil/AM	Ticuna-Yuri	19	GL-UCL
Ticuna S	Brasil/AM	Ticuna-Yuri	12	GL-UCL
Zoe	Brasil/PA	Tupi	9	LGHM-UFPA
Katuena	Brasil/PA	Tupi	7	LGHM-UFPA
Arara I	Brasil/PA	Karíb	10	LGHM-UFPA
Arara L	Brasil/PA	Karíb	11	LGHM-UFPA
Kaiapó	Brasil/PA	Ge	28	LGHM-UFPA
Guarani	Brasil/MS	Guarani	28	LGMH-UFPR

* Os Zenu perderam a sua linguagem, utilizando atualmente o espanhol.



Figura 5. Mapa das Américas mostrando a localização geográfica das populações amostradas nesse trabalho. A localização da população *Nentsi* também é mostrada. Fonte: Wang e colaboradores (2007), modificada (todos os demais mapas utilizados no presente trabalho apresentam essa mesma fonte).

3.2. AMPLIFICAÇÃO, PURIFICAÇÃO E SEQÜENCIAMENTO DOS ÉXONS 2 E 3 E DO ÍNTRON 2 DO GENE *HLA-B*

Todas as técnicas de tipagem do gene *HLA-B* analisaram a variação contida nos éxons 2 e 3. Nós amplificamos o segmento entre os íntron 1 e 3, que contem os éxons 2 e 3, e utilizamos uma bateria de iniciadores internos de seqüenciamento (figura 6). Inicialmente, utilizamos o protocolo do *International Histocompatibiliy Working Group* (IHWG) (FERRARA *et al.* 2003), mas não obtivemos sucesso. Nós desenvolvemos nosso próprio protocolo de amplificação, utilizando alguns dos iniciadores propostos por Ferrara e colaboradores (2003) (tabelas 3, 4 e 6).

Após a PCR, submetemos todos os produtos à eletroforese em gel de agarose 1%, corando-os utilizando brometo de etídio. A visualização e fotografia dos géis foram feitas sob luz ultravioleta. Estimamos a concentração dos produtos visualmente utilizando o marcador *Low DNA Mass Ladder (Invitrogen)*.

Submetemos os produtos de PCR com concentrações entre 20 a 30 ng/µl ou superior à purificação com as enzimas exonuclease I (*GE Healthcare*) e fosfatase alcalina de camarão (*GE Healthcare*). As concentrações dos reagentes que utilizamos no protocolo de purificação foram 1µl de cada solução de enzima para cada 10 µl de produto de PCR.

Nós seqüenciamos os produtos de PCR purificados (tabelas 5 e 6) realizando quatro leituras em cada amostra (tabela 3 e figura 6). Inicialmente, utilizamos os iniciadores seqB1, seqB2, seqB3 e ampBr para seqüenciar as amostras amazônicas. Quando começamos a seqüenciar as amostras provenientes da região andina e das Américas Central e do Norte, tivemos de desenhar os iniciadores seqB6, seqB7 e seqB9. O motivo dessa mudança foi a prevalência de alelos com deleções e inserções nessas populações (figura 6).

Aproximadamente, 1065 reações de seqüenciamento foram submetidas à corrida eletroforética em dois seqüenciadores: o ABI PRISM™ 377 (*Applied Biosystems*) do

laboratório Biologia Molecular de Plantas do Departamento de Botânica do Instituto de Biociências da Universidade de São Paulo (USP-SP) e o ABI PRISM TM 3100 (*Applied Biosystems*) do Departamento de Bioquímica do Instituto de Química da USP-SP. Nós analisamos os dados gerados pelos seqüenciadores utilizando os programas Phred, Phrap e Consed (Ewing e Green, 1998; Ewing, Hillier *et al.*, 1998; Gordon, Abajian *et al.*, 1998). Os sítios polimórficos foram indicados utilizando os códigos de ambigüidade nucleotídica propostos pela IUPAC (*International Union of Pure and Applied Chemistry*).

Para inferir o par de haplótipos presentes em cada indivíduo a partir dos dados genotípicos (que não contêm informação sobre a fase entre os sítios polimórficos), submetemos as seqüências com as posições heterozigotas à ferramenta on-line disponibilizada no dbMHC, que permite a resolução de genótipos *HLA* gerados a partir de PCR-SBT (Helmberg, Dunivin *et al.*, 2004). O primeiro passo realizado por essa ferramenta é delimitar as regiões referentes aos éxons, íntrons, 5' e 3'UTRs (regiões não traduzidas a 5' e 3' respectivamente) contidos na seqüência que submetemos. Após isso os segmentos delimitados no passo anterior são alinhados com as seqüências dos alelos contidos no banco de dados. O terceiro passo consiste na busca do par de alelos do banco de dados cuja combinação de sítios polimórficos seja correspondente à presente na seqüência submetida. Como os alelos *HLA* apresentam diferenças em várias posições ao longo da seqüência dos éxons 2 e 3 e do íntron 2, a combinação de sítios polimórficos permite a definição do genótipo do indivíduo. Nós inferimos a fase com que os polimorfismos estão ocorrendo em cada indivíduo utilizando a informação genotípica.

Tabela 2. Sequências dos oligonucleotídeos iniciadores.

Iniciador ¹	Seqüência (5'-3')
ampBf ¹	GGGAGGAGCGAGGGGACCGCAG
ampBr ¹	GGAGGCCATCCCCGGCGACCTAT
seqB1a ¹	GGGCGCAGGACCYGRGGA
seqB2 ¹	TGTGAGCCCTGGCCC
seqB3 ¹	CGAGGCCGGTGAGTGACC
seqB6 ²	GCGGATCTCGGACCCGG
seqB7 ²	GCTGACCGCGGGGGCC
seqB9 ²	GCTCCCACTCCATGAGG

"amp" e "seq" referem-se respectivamente a iniciadores utilizados na amplificação e seqüenciamento. O iniciador ampBr também foi utilizado nas reações de seqüenciamento.¹ Iniciadores obtidos do protocolo proposto por Ferrara e colaboradores (2003).² Iniciadores desenhados no presente trabalho.

Tabela 3. Concentração dos reagentes utilizados no mix de amplificação.

Reagentes	Concentrações
Tampão*	1X
Dntp	0,2mM
MgCl ₂	1,5mM
DMSO	5%
ampBf	0,2mM
ampBr	O,2mM
Taq pol.	0,1 U/ml
DNA	50 ng

* A composição do tampão 1X é: tris pH 9,0 a 0,08M, $(NH_4)_2SO_2$ a 0,02M e tween 0,1%. A sigla DMSO refere-se ao reagente dimetilsulfoxido. O volume final de reação é de 25µl.

Tabela 4. Volumes dos reagentes utilizados na reação de seqüenciamento.

Reagentes	Volumes
Produto de PCR purificado ¹	3µl
Iniciador ²	0,5 µl
Dye Terminator v.3.1 (Applied Biosystems)	2 µl
Água	4,5 µl

*Os produtos de PCR devem estar com a concentração entre 20-30 ng/ μ l. A concentração das soluções dos iniciadores é de 10 μ M.

Tabela 5. Protocolo de ciclagem.

Amplificação	Purificação	Seqüenciamento
96°C por 3min, 96°C por 25seg, 70°C por 45 seg, 72°C por 30 seg; repete-se os três últimos passos 23 vezes; 96°C por 25 seg, 70°C por 1 min, 72°C por 2 min; repete-se os três	Incubação por uma hora a 37°C, seguida de desnaturação enzimática por 15 min a 80°c.	96°C por 25 seg, 56°C por 1 min, 60°C por 4 min; repete- se os três últimos passos 40 vezes.
últimos passos 17 vezes.		



Figura 6. Esquema do posicionamento dos iniciadores utilizados na amplificação e seqüenciamento. As setas apontam a direção de extensão a partir dos iniciadores. As setas amarelas representam os iniciadores utilizados na amplificação (ampBr também passou a ser utilizado no seqüenciamento). As setas verdes representam os iniciadores utilizados no seqüenciamento das amostras cedidas pelos laboratórios LGHM/UFPA e LGMH/UFPR. As setas vermelhas representam os iniciadores utilizados no seqüenciamento das amostras cedidas pelos guenciamento das amostras cedidas pelo GL-UCL. Os círculos rosas representam as deleções e os círculos laranja representam as inserções.

3.3. MÉTODOS ANALÍTICOS

Para testar a se há desvios das proporções genotípicas esperadas sob equilíbrio de *Hardy-Weinberg* nós utilizamos o método de Guo e Thompson (1992). Testamos os desvios de neutralidade e equilíbrio populacional utilizando os seguintes testes:

A) Teste Ewens-Watteson: Watterson (1978) demonstrou que a taxa de homozigose

 $\left(F = \sum_{i}^{n} p_{i}^{2}\right)$, onde p_{i} é a freqüência de cada um dos *i* alelos amostrados, pode ser utilizada

como medida sumária do padrão de distribuição das freqüências alélicas de uma amostra. Nesse teste, a taxa de homozigose observada (F_{obs}) é comparada com a esperada (F_{esp}). O valor da taxa de homozigose esperada 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 freqüência e vários em baixa, situação compatível com seleção direcional ou expansão populacional. Amostras com $F_{obs} < F_{esp}$, apresentam alelos em freqüências 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 ou iguais ao valor observado (F_{obs}) .

B) $D \ de \ Tajima$ (Tajima, 1983): esse teste compara dois estimadores do parâmetro θ (símbolo usado para denotar o produto $4N\mu$ para os *loci* nucleares, onde N representa o tamanho da população e μ é a taxa de mutação). Um dos estimadores é o π que corresponde ao número médio de diferenças nucleotídicas entre pares de seqüências ($\theta\pi$). O outro estimador (θ S) é o número de sítios segregantes presentes na amostra, normalizados por um valor proporcional ao tamanho da amostra. Tajima (1989) demonstrou que sob evolução neutra o valor esperado da diferença entre esses dois estimadores de θ é zero, criando assim a estatística D que é definida por: $D = \frac{\theta\pi - \theta S}{\sqrt{Var(\theta\pi - \theta S)}}$. Valores positivos de D refletem a

presença de polimorfismos em freqüências intermediárias na amostra, fenômeno causado pela seleção balanceadora ou gargalo populacional recente. Valores negativos são um reflexo da presença de vários polimorfismos de baixa freqüência 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.

Todos os testes acima mencionados foram realizados implementando o pacote de programas Arlequin v.3.1 (Excoffier, Laval *et al.*, 2006)

3.4. NOMENCLATURA DOS ALELOS DO GENE HLA-B

Nós utilizamos o sistema de nomenclatura indicado pelo comitê internacional de nomenclatura de alelos *HLA* (Marsh, Albert *et al.*, 2005). Nesta seção apresentaremos sucintamente esse sistema ao leitor, para uma melhor compreensão dos resultados que apresentaremos posteriormente.

Alelos de genes *HLA* são identificados com base na seqüência de nucleotídeos em sua extensão total, incluindo as regiões codificadoras e não-codificadoras. Cada seqüência (ou haplótipo) única recebe um nome com quatro, seis ou oito dígitos. Todos os alelos recebem um nome com no mínimo quatro dígitos. Os primeiros dois dígitos referem-se aos antígenos reconhecidos sorologicamente, e definem as linhagens alélicas. Por exemplo, B*15 e B*35 são duas linhagens alélicas de *HLA-B*. Os terceiros e quartos dígitos representam os subtipos dentro de uma mesma linhagem, sendo os números dados de acordo com a ordem de descoberta das seqüências. Alelos que diferem nos primeiros quatro dígitos apresentam uma ou mais substituições nucleotídicas não-sinônimas. Por exemplo, B*1504 e B*1505 pertencem à linhagem B*15 e, portanto, possuem os mesmos antígenos reconhecidos sorologicamente, mas diferem em alguns aminoácidos.

Alelos que apresentam diferenças nucleotídicas sinônimas são diferenciados pelos quintos e sextos dígitos (B*520101 e B*520102, por exemplo). Alelos que apresentam polimorfismos localizados nos íntrons e/ou nas regiões 5' ou 3'UTRs, são diferenciados pelos sétimos e oitavos dígitos (B*15010101 e B*15010105, por exemplo).

Boa parte dos alelos do gene *HLA-B* diferem entre si nos éxons 2 e 3. Por esse motivo, os nossos esforços de genotipagem e seqüenciamento foram voltados para esses dois éxons. Entretanto, existem alelos que apresentam éxons 2 e 3 idênticos, diferindo em outras regiões da seqüência. Para esses casos, seguimos a última lista de ambigüidades alélicas disponibilizada no banco de dados do IMGT-HLA (Robinson, Waller *et al.*, 2003). Duas formas de nomenclatura foram utilizadas: 1) dígitos separados por barras, na qual os nomes dos alelos com seqüências dos éxons 2 e 3 idênticas são separados por barras (ex., B*400201/4056); 2) A introdução do código G1, no qual é especificada a linhagem à qual pertencem os alelos com éxons 2 e 3 idênticos, seguida de G1 (ex., B*51G1) (tabela 6).

Vale a pena comentar que essa ambigüidade alélica é distinta da ambigüidade genotípica. Na ambigüidade alélica, os dois haplótipos que compõem o genótipo do indivíduo são identificados, entretanto, existem vários alelos *HLA* compatíveis com os haplótipos observados. Na ambigüidade genotípica, vários haplótipos distintos, que diferem na fase em que ocorrem os polimorfismos, são compatíveis com o padrão genotípico obtido pelo seqüenciamento.

Baseando-se nas distribuições alélicas revisadas em Solberg e colaboradores (2008) e em dados de tipagem do banco de dados do IMGT-HLA (Robinson, Waller *et al.*, 2003), classificamos os alelos encontrados neste trabalho de acordo com sua distribuição nas populações humanas em três grupos: alelos cosmopolitas (alelos que provavelmente estavam presentes na população ancestral dos ameríndios, encontrados atualmente em populações asiáticas e ameríndias), endêmicos (alelos exclusivos do continente americano, frutos de eventos de conversão gênica, recombinação e mais raramente mutações de ponto envolvendo os alelos cosmopolitas) (figura 7) e não-ameríndios (alelos raros em populações ameríndias, presentes na amostra provavelmente por fluxo gênico com populações européias ou, mais raramente, com populações africanas) (tabela 7).
Grupos de alelos	Alelos possíveis	Alelo mais provável ¹
B*07G1	B*070201, B*070206, B*0744, B*0749N,	B*070201
	<i>B*0758, B*0759</i> e <i>B*0761;</i>	
<i>B*070501/0706</i>	<i>B*070501 e B*0706;</i>	B*070501
B*080101/0819N	<i>B*080101</i> e <i>B*0819N</i> ;	B*080101
B*15G1	B*15010101, B*150106, B*150107, B*9502,	B*15010101
	<i>B*9504, B*9540</i> e <i>B*9546;</i>	
B*18G1	<i>B*180101, B*180103</i> e <i>B*1817N;</i>	<i>B*180101</i> ou <i>B*180103</i>
B*270502/270504/2713	B*270502, B*270504 e B*2713;	B*270502
B*35G1	B*350101, B*350103, B*3540N, B*3542,	B*350101
	B*3557;	
<i>B*3543/3567/3579</i>	<i>B*3543, B*3567</i> e <i>B*3579;</i>	B*3543
B*40G1	<i>B*400101, B*400102</i> e <i>B*4055;</i>	B*400101
<i>B*400201/4056</i>	<i>B*400201</i> e <i>B*4056</i> ;	B*400201
B*44G1	B*44020101, B*44020102S, B*4419N, B*4427	B*44020101
<i>B*440301/440303</i>	<i>B*440301</i> e <i>B*440303</i> ;	B*440301
B*480101/4809	<i>B*480101</i> e <i>B*4809;</i>	B*480101
B*51G1	B*510101, B*510105, B*510107, B*5111N,	B*510101
	<i>B*5130, B*5132, B*5148 e B*5151;</i>	
B*550101/550103	<i>B*550101</i> e <i>B*550103;</i>	B*550101
B*580101/5811	<i>B*580101</i> e <i>B*5811;</i>	B*580101

Tabela 6. Lista de alelos com éxons 2 e 3 idênticos que encontrados neste trabalho.

¹. A escolha do alelo mais provável foi baseada em dados de freqüências alélicas de populações espalhados por todo mundo.

Tabela 7. Classificação dos alelos de acordo com a sua distribuição geográfica

Classificação	Alelos
Endêmicos	<i>B</i> *070201, <i>B</i> *1504, <u><i>B</i>*15040102</u> , <i>B</i> *1507, <i>B</i> *1520, <i>B</i> *1530, <i>B</i> *1539, <i>B</i> *350401,
	<i>B</i> *3505, <u><i>B</i>*35050102</u> , <i>B</i> *3506, <i>B</i> *350901, <i>B</i> *3511, <i>B</i> *3512, <i>B</i> *351401, <i>B</i> *3517,
	B*3519, B*352001, B*3521, B* 3523, B*3524, B*3543, <u>B*35(2458)</u> , B*3548,
	B*390202, B*3903, B*3904, B*3905, B*390601, B*390602, B*3908, B*3909,
	<u>B*39090102</u> , B*3911, B*391301, B*391902, B*4004, B*4802, B*480301,
	B*4807, B*5104, B*5110, B*511301, B*511302 e B*520102;
Cosmopolitas	B*15010101, B*1505, B*1508, B*1539, B*270502, B*350101, B*3510,
1	<i>B*39010101, B*400101, B*400201, B*4003, B*44020101, B*480101</i> e
	<i>B*510101;</i>
Não-ameríndios	B*070501, B*080101, B*1401, B*140201, B*1503, B*15170101, B*1801,
	B*2712, B*3549, B*370101, B*380101, B*3937, B*4008, B*4101, B*4201,
	B*4202, B*440301, B*47010101, B*4901, B*5001, B*5108, B*530101,
	<i>B*550101, B*570101, B*5702</i> e <i>B*580101.</i>

Alelos novos foram sublinhados, sendo os nomes não oficiais (dígito entre parênteses refere-se ao código do indivíduo em que esse alelo foi observado).

4.1. COMPARAÇÃO DOS RESULTADOS OBTIDOS PELAS TÉCNICAS PCR-SBT E PCR-SSOP

Nessa primeira seção iremos descrever os resultados da comparação dos dados de genotipagem das técnicas de PCR-SBT (realizada por mim) e PCR-SSOP (realizada no DOMI-GU). Das 145 amostras que foram tipadas com ambas as técnicas, 132 apresentaram resultados idênticos (tabelas 8 e 9). Dentre as 13 discrepâncias de resultados, sete (cinco em *Chipewyan*, um em *Qjibwa* e um em *Huilliche*) resultaram de um polimorfismo no sítio de hibridação do iniciador ampBf, utilizado na técnica PCR-SBT. Esse polimorfismo está presente nos alelos *B*44020101* e *B*4408*, tornando-os nulos na tipagem baseada em seqüenciamento, de modo que os indivíduos que os continham mostravam-se homozigotos na PCR-SBT. A presença do alelo nulo foi notada mesmo antes da comparação pois, aplicando o teste de *Hardy-Weinberg* nos dados de seqüenciamento, encontramos um excesso de homozigotos para população *Chipewyan* (dados não mostrados). Nesse caso, nós utilizamos os genótipos gerados por PCR-SSOP.

Duas amostras da população *Huilliche* identificadas como homozigotas por PCR-SSOP se mostraram heterozigotas no seqüenciamento. Em ambos os indivíduos, o alelo não observado na técnica de PCR-SSOP foi o B*3909. Essa situação é distinta da que apresentamos no parágrafo anterior, pois o alelo B*3909 não é nulo para técnica de PCR-SSOP, aparecendo em outros genótipos do conjunto dos dados. Nessa situação, a clara presença de sítios polimórficos nesses indivíduos nos indicou que o resultado de seqüenciamento estava correto e, por isso, adotamos o genótipo obtido por essa técnica.

Duas amostras da população *Ticuna S* e uma da população *Waunana* tiveram diferenças em um dos alelos inferidos. No caso de *Ticuna S*, o genótipo do indivíduo 2544 inferido por PCR-SBT foi o *B*390201:B*3903* enquanto o inferido por PCR-SSOP foi

B*390202:B*3903. Re-examinando os resultados do seqüenciamento, percebemos que havíamos ignorado a presença de um SNP (polimorfismo de nucleotídeo único) que caracteriza o alelo B*390202, confirmando o resultado de SSOP. (tabela 8). Para o indivíduo 2558, o genótipo obtido por PCR-SBT foi B*3903:B*520102 enquanto que o obtido por PCR-SSOP foi B*3903:B*520101. A diferença entre os alelos B*520101 e B*520102 também é de apenas um SNP. Re-examinando os dados de tipagem nós confirmamos o resultado obtido por PCR-SBT (tabela 8).

Para o indivíduo 2585 da população *Waunana*, o genótipo inferido por PCR-SBT foi *B*400201:B*4004* enquanto o inferido por PCR-SSOP foi *B*400201:B*4064*. A diferença entre os alelos *B*4004* e *B*4064* também é de apenas um SNP. Nós conferimos os resultados obtidos por ambas as técnicas e confirmamos o resultado de PCR-SBT, corrigindo o genótipo inferido por PCR-SSOP (tabela 8).

Em apenas uma amostra (2021, da população *Cabecar*), houve discrepância completa do genótipo inferido pelas duas técnicas. O genótipo inferido pelo seqüenciamento é extremamente comum nessa população, enquanto que o genótipo inferido pela técnica PCR-SSOP é extremamente raro. Além disso, em comunicação pessoal, a aluna Kelly Nunes, do programa de pós-graduação do departamento de Genética de Biologia Evolutiva da USP-SP, que está gerando dados de microssatélites adjacentes ao gene *HLA-B* nessas mesmas amostras, detectou a presença de variantes microssatélites comuns em ameríndios nesse indivíduo (dado não mostrado). Por esse motivo nós supomos que o genótipo obtido pela técnica de PCR-SSOP seja fruto de erro de identificação da amostra. E consideramos o genótipo obtido por seqüenciamento.

A grande concordância entre os resultados obtidos por ambas as técnicas nos deu confiança para unir os dois conjuntos de amostras. As análises que mostraremos nas próximas seções foram realizadas com a base de dados na qual foram reunidos os resultados de PCR-SSOP e PCR-SBT.

População	PCR-SBT $(N)^1$	PCR-SSOP $(N)^2$	Sobreposição (N) ³	Discrepâncias (N) ⁴
Nentsi da Tundra	-	20	-	-
Chipewyan	20	25	20	5
Cree	12	18	12	-
Ojibwa	8	16	8	1
Mixtec	-	20	-	-
Zapotec	-	16	-	-
Mixe	-	20	-	-
Ouiche	-	19	-	-
Cabecar	18	20	18	1
Guavmi	6	17	6	-
Embera	-	14	-	-
Waunana	12	20	12	1
Zenu	14	15	9	-
Kogi	-	14	-	-
Arhuaco	8	17	8	-
Wayuu	3	19	3	-
Ingano	-	15	-	-
Aymara	20	20	20	-
Huilliche	17	17	14	3
Ticuna N	13	18	12	-
Ticuna S	5	10	3	2
Zoe	9	-	-	-
Katuena	7	-	-	-
Arara I	10	-	-	-
Arara L	11	-	-	-
Kaiapó	28	-	-	-
Guarani	28	-	-	-
Total	249	370	145	13

Tabela 8. Número de indivíduos tipados em cada técnica e discrepâncias de resultados.

^{1.} Numéro de amostras tipadas utilizando a técnica PCR-SBT; ^{2.} Número de amostras tipadas pela técnica PCR-SSOP; ^{3.} Número de amostras tipadas por ambas as técnicas. ⁴ Número de amostras com discrepâncias entre os resultados de PCR-SBT e PCR-SSOP.

Tabela 9. Indivíduos com discrepância entre os resultados de PCR-SBT e PCR-SSOP.

Indivíduos	População	Genótipo PCR-SBT	Genótipo PCR-SSOP	Resolução
2382	Chipewyam	B*400201:B*400201	B*400201:B*44020101	B*400201:B*44020101
2384	Chipewyam	B*39010101:B*39010101	B*39010101:B*44020101	B*390101:B*44020101
2390	Chipewyam	B*15010101:B*15010101	B*15010101:B*44020101	B*15010101:B*44020101
2395	Chipewyam	B*15010101:B*15010101	B*15010101:B*44020101	B*15010101:B*44020101
2396	Chipewyam	B*270502:B*270502	B*270502:B*44020101	B*270502:B*44020101
2423	Ojibwa	B*350101:B*350101	B*350101:B*4408	B*350101:B*4408
2126	Huilliche	B*510101:B*510101	B*44020101:B*510101	B*44020101:B*510101
2129	Huilliche	B*3909:B*440301	B*440301:B*440301	B*3909:B*440301
2133	Huilliche	B*39010101:B*3909	B*39010101:B*39010101	B*39010101:B*3909
2544	Ticuna S	B*390201:B*3903	B*390202:B*3903	B*390202:B*3903
2558	Ticuna S	B*3903:B*520102	B*3903:B*520101	B*3903:B*520102
2585	Waunana	B*400201:B*4004	B*400201:B*4064	B*400201:B*4004
2021	Cabecar	B*350101:B*3543	B*180101:B*380101	B*350101:B*3543

4.2. VERIFICAÇÃO DO EQUILÍBRIO DE *HARDY-WEINBERG (HW)*

Observamos desvios significativos do equilíbrio de *HW* nas populações *Kogi* e *Ticuna N* (tabela 10). Na população *Kogi* nós observamos um excesso de heterozigotos com relação ao esperado, enquanto que em *Ticuna N* observamos um excesso de homozigotos. O excesso de homozigotos em *Ticuna N* poderia ser explicado pela presença de um alelo nulo, resultante de um polimorfismo na região de hibridação de um dos iniciadores da PCR. Esse polimorfismo resultaria na não amplificação das amostras com essa mutação em homozigose, enquanto que os indivíduos com a mutação em heterozigose seriam genotipados como homozigotos para o segundo alelo. Como os *Ticuna N* foram genotipados utilizando as técnicas PCR-SSOP e PCR-SBT, que utilizaram iniciadores diferentes, é improvável que esse desvio seja conseqüência da presença de um alelo nulo.

Uma segunda explicação para o excesso de homozigotos em *Ticuna N* seria uma alta taxa de endogamia nessa população. Se o desvio observado fosse causado por endogamia, o genoma como um todo desses indivíduos apresentaria desvios com relação às expectativas de *Hardy-Weinberg*, aumentando a taxa de homozigose de vários *loci*. Sendo essa a provável causa do desvio que observamos nessa população.

O excesso de indivíduos heterozigotos que observamos na amostra da população *Kogi* não pode ser explicado por problemas metodológicos. É provável que esse desvio tenha sido gerado no processo de amostragem dessa população. Como mostraremos na seção 4.4., essa população não apresentou uma taxa de heterozigose maior que a esperada segundo o modelo de evolução neutra, mostrando que apesar da presença de um excesso de indivíduos heterozigotos, não há sinais de atuação de seleção natural.

Populações	H_o^{-1}	H_e^2	p^3
Tundra Nentsi	0,93750	0,89113	0,83290
Chipewyan	0,88000	0,90618	0,24809
Cree	0,83333	0,87619	0,58542
Ojibwa	0,87500	0,88508	0,86199
Mixtec	0,95000	0,88590	0,84140
Zapotec	1,0000	0,95565	1,00000
Mixe	0,90000	0,82949	0,73938
Quiche	0,94737	0,94737	0,87136
Cabecar	0,75000	0,70769	0,97416
Guaymi	0,94118	0,70588	0,45038
Embera	0,78571	0,78307	0,62139
Waunana	0,85000	0,87051	0,20802
Zenu	0,90000	0,85641	0,77387
Kogi	0,85714	0,71958	<u>0,02453</u>
Arhuaco	0,82353	0,75579	0,57514
Wayuu	0,94737	0,90043	0,89372
Ingano	0,86667	0,89195	0,40620
Aymara	0,85000	0,83718	0,89442
Huilliche	0,78947	0,83926	0,38452
Ticuna N	0,52632	0,76956	<u>0,00686</u>
Ticuna S	0,83333	0,87681	0,89525
Zoé	0,77778	0,83007	0,67351
Katuena	0,57143	0,89011	0,07610
Arara I	0,30000	0,47895	0,47982
Arara L	0,76923	0,88923	0,62494
Kaiapó	0,67857	0,83506	0,10446
Guarani	0,78571	0,91494	0,13913

Tabela 10. Resultados do teste de equilíbrio de Hardy-Weinberg.

¹ taxa de heterozigose observada; ² taxa de heterozigose esperada; ³ valor de *p*.

4.3. DESCRIÇÃO DA VARIAÇÃO DO GENE HLA-B

Encontramos um total de 90 alelos, com 41 sítios polimórficos localizados no éxon 2, 16 no íntron 2 (três deles sendo polimorfismos de inserção-deleção) e 38 sítios no éxon 3, totalizando 95 sítios polimórficos em aproximadamente 791 pares de base. Detectamos três novos alelos caracterizados por mutações localizadas no final do íntron 2. Os novos alelos diferem de B*1504, B*3505 e B*3909 nas posições 514 (t \rightarrow g), 496 (g \rightarrow t) e 479 (g \rightarrow t), respectivamente (essas posições são relativas ao primeiro nucleotídeo do éxon 2). Detectamos um quarto alelo caracterizado por uma transição T \rightarrow C no sítio 716 (éxon 3, posição também relativa ao primeiro nucleotídeo do éxon 2) com relação ao alelo *B*3543*, levando à substituição não-sinônima Trp156Arg.

Dentre os 259 indivíduos que seqüenciamos, havia 73 indivíduos com ambigüidade genotípica. As ambigüidades foram resolvidas utilizando o seguinte procedimento: buscamos dentre os indivíduos com genótipos não-ambíguos aqueles que apresentaram alelos compatíveis com algum dos possíveis genótipos dos indivíduos com ambigüidade. Por exemplo, quatro indivíduos Kaiapó possuíam o genótipo B*1504:B*3505, que apresenta o mesmo padrão de posições heterozigotas que o genótipo B*1507:B*3537. O alelo B*1504 foi encontrado em homozigose em dois indivíduos e em heterozigose sem ambigüidade em cinco indivíduos Kaiapó. O alelo B*3505 também foi encontrado em homozigose em dois indivíduos e em heterozigose sem ambigüidade em oito indivíduos Kaiapó. Os alelos B*1507 e B*3537, por outro lado, estavam completamente ausentes dos genótipos não-ambíguos da mesma amostra, permitindo que inferíssemos o genótipos dos indivíduos como sendo B*1504:B*3505. Esse critério foi aplicado em todas as populações, e as comparações entre os genótipos ambíguos com os homozigotos ou genótipos não ambíguos sempre foram realizadas dentro de cada população. Com exceção de um indivíduo Guarani, que foi descartado das análises, todas as ambigüidades foram resolvidas utilizando essa estratégia, não sendo necessária a realização de clonagens para a inferência de fase e resolução das ambigüidades genotípicas. A inferência dos genótipos com ambigüidade dos dados obtidos PCR-SSOP foi realizada independentemente, havendo total correspondência entre as inferências realizadas a partir dos dados obtidos por ambas as técnicas.

4.3.1. Distribuição e Freqüências Alélicas

Nessa seção faremos a descrição das freqüências e distribuição geográfica dos alelos que observamos. Boa parte da descrição que faremos é baseada na classificação dos alelos como endêmicos, cosmopolitas e não-ameríndios (tabela 7). Por esse motivo, gostaríamos de comentar sobre a classificação dos alelos *B*140201*, *B*180101* e *B*4901* como nãoameríndios. Esses alelos estão em freqüências elevadas em populações européias e africanas e são raros no leste e nordeste asiáticos. Por esse motivo, quando os observamos nas populações ameríndias, consideramos uma provável origem por fluxo gênico recente com populações européias ou africanas. Nós observamos esses alelos nos *Nentsi da Tundra*, única populações não-americana que incluímos nesse trabalho, localizada na região central da Sibéria, próxima aos Montes Urais (tabela 11). A presença desses alelos nos *Nentsi da Tundra* não invalida a nossa classificação dos alelos como não-ameríndios, pois há um gradiente de diminuição de freqüência de *B*140201*, *B*180101* e *B*4901* desde a Europa até o leste da Ásia (Solberg, Mack *et al.*, 2008). Por esse motivo, a presença desses alelos nos *Nentsi da Tundra* pode ser explicada pela existência desses gradientes.

Alelos Nentsi da Tundra	Freqüência (2N=50)
<i>B</i> *070201 ^C	26,3%
$B*130201^{A}$	5,3%
$B^*140201^M$	2,6%
<i>B*180101^C</i>	5,3%
$B^{*}270502^{C}$	2,6%
<i>B*350101^C</i>	5,3%
<i>B*39010101^C</i>	7,9%
$B^{*}400201^{C}$	5,3%
$B^{*}44020101^{C}$	10,5%
<i>B*480101^C</i>	18,4%
$B^{*}4901^{M}$	2,6%
<i>B*510101^C</i>	7,9%
Alelos não-compartilhados ⁴	5,3%
Total de cosmopolitas ^C	84,2%
Alelos não-Asiáticos ^M	10,5%

Tabela 11. Distribuição alélica na população Tundra Nentsi.

^A. Alelos que aparentemente não entraram nas Américas; ^C. Alelos que provavelmente estavam presentes na população ancestral dos ameríndios; ^M. Alelos provavelmente de origem européia.

		Éxon 2	Íntron 2	2	Éxon 3
	111111	11111222222222222222	233344455	555555555555555555555555555555555555555	555556666666777777777777777777777777777
	233667992333344 403891240123609	478890001122334444 936894790989680146	866879911 702796824	333334444568	38999145679011123356888 79067315470567827801138
<i>в*070201</i>	TTGCTACAGAGAAG	GGACAGCGGGGAGGACTGG	cctggtgcg	CCCTCGCACGGC	TGTAGCCGAGACGGGGGAGTCAG
B*15010101 B*1504		A.GGCAACAT	t.ca.gt	TTCCCC T	C.AGTCTGCC
B*15040102	.GA.GGATCC	A.GGCAACAT	t.ca.gg	.TTGGCG.T	C.AGTCTGCC
B*1507	.GA.GGATCC	A.GGCAACAT	??????????????????????????????????????	C	C.AGTCTGCC
B*1520 B*1530	GA.GGATCC	A.GGCAACAT	2222222222	T.AG.TCC.	C.AGCT.TCTGCC
B*1539	.GA.GGATCC	A.GGCAACAT	t.ca.gt	G	C.AGTCTGCC
B*350101	.GA.GGAC.CA	ATAACAT	g	T.AG.TCC.	C.AGCT.TCTGCC
B*350401	.GA.GGAC.CA	ATAACAT	g	T.AG.TCC.	
B*3505 R*35050102		ΑΙΑΑCΑΤ ΔΤΔΔCΔΤ	g	· · · · · · · · · · · · · · · · · · ·	
B*3506	.GA.GGAC.CA	ATAACAT	g	T.AG.TCC.	A.T.AGCT.TCTGCC
B*350901	.GA.GGAC.C/	ATAACAT	g	T.AG.TCC.	<mark>A</mark> GCT.TCTGCC
B*3510 R*3511		A.GGIAACAI	g		
B*3512	.GA.GGAC.C/	ATAACAT	q	T.AG.	
B*351401	.GA.GGAC.CA	ATAACAT	???????????????????????????????????????	T.AG.TCC.	A .C.AGCTCTGCC
B*3517	.GA.GGAC.C	ATAACAT	777777777	T.AC.	C.AGCT.TCTGCC
B*352001			y		
B*3521	.GA.GGAC.CA	ATAACAT	g	T.AG.TCC.	C.AGC A .TCT. C GCC
в*3523	.GA.GGAC.C	ATAACAT	???????????????????????????????????????	T.AG T TCC.	C.AGCT.TCTGCC
B×3524 p*2512	.GA.GGAC.CA	ATAACAT	· · · · · ·	T.AG.TCC.	C.AGCT.TCT. C GCC
B*35(2458)	.GA.GGAC.C	ATAACAT		G	
B*3548	.GA.GGAC.CA	ATAACAT	???????????????????????????????????????	T.AGCTCC.	C.AGCT.TCTGCC
B*39010101		.AGAACA	g	G	A.T.AGCT.T.AACGCC
B*390202 B*3903			g		
B*3904	. GA G	.AGAACA	???????????????????????????????????????	Ğ	
в*3905		. A GAACA	g <u></u>	G	A.T.AGCT.T.AACGCC
B*390601	G	.AGAACA	g <mark>t</mark>	TTGGCG.	A.T.AGCT.T.AACGCC
B*3908	G	AGGCAACAT	777777777	G	
B*3909	G	.AGAACA	g	GCT	A.T.AGCT.T.AACGCC
B*39090102	G	.AGAACA	t g		A.T.AGCT.T.AACGCC
B*3911 B*391301	222 c	A. GAACAL		·····G····	
B*391902	CCGAC	.AGAACA	q	G	
<i>B*400201</i>	CACGTAGAA	A.GGCAACAT	cg.g.		<u>A</u>
B*4003 P*4004	CACGTAGAA	A.GGCAACAT	cg.g.		
B*480101	CACGTAGA	GGCAACAT		1.AG.IC.	
B*4802		GGCAACAT	<mark>g</mark>	T.AG.TCC.	C.AGCT.TCTGCC
B*480301		GGCAACAT	<mark>gt</mark>	G	···A·_·····TTT·T·······
B*4007 B*510101	GATG GAC - CA		at	TTGGCG T	
B*5104	.GATGGAC.CA	ATAACAT.A.TGCTC	g	T.AG.T	
B*5110	.GATGGAC.CA	ATAACAT.A.TGCTC	gt	.TTGGCG.T	
B*511301 R*511302	.GATGGAC.CA	ATAACAT.A.TGCTC	gt	.TTGGCG.T	
B*520102	.GATGGAC.CA	A. GGCAACAT.A. TGCTC	at	.TTGGCG.T	
5 520202					
Legenda:	Sequência do alel	0 R*15010101			
Legenda.	Sequencia do alei	0 B 13010101			
	Seqüência do alel	o <i>B* 350101</i>			
	Seqüência do alel	o <i>B*39010101</i>			
	Seqüência do alel	o <i>B*400201</i>			
	Seqüência do alel	o <i>B*480101</i>			
	Sequência de alal	~ P*510101			
	Sequencia do alei				
	(trechos em branco): Mais de um doade	or possivel		
	G (bases em negrito):	Mutações de ponto		nonhara da	manafaraia das 1
A	GUI (treenos sublinhado	s): 1 recnos nao enc	unrados em	nennum dos	possivels doadores

Figura 7. Alinhamento dos alelos cosmopolitas e endêmicos. São mostrados os trechos mínimos de conversão necessários para formação dos alelos endêmicos. Os prováveis alelos cosmopolitas envolvidos nos eventos de conversão são mostrados em cores na lista de alelos. Os alelos B*3909 e B*391902 apresentam trechos de seqüência não compatíveis com nenhum dos possíveis doadores (inclusive entre outros alelos cosmopolitas que não foram mostrados nessa figura). O alelo cosmopolita B*070201 foi utilizado como referência.

As populações *Chipewyan*, *Cree* e *Ojibwa* localizam-se no norte da América do Norte e apresentam a maior proporção de alelos cosmopolitas da amostragem (figuras 8 e 9). Dentre as populações que analisamos, *Chipewyan* é a única não pertencente ao tronco lingüístico ameríndio. Essa população é pertencente ao estoque lingüístico *Na-Dene* e, curiosamente, é a única onde não encontramos nenhum alelo endêmico. (figura 8). O alelo *B*350101*, que se encontra em altas freqüências nas populações do leste e nordeste asiáticos (Solberg, Mack *et al.*, 2008), é o mais comum em *Chipewyan* (16%), seguido de *B*270502* e *B*400201* (ambos com freqüência de 14%)(figura 8). O alelo *B*270502* é encontrado em freqüências elevadas nas populações ao norte do hemisfério norte e o alelo *B*400201* apresenta altas freqüências no nordeste asiático (Solberg, Mack *et al.*, 2008). Encontramos os alelos *B*44020101*, *B*39010101* e *B*510101* com freqüências superiores a 10%, resultando num excesso de alelos em freqüências intermediárias nos *Chipewyan*, como mostraremos nos resultados dos testes de neutralidade (tabela 13).

Diferentemente do que ocorre nos *Chipewyan*, as populações *Cree* e *Ojbwa* apresentam dois alelos em maior freqüência, seguidos de vários em freqüências mais baixas. O alelo mais comum encontrado na população *Cree* foi *B*510101* (30,6%), seguido de *B*350101* (16,7%). *B*510101* é um dos alelos com mais ampla distribuição mundial, encontrado na África, Europa e Ásia (Solberg, Mack *et al.*, 2008). Nós encontramos um alelo endêmico nessa população, *B*390601* (2,8%) (figura 8). Os alelos *B*350101* e *B*400201* são os mais comuns da população *Ojibwa* (25 e 15,6%, respectivamente). O alelo *B*4807* foi o único endêmico que encontramos nessa população (6,3%) (figura 9).

Diferentemente do que ocorre no norte do continente, os alelos endêmicos são freqüentes nas populações *Mixe*, *Mixtec* e *Zapotec*. O alelo mais comum da população *Mixe* é o endêmico *B*390202* (32,5%), seguido de *B*3512* (15%), *B*400201* (15%) e *B*350101* (12,5%). A freqüência total de alelos endêmicos nessa população foi de 60% (figura 9).







Figura 9. Freqüências alélicas das populações *Ojibwa* e *Mixe*. Os expoentes A^E, A^C e A^M indicam os alelos endêmicos, cosmopolitas e não-ameríndios respectivamente.

O alelo B^{*3512} é o mais freqüente da população *Mixtec* (25%), seguido de B^{*3905} (17,5%) e $B^{*350101}$ (12,5%). B^{*3905} é um alelo endêmico e o total de alelos endêmicos observados nos *Mixtec* foi de 62,5% (figura 10). Na população *Zapotec*, os três alelos mais comuns foram B^{*3512} , $B^{*390202}$ e B^{*3905} (todos na freqüência 12,5%). A soma das freqüências dos alelos endêmicos nessa população foi de 75%.

Na Guatemala nós encontramos a população *Quiche*. O alelo mais comum dessa população é o cosmopolita B*350101 (18,4%), seguido do endêmico B*3512 (10,5%). Assim como os *Zapotec*, os *Quiche* apresentam uma alta proporção de alelos não-ameríndios (21,1%). A freqüência total de alelos endêmicos (42,1%) é superior à de cosmopolitas (36,8%) (figura 11).

A freqüência somada dos alelos $B^{*350101}$, B^{*3543} e $B^{*400201}$ nas populações Cabecar e Guaymi é de aproximadamente 90%. O alelo mais comum nos Cabecar é o cosmopolita $B^{*400201}$ (47.5%), seguido do endêmico B^{*3543} (22.5%) e dos cosmopolitas $B^{*350101}$ e $B^{*15010101}$ (10%) (FIGURA 11). Nos Guaymi, o alelo mais comum é o $B^{*350101}$ (44.1%), seguido de B^{*3543} (29.4%) e $B^{*400201}$ (17.6%). O único alelo endêmico que observamos em ambas as populações foi B^{*3543} .

Na população Zenu, o alelo mais comum foi o cosmopolita B*400201 (27,5%), seguido dos endêmicos B*5110 (20%) e B*3543 (15%). Nós encontramos uma alta proporção de alelos não-ameríndios nessa população (22,5%), sendo o alelo europeu B*3549 o mais comum deles (12,5%) (figura 12). Os *Embera* apresentam um perfil alélico completamente distinto dos Zenu, apesar da proximidade geográfica: o endêmico B*3905 ocupa quase metade da amostragem (42,9%), seguido do endêmico B*3510 (17,9%). A freqüência somada dos alelos endêmicos foi de aproximadamente 90% nessa população e não encontramos alelos não-ameríndios (figura 12). Na população *Kogi*, o alelo que ocupa praticamente a metade da amostragem é o B*3543 (42,9%) seguido do cosmopolita B*400201 (32,1%) (figura 13).

		Mixtec Mixe Zapotec Quiche Cabecar	América Central Kogi Zenu Wayuu
Alelos Mixtec	Freqüência (2N=40)	- \	Embera S
B*350101 ^c	12.5%	V V	Vaunana 1 grow 7 G
B*3512 ^E	25.0%	4	
B*351401 ^E	2.5%	Alelos Zapotec	Fregüência (2N=32)
B*3517 ^E	2.5%	B*15010101 ^c	3.1%
B*3523 ^E	10.0%	B*1517 ^M	3.1%
B*380101 ^M	2.5%	B*1530= B*2501010	6.3% 3.494
B*390101010	7.5%	B*3512E	12.5%
B*390202E	2.5%	B*351401E	9.4%
B*3905E	17.5%	B*3517 ^E	6.3%
B*200602E	5.0%	B*3524 ^E	3.1%
B 390002-	5.0%	B*3543 ^E	3.1%
B*4002010	1.5%	B*390202E	12.5%
B*4011 ¹	2.5%	B*3905-	12.5%
B*510101 ^c	2.5%	- B*4027 ^M	3.1%
Total de Endêmicos ^E	65.0%	B*47010101M	3.1%
Total de cosmopolitas ^c	30.0%	B*510101C	3.1%
Alelos não-ameríndios ^M	5.0%	B*520102E	6.3%
		B*5301 ^M	3.1%
		B*580101 ^M	3.1%
		Total de Endêmicos ^E	75.0%
		<u></u>	9.4%
		Alelos nao-amerindios ^w	15.6%

Figura 10. Freqüências alélicas das populações *Mixtec* e *Zapotec*. Os expoentes A^E, A^C e A^M indicam os alelos que são endêmicos, cosmopolitas e não-ameríndios, respectivamente.

	2.2	Mixtec	xe	The second	América Central
Nalas Ouisha	Za	apotec	Par 1	Arhuac	0
Alelos Quicne	Frequencia (2N=38)	– 🔶 Quicl	he Y ! .	Kogi	Wayuu .
B 180101"" D*2705026	0.3%		25	Zenu/	-insi
B 270302°	2.076		Cabecar	104	2
B 350101°	18.4%		/ Guavmi	2.6.	5
B 3512-	10.5%			Embe	ran L
B 301/-	2.6%		/ Waun	ana	1 mg
B*3543*	5.3%			1	I 7 (
B*3548-	2.6%	/	•		
B*380101%	5.3%	/-	Alelos Guaymi		Freqüência (2N=34)
B*390202	5.3%	/ -	B*15010101C		2.9%
B*3905	7.9%	/	B*350101 ^c		44.1%
B*390602E	2.6%	/	B*3543E		29.4%
B*3908⊧	2.6%	/	B*3549 ^M		2.9%
B*4002010	5.3%	/	B*400201C		17.6%
B*4101 ^M	2.6%	/	B*510101C		2.9%
B*4201 ^M	2.6%	/ -	Total de Endêmico	s ^E	29.4%
B*44020101 ^c	5.3%	/ -	Total de cosmopol	itas ^c	67.6%
B*4403 ^M	2.6%	/ -	Alelos não-amerín	dios ^M	2.9%
B*510101 ^C	5.3%	+			
B*511302 ^E	2.6%				1000 X100 0.
B*570101M	2.6%	Alelos Cab	ecar	Freqüência	(2N=34)
Total de Endêmicos ^E	42.1%	- B-1501010	10	10.0%	2
Total de cosmopolitas ^c	36.8%	- B*3501010		15.0%)
Alelos não-ameríndios ^M	21.1%	- B°3543 ^E		22.5%	2
	525 CA2600 5640 50	- B*4002010		47.5%	D
		B"4707"		2.5%	
		B*5/02	-dânsies -F	2.5%	
		I OTAL de Er	aemicos ²	22.5%	<u>) </u>
			smopolitas	(2.5%	<u>) </u>
		Alelos nao	-amerinaios"	5.0%	

Figura 11. Freqüências alélicas nas populações *Quiche, Cabecar* e *Guaymi*. Os expoentes A^{E} , A^{C} e A^{M} indicam os alelos endêmicos, cosmopolitas e não-ameríndios, respectivamente.



Figura 12. Freqüências alélicas das populações *Zenu* e *Embera*. Os expoentes A^E , A^C e A^M indicam os alelos que são endêmicos, cosmopolitas e não-ameríndios, respectivamente.

A população *Waunana*, assim como os *Embera*, apresenta maior parte da amostragem compreendida por alelos endêmicos (65%). Apesar disso, o alelo mais comum é o cosmopolita B*400201 (25%), seguido dos endêmicos B*4004 (17,5%), B*1504 (15%) e B*3905 (15%) (figura 14). Na população *Ingano, o* alelo B*400201 também é o mais freqüente, seguido de B*15010101 (16,7%) e B*3543 (13,3%) (figura 14).

Nos *Arhuaco*, o alelo B^{*3543} também é encontrado em freqüência elevada (44,1%), seguido do alelo novo que denominamos aqui $B^{*35(2468)}$ e do alelo não ameríndio $B^{*180101}$ (ambos na freqüência de 17,6%). A população *Arhuaco* foi a única sul-americana na qual o alelo $B^{*270502}$ foi detectado (figura 13). Até esse resultado, a linhagem B^{*27} seria a única dentre as linhagens presentes nas populações norte-americanas que está ausente nas sul-americanas. Como já comentamos, $B^{*270502}$ é encontrado em freqüências elevadas em populações do norte do hemisfério norte, havendo a possibilidade de que a sua presença na população *Arhuaco* seja devido a fluxo gênico com europeus. Entretanto, não podemos descartar a possibilidade desse alelo ter chegado a América do Sul juntamente com a população ancestral.

O alelo mais comum observado na população *Wayuu* foi *B*400201* (23,7%), seguido de *B*3512* e *B*180101*. Assim como as populações *Arhuaco* e *Zenu*, os *Wayuu* apresentam uma elevada proporção de alelos não ameríndios na amostragem, principalmente *B*180101*. A população *Wayuu* apresenta uma grande quantidade de alelos endêmicos em baixa freqüência (figura 13).

As populações *Aymara* e *Huilliche* habitam respectivamente o norte e sul do Chile. Os alelos mais comuns em ambas as populações são endêmicos das Américas, seguidos de alelos cosmopolitas. Nos *Aymara*, o alelo mais freqüente é *B*3505* (37,5%), seguido do cosmopolita *B*480101* (15%). Nos *Huilliche*, o alelo mais comum é o endêmico *B*3909*, seguido do cosmopolita *B*510101* (figura 15).



Figura 13. Freqüências alélicas das populações *Kogi, Arhuaco* e *Wayuu*. Os expoentes A^{E} , A^{C} e A^{M} indicam os alelos que são endêmicos, cosmopolitas e não-ameríndios, respectivamente.







Figura 15. Freqüências alélicas das populações *Aymara* e *Huilliche*. Os expoentes A^E , A^C e A^M indicam os alelos que são endêmicos e cosmopolitas respectivamente.

Todas as populações sul-americanas cujos resultados apresentamos acima estão localizadas a oeste dos Andes. Dentre as populações localizadas ao leste dos Andes, os *Ticuna* habitam a região mais ocidental. As duas populações Ticuna que estudamos estão separadas pelo Rio Amazonas, uma localizada a norte (*Ticuna N*) e outra a sul (*Ticuna S*). O alelo mais comum em *Ticuna N* e *Ticuna S* é o cosmopolita B*400201 (44,7% e 29,2%, respectivamente). Em *Ticuna N*, o segundo alelo mais freqüente é o endêmico B*350401 (15,8%), enquanto que em *Ticuna S*, o segundo alelo mais freqüente é o endêmico B*3903, seguido de B*350401 (16,7% e 12,5%, respectivamente). Em ambas as populações os alelos endêmicos são mais comuns e há uma baixa evidência de fluxo gênico (figura 16).

O alelo mais freqüente na população *Zoé* é o cosmopolita *B*510101* (33,3%), seguido do endêmico *B*4802* (22,2%) e do cosmopolita *B*400201* (16,7%) (figura 17). O pequeno tamanho amostral da população *Katuena* não nos permitem realizar fortes inferências a respeito das distribuições alélicas. Entretanto, fica claro que essa população apresenta uma alta proporção de alelos endêmicos das Américas (figura 17).

Nós trabalhamos com duas populações da etnia *Arara: Arara do Iriri* e *Arara do Laranjal.* A população *Arara do Laranjal* apresenta a maior freqüência de alelos B^{*15} , pertencendo a essa linhagem os três alelos mais freqüentes: B^{*1504} (23,1%), B^{*1508} (19,2%) e B^{*1520} (15,4%). Desses três, B^{*1508} é o único cosmopolita, presente em populações norte-americanas e asiáticas (figura 18). Os indivíduos da população *Arara do Iriri* são descendentes de um casal de irmãos da tribo *Arara do Laranjal.* O genótipo dos indivíduos fundadores é conhecido, um deles seqüenciado no nosso trabalho, e o segundo inferido a partir do heredograma da tribo. Os genótipos dos fundadores eram $B^{*1520}(B^$

Praticamente todos os alelos observados nos indivíduos da população Kaiapó são endêmicos das Américas. Nessa população os três alelos mais freqüentes são os endêmicos B*3505 (26,8%), B*1504 (23,2%) e B*520102 (17,9%). O único alelo cosmopolita observado nessa população foi B*1508 (3,6%) (figura 18). Dentre as populações a leste dos Andes, os *Guarani* são os mais meridionais. Os alelos endêmicos representam 75% da freqüência total, sendo B*1504 (19,6%), B*3905 (16,1%) e B*4004 (10,7%) os mais comuns (figura 19).



Figura 16. Freqüências alélicas das Populações Ticuna N é Ticuna S. Os expoentes A^E , A^C e A^M indicam os alelos que são endêmicos, cosmopolitas e não-ameríndios, respectivamente.



Figura 17. Freqüências alélicas das populações *Zoé* e *Katuena*. Os expoentes A^E, A^C e A^M indicam os alelos que são endêmicos, cosmopolitas e não-ameríndios, respectivamente.



Figura 18. Freqüências alélicas das populações *Arara do Iriri, Arara do Laranjal* e *Kaiapó*. Os expoentes A^E , A^C e A^M indicam os alelos que são endêmicos, cosmopolitas e não-ameríndios, respectivamente.



Figura 19. Freqüências alélicas da populações *Guarani*. Os expoentes A^E, A^C e A^M indicam os alelos que são endêmicos, cosmopolitas e não-ameríndios, respectivamente.

Depois de observarmos o padrão de distribuição de freqüências alélicas ao longo do continente (figura 20), resolvemos testar a presença de um gradiente norte-sul de aumento de freqüência dos alelos endêmicos nas Américas. Nós calculamos os valores dos coeficientes de determinação (r²) entre a distância em quilômetros (km) de cada população com relação ao Estreito de *Bering* e a freqüência de alelos endêmicos (figura 21). As distâncias entre as populações e o Estreito de *Bering* foram obtidas manualmente utilizando o *software google earth* (4.3.7284.3916 beta), levando em consideração o modelo de colonização pela costa. Nós traçamos linhas paralelas ao Equador entre cada população e a costa do Oceano Pacífico. Depois disso traçamos manualmente linhas ao longo da costa do Pacifico até o estreito de *Bering*.

A unidade utilizada para quantificar a regressão é o coeficiente de determinação (r^2) . O valor de r^2 é uma medida da proporção de uma variável (ex., freqüência de alelos endêmicos) que é explicada pela outra variável (ex., distância com relação ao estreito de *Bering*). Nós confirmamos a presença de uma correlação positiva, porém não muito elevada, entre a distância e a freqüência de alelos endêmicos ($r^2 = 0,351$). Para verificar se valor de r^2 que obtivemos era significativo, realizamos o teste F da análise de variância do modelo. Nos casos em que os valores de F são significativamente diferentes de zero a hipótese nula de que não há correlação é rejeitada. Esse teste mostrou que o valor de r^2 que obtivemos foi significativo (p<0,01).

As populações *Chipewyan, Cree* e *Ojibwa* são as que apresentaram as menores freqüências de alelos endêmicos e distâncias com relação ao Estreito de Bering (figura 21). Quando excluímos essas populações da regressão o valor de r^2 perdeu significância ($r^2 = 0,109$, p>0,10). Esse resultado mostra que o valor significativo de r^2 é reflexo da diferença de freqüência de alelos endêmicos entre as populações norte-americanas com relação às demais. Quando calculamos a regressão considerando apenas as populações localizadas nos dois extremos da distribuição (América do Norte e leste da América do Sul) obtivemos um alto e significativo valor de $r^{correlação}$ ($r^2 = 0,876$, p<0,0001), mostrando que as populações centro-americanas e andinas são as responsáveis pelo enfraquecimento da correlação entre distância e freqüência de endêmicos.



Figura 20. Freqüência dos alelos cosmopolitas, endêmicos e não-ameríndios.



Figura 21. Regressão linear entre a distância em km da Beríngia e a freqüência de alelos endêmicos. * Levamos em consideração o modelo de colonização das Américas pela costa do Oceano Pacífico no calculo da distância (detalhes no texto).

4.4. TESTES DE DESVIO DE NEUTRALIDADE

Os testes de neutralidades foram aplicados a duas bases de dados: 1) uma incluindo todas as amostras de cada população; 2) e uma segunda na qual excluímos os indivíduos miscigenados. Os indivíduos miscigenados são aqueles que apresentaram alelos nãoameríndios (tabelas 7 e 12). Observamos desvios significativos de neutralidade detectados pelo teste de *Ewens-Watterson* apenas em duas populações: *Chipewyan* e *Huilliche*. Nos *Chipewyan*, a taxa de homozigose observada foi significativamente menor que o esperado (tabela 13), tanto na amostra total quanto na amostra sem os indivíduos miscigenados. Esse resultado é compatível com a ação de seleção balanceadora ou gargalo populacional recente. Nos *Huilliche* nós observamos desvios na direção contrária, pois a taxa de homozigose nessa população foi significativamente maior que o esperado (tabela 13), resultado compatível com ação de seleção direcional ou expansão populacional. A significância desse resultado não persistiu após a exclusão dos indivíduos miscigenados (tabela 13).

Com relação aos resultados obtidos com o teste *D de Tajima*, nós observamos desvios significativos com relação ao esperado segundo o modelo neutro em 16 das 29 populações que analisamos (tabela 13). Levando em consideração a amostra total, as populações do leste da América do Sul apresentaram com exceção de *Arara do Laranjal, D de Tajima* significativamente positivos (*Ticuna S* apresentou valores de *D de Tajima* próximos à margem de significância). Dentre as populações localizadas no lado oeste da América do Sul, *Zenu* e *Waunana* foram as únicas com valores de *D de Tajima* positivo e significativo. Na América do Norte, as populações *Chipewyan* e *Cree* apresentaram *D de Tajima* positivo e significância (tabela 13).

Quando excluímos os indivíduos miscigenados, os valores de *D de Tajima* das populações *Zenu* e *Cree* perderam significância, apesar de manterem-se próximos à margem. Nas populações *Chipewyan, Waunana* e *Guarani,* os desvios mantiveram-se significativos. Nós obtivemos *D de Tajima* significativamente positivo após a exclusão dos indivíduos miscigenados nas populações *Zapotec, Cabecar* e *Kogi*. Na população *Arhuaco,* nós obtivemos o resultado contrário: um D de Tajima significativamente negativo após a exclusão

dos indivíduos miscigenados.

Populações	Amostra Total (N)	Nº de indivíduos Miscigenados (N)
Chipewyan	25	5
Cree	18	7
Ojibwa	16	4
Mixtec	20	2
Zapotec	16	6
Mixe	20	2
Quiche	19	7
Cabecar	20	2
Guaymi	17	1
Embera*	14	-
Waunana	20	1
Zenu	20	6
Kogi	14	1
Arhuaco	17	8
Wayuu	19	7
Ingano	15	6
Aymara	20	6
Huilliche	19	6
Ticuna N*	19	-
Ticuna S	12	2
Zoé*	9	-
Katuena*	7	-
Arara I*	10	-
Arara L*	13	-
Kaiapó*	28	-
Guarani	28	3

Tabela 12. Tamanho das amostras totais e sem indivíduos miscigenados.

* Não encontramos evidências de miscigenação

	Amostra Total			1	Amostra sem Miscigenados						
	Ewe	ns-Watte	erson	D de	Tajima	Ewe	Ewens-Watterson			D de Tajima	
	$F_{obs}{}^{l}$	F_{esp}^{2}	p (EW) ³	D de Tajima	$p(\mathrm{D})^4$	$F_{obs}{}^{l}$	F_{esp}^{2}	p (EW) ³	D de Tajima	$p(D)^4$	
Nentsi	0,1367	0,1518	0,4262	1,3885	0,94400*	#	#	#	#	#	
Chipewyan	<u>0,1120</u>	<u>0,1811</u>	<u>0,0136</u>	<u>2,3012</u>	<u>0,9941</u>	<u>0,1375</u>	<u>0,2677</u>	<u>0,0003</u>	<u>2,1836</u>	<u>0,9925</u>	
Cree	0,1482	0,1178	0,8896	<u>1,5189</u>	<u>0,9549</u>	0,2273	0,1867	0,8588	1,3402	0,9368*	
Ojibwa	0,1426	0,1507	0,5140	1,1093	0,8965	0,1458	0,1939	0,0990	1,1178	0,9074	
Mixtec	0,1363	0,1493	0,4493	0,9122	0,8684	0,1512	0,1991	0,1621	0,8526	0,8473	
Zapotec	0,0742	0,0843	0,2490	0,8669	0,8508	0,0950	0,1046	0,3498	<u>1,6443</u>	<u>0,9693</u>	
Mixe	0,1913	0,2082	0,4785	0,5159	0,7668	0,2176	0,2593	0,3593	1,0809	0,8935	
Quiche	0,0776	0,0794	0,5648	0,7023	0,8202	0,1458	0,1483	0,6056	0,6065	0,7868	
Cabecar	0,3100	0,3559	0,4204	0,5827	0,7866	0,3256	0,4934	0,0896	<u>2,9820</u>	<u>1,0000</u>	
Guaymi	0,3149	0,3432	0,4851	0,0951	0,6058	0,3262	0,4009	0,3230	0,3242	0,6893	
Embera	0,2449	0,2381	0,6510	1,0356	0,8862	#	#	#	#	#	
Waunana	0,1513	0,1851	0,2682	2,0228	<u>0,9860</u>	0,1648	0,2034	0,2548	2,1154	<u>0,9887</u>	
Zenu	0,1650	0,2072	0,2509	<u>1,5490</u>	<u>0,9585</u>	0,2245	0,2376	0,5225	1,2635	0,9220*	
Kogi	0,3061	0,3244	0,5131	0,5824	0,7731	0,3521	0,4676	0,2120	<u>2,0656</u>	<u>0,9912</u>	
Arhuaco	0,2664	0,2529	0,6709	-0,4824	0,3598	0,4568	0,4324	0,6810	-2,2466	<u>0,0014</u>	
Wayuu	0,1233	0,1208	0,6378	1,0443	0,8875	0,1944	0,1939	0,6107	0,8440	0,8448	
Ingano	0,1378	0,1314	0,6988	0,5619	0,7615	0,2778	0,2838	0,5752	1,3037	0,9343*	
Aymara	0,1838	0,1360	0,9184	0,2661	0,6763	0,2579	0,2067	0,8891	0,0038	0,5700	
Huilliche	0,1828	0,1206	<u>0,9655</u>	0,9106	0,8624	0,2917	0,2620	0,7396	1,0501	0,8941	
Ticuna N	0,2507	0,2299	0,7164	1,6292	0,9637	#	#	#	#	#	
Ticuna S	0,1597	0,1693	0,5174	1,3976	0,9465*	0,1612	0,1862	0,3354	1,4001	0,9471*	
Zoé	0,2161	0,2832	0,1542	2,0326	<u>0,9923</u>	#	#	#	#	#	
Katuena	0,1735	0,2130	0,1631	1,8509	0,9879	#	#	#	#	#	
Arara I	0,5450	0,7315	0,2086	2,4450	0,9967	#	#	#	#	#	
Arara L	0,1653	0,1623	0,6727	0,8272	0,8407	#	#	#	#	#	
Kaiapó	0,1799	0,2594	0,1269	2,2935	<u>0,9934</u>	#	#	#	#	#	
Guarani	0,1014	0,1133	0,3971	2,1987	0,9897	0,1120	0,1160	0,5429	<u>2,2382</u>	<u>0,9907</u>	

Tabela 13. Resultados dos testes de neutralidade.

¹ valor observado da taxa de homozigose; ² valor esperado da taxa de homozigose. ³ Valor de p do teste de *Ewens-Watterson*. ⁴ Valor de p do teste de *D de Tajima*. Valores significativos são mostrados em negrito e sublinhados. *Valores próximos às margens de significância. # Valores idênticos ao da amostra total uma vez que não achamos evidências de miscigenação.

5. DISCUSSÃO

Turnover e as distribuições alélicas

Um dos objetivos do presente trabalho foi caracterizar a variação do gene *HLA-B* em populações nativas do continente americano, tendo como ênfase a hipótese do *turnover* alélico. Nossos resultados representam uma expressiva ampliação do conhecimento do gene *HLA-B* nas populações ameríndias e mostram que o padrão de alta freqüência de alelos endêmicos que foram propostos para algumas populações sul-americanas é geral para a América do Sul.

Corroborando resultados já descritos na literatura, nós encontramos uma clara diferenciação entre as populações localizadas na América do Norte e as demais, no que diz respeito à presença dos alelos endêmicos. Nas análises em que populações norte-americanas foram removidas, não há uma correlação estatisticamente significativa entre localização geográfica e freqüências de alelos endêmicos (figura 21). Esse resultado sugere que as populações nativas do continente podem ser divididas em dois grupos: populações neárticas (que praticamente não possuem alelos endêmicos e estão localizadas a norte do Trópico de Câncer) e neotropicais (que apresentam alelos endêmicos em freqüências elevadas, localizadas ao sul do Trópico de Câncer).

Como as populações norte-americanas presentes no nosso estudo restringemse ao território canadense, havia a possibilidade de ocorrência de um gradiente de aumento de freqüência de alelos endêmicos dentro da América do Norte. Para verificar essa questão nós buscamos na literatura trabalhos que analisaram a variação do gene *HLA-B* em outras populações nativas da América do Norte. As populações *Havasupai* e *Zuni*, localizadas no sudoeste dos Estados Unidos não mostraram qualquer evidência da presença de alelos endêmicos (Watkins, Mcadam *et al.*, 1992; Parham, Arnett *et al.*, 1997); os *Pima* e *Sioux*, também localizadas no sudoeste dos Estados Unidos, apresentaram 27% e 2% de alelos endêmicos, respectivamente (Leffell, Fallin *et al.*, 2004; Mack, Castro *et al.*, 2007); os *Seri*, população nativa do noroeste do México, apresentou 9,1% de alelos endêmicos (Williams, Meenagh *et al.*, 2001). Exceto pelos *Pima*, as populações ameríndias do sul dos Estados Unidos e norte do México apresentam baixas freqüências de alelos endêmicos, assim como as populações nativas do território canadense que analisamos. Entre as populações do sul México, as freqüências alélicas se invertem, havendo uma maior prevalência de alelos endêmicos (figura 20). Esses resultados indicam que a presença de populações com uma maior freqüência de alelos endêmicos está realmente associada ao começo da região climática neotropical. Esse padrão corrobora as suposições de que o ambiente patogênico tropical seja a fonte das pressões seletivas que aumentaram as freqüências dos alelos endêmicos do neotrópico.

As populações neotropicais não são homogêneas no que diz respeito à composição alélica, havendo alelos de distribuições mais amplas ou outros com distribuições mais restritas. Por exemplo, o alelo *B*3905* ocorre desde as populações mexicanas aos *Guarani*. Na América do Sul os alelos *B*1504*, *B*350401*, *B*3505*, *B*3909*, *B*4004* e *B*520102* também possuem distribuições amplas, e também apresentam freqüências elevadas nas populações individuais em que ocorrem. Essa observação é contrária às interpretações apresentadas nos trabalhos que propuseram a hipótese do *turnover* alélico, onde cada população analisada apresentou perfis alélicos distintos (Belich, Madrigal *et al.*, 1992; Watkins, Mcadam *et al.*, 1992; Cadavid e Watkins, 1997; Parham, Arnett *et al.*, 1997). É provável que as diferenças que apresentamos sejam resultado da amostragem mais ampla que utilizamos o que nos

permitiu detectar maiores compartilhamentos de alelos do que os propostos originalmente.

De acordo com os nossos resultados, as maiores diferenças de perfis alélicos não ocorrem entre as populações individuais, mas sim entre grupos de populações. Por exemplo, as populações *Mixtec*, *Zapotec* e *Mixe*, localizadas no sul do México, possuem os alelos endêmicos *B*3512*, *B*390202* e *B*3905* em freqüências elevadas (cada um desses alelos é freqüente em pelo menos duas dessas populações). Já as populações *Arara do Laranjal, Kaiapó* e *Guarani* compartilham os alelos endêmicos *B*1504*, *B*350401*, *B*3505*, *B*3905* e o cosmopolita *B*1508*.

Esses padrões de compartilhamento alélico indicam que, provavelmente, o aumento de freqüência dos alelos endêmicos e perda dos alelos asiáticos não ocorreram em populações isoladas. Consideramos que é mais parcimonioso supor que esses alelos tenham surgido e aumentado de freqüência em diferentes "estoques" da população ancestral da região neotropical. Cada "estoque" teria colonizado determinada região, tal como o planalto mexicano ou a região amazônica, o que explicaria o compartilhamento de determinados conjuntos de alelos endêmicos dentro dessas regiões. Essa idéia é muito parecida com a do *turnover* original, exceto pelo fato de que, na nossa suposição, o aumento de freqüência dos alelos tenha ocorrido nos estoques ancestrais e não nas populações individuais. Essa hipótese é reforçada pelos resultados de testes de neutralidade que discutiremos no próximo tópico.

Turnover e os resultados dos testes de neutralidade

Como comentamos no item anterior, há uma extrema diferença entre as freqüências de alelos endêmicos em populações norte-americanas (neárticas) e as demais (neotropicais). Enquanto as populações neárticas mantiveram o perfil alélico asiático, nas neotropicais esses alelos foram substituídos por micro-recombinantes dos alelos asiáticos (figura 7). Os eventos de conversão gênica que geraram os alelos endêmicos ocorreram preferencialmente no início do éxons 3. De acordo com a predição funcional para diferentes regiões das moléculas HLA feita por Saper e colaboradores (1991), os trechos envolvidos nas conversões dos alelos endêmicos não teriam um papel crucial na capacidade de ligação dos peptídeos antigênicos. Entretanto, trabalhos mais recentes demonstraram que essas regiões podem ser fundamentais para a ligação dos peptídeos. Bade-Doeding e colaboradores (2007) demonstraram que as posições chave para ancoragem dos peptídeos podem variar de acordo com a linhagem HLA analisada. Por exemplo, em alelos pertencentes à linhagem B^*41 os aminoácidos chaves são diferentes dos preditos por Saper e colaboradores (1991) e coincidem com os trechos envolvidos nos eventos de conversão do éxon 3, característicos dos alelos endêmicos. Yagüe e colaboradores (2000) demonstraram que apesar do alelo B*3905 apresentar apenas uma diferença nucleotídica (G \rightarrow T) com relação ao seu provável alelo ancestral, o cosmopolita $B^{*390101}$, a substituição levou a uma profunda mudança nas propriedades de apresentação. Se uma mudança nucleotídica teve tamanho impacto na capacidade de apresentação do alelo B*3905, é plausível supor que os eventos de conversão que envolveram vários aminoácidos nos alelos endêmicos provavelmente tiveram um grande impacto funcional.

As diferenças funcionais entre os alelos endêmicos e seus prováveis ancestrais asiáticos sustentam a hipótese de que o processo de substituição dos alelos antigos tenha sido promovido pela seleção natural. Entretanto, não observamos relação entre a presença de alelos endêmicos em altas freqüências e desvios de neutralidade utilizando os testes de *Ewens-Watterson* e *D de Tajima*.

Encontramos uma diferença marcante no número de populações que desviaram das expectativas neutras utilizando os dois testes. Apenas duas populações desviaram com relação ao esperado sob neutralidade utilizando o teste de *Ewens-Watterson* enquanto que, utilizando a amostra total, dez populações apresentaram desvios significativos para o teste *D de Tajima* (tabela 13). Excluindo-se os indivíduos provavelmente miscigenados, o número de populações com valores significativos de *D de Tajima* subiu para 12, havendo duas que perderam significância (tabela 13).

Garrigan e Hedrick (2003) mostraram que várias populações humanas estudadas para o gene *HLA-A* não apresentavam desvios significativos com relação ao esperado pelo teste de *Ewens-Watterson*, enquanto desvios positivos e significativos de *D de Tajima* foram obtidos. Valores significativamente positivos de *D de Tajima* são obtidos para amostras que apresentam alelos com altas divergências nucleotídicas enquanto que o grau de diferenciação entre os alelos não é levada em conta no teste de *Ewens-Watterson*, que só utiliza as freqüências alélicas. Segundo Garrigan e Hedrick (2003), os altos níveis de divergência nucleotídica observados entre os alelos de *HLA-A* são mais compatíveis com a atuação de seleção balanceadora por milhões de anos do que com seleção recente nas populações individuais.

Aplicando esse mesmo raciocínio ao nosso resultado com *HLA-B*, chegamos à conclusão que os desvios significativos que captamos com o teste *D de Tajima* resultam da presença de alelos com altos níveis de divergência. Há várias linhagens alélicas, divergentes ao nível molecular, presentes nas populações ameríndias. A existência dessas linhagens divergentes explica o valor de *D de Tajima* significativo. Uma vez que o processo que resulta em diferenças entre linhagens, envolvendo múltiplas substituições, se dá em longas escalas de tempo, concluímos que o sinal de

seleção balanceadora captado pelo *D de Tajima* é anterior à entrada no continente americano, e talvez à especiação humana.

Apesar dessa tendência geral a valores de D positivos, encontramos diversas populações com valores de D que não desviavam significativamente da neutralidade. Uma provável explicação seria a presença de indivíduos miscigenados na amostra. As amostras das populações *Zapotec, Cabecar, Kogi* e *Arhuaco* adquiriram valores significativos de D, quando delas excluímos os indivíduos supostamente miscigenados, enquanto as populações *Cree* e *Zenu* perderam significância quando realizamos o mesmo procedimento (tabela 13). Muitos dos alelos presentes na amostra devido a fluxo gênico com europeus pertencem a linhagens raras nas populações ameríndias. Conseqüentemente, a presença desses alelos introduz sítios polimórficos de baixas freqüências, que contribuem para o componente θS do D de *Tajima*, tornando-o mais negativo (ver seção 3.3. METODOS ANALÍTICOS). Quando retiramos os indivíduos que contêm alelos não-ameríndios, retiramos também sítios polimórficos de baixa de freqüência da amostra e dessa forma o D de *Tajima* torna-se mais positivo, explicando a significância que observamos nas populações *Zapotec, Cabecar* e *Kogi*.

Os alelos não-ameríndios $B^{*570101}$, B^{*3549} e $B^{*180101}$ encontram-se em freqüências elevadas nas populações *Cree, Zenu* e *Arhuaco*. Como conseqüência, a exclusão dos indivíduos com esses alelos removeu sítios polimórficos de freqüência intermediária na amostra, diminuindo o componente $\theta \pi$ do *D de Tajima*, tornando essa estatística mais negativa.

A discussão acima mostra que é incorreto interpretar os resultados significativos do teste *D de Tajima* para *HLA-B* como indicativos de que as populações estão ou estiveram sobre seleção em tempos recentes. Os resultados do teste de *Ewens-Watterson* são menos afetados por eventos remotos e, portanto, mais

adequados na captação de sinais seletivos ou demográficos mais recentes (Garrigan e Hedrick, 2003). Esse teste utiliza a taxa de homozigose como estatística sumária, captando informação sobre as freqüências alélicas da amostra. Com exceção das populações *Chipewyan* e *Huilliche*, todas as demais apresentaram distribuição de freqüências alélicas compatíveis com as expectativas neutras. Em *Chipewyan* observamos uma taxa de homozigose menor que a esperada. Isso significa que os alelos estão distribuídos com freqüências muito parecidas, uma situação compatível com seleção balanceadora. A população *Huilliche* apresenta uma taxa de homozigose maior que a esperada. Esse resultado é o reflexo da alta freqüência do alelo *B*3909* com relação aos demais, situação compatível com a atuação de seleção a favor desse alelo.

De acordo com esses resultados, as populações com maiores freqüências de alelos endêmicos não apresentam mais evidências de desvio da neutralidade, quando comparadas às populações com menor freqüência de alelos endêmicos. Uma explicação para esse resultado é que o baixo poder estatístico do teste de *Ewens-Watterson* (Garrigan e Hedrick, 2003) não permitiu a detecção de diferenças entre populações com mais ou menos alelos endêmicos, mesmo se elas existissem.

Uma segunda hipótese é que a seleção promotora da substituição dos alelos asiáticos não estabeleceu padrões tais como a presença de vários alelos em freqüências intermediárias ou favorecimento de um único alelo. Esses padrões são os responsáveis pela rejeição da hipótese nula de neutralidade e equilíbrio. É possível que diferenças entre os coeficientes seletivos dos alelos endêmicos individuais criem padrões populacionais que não refutem a neutralidade.

Uma terceira alternativa está atrelada à hipótese que lançamos no item anterior, de acordo com a qual o evento de substituição alélica não ocorreu nas populações individuais, e sim nos "estoques" ancestrais de diferentes regiões do continente. Segundo nossa hipótese, a seleção teria atuado nos "estoques" ancestrais, o amazônico por exemplo, promovendo aumento de freqüência dos alelos endêmicos e substituição dos alelos antigos. Entretanto, a deriva genética associada à fundação de grupos populacionais e pequenos tamanhos efetivos teriam levado as freqüências alélicas a padrões compatíveis com neutralidade.

Conquistas européias

As populações nativo-americanas sofreram um severo gargalo populacional, com estimativas de até 90%, após os primeiros contatos com os conquistadores europeus (Naranjo, 1995). Os relatos da conquista do continente, vindo tanto dos colonizadores quanto dos colonizados, apontam as doenças trazidas pelos europeus como as principais responsáveis pela drástica redução populacional sofrida pelas populações nativas do continente (Bethell, 1999; Hemming, 2007). Dentre as principais epidemias, varíola, sarampo e/ou rubéola e gripe foram as que ceifaram mais vidas dentro de todo o Novo Mundo (Settipane e Russo, 1995). Essas três doenças são causadas por vírus e a resposta imune adaptativa gerada contra esse tipo de patógenos é via moléculas HLA de classe I, incluindo as moléculas HLA-B.

Apesar do impacto que causaram em todo o continente, as epidemias européias não tornaram homogêneas as distribuições alélicas das populações nativoamericanas. Os nativos norte-americanos são um exemplo. Tanto os *Chipewyan* quanto os *Cree* sofreram com as epidemias de gripe, varíola e sarampo trazidas com o estabelecimento da companhia *Hudson*'s *Bay* no Canadá (Ray, 1976). Apesar disso, essas populações apresentam discrepâncias nos níveis de taxa de heterozigose (tabela 13) e composição dos alelos (figura 8). Isso nos indica que as epidemias trazidas pelos europeus não tiveram um impacto seletivo na variação do gene *HLA-B*. Existem associações bem documentadas entre alelos de genes *HLA* e proteção a doenças de progressão lenta ou crônica, tal como AIDS e malária. O mesmo não ocorre com agentes infecciosos de contágio e progressão rápidos, tal como os vírus que assolaram as populações nativo-americanas. Existem relatos, tanto da América espanhola quanto da portuguesa de tribos inteiras dizimadas em questão de poucos dias pelas doenças trazidas pelos europeus (Bethell, 1999; Hemming, 2007). A resposta imune adaptativa requer alguns dias para ser montada, sendo provável que não houve tempo para a montagem de uma resposta imune contra os vírus trazidos pelos conquistadores. Por esse motivo, acreditamos que as epidemias dos últimos 500 anos causaram à variação do gene *HLA-B* um forte gargalo populacional, sem efeito seletivo.

Alelos endêmicos no Velho mundo, um desafio ao modelo de tunover?

Toda a teoria do *turnover* e as considerações que apresentamos nas seções anteriores dessa discussão são baseadas na premissa da existência dos alelos endêmicos. Entretanto, trabalhos recentes demonstraram a presença de alguns alelos tidos como endêmicos das Américas em populações do leste Asiático (Solberg, Mack *et al.*, 2008). Entre esses alelos encontramos B^{*3505} , o segundo mais freqüente entre populações amazônicas (figuras 18 e 19). O alelo B^{*3505} provavelmente surgiu por uma conversão envolvendo o alelo $B^{*350101}$ como receptor e os alelos $B^{*400201}$ ou $B^{*480101}$ como doadores (figura 7). Há três possibilidades para explicar a distribuição disjunta de B^{*3505} : 1) Trata-se de um alelo cosmopolita, que teria vindo à América do Sul durante a colonização do continente a 20.000 anos e que posteriormente se perdeu nas populações norte e centro-americanas e asiáticas; 2) Trata-se de um alelo homoplásico, originado por dois eventos independentes de

conversão gênica na Ásia e na América do Sul; 3) O compartilhamento desse alelo se deu por fluxo gênico entre as populações do sudeste asiático e a América do Sul.

Não há nenhum estudo publicado com populações nativas norte-americanas que tenha encontrado qualquer evidência da presença desse alelo. Igualmente, não encontramos nenhum indício da presença de B^{*3505} nas populações norte-americanas que estudamos. O alelo B^{*3505} nunca foi descrito em populações que habitam as ilhas do oceano Pacífico (Solberg, Mack *et al.*, 2008), dessa forma, a hipótese de fluxo gênico trans-Pacífico fica enfraquecida. Dado o elevado desequilíbrio de ligação observado na região do MHC, a dúvida entre ancestralidade comum ou homoplasia dos alelos B^{*3505} sul-americano e asiático seria solucionada averiguando em que haplótipos os alelos ocorrem nesses dois continentes. Devido à distribuição disjunta, a alta freqüência desse alelo nas populações sul-americanas e a presença dos prováveis alelos ancestrais em ambos os continentes, é provável que o alelo B^{*3505} encontrado na América do Sul seja homoplásico com relação ao asiático, tendo aumentado de freqüência independentemente nesses dois continentes.

6. CONCLUSÕES

- A grande concordância entre os resultados obtidos pelas técnicas de PCR-SBT e PCR-SSOP nos permitiu a construção de um conjunto de dados de alta confiabilidade.
- 2) De acordo com nossos resultados de regressão linear entre distância do Estreito de Bering e freqüência de alelos endêmicos, nós associamos a ocorrência de populações com altas freqüências de alelos endêmicos ao início da região climática neotropical, corroborando a hipótese de que o ambiente tropical seja fonte das pressões seletivas que promoveram a mudança do perfil alélico observado nas populações nativas da América Latina;
- 3) Concluímos que as maiores diferenças de perfis alélicos não ocorrem entre as populações individuais, mas entre grupos de populações e, conseqüentemente, consideramos que a substituição dos alelos asiáticos pelos ameríndios não ocorreu em populações isoladas, mas nos estoques ancestrais de pelo menos duas regiões do neotrópico: sul do México e leste da América do Sul;
- 4) Os resultados obtidos com o teste de *D de Tajima* não foram informativos sobre os processos de mudança de perfis alélicos que ocorreram nas populações ameríndias, pois os sinais de seleção balanceadora captados por esse teste referem-se a eventos anteriores à entrada das populações no continente americano;
- 5) A ausência de desvios significativos com relação à expectativa de neutralidade e equilíbrio populacional, associada à nossa conclusão do item 3, são uma evidência que a seleção teria atuado nos "estoques" ancestrais, promovendo aumento de freqüência dos alelos endêmicos e substituição dos alelos antigos, e que a deriva genética associada à fundação das populações individuais e os
pequenos tamanhos efetivos mantidos por essas populações teriam levado as freqüências alélicas a padrões compatíveis com neutralidade;

6) Acreditamos que as epidemias dos últimos 500 anos resultaram num gargalo populacional que afetou o genoma inteiro das populações, inclusive o gene *HLA-B*. Entretanto, é provável que essas epidemias não tenham representado um fator seletivo sobre genes do sistema imune adaptativo, como o *HLA-B*.

7. REFERENCIAS BIBLIOGRÁFICAS

ABBAS, A.; LICHTMAN, A. *Cellular and Molecular Immunology*. 5. ed. Philadelphia: Saunders, 2005.

AYALA, F. et al. Molecular genetics of speciation and human origins. Proc Natl Acad Sci USA [S.I.], v. 91, n. 15, p. 6787-94, Jul 1994.

BADE-DOEDING, C. *et al.* Amino acid 95 causes strong alteration of peptide position Pomega in HLA-B*41 variants. *Immunogenetics* [S.I.], v. 59, n. 4, p. 253-9, Apr 2007.

BELICH, M. *et al.* Unusual HLA-B alleles in two tribes of Brazilian Indians. *Nature* [S.I.], v. 357, n. 6376, p. 326-9, May 1992.

BETHELL, L. *História da América Latina*. São Paulo: Editora da Universidade de São Paulo, 1999.

BJORKMAN, P. et al. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* [S.I.], v. 329, n. 6139, p. 506-12, 1987.

BORGHANS, J. et al. MHC polymorphism under host-pathogen coevolution. *Immunogenetics* [S.I.], v. 55, n. 11, p. 732-9, Feb 2004.

BUBB, K. *et al.* Scan of human genome reveals no new Loci under ancient balancing selection. *Genetics* [S.I.], v. 173, n. 4, p. 2165-77, Aug 2006.

BUSTAMANTE, C. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* [S.I.], v. 437, n. 7062, p. 1153-7, Oct 2005.

CADAVID, L.; WATKINS, D. Heirs of the jaguar and the anaconda: HLA, conquest and disease in the indigenous populations of the Americas. *Tissue Antigens* [S.I.], v. 50, n. 6, p. 702-11, Dec 1997.

CARRINGTON, M. *et al.* HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* [S.I.], v. 283, n. 5408, p. 1748-52, Mar 1999.

DE BOER, R. *et al.* Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics* [S.I.], v. 55, n. 11, p. 725-31, Feb 2004.

DOHERTY, P.; ZINKERNAGEL, R. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* [S.I.], v. 256, n. 5512, p. 50-2, Jul 1975a.

. H-2 compatibility is required for T-cell-mediated lysis of target cells infected with lymphocytic choriomeningitis virus. *J Exp Med* [S.I.], v. 141, n. 2, p. 502-7, Feb 1975b.

EWENS, W. The sampling theory of selectively neutral alleles. *Theor Popul Biol* [S.I.], v. 3, n. 1, p. 87-112, Mar 1972.

EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* [S.I.], v. 8, n. 3, p. 186-94, Mar 1998.

EWING, B. *et al.* Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* [S.I.], v. 8, n. 3, p. 175-85, Mar 1998.

EXCOFFIER, L. *et al.* Arlequin ver. 3.1: An integrated software package for population genetics analysis. *Bioinformatics online* [S.I.], v. 1, p. 47-50, 2006.

FAGUNDES, N. *et al.* Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* [S.I.], v. 82, n. 3, p. 583-92, Mar 2008.

GARRIGAN, D.; HEDRICK, P. Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* [S.I.], v. 57, n. 8, p. 1707-22, Aug 2003.

GORDO, I.; CHARLESWORTH, B. Genetic linkage and molecular evolution. *Curr Biol* [S.I.], v. 11, n. 17, p. R684-6, Sep 2001.

GORDON, D. *et al.* Consed: a graphical tool for sequence finishing. *Genome Res* [S.I.], v. 8, n. 3, p. 195-202, Mar 1998.

GUO, S.; THOMPSON, E. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* [S.I.], v. 48, n. 2, p. 361-72, Jun 1992.

HELMBERG, W. *et al.* The sequencing-based typing tool of dbMHC: typing highly polymorphic gene sequences. *Nucleic Acids Res* [S.I.], v. 32, n. Web Server issue, p. W173-5, Jul 2004.

HEMMING, J. *Ouro Vermelho*. São Paulo: Editora da Universidade de São Paulo, 2007.

HILL, A. *et al.* Genetic analysis of host-parasite coevolution in human malaria. *Philos Trans R Soc Lond B Biol Sci* [S.I.], v. 352, n. 1359, p. 1317-25, Sep 1997.

HOPKINS, D. spects of the paleogeography of Beringia during the late Pleistocene. In: HOPKINS, D. *et al* (Ed.). *Paleoecology of Beringia*. New York: Academic Press, 1982. p. 3-28.

HUGHES, A.; NEI, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* [S.I.], v. 335, n. 6186, p. 167-70, Sep 1988.

HUTTLEY, G. *et al.* A scan for linkage disequilibrium across the human genome. *Genetics* [S.I.], v. 152, n. 4, p. 1711-22, Aug 1999.

JANEWAY, C. J. et al. Imunobiology. 5. ed. New York: Garland, 2001.

KASLOW, R. *et al.* Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med* [S.I.], v. 2, n. 4, p. 405-11, Apr 1996.

KITCHEN, A. *et al.* A three-stage colonization model for the peopling of the Americas. *PLoS ONE* [S.I.], v. 3, n. 2, p. e1596, 2008.

KOSTYU, D.; AMOS, D. Mysteries of the Amerindians. *Tissue Antigens* [S.I.], v. 17, n. 1, p. 111-23, Jan 1981.

LAHIRI, D.; NURNBERGER, J. J. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res* [S.I.], v. 19, n. 19, p. 5444, Oct 1991.

LEFFELL, M. *et al.* HLA alleles and haplotypes among the Lakota Sioux: report of the ASHI minority workshops, part III. *Hum Immunol* [S.I.], v. 65, n. 1, p. 78-89, Jan 2004.

MACK, S. et al. Pima from Arizona. In: HANSEN (Ed.). Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference. Seattle: WA: IHWG Press, 2007. p. 634-638.

MALE, D. et al. Immunology. Elsevier, 2006.

MARSH, S. *et al.* Nomenclature for Factors of the HLA System, 2004. *Hum Immunol* [S.I.], v. 66, n. 5, p. 571-636, May 2005.

MEYER, D. et al. Single Locus Polymorphism of Classical HLA Genes. In: HANSEN, J. (Ed.). Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference. Seattle: WA: IHWG, 2007. p. 653-704.

MEYER, D.; THOMSON, G. How selection shapes variation of the human major histocompatibility complex: a review. *Ann Hum Genet* [S.I.], v. 65, n. Pt 1, p. 1-26, Jan 2001.

MITHUN, M. *The languages of native North America*. Cambridge: Cambridge University Press, 1999.

NARANJO, P. Epidemic hecatombin the New World. In: GA, S. (Ed.). *Columbus and the New World: medical implications*. Providence: Oceanside Publications, 1995. p. 19-23.

NIELSEN, R. Molecular signatures of natural selection. *Annu Rev Genet* [S.I.], v. 39, p. 197-218, 2005.

PARHAM, P. MHC class I molecules and KIRs in human history, health and survival. *Nat Rev Immunol* [S.I.], v. 5, n. 3, p. 201-14, Mar 2005.

PARHAM, P. *et al.* Episodic evolution and turnover of HLA-B in the indigenous human populations of the Americas. *Tissue Antigens* [S.I.], v. 50, n. 3, p. 219-32, Sep 1997.

PRUGNOLLE, F. et al. Pathogen-driven selection and worldwide HLA class I diversity. Curr Biol [S.I.], v. 15, n. 11, p. 1022-7, Jun 2005.

RAY, A. Diffusion of diseases in the western interior of Canada, 1830-1850. *Geogr Rev* [S.I.], v. 66, p. 139-57, 1976.

ROBINSON, J. *et al.* IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res* [S.I.], v. 31, n. 1, p. 311-4, Jan 2003.

SABETI, P. *et al.* Positive natural selection in the human lineage. *Science* [S.I.], v. 312, n. 5780, p. 1614-20, Jun 2006.

SALZANO, F. Molecular variability in Amerindians: widespread but uneven information. *An Acad Bras Cienc* [S.I.], v. 74, n. 2, p. 223-63, Jun 2002.

SAMBROOK, J.; RUSSELL, D. *Molecular cloning: a laboratory manual*. 3. ed. New York: Cold Spring Harbor laboratory Press, 2001. (3).

SAPER, M. *et al.* Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution. *J Mol Biol* [S.I.], v. 219, n. 2, p. 277-319, May 1991.

SETTIPANE, G.; RUSSO, T. Mechanism of desease transmission to native Indians. In: GA, S. (Ed.). *Columbus and the New World: medical implications*. Providence: Oceanside Publications, 1995. p. 25-27.

SOLBERG, O. *et al.* Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum Immunol* [S.I.], v. 69, n. 7, p. 443-64, Jul 2008.

TAJIMA, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* [S.I.], v. 105, n. 2, p. 437-60, Oct 1983.

TAKAHATA, N.; NEI, M. Allelic genealogy under overdominant and frequencydependent selection and polymorphism of major histocompatibility complex loci. *Genetics* [S.I.], v. 124, n. 4, p. 967-78, Apr 1990.

TARAZONA-SANTOS, E. *et al.* Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* [S.I.], v. 68, n. 6, p. 1485-96, Jun 2001.

VALLENDER, E.; LAHN, B. Positive selection on the human genome. *Hum Mol Genet* [S.I.], v. 13 Spec No 2, p. R245-54, Oct 2004.

WANG, S. *et al.* Genetic variation and population structure in native Americans. *PLoS Genet* [S.I.], v. 3, n. 11, p. e185, Nov 2007.

WATKINS, D. *et al.* New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. *Nature* [S.I.], v. 357, n. 6376, p. 329-33, May 1992.

WATTERSON, G. The Homozygosity Test of Neutrality. *Genetics* [S.I.], v. 88, n. 2, p. 405-417, Feb 1978.

WILLIAMS, F. *et al.* Analysis of the distribution of HLA-B alleles in populations from five continents. *Hum Immunol* [S.I.], v. 62, n. 6, p. 645-50, Jun 2001.

YAGÜE, J. *et al.* Peptide specificity of the Amerindian B*3905 allotype: molecular insight into selection mechanisms driving HLA class I evolution in indigenous populations of the Americas. *Tissue Antigens* [S.I.], v. 56, n. 5, p. 385-91, Nov 2000.