Rodrigo dos Santos Francisco

Unidades de Seleção nos Genes *HLA*

Units of Selection in HLA Genes

Versão simplificada e corrigida, o original encontra-se disponível no Instituto de Biociências

São Paulo 2013

Rodrigo dos Santos Francisco

Unidades de Seleção nos Genes HLA

Units of Selection in HLA Genes

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

Orientador(a): Diogo Meyer

São Paulo

RESUMO

Os genes HLA (Antígenos leucocitários humanos) estão localizados no Complexo Principal de Histocompatibilidade humano (o MHC), e possuem os maiores níveis de variação do genoma, com milhares de alelos, altas taxas de heterozigose e diversidade nucleotídica. No presente estudo, nosso objetivo foi a identificação dos principais alvos da seleção natural nos genes HLA. Para isso, propusemos duas abordagens que resultaram na redação de dois manuscritos. Na primeira abordagem, nós testamos a hipótese de que os principais alvos da atuação da seleção natural nas moléculas HLA seriam os aminoácidos que compões os sítios que ancoram os peptídeos antigênicos (os bolsões B e F da região de ligação de peptídeos (PBR)). Para isso, utilizamos um conjunto de dados de 6.435 e 6.409 indivíduos genotipados para os genes HLA-A e -B respectivamente, gerados para o 13º Workshop Internacional de Histocompatibilidade (IHW) e pertencentes a 55 populações espalhadas por todos os continentes. Nós estimamos a diversidade nucleotídica (π) das sequências que codificam para os bolsões B e F e comparamos esses dados com os obtidos de outros bolsões da PBR. Concomitantemente, utilizamos a classificação de alelos dos locos HLA-A e -B em supertipos, que são agrupamentos alélicos com similaridades nos perfis de ligação de peptídeos, devido a semelhanças em aminoácidos específicos dos bolsões B e F. Nós descrevemos os padrões observados de variação dos supertipos e desenvolvemos um teste de hipótese onde comparamos os estimadores observados de diferenciação populacional (Gst) e diversidade genética (taxa de heterozigose (He) e número de alelos (k)) com os obtidos a partir de 10.000 réplicas constituídas por agrupamentos aleatórios de alelos.

O bolsão B foi a região que apresentou os maiores níveis de diversidade no gene HLA-B (p <0,00001, teste de soma de ranques de Mann-Whitney) e boa parte de sua variação está estruturada entre os supertipos desse mesmo loco. Além disso, os padrões observados de variação nos supertipos de HLA-B não foram reproduzidos pelos agrupamentos aleatórios de alelos (com 98 % das simulações apresentando valores de Gst menores do que os observados, utilizando as amostras africanas, europeias e asiáticas). Esse resultado indicou que os supertipos e consequentemente as especificidades do bolsão B estão significativamente estruturas entre as populações, um indicativo de adaptações locais aos patógenos específicos de diferentes regiões geográficas. Esses mesmos padrões não foram reproduzidos na análise do loco HLA-A, pois boa parte da variação no PBR desse gene não está localizada nos bolsões B e F. Além disso, as simulações envolvendo os supertipos de HLA-A reproduziram mais frequentemente os padrões observados de variação, indicando que os bolsões B e F não são os principais alvos da seleção nesse gene, ou que os níveis de seleção em HLA-A sejam menores dos atuantes em HLA-B.

Na segunda abordagem, o nosso principal objetivo foi a identificação de genes que contribuíram para a adaptação local das populações nativas das Américas, que teria ocorrido durante o recente processo de colonização desse continente. Nós sequenciamos os exons 2 e 3 dos *loci* de classe I *HLA-B* e -*C* e o exon 2 do loco de classe II -*DRB1* em 635, 524 e 568 indivíduos, respectivamente, pertencentes a 32 populações nativas do continente Americano. Os dados de sequência foram utilizados na estimativa das frequências alélicas, taxa de heterozigose (He), grau de compartilhamento de alelos entre populações (medido pela distância de Prevosti) e desvios de neutralidade utilizando o teste D de Tajima. Nós também comparamos os padrões de variação das taxas de heterozigose obtidas a partir dos loci *HLA* ao longo do continente com os obtidos a partir de um conjunto de 61 microssatélites espalhados ao longo do genoma, permitindo-nos a diferenciação dos padrões provavelmente gerados pela história demográfica ou seleção natural.

O loco HLA-B apresentou o maior número de pares de populações em que não observamos compartilhamento de alelos (44 pares contra 4 e 6 para os *loci* HLA-C e -DRB1,

respectivamente) sendo que a região leste da América do Sul (SAE) foi a que apresentou os menores níveis de compartilhamento de alelos com outras regiões das Américas (39 dos 44 pares de populações continham uma população SAE). Essa maior diferenciação do gene HLA-B nas populações SAE é uma consequência da presença de alelos exclusivos dessa região, originados por eventos de conversão genica e/ou recombinação envolvendo alelos presentes em outras regiões do continente. As populações SAE também apresentaram níveis elevados de variação para o gene HLA-B, resultado evidenciado pela falta de correlação entre a diminuição da taxa de heterozigose e o aumento da distância em relação ao Estreito de Bering ($r^2 = -0,1117, p > 0,05$), o que contrasta com a tendência geral observada nos microssatélites e genes HLA -C e -DRB1 $(r^2 = -0.1957, -0.2261 \text{ e} -0.2637, \text{ respectivamente } (p < 0.05))$. Finalizando, as populações SAE apresentaram valores de D de Tajima maiores (p <0,001, teste de soma de ranques de Mann-Whitney) e mais significativos (p < 0.0000005, aplicando um teste binomial exato) no loco HLA-B, quando comparadas às populações das outras regiões. Essas diferenças entre regiões geográficas não foram observadas nos genes HLA-C e -DRB1, corroborando a explicação seletiva para o aumento da frequência dos alelos de HLA-B originados por conversão gênica/recombinação em resposta aos novos desafios ambientais das regiões tropicais na América do Sul.

As conclusões obtidas a partir de ambas as abordagens do presente trabalho apontam o gene *HLA-B* como o principal alvo da seleção natural, uma vez que esse loco concentra as maiores evidências de atuação de seleção natural recente quando comparado aos demais genes *HLA* analisados. Nós também demonstramos com as análises intragênicas que o bolsão B do PBR de *HLA-B* concentra por boa parte das diferenças observadas entre as populações, implicando em diferenças nos perfis de apresentação de peptídeos entre essas mesmas populações, o que pode ser interpretado como um indicativo de adaptações locais aos conjuntos de patógenos presentes em distintas regiões geográficas.

ABSTRACT

The Classical HLA genes (Human Leucocyte Antigens) are located in the human Major Histocompatibility Complex (the MHC) and present the highest levels of variation on the Human genome, with thousands of alleles associated with high levels of heterozygosis and nucleotide diversity. In the present study, our goal was the identification of the main targets of natural selection on the HLA genes. We proposed two different approaches to address this issue resulting in two manuscripts. At the first approach, we performed an intragenic analysis, verifying if the amino acids at the peptide-binding region (PBR) anchor positions (the B and F pockets) exhibit higher evidences of evolution under natural selection when compared with the remaining regions of the HLA molecules. To do so, we used a dataset generated for the 13th International Histocompatibility Workshop (IHW), composed by 6,435 and 6,409 individuals genotyped for HLA-A and -B respectively, belonging to 55 populations scattered along all the continents. We measured the levels of nucleotide diversity (π) of the sequences coding for the B and F pockets and compared them with the remaining PBR pockets. Concomitantly, we applied the supertype classification which consists in groups of HLA-A and -B alleles which bind overlapping sets of peptides, as a consequence of sharing specific amino acids at B and F pockets and described the patterns of supertype variation in the observed data. Next, we developed a hypothesis test in which the observed patterns of population differentiation (Gst) and variability (heterozygosity (He) and number of alleles (k)), using the supertype definition, were compared with 10,000 replicates of random assigned groups of alleles.

At the HLA-B locus, the B pocket presented the highest levels of variation (p < 0.00001, Wilcoxon rank sum test) and concentrated most of the differences between supertypes. Our simulations results revealed that the reassignment of alleles into random groups could not reproduce the observed patterns of population differentiation (with 98% of the simulations presenting Gst values smaller than the observed, using the African, European and Asiatic samples), indicating that supertypes and more specifically the B pocket specificities are significantly structured among populations, which could be an indicative of adaptations to local pathogens. We did not observe the same patterns at the HLA-A locus which presented relative lower levels of variation at B and F pockets when compared with the remaining PBR regions, and simulated values of Gst and He which often reproduced the observed data.

At the second approach, our main objective was the identification of genes that contributed for local adaptation on Native American Populations because of the relatively recent colonization of the new American environments. We sequenced the exons 2 and 3 of the *HLA-B* and *-C* class I loci and the exon 2 of the *-DRB1* class II locus in 635, 524 and 568 individuals, respectively, belonging to 32 different Native American Populations scattered along all the Americas. We estimated the allele frequencies, expected heterozygosity (He), degree of allelic sharing between populations (measured by the Prevost's Distance) and departure from neutral expectation using the Ewens-Watterson (EW) and Tajima's D test. Concomitantly, we used a dataset of 61 microsatellites scattered along the genome as a demographic control, comparing the degree of variation of the heterozygosity along the continent.

The *HLA-B* locus showed the highest number of pairs of populations in which we did not observe any sharing of alleles (44 pairs against 4 and 6 for *HLA-C* and *-DRB1* loci, respectively) and the Eastern South American (SAE) region was the one presenting the smallest levels of allelic sharing with other American regions at this locus (39 out the 44 pair of populations contained a SAE population). The presence of exclusive gene conversion and/or recombination alleles accounts for the higher differentiation of SAE populations at the *-B* locus. The *-B* locus also exhibited a higher level of variation at SAE populations which was evidenced

by the lack of correlation between the decrease of variation and the increase of distance from Bering Strait ($r^2 = -0.1117$, p > 0.05), a general trend observed at the microsatellites marker and HLA-C and DRB1 loci ($r^2 = -0.1957$, -0.2261 and -0.2637, respectively (p < 0.05)). Finally, The SAE populations presented the highest (p < 0.001, Wilcoxon rank sum test) and most significant (p < 0.000005, applying an exact binomial test) positive Tajima's D values for the -B locus when compared with the other regions. We did not observe such differences between geographic regions at the HLA-C and -DRB1 loci, corroborating the adaptive explanation for the increase on the frequency of the HLA-B gene conversion/recombination alleles as a response for the new environmental challengers of the tropical regions in South America.

The combined conclusions obtained from both approaches indicate that HLA-B locus seems to be the main target of natural selection, collecting the highest evidences of recent evolution when compared with the other HLA loci analyzed in the present study. We also demonstrated with the intragenic analysis that at HLA-B PBR, the B pocket seems to be the portion of the molecule whose variation accounts for most of the differences observed between populations, which can be direct associated with the immunological function of HLA-B, since this pocket was one of the more important containments for the peptide presentation.

Introdução Geral

O entendimento da contribuição relativa dos processos estocásticos e da seleção natural na evolução e manutenção da variação genética é um dos principais objetivos da pesquisa em biologia evolutiva (Piertney and Oliver 2006). Dentre os *loci* que apresentam sinais de evolução por seleção natural, podemos citar os genes *HLA* (*Human Leukocyte Antigen*), que apresentam níveis de diversidade incompatíveis com o esperado de acordo com o modelo de evolução neutra em humanos. Além das evidências populacionais, os resultados de estudos de associação com doenças infecciosas indicam que os genes *HLA* estão evoluindo sob seleção natural dirigida por patógenos (Meyer and Thomson 2001).

Apesar do grande número de estudos, ainda há debate a respeito de diversos aspectos do processo seletivo que molda a variação dos genes HLA. Nosso objetivo no presente trabalho foi entender qual ou quais foram os principais alvos da atuação da seleção natural dentro desse sistema gênico. Alvos plausíveis incluem desde aminoácidos específicos, até genes inteiros. Esses dois níveis de análise, o intra e intergênico, motivaram a organização do presente trabalho em dois capítulos no formato de manuscritos. Porém, antes de entrarmos nas questões especificas exploradas em cada um desses capítulos, nós faremos uma breve contextualização do tema, descrevendo a estrutura e função dos genes HLA e apresentando as evidências de que esses genes estão evoluindo sob a ação da seleção natural.

1.1. Genes HLA: Estrutura, Função e Evidências da atuação da Seleção Natural.

Os genes HLA encontram-se no MHC (o Complexo Principal de Histocompatibilidade), localizado no braço curto do cromossomo 6 (6p21.3) (Figura 1). O MHC é uma das regiões mais bem conhecidas do genoma humano, apresentando altos níveis de diversidade e uma alta concentração gênica (The MHC sequencing consortium 1999). Esse complexo é organizado em três regiões, definidas de acordo com a função dos genes localizados em cada uma delas: classe I, II e III.

Na região de classe I estão localizados os genes HLA-A, -B e -C, também conhecidos como genes HLA de classe I clássicos, que codificam respectivamente as moléculas HLA-A, -B e -C. As moléculas HLA de classe I são formadas por duas subunidades: a cadeia α e a β 2-microglobulina. A cadeia α é codificada pelos genes HLA de classe I e apresenta três domínios extracelulares: α 1, α 2 e α 3. Os domínios α 1 e α 2 formam a fenda de ligação de peptídeos (Saper, Bjorkman, and Wiley 1991).

Na região MHC de classe II estão localizados os genes HLA-DPA1 e -DPB1 que codificam a molécula HLA-DP; os genes HLA-DQA1 e -DQB1 que codificam a molécula HLA-DQ; e os genes HLA-DRA e o complexo dos genes -DRB (HLA-DRB1, -DRB3, -DRB4 e -DRB5) que codificam as moléculas HLA-DR. As moléculas de classe II são também formadas por duas subunidades: as cadeias α e β , entretanto, diferentemente das moléculas de classe I, ambas as subunidades são codificadas por diferentes genes HLA (ex: as cadeias α e β da molécula HLA-DRB1 são codificadas pelos genes HLA-DRA e HLA-DRB1, respectivamente. Cada uma das subunidades das moléculas de classe II apresentam dois domínios: a cadeia α com α 1 e α 2; e a cadeia β com β 1 e β 2. Os domínios α 1 e β 1 formam a fenda de ligação do peptídeo (Figura 2).

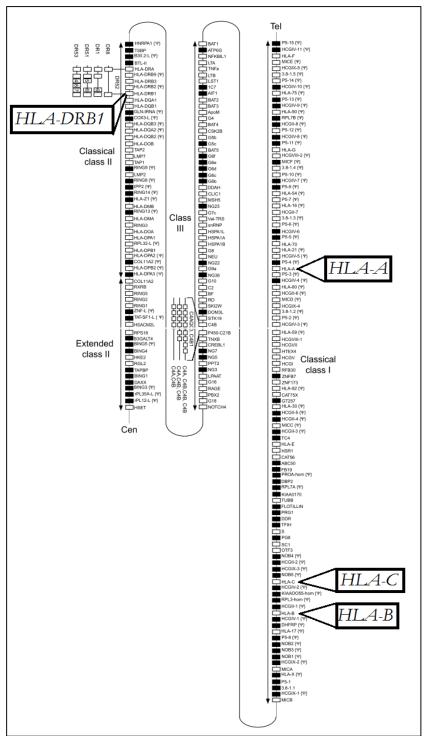


Figura 1. Mapa esquemático do MHC humano: as caixas indicam os genes que foram estudados no presente trabalho *HLA-A*, *-B* e *-C* e *-DRB*. Fonte: (The MHC sequencing consortium 1999).

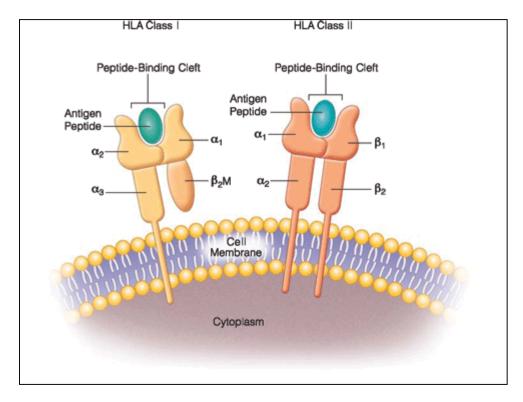


Figura 2. Imagem ilustrando as estruturas das moléculas HLA de classe I e classe II: (veja descrição no texto).

Inicialmente, acreditava-se que a função dos genes *HLA* estava correlacionada apenas com a compatibilidade tecidual observada durante transplantes (PONTES e PORTO 2007). A verdadeira função fisiológica desses genes somente foi compreendida posteriormente: as moléculas codificadas pelos genes *HLA-A*, *-B* e *-C* apresentam peptídeos citoplasmáticos na superfície da maioria das células nucleadas aos linfócitos T CD8+ (Doherty and Zinkernagel 1975) e também são ligantes dos receptores KIR (Killer cell immunoglobulin-like receptors) expressos pelas células natural killers, ou NK (Parham 2005). Já as moléculas de classe II são constitutivamente expressas pelas chamadas células apresentadoras de antígeno profissionais (APC) (Watts 1997; Varney, Gavrilidis e Tait 1999) e apresentam peptídeos oriundos do exterior das células (que foram previamente fagocitados ou pinocitados) aos linfócitos T CD4+ (Varney et al., 1999).

Normalmente as moléculas HLA são expressas associadas a peptídeos oriundos de proteínas próprias. Porém, na presença de proteínas não próprias, moléculas HLA com

peptídeos antigênicos são levadas à superfície celular onde são reconhecidas por linfócitos T (Abbas e Lichtman 2005). A sobrevivência dos patógenos está atrelada à presença de variantes antigênicas que permitem a evasão da apresentação pelas moléculas HLA, enquanto que a sobrevivência dos hospedeiros está atrelada à capacidade de apresentação de antígenos dos patógenos.

A consequência do elo evolutivo entre hospedeiros e patógenos faria com que as moléculas HLA e antígenos coevoluam. Os patógenos tenderiam a acumular mutações de escape, que são aquelas que evitam a apresentação de seus peptídeos pelas moléculas HLA mais comuns na população de hospedeiros. Por outro lado, novas moléculas HLA, capazes de apresentar de forma eficiente essas mutações de escape, seriam adaptativas e aumentariam de frequência na população de hospedeiros, recomeçando o ciclo. Portanto, os microrganismos patogênicos seriam a fonte da seleção atuante nos genes HLA. Essa hipótese evolutiva é conhecida como seleção dirigida por patógenos (Gillespie 1977).

Simulações computacionais mostraram que a seleção dirigida por patógenos explica de maneira satisfatória várias das características observadas nos dados reais para genes HLA, tal como o grande número de alelos, altas taxas de heterozigose e longa persistência dos alelos nas populações ao longo das gerações (Borghans, Beltman, and De Boer 2004). Esse mecanismo evolutivo também seria capaz de explicar as altas taxas de identidade por descendência descritas para genes os genes HLA (Albrechtsen, Moltke e Nielsen 2010).

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 et al. 1997). Alelos de *HLA-B* foram associados a uma maior ou menor progressão à infecção por HIV (Carrington et al. 1999; Kaslow et al. 1996). Além disso, Prugnolle et al. (2007), Qutob et al. (2011) e

Sanchez-Mazas et al. (2012) mostraram que populações localizadas em regiões com grande diversidade de patógenos, possuem em média, uma maior diversidade nos loci HLA.

Todas esses resultados indicam que os genes HLA estão evoluindo sob ação da seleção natural e que a provável fonte da pressão seletiva atuante nesses loci vem da interação com os patógenos presentes no ambiente.

Baseando-se nessas premissas nós desenvolvemos as duas principais questões que motivaram o desenvolvimento do presente trabalho:

- 1) As regiões das moléculas HLA que interagem de forma direta com os antígenos patogênicos foram os alvos principais da atuação da seleção natural?
- 2) Quais foram os genes *HLA* mais impactados durante a ocupação de novos ambientes pelas populações humanas?

Considerações Finais

Os genes HLA apresentam os mais altos níveis de variação do genoma e uma vasta gama de evidências apontam a seleção natural dirigida por patógenos como o principal fator evolutivo contribuindo para a manutenção dessa variação. Partindo desse pressuposto, nossa motivação consistiu na identificação dos principais alvos da atuação da seleção natural nos genes HLA. No primeiro capítulo, nosso objetivo foi a caracterização das regiões que apresentam as mais fortes evidências da atuação de seleção dentro dos genes HLA-A e -B. No segundo capítulo, nos perguntamos quais dentre os genes HLA-B, -C e -DRB1 evoluíram mais rapidamente durante a ocupação relativamente recente do continente Americano.

As diferenças nos regimes seletivos atuantes nas distintas porções das moléculas HLA foram inicialmente exploradas por Hughes e Nei (1988) que demonstraram que os códons dos aminoácidos que compões a fenda de ligação dos peptídeos apresentam uma maior taxa de variação não sinônima quando comparados aos demais códons dos genes HLA. Entretanto, esse estudo não levou em consideração as diferenças entre as regiões da fenda de ligação dos peptídeos. Como descrevemos na introdução do primeiro manuscrito, há evidências que apontam os bolsões B e F como os mais críticos na restrição do repertório de peptídeos apresentados pelas moléculas HLA de classe I. Além disso, Sidney e colaboradores (1996), Sette e Sidney (1999) e Sidney e colaboradores (2008) demonstraram que os alelos dos genes HLA-A e -B poderiam ser classificados em quatro e cinco supertipos, respectivamente, baseando-se no grau de compartilhamento do repertório de apresentação de peptídeos, uma propriedade definida basicamente pela variação nos bolsões B e F. Baseando-se nessas características, nos perguntamos se os bolsões B e F apresentariam níveis mais elevados de variação e se os supertipos estariam

pouco estruturados entre as populações humanas, sinais condizentes com a atuação da seleção balanceadora.

Essas hipóteses foram parcialmente corroboradas na análise do gene *HLA-B*. O sítio de ligação de peptídeos desse gene apresentou níveis significativamente mais elevados de variação quando comparado à mesma região em *HLA-A*. Também demonstramos que o bolsão B concentra boa parte da variação observada em *HLA-B* e a maioria das diferenças entre os supertipos desse loco. Por outro lado, o bolsão F apresentou os níveis mais baixos de variação em ambos os *loci* e boa parte das diferenças entre os supertipos de *HLA-A*.

Esses resultados indicam uma provável restrição evolutiva na variação do bolsão F que pode ter sido imposta pelos padrões de geração e transporte dos peptídeos até a luz do reticulo endoplasmático, realizados pelo proteossomo e proteínas TAP, respectivamente. Essa interpretação é corroborada pelo resultado de que boa parte dos peptídeos extraídos de moléculas HLA de classe I apresentam extremidades carboxila hidrofóbicas e essa limitação pode ser a força por trás dos baixos níveis de diversidade no bolsão F. Por outro lado, a restrição da variação no bolsão F provavelmente tornou vantajosa a maximização da variação no bolsão B, pois indivíduos contendo alelos que codificam bolsões B distintos teriam uma capacidade mais ampla de ligação de peptídeos e consequentemente, de desencadear uma resposta imunológica.

Os elevados níveis de variação no bolsão B de *HLA-B* corroboram essa ideia, indicando que esse bolsão provavelmente é o alvo da atuação da seleção balanceadora nesse loco. Em *HLA-A*, por outro lado, o bolsão B não é a região com os mais altos níveis de variação indicando que outras regiões da fenda podem contribuir para a ancoragem dos peptídeos na molécula, ou que o repertório de apresentação de peptídeos pelas moléculas HLA-A seja mais restrito e/ou conservado quando comparado a HLA-B. Os resultados com as simulações envolvendo os supertipos corroboram as interpretações obtidas a partir

das análises moleculares, uma vez que as simulações não foram capazes de reproduzir os padrões observados de variação e diferenciação populacional que obtivemos com os supertipos de *HLA-B*. Esses resultados permitem a interpretação de que as especificidades de ligação de peptídeo das moléculas HLA-B encontram-se mais estruturadas entre as populações do que seria esperado pelo acaso o que pode ser um reflexo de adaptações locais contra patógenos.

Os nossos resultados permitem a inferência de que as moléculas HLA-B provavelmente possuem os mais amplos repertórios de apresentação de peptídeos. Essa maior diversidade na capacidade de apresentação seria uma consequência dos níveis mais elevados de variação no bolsão B associados à estruturação dessa variação entre os supertipos, fazendo dessas moléculas a principal linha de defesa contra novos agentes patogénicos ou mutações de escape. Essa interpretação está completamente de acordo os resultados que apresentamos no segundo capitulo dessa tese.

O gene *HLA-B* apresenta vários alelos originados por conversão gênica e/ou recombinação, presentes em altas frequências nas regiões tropicais das Américas, especialmente no leste da América do Sul, sendo praticamente ausentes nas populações nativas da América do Norte. Esse perfil alélico distinto foi descoberto no início da década de 90 e a relevância deste achado deve-se ao contraste com outros marcadores genéticos, incluindo outros genes de *HLA*, cuja diversidade parece ser um subconjunto do encontrado em outras regiões das Américas e nordeste da Ásia. Para explicar esse padrão incomum, Parham e colaboradores (1997) propuseram a hipótese *turnover* de alelos que aponta a seleção imposta pelos patógenos endêmicos das regiões tropicais das Américas como a principal força que contribuiu para a substituição dos alelos ancestrais e aumento da frequência dos novos alelos produtos de recombinação e conversão de gênica. Apesar de intuitiva, essa hipótese nunca foi formalmente testada, não havendo estudos

populacionais que tenham comparado as frequências dos novos alelos de HLA-B com desvios do esperado segundo o modelo de neutralidade e equilíbrio populacional.

No segundo capitulo dessa tese nós revisitamos esse modelo e demonstramos que as populações do leste América do Sul (SAE) apresentam as maiores frequências dos alelos de *HLA-B* gerados por conversão gênica/recombinação. Também demostramos que boa parte da variação observada nos *loci HLA-C* e *-DRB1* foi constituída por um subconjunto do observado nas demais regiões do continente Americano.

Corroborando com os dados presentes na literatura, demonstramos que as populações SAE apresentam níveis de variação obtidos a partir de 61 microssatélites espalhados pelo genoma e dos genes *HLA-C* e *-DRB1*, significativamente inferiores ao observado nas demais regiões do continente. Esse resultado está de acordo com a interpretação de que as populações SAE experimentaram elevadas taxas de deriva genética, causando uma diminuição global dos níveis de variação do genoma. Entretanto, contrastando com esse padrão geral, demonstramos que os níveis de diversidade do gene *HLA-B* não sofreram esse mesmo decaimento, indicando que a seleção natural provavelmente contribuiu na manutenção da diversidade nesse loco no leste da América do Sul.

Esse resultado indicou uma provável relação entre a presença dos novos alelos de HLA-B e a atuação da seleção natural. Essa reação foi confirmada com a aplicação do teste D de Tajima, uma vez que as populações do leste da América do Sul apresentaram para o gene HLA-B os mais elevados e significativos valores para essa estatística. Por outro lado, não observamos esse mesmo padrão nos genes HLA-C e -DRB1. Valores significativos e positivos de D de Tajima são obtidos de populações que apresentam alelos divergentes em frequências intermediárias. Como já discutimos, as populações SAE apresentam as mais altas frequências dos alelos de HLA-B originados por eventos de conversão gênica/recombinação. Esses alelos são muito divergentes, diferindo um dos

outros em múltiplas posições e a sua presença em frequências intermediárias explica os valores elevados de Tajima D que obtivemos. O aumento da frequência dos alelos gerados por conversão gênica/recombinação em detrimento de alelos gerados por mutações de ponto foi provavelmente a consequência da combinação da maior velocidade geração de novos alelos associada à maior probabilidade de criação de novas especificidades peptídicas. Dentre os genes que analisamos, HLA-B foi o que teve o maior número de alelos que foram aumentados de frequência desde que o leste da América do Sul foi ocupado.

As conclusões obtidas a partir de ambas as abordagens do presente trabalho apontam o gene *HLA-B* como o principal alvo da seleção natural, uma vez que esse loco concentra as maiores evidências de atuação de seleção natural recente quando comparado aos demais genes *HLA* analisados. Nós também demonstramos com as análises intragênicas que o bolsão B do sítio de ligação de peptídeos de *HLA-B* concentra por boa parte das diferenças observadas entre as populações, implicando em diferenças nos perfis de apresentação de peptídeos entre essas mesmas populações, o que pode ser interpretado como um indicativo de adaptações locais aos conjuntos de patógenos presentes em distintas regiões geográficas.

Referencias Bibliograficas

- Albrechtsen, Anders, Ida Moltke, and Rasmus Nielsen. 2010. "Natural selection and the distribution of identity-by-descent in the human genome." *Genetics* 186(1): 295–308. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2940294& tool=pmcentrez&rendertype=abstract (August 7, 2013).
- Apanius, V, D Penn, P R Slev, L R Ruff, and W K Potts. 1997. "The nature of selection on the major histocompatibility complex." *Critical reviews in immunology* 17(2): 179–224. http://www.ncbi.nlm.nih.gov/pubmed/9094452 (August 19, 2012).
- Bamshad, Michael, and Stephen P Wooding. 2003. "Signatures of natural selection in the human genome." *Nature reviews. Genetics* 4(2): 99–111. http://www.ncbi.nlm.nih.gov/pubmed/12560807 (September 20, 2013).
- Borghans, José A M, Joost B Beltman, and Rob J De Boer. 2004. "MHC polymorphism under host-pathogen coevolution." *Immunogenetics* 55(11): 732–9. http://www.ncbi.nlm.nih.gov/pubmed/14722687 (August 14, 2013).
- Buhler, Stéphane, and Alicia Sanchez-Mazas. 2011. "HLA DNA sequence variation among human populations: molecular signatures of demographic and selective events." *PloS one* 6(2): e14643. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3051395& tool=pmcentrez&rendertype=abstract (September 12, 2011).
- Bustamante, Carlos D, Adi Fledel-Alon, Scott Williamson, Rasmus Nielsen, Melissa Todd Hubisz, Stephen Glanowski, David M Tanenbaum, Thomas J White, John J Sninsky, Ryan D Hernandez, Daniel Civello, Mark D Adams, Michele Cargill, and Andrew G Clark. 2005. "Natural selection on protein-coding genes in the human genome." *Nature* 437(7062): 1153–7. http://www.ncbi.nlm.nih.gov/pubmed/16237444 (July 18, 2011).
- Campos-Lima, P O, V Levitsky, M P Imreh, R Gavioli, and M G Masucci. 1997. "Epitope-dependent selection of highly restricted or diverse T cell receptor repertoires in response to persistent infection by Epstein-Barr virus." *The Journal of experimental medicine* 186(1): 83–9. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2198955& tool=pmcentrez&rendertype=abstract (August 19, 2012).

- Carrington, M, G W Nelson, M P Martin, T Kissner, D Vlahov, J J Goedert, R Kaslow, S Buchbinder, K Hoots, and S J O'Brien. 1999. "HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage." *Science (New York, N.Y.)* 283(5408): 1748–52. http://www.ncbi.nlm.nih.gov/pubmed/10073943 (August 19, 2012).
- Chakraborty, Sajib, Taibur Rahman, Rajib Chakravorty, Alison Kuchta, Atai Rabby, and Munsi Sahiuzzaman. 2013. "HLA supertypes contribute in HIV type 1 cytotoxic T lymphocyte epitope clustering in Nef and Gag proteins." *AIDS research and human retroviruses* 29(2): 270–8. http://www.ncbi.nlm.nih.gov/pubmed/23061377 (September 13, 2013).
- Cordery, Damien V, Allison Martin, Janaki Amin, Anthony D Kelleher, Sean Emery, and David A Cooper. 2012. "The influence of HLA supertype on thymidine analogue associated with low peripheral fat in HIV." *AIDS (London, England)* 26(18): 2337–44. http://www.ncbi.nlm.nih.gov/pubmed/23032422 (September 13, 2013).
- Doherty, P C, and R M Zinkernagel. 1975. "A biological role for the major histocompatibility antigens." *Lancet* 1(7922): 1406–9. http://www.ncbi.nlm.nih.gov/pubmed/49564 (August 19, 2012).
- Ewens, W J. 1972. "The sampling theory of selectively neutral alleles." *Theoretical population biology* 3(1): 87–112. http://www.ncbi.nlm.nih.gov/pubmed/4667078 (August 26, 2013).
- Excoffier, Laurent, and Heidi E L Lischer. 2010. "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." *Molecular ecology resources* 10(3): 564–7. http://www.ncbi.nlm.nih.gov/pubmed/21565059 (July 8, 2011).
- Garrigan, Daniel, and Philip W Hedrick. 2003. "Perspective: detecting adaptive molecular polymorphism: lessons from the MHC." *Evolution; international journal of organic evolution* 57(8): 1707–22. http://www.ncbi.nlm.nih.gov/pubmed/14503614.
- Gilchuk, Pavlo, Charles T Spencer, Stephanie B Conant, Timothy Hill, Jennifer J Gray, Xinnan Niu, Mu Zheng, John J Erickson, Kelli L Boyd, K Jill McAfee, Carla Oseroff, Sine R Hadrup, Jack R Bennink, William Hildebrand, Kathryn M Edwards, James E Crowe, John V Williams, Søren Buus, Alessandro Sette, Ton N M Schumacher, Andrew J Link, and Sebastian Joyce. 2013. "Discovering naturally processed antigenic

- determinants that confer protective T cell immunity." *The Journal of clinical investigation* 123(5): 1976–87.
- http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3635741& tool=pmcentrez&rendertype=abstract (August 8, 2013).
- Gillespie, J. H. 1977. "Sampling theory for alleles in a random environment." *Nature* 266(5601): 443–5. http://www.ncbi.nlm.nih.gov/pubmed/859613 (August 19, 2012).
- Gonzalez-Galarza, Faviel F, Stephen Christmas, Derek Middleton, and Andrew R Jones. 2011. "Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations." *Nucleic acids research* 39(Database issue): D913–9. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3013710& tool=pmcentrez&rendertype=abstract (August 7, 2013).
- Guo, S W, and E A Thompson. 1992. "Performing the exact test of Hardy-Weinberg proportion for multiple alleles." *Biometrics* 48(2): 361–72. http://www.ncbi.nlm.nih.gov/pubmed/1637966 (August 19, 2012).
- Hammer, Michael F, Felisa Blackmer, Dan Garrigan, Michael W Nachman, and Jason A Wilder. 2003. "Human population structure and its effects on sampling Y chromosome sequence variation." *Genetics* 164(4): 1495–509.

 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1462677& tool=pmcentrez&rendertype=abstract (September 18, 2013).
- Hill, A V, A Jepson, M Plebanski, and S C Gilbert. 1997. "Genetic analysis of host-parasite coevolution in human malaria." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 352(1359): 1317–25. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1692024& tool=pmcentrez&rendertype=abstract (August 3, 2012).
- Hraber, Peter, Carla Kuiken, and Karina Yusim. 2007. "Evidence for human leukocyte antigen heterozygote advantage against hepatitis C virus infection." *Hepatology (Baltimore, Md.)* 46(6): 1713–21. http://www.ncbi.nlm.nih.gov/pubmed/17935228 (August 19, 2012).
- Hughes, A L, and M Nei. 1988. "Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection." *Nature* 335(6186): 167–70. http://www.ncbi.nlm.nih.gov/pubmed/3412472 (September 26, 2011).
- Jaruzelska, J, E Zietkiewicz, M Batzer, D E Cole, J P Moisan, R Scozzari, S Tavaré, and D Labuda. 1999. "Spatial and temporal distribution of the

- neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy." *Genetics* 152(3): 1091–101. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1460666& tool=pmcentrez&rendertype=abstract (September 18, 2013).
- Jost, Lou. 2008. "G ST and its relatives do not measure differentiation." *Molecular Ecology* 17(18): 4015–4026. http://doi.wiley.com/10.1111/j.1365-294X.2008.03887.x (June 16, 2011).
- Karlsson, Ingrid, Lea Brandt, Lasse Vinner, Ingrid Kromann, Lars Vibe Andreasen, Peter Andersen, Jan Gerstoft, Gitte Kronborg, and Anders Fomsgaard. 2013. "Adjuvanted HLA-supertype restricted subdominant peptides induce new T-cell immunity during untreated HIV-1-infection." *Clinical immunology (Orlando, Fla.)* 146(2): 120–30. http://www.ncbi.nlm.nih.gov/pubmed/23314272 (September 13, 2013).
- Karlsson, Ingrid, Henrik Kløverpris, Kristoffer Jarlov Jensen, Anette Stryhn, Søren Buus, Annika Karlsson, Lasse Vinner, Philip Goulder, and Anders Fomsgaard. 2012. "Identification of conserved subdominant HIV Type 1 CD8(+) T Cell epitopes restricted within common HLA Supertypes for therapeutic HIV Type 1 vaccines." *AIDS research and human retroviruses* 28(11): 1434–43. http://www.ncbi.nlm.nih.gov/pubmed/22747336 (September 13, 2013).
- Kaslow, R A, M Carrington, R Apple, L Park, A Muñoz, A J Saah, J J Goedert, C Winkler, S J O'Brien, C Rinaldo, R Detels, W Blattner, J Phair, H Erlich, and D L Mann. 1996. "Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection." *Nature medicine* 2(4): 405–11. http://www.ncbi.nlm.nih.gov/pubmed/8597949 (August 19, 2012).
- Knowles, L. L. 2003. "The burgeoning field of statistical phylogeography." *Journal of Evolutionary Biology* 17(1): 1–10. http://doi.wiley.com/10.1046/j.1420-9101.2003.00644.x (July 22, 2011).
- Kuniholm, M H, K Anastos, A Kovacs, X Gao, D Marti, A Sette, R M Greenblatt, M Peters, M H Cohen, H Minkoff, S J Gange, C L Thio, M A Young, X Xue, M Carrington, and H D Strickler. 2013. "Relation of HLA class I and II supertypes with spontaneous clearance of hepatitis C virus." *Genes and immunity* 14(5): 330–5.

- http://www.ncbi.nlm.nih.gov/pubmed/23636221 (September 13, 2013).
- Lazaryan, Aleksandr, Elena Lobashevsky, Joseph Mulenga, Etienne Karita, Susan Allen, Jianming Tang, and Richard A Kaslow. 2006. "Human leukocyte antigen B58 supertype and human immunodeficiency virus type 1 infection in native Africans." *Journal of virology* 80(12): 6056–60. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1472610& tool=pmcentrez&rendertype=abstract (September 13, 2013).
- Meyer, D, and G Thomson. 2001. "How selection shapes variation of the human major histocompatibility complex: a review." *Annals of human genetics* 65(Pt 1): 1–26. http://www.ncbi.nlm.nih.gov/pubmed/11415519 (December 16, 2011).
- Naugler, Christopher, and Robert Liwski. 2008. "An evolutionary approach to major histocompatibility diversity based on allele supertypes." *Medical hypotheses* 70(5): 933–7. http://www.ncbi.nlm.nih.gov/pubmed/18063318 (October 8, 2011).
- Nei, M, and R K Chesser. 1983. "Estimation of fixation indices and gene diversities." *Annals of human genetics* 47(Pt 3): 253–9. http://www.ncbi.nlm.nih.gov/pubmed/6614868 (September 13, 2013).
- Nei, Masatoshi. 1987. "Molecular Evolutionary Genetics."
- Ovsyannikova, Inna G, Robert A Vierkant, V Shane Pankratz, Megan M O'Byrne, Robert M Jacobson, and Gregory A Poland. 2009. "HLA haplotype and supertype associations with cellular immune responses and cytokine production in healthy children after rubella vaccine." *Vaccine* 27(25-26): 3349–58. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2693336& tool=pmcentrez&rendertype=abstract (September 13, 2013).
- Parham, P, K L Arnett, E J Adams, A M Little, K Tees, L D Barber, S G Marsh, T Ohta, T Markow, and M L Petzl-Erler. 1997. "Episodic evolution and turnover of HLA-B in the indigenous human populations of the Americas." *Tissue antigens* 50(3): 219–32. http://www.ncbi.nlm.nih.gov/pubmed/9331945 (December 16, 2011).
- Parham, Peter. 2005. "MHC class I molecules and KIRs in human history, health and survival." *Nature reviews. Immunology* 5(3): 201–14. http://www.ncbi.nlm.nih.gov/pubmed/15719024 (July 15, 2012).

- Piertney, S B, and M K Oliver. 2006. "The evolutionary ecology of the major histocompatibility complex." *Heredity* 96(1): 7–21. http://www.ncbi.nlm.nih.gov/pubmed/16094301 (June 20, 2011).
- Prevosti, A., J. Ocaña, and G. Alonso. 1975. "Distances between populations of Drosophila subobscura, based on chromosome arrangement frequencies." *Theoretical and Applied Genetics* 45(6): 231–241. http://link.springer.com/10.1007/BF00831894 (August 16, 2013).
- Prugnolle, Franck, Andrea Manica, Marie Charpentier, Jean François Guégan, and François Balloux. 2007. "UKPMC Funders Group Pathogen-driven selection and worldwide HLA class I diversity." *Genetics* 15(11): 1022–1027.
- Prugnolle, Franck, Andrea Manica, Marie Charpentier, Jean François Guégan, Vanina Guernier, and François Balloux. 2005. "Pathogendriven selection and worldwide HLA class I diversity." *Current biology: CB* 15(11): 1022–7. http://www.ncbi.nlm.nih.gov/pubmed/15936272 (July 30, 2012).
- Qutob, Nouar, Francois Balloux, Towfique Raj, Hua Liu, Sophie Marion de Procé, John Trowsdale, and Andrea Manica. 2011. "Signatures of historical demography and pathogen richness on MHC class I genes." *Immunogenetics*. http://www.ncbi.nlm.nih.gov/pubmed/21947542 (November 30, 2011).
- Reich, David, Nick Patterson, Desmond Campbell, Arti Tandon, Stéphane Mazieres, Nicolas Ray, et al. 2012. "Reconstructing Native American population history." *Nature*: 1–6. http://www.nature.com/doifinder/10.1038/nature11258 (July 12, 2012).
- Robinson, James. 2003. "IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex." *Nucleic Acids Research* 31(1): 311–314. http://nar.oxfordjournals.org/content/31/1/311.short (July 6, 2011).
- Robinson, James, Kavita Mistry, Hamish McWilliam, Rodrigo Lopez, Peter Parham, and Steven G E Marsh. 2011. "The IMGT/HLA database." *Nucleic acids research* 39(Database issue): D1171–6. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3013815& tool=pmcentrez&rendertype=abstract (September 1, 2013).
- Sanchez-Mazas, Alicia, Jean-François Lemaître, and Mathias Currat. 2012. "Distinct evolutionary strategies of human leucocyte antigen loci in

- pathogen-rich environments." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 367(1590): 830–9. http://www.ncbi.nlm.nih.gov/pubmed/22312050 (August 19, 2012).
- Saper, M A, P J Bjorkman, and D C Wiley. 1991. "Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution." *Journal of molecular biology* 219(2): 277–319. http://www.ncbi.nlm.nih.gov/pubmed/2038058 (December 16, 2011).
- Sette, a, and J Sidney. 1999. "Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism." *Immunogenetics* 50(3-4): 201–12. http://www.ncbi.nlm.nih.gov/pubmed/10602880.
- Shehzadi, Abida, Shahid Ur Rehman, and Tayyab Husnain. 2012. "Selection of epitope-based vaccine targets of HCV genotype 1 of Asian origin: a systematic in silico approach." *Bioinformation* 8(20): 957–62. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3524940& tool=pmcentrez&rendertype=abstract (September 13, 2013).
- Sidney, John, Howard M Grey, Ralph T Kubo, and Alessandro Sette. 1996. "Practical, biochemical and evolutionary implications of the discovery of HLA class I supermotifs." *Immunology today* 17(6): 261–266.
- Sidney, John, Bjoern Peters, Nicole Frahm, Christian Brander, and Alessandro Sette. 2008. "HLA class I supertypes: a revised and updated classification." *BMC immunology* 9: 1. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2245908& tool=pmcentrez&rendertype=abstract (July 17, 2011).
- Slatkin, M, and L Excoffier. 1996. "Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm." *Heredity* 76 (Pt 4): 377–83. http://www.ncbi.nlm.nih.gov/pubmed/8626222 (August 26, 2013).
- Solberg, Owen D, Steven J Mack, Alex K Lancaster, Richard M Single, Yingssu Tsai, Alicia Sanchez-Mazas, and Glenys Thomson. 2008. "Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies." *Human immunology* 69(7): 443–64. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2632948& tool=pmcentrez&rendertype=abstract (June 13, 2011).
- Tajima, F. 1983. "Evolutionary relationship of DNA sequences in finite populations." *Genetics* 105(2): 437–60.

- http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1202167&tool=pmcentrez&rendertype=abstract.
- Takahata, N, Y Satta, and J Klein. 1992. "Polymorphism and balancing selection at major histocompatibility complex loci." *Genetics* 130(4): 925–38. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1204941& tool=pmcentrez&rendertype=abstract (September 13, 2012).
- The MHC sequencing consortium. 1999. "Complete sequence and gene map of a human major histocompatibility complex." 401(October): 921–923.
- Trachtenberg, Elizabeth, Bette Korber, Cristina Sollars, Thomas B Kepler, Peter T Hraber, Elizabeth Hayes, Robert Funkhouser, Michael Fugate, James Theiler, Yen S Hsu, Kevin Kunstman, Samuel Wu, John Phair, Henry Erlich, and Steven Wolinsky. 2003. "Advantage of rare HLA supertype in HIV disease progression." *Nature medicine* 9(7): 928–35. http://www.ncbi.nlm.nih.gov/pubmed/12819779 (September 13, 2013).
- Wang, Sijia, Cecil M Lewis, Mattias Jakobsson, Sohini Ramachandran, Nicolas Ray, Gabriel Bedoya, Winston Rojas, Maria V Parra, Julio A Molina, Carla Gallo, Guido Mazzotti, Giovanni Poletti, Kim Hill, Ana M Hurtado, Damian Labuda, William Klitz, Ramiro Barrantes, Maria Cátira Bortolini, Francisco M Salzano, Maria Luiza Petzl-Erler, Luiza T Tsuneto, Elena Llop, Francisco Rothhammer, Laurent Excoffier, Marcus W Feldman, Noah A Rosenberg, and Andrés Ruiz-Linares. 2007. "Genetic variation and population structure in native Americans." *PLoS genetics* 3(11): e185. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2082466& tool=pmcentrez&rendertype=abstract (August 7, 2013).
- Watterson, G A. 1978. "The homozygosity test of neutrality." *Genetics* 88(2): 405–17. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1213809& tool=pmcentrez&rendertype=abstract (December 16, 2011).
- Wright, Sewall. 1978. Evolution and the Genetics of Populations. Chicago: University of Chicago.
- Xavier Eurico de Alencar, Liciana, Ulisses de Mendonça Braga-Neto, Eduardo José Moura do Nascimento, Marli Tenório Cordeiro, Ana Maria Silva, Carlos Alexandre Antunes de Brito, Maria da Paz Carvalho

da Silva, Laura Helena Vega Gonzales Gil, Silvia Maria Lucena Montenegro, and Ernesto Torres de Azevedo Marques. 2013. "HLA-B*44 Is Associated with Dengue Severity Caused by DENV-3 in a Brazilian Population." *Journal of tropical medicine* 2013: 648475. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3684019& tool=pmcentrez&rendertype=abstract (September 13, 2013).