



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA

DISSERTAÇÃO DE MESTRADO
**AVALIAÇÃO DA HISTÓRIA EVOLUTIVA DO GENE *HLA-G* POR MEIO DE
POLIMORFISMOS DE BASE ÚNICA E DA INSERÇÃO *ALU*YHG**

Aluno: Kaison Ernane dos Santos
Orientador: Prof. Dr. Erick da Cruz Castelli

Goiânia, outubro de 2013.

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR AS TESES E DISSERTAÇÕES ELETRÔNICAS NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: **Dissertação** **Tese**

2. Identificação da Tese ou Dissertação

Nome completo do autor: Kaison Ernane dos Santos

Título do trabalho: Avaliação da história evolutiva do gene *HLA-G* por meio de polimorfismos de base única e da inserção *Alu*YG.

3. Informações de acesso ao documento:

Concorda com a liberação total do documento SIM NÃO¹

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

Kaison Ernane dos Santos

Assinatura do (a) autor (a)

Data: 03 / 01 / 2016

¹ Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA

DISSERTAÇÃO DE MESTRADO
**AVALIAÇÃO DA HISTÓRIA EVOLUTIVA DO GENE *HLA-G* POR MEIO DE
POLIMORFISMOS DE BASE ÚNICA E DA INSERÇÃO *ALU*YHG**

Dissertação de Mestrado realizada sob a orientação do Prof. Dr. Erick da Cruz Castelli, apresentada ao Programa de Pós-graduação em Biologia (área de concentração Biologia Celular e Molecular) como pré-requisito para a obtenção do título de Mestre em Biologia.

Aluno: Kaison Ernane dos Santos
Orientador: Prof. Dr. Erick da Cruz Castelli

Goiânia, outubro de 2013.

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Ernane dos Santos, Kaison
Avaliação da história evolutiva do gene HLA-G por meio de Polimorfismos de Base Única e da inserção AluYHG [manuscrito] / Kaison Ernane dos Santos. - 2013.
LXVIII, 68 f.: il.

Orientador: Prof. Dr. Erick da Cruz Castelli.
Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Ciências Biológicas (ICB), Programa de Pós-Graduação em Biologia, Goiânia, 2013.

Bibliografia. Anexos.

Inclui siglas, abreviaturas, tabelas, lista de figuras, lista de tabelas.

1. Complexo Principal de Histocompatibilidade. 2. Antígenos Leucocitários Humanos. 3. HLA-G. 4. AluYHG. 5. Desequilíbrio de Ligação. I, da Cruz Castelli, Erick, orient. II. Título.

CDU 57



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA

ATA DE DEFESA PÚBLICA DE DISSERTAÇÃO
Nº 408

Ata da reunião da banca examinadora da defesa de dissertação em nível de mestrado – Aos vinte e cinco dias do mês de Novembro do ano de 2013 (25/11/2013), às 14h00min, reuniram-se os componentes da Banca Examinadora: Prof. Dr. Erick da Cruz Castelli ; Prof^ª. Dra. Thannya Nascimento Soares; Prof^ª. Dra. Daniela Melo e Silva, para sob a presidência do primeiro, e em sessão pública realizada no Mini Auditório do Instituto de Ciências Biológicas IV, da Universidade Federal de Goiás, procederem à avaliação da defesa de dissertação intitulada: “**AVALIAÇÃO DA HISTÓRIA EVOLUTIVA DO GENE HLA-G POR MEIO DE POLIMORFISMOS DE BASE ÚNICA E DA INSERÇÃO DE Alu YHG**” em nível de mestrado, de autoria de **KAISSON ERNANE DOS SANTOS**, discente do Programa de Pós-Graduação em Biologia da Universidade Federal de Goiás. A sessão foi aberta pelo presidente da banca examinadora Prof. Dr. Erick da Cruz Castelli que fez a apresentação formal dos membros da banca. A palavra a seguir foi concedida ao autor da dissertação que, em...25:..... minutos procedeu à apresentação de seu trabalho. Terminada a apresentação, cada membro da banca arguiu a examinada, tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se à avaliação da defesa. Tendo-se em vista o que consta na Resolução nº. 572/2002 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta o Programa de Pós-Graduação em Biologia e procedidas às correções recomendadas, a dissertação foi...aprovada..... por unanimidade, considerando-se

33 integralmente cumprido o requisito para fins de obtenção do título de MESTRE
34 EM BIOLOGIA, na área de concentração em Biologia Celular e Molecular pela
35 Universidade Federal de Goiás. A conclusão do curso dar-se-á quando da entrega
36 na secretária do programa da versão definitiva da dissertação, com as devidas
37 correções, com as devidas correções, no prazo de trinta dias, a contar da data da
38 defesa, procedidas as modificações sugeridas pela banca avaliadora. Cumpridas
39 as formalidades de pauta, às ..16..h e ..30.. min., a presidência da mesa encerrou
40 esta sessão de defesa de dissertação de mestrado e para constar eu, Gleizilene
41 Braz Pereira dos Santos, Assistente em Administração, da Universidade Federal
42 de Goiás, lavrei a presente ata, que após lida e aprovada, será assinada pelos
43 membros da Banca Examinadora em três vias de igual teor.

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

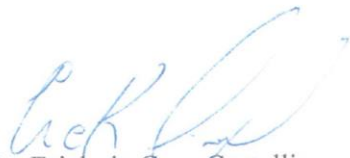
63


64

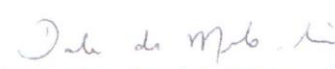
65

66

67


Prof. Dr. Erick da Cruz Castelli
Universidade Estadual Paulista


Prof.ª Dra. Thannya Nascimento Soares
Universidade de Federal de Goiás


Prof.ª Dra. Daniela Melo e Silva
Universidade de Federal de Goiás

*A*os meus pais, José Carlos e

*Líramênia, e ao meu irmão,
Kleisson, dedico...*

*P*rocure ser um homem de
valor, em vez de ser um
homem de sucesso.

Albert Einstein

Agradeço a Deus pela dádiva da vida. Aos meus pais pelo amor incondicional e pela educação que recebi: à minha mãe que por várias vezes se privou do próprio conforto para poder agradecer seus filhos, que sempre estava aqui quando mais precisei, com quem sempre contei em todos os momentos da minha vida, e que sempre foi exemplo de dedicação à família; a meu pai, que com seu jeito alegre e extrovertido, mas responsável ao extremo, dedicado ao bem estar dos outros e quase nunca ao seu próprio, nunca nos deixou faltar nada, nos ofertando as oportunidades que a vida não lhe deu, projetando em seus filhos aquilo que gostaria de ter tido. A eles que são o meu exemplo do que “quero ser quando crescer”.

A meu irmão que por várias vezes deixou de fazer suas próprias vontades para permitir que eu estudasse ou tivesse algum privilégio. Muito obrigado pela compreensão.

Ao meu orientador, Prof. Dr. Erick da Cruz Castelli, pela oportunidade de ingressar na pós-graduação, pelo voto de confiança que me deu ao me ofertar uma carta de orientação, pelo exemplo em pesquisa e pela paciência durante a sua orientação. Espero ter correspondido às expectativas em mim depositadas.

Aos meus avós por todo o incentivo, apoio, suporte, energia positiva e por representarem meu porto seguro.

Ao meu tio Nelson que foi, de longe, o maior incentivador da realização da minha pós-graduação e a quem considero como o grande exemplo de vida e luta profissional e familiar.

A minha melhor amiga, antes de tudo, Natália pelo incentivo, apoio e bondade durante os últimos 7 anos.

Aos meus queridos amigos de laboratório: Amanda, Andréia, Athamy, Iane, Denise, Karla, Lais, Leandro, Lya, Mariana, Moisés, Pedro, Thállita e Thiago pelo convívio, pelo auxílio na bancada, pelos momentos de descontração e de ‘nerdices’ e por ajudar a fazer do laboratório uma extensão de minha casa.

Ao Prof. Dr. Paulo César Gedhini, por nos acolher tão bem e nos oferecer um espaço físico dentro do Laboratório de Farmacologia Bioquímica e Molecular.

Aos prezados Ana Carolina Arcanjo, Silviene Oliveira, Philippe Moureau, Andre Garcia, Audrey Sabbagh, Eduardo Donadi, Celso Mendes-Junior, Juliana Massaro e Gustavo Palomino pela grande ajuda durante a escrita do artigo e por parte das amostras avaliadas nesse trabalho. Muito obrigado pela grande ajuda.

Ao Programa de Pós-Graduação em Biologia da Universidade Federal de Goiás e ao seu corpo docente pelos ensinamentos.

Aos membros da banca por se disporem a oferecer seu tempo para contribuir com a minha formação.

A CAPES por fornecer o apoio financeiro e a bolsa de mestrado. E a todos aqueles que de forma direta ou indireta contribuíram para que meu trabalho pudesse ser concluído.

SUMÁRIO

LISTA DE FIGURAS.....	9
LISTA DE TABELAS.....	11
LISTA DE ABREVIATURAS E SIGLAS.....	12
RESUMO.....	13
ABSTRACT.....	14
1. INTRODUÇÃO.....	15
1.1 - FUNCIONAMENTO, ESTRUTURA E HISTÓRICO DO COMPLEXO PRINCIPAL DE HISTOCOMPATIBILIDADE HUMANO.....	15
1.2 - O GENE HLA-G.....	20
1.2.1. CARACTERÍSTICAS GERAIS.....	20
1.2.2. SELEÇÃO NATURAL E O GENE <i>HLA-G</i>	24
1.3 - A SEQUÊNCIA <i>AluYHG</i>	26
1.4. O PROJETO 1000Genomes CONSORTIUM.....	29
2. OBJETIVOS.....	30
2.1. OBJETIVO GERAL.....	30
2.2. OBJETIVOS ESPECÍFICOS.....	30
3. JUSTIFICATIVA.....	31
4. MATERIAL E MÉTODOS.....	32
4.1. CARACTERIZAÇÃO DA AMOSTRA.....	32
4.2. CARACTERIZAÇÃO DA INSERÇÃO <i>ALUYHG</i>	33
4.3. VARIABILIDADE DA REGIÃO CODIFICADORA E 3'NT DO GENE HLA-G.....	35
4.4. NOMENCLATURA DOS HAPLÓTIPOS HLA-G.....	36
4.5. FORMATAÇÃO DOS BANCOS DE DADOS.....	37
4.6. ANÁLISE DOS DADOS.....	38
5. RESULTADOS.....	41
5.1. FREQUÊNCIAS ALÉLICAS E GENOTÍPICAS DA INSERÇÃO <i>ALUYHG</i>	41
5.2. DESEQUILÍBRIO DE LIGAÇÃO.....	42
5.3. HAPLÓTIPOS HLA-G/ <i>ALUYHG</i> E RELAÇÃO ENTRE OS HAPLÓTIPOS.....	44
6. DISCUSSÃO.....	52

7. CONCLUSÃO	60
8. REFERÊNCIAS.....	61
ANEXO	70

LISTA DE FIGURAS

Figura 1	Estrutura da região do MHC humano, representando os genes MHCdeclassel,IIeIII(Shiina <i>et al.</i> , 2009).....	17
Figura 2	Estrutura comparativa das moléculas de MHC de classe I e classe II (Abbas, 2010).....	18
Figura 3	Isoformas solúveis e ligadas à membrana celular da molécula HLA-G. As cores dos éxons são as mesmas encontradas em cada domínio da molécula produzida. (Donadi <i>et al.</i> , 2011).....	22
Figura 4	Mapa das posições das inserções <i>Alu</i> dentro do MHC, evidenciando a inserção <i>AluyHG</i> . Adaptado de (Kulski e Dunn, 2005).....	28
	..	
Figura 5	Gel de agarose apresentando bandas específicas para a inserção <i>AluyHG</i>	35
Figura 6	Perfil de LD entre os polimorfismo da região 3'NT do gene <i>HLA-G</i> e a inserção <i>AluyHG</i> considerando dados de genótipos de indivíduos de quatro populações distintas (brasileira, senegalesa, congoleza e francesa).....	43
Figura 7	Perfil de LD entre os polimorfismos da região codificadora e 3'NT do gene <i>HLA-G</i> , considerando dados de genótipos de indivíduos brasileiros e das 14 populações disponibilizadas pelo Projeto <i>1000Genomes</i>	44
Figura 8	Rede de haplótipos construída a partir dos dados de polimorfismos da região 3'NT do gene <i>HLA-G</i> e a inserção <i>AluyHG</i> considerando os dados das populações brasileira, senegalesa, congoleza e francesa. As cores diferentes indicam linhagens diferentes desses haplótipos.....	50
Figura 9	Rede de haplótipos construída a partir dos dados de polimorfismos das regiões codificadora e 3'NT do gene <i>HLA-G</i>	

a partir dos dados disponibilizados pelo Projeto *1000Genomes*.
As linhagens têm as mesmas cores da Figura 7 e foram
nomeadas de acordo com (Castelli *et al.*,
2011).....

LISTA DE TABELAS

Tabela 1	Número de alelos identificados para os principais genes HLA de classe I e II (IMGT/HLA - <i>Database</i> 3.14.0, novembro de 2013).....	19
Tabela 2	Reagentes e concentrações utilizadas para a amplificação do elemento <i>Alu</i> YG.....	34
Tabela 3	Principais haplótipos da região 3'NT do gene <i>HLA-G</i>	37
Tabela 4	Frequências da ausência (<i>Alu</i> YG*1) e presença (<i>Alu</i> YG*2) da inserção <i>Alu</i> YG em populações mundiais.....	42
Tabela 5	Haplótipos entre polimorfismos da região 3'NT do gene <i>HLA-G</i> e a inserção <i>Alu</i> YG em quatro populações distintas: brasileiros, senegaleses, congolese e franceses.....	46
Tabela 6	Frequências dos haplótipos considerando polimorfismos da região codificadora e 3'NT do gene <i>HLA-G</i> na população mundial.....	48

LISTA DE ABREVIATURAS E SIGLAS

.vcf	(<i>Variant Call Format</i>)
%	Por cento
µL	Microlitro (10 ⁻⁶ litro)
APC	Célula Apresentadora de Antígeno
EDTA	Acido Etileno Diamino Tetracético
HLA	Antígeno Leucocitário Humano (do inglês, <i>human leukocyte antigens</i>)
IMGT	<i>International Immunogenetics Database</i>
Kb	Kilobase (10 ³ bases)
LTA	Linfotoxina alfa
Mb	Megabase (10 ⁶ bases)
MAF	Frequência alélica mínima (do inglês, <i>minimum allele frequency</i>)
MHC	Complexo Principal de Histocompatibilidade (do inglês, <i>Major Histocompatibility Complex</i>)
miRNA	MicroRNA
mL	Mililitro (10 ⁻³ Litro)
mRNA	RNA mensageiro
ng	Nanograma (10 ⁻⁹ grama)
NK	Célula <i>Natural Killer</i>
NT	Não traduzida
°C	Graus Celsius
Pb	Pares de bases
PCR	Reação em Cadeia da Polimerase
SNP	Polimorfismo de Base Única (do inglês, <i>Single Nucleotide Polymorphism</i>)
SP	São Paulo
TCR	Receptor de Célula T
TNF	Fator de Necrose Tumoral
UV	Ultravioleta
V	Volts

RESUMO

O Complexo Principal de Histocompatibilidade (MHC) é formado principalmente por genes que participam da resposta imunológica adaptativa. Entre esses genes encontramos o grupo denominado de Antígenos Leucocitários Humanos (HLA), que são responsáveis pela apresentação de antígenos específicos às células efetoras do sistema imunológico. Os genes HLA de classe I clássicos (*HLA-A*, *-B* e *-C*), responsáveis pela apresentação antigênica aos linfócitos T citotóxicos, são considerados como os mais polimórficos do genoma humano e de outros vertebrados. A variabilidade desses genes e elevada heterozigose é mantida por seleção mediada por microrganismos. Diferentemente dos genes clássicos, os genes HLA de classe I não clássicos (*HLA-G*, *-E* e *-F*) apresentam variabilidade reduzida e como função principal a tolerância imunológica, por meio de sua interação com receptores inibitórios presentes nas células NK e T. O *HLA-G* é o mais estudado entre esses genes e, devido sua importância como molécula imunomoduladora e sua importância em situações como gestação, e considerando evidências anteriores de seleção natural mantendo uma elevada heterozigose nas regiões regulatórias do *HLA-G*, avaliamos a presença de uma inserção *Alu* (*Alu*YG) próxima a este gene correlacionando os achados com a variabilidade contida nas suas regiões codificadora e 3' não traduzida. A inserção *Alu*YG mostrou-se em desequilíbrio de ligação (LD) com os polimorfismos do gene *HLA-G*. Especificamente, o elemento inserido apresentou-se em LD com um haplótipo denominado G*01:01:01:01/UTR-1, considerado como um haplótipo de alta produção da molécula de *HLA-G*. Esse haplótipo aparentemente é o mais jovem entre humanos, apesar de sua elevada frequência nas populações estudadas até o momento.

Palavras-Chave: Complexo Principal de Histocompatibilidade, Antígenos Leucocitários Humanos, *HLA-G*, *Alu*YG, Desequilíbrio de Ligação, Haplótipos.

ABSTRACT

The Major Histocompatibility Complex is mainly composed by genes of the adaptive immune response. In humans, part of this complex is known as the Human Leukocyte Antigens (HLA), whose genes are responsible for specific antigen presentation to effector immune cells. The classical class I HLA genes (HLA-A, -B and -C) are responsible for antigen presentation to T CD8+ cells and they constitute the most polymorphic genes in the human genome. This variability is maintained by selection mediated by microorganisms. In contrast to their classical counterparts, the non classical class I genes (HLA-G, -E and -F) present low variability and are associated with immune tolerance due to the interaction with NK and T cells inhibitor receptors. HLA-G is the most studied non classical gene, which is associated with immune response modulation, mainly during pregnancy. Considering that natural selection is acting on the HLA-G regulatory regions maintaining high heterozygosity in this region, we evaluated a nearby Alu insertion (AluyHG) correlating this Alu element with coding and 3'UTR HLA-G polymorphisms. The AluyHG insertion was particularly associated with the HLA-G haplotype known as G*01:01:01:01/UTR-1, considered a high-expressing HLA-G haplotype. The G*01:01:01:01/UTR-1/*AluyHG* haplotype would be the most recent HLA-G haplotypes, in spite of its high frequency in worldwide populations.

Key Words: Major Histocompatibility Complex, Human Leukocyte Antigens, *HLA-G*, *AluyHG*, Linkage Disequilibrium, Haplotypes.

1. INTRODUÇÃO

1.1 - FUNCIONAMENTO, ESTRUTURA E HISTÓRICO DO COMPLEXO PRINCIPAL DE HISTOCOMPATIBILIDADE HUMANO

Em vertebrados, a resposta imunitária pode ser dividida em resposta inata e adaptativa (Parkin e Cohen, 2001). A primeira ocorre de maneira imediata ao dano sofrido pelos tecidos ou ao contato com microrganismos e tem baixa especificidade, i.e., não é direcionada a um determinado microrganismo (Delves e Roitt, 2000a; 2000b). A resposta imune inata é conservada em vários organismos (Delves e Roitt, 2000a; 2000b), usualmente envolvendo as mesmas moléculas e células. No entanto, a resposta adaptativa envolve o reconhecimento de antígenos por receptores específicos nas células T e B, o que permite o desenvolvimento de uma resposta mais eficiente, específica e duradoura (Parkin e Cohen, 2001).

O primeiro estágio para que a resposta imune adaptativa ocorra é a apresentação dos antígenos específicos às células T (Klein e Sato, 2000a; 2000b; Parkin e Cohen, 2001), por meio de glicoproteínas transmembrânicas responsáveis por acomodar pequenos peptídeos em sua estrutura e expô-los na superfície celular (Klein e Sato, 2000a; 2000b). Essas glicoproteínas, em vertebrados, estão codificadas em genes agrupados em uma região genômica formando um complexo ou família gênica, denominado Complexo Principal de Histocompatibilidade (MHC, do inglês *Major Histocompatibility Complex*).

Sua descoberta deu-se durante ensaios de transplantes alogênicos realizados em camundongos por George Snell e colaboradores ainda na década de 40 (Abbas, 2010). Durante seus experimentos esses pesquisadores intercruzavam linhagens de camundongos por pelo menos 20 gerações a fim de obterem animais geneticamente idênticos divididos em dois grupos com características genéticas distintas. Então, tecido da pele retirado cirurgicamente era transplantado em animais singênicos e alogênicos. Após a recuperação notava-se que apenas havia resposta de rejeição no segundo grupo, o que indicava a participação de proteínas de compatibilidade na resposta imunitária desses indivíduos (Abbas, 2010).

Em humanos, a descoberta do MHC ocorreu por volta do ano de 1958 por meio do trabalho realizado por três pesquisadores, Jean Dausset, Jon van Rood e Rose Payne. Esses pesquisadores trabalhavam na identificação de anticorpos no soro de pacientes que recebiam múltiplas transfusões sanguíneas. Entretanto, o crédito pela descoberta do primeiro Antígeno Leucocitário Humano (HLA, do inglês *Human Leucocyte Antigen*), e por consequência do MHC humano, foi de Jean Dausset, que recebeu um prêmio Nobel em 1980 (Thorsby, 2009).

A principal função do MHC em nossa espécie é referida como a apresentação de antígenos aos linfócitos T durante o desenvolvimento da resposta imune adaptativa (Klein e Sato, 2000a; 2000b; Parkin e Cohen, 2001). Entretanto, estudos apontam que, além da deflagração da resposta adaptativa, algumas dessas moléculas estão também relacionadas à modulação do sistema imune humano (Moscoso *et al.*, 2006; Carosella, 2011; Donadi *et al.*, 2011). Alguns estudos apontam, ainda, o possível papel dessas moléculas na escolha sexual de parceiros em nossa espécie (Wedekind *et al.*, 1995; Jacob *et al.*, 2002).

Na espécie humana, o MHC está localizado no cromossomo 6 (6p21.3), apresentando aproximadamente 3,6 megabases (Mb) e 224 genes (Klein e Sato, 2000a; Garcia-Obregon *et al.*, 2011). Essa família de genes é, didaticamente, dividida em três grupos denominados genes de classe I, II e III (Horton *et al.*, 2004) (Figura 1). Na região de classe I e II está o sistema conhecido como HLA (do inglês, *Human Leukocyte Antigens*) (Shiina *et al.*, 2009). As moléculas de classe I e II, ou HLA de classe I e II, são responsáveis pela apresentação de antígenos às células T do sistema imune (Parkin e Cohen, 2001). Os genes de HLA classe I podem ainda ser divididos, de acordo com a sua função, em genes clássicos (Ia), representados pelos genes *HLA-A*, *-B* e *-C*, envolvidos na apresentação de antígenos aos linfócitos T (Klein e Sato, 2000a; 2000b), e genes não clássicos (Ib), representados pelos genes *HLA-E*, *-F* e *-G*, envolvidos na modulação do sistema imunitário (Geraghty *et al.*, 1992; Klein e Sato, 2000a; 2000b).

Estruturalmente, as moléculas de classe I, tanto clássicas como não clássicas, são formadas pela união de duas outras moléculas: uma cadeia α , codificada pelos genes MHC de classe I, e uma molécula β_2 -microglobulina, codificada por um gene localizado no cromossomo 15, fora do MHC (Klein e Sato, 2000a) (Figura 2). A cadeia α é dividida em cinco domínios: $\alpha 1$ e $\alpha 2$ (formam a fenda de ligação ao peptídeo), $\alpha 3$ (domínio extracelular do tipo imunoglobulina), um domínio transmembrana e um domínio citoplasmático (Klein e Sato, 2000a). Possivelmente essa semelhança estrutural ocorre devido ao fato de os genes desse complexo terem se originado por duplicações imperfeitas de blocos gênicos (Kulski *et al.*, 2000).

Por outro lado, as moléculas de classe II são formadas pela união de duas cadeias (α e β) ambas codificadas por genes situados no MHC humano (Klein e Sato, 2000a) (Figura 2). Cada uma das cadeias pode ser dividida em quatro domínios: $\alpha 1$ e $\beta 1$ que formam a fenda de ligação à peptídeos; $\alpha 2$ e $\beta 2$, extracelulares; um domínio transmembrana; e um domínio citoplasmático (Klein e Sato, 2000a).

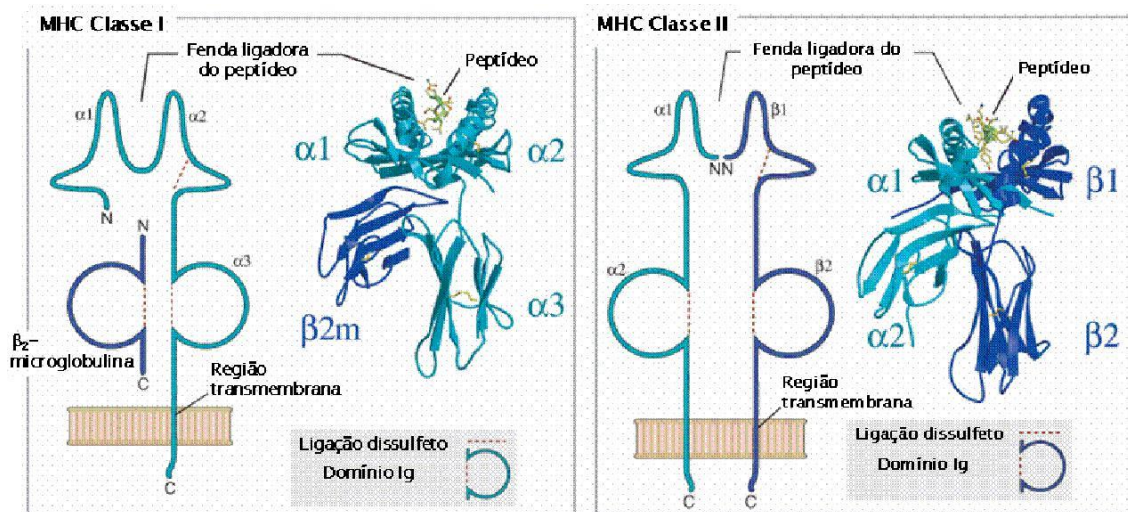


Figura 2 - Estrutura comparativa das moléculas de MHC de classe I e classe II (Abbas, 2010).

Devido ao seu papel de ligação e apresentação de peptídeos diversos, oriundos tanto de patógenos quanto do próprio organismo, tem-se considerado a presença de forças seletivas mantendo uma alta diversidade e heterozigotidade nos genes HLA classe I e II, aumentando o repertório de moléculas de classe I e II em

uma população, permitindo, assim, a apresentação de uma maior quantidade de epítomos (Meyer e Thomson, 2001; Segal e Hill, 2003; Piertney e Oliver, 2006; Solberg *et al.*, 2008). Nas moléculas de classe I e II, o polimorfismo está principalmente confinado aos sítios de ligação de peptídeos. Moléculas codificadas por diferentes alelos de um mesmo gene HLA apresentam diferentes sítios de ligação a peptídeos. Este polimorfismo é a razão da especificidade de ligação de cada molécula HLA a determinados peptídeos antigênicos (Parkin e Cohen, 2001).

Provavelmente devido a esse fato, o MHC apresenta-se como a região mais variável do genoma humano e da maioria dos vertebrados (Penn *et al.*, 2002). No entanto, este polimorfismo acentuado concentra-se em alguns genes importantes para a apresentação antigênica, como os genes clássicos de classe I, que em conjunto apresentam 7553 alelos ou haplótipos codificadores distintos (Tabela 1). Diversos estudos demonstraram a seleção a favor da variabilidade nos genes de classe Ia, que seriam os responsáveis pela manutenção da resposta imunitária adaptativa como descrito anteriormente (Meyer e Thomson, 2001; Segal e Hill, 2003; Piertney e Oliver, 2006; Solberg *et al.*, 2008; Donadi *et al.*, 2011). Nas espécies até hoje estudadas, essa variabilidade é interessante considerando a necessidade de combater patógenos. Por outro lado, os genes não clássicos de classe I, embora vizinhos dos genes mais variáveis do genoma humano, apresentam poucos alelos. Um estudo prévio mostrou que o gene de classe Ib *HLA-G*, por exemplo, apresenta este perfil de elevada heterozigose somente nas suas regiões regulatórias (Castelli *et al.*, 2011).

Tabela 1 - Número de alelos identificados para os principais genes HLA de classe I e II (IMGT/HLA - *Database* 3.14.0, outubro de 2013).

CLASSE I		CLASSE II		
Loco	Alelos	Loco		Alelos
Clássicos (Ia)		<i>HLA-DRA</i>	cadeia α do HLA-DR	7
<i>HLA-A</i>	2432	<i>HLA-DRB</i>	cadeia β 1 do HLA-DR	1476
<i>HLA-B</i>	3086	<i>HLA-DQA1</i>	cadeia α do HLA-DQ	51
<i>HLA-C</i>	2035	<i>HLA-DQB1</i>	cadeia β do HLA-DQ	459
		<i>HLA-DPA1</i>	cadeia α do HLA-DP	37
Não-Clássicos (Ib)		<i>HLA-DPB1</i>	cadeia β do HLA-DP	193
<i>HLA-E</i>	13	<i>HLA-DOA</i>	cadeia α do HLA-DO	12
<i>HLA-F</i>	22	<i>HLA-DOB</i>	cadeia β do HLA-DO	13
<i>HLA-G</i>	50	<i>HLA-DMA</i>	cadeia α do HLA-DM	7
		<i>HLA-DMB</i>	cadeia β do HLA-DM	13

No entanto, a capacidade de cada variante do MHC de acomodar peptídeos específicos tem sido estudada como um dos motivos que explicam a razão pela qual algumas variantes MHC estão associadas com susceptibilidade a doenças autoimunes e neoplásicas. Para tumores, por exemplo, postula-se que tipos parecidos de lesões expressam antígenos tumorais semelhantes e que algumas variantes MHC não seriam capazes de apresentar peptídeos oriundos de tais antígenos, pré-dispondo o indivíduo ao desenvolvimento de determinada neoplasia por falha da imunovigilância tumoral. Para doenças autoimunes, por exemplo, é bem conhecida teoria do Epítopo Compartilhado e artrite reumatoide, que prediz que diferentes moléculas de HLA podem carregar sequências de aminoácidos conservadas, modificando sua interação com os receptores da célula T (Gregersen *et al.*, 1987; De Almeida *et al.*, 2010). Assim, mesmo que haja diferentes moléculas capazes de apresentar diferentes antígenos, a estrutura molecular da interação entre HLA e receptor na célula T pode dirigir o tipo de resposta dada pelas células efectoras do sistema imunitário (Gregersen *et al.*, 1987).

Não obstante, a grande diversidade genética encontrada no HLA faz com que haja uma diferença na predisposição a enfermidades comparando-se grupos geneticamente distintos, fato este evolutivamente importante, pois torna mais difícil a ocorrência de grandes epidemias que poderiam dizimar a espécie (Van Rood, 1993).

1.2 - O GENE HLA-G

1.2.1. CARACTERÍSTICAS GERAIS

Em contraste com a expressão ubíqua das moléculas de HLA de classe I clássicas bem como com sua grande variabilidade genética, as moléculas HLA de classe I não clássicas, de maneira geral, ou apresentam sua expressão restrita aos tecidos em que há a necessidade de modulação da resposta imune (caso da molécula HLA-G) ou expressão ubíqua, porém reduzida (caso da molécula HLA-E), e uma variabilidade genética reduzida (Donadi *et al.*, 2011; Veiga-Castelli *et al.*, 2012). Apesar da sua menor influência na apresentação antigênica aos linfócitos T,

sua função na resposta imune tem sido intensamente estudada desde a descoberta do primeiro gene não clássico (*HLA-G*) por Geraghty e colaboradores entre 1987 e 1992 (Geraghty *et al.*, 1987; Geraghty *et al.*, 1992). Entre eles, o *HLA-G* tem sido alvo da maioria das pesquisas recentes.

Esse gene possui a mesma estrutura apresentada pelos genes de classe I (Donadi *et al.*, 2011). O primeiro éxon 1 traduzido codifica para o peptídeo sinal (que é utilizado na montagem da molécula de HLA-E, por exemplo), o segundo e terceiro éxons traduzidos codificam para a fenda de ligação a peptídeos (domínios $\alpha 1$ e $\alpha 2$, respectivamente), o quarto éxon traduzido codifica o domínio extracelular chamado $\alpha 3$, o quinto e sexto éxons traduzidos codificam a porção transmembrana e citoplasmática da molécula de HLA-G (Carosella *et al.*, 2008). Uma região 3' não traduzida (3'NT) é formada após um códon de parada prematuro localizado no sexto éxon traduzido e o éxon subsequente (Castelli *et al.*, 2010).

Esse gene produz uma molécula estruturalmente semelhante às moléculas produzidas pelos genes clássicos de classe I, entretanto apresenta algumas diferenças (Donadi *et al.*, 2011). Entre essas diferenças estão: (a) uma limitada variabilidade genética e proteica (Tabela 1), (b) apresentação de isoformas ligadas à membrana e solúveis, formadas por edição alternativa do transcrito primário (Figura 3) (Ishitani e Geraghty, 1992; Donadi *et al.*, 2011), (c) estrutura molecular única, com a presença de uma cauda citoplasmática curta devido à presença de um códon de parada prematuro (Donadi *et al.*, 2011), (d) é responsável pela modulação da resposta imune (tolerância imunológica) (Carosella *et al.*, 2008) e (e) apresenta expressão restrita a alguns tecidos (Donadi *et al.*, 2011).

Os estudos sobre esse gene têm revelado que os polimorfismos mais importantes para a função biológica dessas moléculas estão presentes em locais diferentes quando comparados aos genes clássicos, i.e., na fenda de ligação a peptídeos. Aqui, os principais polimorfismos encontrados na forma de SNP (do inglês, *Single Nucleotide Polymorphism*) estão presentes nas regiões promotora 5', influenciando sua transcrição (Moreau *et al.*, 2009), e na região 3'NT, influenciando a estabilidade do RNA mensageiro e sua tradução (Castelli *et al.*, 2010).

Ainda, diferentemente dos genes clássicos, em que os sítios de variação da região codificadora ocorrem principalmente nos éxons 2 e 3, os polimorfismos da região codificadora do *HLA-G* apresentam-se distribuídos, principalmente, entre os éxons 2, 3 e 4 bem como em seus íntrons (Donadi *et al.*, 2011). No entanto, a

maioria desses sítios polimórficos dentro de sequências exônicas constituem mutações sinônimas, gerando uma baixa diversidade proteica (Donadi *et al.*, 2011).

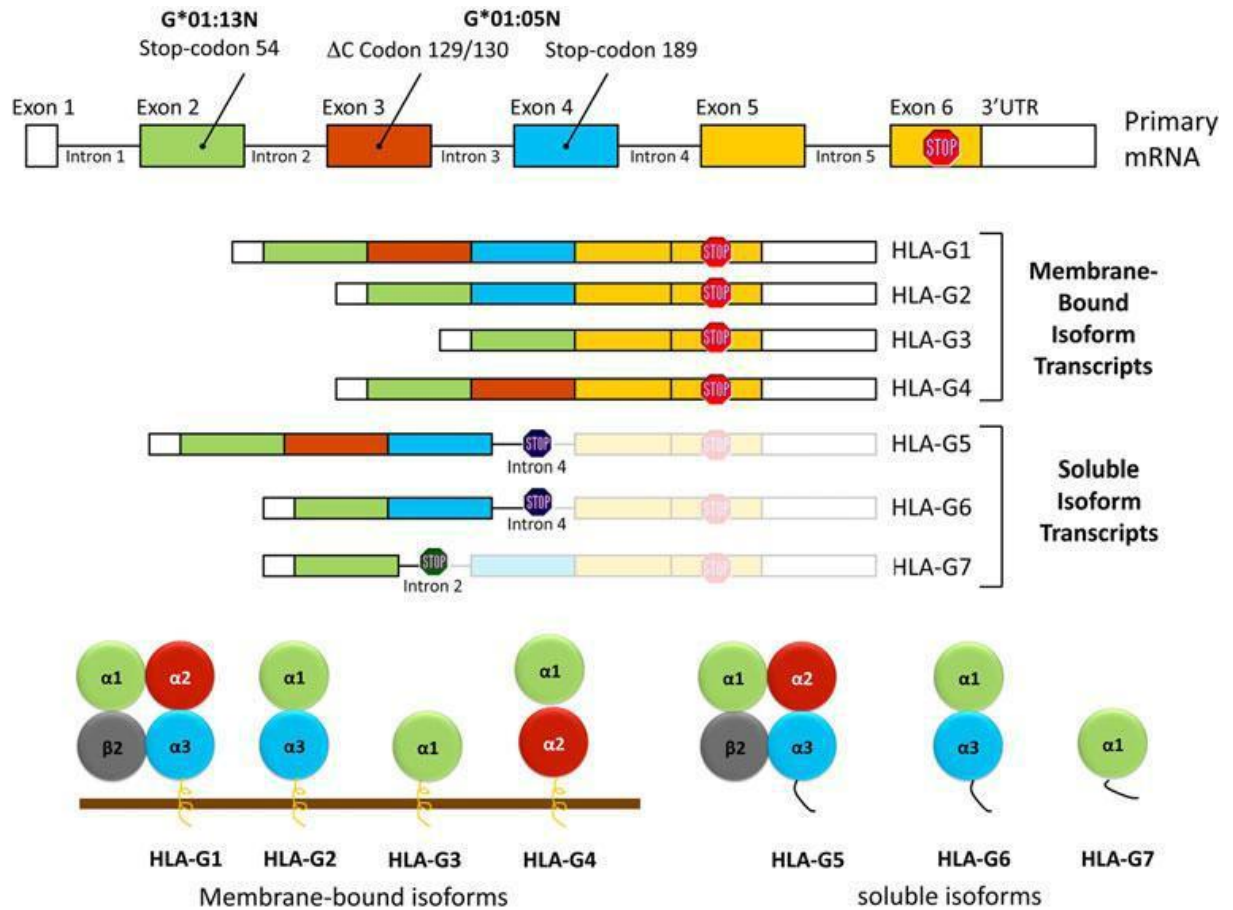


Figura 3 - Isoformas solúveis e ligadas à membrana celular do molécula HLA-G. As cores dos éxons são as mesmas encontradas em cada domínio da molécula produzida (Donadi *et al.*, 2011).

Recentemente, o perfil de polimorfismos das regiões regulatórias tem sido avaliado em outros genes não clássicos, notadamente o *HLA-E*, que tem apresentado uma variabilidade mais reduzida do que a encontrada no *HLA-G* (Veiga-Castelli *et al.*, 2012). É inegável, entretanto, a importância dos polimorfismos que ocorrem nessa região para a função desempenhada por essas moléculas, já que esses sítios de variação podem gerar isoformas solúveis e ligadas à membrana celular, modificando sua forma de interação com as células efectoras do sistema imune (Donadi *et al.*, 2011).

Com relação aos polimorfismos que fazem parte da região 3'NT do gene *HLA-G*, muitos esforços foram depositados para o entendimento da variabilidade da

região em si (Castelli *et al.*, 2010), bem como sua importância para a produção da molécula de HLA-G (Tan *et al.*, 2007; Yie *et al.*, 2008; Castelli *et al.*, 2009; Veit e Chies, 2009; Donadi *et al.*, 2011).

A região 3'NT do *HLA-G* foi pioneiramente caracterizada em brasileiros do Estado de São Paulo (Castelli *et al.*, 2010). Neste estudo, oito polimorfismos foram encontrados na região 3'NT, caracterizando sete haplótipos frequentes denominados UTR-1 a UTR-7. Posteriormente, outros grupos populacionais brasileiros foram explorados, indicando que o perfil observado em São Paulo também era observado em outras regiões (Mendes *et al.*, 2007; Lucena-Silva *et al.*, 2012). Em seguida, este perfil brasileiro tornou-se referência mundial, visto que diversos outros trabalhos mostraram que os mesmos polimorfismos e haplótipos eram frequentemente encontrado em diversas populações mundiais (Ober e Aldrich, 1997; Hviid *et al.*, 1999; Hviid *et al.*, 2004; Tan *et al.*, 2005; Hviid, 2006; Larsen e Hviid, 2009; Cervera *et al.*, 2010; Castelli *et al.*, 2011; Donadi *et al.*, 2011; Lucena-Silva *et al.*, 2012; Courtin *et al.*, 2013; Di Cristofaro *et al.*, 2013; Garcia *et al.*, 2013; Martinez-Laso *et al.*, 2013; Sabbagh *et al.*, 2013).

É bem caracterizado que polimorfismos presentes na região 3'NT podem influenciar na tradução de um determinado RNA mensageiro. Os principais mecanismos que regulam a tradução e que agem sobre a região 3'NT são: a ligação de proteínas específicas que dirigem a localização e a degradação da molécula de RNA mensageiro, e, nesse contexto, um dado polimorfismo poderá modificar essas regiões aumentando ou diminuindo a afinidade dessas proteínas àquele sítio (Kuersten e Goodwin, 2003); a taxa de decaimento do RNA mensageiro maduro, provocada por variantes deletérias (Kuersten e Goodwin, 2003); a ligação a microRNAs (miRNA) específicos, que pode levar o RNA mensageiro à degradação por diferentes vias (Bartel, 2004).

Em relação ao gene *HLA-G*, foi demonstrado que a presença de uma Guanina na posição +3142 pode favorecer a ligação a determinados microRNAs (miR-148a, miR-148b, miR-152, por exemplo), que podem dirigir à degradação desse RNA mensageiro (Tan *et al.*, 2007). Além disso, a presença de um fragmento de inserção/deleção de 14 pares de base (14pb), geralmente encontrado em desequilíbrio de ligação (LD) com o polimorfismo +3142, também pode influenciar a estabilidade do RNA mensageiro maduro (Hviid *et al.*, 2003). Nesse contexto, a presença do fragmento de 14pb tem sido associada a uma baixa produção da

molécula de HLA-G em trofoblastos (Hviid *et al.*, 2003). Por outro lado, a presença da sequência de 14pb pode levar a uma edição alternativa do RNA mensageiro, que culmina com a retirada de 92 nucleotídeos que incluem alguns polimorfismos, como o 14pb, +3003 e o +3010 (Hviid *et al.*, 2003). Assim, esses autores comentam que esse evento gera uma molécula menor e mais estável de RNA mensageiro. Ainda, a presença de uma Adenina na posição +3187 aumenta um motivo rico em AU (Adenina e Uracila), diminuindo a estabilidade do RNA mensageiro e favorecendo sua degradação (Yie *et al.*, 2008).

A caracterização dessas moléculas passa a ter importância na função biológica desempenhada pelo HLA-G no contexto da resposta imune. Os estudos realizados até agora apontam para a característica de modulação da resposta imunológica apresentada pelas moléculas de HLA-G, em contrapartida de seus correlatos clássicos (Donadi *et al.*, 2011). Várias são as peculiaridades que fazem desse gene um importante regulador do sistema imunológico: (a) a cauda citoplasmática mais curta prolonga a meia vida da molécula na membrana biológica (Park *et al.*, 2001; Park e Ahn, 2003), o que permite a interação com várias células pertencentes ao sistema imune e por mais tempo; (b) os domínios extracelulares interagem com receptores inibitórios nos linfócitos, como CD8, LILRB1 e LILRB2, e células NK (do inglês, *Natural Killer*), como KIR2DL4 (Gao *et al.*, 2000; Shiroishi *et al.*, 2003; Shiroishi *et al.*, 2006). Assim, as moléculas produzidas por esse gene passam a ter grande importância em contextos fisiológicos em que há a necessidade de modulação fina do sistema imune, como por exemplo, durante a gestação, em doenças autoimunes, transplantes e doenças crônicas, como tumores (Donadi *et al.*, 2011).

1.2.2. SELEÇÃO NATURAL E O GENE *HLA-G*

Evolutivamente, o gene *HLA-G* comporta-se de maneira diversa de seus correlatos clássicos. Segundo Castelli *et al.* (2011), há evidência de seleção purificadora atuando sobre a região codificadora desse gene (Castelli *et al.*, 2011), enquanto que nos genes clássicos de classe I há evidência de seleção balanceadora atuando sobre essa região (Solberg *et al.*, 2008; Donadi *et al.*, 2011). A baixa

variabilidade do gene *HLA-G* ganha grande importância quando passamos a analisar a função desempenhada pelas moléculas não clássicas de classe I (Castelli *et al.*, 2011; Donadi *et al.*, 2011). Sua função precípua não é a apresentação antigênica e sim a modulação da resposta imune, por meio da interação com receptores inibitórios menos variáveis do que os receptores para reconhecimento de antígenos (Contini *et al.*, 2003a; Contini *et al.*, 2003b).

Entretanto, as duas regiões regulatórias (5'URR e 3'NT) do gene *HLA-G* apresentam evidência de seleção balanceadora em diferentes populações com história de formação distintas (Tan *et al.*, 2005; Mendes *et al.*, 2007; Castelli *et al.*, 2011; Donadi *et al.*, 2011; Veit *et al.*, 2012; Mendes-Junior *et al.*, 2013). De fato, o perfil de distribuição de haplótipos desse gene aponta para o fato de que há poucos haplótipos em altas frequências em torno do mundo (Donadi *et al.*, 2011), geralmente ocorrendo uma heterozigose entre um haplótipo de alta produção e um de baixa produção da molécula de HLA-G (Mendes *et al.*, 2007; Castelli *et al.*, 2010; Castelli *et al.*, 2011; Donadi *et al.*, 2011). De fato, é observado que o haplótipo conhecido como UTR-1, descrito inicialmente no Brasil (Castelli *et al.*, 2010), considerado como de alta produção da molécula de HLA-G (Martelli-Palomino *et al.*, 2013), é frequentemente encontrado em heterozigose com um haplótipo considerado de baixa produção dessa molécula, conhecido como UTR-2.

Funcionalmente, a heterozigose tem importância devido ao fato de a resposta imunitária ter que atuar ora de maneira ávida, ora de maneira permissiva, i.e., durante a infecção por qualquer patógeno, por exemplo, há a necessidade de forte atuação das células efetoras imunológicas, com o intuito de eficazmente combater a infecção. Nesse sentido, a baixa produção da molécula de HLA-G, que tem função modulatória e, portanto, de aumentar a tolerância imunológica (Donadi *et al.*, 2011), torna-se importante a fim de permitir a correta atuação das células efetoras. Por outro lado, há casos em que há necessidade de modulação e tolerância imunológica, como, por exemplo, durante a gestação (Hviid, 2006). Nesse contexto, para que haja a aceitação materno-fetal, as células efetoras do sistema imunitário devem ser reguladas e a presença de haplótipos de alta produção da molécula de HLA-G garante a produção dessa molécula. Assim, a presença de força seletiva em favor da manutenção de heterozigose nessa região toma grande importância.

Esse mesmo perfil também é observado para a região promotora desse gene (Tan *et al.*, 2005). Novamente, há a presença de alguns polimorfismos associados à

alta produção da molécula de HLA-G em heterozigose com aqueles associados à baixa produção dessa molécula (Donadi *et al.*, 2011), também favorecendo a dualidade de respostas necessárias em condições fisiológicas normais. Além disso, já foi descrito que há forte LD entre sítios polimórficos da região promotora e da região 3'NT, o que aponta para a importância dessas duas regiões atuando em conjunto para a correta regulação da produção da molécula de HLA-G em locais e tempos em que há a necessidade (Donadi *et al.*, 2011). Entretanto, mais estudos são necessários para que se tenha o conhecimento completo de como a seleção molda as frequências dos haplótipos desse gene em torno do mundo.

1.3 - A SEQUÊNCIA *Alu*YG

Na tentativa de elucidar alguns mecanismos evolutivos envolvidos na formação dos genes HLA de classe I, realizaram-se estudos da presença ou ausência de inserções *Alu* presentes nesse complexo gênico (Kulski e Dunn, 2005). Em humanos, as sequências *Alu* são os mais abundantes representantes de uma família de retrotransposons chamados de SINE (do inglês, *Short Interspersed Nuclear Element*), e receberam esse nome devido à presença de um sítio de restrição para a enzima *Alu* em sua estrutura (Houck *et al.*, 1979). Os elementos *Alu* são compostos por aproximadamente 300pb, divididos em dois braços praticamente idênticos, separados por uma região rica em adeninas (A) (Batzer e Deininger, 2002). O motivo rico em Adenina presente no centro da estrutura geralmente possui a repetição A₅TACA₆ (Batzer e Deininger, 2002). As sequências *Alu* possuem uma cauda terminal oligo-(dA) de tamanho variável e repetições curtas de Timinas e Adeninas (T + A) flanqueando a sequência, refletindo sua inserção cromossomal (Jurka, 1997). Como parte da sequência, há a presença de um promotor para a enzima RNA polimerase III (Batzer e Deininger, 2002), necessária à sua retrotranscrição.

Esses elementos móveis são derivados, ancestralmente, do RNA 7SL, envolvido na ligação com o peptídeo sinal para a translocação de proteínas ao retículo endoplasmático (Ullu e Tschudi, 1984). Essa função tem sido atribuída,

ainda, a alguns elementos *Alu* distribuídos no genoma humano (Weichenrieder *et al.*, 2001).

O seu mecanismo de transposição envolve a retrotranscrição de um produto originado a partir da ação da enzima RNA polimerase III, provavelmente pela ação de uma transcriptase reversa codificada por uma sequência do tipo LINE (do inglês, *Long interspersed Nuclear Elements*) (Mathias *et al.*, 1991), já que os elementos *Alu* perderam, ao longo do processo evolutivo, a sequência que codifica essa enzima (Batzer e Deininger, 2002). Esses autores comentam, ainda, a importância de regiões conservadas (TTAAA) flanqueando a sequência *Alu* necessárias ao reconhecimento por endonucleases para que ocorra o processo de integração ao genoma.

Há cerca de 1,1 milhão de cópias de sequências *Alu*, o que corresponde a aproximadamente 10% de todo o genoma haplóide, presentes no genoma humano (Kriegs *et al.*, 2007). Entre as sequências presentes na espécie humana existem aquelas mais antigas, compartilhadas com outros primatas, chamadas de *AluJ* e *AluS* (Jurka e Smith, 1988) e algumas sequências específicas da espécie humana, notadamente as de inserções mais recentes, chamadas *AluY* (Batzer *et al.*, 1994). Para a diferenciação das famílias de inserções *Alu* há consenso em utilizarem-se diferenças nas suas sequências, como algumas mutações que são características para cada família e acumuladas durante o processo evolutivo (Jurka e Smith, 1988; Batzer *et al.*, 1996; Kapitonov e Jurka, 1996).

Dentre as famílias *Alu* descritas acima, a *AluY* mostra-se a mais útil no estudo populacional, devido ao fato de algumas sequências dessa família ainda não terem alcançado sua completa fixação no genoma (Kulski *et al.*, 2001), sendo dimórficos entre a população, ou seja, ou estão presentes ou ausentes (Dunn *et al.*, 2002). Ainda, considera-se sua característica de terem identidade por descendência, ou seja, devido ao fato de não haver mecanismo que retire a inserção *Alu* do genoma, dois indivíduos compartilhando uma inserção em um mesmo locus terá, provavelmente, a mesma origem evolutiva (Deininger e Batzer, 1999). Além disso, sua avaliação é um procedimento rápido e fácil por meio da reação em cadeia da polimerase (PCR) e posterior eletroforese (Roy-Engel *et al.*, 2001), apresentando uma taxa de nova inserção de uma a cada 20 nascimentos (Cordaux *et al.*, 2006).

Uma série de estudos foram realizados avaliando a presença de inserções *Alu* localizadas no MHC humano. Como resultado foram encontradas cinco

inserções margeando alguns genes, dentre as quais: *AluM1CB*, encontrado no primeiro íntron do gene *M1CB* (Kulski *et al.*, 2002); *AluTF*, localizado entre os genes *HLA-B* e *HLA-E*, a cerca de 505 kb centromérico em relação ao último (Dunn *et al.*, 2003); *AluHG* localizado a aproximadamente 20 kb da ponta 3' do gene *HLA-G* (Kulski *et al.*, 2001); *AluHJ*, localizado a aproximadamente 190 kb centromérico em relação a *HLA-G* (Dunn *et al.*, 2002); *AluHF*, localizado a aproximadamente 130 kb telomérico em relação a *HLA-G* (Dunn *et al.*, 2002) (Figura 4).

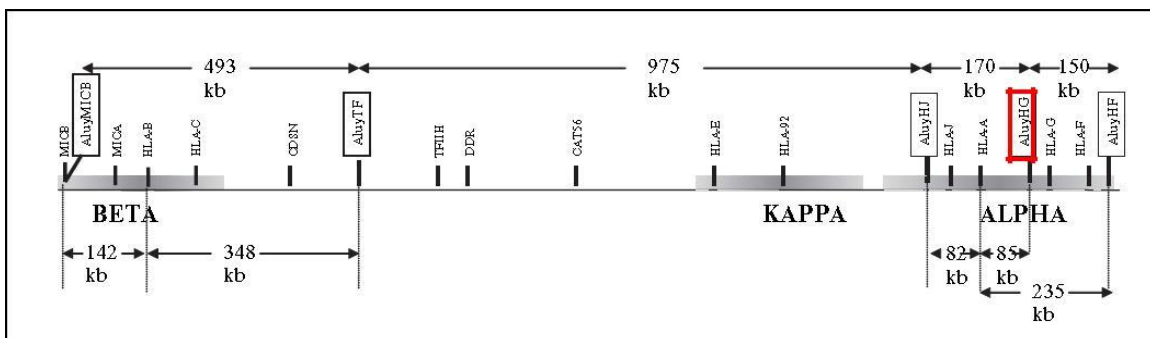


Figura 4. Mapa das posições das inserções *Alu* dentro do MHC, evidenciando a inserção *AluHG*. Adaptado de (Kulski e Dunn, 2005).

Dentre essas inserções a que se mostra mais próxima ao gene não clássico *HLA-G* é a inserção *AluHG*. Desta forma, este elemento torna-se bastante útil no estudo de sua relação com esse gene. Kulski e colaboradores em 2001 mostraram que esse polimorfismo apresentava desequilíbrio de ligação com o gene *HLA-A* e que apresentava frequências compatíveis com o grupo alélico *HLA-A*02* desse mesmo gene (Kulski *et al.*, 2001). Entretanto, apesar de sua proximidade ao gene *HLA-G*, nenhum estudo teve como foco a análise da relação entre a presença dessa inserção e polimorfismos desse gene.

1.4. O PROJETO 1000Genomes CONSORTIUM

O projeto *1000Genomes* (The 1000Genomes Project Consortium, 2010; The 1000Genomes Project Consortium, 2012) é um consórcio internacional que, utilizando técnicas de sequenciamento de nova geração, avaliou o genoma completo de 1092 indivíduos oriundos de 14 populações diferentes. A comparação dos dados obtidos por este projeto torna-se uma valiosa fonte para estudos de associação e identificação de genes candidatos para diferentes doenças (Harrow *et al.*, 2012), resposta individual e metabolização de fármacos (Allen, 2005) e padrões de diversidade gênica ao longo de diferentes populações. Neste aspecto, o sequenciamento em larga escala de diferentes indivíduos pode gerar dados mais completos e complementares aos dois consórcios iniciais que se propuseram a sequenciar o genoma humano de poucos indivíduos (International Human Genome Sequencing Consortium, 2001; Venter *et al.*, 2001; Venter, 2003; International Human Genome Sequencing Consortium, 2004).

Em nenhum dos projetos de sequenciamento completo do genoma humano a população brasileira foi avaliada. Neste aspecto, essa população torna-se uma excelente fonte de estudos, tendo em vista que constitui uma das populações mais heterogêneas do mundo, fruto de mais de cinco séculos de miscigenação entre as populações de quatro diferentes continentes: 1) Americano: população nativa de ameríndios; 2) Africano: escravos vindos para o Brasil entre o século 16 e o ano de 1850; 3) Europeu: principalmente portugueses que colonizaram o Brasil, seguidos por italianos, espanhóis e alemães (Pimenta *et al.*, 2006); 4) Asiático: japoneses que chegaram ao Brasil principalmente no período pós 2ª Guerra Mundial.

Assim, concatenar dados disponibilizados pelo Projeto *1000Genomes* com dados obtidos de populações brasileiras permite a observação de distribuição de frequências de alelos e haplótipos em torno do mundo.

2. OBJETIVOS

2.1. OBJETIVO GERAL

Analisar a frequência da presença da inserção *Alu*YG em amostras de três continentes (Brasileiros, Franceses e Africanos) e correlacionar este elemento com a variabilidade da região 3'NT do gene *HLA-G*.

2.2. OBJETIVOS ESPECÍFICOS

- Avaliar a presença da inserção *Alu*YG em uma amostra de brasileiros (estado de São Paulo), franceses e africanos;
- Avaliar a relação entre o elemento *Alu*YG e polimorfismos na região 3'NT do *HLA-G*;
- Inferir haplótipos *HLA-G/Alu*YG;
- Avaliar a distribuição mundial dos haplótipos de *HLA-G* considerando dados brasileiros e do projeto 1000Genomes;

3. JUSTIFICATIVA

O gene *HLA-G* tem grande importância na fisiologia do sistema imune humano, principalmente no contexto da gravidez, transplantes, tumores e doenças autoimunes (Donadi *et al.*, 2011). Entretanto, alguns tópicos relacionados a esse gene ainda permanecem obscuros, como perfil de expressão de todas as variantes proteicas, regulação da transcrição e tradução, bem como mecanismos evolutivos adjacentes à formação desse gene no curso da especiação humana. Assim, entender os mecanismos adjacentes à evolução desse gene pode permitir o entendimento da distribuição de poucos haplótipos em torno do mundo.

A identificação de sequências *Alu* próximas aos genes não clássicos e, em especial, uma localizada a 20kb de distância da região 3'NT do gene *HLA-G*, permite o estudo de sua relação com esse gene. Devido à proximidade dessas inserções e os genes de HLA, estudos realizados na última década têm buscado identificar perfis de desequilíbrio de ligação (LD, do inglês *Linkage Disequilibrium*) entre ambos..

Além disso, a proximidade entre o elemento *Alu*_{HG} e o gene *HLA-G* pode permitir a identificação de haplótipos associados à inserção *Alu*_{HG}, o que pode permitir sua fácil identificação para futuros estudos por meio de reações simples de PCR. A abordagem utilizada até o momento para a definição e triagem de haplótipos é o sequenciamento, técnica mais cara e mais trabalhosa em comparação à técnica de PCR. Portanto, em estudos em que é necessária apenas a identificação de alelos de alta e/ou baixa produção da molécula de *HLA-G*, a utilização de um marcador torna-se bastante útil.

4. MATERIAL E MÉTODOS

4.1. CARACTERIZAÇÃO DA AMOSTRA

Para a caracterização da presença da inserção *Alu*YHG, foram utilizadas 641 amostras originárias de três diferentes continentes: 165 brasileiros (BRA), 161 congolezes (CNG), 193 senegaleses (SEE) e 122 franceses (FRA).

A população brasileira foi composta por doadores de medula óssea, não relacionados e saudáveis oriundos do Hemocentro de Ribeirão Preto - SP, selecionados de forma aleatória. Estas amostras foram utilizadas em estudos prévios que avaliaram a variabilidade de outros genes do complexo HLA, incluindo os genes *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1* e *HLA-G*, realizados pelo grupo (Castelli *et al.*, 2010; Castelli *et al.*, 2011). Este protocolo experimental foi aprovado pelo comitê de Ética em pesquisa da Faculdade de Medicina de Ribeirão Preto - SP, Universidade de São Paulo (FMRP-USP) sob o protocolo 12398/2004, autorizando sua utilização para estudos envolvendo a variabilidade dos genes *HLA-G* e *HLA-E*, estando as amostras disponíveis para estudo.

A população congoleza foi composta por indivíduos saudáveis não relacionados oriundos da província de Badundu da República Democrática do Congo, gentilmente cedidas pelo Dr. Andre Garcia. O Ministério do Congo aprovou a utilização dessas amostras para estudos de variabilidade de genes do complexo HLA, estando essas amostras disponíveis para o estudo.

A população senegalesa foi composta por indivíduos saudáveis não relacionados oriundos da área de Niakhar, Senegal, também gentilmente cedidas pelo Dr. Andre Garcia. O protocolo experimental foi aprovado pelo comitê de Ética local para a utilização dessas amostras em estudos de variabilidade genética dos genes de HLA, estando essas amostras disponíveis para o presente estudo.

A população francesa foi composta por indivíduos saudáveis não relacionados oriundos de Paris, França, gentilmente cedidas pelo Dr. Philippe Moreau. O Governo Francês autorizou o uso dessas amostras para estudo de variabilidade genética dos genes de HLA, estando essas amostras também disponíveis para o estudo.

Todas as 641 amostras utilizadas já haviam sido avaliadas quanto à variabilidade da região 3'NT do gene *HLA-G* em estudos já publicados anteriormente (Castelli *et al.*, 2010; Castelli *et al.*, 2011; Courtin *et al.*, 2013; Garcia *et al.*, 2013; Martelli-Palomino *et al.*, 2013; Sabbagh *et al.*, 2013). Os dados de variabilidade foram gentilmente cedidos pelos pesquisadores responsáveis.

4.2. CARACTERIZAÇÃO DA INSERÇÃO ALUYHG

A inserção *AluyHG* foi caracterizada por procedimento de PCR (do inglês, *Polymerase Chain Reaction*) já descrito previamente (Kulski *et al.*, 2001). O DNA de cada amostra, já previamente extraído por kits comerciais ou métodos caseiros do tipo *salting-out*, foi amplificado por meio dos iniciadores AluyHGF-CAGGACAACCAGTAAAGATGCTGG e AluyHGR-GCTTCAGTTAACATGCAAGTTTATGCC (Kulski *et al.*, 2001).

A reação de PCR foi realizada em volume final de 25µL. Os reagentes utilizados para a reação de amplificação foram misturados em tubo separado conforme especificado na Tabela 2. Em cada tubo para PCR de 0,2mL, previamente identificado, foi adicionado 23,5µL da mistura de reagentes (Tabela 2) e mais 1,5µL de amostra de DNA diluído a 100ng/µL. Em cada bateria de amplificação foi adicionado um controle negativo apenas com a mistura de reagentes, portanto sem DNA, para a confirmação da ausência de contaminação dos reagentes por fragmentos de DNA que poderiam ser amplificados pelo conjunto de iniciadores utilizados.

As condições de ciclagem consistiram em: (a) desnaturação inicial do DNA a 94°C por 3 minutos; (b) 30 ciclos de 94°C de desnaturação por 30 segundos, 58°C de temperatura de anelamento por 30 segundos, 72°C de temperatura de extensão por 50 segundos; (c) um ciclo de extensão final a 72°C por 5 minutos. O produto obtido foi armazenado em freezer a -20°C até sua utilização como descrito adiante.

A amplificação correta foi, avaliada em gel de agarose a 1,5% corado com brometo de etídeo. Para cada amostra foi aplicado um volume de 6µL de produto de PCR junto com 1µL de tampão de carregamento 6x *Loading Dye*TM (Fermentas - Vilnius, Lituânia).

Tabela 2 - Reagentes e concentrações utilizadas para a amplificação do elemento *AluyHG*.

Componentes da reação	[] Solução de Uso	1X (µL)	[] Final
Água	-	16,05	-
Tampão de amplificação da DNA polimerase	10X	2,0	0,8X
MgCl₂	25mM	0,75	1,50mM
Solução de dNTPs	5mM	0,25	0,20mM
Iniciador AluyHGF	10pM	1,5	15pmol
Iniciador AluyHGR	10pM	1,5	15pmol
DMSO	100%	1,25	5% v/v
DNA Polimerase (Platinum – Invitrogen)	5U/µL	0,2	1U
Volume Total	-	23,5	-
Volume de Amostra (100 ng/µL)	-	1,5	-

A eletroforese foi realizada durante 1 hora mantendo-se a voltagem fixada em 90 V. Os géis foram expostos à luz ultravioleta (UV) e, posteriormente, fotografados em um fotodocumentador. Em cada corrida foi adicionada uma escada alélica *GeneRuler™* (Fermentas – Vilnius, Lituânia) com tamanho específico para conferência do tamanho aproximado dos produtos de amplificação gerados pela reação de PCR, bem como foi adicionado o controle negativo como um parâmetro de ausência de contaminação.

O tamanho esperado de cada fragmento de amplificação foi de 218pb para a ausência da inserção *AluyHG* e 540pb para a presença da inserção *AluyHG* (Figura 5), pois os iniciadores foram desenhados para amplificar regiões que flanqueiam o local de inserção dessa sequência em estudo.

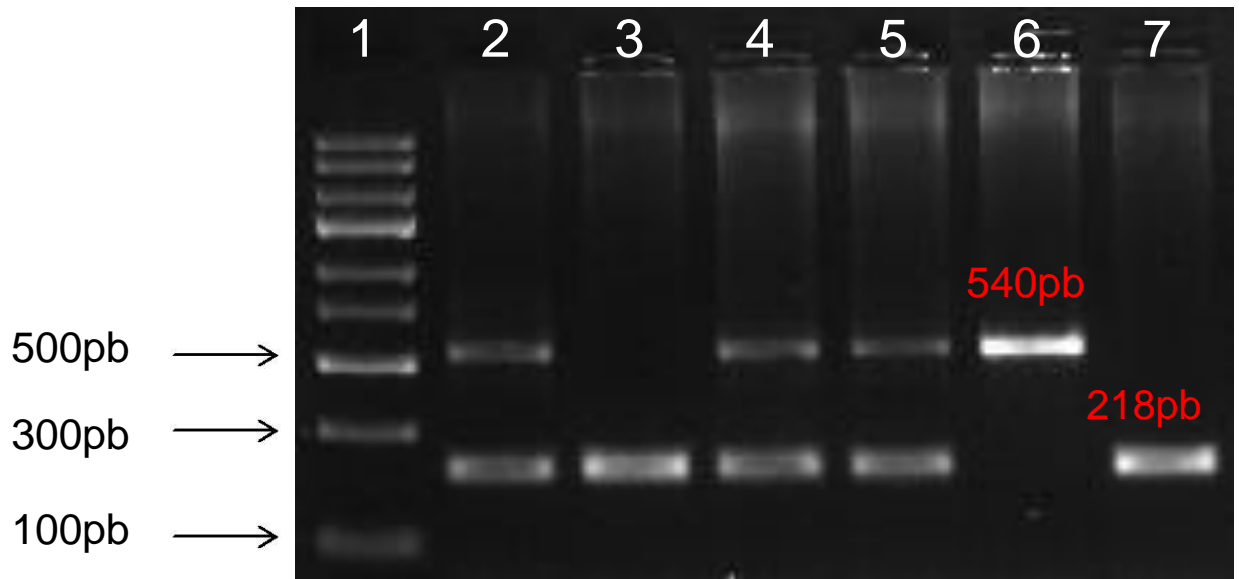


Figura 5 - Gel de agarose apresentando bandas específicas para a inserção *Alu*HG. A banda superior (540pb) reflete a presença da inserção, enquanto que a banda inferior (218pb) reflete a ausência de inserção. Corrida 1: marcador de peso molecular, corridas 2, 4 e 5: heterozigotos para o elemento *Alu*HG, corridas 3 e 7: homozigotos para ausência do elemento *Alu*HG, corrida 6: homozigoto para a presença do elementos *Alu*HG.

4.3. VARIABILIDADE DA REGIÃO CODIFICADORA E 3'NT DO GENE HLA-G

A variabilidade da região codificadora do gene *HLA-G* na amostra de brasileiros utilizada neste estudo foi previamente avaliada e publicada em um artigo do grupo de pesquisa (Castelli *et al.*, 2011). Esta avaliação considerou a região compreendida pelos éxons 1 a 4 e os íntrons 1, 2 e parte do íntron 3. Desta forma, os pontos de variação detectados nesse estudo ocorreram na posição +15, +36, +99, +126, +130, +147, +188, +292, +297, +372, +706, +748, +755, +814, +1016, +1019, +1054, +1590, +1799 e +1827, considerando +1 como a Adenina do primeiro ATG traduzido. As amostras francesas e africanas não foram avaliadas quanto à variabilidade da região codificadora do *HLA-G*.

A variabilidade da região 3'NT do *HLA-G*, por sua vez, foi previamente analisada para todos os grupos populacionais aqui estudados e os dados publicados (Castelli *et al.*, 2010; Castelli *et al.*, 2011; Courtin *et al.*, 2013; Garcia *et al.*, 2013; Martelli-Palomino *et al.*, 2013; Sabbagh *et al.*, 2013), seguindo uma metodologia descrita por nosso grupo (Castelli *et al.*, 2010).

Os dados de variabilidade das 14 populações disponibilizadas pelo projeto *1000Genomes* foram adquiridos diretamente no *browser* do projeto, considerando as mesmas regiões previamente analisadas para as amostras brasileiras, africanas e europeias.

4.4. NOMENCLATURA DOS HAPLÓTIPOS HLA-G

A nomenclatura dos haplótipos da região 3'NT do gene *HLA-G* foi dada de acordo com o descrito em estudos anteriores (Castelli *et al.*, 2010; Castelli *et al.*, 2011; Lucena-Silva *et al.*, 2012). A tabela 3 indica os haplótipos já descritos para a região 3'NT, o nome pelo qual este haplótipo é conhecido, e os polimorfismos associados com cada haplótipo.

A nomenclatura dos haplótipos de região codificadora, por sua vez, é regulamentada pelo International Immunogenetics Database (IMGT/HLA). Usando como exemplo o alelo HLA-G*01:01:01:01, este padrão de nomenclatura segue as seguintes regras:

- (a) anterior ao asterisco está o nome do gene (HLA-G, no caso);
- (b) o separador é sempre o asterisco (*);
- (c) o primeiro número indica o grupo alélico, dado pelo tipo sorológico do antígeno produzido pelo gene (neste caso, G*01);
- (d) o separador (:) separa cada nível da nomenclatura;
- (e) o segundo número indica uma proteína específica do gene por conta da presença de mutações não sinônimas;
- (f) o terceiro número indica substituições sinônimas na região codificadora (sem a produção de variantes proteicas);
- (g) e o quarto número indica mudanças de nucleotídeos nas regiões regulatórias do gene.

(h) um sufixo no final do nome indica mudanças na expressão dessa molécula, como, por exemplo, HLA-G*01:05N, em que o sufixo “N” indica que esse alelo é nulo.

Tabela 3 - Principais haplótipos da região 3’NT do gene *HLA-G*.

Haplótipo	Polimorfismos da região 3’NT							
	14pb	+3003	+3010	+3027	+3035	+3142	+3187	+3196
UTR-1	Del ^a	T	G	C	C	C	G	C
UTR-2	Ins ^b	T	C	C	C	G	A	G
UTR-3	Del	T	C	C	C	G	A	C
UTR-4	Del	C	G	C	C	C	A	C
UTR-5	Ins	T	C	C	T	G	A	C
UTR-6	Del	T	G	C	C	C	A	C
UTR-7	Ins	T	C	A	T	G	A	C
UTR-8	Ins	T	G	C	C	G	A	G
UTR-9	Ins	T	G	C	T	G	A	C
UTR-10	Del	T	C	C	C	G	A	G
UTR-11	Del	C	C	C	C	G	A	C
UTR-13	Del	T	C	C	T	G	A	C
UTR-14	Del	T	G	C	C	G	G	C
UTR-15	Ins	T	C	C	C	G	A	C
UTR-16	Ins	T	C	C	T	G	A	G

Os haplótipos foram nomeados de acordo com o descrito anteriormente em literatura (Castelli *et al.*, 2010; Castelli *et al.*, 2011; Lucena-Silva *et al.*, 2012).

^a Del - Deleção

^b Ins - Inserção

4.5. FORMATAÇÃO DOS BANCOS DE DADOS

Inicialmente, os dados de variabilidade do gene *HLA-G* foram obtidos no *browser* oficial do Projeto *1000Genomes* (<http://browser.1000genomes.org/index.html>). Os arquivos *.vcf* obtidos foram convertido para o formato *genepop* (Rousset, 2008) usando o programa *PGDSpider 2.0.19* (Lischer e Excoffier, 2012). Os dados obtidos foram analisados separadamente e posteriormente concatenados aos dados disponibilizados para o gene *HLA-G* das 641 amostras desse estudo, incluindo as avaliações de *Alu*YHG. Entretanto, dados referentes a polimorfismos do tipo INDEL geralmente não são

disponibilizados pelo projeto 1000Genomes, o que neste caso inclui a *Alu*YHG e o polimorfismo de 14pb da região 3'NT do gene *HLA-G*. No caso destes marcadores, os bancos de dados continham a informação de “ausência” do dado para todos os registros provenientes do projeto 1000Genomes.

Este banco de dados único (ou frações deste banco conforme a necessidade), foi inicialmente convertido em um arquivo .ped (Plink Ped format), para elaboração de um plot de LD pelo *software* Haploview 4.2 (Barrett *et al.*, 2005). Em seguida, os bancos de dados foram recodificados em um formato compatível com os *softwares* PHASE 2.1.1 (Stephens *et al.*, 2001; Stephens e Donnelly, 2003), com auxílio de macros criados no Microsoft Word 2010. Estes últimos bancos de dados foram utilizados para inferência de haplótipos conforme descrito adiante.

Finalmente, os dados de haplótipos foram convertidos em um formato de texto plano compatível com o *software* Network 4.6.1.1 (Bandelt *et al.*, 1999) para elaboração de uma rede de interação envolvendo os haplótipos inferidos.

4.6. ANÁLISE DOS DADOS

As frequências alélicas e genotípicas da presença ou ausência da inserção *Alu*YHG foram estimadas por contagem direta. A aderência das frequências genotípicas em relação às proporções teóricas de Hardy-Weinberg foi avaliada pelo teste exato de Guo e Thompson (Guo e Thompson, 1992), implementado no *software* Genepop® 4.2 (Raymond e Rousset, 1995).

Para a análise do desequilíbrio de ligação (LD) envolvendo os polimorfismos do gene *HLA-G* e a inserção *Alu*YHG foi utilizada a metodologia de D' e LOD, utilizando-se o *software* Haploview® (Barrett *et al.*, 2005). As imagens de LD foram geradas por esse programa utilizando-se os pontos de variação cuja frequência alélica mínima (MAF) fosse 0,01, em que as áreas em vermelho escuro indicam forte LD ($LOD \geq 2$, $D' = 1$), áreas em rosa indicam moderado LD ($LOD \geq 2$, $D' < 1$), áreas em azul indicam fraco LD ($LOD < 2$, $D' = 1$) e áreas em branco indicam ausência de LD ($LOD < 2$, $D' < 1$).

Foram utilizadas duas abordagens para a análise do LD entre as populações do estudo. A primeira teve como foco os dados referentes à inserção *Alu*YHG e os

polimorfismos presentes na região 3'NT do gene *HLA-G* buscando a identificação de haplótipos entre essas duas regiões distantes entre si cerca de 20kb. Assim, nessa primeira abordagem, foram utilizadas as populações brasileira, congoleza, senegalesa e francesa. A segunda abordagem, por sua vez, teve como foco o próprio gene *HLA-G* com a utilização dos dados das 14 populações do Projeto *1000Genomes* disponibilizados publicamente e a população brasileira. Foram avaliados, nesse contexto, polimorfismos presentes na região codificadora e na região 3'NT do referido gene.

Dada uma associação positiva entre os alelos dos pontos de variação encontrados, mas a fase gamética desconhecida, foi realizada uma inferência computacional dos haplótipos definindo-se a provável constituição de cada um dos cromossomos dos indivíduos analisados.

Para a inferência dos haplótipos, dois métodos probabilísticos foram empregados: o *software* PHASE v2 (Stephens *et al.*, 2001; Stephens e Donnelly, 2003) que implementa um método Bayesiano para reconstrução do haplótipo mais provável; e algoritmo de máxima verossimilhança implementado no *software* PL-EM (Qin *et al.*, 2002), que calcula pela maximização da expectativa para cada um dos haplótipos presentes nas amostras. Foram realizadas 10 corridas independentes para cada um dos métodos e os resultados foram então comparados entre si. Para este procedimento foi utilizado um script em Perl denominado HaploRunner (desenvolvido por E. C. Castelli – disponível em <http://www.castelli-lab.net>), versão 1.1b. Este script executou as 10 corridas independentes de cada algoritmo, comparando os resultados obtidos em todas as corridas e entre ambos os métodos. Foram aceitas somente as inferências de haplótipos que atenderam a dois requisitos: a) tiveram probabilidade de inferência superior a 90%; b) obtiveram o mesmo haplótipo inferido em todas as corridas para os dois programas. Para o método PHASE foram utilizados os seguintes parâmetros: *number of iterations*: 1000; *thinning interval*: 1; *burn-in value*: 1000 e valores *seed* diferentes para cada corrida, conforme descrito em um trabalho prévio (Castelli *et al.*, 2011). Para o algoritmo PL-EM foram utilizados os seguintes parâmetros: *Top value*: 0; *Parsize value*: 2; *Buffer*: 1300 e *Round value*: 200.

Apesar dos dados do Projeto *1000Genomes* estarem em fase, os dados brasileiros não estavam por se tratar de sequenciamento de produto de PCR. Desta forma, a mesma abordagem de inferência foi realizada para todas as 14 populações

agrupadas e os dados brasileiros. A compatibilidade entre as fases obtidas computacionalmente e as descritas pelo projeto 1000Genomes foi maior que 99%.

As relações entre os haplótipos obtidos foram inferidas construindo-se uma *network*, utilizando-se para tanto o algoritmo *median-joining* implementado no programa Network® 4.6.1.0 (<http://www.fluxus-engineering.com/sharenet.htm>). Foram utilizados os haplótipos com frequências superiores a 1% e excluídos os haplótipos recombinantes a fim de facilitar a leitura da Network. As imagens geradas pelo *software* foram reconstruídas utilizando-se programa de edição vetorial.

5. RESULTADOS

5.1. FREQUÊNCIAS ALÉLICAS E GENOTÍPICAS DA INSERÇÃO *ALUYHG*

Após a análise dos 641 indivíduos, compostos por 165 brasileiros (BRA), 161 congolezes (CNG), 193 senegaleses (SEE) e 122 franceses (FRA), foram contabilizadas as frequências do alelo de ausência de inserção *AluyHG* (*AluyHG**1) e de presença de inserção *AluyHG* (*AluyHG**2), como demonstrado na Tabela 4. Em todas essas populações as frequências genotípicas foram compatíveis com o esperado pelo teorema de Hardy-Weinberg. A Tabela 3 mostra, ainda, dados sobre o polimorfismo *AluyHG* em diferentes populações estudadas até agora. Para facilitar a compreensão, as populações foram divididas em relação ao seu componente de ancestralidade em cinco grandes grupos: (a) África, composto pelas populações *Sekele San*, *Bantu*, *!Kung San*, *Khoi*, senegaleses e congolezes; (b) Ásia, composto pelas populações tailandesa, japoneses, mongóis, malaio chineses, *Hunan* da etnia Han, mongóis da etnia Han, mongóis da etnia Mongol, *Guangdong* da etnia Han; (c) Oceania, composto por australianos de origem europeia; (d) Europa, composto pelos franceses; (e) América, composto por brasileiros.

As frequências do alelo *AluyHG**2 variam entre as populações estudadas. As populações de origem asiática, incluindo as populações malaio chinesa, *Hunan* da etnia Han, mongóis da etnia Han, mongóis da etnia Mongol e *Guangdong* da etnia Han apresentaram altas frequências desse alelo quando comparadas às outras populações estudadas até o momento. Em contrapartida, as populações de origem africana, incluindo as populações congoleza e senegalesa (alvos do presente estudo), *Bantu*, *Khoi*, *Sekele San* e *!Kung San* geralmente apresentam as frequências mais baixas quando comparadas às outras populações estudadas até agora. As populações brasileira e francesa (alvos do presente estudo) apresentaram frequências intermediárias às frequências de populações asiáticas e africanas para o alelo *AluyHG**2 (Tabela 4).

TABELA 4 - Frequências da ausência (*AluyHG*1*) e presença (*AluyHG*2*) da inserção *AluyHG* em populações mundiais.

Continente	População	N	<i>AluyHG*1</i>	<i>AluyHG*2</i>	Referência
África	<i>Sekele San</i>	60	0,9670	0,0330	(Kulski e Dunn, 2005)
	<i>Bantu</i>	50	0,9400	0,0600	(Kulski e Dunn, 2005)
	<i>!Kung San</i>	42	0,9270	0,0730	(Kulski e Dunn, 2005)
	<i>Khoi</i>	43	0,8690	0,1310	(Kulski e Dunn, 2005)
	<i>Senegaleses</i>	193	0,8964	0,1036	Este trabalho
	<i>Congoloses</i>	161	0,8944	0,1056	Este trabalho
Ásia	Tailandeses	192	0,7080	0,2920	(Dunn <i>et al.</i> , 2006)
	Japoneses	99	0,7300	0,2700	(Kulski e Dunn, 2005)
	Mongóis	41	0,7800	0,2200	(Kulski e Dunn, 2005)
	Malaios Chineses	50	0,4400	0,5600	(Dunn <i>et al.</i> , 2007)
	<i>Hunan</i> da etnia Han	147	0,5646	0,4354	(Tian <i>et al.</i> , 2008)
	Mongóis da etnia Han	104	0,6106	0,3894	(Tian <i>et al.</i> , 2008)
	Mongóis da etnia Mongol	87	0,6322	0,3678	(Tian <i>et al.</i> , 2008)
<i>Guangdong</i> da etnia Han	107	0,6121	0,3879	(Tian <i>et al.</i> , 2008)	
Oceania	Australianos	105	0,6990	0,3010	(Kulski e Dunn, 2005)
Europa	Franceses	122	0,7787	0,2213	Este trabalho
América	<i>Coorte Brasileiro 1</i> ^a	165	0,7333	0,2667	Este trabalho
	<i>Coorte Brasileiro 2</i> ^b	101	0,7570	0,2430	Silva ACA ^d
	<i>Coorte Brasileiro 3</i> ^c	61	0,7000	0,3000	Silva ACA ^d

Nota: N: número de indivíduos;

^aBrasileiros de Ribeirão Preto - SP;

^bBrasileiros de Brasília - DF;

^cPopulação Kalunga (População brasileira afrodescendente);

^dSilva ACA, et al. (comunicação pessoal).

5.2. DESEQUILÍBRIO DE LIGAÇÃO

A presença de uma associação significativa entre os pontos de variação do gene *HLA-G* e a inserção *AluyHG* foi avaliada mediante avaliação do desequilíbrio de ligação (D') (Lewontin, 1964), utilizando o *software* Haploview 4.1 (Barrett *et al.*, 2005), conforme previamente descrito. As figuras 6 e 7 mostram os padrões de LD obtidos utilizando-se os pontos de variação que apresentaram frequência alélica mínima de 1%.

Foram realizadas duas abordagens distintas para a mensuração do LD entre as regiões citadas acima. A primeira abordagem analisou o padrão de desequilíbrio

de ligação entre a inserção *Alu*YHG e os polimorfismos presentes na região 3'NT do gene *HLA-G* considerando os dados de genótipos de indivíduos de quatro populações distintas (brasileiros, senegaleses, congolezes e franceses). Essa análise revelou a provável existência de um bloco único de segregação que engloba a região 3'NT do *HLA-G* e estende-se por pelo menos 20kb de distância dessa região, alcançando o sítio de inserção do retroelemento *Alu*YHG (Figura 6).

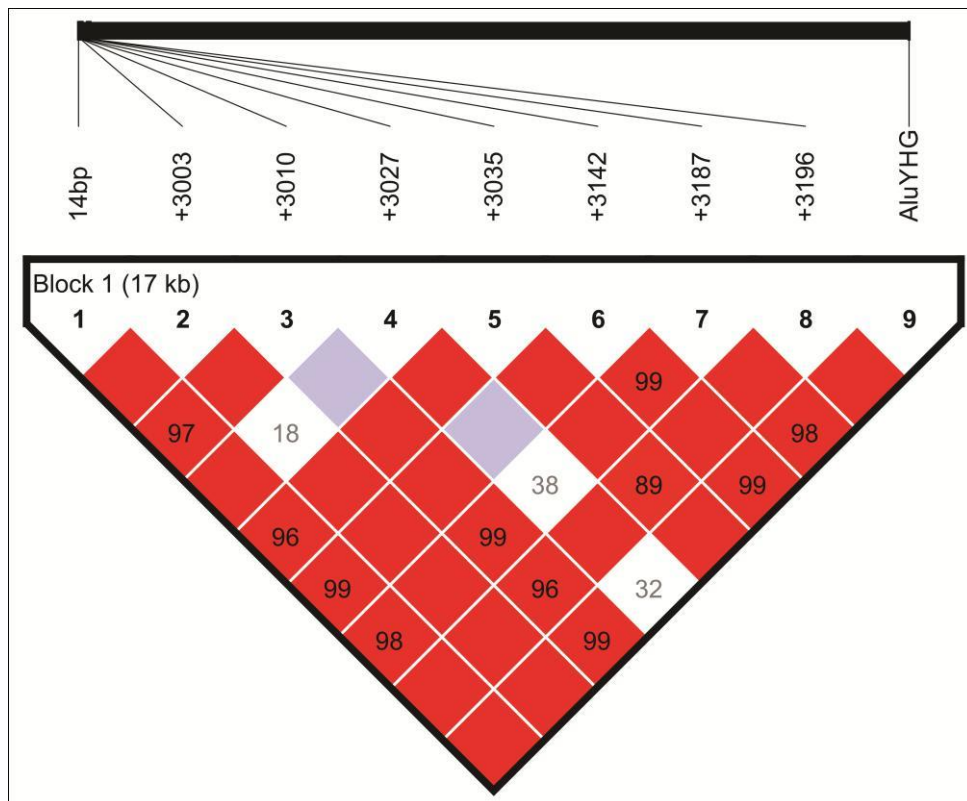


Figura 6 - Perfil de LD entre os polimorfismos da região 3'NT do gene *HLA-G* e a inserção *Alu*YHG considerando dados de genótipos de indivíduos de quatro populações distintas (brasileira, senegalesa, congoleza e francesa).

A segunda abordagem avaliou o padrão de LD entre os polimorfismos das regiões codificadora e 3'NT do gene *HLA-G*, considerando os dados de 15 populações distintas, i.e., a brasileira (Castelli *et al.*, 2011) e as 14 populações disponibilizadas pelo Projeto *1000Genomes*. Da mesma forma, nota-se a provável existência de um bloco de segregação único que engloba desde o éxon 1 até a região 3'NT do gene *HLA-G*, sem evidências de sítios de recombinação existentes entre essas duas regiões (Figura 7).

Extrapolando os dados obtidos (Figura 6 e 7), podemos concluir que o padrão de LD observado estende-se desde a região codificadora do gene *HLA-G* até pelo menos 20kb de distância, chegando ao sítio de inserção do retroelemento *Alu*YG.

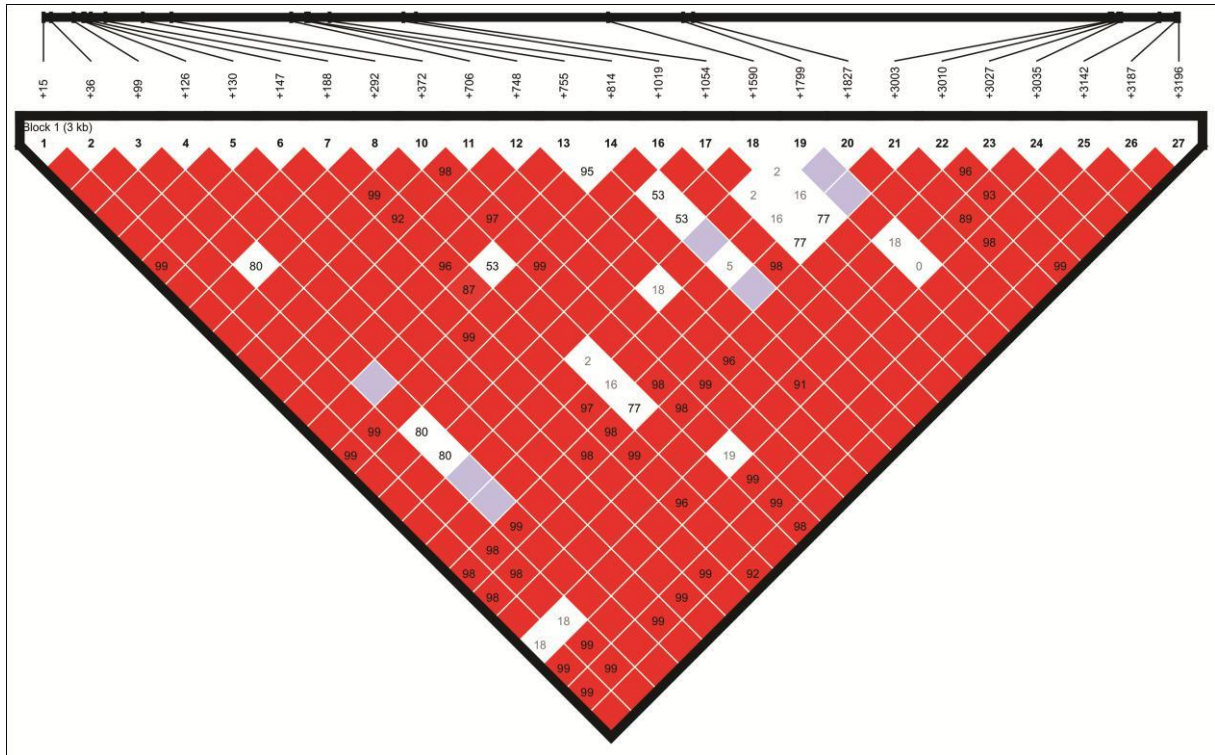


Figura 7 - Perfil de LD entre os polimorfismos da região codificadora e 3'NT do gene *HLA-G*, considerando dados de genótipos de indivíduos brasileiros e das 14 populações disponibilizadas pelo Projeto *1000Genomes*.

5.3. HAPLÓTIPOS *HLA-G/ALUYHG* E RELAÇÃO ENTRE OS HAPLÓTIPOS

Dada a associação positiva entre todos os polimorfismos do gene *HLA-G* e a inserção *Alu*YG (Figura 6), mas fase gamética desconhecida, haplótipos foram inferidos por métodos probabilísticos como citado na seção de materiais e métodos. A inferência de haplótipos foi realizada em 1733 indivíduos (1092 amostras do Projeto *1000Genomes* e 641 amostras do presente trabalho), divididos, assim como para a análise do LD, em duas abordagens.

Os dados de 641 indivíduos (brasileiros, franceses, senegaleses e congoleses) foram usados para a inferência de haplótipos considerando os polimorfismos da região 3'NT do *HLA-G* e do elemento *AluYHG*. Destes, 628 inferências passaram nos critérios de qualidade previamente descritos (aproximadamente 97,97%). A análise revelou a presença de 14 haplótipos distintos, considerando-se 8 pontos de variação da região 3'NT e a inserção *AluYHG*, entre as quatro populações. As frequências haplotípicas variaram entre 0,31% e 27,98% (Tabela 5). A probabilidade média de inferência de cada par de haplótipos foi de 0,9952 para o método PHASE e de 0,9990 para o método PL-EM. A Tabela 4 apresenta o conjunto de haplótipos encontrados bem como suas frequências para as quatro populações estudadas. O nome dos haplótipos referentes à região 3'NT do gene *HLA-G* foram dados de acordo com estudos já realizados (Castelli *et al.*, 2010; Lucena-Silva *et al.*, 2012), conforme tabela 3.

A população congoleza foi a que apresentou a maior quantidade de haplótipos distintos (11 no total) em relação às outras, enquanto que senegaleses apresentaram a menor quantidade de haplótipos (8 no total). A presença do elemento *AluYHG* está associada à presença do haplótipo de 3'NT denominado UTR-1. De fato, houve apenas uma ocorrência da presença do alelo *AluYHG*2* em ligação com outro haplótipo, denominado UTR-3, na população francesa. Assim, as frequências do alelo *AluYHG*2* e do haplótipo UTR-1 são compatíveis entre si, indicando uma possível utilização desse polimorfismo como um marcador molecular para o haplótipo UTR-1.

Além disso, o haplótipo UTR-1/*AluYHG*2* foi o mais frequente no Brasil e o segundo mais comum na França, tendo, em contrapartida, frequências mais baixas nos países de origem africana (Tabela 5). Apesar dessa associação, o haplótipo UTR-1 associado à ausência da inserção *AluYHG* foi encontrado em todas as populações analisadas, em baixas frequências, com as maiores frequências apresentadas pela população francesa. Ainda, não foram encontrados haplótipos UTR-7 nas duas populações africanas (Tabela 5).

Tabela 5 - Haplótipos entre polimorfismos da região 3'NT do gene *HLA-G* e a inserção *Alu*YG em quatro populações distintas: brasileiros, senegaleses, congolezes e franceses.

Haplótipos 3'NT ^b	Alelo <i>Alu</i> YG	Brasileiros ^a	Congolezes	Senegaleses	Franceses
		N ^c	152 ^d	161	193
		Frequências			
UTR-1	Presente	0,2500	0,1056	0,1036	0,2131
UTR-1	Ausente	0,0230	0,0342	0,0052	0,0533
UTR-2	Ausente	0,2500	0,2112	0,3575	0,2664
UTR-3	Presente	-	-	-	0,0082
UTR-3	Ausente	0,1250	0,3043	0,2798	0,1352
UTR-4	Ausente	0,1282	0,1056	0,0518	0,1434
UTR-5	Ausente	0,0757	0,0932	0,1321	0,0369
UTR-6	Ausente	0,1020	0,1304	0,0648	0,0861
UTR-7	Ausente	0,0428	-	-	0,0533
UTR-8	Ausente	0,0033	0,0031	-	-
UTR-15	Ausente	-	0,0124	-	0,0041
UTR-16	Ausente	-	-	0,0052	-
UTR-17 ^e	Ausente	-	0,0093	-	-
UTR-13	Ausente	-	0,0031	-	-

^a Brasileiros de Ribeirão Preto, São Paulo, Brasil.

^b Os haplótipos foram nomeados de acordo com (Castelli *et al.*, 2010; Lucena-Silva *et al.*, 2012).

^c Número de indivíduos.

^d O tamanho da amostra difere do original devido ao fato de os dados da região 3'NT do gene *HLA-G* não estarem disponíveis para 13 amostras.

^e Estes haplótipos de 3'NT não foram detectados por (Castelli *et al.*, 2010; Lucena-Silva *et al.*, 2012).

Em seguida, haplótipos foram inferidos a partir dos dados dos polimorfismos do gene *HLA-G* (regiões codificadora e 3'NT) para as 14 populações do Projeto *1000Genomes* e concatenados aos dados já publicados da população brasileira (Castelli *et al.*, 2011). A inferência foi realizada totalizando 1092 pares de haplótipos. A comparação das fases obtidas computacionalmente com as fases definida pelo sequenciamento de nova geração resultou nos mesmos haplótipos em 99% das amostras. A Tabela 6 apresenta um total de 32 haplótipos entre as regiões codificadora e 3'NT encontrados considerando-se as 14 populações estudadas, com

frequências variando entre 0,5% e 42,5%. A probabilidade média de inferência para o método PHASE foi de 0,9738, enquanto que para o método PL-EM essa probabilidade foi de 0,9778.

O perfil de distribuição de haplótipos do *HLA-G* (regiões codificadora e 3'NT) em torno do mundo foi o mesmo descrito no Brasil (Castelli *et al.*, 2010; Castelli *et al.*, 2011) e em outros estudos (Hviid *et al.*, 2004; Hviid, 2006; Larsen e Hviid, 2009; Jassem *et al.*, 2012; Martinez-Laso *et al.*, 2013). O haplótipo UTR-1 foi encontrado em todas as populações analisadas, principalmente associado ao haplótipo de região codificadora denominado G*01:01:01:01, e em haplótipos recombinantes em que a porção final da sequência da região codificadora apresenta a mesma sequência de polimorfismos do haplótipo G*01:01:01:01. Poucos haplótipos recombinantes, tal como G*01:01:01:04/UTR-1, foram encontrados, principalmente na população africana (Tabela 6). No Brasil, todos os haplótipos UTR-1 estavam associados ao haplótipo codificador G*01:01:01:01. As frequências desse haplótipo variaram de 12,18% para a população Yorubá (africana) a 42,50% para a população *Han do sul da China* (asiática).

Dado o fato de que o mesmo perfil de haplótipos da região codificadora/3'NT encontrado no Brasil também foi encontrado para outras populações mundiais, e que o perfil de LD mostra evidência de associação entre a região 3'NT e o sítio de inserção da sequência *Alu*YG, o mesmo padrão de associação entre a região 3'NT/*Alu*YG encontrado para as populações brasileira, senegalesa, congoleza e francesa foi extrapolado para a população mundial. Soma-se a isso o fato de as frequências do haplótipo G*01:01:01:01/UTR-1 serem compatíveis à frequência da inserção *Alu*YG. Assim, podemos estimar a presença de um haplótipo estendido G*01:01:01:01/UTR-1/*Alu*YG.

Nota-se também a ausência do haplótipo UTR-6 (e de seus haplótipos codificadores associados) nas populações de origem asiática, bem como a ausência do haplótipo UTR-7 (e de seus codificadores associados) nas populações de origem africana. Além disso, as populações africanas são as que apresentam as maiores frequências para o haplótipo UTR-3 e seus codificadores associados (Tabela 6).

Tabela 6 - Frequências dos haplótipos considerando polimorfismos da região codificadora e 3'NT do gene *HLA-G* na população mundial.

População		Europa				Ásia				África		América					
		89	93	14	98	100	97	89	88	97	55	60	66	85	61	108	
Haplótipos 3'NT ^b	Haplótipos Codificadora ^c	N	89	93	14	98	100	97	89	88	97	55	60	66	85	61	108
UTR-1	G*01:01:01:01		0,3371	0,3602	0,3214	0,2908	0,4250	0,2947	0,2619	0,1218	0,2391	0,2778	0,2373	0,2727	0,3941	0,2182	0,2360
	G*01:01:01:04		-	-	-	-	-	-	-	0,0256	0,0109	0,0093	-	0,0076	-	0,0091	-
	G*01:01:03 / G*01:01:01:01 ^d		-	-	-	-	0,0050	-	-	-	-	-	-	-	-	-	-
UTR-2	G*01:01:02:01		0,2303	0,1774	0,3571	0,2041	0,0400	0,1105	0,1607	0,1410	0,1848	0,1111	0,1780	0,2127	0,2059	0,2091	0,1530
	G*01:01:02:02		-	-	-	0,0051	-	-	-	0,0064	0,0380	-	-	-	-	-	-
	G*01:06		0,0618	0,0269	0,0357	0,0663	0,0100	0,0263	0,0060	-	0,0054	0,0278	0,0424	0,0227	0,0471	0,0091	0,0510
	G*01:05N		-	0,0108	-	0,0255	0,0150	0,0421	0,0060	0,0962	0,0543	-	0,0085	0,0303	0,0059	0,0636	0,0420
	G*01:05N (+188 C) ^e		-	-	-	0,0153	-	0,0053	-	0,0128	0,0217	-	0,0085	0,0152	-	0,0182	-
	G*01:01:03:01 / G*01:01:02:01 ^d		0,0056	-	-	-	-	-	0,0060	-	-	-	-	-	-	0,0182	-
UTR-3	G*01:01:14		-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0050
	G*01:04:01		0,0449	0,0591	0,0714	0,0969	0,2550	0,2474	0,4524	0,0256	0,0109	0,1296	0,1780	0,1515	0,0647	0,0364	0,0460
	G*01:04:03		-	-	-	-	0,0050	0,0053	0,0179	-	-	-	-	-	-	-	0,0050
	G*01:04:04		0,0112	0,0054	-	0,0306	-	-	-	0,2244	0,0978	0,0556	0,0339	0,0076	0,0235	0,1000	0,0320
	G*01:04:05		-	-	-	-	-	-	-	0,0064	0,0109	0,0093	0,0169	-	-	0,0091	-
	G*01:01:08		-	-	-	-	-	0,0053	0,0119	-	-	-	-	-	-	-	-
	G*01:04:04 (+188 C) ^e		-	-	-	-	-	-	-	0,0064	-	-	-	-	-	-	-
G*01:01:03 / G*01:04:01 ^d		-	-	-	-	0,0050	0,0158	0,0119	-	-	-	-	-	-	-	-	
UTR-4	G*01:01:01:05		0,1180	0,2849	0,1071	0,1429	0,0200	0,0474	0,0060	0,1154	0,0761	0,1296	0,1356	0,0758	0,1529	0,0545	0,0970
	G*01:01:09		-	-	-	-	-	-	-	0,0321	0,0272	-	0,0085	0,0076	-	0,0091	0,0050

UTR-5	G*01:03	0,0281	0,0161	-	0,0306	-	0,0263	0,0179	0,0897	0,0870	0,1111	0,0763	0,0985	0,0353	0,1364	0,0880
	G*01:04:01	0,0056	-	-	0,0051	-	-	-	-	-	0,0093	-	-	-	-	-
	G*01:01:08	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0090
UTR-6	G*01:01:01:04	0,0618	0,0108	0,0714	0,0204	-	-	-	0,0833	0,1359	0,0833	0,0678	0,0379	0,0118	0,1000	0,0740
	G*01:01:01:01	-	-	-	0,0051	-	-	-	-	-	-	-	-	0,0059	-	0,0090
	G*01:01:01:05	-	-	-	0,0051	-	-	-	0,0128	-	-	-	-	-	-	0,0149
	G*01:01:01:04 (+1019 C) ^e	-	-	-	0,0051	-	-	-	-	-	-	-	-	-	-	-
UTR-7	G*01:01:03:01	0,0899	0,0484	0,0357	0,0459	0,2200	0,1684	0,0471	-	-	0,0463	0,0085	0,0455	0,0471	0,0091	0,0370
	G*01:01:05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0050
UTR-8	G*01:06	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0140
UTR-9	G*01:01:08	0,0056	-	-	0,0051	-	0,0053	-	-	-	-	-	-	-	-	-
UTR-18 ^f	G*01:01:02:01	-	-	-	-	-	-	-	-	-	-	-	0,0076	-	-	-
UTR-19 ^f	G*01:01:03:01	-	-	-	-	-	-	-	-	-	-	-	-	0,0059	-	-

N - número de indivíduos

^a Brasileiros de Ribeirão Preto, São Paulo. A variabilidade do gene *HLA-G* foi publicada previamente (Castelli *et al.*, 2011).

^b Os haplótipos da região 3'NT do gene *HLA-G* foram nomeados de acordo com (Castelli *et al.*, 2010; Lucena-Silva *et al.*, 2012).

^c Os haplótipos da região codificadora foram nomeados de acordo com o padronizado pelo IMGT/HLA (<http://www.ebi.ac.uk/imgt/hla/>).

^d Haplótipos recombinantes.

^e Haplótipo ancestral mais provável e a mutação que define esse novo haplótipo.

^f Esses haplótipos não foram detectados previamente (Castelli *et al.*, 2010; Lucena-Silva *et al.*, 2012).

Para avaliar a similaridade entre os haplótipos encontrados por meio das duas abordagens citadas, duas redes foram construídas (Figuras 8 e 9). A partir da análise da primeira rede (Figura 8), construída a partir dos dados de polimorfismos da região 3'NT do gene *HLA-G* e da inserção *Alu*yHG, pudemos definir 5 linhagens de acordo com o descrito na literatura (Castelli *et al.*, 2011). Notamos que a linhagem que carrega o haplótipo *Alu*yHG*2 aparece na extremidade da rede e em alta frequência. Apesar disso, esse haplótipo pode ser um dos mais recentes na história evolutiva humana.

Por meio da análise da segunda rede (Figura 9), construída a partir dos polimorfismos das regiões codificadora e 3'NT do gene *HLA-G*, as 5 principais linhagens já anteriormente também estão representadas. Novamente, a rede mostra o haplótipo G*01:01:01:01/UTR-1 na extremidade e em alta frequência. Considerando a associação entre *Alu*yHG e o haplótipos UTR-1, é provável que o haplótipos estendido G*01:01:01:01/UTR1/*Alu*yHG seja um dos mais recentes no curso da história evolutiva humana, apesar de sua elevada frequência.

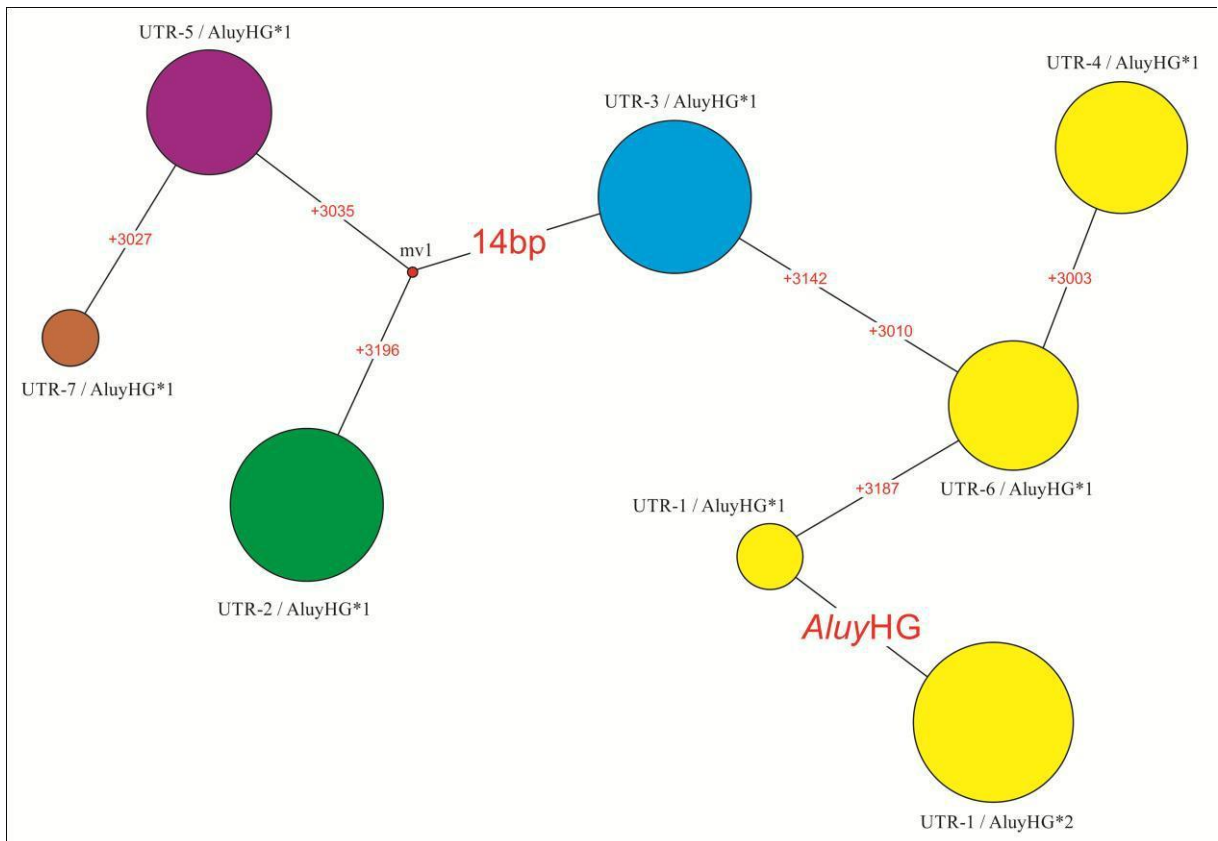


Figura 8 - Rede de haplótipos construída a partir dos dados de polimorfismos da região 3'NT do gene *HLA-G* e a inserção *Alu*yHG considerando os dados das

populações brasileira, senegalesa, congoleza e francesa. As cores diferentes indicam linhagens diferentes desses haplótipos.

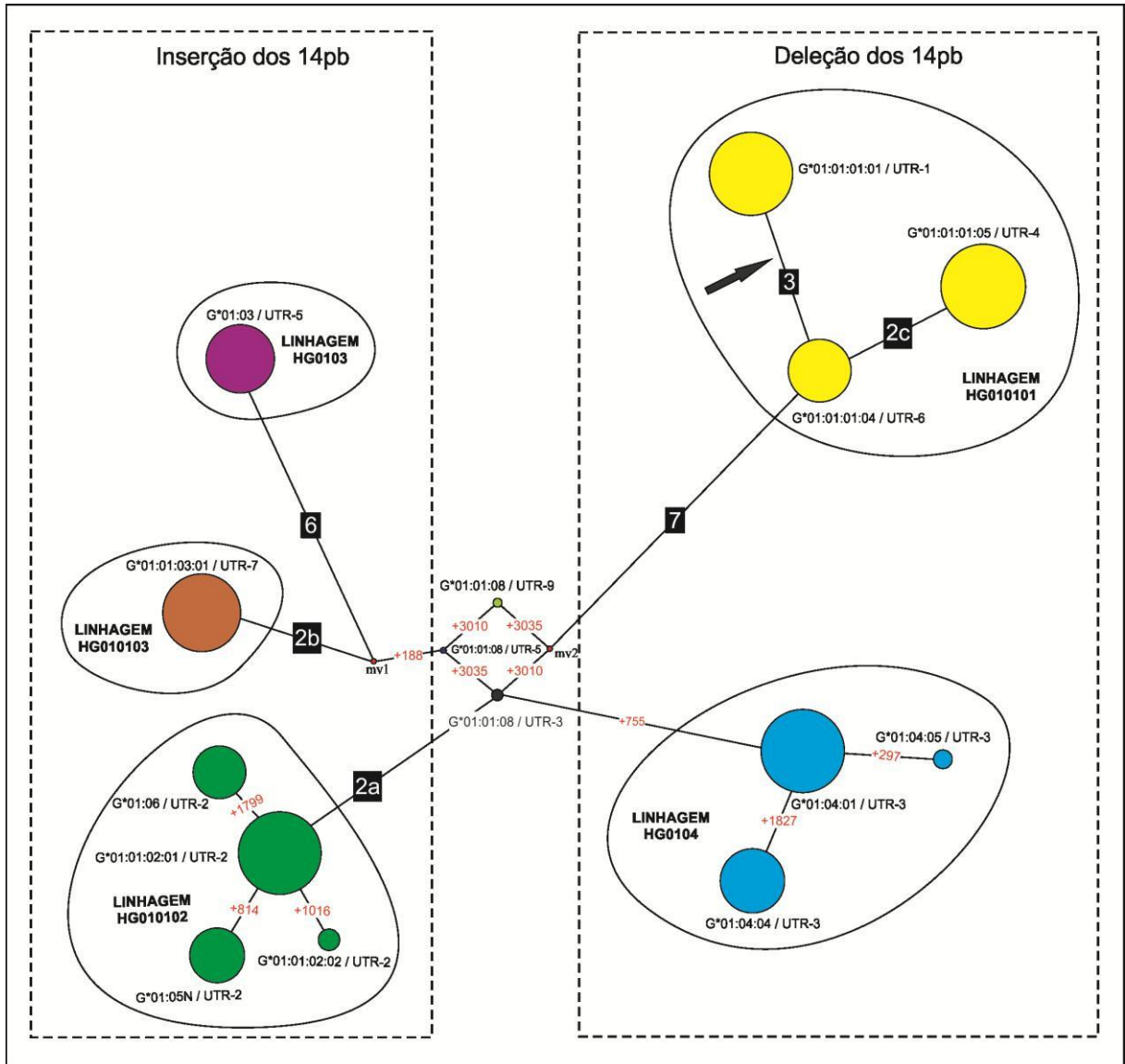


Figura 9 - Rede de haplótipos construída a partir dos dados de polimorfismos das regiões codificadora e 3'NT do gene *HLA-G* a partir dos dados disponibilizados pelo Projeto *1000Genomes*. As linhagens têm as mesmas cores da Figura 7 e foram nomeadas de acordo com (Castelli *et al.*, 2011).

6. DISCUSSÃO

O presente trabalho consiste no primeiro a avaliar a relação entre a inserção *Alu*YG e polimorfismos presentes no gene *HLA-G*. Foram caracterizadas 641 amostras de quatro populações distintas, incluindo uma amostra da população brasileira do estado de São Paulo, e amostras do Congo, Senegal e França, quanto à presença ou ausência do retroelemento *Alu*YG e sua relação com sítios polimórficos encontrados nas regiões codificadora e 3'NT do gene *HLA-G*. Os resultados obtidos foram comparados com dados disponibilizados pelo Projeto *1000Genomes*, permitindo uma avaliação global da distribuição dos haplótipos de *HLA-G* no mundo.

Estudos anteriores relataram a presença da inserção *Alu*YG em altas frequências nas populações asiáticas (Tabela 4), principalmente entre chineses (Dunn *et al.*, 2007; Tian *et al.*, 2008). No Brasil, as frequências do alelo de inserção, *Alu*YG*2, foram similares as frequências já descritas para populações europeias, japonesas e tailandesas (Tabela 4). Esta similaridade pode ser explicada pela história de formação da população brasileira do estado de São Paulo, que apresenta como principal componente de ancestralidade as populações de origem europeia (Ferreira *et al.*, 2006; Muniz *et al.*, 2008). Além disso, as frequências do alelo *Alu*YG*2 (ausência do elemento *Alu*YG) foram similares às encontradas para a população de Brasília - DF (Capital Federal, distante aproximadamente 706 km de Ribeirão Preto - SP) e também para a população Kalunga (população afrodescendente habitante do Estado de Goiás - GO) (Silva ACA, comunicação pessoal). A população Kalunga, considerada afrodescendente, apresentou frequência do alelo *Alu*YG*2 mais elevada do que a encontrada para a população de Ribeirão Preto - SP (Tabela 4). Em geral, as populações africanas apresentaram as menores frequências deste elemento. Acredita-se que este fato deva-se à miscigenação que a população Kalunga sofreu ao longo do seu processo de formação.

Por outro lado, as frequências do alelo *Alu*YG*2 na população francesa apresentam-se intermediárias às encontradas para os grupos brasileiro e africanos. Entretanto, apesar de serem baixas as frequências de inserção desse retroelemento nas populações africanas, o fato de podermos encontrar cromossomos

apresentando tal inserção nesse continente indica que provavelmente o evento de inserção ocorreu antes da dispersão do homem para outros continentes. Tal hipótese é baseada no fato de que o evento de transposição do elemento *Alu*HG é considerado um processo aleatório, ou seja, não há predileção de inserção em uma sequência específica de um cromossomo (Jurka, 1997). Assim, é pouco provável que uma mesma sequência de transposição seja inserida exatamente no mesmo loco por dois eventos diversos de inserção que ocorreram também em tempos diversos (Kulski e Dunn, 2005). Além disso, não há mecanismo conhecido responsável por retirar a sequência *Alu* após sua inserção em um determinado loco. Portanto, podemos inferir que aqueles indivíduos que apresentam a mesma sequência *Alu* no mesmo loco apresentam um ancestral comum, devido à sua característica de identidade por descendência (Kulski e Dunn, 2005), e que o processo de transposição ocorreu ainda na África antes da dispersão do homem a outros continentes. Essas baixas frequências no continente africano indicam que a inserção do elemento *Alu*HG é recente na história humana, mas antiga o suficiente para que fosse difundido pelos outros continentes durante a dispersão do homem.

Entretanto, vale ressaltar que o processo de transposição pode ser dirigido por sequências específicas apresentadas em certos cromossomos (Jurka, 1997). Além disso, a frequência de determinados haplótipos presentes à época do evento inicial de transposição também pode facilitar a sua inserção em determinados haplótipos, notadamente os mais frequentes, sem, entretanto, desobedecermos à premissa de que a inserção é um evento aleatório. Assim, apesar do possível direcionamento do evento de transposição para haplótipos mais frequentes, continua pouco provável que dois eventos distintos insiram uma mesma sequência de transposição num mesmo loco.

As frequências do alelo *Alu*HG*2 apontam para o fato de que o possível efeito fundador sofrido durante a formação das populações humanas a partir da sua saída do continente africano (Henn *et al.*, 2012) pode ser responsável pelo aumento dessas frequências nas populações consideradas não africanas (Tabela 4). De fato, quando comparamos as frequências do elemento *Alu*HG com o evento de dispersão do homem, bem como suas rotas migratórias (Henn *et al.*, 2012), observamos um aumento gradual das frequências com o aumento do tempo e da distância da dispersão do homem à partir da África para os outros continentes. Assim, as frequências aumentadas do alelo *Alu*HG*2 em continentes não africanos

pode ser consequência do isolamento pela distância dessas populações em relação às populações africanas, além da possível pressão seletiva atuando sobre essa região. As frequências das quatro populações avaliadas pelo presente estudo corroboram tal hipótese, indicando a possibilidade de sucessivos efeitos fundadores atuando sobre esse loco, pois notamos frequências baixas do alelo de inserção desse elemento no continente africano, frequências intermediárias na população europeia e frequências ligeiramente mais altas na população brasileira (Tabela 4).

Frequências aumentadas desse retroelemento apresentadas por populações de origem asiática, notadamente as chinesas, podem representar um reflexo cultural dessas populações, incluindo maior incidência de endogamia, principalmente em se tratando de grupos étnicos fechados dentro dessas populações (Chen *et al.*, 2007), somada à possível pressão seletiva diferenciada apresentada nessa região do mundo.

Embora a teoria da evolução neutra possa explicar a distribuição do retroelemento *Alu*YG, devido à sua localização ser intergênica, a região cromossômica adjacente é um dos principais alvos de seleção do genoma humano (Solberg *et al.*, 2008). Tal característica ocorre devido ao fato de essa região apresentar genes relacionados à função do sistema imune, sendo moldada por exposição a patógenos específicos (Klein e Sato, 2000a; 2000b). De fato, há evidências de sinais de seleção balanceadora atuando sobre regiões regulatórias do gene *HLA-G* (Tan *et al.*, 2005; Castelli *et al.*, 2011), notadamente em sua região 3'NT (Castelli *et al.*, 2011), distante apenas 20kb do sítio de inserção *Alu*YG. Além disso, a assinatura de seleção balanceadora que atua sobre essas regiões do gene *HLA-G* pode ser devida à seleção balanceadora atuando em outros genes da mesma região cromossômica (Gaudieri *et al.*, 2000), principalmente sobre o gene *HLA-A* (responsável por apresentação antigênica e um dos mais variáveis do genoma humano), localizado após a inserção *Alu*YG. Assim, como a inserção *Alu*YG encontra-se entre esses dois importantes genes do sistema imune e já foi demonstrado forte LD entre essa inserção e grupos alélicos específicos de *HLA-A*, notadamente o grupo *HLA-A*02* (Kulski *et al.*, 2001), acreditamos que essa região, apesar de intergênica, pode sofrer a influência da seleção natural atuando em genes localizados em sua periferia.

Considerando sua proximidade ao gene *HLA-G* (cerca de 20kb) e sua relação já demonstrada com certos grupos alélicos do gene *HLA-A* (Kulski *et al.*, 2001),

avaliamos o perfil de desequilíbrio de ligação entre a inserção *Alu*YG e os polimorfismos presentes na região 3'NT do gene *HLA-G* (Figura 6), que apresenta grande importância na regulação pós-transcricional desse gene e sinais de seleção balanceadora mantendo alta heterozigosidade (Castelli *et al.*, 2011; Martinez-Laso *et al.*, 2013). Devido ao fato de o evento de transposição da inserção *Alu*YG ser um evento antigo, provavelmente ocorrido antes da dispersão do homem à partir da África, e também considerando sua distância em relação ao gene *HLA-G*, era esperado que eventos de recombinação gerassem diferentes haplótipos de 3'NT associados com a inserção do *Alu*YG. Entretanto, observamos que a inserção *Alu*YG somente foi associada a apenas um haplótipo de 3'NT, denominado UTR-1 (Tabela 4).

O haplótipo UTR-1 foi recentemente considerado como um haplótipo associado com alta produção da molécula de HLA-G (Martelli-Palomino *et al.*, 2013) e apresenta altas frequências nas populações mundiais (Castelli *et al.*, 2011; Donadi *et al.*, 2011). Este haplótipo possui um polimorfismo do tipo *indel* caracterizado pela deleção de 14pb na região 3'NT, previamente associada com uma maior produção de HLA-G (Hviid *et al.*, 2003; Rousseau *et al.*, 2003; Castelli *et al.*, 2011; Donadi *et al.*, 2011; Martelli-Palomino *et al.*, 2013), além de outros polimorfismos que aumentam a estabilidade do mRNA e diminuem sua afinidade por microRNAs (Martelli-Palomino *et al.*, 2013). As frequências do haplótipo UTR-1 foram bastante similares às frequências encontradas para a inserção *Alu*YG (Tabelas 4 e 5), indicando que essa inserção pode ser considerada um marcador para esse haplótipo. Novamente, observamos que as frequências do haplótipo UTR-1/*Alu*YG*2 aumentam periodicamente em relação ao aumento da distância e do tempo de dispersão do homem a partir da África, como já descrito. Entretanto, a hipótese de sinais de seleção atuando sobre a região de inserção faz-se mais plausível sobre o olhar de sua relação com o gene *HLA-G*. Assim, o aumento de frequências tanto do haplótipo UTR-1 e da inserção *Alu*YG pode ser reflexo dessas duas forças atuando sobre essas regiões.

Apesar da forte associação encontrada entre o elemento *Alu*YG e o haplótipo UTR-1, observamos a presença desse haplótipo associado com a ausência da inserção *Alu*YG em todas as populações avaliadas, em baixas frequências. Teoricamente, esse haplótipo (UTR-1/*Alu*YG*1) pode ser considerado como o haplótipo ancestral, mostrando que o evento de inserção, apesar de

bastante antigo como já demonstrado, ainda não conseguiu fixar a inserção ao haplótipo de UTR-1. Entretanto, possíveis eventos de recombinação entre haplótipos UTR-1/*Alu*YG*2 e qualquer outro haplótipo que não apresenta a inserção não pode ser descartada. Tal fato pode ser comprovado pela presença de um cromossomo apresentando o haplótipo UTR-3/*Alu*YG*2 na França (Tabela 5). Além disso, era esperado que o haplótipo UTR-1/*Alu*YG*1 fosse encontrado em frequências elevadas no continente africano, considerado o continente ancestral. Entretanto, era esperada a presença de haplótipos diferentes acompanhados pela inserção *Alu*YG em todas as populações analisadas, o que torna pouco provável que os haplótipos UTR-1/*Alu*YG*1 sejam recombinantes. Desta forma, é provável que o haplótipo UTR-1/*Alu*YG*1 seja o haplótipo ancestral da inserção *Alu*YG.

A distribuição de haplótipos de 3'NT do *HLA-G* em torno do mundo é praticamente a mesma já definida em estudos prévios para a população brasileira (Castelli *et al.*, 2011; Lucena-Silva *et al.*, 2012), com exceção ao haplótipo UTR-7. O haplótipo UTR-7 não foi encontrado no continente africano, indicando que sua formação pode ser um evento recente, após a dispersão do homem a partir da África e que o fluxo gênico ainda não a introduziu nesse continente ou sua frequência é muita baixa para ser encontrado considerando a amostra analisada. Entretanto, mais estudos são necessários para se desvendar o verdadeiro significado da sua ausência nesse continente, já que a seleção direcionada por patógenos também pode moldar suas frequências em torno do mundo.

O haplótipo UTR-1, de maneira geral, ocorre em cromossomos também portando o haplótipo de região codificadora G*01:01:01:01 (Castelli *et al.*, 2011). Considerando que o mesmo perfil de associação entre haplótipos foi observado no mundo todo, bem como a presença de blocos únicos de segregação englobando o gene *HLA-G* e o elementos *Alu*YG, é provável que o haplótipo G*01:01:01:01/UTR-1/*Alu*YG também ocorra no mundo todo acompanhando as frequências observadas para o haplótipo G*01:01:01:01/UTR-1. Curiosamente, o haplótipo G*01:01:01:01/UTR-1 tem baixas frequências no continente africano e frequências mais altas nas outras partes do mundo, seguindo o mesmo modelo já descrito tanto para a inserção *Alu*YG como para o haplótipo de UTR-1 (Tabela 6).

O haplótipo G*01:01:01:01/UTR-1 foi recentemente associado a alta produção da molécula de *HLA-G* por meio da combinação de características, que incluem: (a) a transcrição de um RNA mensageiro mais estável (Yie *et al.*, 2008); (b) baixa

afinidade por microRNAs, responsáveis por mediar a degradação do RNA mensageiro por meio de um complexo enzimático (Tan *et al.*, 2007; Castelli *et al.*, 2009; Manaster *et al.*, 2012); (c) uma região regulatória 5' única, inclusive apresentando fortes sítios promotores para o hormônio progesterona (Hviid *et al.*, 1999; Solier *et al.*, 2001; Tan *et al.*, 2005; Castelli *et al.*, 2011; Martinez-Laso *et al.*, 2013). Desta forma, é razoável propor que, devido às características de alta produção da molécula de HLA-G apresentada pelo haplótipo G*01:01:01:01/UTR-1, as frequências aumentadas desse haplótipo e, em consequência, da inserção *Alu*YG, nas populações modernas, pode ser devido ao efeito da seleção natural atuando sobre essa região.

A presença do haplótipo UTR-1 geralmente é acompanhada de outro haplótipo de baixa produção da molécula de HLA-G, notadamente o UTR-2. Esse fato está em acordo com as evidências de seleção balanceadora já observadas para a região 3'NT desse gene, ou seja, favorecendo a heterozigose (Castelli *et al.*, 2011). Fisiologicamente, essa dualidade passa a ser favorável ao indivíduo, que estaria melhor adaptado para responder a diferentes situações adequadamente, como durante a gestação ou doenças autoimunes (necessidade de inibição do sistema imunitário) ou infecções crônicas (necessidade de uma resposta imunológica mais ativa). Assim, teoricamente, indivíduos que apresentam alelos para alta e baixa produção da molécula de HLA-G, considerada como de modulação do sistema imune, têm um perfil de atuação das células do sistema imune mais preciso (Donadi *et al.*, 2011).

De maneira global, observamos que as frequências da inserção *Alu*YG foram compatíveis com as frequências de UTR-1 nas populações analisadas nesse trabalho, provavelmente devido ao efeito carona atuando entre essas duas regiões. Estendendo nossa análise para todo o gene *HLA-G*, em que notamos forte desequilíbrio de ligação entre os pontos de variação pertencentes às suas regiões codificadora e 3'NT, podemos observar que há um bloco de haplótipos definido que engloba todo o gene *HLA-G* e pelo menos 20kb de nucleotídeos além da região 3'NT, local onde está inserida a sequência *Alu*YG (Figuras 6 e 7). Soma-se a isso o achado de Kulski *et al.* (2001) de que o forte LD também pode ser encontrado até o gene *HLA-A*, mostrando que esse bloco de haplótipos pode chegar a 200kb. Portanto, os eventos que atuam sobre o gene *HLA-G* e aumentam as frequência de determinados haplótipos, notadamente o G*01:01:01:01/UTR-1, também atuam

sobre a região de inserção *Alu*YG. Acreditamos que esse conservado bloco de haplótipos pode ser uma consequência do importante papel desempenhado pelo gene *HLA-G* na modulação do sistema imune e na tolerância imunológica, tão importante, por exemplo, durante a gravidez (Hviid, 2006; Larsen e Hviid, 2009; Donadi *et al.*, 2011), pois a manutenção de moléculas ativas na imunomodulação é de grande importância para a fisiologia humana.

As frequências mais altas para o haplótipo G*01:01:01:01/UTR-1 foram encontradas nas populações asiáticas, principalmente na população Chinesa do grupo étnico Han, o que está em acordo com as altas frequências da inserção *Alu*YG na Ásia (Dunn *et al.*, 2007; Tian *et al.*, 2008). Essas populações não apresentam o haplótipo UTR-6 ou seus haplótipos codificadores associados. Acreditamos que esse haplótipo possa ter sido perdido no curso evolutivo de formação dessa população por deriva genética ou por pressões seletivas, i.e., as mesmas razões para o aumento das frequências do haplótipo G*01:01:01:01/UTR-1. Em contraste, essas populações, assim como as africanas, apresentam altas frequências do haplótipo UTR-3 (Tabela 5), o que também pode ser observado para as populações senegalesa e congolosa analisadas nesse estudo (Tabela 5).

As frequências do haplótipo UTR-3 são baixas no continente africano, provavelmente também devido à atuação de efeito fundador somado às pressões seletivas. Essas populações não apresentam o haplótipo UTR-7 assim como seus codificadores associados (Tabelas 4 e 5). A UTR-7 foi, recentemente, associada com baixas produções da molécula solúvel de *HLA-G* (Martelli-Palomino *et al.*, 2013). A sua ausência nessas populações pode significar a presença de forças seletivas negativas atuando sobre esse haplótipo ou mesmo que ele seja mais recente, como já discutido. Entretanto, mais estudos são necessários para se desvendar esse cenário.

Não obstante, o fato de praticamente não haver recombinante (99% do alelo *Alu*YG foi associado com UTR-1 considerando-se o Brasil, Congo, Senegal e França, e 100% do alelo *Alu*YG foi associado ao haplótipo G*01:01:01:01/UTR-1, (considerando somente o Brasil) e a presença de baixas frequências de *Alu*YG na África, nos permite dizer que, apesar de apresentar altas frequências nas populações modernas, o haplótipo G*01:01:01:01/UTR-1 pode ser um dos mais recentes haplótipos de *HLA-G*. Caso contrário, seria esperada uma alta taxa de recombinação entre essas duas regiões. Da mesma forma, a análise das redes

(Figuras 8 e 9) sugere que, caso o haplótipo G*01:01:01:01/UTR-1 fosse antigo, seria esperado que outros haplótipos também compartilhassem a presença da inserção *Alu*YG. De fato, é bastante provável que a inserção tenha ocorrido cedo durante a emergência do haplótipo G*01:01:01:01/UTR-1 ainda na África, e que esse padrão permaneceu o mesmo desde então, ocorrendo tão somente o aumento de frequências até os dias atuais. Contudo, esse haplótipo já havia sido considerado como possivelmente um dos mais recentes haplótipos de *HLA-G* devido à ausência de uma inserção de 14pb presente na sua região 3'NT (Castro *et al.*, 2000). Esses pesquisadores mostraram que essa sequência estava presente em primatas e, teoricamente, os haplótipos modernos que possuem os 14pb em sua sequência são mais antigos em relação aos que não a apresentam.

As mesmas linhagens para o haplótipo estendido de *HLA-G* descritas por Castelli *et al.* (2011) foram encontradas pelo presente trabalho. Além disso, essas linhagens também foram encontradas quando utilizamos apenas os dados de polimorfismos da região 3'NT, somados ao dado da inserção *Alu*YG (Figura 8). Assim, novamente a identificação de associação entre a UTR-1 e a inserção *Alu*YG corrobora o fato de esse haplótipo ser considerado um dos mais recentes na história evolutiva desse gene. De fato, como não existe nenhum mecanismo conhecido que retire exatamente a sequência *Alu* de uma dada localização cromossômica, é pouco provável que a raiz da rede seja a linhagem HG010101, ou os haplótipos derivados desse também apresentariam a inserção *Alu*YG.

7. CONCLUSÃO

As frequências da inserção *Alu*YG nas quatro populações analisadas no presente manuscrito são menores nas populações africanas (congoleses e senegaleses) em comparação com as populações brasileira e francesa, indicando possível efeito do fundador atuando nessa região aumentando as frequências dessa inserção ao longo do tempo. Além disso, a existência dessa inserção em populações africanas, mesmo que em baixas frequências, indica que o evento de transposição ocorreu ainda nesse continente, antes da dispersão do homem aos outros continentes.

O perfil de Desequilíbrio de Ligação (LD) encontrado entre a inserção *Alu*YG e pontos de variação na região 3'NT do gene *HLA-G*, além do LD entre os pontos de variação da região 3'NT e da região codificadora desse gene, incluindo-se os dados disponibilizados pelo *1000Genomes Project Consortium*, sugere a existência de um bloco único de segregação abrangendo todo o gene *HLA-G* e estendendo-se por, pelo menos, 20-kb à partir da região 3'NT, indicando que essa região sofreu poucos eventos de recombinação desde a dispersão do homem à partir da África.

O presente trabalho corrobora a evidência de que o haplótipo G*01:01:01:01/UTR-1 pode ser o mais recente entre os mais frequentes haplótipos encontrados para o gene *HLA-G*. As frequências do haplótipo G*01:01:01:01/UTR-1/*Alu*YG em torno do mundo (e também para outros haplótipos do gene *HLA-G*) pode ser consequência de consecutivos efeitos de fundador atuando sobre esse gene desde a saída do homem do continente africano, bem como de seleção atuando sobre essas frequências.

8. REFERÊNCIAS

- ABBAS, A. K. ET AL. Cellular and Molecular Immunology. Philadelphia: Saunders/Elsevier, 2010. 566.
- ALLEN, P. D. Anesthesia and the human genome project: the quest for accurate prediction of drug responses. **Anesthesiology**, v. 102, n. 3, p. 494-5, Mar 2005.
- BANDELT, H. J.; FORSTER P.; ROHL A. Median-joining networks for inferring intraspecific phylogenies. **Mol Biol Evol**, v. 16, n. 1, p. 37-48, Jan 1999.
- BARRETT, J. C.; FRY B.; MALLER J. et al. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, v. 21, n. 2, p. 263-5, Jan 15 2005.
- BARTEL, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. **Cell**, v. 116, n. 2, p. 281-97, Jan 23 2004.
- BATZER, M. A.; DEININGER P. L. Alu repeats and human genomic diversity. **Nat Rev Genet**, v. 3, n. 5, p. 370-9, May 2002.
- BATZER, M. A.; DEININGER P. L.; HELLMANNBLUMBERG U. et al. Standardized nomenclature for Alu repeats. **Journal of Molecular Evolution**, v. 42, n. 1, p. 3-6, Jan 1996.
- BATZER, M. A.; STONEKING M.; ALEGRIA-HARTMAN M. et al. African origin of human-specific polymorphic Alu insertions. **Proc Natl Acad Sci U S A**, v. 91, n. 25, p. 12288-92, Dec 6 1994.
- CAROSELLA, E. D. The tolerogenic molecule HLA-G. **Immunol Lett**, v. 138, n. 1, p. 22-4, Jul 2011.
- CAROSELLA, E. D.; FAVIER B.; ROUAS-FREISS N. et al. Beyond the increasing complexity of the immunomodulatory HLA-G molecule. **Blood**, v. 111, n. 10, p. 4862-70, May 15 2008.
- CASTELLI, E. C.; MENDES-JUNIOR C. T.; DEGHAIDE N. H. et al. The genetic structure of 3'untranslated region of the HLA-G gene: polymorphisms and haplotypes. **Genes Immun**, v. 11, n. 2, p. 134-41, Mar 2010.
- CASTELLI, E. C.; MENDES-JUNIOR C. T.; VEIGA-CASTELLI L. C. et al. A comprehensive study of polymorphic sites along the HLA-G gene: implication for gene regulation and evolution. **Mol Biol Evol**, v. 28, n. 11, p. 3069-86, Nov 2011.
- CASTELLI, E. C.; MOREAU P.; OYA E CHIROMATZO A. et al. In silico analysis of microRNAs targeting the HLA-G 3' untranslated region alleles and haplotypes. **Hum Immunol**, v. 70, n. 12, p. 1020-5, Dec 2009.

CASTRO, M. J.; MORALES P.; MARTINEZ-LASO J. et al. Evolution of MHC-G in humans and primates based on three new 3'UT polymorphisms. **Hum Immunol**, v. 61, n. 11, p. 1157-63, Nov 2000.

CERVERA, I.; HERRAIZ M. A.; PENALOZA J. et al. Human leukocyte antigen-G allele polymorphisms have evolved following three different evolutionary lineages based on intron sequences. **Hum Immunol**, v. 71, n. 11, p. 1109-15, Nov 2010.

CHEN, S.; HU Q.; LIU Z. et al. The distribution of HLA alleles revealed a founder effect in the geographically isolated Chinese population, Drung. **Mol Immunol**, v. 44, n. 8, p. 2017-22, Mar 2007.

CONTINI, P.; GHIO M.; POGGI A. et al. Soluble HLA-G1/CD8 ligation triggers apoptosis in CD8(+) T lymphocytes and NK cells by FAS/FASL interaction and inhibits cytotoxic T-cell activity. **Tissue Antigens**, v. 62, n. 4, p. 347-347, Oct 2003a.

CONTINI, P.; GHIO M.; POGGI A. et al. Soluble HLA-A,-B,-C and -G molecules induce apoptosis in T and NKCD8(+) cells and inhibit cytotoxic T cell activity through CD8 ligation. **European Journal of Immunology**, v. 33, n. 1, p. 125-134, Jan 2003b.

CORDAUX, R.; HEDGES D. J.; HERKE S. W. et al. Estimating the retrotransposition rate of human Alu elements. **Gene**, v. 373, p. 134-7, May 24 2006.

COURTIN, D.; MILET J.; SABBAGH A. et al. HLA-G 3' UTR-2 haplotype is associated with Human African trypanosomiasis susceptibility. **Infect Genet Evol**, Mar 26 2013.

DE ALMEIDA, D. E.; LING S.; PI X. et al. Immune dysregulation by the rheumatoid arthritis shared epitope. **J Immunol**, v. 185, n. 3, p. 1927-34, Aug 1 2010.

DEININGER, P. L.; BATZER M. A. Alu repeats and human disease. **Mol Genet Metab**, v. 67, n. 3, p. 183-93, Jul 1999.

DELVES, P. J.; ROITT I. M. The immune system. First of two parts. **N Engl J Med**, v. 343, n. 1, p. 37-49, Jul 6 2000a.

DELVES, P. J.; ROITT I. M. The immune system. Second of two parts. **N Engl J Med**, v. 343, n. 2, p. 108-17, Jul 13 2000b.

DI CRISTOFARO, J.; EL MOUJALLY D.; AGNEL A. et al. HLA-G haplotype structure shows good conservation between different populations and good correlation with high, normal and low soluble HLA-G expression. **Hum Immunol**, v. 74, n. 2, p. 203-6, Feb 2013.

DONADI, E. A.; CASTELLI E. C.; ARNAIZ-VILLENA A. et al. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. **Cell Mol Life Sci**, v. 68, n. 3, p. 369-95, Feb 2011.

DUNN, D. S.; CHOY M. K.; PHIPPS M. E. et al. The distribution of major histocompatibility complex class I polymorphic Alu insertions and their associations with HLA alleles in a Chinese population from Malaysia. **Tissue Antigens**, v. 70, n. 2, p. 136-43, Aug 2007.

DUNN, D. S.; INOKO H.; KULSKI J. K. Dimorphic Alu element located between the TFIID and CDSN genes within the major histocompatibility complex. **Electrophoresis**, v. 24, n. 16, p. 2740-2748, Aug 2003.

DUNN, D. S.; INOKO H.; KULSKI J. K. The association between non-melanoma skin cancer and a young dimorphic Alu element within the major histocompatibility complex class I genomic region. **Tissue Antigens**, v. 68, n. 2, p. 127-34, Aug 2006.

DUNN, D. S.; NARUSE T.; INOKO H. et al. The association between HLA-A alleles and young Alu dimorphisms near the HLA-J, -H, and -F genes in workshop cell lines and Japanese and Australian populations. **Journal of Molecular Evolution**, v. 55, n. 6, p. 718-726, Dec 2002.

FERREIRA, L. B.; MENDES-JUNIOR C. T.; WIEZEL C. E. et al. Genomic ancestry of a sample population from the state of Sao Paulo, Brazil. **Am J Hum Biol**, v. 18, n. 5, p. 702-5, Sep-Oct 2006.

GAO, G. F.; WILLCOX B. E.; WYER J. R. et al. Classical and nonclassical class I major histocompatibility complex molecules exhibit subtle conformational differences that affect binding to CD8alpha. **J Biol Chem**, v. 275, n. 20, p. 15232-8, May 19 2000.

GARCIA-OBREGON, S.; ALFONSO-SANCHEZ M. A.; GOMEZ-PEREZ L. et al. Microsatellites and Alu elements from the human MHC in Valencia (Spain): analysis of genetic relationships and linkage disequilibrium. **Int J Immunogenet**, v. 38, n. 6, p. 483-91, Dec 2011.

GARCIA, A.; MILET J.; COURTIN D. et al. Association of HLA-G 3'UTR polymorphisms with response to malaria infection: A first insight. **Infect Genet Evol**, Mar 14 2013.

GAUDIERI, S.; DAWKINS R. L.; HABARA K. et al. SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. **Genome Res**, v. 10, n. 10, p. 1579-86, Oct 2000.

GERAGHTY, D. E.; KOLLER B. H.; HANSEN J. A. et al. The Hla Class-I Gene Family Includes at Least 6 Genes and 12 Pseudogenes and Gene Fragments. **Journal of Immunology**, v. 149, n. 6, p. 1934-1946, Sep 15 1992.

GERAGHTY, D. E.; KOLLER B. H.; ORR H. T. A human major histocompatibility complex class I gene that encodes a protein with a shortened cytoplasmic segment. **Proc Natl Acad Sci U S A**, v. 84, n. 24, p. 9145-9, Dec 1987.

GREGERSEN, P. K.; SILVER J.; WINCHESTER R. J. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. **Arthritis Rheum**, v. 30, n. 11, p. 1205-13, Nov 1987.

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

HARROW, J.; FRANKISH A.; GONZALEZ J. M. et al. GENCODE: the reference human genome annotation for The ENCODE Project. **Genome Res**, v. 22, n. 9, p. 1760-74, Sep 2012.

HENN, B. M.; CAVALLI-SFORZA L. L.; FELDMAN M. W. The great human expansion. **Proc Natl Acad Sci U S A**, v. 109, n. 44, p. 17758-64, Oct 30 2012.

HORTON, R.; WILMING L.; RAND V. et al. Gene map of the extended human MHC. **Nat Rev Genet**, v. 5, n. 12, p. 889-99, Dec 2004.

HOUCK, C. M.; RINEHART F. P.; SCHMID C. W. A ubiquitous family of repeated DNA sequences in the human genome. **J Mol Biol**, v. 132, n. 3, p. 289-306, Aug 15 1979.

HVIID, T. V. HLA-G in human reproduction: aspects of genetics, function and pregnancy complications. **Hum Reprod Update**, v. 12, n. 3, p. 209-32, May-Jun 2006.

HVIID, T. V.; HYLENIUS S.; RORBYE C. et al. HLA-G allelic variants are associated with differences in the HLA-G mRNA isoform profile and HLA-G mRNA levels. **Immunogenetics**, v. 55, n. 2, p. 63-79, May 2003.

HVIID, T. V.; RIZZO R.; CHRISTIANSEN O. B. et al. HLA-G and IL-10 in serum in relation to HLA-G genotype and polymorphisms. **Immunogenetics**, v. 56, n. 3, p. 135-41, Jun 2004.

HVIID, T. V.; SORENSEN S.; MORLING N. Polymorphism in the regulatory region located more than 1.1 kilobases 5' to the start site of transcription, the promoter region, and exon 1 of the HLA-G gene. **Hum Immunol**, v. 60, n. 12, p. 1237-44, Dec 1999.

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860-921, Feb 15 2001.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. **Nature**, v. 431, n. 7011, p. 931-45, Oct 21 2004.

ISHITANI, A.; GERAGHTY D. E. Alternative splicing of HLA-G transcripts yields proteins with primary structures resembling both class I and class II antigens. **Proc Natl Acad Sci U S A**, v. 89, n. 9, p. 3947-51, May 1 1992.

JACOB, S.; MCCLINTOCK M. K.; ZELANO B. et al. Paternally inherited HLA alleles are associated with women's choice of male odor. **Nat Genet**, v. 30, n. 2, p. 175-9, Feb 2002.

JASSEM, R. M.; SHANI W. S.; LOISEL D. A. et al. HLA-G polymorphisms and soluble HLA-G protein levels in women with recurrent pregnancy loss from Basrah province in Iraq. **Hum Immunol**, v. 73, n. 8, p. 811-7, Aug 2012.

JURKA, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. **Proc Natl Acad Sci U S A**, v. 94, n. 5, p. 1872-7, Mar 4 1997.

JURKA, J.; SMITH T. A fundamental division in the Alu family of repeated sequences. **Proc Natl Acad Sci U S A**, v. 85, n. 13, p. 4775-8, Jul 1988.

KAPITONOV, V.; JURKA J. The age of Alu subfamilies. **J Mol Evol**, v. 42, n. 1, p. 59-65, Jan 1996.

KLEIN, J.; SATO A. The HLA system. First of two parts. **N Engl J Med**, v. 343, n. 10, p. 702-9, Sep 7 2000a.

KLEIN, J.; SATO A. The HLA system. Second of two parts. **N Engl J Med**, v. 343, n. 11, p. 782-6, Sep 14 2000b.

KRIEGS, J. O.; CHURAKOV G.; JURKA J. et al. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. **Trends Genet**, v. 23, n. 4, p. 158-61, Apr 2007.

KUERSTEN, S.; GOODWIN E. B. The power of the 3' UTR: translational control and development. **Nat Rev Genet**, v. 4, n. 8, p. 626-37, Aug 2003.

KULSKI, J. K.; DUNN D. S. Polymorphic Alu insertions within the Major Histocompatibility Complex class I genomic region: a brief review. **Cytogenet Genome Res**, v. 110, n. 1-4, p. 193-202, 2005.

KULSKI, J. K.; DUNN D. S.; HUI J. et al. Alu polymorphism within the MICB gene and association with HLA-B alleles (vol 53, pg 975, 2002). **Immunogenetics**, v. 54, n. 5, p. 365-365, Aug 2002.

KULSKI, J. K.; GAUDIERI S.; DAWKINS R. L. Using Alu J elements as molecular clocks to trace the evolutionary relationships between duplicated HLA class I genomic segments. **Journal of Molecular Evolution**, v. 50, n. 6, p. 510-519, Jun 2000.

KULSKI, J. K.; MARTINEZ P.; LONGMAN-JACOBSEN N. et al. The association between HLA-A alleles and an Alu dimorphism near HLA-G. **Journal of Molecular Evolution**, v. 53, n. 2, p. 114-123, Aug 2001.

LARSEN, M. H.; HVIID T. V. Human leukocyte antigen-G polymorphism in relation to expression, function, and disease. **Hum Immunol**, v. 70, n. 12, p. 1026-34, Dec 2009.

LEWONTIN, R. C. The Interaction of Selection and Linkage. II. Optimum Models. **Genetics**, v. 50, p. 757-82, Oct 1964.

LISCHER, H. E.; EXCOFFIER L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. **Bioinformatics**, v. 28, n. 2, p. 298-9, Jan 15 2012.

LUCENA-SILVA, N.; MONTEIRO A. R.; DE ALBUQUERQUE R. S. et al. Haplotype frequencies based on eight polymorphic sites at the 3' untranslated region of the HLA-G gene in individuals from two different geographical regions of Brazil. **Tissue Antigens**, v. 79, n. 4, p. 272-8, Apr 2012.

MANASTER, I.; GOLDMAN-WOHL D.; GREENFIELD C. et al. MiRNA-mediated control of HLA-G expression and function. **PLoS One**, v. 7, n. 3, p. e33395, 2012.

MARTELLI-PALOMINO, G.; PANCOTO J.A.; MUNIZ Y.C.N. et al. Polymorphic sites at the 3' untranslated region of the HLA-G gene are associated with differential HLA-G soluble levels in the Brazilian and French population. **Plos One**, v. (in press), 2013.

MARTINEZ-LASO, J.; HERRAIZ M. A.; PENALOZA J. et al. Promoter sequences confirm the three different evolutionary lineages described for HLA-G. **Hum Immunol**, v. 74, n. 3, p. 383-8, Mar 2013.

MATHIAS, S. L.; SCOTT A. F.; KAZAZIAN H. H., JR. et al. Reverse transcriptase encoded by a human transposable element. **Science**, v. 254, n. 5039, p. 1808-10, Dec 20 1991.

MENDES-JUNIOR, C. T.; CASTELLI E. C.; MEYER D. et al. Genetic diversity of the HLA-G coding region in Amerindian populations from the Brazilian Amazon: a possible role of natural selection. **Genes Immun**, Oct 3 2013.

MENDES, C. T.; CASTELLI E. C.; SIMOES R. T. et al. HLA-G 14-bp polymorphism at exon 8 in Amerindian populations from the Brazilian Amazon. **Tissue Antigens**, v. 69, n. 3, p. 255-260, Mar 2007.

MEYER, D.; THOMSON G. How selection shapes variation of the human major histocompatibility complex: a review. **Annals of Human Genetics**, v. 65, p. 1-26, Jan 2001.

MOREAU, P.; FLAJOLLET S.; CAROSELLA E. D. Non-classical transcriptional regulation of HLA-G: an update. **Journal of Cellular and Molecular Medicine**, v. 13, n. 9B, p. 2973-2989, Sep 2009.

MOSCOSO, J.; SERRANO-VELA J. I.; PACHECO R. et al. HLA-G, -E and -F: allelism, function and evolution. **Transpl Immunol**, v. 17, n. 1, p. 61-4, Dec 2006.

MUNIZ, Y. C.; FERREIRA L. B.; MENDES-JUNIOR C. T. et al. Genomic ancestry in urban Afro-Brazilians. **Ann Hum Biol**, v. 35, n. 1, p. 104-11, Jan-Feb 2008.

OBER, C.; ALDRICH C. L. HLA-G polymorphisms: neutral evolution or novel function? **J Reprod Immunol**, v. 36, n. 1-2, p. 1-21, Nov 30 1997.

PARK, B.; AHN K. An essential function of tapasin in quality control of HLA-G molecules. **J Biol Chem**, v. 278, n. 16, p. 14337-45, Apr 18 2003.

PARK, B.; LEE S.; KIM E. et al. The truncated cytoplasmic tail of HLA-G serves a quality-control function in post-ER compartments. **Immunity**, v. 15, n. 2, p. 213-24, Aug 2001.

PARKIN, J.; COHEN B. An overview of the immune system. **Lancet**, v. 357, n. 9270, p. 1777-89, Jun 2 2001.

PENN, D. J.; DAMJANOVICH K.; POTTS W. K. MHC heterozygosity confers a selective advantage against multiple-strain infections. **Proc Natl Acad Sci U S A**, v. 99, n. 17, p. 11260-4, Aug 20 2002.

PIERTNEY, S. B.; OLIVER M. K. The evolutionary ecology of the major histocompatibility complex. **Heredity**, v. 96, n. 1, p. 7-21, Jan 2006.

PIMENTA, J. R.; ZUCCHERATO L. W.; DEBES A. A. et al. Color and genomic ancestry in Brazilians: a study with forensic microsatellites. **Hum Hered**, v. 62, n. 4, p. 190-5, 2006.

QIN, Z. S.; NIU T.; LIU J. S. Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. **Am J Hum Genet**, v. 71, n. 5, p. 1242-7, Nov 2002.

RAYMOND, M.; ROUSSET F. Genepop (Version-1.2) - Population-Genetics Software for Exact Tests and Ecumenicism. **Journal of Heredity**, v. 86, n. 3, p. 248-249, May-Jun 1995.

ROUSSEAU, P.; LE DISCORDE M.; MOUILLOT G. et al. The 14 bp deletion-insertion polymorphism in the 3' UT region of the HLA-G gene influences HLA-G mRNA stability. **Hum Immunol**, v. 64, n. 11, p. 1005-10, Nov 2003.

ROUSSET, F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. **Mol Ecol Resour**, v. 8, n. 1, p. 103-6, Jan 2008.

ROY-ENGEL, A. M.; CARROLL M. L.; VOGEL E. et al. Alu insertion polymorphisms for the study of human genomic diversity. **Genetics**, v. 159, n. 1, p. 279-90, Sep 2001.

SABBAGH, A.; COURTIN D.; MILET J. et al. Association of HLA-G 3' untranslated region polymorphisms with antibody response against Plasmodium falciparum antigens: preliminary results. **Tissue Antigens**, v. 82, n. 1, p. 53-8, Jul 2013.

SEGAL, S.; HILL A. V. S. Genetic susceptibility to infectious disease. **Trends in Microbiology**, v. 11, n. 9, p. 445-448, Sep 2003.

- SHIINA, T.; HOSOMICHI K.; INOKO H. et al. The HLA genomic loci map: expression, interaction, diversity and disease. **Journal of Human Genetics**, v. 54, n. 1, p. 15-39, Jan 2009.
- SHIROISHI, M.; KUROKI K.; RASUBALA L. et al. Structural basis for recognition of the nonclassical MHC molecule HLA-G by the leukocyte Ig-like receptor B2 (LILRB2/LIR2/ILT4/CD85d). **Proc Natl Acad Sci U S A**, v. 103, n. 44, p. 16412-7, Oct 31 2006.
- SHIROISHI, M.; TSUMOTO K.; AMANO K. et al. Human inhibitory receptors Ig-like transcript 2 (ILT2) and ILT4 compete with CD8 for MHC class I binding and bind preferentially to HLA-G. **Proc Natl Acad Sci U S A**, v. 100, n. 15, p. 8856-61, Jul 22 2003.
- SOLBERG, O. D.; MACK S. J.; LANCASTER A. K. et al. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. **Hum Immunol**, v. 69, n. 7, p. 443-64, Jul 2008.
- SOLIER, C.; MALLET V.; LENFANT F. et al. HLA-G unique promoter region: functional implications. **Immunogenetics**, v. 53, n. 8, p. 617-25, Oct 2001.
- STEPHENS, M.; DONNELLY P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. **Am J Hum Genet**, v. 73, n. 5, p. 1162-9, Nov 2003.
- STEPHENS, M.; SMITH N. J.; DONNELLY P. A new statistical method for haplotype reconstruction from population data. **Am J Hum Genet**, v. 68, n. 4, p. 978-89, Apr 2001.
- TAN, Z.; RANDALL G.; FAN J. et al. Allele-specific targeting of microRNAs to HLA-G and risk of asthma. **Am J Hum Genet**, v. 81, n. 4, p. 829-34, Oct 2007.
- TAN, Z.; SHON A. M.; OBER C. Evidence of balancing selection at the HLA-G promoter region. **Hum Mol Genet**, v. 14, n. 23, p. 3619-28, Dec 1 2005.
- The 1000Genomes Project Consortium. A map of human genome variation from population-scale sequencing. **Nature**, v. 467, n. 7319, p. 1061-73, Oct 28 2010.
- The 1000Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. **Nature**, v. 491, n. 7422, p. 56-65, Nov 1 2012.
- THORSBY, E. A short history of HLA. **Tissue Antigens**, v. 74, n. 2, p. 101-16, Aug 2009.
- TIAN, W.; WANG F.; CAI J. H. et al. Polymorphic insertions in 5 Alu loci within the major histocompatibility complex class I region and their linkage disequilibria with HLA alleles in four distinct populations in mainland China. **Tissue Antigens**, v. 72, n. 6, p. 559-67, Dec 2008.

ULLU, E.; TSCHUDI C. Alu Sequences Are Processed 7sl Rna Genes. **Nature**, v. 312, n. 5990, p. 171-172, 1984.

VAN ROOD, J. J. The impact of the HLA-system in clinical medicine. **Schweiz Med Wochenschr**, v. 123, n. 3, p. 85-92, Jan 23 1993.

VEIGA-CASTELLI, L. C.; CASTELLI E. C.; MENDES C. T., JR. et al. Non-classical HLA-E gene variability in Brazilians: a nearly invariable locus surrounded by the most variable genes in the human genome. **Tissue Antigens**, v. 79, n. 1, p. 15-24, Jan 2012.

VEIT, T. D.; CAZAROLLI J.; SALZANO F. M. et al. New evidence for balancing selection at the HLA-G locus in South Amerindians. **Genet Mol Biol**, v. 35, n. 4 (suppl), p. 919-23, Dec 2012.

VEIT, T. D.; CHIES J. A. Tolerance versus immune response -- microRNAs as important elements in the regulation of the HLA-G gene expression. **Transpl Immunol**, v. 20, n. 4, p. 229-31, Mar 2009.

VENTER, J. C. A part of the human genome sequence. **Science**, v. 299, n. 5610, p. 1183-4, Feb 21 2003.

VENTER, J. C.; ADAMS M. D.; MYERS E. W. et al. The sequence of the human genome. **Science**, v. 291, n. 5507, p. 1304-51, Feb 16 2001.

WEDEKIND, C.; SEEBECK T.; BETTENS F. et al. MHC-dependent mate preferences in humans. **Proc Biol Sci**, v. 260, n. 1359, p. 245-9, Jun 22 1995.

WEICHENRIEDER, O.; STEHLIN C.; KAPP U. et al. Hierarchical assembly of the Alu domain of the mammalian signal recognition particle. **RNA**, v. 7, n. 5, p. 731-40, May 2001.

YIE, S. M.; LI L. H.; XIAO R. et al. A single base-pair mutation in the 3'-untranslated region of HLA-G mRNA is associated with pre-eclampsia. **Mol Hum Reprod**, v. 14, n. 11, p. 649-53, Nov 2008.

ANEXO

Insights on the *HLA-G* Evolutionary History Provided by a Nearby *Alu* Insertion

Kaísson E. Santos,¹ Thálitta H.A. Lima,¹ Leandro P. Felício,¹ Juliana D. Massaro,² Gustavo M. Palomino,² Ana Carolina A. Silva,³ Silviene F. Oliveira,³ Audrey Sabbagh,^{4,5} Andre Garcia,^{4,5} Philippe Moreau,^{6,7} Eduardo A. Donadi,² Celso T. Mendes-Junior,⁸ and Erick C. Castelli^{*,9}

¹Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia-GO, Brasil

²Divisão de Imunologia Clínica, Departamento de Clínica Médica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, USP, Ribeirão Preto, SP, Brasil

³Laboratório de Genética, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília-DF, Brasil

⁴Institut de Recherche pour le Développement, UMR 216 Mère et enfant face aux infections tropicales, Université Paris Descartes, 4, avenue de l'Observatoire, Paris, France

⁵Faculté de Pharmacie, Université Paris Descartes, Sorbonne Paris Cité, 4, avenue de l'Observatoire, Paris, France

⁶Commissariat à l'Énergie Atomique et aux Énergies Alternatives, Institut des Maladies Emergentes et des Thérapies Innovantes, Service de Recherches en Hémato-Immuno-génétique, Hôpital Saint-Louis, Paris, France

⁷Université Paris-Diderot, Sorbonne Paris-Cité, UMR ES, Institut Universitaire d'Hématologie, Hôpital Saint-Louis, Paris, France

⁸Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brasil

⁹Departamento de Patologia, Faculdade de Medicina de Botucatu, Universidade Estadual Paulista – UNESP, Botucatu, SP, Brasil

*Corresponding author: E-mail: castelli@fmb.unesp.br

Associate editor: Meredith Yeager

Abstract

The *AluYHG* element belongs to the *AluYb8* subfamily. It is a polymorphic insertion, located approximately 20 kb from the *HLA-G* 3'-untranslated region (3'-UTR), which has been used for evolution studies because it exhibits identity for descendants and it is still polymorphic in the human genome. To understand the evolutionary mechanisms acting on *HLA-G*, we evaluated the presence or absence of the *AluYHG* element, associating this variable site with others observed at *HLA-G* coding, 3'-UTR, or both regions in four distinct populations (Brazilian, French, Congolese, and Senegalese). The results were compared with the 1000Genomes Consortium data. The worldwide *AluYHG* frequencies showed an increment, starting lower in Africa and increasing following distance and time of human dispersion out of Africa. The same haplotype pattern was observed in all populations, indicating that most of the *HLA-G* haplotypes already detected were originated earlier in Africa, before *Homo sapiens* dispersion. The *AluYHG* insertion was associated with the G*01:01:01:01/UTR-1 haplotype, with rare recombinants. Despite its high frequency in worldwide populations, the G*01:01:01:01/UTR-1 haplotype should be very recent. The low frequency of recombinants indicates that the rate of recombination at the *HLA-G* gene is very low.

Key words: *AluYHG*, *HLA-G*, 3'-untranslated region, haplotypes, worldwide diversity, human evolution.

Introduction

The human major histocompatibility complex (MHC) comprises at least 224 genes at chromosome 6p21.3, coding the human leukocyte antigens (HLAs) that have a key role on the immune system. Classical class I genes (*HLA-A*, *HLA-B*, and *HLA-C*) encode molecules that present antigen peptides to T CD8+ cells, whereas the nonclassical class I genes (*HLA-G*, *HLA-E*, and *HLA-F*) have been primarily associated with the modulation of the immune system cells (Klein and Sato 2000; Hviid 2006; Donadi et al. 2011).

HLA-G has been considered to be an immune modulatory molecule, predominantly expressed at the maternal–fetal interface and has primarily been associated with maternal–fetal tolerance (Hviid 2006; Carosella et al. 2008; Berger et al. 2010). The complete *HLA-G* molecule presents the same

extracellular structure of the classical *HLA* counterparts; however, its major function is not antigen presentation. *HLA-G* inhibits the cytotoxic activity of T CD8+ and NK cells through direct interaction with leukocyte receptors, such as ILT-2 (LILRB1 and CD158j), ILT-4 (LILRB2 and CD158d), and KIR2DL4 (CD158d) (Colonna 1997; Ponte et al. 1999; Rajagopalan and Long 1999; Cao et al. 2000; Contini et al. 2003; Shiroishi et al. 2006; Donadi et al. 2011). *HLA-G* gene and molecule expression patterns differ in relation to classical *HLA* class I molecules in many aspects, including 1) restricted tissue expression in nonpathological conditions (Lee et al. 1995), being expressed at the maternal–fetal interface in the extravillous cytotrophoblast cells (Berger et al. 2010), thymus, cornea, proximal nail matrix, pancreas, and hematopoietic stem cells (Crisa et al. 1997; Mallet et al. 1999;

