

Universidade de Brasília
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Biologia Animal

Dinâmica de Marcadores Genéticos na Região do Complexo de Histocompatibilidade
Principal Humano em Populações do Centro-Oeste do Brasil

Ana Carolina Arcanjo Silva

Brasília

2012

Dinâmica de Marcadores Genéticos na Região do Complexo de Histocompatibilidade
Principal Humano em Populações do Centro-Oeste do Brasil

Dissertação apresentada ao Programa de
Pós-Graduação em Biologia Animal
como requisito parcial para obtenção do
grau de mestre em Biologia Animal.

Ana Carolina Arcanjo Silva
Orientadora: Dra. Silviene Fabiana de Oliveira

Brasília
2012

Trabalho desenvolvido no Laboratório de Genética da Universidade de Brasília–UnB e no Laboratório de Imunologia Molecular da Faculdade de Medicina de Ribeirão Preto, no Hospital das Clínicas da Universidade de São Paulo - USP, com o suporte financeiro da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

“O sabor de uma vitória depende da intensidade da luta.”

Autor Desconhecido

Agradecimentos

À querida amiga, mestre, modelo e orientadora, Dra. Silviene Fabiana de Oliveira, por todo o carinho, amizade, compreensão e, sobretudo, por transformar a minha vida e fazer de mim uma cientista.

Aos membros da banca examinadora, por aceitarem o convite e doarem seu tempo e conhecimento para o melhoramento deste trabalho.

À Dra. Neifi Hassan Saloum Deghaide, pela atenção e carinho enquanto estive trabalhando em Ribeirão Preto.

Ao Laboratório de Imunologia Molecular do Hospital das Clínicas da USP de Ribeirão Preto, especialmente ao Dr. Eduardo Antônio Donadi por abrir as portas para mim, e ao Biologista Rubens Donizeti da Silveira por todo o apoio técnico dispensado a mim.

Ao Dr. Jose Angel Pena, da Universidade do País Basco, por ter aceitado fazer parte deste trabalho, fornecendo materiais essenciais para o mesmo.

Aos responsáveis pelo sequenciador do laboratório de Biologia Molecular da UnB, Adriane Oliveira e Marcus Teixeira, pela atenção e cuidado dispensados.

Aos colegas do Laboratório de Genética Humana, na forma de professores de práticas (Carol, Dudu, Rafa, Neide, Gustavo, Gabriel e Mila) e estagiários que sofreram comigo (Tati, Gabi, Luana, Raphael e Mari).

À Dra. Cynthia Maria Kyaw, por me ouvir chorando e reclamando das PCRs, acalmando e sugerindo novas abordagens para meus experimentos, com enorme carinho.

Às minhas queridas Laaa, Thais (a Rural), Lena, Karla (com K), Póka e Lalauzinha (minha Flor), Vitão e Lola, meus filhos postiços.

À minha Lena, que esteve comigo desde a seleção do Mestrado até os vários fins-de-semana no laboratório, ajudando com PCRs e géis de poliacrilamida, evitando que eu morresse de dor por causa da tendinite.

Às queridas Tata, Tabata, Rafa e Kami, por cuidarem da minha branquela enquanto ela morava em São Carlos.

À Dona Augusta Barbosa e Silva (*in memoriam*), madrinha e avó, que faleceu nos últimos meses desse mestrado. Saudades eternas...

À minha família de origem, Miguel, Lucimar, Augusto e Gustavo, por entenderem minha ausência em casa e me incentivarem a ir cada vez mais longe. A meu pai, por me ensinar a disciplina e que o maior bem que podemos obter na vida é o conhecimento; à minha mãe, por me dar colo e me ensinar que sempre devemos seguir o nosso coração, por se interessar pelo que eu faço, mesmo sem entender muita coisa do que digo.

Ao meu querido irmão, Gustavo Arcanjo, pelo suporte TI e paciência com minhas perguntas ignorantes sobre o funcionamento de um computador. Se não fosse por você, creio que metade das análises estatísticas não poderiam ter sido concluídas.

À minha irMÃE Ana Paula, e à segunda família que ela me deu: Amanda, Isadorinha e Paulo Santos, meu cunhadinho querido.

Aos meus sogros, Walter e Elayne, por me fazerem sentir em casa em cada momento com eles.

Por fim, e não menos importante, à família que eu escolhi: Helena, Louis e Mingau (*in memoriam*), Mika, Pedro Leopoldo e Chanel, meus amores.

Índice

Índice de Tabelas 9

Índice de Figuras 10

Resumo 11

Abstract 12

Introdução 13

Marcadores genéticos 14

Inserções *Alu* 14

STRs (Short Tandem Repeats) 16

Objeto de Estudo 17

Objetivos 23

Objetivos específicos 23

Material e Métodos 24

Populações escolhidas para estudo 24

Aspectos éticos 26

Material biológico 26

Marcadores moleculares 26

Amplificação dos marcadores 27

Genotipagem dos marcadores 31

Inserções *Alu* 31

Microssatélites TNF 33

Eletroforese capilar 33

Identificação dos alelos 34

Análises estatísticas 36

Resultados e Discussão 39

Microssatélites TNF 39

Inserções *Alu* 46

Desequilíbrio de ligação: microssatélites TNF e inserções <i>Alu</i>	47
Microssatélites TNF: haplótipos	48
Haplótipos das inserções <i>Alu</i>	50
Microssatélites TNF e Inserções <i>Alu</i>	54
Diferenciação populacional	55
Teste de neutralidade	62

Conclusões 63

Referências 64

Anexos 74

Anexo 01	74 a 76
Anexo 02	77
Anexo 03	80
Anexo 04	83
Anexo 05	92

Índice de Tabelas

Tabela 01. Iniciadores, alelos e referências para os marcadores *Alu* utilizados no presente trabalho.

Tabela 02. Concentração dos reagentes para as reações de PCR para os marcadores *Alu* estudados.

Tabela 03. Iniciadores, número de alelos, variação no tamanho dos alelos, motivo de repetição e tamanho de fragmentos obtidos para os microssatélites TNF.

Tabela 04. Sequências selecionada para marcação com fluorescência, tamanho do fragmento gerado após amplificação e respectivo fluoróforo.

Tabela 05. Frequência dos alelos dos microssatélites TNF, teste de aderência ao Equilíbrio de Hardy-Weinberg, heterozigose observada e esperada para cada *locus* e teste de deficiência e excesso de heterozigotos em Brasília e Kalunga.

Tabela 06. Frequência dos alelos das inserções *Alu*, teste de aderência ao Equilíbrio de Hardy-Weinberg, heterozigose observada e esperada para cada locus e teste de deficiência e excesso de heterozigotos em Brasília e Kalunga.

Tabela 07. Haplótipos TNF compartilhados em Brasília e Kalunga, e a incidência deles em cada população.

Tabela 08. Listagem de todos os haplótipos encontrados nas duas populações estudadas, bem como suas frequências e desvio-padrão da estimativa de frequência. Os valores marcados indicam haplótipos com frequência maior ou igual a 10%.

Tabela 09. Haplótipos compartilhados entre as populações de Brasília e Kalunga e a incidência dos mesmos.

Tabela 10. Índice de diferenciação populacional (F_{ST}) entre Brasília e Kalunga, calculado a partir dos *loci* estudados.

Índice de Figuras

Figura 01. Estrutura dimérica geral de um elemento *Alu*. Retirado de Strachan & Read (1999).

Figura 02. Estrutura básica de um VNTR, sendo este um STR com motivo de dois pares de base, CG.

Figura03. **A.** Cromossomo 6 humano e a localização dos genes do MHC; **B.** Distribuição de parte dos genes do MHC humano.

Figura 04. Representação do número de alelos descritos para genes do Complexo de Histocompatibilidade Principal Humano. Cada ponto representa até 50 alelos. Adaptado de Trowsdale (2011).

Figura 05. Mapa indicando a localização geográfica das duas populações estudadas. Alfinete vermelho: Brasília; Alfinete azul: Kalunga.

Figura 06. Localização de marcadores *Alu* e microssatélites ao longo da região do MHC, bem como as distâncias entre eles.

Figura 07. Localização dos microssatélites TNF α -e na região dos genes TNF α e TNF β . IR são os iniciadores para cada microssatélite, como mencionado na Tabela 04. Figura adaptada de Udalova *et al.* (1993).

Figura 08. Exemplo de gel de poliacrilamida a 6%, corado com nitrato de prata, para a genotipagem de marcador do tipo *Alu*, no caso marcador AluHJ. O marcador de peso molecular utilizado, ao centro do gel, é o 100bp (Invitrogen).

Figura 09. Exemplo de eletroferogramas obtidos para os marcadores TNFa (A), TNFb (B), TNFd (C) e TNFe (D). Os *bins* estão representados como faixas acimentadas, que denotam o campo de aceitação para cada alelo.

Figura 10. Representação gráfica da diferenciação populacional (F_{ST}) entre Brasília e Kalunga com relação a outros grupos populacionais - Irlanda, Caucasianos, Escócia, Croácia, Valência, Bascos, Coréia, Curitiba e Ameríndios, com relação ao locus TNFa.

Figura 11. Representação gráfica da diferenciação populacional (F_{ST}) Brasília e Kalunga, com relação a outros grupos populacionais - Irlanda, Caucasianos, Escócia, Croácia e Coréia.

Figura 12. Representação gráfica das análises realizadas para diferenciação populacional baseadas em F_{ST} para Brasília, Kalunga, Irlanda, Caucasianos, Croácia e Coréia.

Figura 13. Matriz de diferenciação populacional baseado em F_{ST} para populações dos diversos continentes (Brasília, Kalunga, Australianos, Valência, Japoneses, Tailandeses, Chineses, Malaio-Chineses, Mongóis, Bantu, Sekele e !Kung), analisadas para o conjunto de marcadores *Alu* presentes neste trabalho.

Figura 14. Diferenciação populacional baseada em F_{ST} com relação à frequência dos haplótipos encontrados em cada população, nas populações Brasília, Kalunga, Australianos, Valência, Tailandeses, Mongóis, Chineses Hunan, Bantu, Sekele e !Kung.

Figura 15. Média das diferenças intra e inter-populacionais para cada par de populações analisadas, bem como a distância de Nei estimada pela média de diferenças encontradas entre cada par de populações para as populações Brasília, Kalunga, Australianos, Valência, Tailandeses, Mongóis, Chineses Hunan, Bantu, Sekele e !Kung.

Resumo

Teoricamente, regiões genômicas com grande importância funcional, como genes e regiões regulatórias, são as mais afetadas pela seleção natural. O MHC humano (6p23.1) é uma dessas regiões pois contém a maioria dos genes relacionados à regulação e ação do sistema imune em humanos. Neste trabalho foram escolhidos marcadores genéticos localizados no MHC humano de dois tipos – STRs, que apresentam alta taxa de mutação, e inserções *Alu*, com baixa taxa de mutação. Foram analisados 165 indivíduos de duas populações, uma urbana com alto grau de miscigenação (Brasília, 105 indivíduos) e uma comunidade afroderivada remanescente de quilombo (Kalunga, 60 indivíduos). Estes foram genotipados em relação aos marcadores AluMICB, AluTF, AluHJ, AluHG e AluHF (no MHC de Classe I), bem como para os microssatélites TNFa, TNFb, TNFc, TNFd, TNFe (no MHC de Classe III). Os fragmentos foram amplificados por meio de PCR e genotipados em gel de poliacrilamida a 6% (todos os *Alu*), gel de poliacrilamida desnaturante a 12% (TNFc) e eletroforese capilar em sequenciador automático ABI3130 (TNFa-e). Os dados obtidos foram analisados em três grupos: haplótipos *Alu*, haplótipos microssatélites e haplótipos contendo todos os 10 marcadores estudados, cujas inferências foram feitas com o algoritmo PHASE. Os marcadores TNFa e TNFd se mostraram fora do Equilíbrio de Hardy-Weinberg em ambas populações estudadas, mesmo após Correção de Bonferroni. Brasília e Kalunga apresentaram déficit global de heterozigotos. Em Kalunga o marcador AluHG apresentou excesso de heterozigotos mesmo sem estar fora do EHW, o que pode indicar ação de mecanismo evolutivo devido à sua proximidade com o gene HLA-G. Mesmo fora do EHW em Kalunga, o marcador TNFd não mostrou déficit ou excesso de heterozigotos, o que também pode indicar atuação de mecanismo evolutivo. Os haplótipos analisados mostraram diferenças de comportamento entre os tipos de marcadores moleculares. Conforme as análises feitas com base nos haplótipos *Alu*, as populações são similares e apresentam pouca diferenciação populacional, além de compartilharem os haplótipos mais frequentes. Para os haplótipos TNF, entretanto, as populações se mostram moderadamente diferentes e compartilham apenas 21 dos 190 haplótipos possíveis. O Teste de Neutralidade de Ewens-Watterson indicou desvio do esperado em condições de neutralidade para os *loci* TNFa, TNFb e TNFd isoladamente, mas não indicou desvio da neutralidade para nenhum dos conjuntos haplotípicos estudados.

Abstract

Genetic markers are used in population genetics to address different questions, like genetic counseling or the understanding of evolutionary mechanisms. Especially, genetic markers located in important genomic regions, such as genes or regulatory sequences, are most commonly affected by natural selection. The human MHC (6p23.1) is a region of great importance due to the presence of genes related to action and regulation of the human immune system, thus, several studies have reported that natural selection might be acting in such region. In this work, two types of genetic markers were selected in the MHC region, one of them being *Alu* insertions, which have low rates of mutation; and the other being STRs, which have rather high mutation rates. 165 samples from two distinct populations, an urban admixed population (Brasília, 105 samples) and an afro-derived rural population (Kalunga, 60 individuals) were studied for the *loci* *Alu*MICB, *Alu*TF, *Alu*HJ *Alu*HG e *Alu*HF (Class I MHC) and *TNFA*, *TNFB*, *TNFC*, *TNFD*, *TNFE* (Class III MHC). The fragments were amplified by PCR and genotyped in 6% polyacrylamide gel (*Alu* insertions), 12% denaturing polyacrylamide gel (*TNFC*) and capillary electrophoresis in an ABI3130 automatic sequencer (*TNFA*-e). Data were analyzed in three groups: *Alu* haplotypes, microsatellite haplotypes and haplotypes containing all *loci* here studied. The haplotypes were inferred with the PHASE algorithm. Both *TNFA* and *TNFD* *loci* departed from Hardy-Weinberg Equilibrium (HWE) in both populations, even after Bonferroni's Correction. Brasília and Kalunga showed global heterozygote deficit. *Alu*HG showed heterozygote excess even in accordance with HWE, which may indicate influence of evolutionary mechanisms due to its proximity to HLA-G gene. Aside being apart of HWE in Kalunga, *TNFD* *locus* did not show excess or deficit of heterozygotes. The analyzed haplotypes showed behavioral differences among the two sets of markers. In the analysis based on the *Alu* haplotypes, both populations are similar and show little populational divergence. For the *TNF* haplotypes, however, both populations showed moderate differences and share only 21 of the 190 possible haplotypes. The Ewens-Watterson Neutrality test indicated that the *TNFA*, *TNFB* and *TNFD* *loci* do not behave neutrally when analyzed separately, but did not show departure of neutrality for any of the haplotypes analyzed.

Introdução

A genética de populações tem como objetivo avaliar a composição genética de populações naturais dos mais variados organismos visando melhor entender a dinâmica dos mecanismos evolutivos. No caso de populações humanas, aborda tanto questões teóricas quanto práticas. Desta forma, busca, por exemplo, o melhor entendimento da evolução humana e a reconstrução histórica a partir de dados genéticos, assim como contribui para o aconselhamento genético de pais e outros parentes de pacientes com doenças hereditárias; o mapeamento genético e a identificação de genes que possam predispor ou conferir alguma susceptibilidade ou resistência a doenças; e a identificação humana, tanto para fins civis quanto penais (Hartl & Clark, 2010).

Ao longo da história da genética de populações, diversas técnicas de acesso a dados foram desenvolvidas e novas regiões do genoma passaram a ser acessadas, levando à descrição de uma enorme gama de marcadores genéticos. A busca por um melhor conhecimento da variabilidade genética humana, assim como a aplicação desse conhecimento na genética médica e outros setores do conhecimento, permeou o surgimento de projetos como o *International HapMap Project* (<http://hapmap.ncbi.nlm.nih.gov/>), que visa um melhor conhecimento da variabilidade para assessorar a genética médica e a indústria farmacêutica; o *Human Genome Organisation* (HUGO - <http://www.hugo-international.org/>) e o *Genographic Project* (<https://genographic.nationalgeographic.com/genographic/index.html>), este relacionado a evolução humana e aquele à coordenação internacional das iniciativas de pesquisa sobre o genoma humano.

Marcadores Genéticos

O termo marcador genético pode se referir a um gene, um sítio de restrição ou qualquer outra sequência do DNA que apresente diferentes versões alélicas para aquele *locus* (Edwards *et al.*, 1991). A genética de populações utiliza, hoje, marcadores genéticos de diferentes naturezas e com diferentes taxas de mutação. Por exemplo, inserções *Alu* apresentam taxa de uma inserção a cada vinte nascimentos (Cordeaux *et al.*, 2006), enquanto STRs exibem taxas de mutação tão altas quanto $1,2 \times 10^{-3}$ mutações por *locus* por gameta por geração (Weber & Long, 1993). Dependendo do padrão de herança, do número e frequência de alelos, diferentes marcadores podem mostrar diferenças drásticas sobre a quantidade de informação que pode ser explorada sobre a relação de indivíduos, populações e filogenia (Wang, 2006).

Inserções Alu

Transposons são sequências de DNA que podem mudar sua localização no genoma (Rogers, 1983). Também conhecidos como elementos transponíveis, são capazes de criar rearranjos inovadores no genoma, sendo agentes de variação genética (Hartl & Clark, 2010). Existem várias famílias de elementos transponíveis anotados no genoma humano, mas a maioria delas é muito antiga e não é mais capaz de retrotransposição ativa. Algumas delas, no entanto, permanecem ativas e são grande fonte de variabilidade entre seres humanos, influenciando inclusive a arquitetura do genoma (Stewart *et al.*, 2011). Dentre eles, os SINE (*short interspersed nuclear elements*) e LINE (*long interspersed nuclear elements*). Estes são transposons de classe I que sofrem transposição durante um processo replicativo, por meio de um intermediário de RNA, que depois é transcrito reversamente em DNA de fita dupla por meio de uma *transcriptase reversa* (Hartl & Clark, 2010). Os elementos SINE não codificam

sua própria *transcriptase reversa* e supostamente são mobilizados pelas enzimas codificadas pelos elementos LINE que residem no genoma humano (Novick *et al.*, 1996)

O mais abundante SINE em humanos é inserção *Alu*, assim denominada devido à presença de um sítio alvo de restrição para a enzima *Alu* (Houck *et al.*, 1979). Esses elementos têm cerca de 300 pb, com dois braços praticamente idênticos, divididos por uma região rica em adenina (Figura 01). Estima-se que exista aproximadamente 1,1 milhão de cópias desse tipo de sequência no genoma humano (Kriegs *et al.*, 2007). Há indícios de que esses elementos continuaram a se mover durante a evolução humana, mesmo depois da separação entre os humanos e demais primatas, o que é evidenciado pela existência de elementos *Alu* específicos de *Homo sapiens* (Callinan & Batzer, 2006). As inserções *Alu* são ferramentas altamente úteis para a análise da reconstrução histórica de populações humanas e dos eventos demográficos pertinentes a estas. Isto porque a inserção desses elementos é um tipo de evento mutacional estável, único e unidirecional (Shedlock & Okada, 2000; Salem *et al.*, 2005).

Marcadores do tipo LINE e SINE são interessantes para a genética de populações por dois motivos: (a) é extremamente pequena a probabilidade de que uma determinada sequência ocorra no mesmo local em duas pessoas diferentes, salvo quando estas pessoas compartilham um ancestral comum —uma vez que a taxa de inserção dos elementos *Alu* é muito baixa e há muitos locais em que esses elementos podem se inserir (Batzer & Deininger, 1991) – e (b) o estado ancestral do locus é a ausência de inserção (Batzer *et al.*, 1994; Whitherspoon *et al.*, 2006). Além disso, diferentes inserções *Alu* cujas distribuições geográficas são conhecidas podem ser utilizadas em conjunto para indicar afinidade geográfica de uma amostra (Cordeaux *et al.*, 2007).



Figura 01. Estrutura dimérica geral de um elemento *Alu*. Retirado de Strachan & Read (1999).

Atualmente, os elementos *Alu* são classificados em pelo menos três grandes famílias, denominadas segundo o tempo de surgimento estimado: J para famílias antigas, S para intermediárias e Y para as mais novas (Batzer *et al.*, 1996). Os membros da família *AluY* são possíveis candidatos para investigar as origens dos haplótipos humanos ancestrais, grupos étnicos e associação a doenças (Kulski *et al.*, 2001). Eles surgiram na história evolutiva de primatas há aproximadamente 19 milhões de anos, e atualmente há seis subfamílias descritas para *AluY* no genoma humano (Batzer *et al.*, 1990). Dentre elas, *AluYb8*, a subfamília em que se encontram as inserções estudadas neste trabalho (Kulski *et al.*, 2001), é encontrada exclusivamente em humanos e provavelmente ainda apresenta mobilidade (Zietkewicz *et al.*, 1988).

STRs (Short Tandem Repeats)

O genoma humano é rico em trechos de DNA repetitivo, entre eles os VNTR (*variable number of tandem repeats*), que são separados em classes de acordo com seu tamanho: satélites, minissatélites e microsatélites (Figura 02). Os VNTR são sequências de repetições de um determinado motivo (de 2 a 32 pb), seguidos em tandem. Os microsatélites (ou STRs – do inglês *short tandem repeats*) possuem motivos básicos de um a seis pares de base (pb), são abundantes em todo o genoma (de Koenig *et al.*, 2011) e juntamente com outros elementos repetitivos, compreendem cerca de dois terços do genoma humano total. A

herança dessas regiões genômicas segue o padrão mendeliano e apresentam alto polimorfismo, o que os torna bons marcadores genéticos (Koreth *et al.*, 1996; Foissac *et al.*, 1997; Oliveira, 2007).

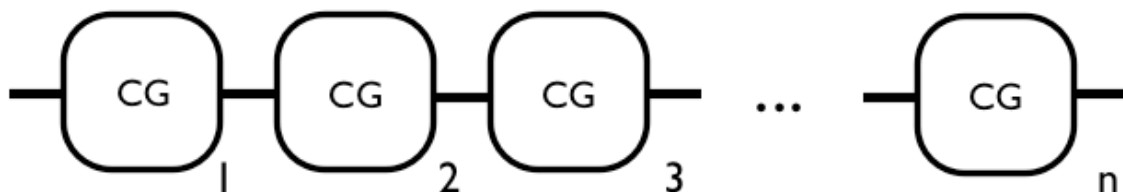


Figura 02. Estrutura básica de um VNTR, sendo este um STR com motivo de dois pares de base, CG.

Na busca de um melhor entendimento da dinâmica de populações, a análise da variabilidade genética é central. E ainda, a utilização de distintas classes de marcadores assim como diferentes análises dos dados podem mostrar resultados divergentes para uma mesma população. Dessa forma, torna-se interessante estudar diferentes tipos de marcadores para que se tenha uma visão mais ampla dos processos, em especial quando na presença de mistura genética e diferenciação populacional ocorridos na formação da população de interesse (Barcelos, 2006). Assim, este trabalho se propõe a descrever o comportamento de regiões genômicas no Complexo de Histocompatibilidade Principal Humano (MHC, do inglês, *Major Histocompatibility Complex*) a partir da análise de marcadores genéticos, de diferentes categorias, situados nessa região.

Objeto de estudo

O complexo de histocompatibilidade principal humano é composto por aproximadamente 224 genes que se estendem por 3,6 megabases (Mb) do cromossomo 6 (6p21.3) (Figura 03A). Alguns dos genes desse complexo codificam glicoproteínas

transmembrânicas que se associam a fragmentos peptídicos importantes na atuação do sistema imune. Estes são apresentados às células do sistema imunitário, principalmente linfócitos T, desempenhando papel importante na resposta imunitária adaptativa, assim como na susceptibilidade a diversas doenças auto-imunes e degenerativas (Klein & Sato, 2000b; a).

Historicamente, os genes do MHC foram divididos em três classes distintas de acordo com a funcionalidade de suas proteínas. Na região telomérica encontram-se os genes de classe I e na região centromérica os genes de classe II. Na região central, entre as classes I e II, está a região denominada MHC central ou de classe III, que se estende por aproximadamente 700kb e contém mais de 60 dos 224 *loci* identificados no complexo (Figura 03B).

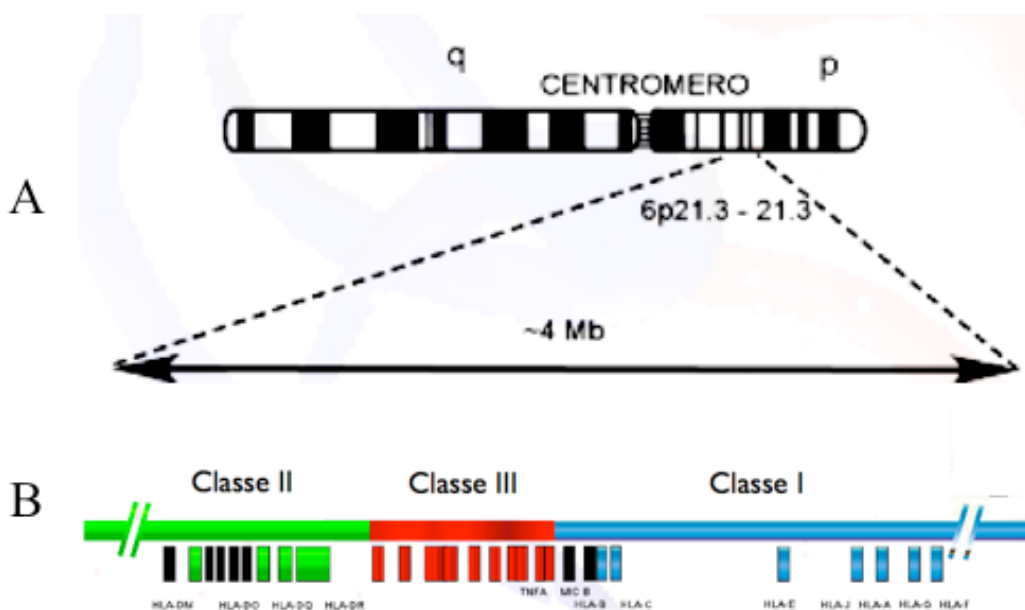


Figura 03. A. Cromossomo 6 humano e a localização do Complexo de Histocompatibilidade Principal; B. Divisão das classes no MHC.

Os genes de classe I dividem-se em dois subgrupos: os *loci* de classe Ia (ou clássicos – *HLA-A*, *-B* e *-C*), expressos constitutivamente e caracterizados por uma alta

variabilidade; e os *loci* de classe Ib (ou não-clássicos – *HLA-G*, *-E*, *-F*), mais conservados e com expressão restrita a poucos tecidos (Castelli, 2008). Dentre os genes de classe III encontra-se o *locus* TNF, que codifica o fator de necrose tumoral (*TNFA*) e a linfotoxina (*TNFB*), além de genes codificadores das proteínas do sistema complemento, dentre outros (Udalova et al., 1993) (Figura 03B).

Os genes do MHC de classe I e II regulam o equilíbrio entre respostas agressivas a um patógeno invasor enquanto minimizam a destruição de tecidos do hospedeiro (Trowsdale, 2011). Por esse motivo, o MHC tem sido relacionado a diversas doenças inflamatórias e autoimunes, principalmente em estudos de associação em genoma total (GWAS – *Genome-Wide Association Studies*), além de ser uma das regiões mais estudadas do genoma (Horton et al., 2008).

Embora alvo de muito estudo, ainda não é sabido o mecanismo de interação dos genes do MHC e sua relação com diversas doenças, tanto autoimunes quanto inflamatórias. Parte desse desconhecimento deve-se ao forte desequilíbrio de ligação que existe entre genes e sequências não codificadoras ao longo de todo o Complexo de Histocompatibilidade Principal Humano (Trowsdale, 2011).

O MHC é considerado a região mais variável do genoma humano, sendo essa variabilidade caracterizada predominantemente por substituições de nucleotídeo único (SNP), microssatélites e inserção/deleção (indels) na forma de retroelementos, tais como as inserções *Alu* (Jongeneel et al., 1991; Walsh et al., 2003; Malkki et al., 2005; Kulski & Dunn, 2005; Deghaide et al., 2009; Yao et al., 2009). Essa variabilidade pode estar associada a diversas doenças, especialmente aquelas com forte componente imunológico, susceptibilidade a infecções, neoplasias e expressão diferencial dos genes do complexo (Deininger & Batzer, 1999; Deghaide et al., 2009).

Porém, essa variabilidade genética não se distribui uniformemente ao longo do Complexo, especialmente na região de classe I (Figura 04). Os *loci HLA-B* e *HLA-C* são caracterizados por uma alta variabilidade (2338 alelos distintos somente para o *locus HLA-B* – em <http://www.ebi.ac.uk/imgt/hla/stats.html>, acessado em 07/02/2012) e parecem sofrer forte pressão seletiva em favor dela (seleção balanceadora). Foram descritos dois marcadores do tipo inserção *Alu*, *AluYMICB* e *AluYTF*, que estão nas extremidades desta região altamente variável (Kulski & Dunn, 2005). Os *loci HLA-G*, *HLA-E* e *HLA-F* (Figura 03B) são caracterizados por uma baixa variabilidade quando comparados aos demais genes de classe I, e parecem sofrer ação de seleção purificadora - i.e., nota-se a preservação das sequências de DNA e proteínas, redução da variabilidade e uma taxa de mutações não-sinônima menor do que a esperada (Castro *et al.*, 2000a). Nesta região mais conservada foram descritos três marcadores genéticos do tipo inserção *Alu*, *AluYHJ*, *AluYHG* e *AluYHF* (Kulski & Dunn, 2005).

Curiosamente, ao centro desta região virtualmente invariável, encontra-se o *locus HLA-A*, considerado uma área de intensa variabilidade dentro de uma região conservada de classe I. Por fim, na região de classe III, considerada a região de maior densidade gênica do genoma (1 gene a cada 14.500 bases, segundo Shiina *et al.*, 2004), foram descritos cinco microssatélites (a-e) margeando o *locus TNF* (Fator de Necrose Tumoral, do inglês *Tumor Necrosis Factor*). Alguns desses microssatélites apresentam mais de sete alelos e são muito informativos para estudos populacionais (Jongeneel *et al.*, 1991).

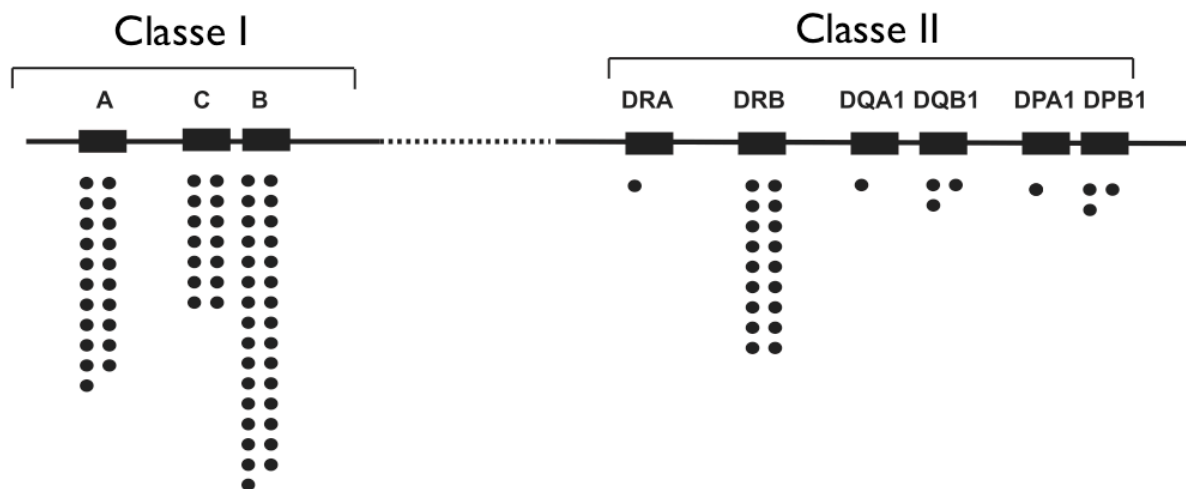


Figura 04. Representação do número de alelos descritos para genes do Complexo de Histocompatibilidade Principal Humano. Cada ponto representa até 50 alelos. Adaptado de Trowsdale (2011).

Os genes que se encontram na região TNF são responsáveis pela produção de mediadores inflamatórios durante a resposta imune aguda. O *TNFA*, por exemplo, é extremamente importante na contenção de infecções localizadas, e, caso seja produzido em infecções sistêmicas, causa a conhecida sepse, ou choque séptico (Janeway, 2007). Alguns trabalhos têm mostrado que determinados alelos dos microssatélites situados nos genes TNF estão relacionados à expressão diferenciada de *TNFA* e *TNFB* e que essa relação seria marcada justamente pela proximidade dos microssatélites com os genes em questão (Hajeer & Hutchinson, 2000). Derkx *et al.* (1995) mostraram que os alelos a2, a6 e a10 estão relacionados a uma produção mais baixa de *TNFA* que os alelos a4 e a11 em culturas de células sanguíneas estimuladas com endotoxinas. Por outro lado, Pociot *et al.* (1993) mostraram que os alelos a2 e c2 estão associados a uma alta produção de *TNFA* e os alelos a6 e c1 à baixa produção de *TNFA* em células estimuladas com lipopolissacarídeos ou fitohemaglutinina *in vitro*.

Em estudo com indivíduos que foram submetidos a transplante de coração, Turner *et al.* (1995) concluíram que o alelo d3 diminui a produção de citocinas em cultura de linfócitos estimulados com concavalina-A. E ainda, diversos trabalhos têm reportado associações dos genes do *TNF* e os microssatélites contidos na região dos genes *TNF* à progressão de doenças como AIDS (atuando como agente inflamatório na retinite causada por citomegalovírus; conferindo proteção ao desenvolvimento da síndrome de lipodistrofia), esclerose múltipla, diabetes melitos e alteração na expressão dos genes da região (McDonnell *et al.*, 1999; Hajeer & Hutchinson, 2000; Rassi *et al.*, 2006; Deghaide *et al.* 2009).

Diante da importância do MHC para a manutenção da saúde humana e do peculiar comportamento dessa região genômica, que abriga tanto sítios altamente conservados quanto sítios extremamente variáveis; além das características evolutivas dos marcadores selecionados (evolução rápida e evolução lenta), pergunta-se como esses marcadores se comportam em populações com histórico de formação e composição genética distintas. Buscando resposta a essa pergunta, este trabalho se propõe a reportar o comportamento desses marcadores em duas populações do Centro-Oeste brasileiro: Distrito Federal e Kalunga, as quais foram previamente analisadas para marcadores genéticos neutros (Barcelos, 2006; Barcelos *et al.*, 2006; Gontijo, 2008; Ribeiro, 2009).

Objetivos

O objetivo geral desse trabalho foi avaliar como marcadores genéticos de evolução rápida e de evolução lenta situados no MHC humano se comportam em populações com histórico de formação e composição genética distintos. Buscando atingir esse objetivo geral, foram objetivos específicos:

1. Analisar a distribuição alélica de marcadores genéticos de evolução lenta (inserções AluMICB, AluHJ, AluHG, AluTF, AluHF (região de classe I)) e de evolução rápida (microssatélites TNFa, TNFb, TNFc, TNFd e TNFe (região de classe III)) em uma população brasileira urbana miscigenada (Brasília-DF) e uma afroderivada rural (o remanescente de quilombo Kalunga-GO);
2. Analisar a estruturação genética dessas populações por meio de parâmetros como a ocorrência de desequilíbrio de ligação entre os *loci* analisados, a comparação das heterozigoses esperadas e observadas, deficiência e excesso de heterozigotos, índice de diferenciação populacional (F_{ST}) e análise de estruturação populacional;
3. Analisar a distribuição dos haplótipos gerados por essas duas classes de marcadores e do conjunto completo nas duas populações e realizar análises comparativas entre as duas populações em questão e destas com dados disponíveis na literatura;
4. Comparar os parâmetros de diversidade intra e interpopulacionais obtidos para essas duas populações com os dados resultantes da análise desses marcadores em outras populações ao redor do mundo;
5. Avaliar a ocorrência de desvios da neutralidade para os marcadores genéticos de evolução lenta, de evolução rápida e dos dois conjuntos associados.

Material e Métodos

Populações escolhidas para estudo

Foram selecionadas duas populações do Centro-Oeste brasileiro: Distrito Federal e Kalunga. Essas duas populações apresentam ancestralidade e modo de vida distintos e foram previamente analisadas para marcadores genéticos neutros (autossômicos, mitocondriais, do cromossomo X e cromossomo Y).

O povoamento do Distrito Federal ocorreu artificialmente, incentivado por benefícios governamentais para a construção da cidade de Brasília (1956 até 1960). A construção da Capital foi o principal atrativo para o processo de migração, originando-se principalmente das regiões Norte, Sul e Sudeste. Estudos anteriores mostraram que o Distrito Federal é representativo do Brasil como um todo por causa da intensa migração de pessoas de todo o país (Barcelos *et al.*, 2006).

Kalunga, por sua vez, é uma população remanescente de quilombo, rural e isolada. Segundo a definição do Governo Brasileiro, quilombos são grupos étnico-raciais, segundo critérios de auto-atribuição, com trajetória histórica própria, dotados de relações territoriais específicas, com presunção de ancestralidade negra relacionada com a resistência à opressão histórica sofrida (Ministério da Cultura, portaria 6/2004, artigo 2o, retirado de <http://www.cpisp.org.br/htm/leis/page.aspx?LeiID=141>). De acordo com o relato de moradores mais antigos, Kalunga foi fundada por escravos que fugiram do trabalho escravo nas minas de ouro no interior do Brasil, bem como pelos escravos que foram abandonados ao fim da exploração aurífera, no final do século XVIII. O quilombo ocupa uma área de aproximadamente 258.000 hectares na zona rural dos municípios de Cavalcante, Teresina e Monte Alegre, no nordeste do Estado de Goiás (Figura 05). Kalunga foi oficialmente

reconhecida como Remanescente de Quilombo em 2000 e é considerada patrimônio cultural do estado de Goiás (Ribeiro, 2009).



Figura 05. Mapa indicando a localização geográfica das duas populações estudadas. Alfinete vermelho: Brasília; Alfinete azul: Kalunga.

Ambas populações foram bastante estudadas e estão bem caracterizadas por marcadores genéticos autossômicos e uniparentais neutros. Os trabalhos envolvendo esses marcadores evidenciaram que as duas populações são, de fato, diferentes em termos de frequência alélica, composição e estruturação genética (Barcelos *et al.*, 2006; Barcelos, 2006; Godinho, 2008; Gontijo, 2008; Arcanjo *et al.*, 2009; Dalton *et al.*, 2009; Oliveira *et al.*, 2009; Ribeiro *et al.*, 2009; Ribeiro, 2009; Silva, 2010; Trindade-Filho, 2010).

Neste trabalho foram analisadas, ao todo, 165 amostras, das quais 105 amostras de indivíduos do Distrito Federal e 60 de Kalunga. Todos os indivíduos estudados apresentam índice de parentesco inferior a 1/16.

Aspectos éticos

A coleta e a utilização das amostras empregados no presente trabalho foram aprovadas pelo Comitê de Ética em Pesquisa da Faculdade de Saúde da Universidade de Brasília e pelo Conselho Nacional de Ética em Pesquisa. Os pareceres, de números CEP 151/07 (Kalunga); CEP 028/2001 e CONEP 1259/2001 (Brasília) podem ser visualizados no Anexo 1.

Material biológico

As amostras de sangue foram obtidas por punção venosa de voluntários, em tubos de coleta com sistema à vácuo e EDTA como anticoagulante. Os voluntários, todos maiores de 18 anos, assinaram um Termo de Consentimento Livre e Esclarecido que se encontram nos arquivos do Laboratório de Genética do Instituto de Biologia da Universidade de Brasília.

A coleta do sangue dos voluntários de Brasília se deu entre 2002 e 2008, e de Kalunga em 2001/2002 e 2007. As amostras de sangue coletadas fazem parte do Banco de Sangue do Laboratório de Genética do Instituto de Biologia da Universidade de Brasília.

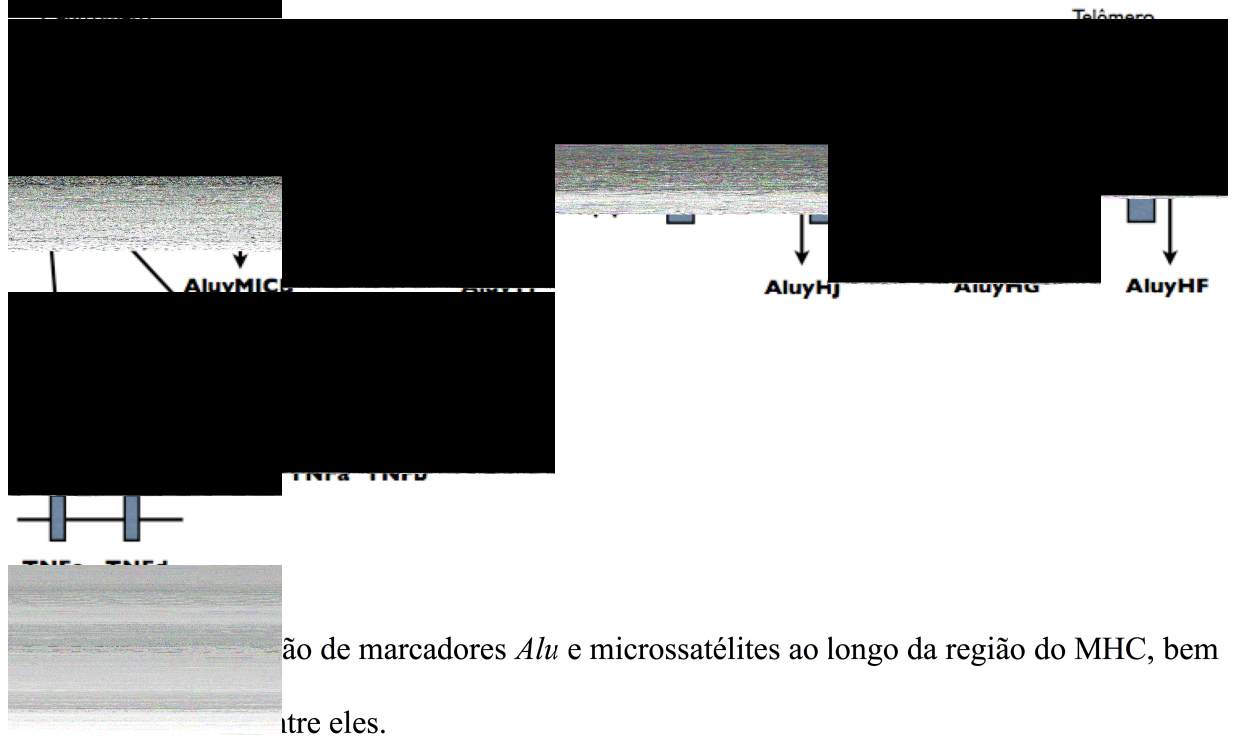
A extração de DNA foi feita preferencialmente da placa leucocitária por conter a maior proporção de células nucleadas. O DNA das amostras provenientes da população de Kalunga foi extraído com o kit comercial Illustra® da Amersham, enquanto o DNA proveniente das amostras da população de Brasília foi extraído pelo método de salting-out (Miller *et al.*, 1988, com modificações descritas em Godinho, 2008).

Marcadores moleculares

Como já comentado anteriormente, foram selecionados dois conjuntos de

evolução lenta
simos aos
HLA-G,
HLA-E e *HLA-F*. O conjunto de marcadores genéticos de evolução rápida foi representado por

localizados na região do gene de Fator de Necrose Tumoral: TNFa,
e TNFe, também referidos como TNFa-e (Figura 06).



Amplificação dos marcadores

Os fragmentos correspondentes às inserções *Alu* foram amplificados utilizando a técnica da Reação em Cadeia da Polimerase (PCR - do inglês *Polymerase Chain Reaction*) realizada em um termociclador *Veriti*TM (Applied Biosystems). Os iniciadores empregados

foram obtidos na literatura (Tabela 01). A mistura de reagentes variou entre os marcadores (Tabela 02), assim como a temperatura de anelamento (Tabela 01). O ciclo de PCR básico foi de 10 minutos de desnaturação inicial a 95°C, desnaturação por 30 segundos a 95°C, anelamento dos iniciadores em temperaturas variadas por tempos variáveis, extensão a 72°C por 45 segundos e extensão final por 72°C por 10 minutos.

Tabela 01. Iniciadores, alelos e referências para os marcadores *Alu* utilizados no presente trabalho.

Marcador	Iniciadores	Alelos	Referência	Temperatura e tempo de anelamento dos iniciadores
AluyHF	F5' GCCTCATGGCCTGAATCTGCCAGTGTCTT 3' R5' GTAACCTGACCTGCCCTCTATAGTATAGTCT 3'	457pb 612pb	Dunn <i>et al.</i> , 2002	62°C - 30''
AluyHG	F5' CAGGACAACCAGTAAAGATGCTGG 3' R5' GCTTCAGTTAACATGCAAGTTTATGCC 3'	218pb 540pb	Kulski <i>et al.</i> , 2001	64°C - 30''
AluyHJ	F 5' AAGAACCATAACTCACTTG 3' R 5' TGTGTCCAGGTTAAACTTCAG 3'	162pb 494pb	Dunn <i>et al.</i> , 2002	58°C - 45''
AluyMICB	F 5' GCCTTCCAATGCCATTCACAG 3' R 5' CTCAGCCCTGCTTTCCCATCT 3'	503pb 665pb	Kulski <i>et al.</i> , 2002	65°C - 45''
AluyTF	F 5' GTGCCTGGTAAAAATTTAAAGAGCTGTA 3' R 5' TGCACCCGGCCTAAAACCACTGGTT 3'	426pb 711pb	Dunn <i>et al.</i> , 2003	65°C - 45'' 63°C - 45''

Tabela 02. Concentração dos reagentes nas reações de PCR para os marcadores *Alu* estudados.

Reagente	AluyHF	AluyHG	AluyHJ	AluyMICB	AluyTF
Tampão	1,0X	0,8X	0,8X	0,8X	0,8X
dNTP	0,8mM	0,6mM	0,8mM	0,8mM	0,6mM
Iniciador	0,4 μ M	0,3 μ M	0,4 μ M	0,6 μ M	0,4 μ M
MgCl ₂	2,0mM	2,5mM	2,5mM	2,5mM	2,0mM
<i>Taq</i>	1,5U	1,5U	1,5U	1,5U	1,5U
DNA	>20ng	>20ng	>20ng	>20ng	>20ng

A amplificação dos microssatélites *TNF* foi feita também pela técnica da PCR, em um termociclador Veriti™. Os iniciadores usados foram retirados de Udalova et al. (1993) (Tabela 03 e Figura 07). Com exceção do TNFc, os microssatélites foram amplificados com iniciadores marcados com fluoróforo fabricados pela Applied Biosystems (Tabela 04). Os parâmetros da reação de PCR para todos os microssatélites foram: desnaturação inicial a 94°C por cinco minutos; 32 ciclos de desnaturação a 94°C por 30 segundos, anelamento dos iniciadores a 59°C por 90 segundos e extensão do fragmento a 72°C por 60 segundos; e um ciclo de extensão final de 72°C por 10 minutos.

Tabela 03. Iniciadores*, número de alelos, variação no tamanho dos alelos, motivo de repetição e tamanho de fragmentos obtidos** para os microssatélites TNF.

Microssatélite	Iniciadores	Número de Alelos	Motivo de Repetição	Tamanho de Fragmento
TNFa*	IR1 - F 5' GCACTCCAGCCTAGGCCACAGA 3' IR4 - R 5' CCTCTCTCCCCTGCAACACACA 3'	14	(GT) _n	99-125 pb
TNFb*	IR2 - F 5' GCCTCTACATTTTCATCCAGCCACA 3' IR5 - R 5' GTGTGTGTTGCAGGGGAGAGAG 3'	7	(GA) _n	125-131 pb
TNFc*	IR6 - F 5' GGTTTCTCTGACTGCATCTTGTCC 3' IR7 - R 5' TCATGGGGAGAACCTGCAGAGAA 3'	2	(GA) _n	159-161 pb
TNFd*	IR11 - F 5' AGATCCTTCCCTGTGAGTTCTGCT 3' IR12 - R 5' GTGCCTGGTTCTGGAGCCTCTC 3'	7	(GA) _n	124-136 pb
TNFe*	IR13 - F 5' CATAGTGGGACTCTGTCTCCAAAG 3' IR14 - R 5' TGAGACAGAGGATAGGAGAGACAG 3'	3	(GA) _n	98-102 pb

*Udalova et al., 1993; ** Hajeer & Hutchinson, 2000.

Tabela 04. Sequências selecionada para marcação com fluorescência, tamanho do fragmento gerado após amplificação e respectivo fluoróforo.

Microssatélite	Sequência Marcada	Tamanho do Fragmento	Fluoróforo
TNFa	IR4 - R 5' CCTCTCTCCCCTGCAACACACA 3'	99 a 125 pb	VIC
TNFb	IR5 - R 5' GTGTGTGTTGCAGGGGAGAGAG 3'	125 a 131 pb	NED
TNFd	IR 11 - F 5' AGATCCTTCCCTGTGAGTTCTGCT 3'	124 a 136 pb	6-FAM
TNFe	IR14 - R 5' TGAGACAGAGGATAGGAGAGACAG 3'	98 a 102 pb	NED

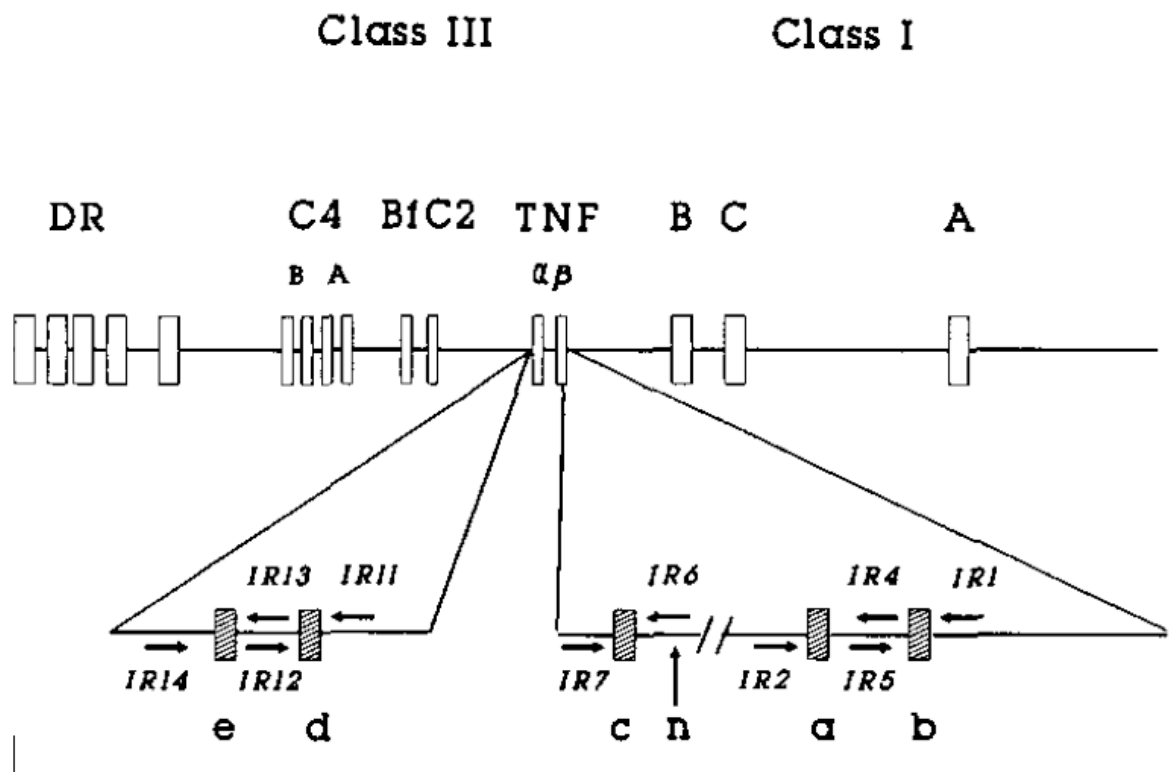


Figura 07. Localização dos microssatélites *TNFA*-e na região dos genes *TNFA* e *TNFB*. IR são os iniciadores para cada microssatélite, como mencionado na Tabela 04. Figura adaptada de Udalova *et al.* (1993).

Genotipagem dos marcadores

Inserções *Alu*

O tamanho do fragmento amplificado convencional de uma inserção *Alu* é de 300pb, aproximadamente. No entanto, *Alu*yMICB e *Alu*yHF têm inserções de 162pb e 155pb, respectivamente. A saber: ambos os alelos de *Alu*yMICB contém um elemento *Alu*Ya completo, com 96,8% de similaridade com os elementos *Alu* da família *Alu*Y. A diferença entre os alelos dá-se ao fato de que um deles contém uma sequência parcial de um segundo elemento *Alu* de 162pb, gerado por duplicação ou inserção (Kulski *et al.*, 2002). A sequência de inserção do elemento *Alu*yHF é um fragmento parcial de elemento *Alu* que tem 91% de

identidade com a sequência consenso da família AluYb8 e 95,7% de identidade com a sequência consenso da família AluYa5/AluYa8 (Dunn *et al.*, 2002).

Após a amplificação, os fragmentos foram separados por eletroforese vertical em gel de poliacrilamida a 6% não desnaturante e corados com nitrato de prata. Em todos os géis foi utilizado marcador de peso molecular 100pb ou 1kb plus (Invitrogen). Por conveniência, o alelo com inserção foi denominado Alelo 1 e o alelo com ausência de inserção, Alelo 2.

Microsatélites TNF

Os dados dos marcadores TNFc, TNFd e TNFe para a população de Brasília já estavam disponíveis no banco de dados do Laboratório de Genética. Esses dados foram gerados por Silva (2010) durante dissertação de mestrado.

A detecção dos microsatélites TNFa, TNFb, TNFc, TNFd e TNFe em Kalunga e TNFa e TNFb em Brasília também foi feita inicialmente em gel desnaturante de poliacrilamida 12%, corados por impregnação pela prata, e parte do processo foi realizado no Laboratório de Imunologia Molecular do Hospital das Clínicas de Ribeirão Preto, da Universidade de São Paulo. A genotipagem dos TNFs foi feita de acordo com Deghaide *et al.* (2009). Os parâmetros de eletroforese para a genotipagem dos TNFs são muito específicos em decorrência da diferença de tamanho entre os alelos, que é de um (TNFb) ou dois (demais TNFs) pares de base (Tabela 03). Nesse caso, a amperagem da fonte foi fixada em 20mA, enquanto os parâmetros como voltagem e potência eram liberados, com o intuito de que a fonte os ajustasse para manter a amperagem fixa. O tempo da eletroforese foi de duas horas e meia para TNFa, TNFc, TNFd e TNFe, e de cinco horas (TNFb).

Na busca por alternativas e melhoria da genotipagem desses marcadores, foi padronizada na UnB a genotipagem dos microsatélites TNFa, TNFb, TNFd e TNFe por eletroforese capilar em um sequenciador automático ABI 3130XL, da Applied Biosystems

(LIFE Technologies). Para o TNFc, que apresenta apenas dois alelos, a genotipagem por eletroforese convencional foi mantida.

Eletroforese Capilar

Segundo o fornecedor, o sucesso do sistema de detecção de alelos de microssatélites da Applied Biosystems reside na combinação de fluorescências diferenciais para diferentes tamanhos de fragmentos. Inclusive, essa combinação, associada ao *software* GeneMapper™, reduz o tempo e custo dos experimentos com microssatélites em até 33%.

Considerando os tamanhos dos fragmentos, foi pré-estabelecida uma genotipagem multiplex, isto é, a eletroforese capilar foi realizada simultaneamente para mais de um marcador. Como vários fragmentos se sobrepõem (Figura 07), foi necessário montar dois sistemas de eletroforese multiplex diferentes: uma para genotipar os microssatélites TNFa e TNFb, e outra para TNFd e TNFe.

Os iniciadores foram marcados em apenas uma das extremidades, devido à sobreposição de sequências, utilizando três fluoróforos, 6-FAM (azul), VIC (verde) e NED (amarelo). Os iniciadores e a marcação foram feitas pela Applied Biosystems conforme explicitados na Tabela 04.

Após a amplificação, os produtos resultantes das reações com cada par de iniciador foi diluído a uma proporção 1/10 e os multiplex foram montados. A cada poço da placa de corrida foi adicionado 1µL do multiplex juntamente com 9µL de Hi-Di™ formamida e 0,0975µL do marcador molecular GeneLiz500(-250)™ (Applied Biosystems).

A eletroforese capilar foi feita no sequenciador automático ABI 3130XL, utilizando o polímero POP-7 e o programa padrão para microssatélites. Após a eletroforese, os arquivos .fsa correspondentes aos sinais de cada amostra foram submetidos à análise pelo software GeneMapper™ para o procedimento de identificação dos alelos (*allele calling*).

Identificação dos alelos (allele calling)

O procedimento de identificação dos alelos pode ser manual ou automática. No primeiro caso, o pesquisador especifica o tamanho de cada alelo que pode ser encontrado para determinado marcador. Muitas vezes, porém, os fragmentos podem não estar completamente amplificados (bandas *stutter*) e, ainda assim, emitirem o sinal fluorescente, que será detectado pelo sequenciador. Em um gel de poliacrilamida corado com nitrato de prata, boa parte desses fragmentos semi-amplificados se agrupa à banda de interesse e fica mascarada pelos fragmentos completamente amplificados, não interferindo tanto na genotipagem dos alelos. No caso do eletroferograma, entretanto, as bandas semi-amplificadas emitem sinal fluorescente significativo, o que pode causar confusão no momento de identificar cada alelo. Outro problema é que os picos do eletroferograma dificilmente conseguirão estimar um fragmento exatamente do tamanho esperado. Um bom exemplo está explícito na figura 09, onde pode-se ver que os tamanhos dos fragmentos para o microssatélite TNFa (parte superior esquerda) são 99,34pb e 114,62pb, respectivamente.

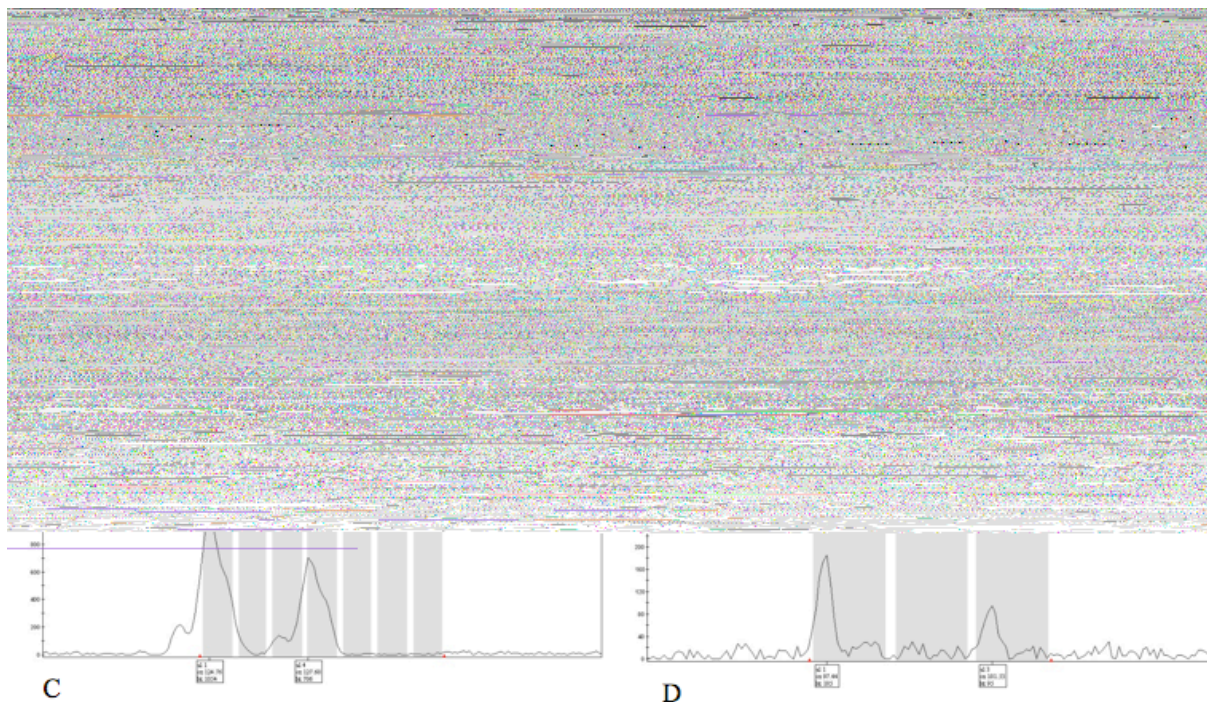


Figura 08. Exemplo de eletroferogramas obtidos para os marcadores TNFa (A), TNFb (B), TNFd (C) e TNFe (D). Os *bins* estão representados como faixas aczentadas, que denotam o campo de aceitação para cada alelo.

Neste trabalho foi feita a opção da identificação automática, na qual são pré-definidos os padrões de menor fragmento amplificado, maior fragmento amplificado e tamanho do motivo da repetição. A partir desses parâmetros o GeneMapper™ estima quais seriam os melhores intervalos de consideração para cada alelo. Esse intervalo é chamado de *bin* (Figura 08) e pode ser manualmente ajustado para evitar alelos sem identidade.

A nomenclatura dos alelos dos microssatélites TNF encontrada na literatura não é consensual. Dessa forma, é complexo determinar, por exemplo, se o alelo a2 deste trabalho apresenta o mesmo número de pares de base que o alelo a2 considerado em Gallagher *et al.* (1997), Marsh *et al.* (2003) ou vários outros trabalhos que não indicam o tamanho do fragmento relativo ao número do alelo. Portanto, as comparações que constam aqui se referem exclusivamente a trabalhos que indicam o número de pares de base para cada alelo com

relação aos marcadores TNFa, TNFb e TNFd. Como os TNFc e TNFe têm pequeno número de alelos (dois e três, respectivamente), os dados que constam na literatura foram utilizados, exceto quanto o número de alelos encontrados supera ou é inferior a três (salvo se indicado o nº de pares de base).

Análises estatísticas

Os testes estatísticos foram feitos em duas etapas, dividindo o conjunto amostral em três conjuntos menores: todos os genótipos para os marcadores de inserção *Alu*, todos os genótipos para os marcadores microssatélites, e o conjunto total de dados. Na primeira etapa os dados de cada marcador foram analisados individualmente e em conjunto, e na segunda etapa foram inferidos haplótipos para cada conjunto de marcadores e para os dois conjuntos.

As análises estatísticas escolhidas para o refinamento dos dados foram feitas sob a ótica da genética de populações, abordando as seguintes estimativas: frequências alélicas, heterozigose observada e esperada, aderência ao equilíbrio de Hardy-Weinberg (Guo & Thompson, 1982), desequilíbrio de ligação entre os marcadores, diferenciação populacional e diversidade alélica. Os gráficos presentes na dissertação foram feitos com o *software* Arlequin 3.5 (Excoffier & Lischer, 2010), através de *scripts* do pacote estatístico R fornecidos com o programa.

As frequências alélicas para cada um dos marcadores selecionados foram estimadas por contagem direta e submetidas à análise pelo software PyPop 0.7.0 (Lancaster *et al.*, 2003). Este software foi utilizado também para avaliar o comportamento de cada locus com relação ao Teste de Neutralidade de Ewens-Watterson (Watterson, 1978).

A inferência de haplótipos é uma estratégia muito utilizada em substituição a análises familiares, que são muito mais dispendiosas em termos tanto de consumo quanto de tempo. Para tanto são utilizados programas computacionais desenvolvidos para esse fim,

sendo os principais o PHASE v2.1 (Stephens *et al.*, 2001), EM e ELB, esse dois últimos são parte do software Arlequin 3.5 (Excoffier & Lischer, 2010). Porém, os dados gerados por esses programas apresentam falhas. Por isso, antes de estimar a fase gamética dos genótipos, foi feita uma comparação dos dados gerados por esses três programas.

Considerando que a região abarcada pelas inserções *Alu* aqui analisadas é grande, isto é, a distância entre os dois marcadores mais distantes, AluMICB – AluHF, é de 1,6Mb (figura 06). A abordagem utilizada foi a mesma de Castelli *et al.* (2010) na reconstrução de haplótipos de *HLA-A*, *-B*, *-DRB1* e *-DQB1*. Essa análise foi realizada empregando 30 indivíduos (trios de pai/mãe/filho) genotipados para os marcadores do tipo *Alu*. Os parâmetros *default* dos três softwares foram alterados segundo sugerido por Castelli *et al.* (2010): PHASE – 5000 iterações, intervalo de rarefação igual a 1, valor de *burn-in* igual a 100 e Δ -*value* igual a 1; EM – *e-value* igual a $1e^{-7}$, número máximo de 5000 iterações, 100 pontos de início e desvio-padrão a 50; ELB – α -*value* igual a 0,01, *e-value* igual a 0,01, intervalo de amostragem igual a 500 e 100.000 passos de *burn-in*. Além disso, foram executadas 10 corridas independentes para cada conjunto de marcadores, buscando obter as melhores estimativas.

Após o final do procedimento, os haplótipos estimados pelos *softwares* foram comparados com os determinados pela análise laboratorial e a porcentagem de acerto de cada algoritmo foi estimada. Como será explicitado em resultados, o melhor desempenho foi observado com o programa PHASE v2.1 (Stephens *et al.*, 2001). Dessa forma, a inferência de haplótipos foi realizada com o programa PHASE v2.1 (Stephens *et al.*, 2001) para os três conjuntos de marcadores (*Alu*, microssatélites e todos juntos) em 10 corridas independentes, com os mesmos parâmetros supracitados.

Utilizando o banco de dados na forma de haplótipo, foram estimadas as diferenças haplotípicas inter e intra-populacional (Slatkin, 1995) e distância baseada na média de diferenças alélicas de Nei (Nei & Li, 1979) utilizando o programa Arlequin 3.5 (Excoffier &

Lischer, 2010). Com objetivo de comparar com outras populações, foram utilizados dados dos microssatélites *TNF* retirados de McDonnell *et al.*(1999), Kim *et al.*(2000), Peña *et al.*(2002), Grubic *et al.*(2006), dos Santos (2009) e Sens-Abuázar *et al.*(2009). Os dados utilizados para as estimativas de diferenciação populacional relativas a outras populações já estudadas para as inserções *Alu* foram retiradas de García-Obregón *et al.* (2011).

Resultados e Discussão

Microsatélites TNF

A população de Brasília apresentou aderência ao Equilíbrio de Hardy-Weinberg (EHW) para a maioria dos marcadores estudados, exceto TNFa e TNFd (Tabela 05). Após utilizar a Correção de Bonferroni, ambos marcadores continuaram fora do EHW na população em questão. Esse resultado é diferente do encontrado para esses marcadores em outras populações, o que pode indicar problemas na genotipagem desses marcadores em Brasília e Kalunga. Os testes de deficiência (HD) e excesso (HE) de heterozigotos indicaram que, na população de Brasília, o *locus* TNFa se encontra com déficit de heterozigotos (p-valor: 0,0000+-0,0000), enquanto o *locus* TNFd se encontra com excesso (p-valor: 0,0355+-0,0040). Considerando todos os marcadores em conjunto, entretanto, Brasília apresentou um déficit global de heterozigotos (p-valor: 0,0001+-0,0001). Tais desvios são também observáveis pela análise da heterozigose, que mostra uma diferença entre o valor de heterozigose observada e esperada para os marcadores TNFa e d.

A população de Kalunga, por sua vez, não mostrou aderência ao EHW para os marcadores TNFa, TNFc, TNFd e TNFe. Porém, após aplicar a Correção de Bonferroni (p-valor menor que 0,005), apenas TNFd se manteve fora do EHW. Os testes de deficiência e excesso de heterozigoto na população de Kalunga indicaram que os marcadores TNFa, TNFc e TNFe apresentaram deficiência de heterozigotos, como pode ser visualizado na tabela 05. Apesar do marcador TNFd estar fora do EHW nessa população, os testes de deficiência e excesso de heterozigotos mostraram-se não significativos, o que pode refletir a ação de algum mecanismo evolutivo nesse *locus* específico. Com relação a estimativa de heterozigose, os desvios são observados também para os mesmos marcadores.

Tabela 05. Frequência dos alelos dos microssatélites TNF, teste de aderência ao Equilíbrio de Hardy-Weinberg, heterozigose observada e esperada para cada *locus* e teste de deficiência e excesso de heterozigotos em Brasília e Kalunga.

Alelo	Brasília					Kalunga				
	Tnf _a	Tnf _b	Tnf _c	Tnf _d	Tnf _e	Tnf _a	Tnf _b	Tnf _c	Tnf _d	Tnf _e
n	103	100	105	105	105	60	60	60	59	56
0	0,019	-	-	-	-	0,117	-	-	-	-
1	0,024	0,138	0,905	0,048	0,095	0,042	0,167	0,442	0,200	0,183
2	0,157	0,038	0,095	0,167	0,033	0,208	0,083	0,558	0,383	0,075
3	0,171	0,190	-	0,300	0,871	0,125	0,108	-	0,175	0,675
4	0,219	0,310	-	0,343	-	0,042	0,142	-	0,075	-
5	0,067	0,243	-	0,095	-	0,017	0,233	-	0,033	-
6	0,067	0,024	-	0,043	-	0,117	0,100	-	0,117	-
7	0,095	0,010	-	0,005	-	0,250	0,167	-	n/o	-
8	0,062	-	-	-	-	0,008	-	-	-	-
9	0,019	-	-	-	-	0,050	-	-	-	-
10	0,048	-	-	-	-	0,008	-	-	-	-
11	0,014	-	-	-	-	0,017	-	-	-	-
12	n/o	-	-	-	-	n/o	-	-	-	-
13	0,019	-	-	-	-	n/o	-	-	-	-

Alelo	Brasilia					Kalunga				
	Tnf _a	Tnf _b	Tnf _c	Tnf _d	Tnf _e	Tnf _a	Tnf _b	Tnf _c	Tnf _d	Tnf _e
Ho	0,553	0,770	0,171	0,838	0,238	0,683	0,867	0,350	0,661	0,321
He	0,871	0,770	0,173	0,755	0,232	0,852	0,848	0,497	0,760	0,436
p _{HW}	0,000	0,206	1,000	0,000	0,264	0,007	0,303	0,034	0,000	0,017
d.p.	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
p _{HD}	0,000	0,692	0,632	0,943	0,723	0,001	0,607	0,019	0,084	0,003
d.p.	0,000	0,018	-	0,008	-	0,001	0,013	-	0,005	-
p _{HE}	1,000	0,270	0,775	0,051	0,323	0,994	0,396	0,995	0,909	0,998
d.p.	0,000	0,022	-	0,010	-	0,004	0,013	-	0,008	-

n: número de amostras analisadas; Ho: heterozigose observada; He: heterozigose esperada; p: p-valor do teste de Hardy-Weinberg; d.p.: desvio-padrão de p; HD: deficiência de heterozigotos; HE: excesso de heterozigotos; n/o: não-observado.

Como já comentado em Material e Métodos, há uma escassez de dados na literatura comparáveis aos dados aqui apresentados, em especial devido a diferenças na nomenclatura e na forma de genotipagem dos marcadores. Diante disso, torna-se necessário a comparação com populações geneticamente distantes, visando uma avaliação do comportamento desses marcadores nas populações aqui avaliadas.

Com relação ao número de alelos para os microsatélites TNF, a maior diversidade alélica foi encontrado em Brasília, pois nessa população somente o alelo a12 não foi observado. Em Kalunga não foram encontrados os alelos a12 e a13 o alelo d7. O alelo a12 também não foi encontrado na população Basca (Peña *et al.*, 2002) e foi o menos frequente em Curitiba (Sans-Abuázar *et al.*, 2009) e em indivíduos afro-descendentes do Paraná (dos Santos, 2009).

Ainda, para o microsatélite TNFa, os alelos mais frequentes e os mais raros em Brasília e Kalunga diferiram. Em Brasília o alelo mais frequente e o mais raro foram, respectivamente, a4 (107pb) e a11 (121pb), enquanto que em Kalunga foram o a7(113pb) e os a8 (115pb) e a10 (119pb), respectivamente. Observa-se ainda que diferentes populações apresentam diferentes alelos mais comuns e raros para este marcador. Por exemplo, em Curitiba, os alelos mais frequentes foram a1 (25%) e a5 (16,1%) e os menos frequentes foram a7 (0,7%) e a12 (0,4%) (Sans-Abuázar *et al.*, 2009), enquanto em afro-descendentes do Paraná os alelos mais frequentes foram a5 (20,9%) e a1 (18,9%) (dos Santos, 2009).

Levando em consideração, agora, o microsatélite TNFb, os alelos mais comuns diferiram comparando Brasília e Kalunga - b4 e b5, respectivamente, assim como os mais raros - b7 e b2. Os alelos do TNFc apresentaram frequências muito distintas entre as duas populações, assim como o TNFd, sendo que o alelo mais comum em Brasília, o d4, é um dos mais raros em Kalunga. O marcador TNFe foi o que apresentou maiores similaridades quanto às frequências alélicas nas duas populações analisadas.

Na figura 09 está apresentada a avaliação de diferenciação genética populacional entre Brasília, Kalunga e outras populações mundiais com relação ao microssatélite TNFa. Esse marcador é o que apresenta o maior volume de dados na literatura. Observa-se que Brasília apresentou pouca diferenciação com relação à Kalunga e mostrou diferenciação moderada com relação às outras populações analisadas, em especial Coreia e Irlanda. Kalunga, em geral, mostrou maior diferenciação populacional do que Brasília com relação às mesmas populações analisadas, com exceção dos Bascos.

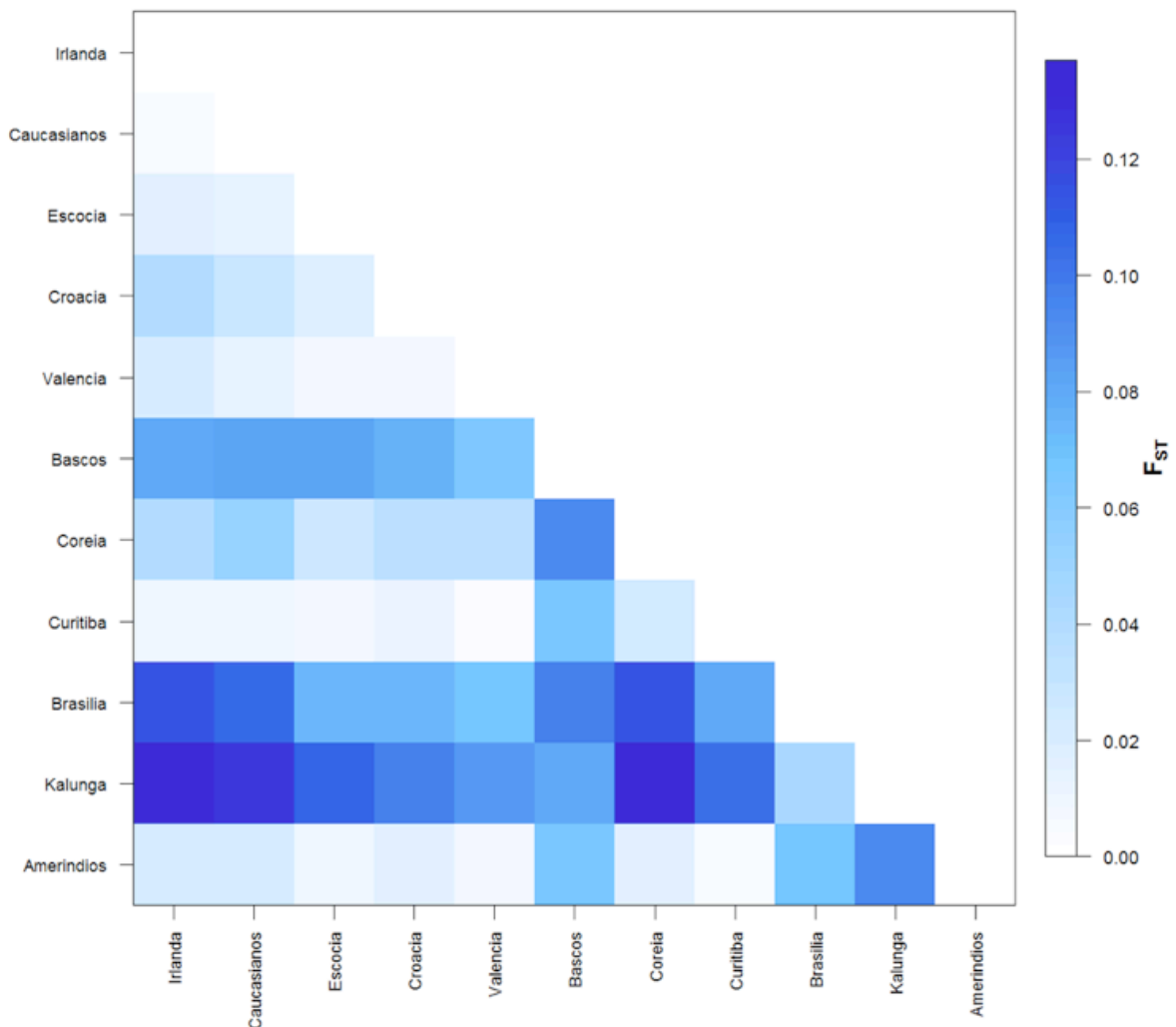


Figura 09. Representação gráfica da diferenciação populacional (F_{ST}) entre Brasília e Kalunga com relação a outros grupos populacionais - Irlanda, Caucasianos, Escócia, Croácia, Valência, Bascos, Coreia, Curitiba e Ameríndios, com relação ao locus TNFa.

Considerando o marcador TNFb (Figura 10), Brasília apresentou pequena diferenciação populacional com relação à população da Escócia e de Kalunga e moderada com relação às outras populações estudadas. Kalunga, por sua vez, apresentou pequena diferenciação populacional com relação a todas as populações analisadas, sendo os maiores índices observados na comparação Coreia - Brasília.

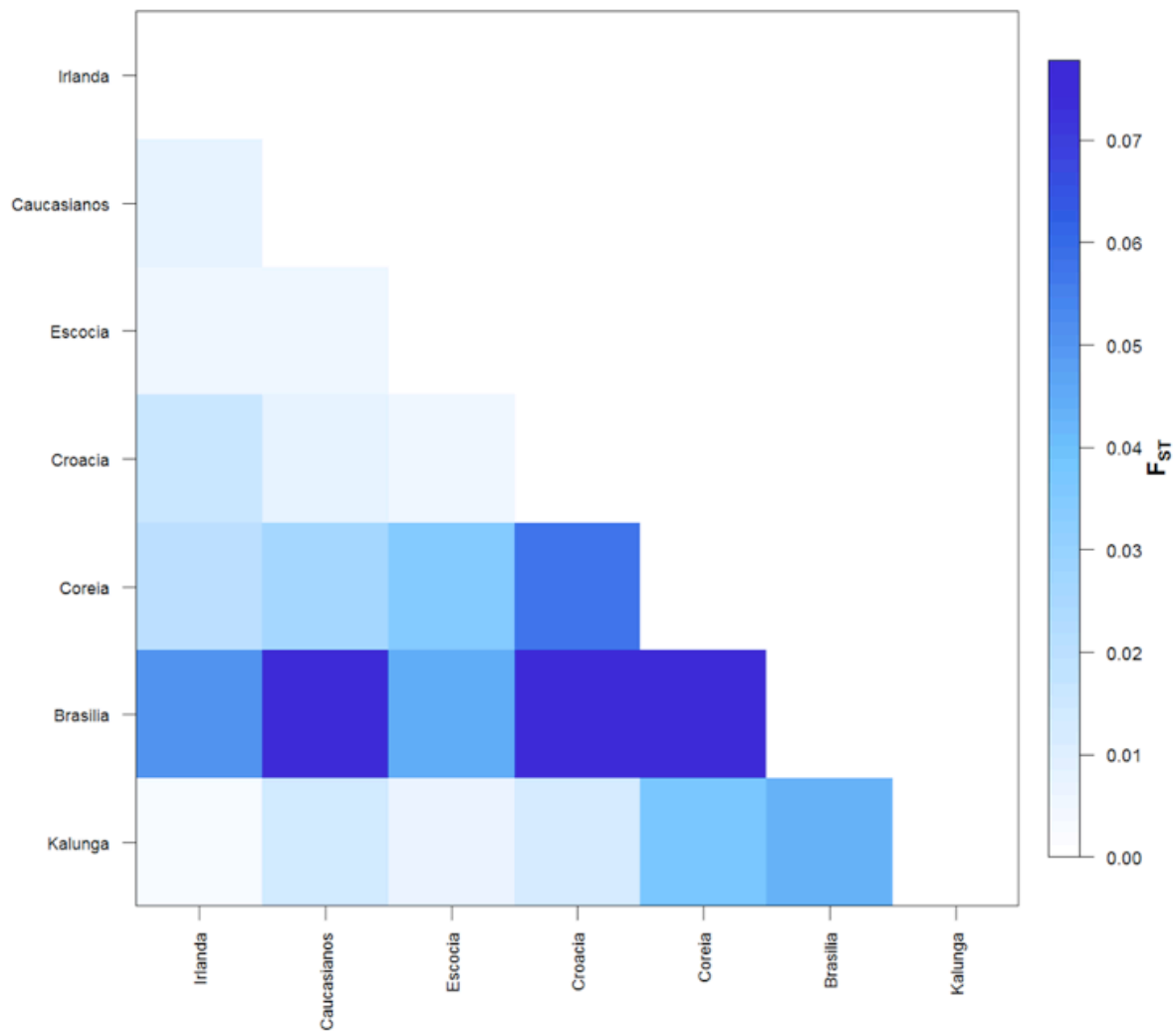


Figura 10. Representação gráfica da diferenciação populacional (F_{ST}) Brasília e Kalunga, com relação a outros grupos populacionais - Irlanda, Caucasianos, Escócia, Croácia e Coreia.

O gráfico gerado pela análise das frequências alélicas para o microssatélite TNFd (figura 11) mostra que Brasília apresentou pequena diferenciação populacional para a maioria

das populações analisadas. Em contraste, o par Kalunga e Croácia apresentou grande diferenciação populacional em relação aos outros pares de populações analisadas. Brasília apresentou maior diferenciação também em relação à população da Croácia.

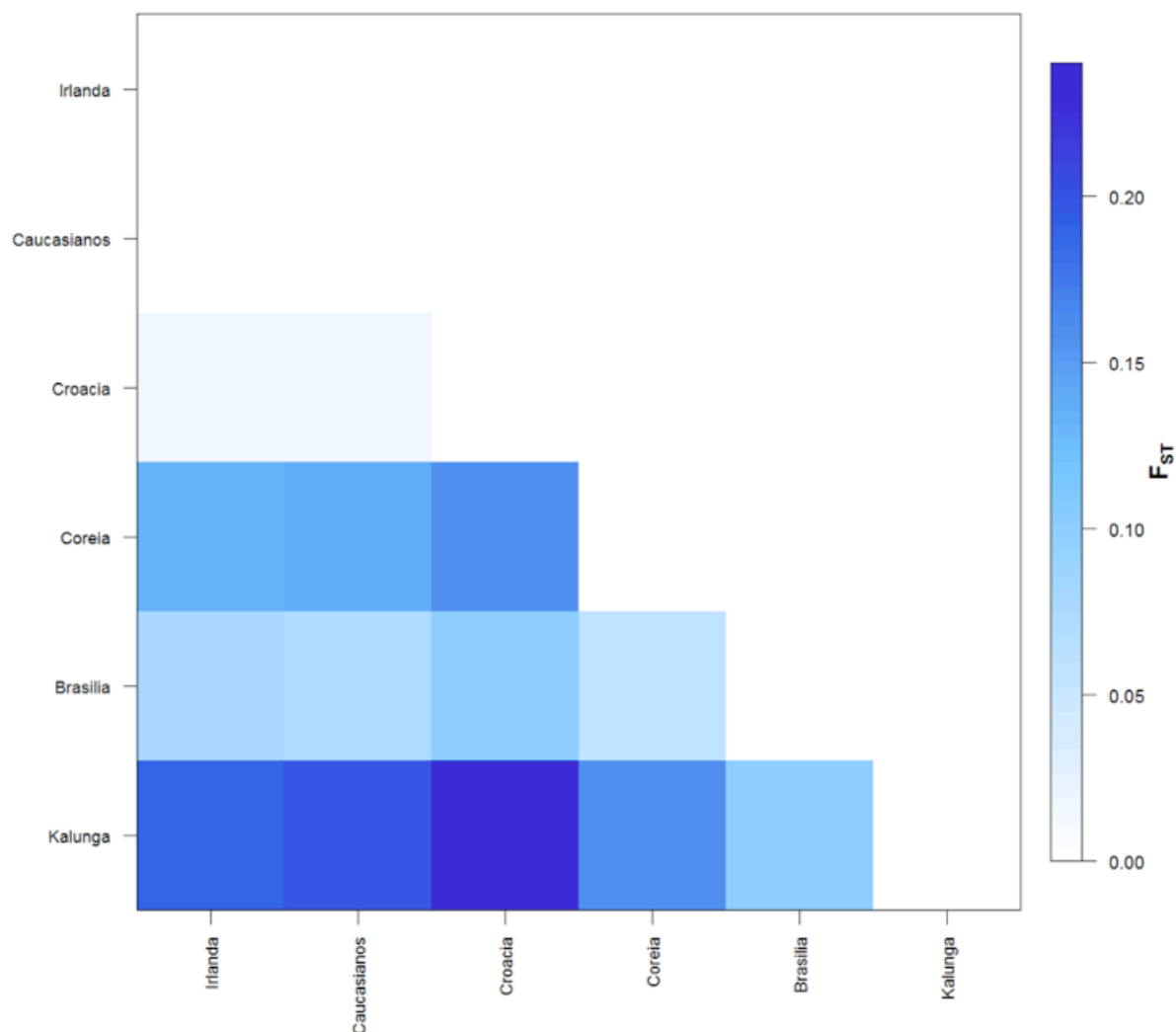


Figura 11. Representação gráfica das análises realizadas para diferenciação populacional baseadas em F_{ST} para Brasília, Kalunga, Irlanda, Caucasianos, Croácia e Coréia.

Inserções Alu

Todos os marcadores *Alu* aqui analisados apresentaram distribuição genotípica compatível com o EHW (tabela 06). Apesar disso, a população de Kalunga apresentou excesso de heterozigotos para o marcador AluHG (p-valor: 0,0467+-0,0020). Este fato se

reflete no pequeno valor de p para o teste exato de Equilíbrio de Hardy-Weinberg (p-valor: 0,0633+-0,0002) muito próximo do valor pré-estabelecido para a significância do teste ($p = 0,05$). Considerando o conjunto total de marcadores, Kalunga apresentou déficit de heterozigotos, com p-valor = 0,0010+-0,0004. Os maiores valores de heterozigose observada foram encontrados para o marcador AluHG em ambas populações.

A inserção AluHG está muito próxima do gene *HLA-G*. Recentemente, estudos têm demonstrado que o gene *HLA-G*, em especial suas regiões 5'URR (*upstream regulatory region*) e 3'UTR (*untranslated terminal region*), não se comportam como regiões neutras do genoma (Tan *et al.*, 2009; Castelli *et al.*, 2011). A avaliação da região 3'UTR do gene *HLA-G* nas populações de Brasília e Kalunga detectou um forte indicativo de desvio da neutralidade, utilizando o Teste de Neutralidade de Ewens-Watterson (Toledo, 2011). Dessa forma, o excesso de heterozigotos encontrados para o marcador AluHG em Kalunga provavelmente é um reflexo da seleção que está atuando no gene *HLA-G*, em um processo conhecido como “efeito carona”, mesmo que esse marcador não tenha demonstrado desvio da neutralidade em ambas Brasília e Kalunga.

Com relação às frequências alélicas, observou-se grande similaridade entre as populações. O alelo ausência de inserção (alelo 2) é o mais frequente para todos os marcadores em ambas populações. O mesmo comportamento foi observado nas demais populações já analisadas para esse conjunto de marcadores (Dunn *et al.*, 2003; Kulski & Dunn, 2005; Tian *et al.*, 2008; Yao *et al.*, 2009; García-Obregón *et al.*, 2011), isto é, em todos os estudos, incluindo o nosso, o alelo mais frequente é a ausência de inserção, em frequências que variam de 57,8% (para AluMICB e AluTF em japoneses) até 96,7% (para AluTF e AluHG em africanos Sekele San). Em Brasília e Kalunga, os alelos mais frequentes ficaram entre 70% (AluHG em Kalunga) e 87,5% (AluTF também em Kalunga)(Tabela 06). Como a condição ancestral do *locus* é a ausência de inserção, esse resultado era esperado em todas as

populações já estudadas, principalmente pelo fato dessas inserções serem recentes no genoma humano.

Tabela 06. Frequência dos alelos das inserções *Alu*, teste de aderência ao Equilíbrio de Hardy-Weinberg, heterozigose observada e esperada para cada *locus* e teste de deficiência e excesso de heterozigotos em Brasília e Kalunga.

Alelo	Brasília					Kalunga				
	Alu MICB	Alu TF	Alu HJ	Alu HG	Alu HF	Alu MICB	Alu TF	Alu HJ	Alu HG	Alu HF
1	0,233	0,238	0,243	0,243	0,248	0,208	0,125	0,158	0,300	0,133
2	0,729	0,762	0,757	0,757	0,752	0,792	0,875	0,842	0,700	0,867
Ho	0,366	0,419	0,390	0,429	0,362	0,350	0,250	0,217	0,533	0,233
He	0,369	0,365	0,370	0,370	0,374	0,333	0,221	0,269	0,424	0,233
p _{HW}	1,000	0,175	0,789	0,115	0,794	1,000	0,579	0,146	0,063	1,000
d.p.	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000
p _{HD}	0,566	0,972	0,801	0,978	0,458	0,784	1,000	0,146	0,994	0,684
d.p.	-	-	-	-	-	-	-	-	-	-
p _{HE}	0,646	0,097	0,383	0,078	0,733	0,513	0,369	0,971	0,037	0,737
d.p.	-	-	-	-	-	-	-	-	-	-
n	101	105	105	105	105	61	61	61	61	61

n: número de amostras analisadas; Ho: heterozigose observada; He: heterozigose esperada; p: p-valor do teste de Hardy-Weinberg; d.p.: desvio-padrão de p; HD:

deficiência de heterozigotos; HE: excesso de heterozigotos; n/o: não-observado.

Desequilíbrio de ligação: microssatélites TNF e inserções Alu

Como já comentado anteriormente, os dez marcadores analisados nesse trabalho encontram-se na região dos genes do complexo de histocompatibilidade principal. A análise de desequilíbrio de ligação considerando todos os marcadores na população de Brasília

apresentou resultado estatisticamente significativo para três pares de marcadores: AluHG/AluHF (p-valor: 0,0132+-0,0021), TNFa/AluMICB (p-valor: 0,0014+-0,0009) e TNFa/TNFc (p-valor: 0,0139+-0,0049). A população de Kalunga, por sua vez, apresentou desequilíbrio de ligação para quatro pares de marcadores: TNFa/TNFd (p-valor: 0,0000+-0,0000), TNFd/TNFc (p-valor: 0,0124+-0,0050), TNFa/AluTF (p-valor: 0,0192+-0,0022) e TNFd/AluHG (p-valor: 0,0266+-0,00448).

Considerando a proximidade física entre os marcadores TNF, o forte desequilíbrio de ligação existente dentro do MHC (Trowsdale, 2011) e conhecimento de que essas populações são fruto de miscigenação recente (Barcelos *et al.*, 2006; Barcelos, 2006; Godinho, 2008; Gontijo, 2008; Arcanjo *et al.*, 2009; Dalton *et al.*, 2009; Oliveira *et al.*, 2009; Ribeiro *et al.*, 2009; Ribeiro, 2009; Silva, 2010; Trindade-Filho, 2010), poderia-se esperar que o número de *loci* ligados fosse maior (Pfaff *et al.*, 2001) do que o encontrado neste trabalho. No entanto, o trabalho de Miretti *et al.* (2005) mostrou que o desequilíbrio de ligação no MHC, embora forte, não é uniformemente distribuído ao longo do complexo, principalmente pelo alto número de *hotspots* de recombinação encontrados. Dessa forma, são encontrados blocos de desequilíbrio de ligação, separados por pontos de recombinação, que respaldam os resultados encontrados neste trabalho.

Microsatélites TNF: haplótipos

O número de haplótipos possíveis considerando todos os alelos já descritos para os microsatélites TNFa-e é de 4116. O número de haplótipos estimados, isto é, o número de combinações possíveis considerando os alelos encontrados em cada população, foi de 3822 para Brasília e 3072 para Kalunga, o que tem relação direta com o maior número de alelos encontrados para os *loci* TNFa e TNFd em Brasília. Utilizando a inferência haplotípica obtida com o uso do algoritmo PHASE, Brasília apresentou 112 haplótipos mais prováveis e

Kalunga 92, sendo 21 deles compartilhados entre as populações (tabela 07). A lista dos haplótipos encontrados para cada população pode ser encontrada nos Anexos 02 e 03. Em Brasília, os haplótipos mais comuns foram h5 (2 5 1 3 3 - 4,17%), h24 (4 4 1 3 3 - 5,24%) e h69 (4 3 1 4 3 - 4,67%), enquanto em Kalunga o haplótipo h114 (7 5 2 2 3 - 4,17%) foi o mais frequente.

Tabela 07. Haplótipos TNF (TNFa:TNFb:TNFc:TNFd:TNFe) compartilhados em Brasília e Kalunga, e a incidência deles em cada população.

Haplótipo	Brasília (n=210)	Kalunga (n=120)
h5(2 5 1 3 3)	9	2
h19(9 5 1 2 3)	1	1
h22(6 3 1 2 3)	2	2
h26(3 4 1 1 3)	1	1
h27(7 4 1 2 3)	1	1
h29(2 5 1 2 3)	2	1
h31(7 3 1 1 3)	1	1
h56(7 1 1 1 3)	3	2
h58(4 1 2 3 3)	1	1
h59(4 5 1 4 1)	2	1
h64(2 5 1 3 1)	1	1
h67(0 1 1 4 3)	1	1
h68(3 5 1 6 3)	4	1
h74(7 3 1 2 3)	1	3
h78(3 2 2 4 3)	1	1
h80(3 1 2 3 3)	3	1
h82(7 1 1 2 3)	2	2
h95(2 1 1 6 3)	2	2
h101(3 4 1 1 1)	1	1
h104(7 2 2 2 3)	1	1
h111(2 6 1 5 3)	1	2

n: número de cromossomos analisados.

Os haplótipos compartilhados entre Brasília e Kalunga (Tabela 07) somaram 19,52% e 24,17% dos haplótipos encontrados em cada população, respectivamente. Em Kalunga todos os outros haplótipos foram observados em frequência menor ou igual a 2,5%. Dos haplótipos mais frequentes em uma população, somente h5 foi compartilhado por elas e, mesmo assim, apresentando uma baixa frequência em Kalunga (1,67%).

Diante dos desvios do EHW apresentado por alguns marcadores, bem como o excesso e deficiência de heterozigotos apresentado por outros, mostrou-se necessário investigar possíveis dinâmicas que possam estar ocorrendo nas populações. A média do número de pares de marcadores diferentes entre as duas populações foi levemente maior, 3,46, do que a diferença média dentro de cada população (Brasília = 2,79; Kalunga = 3,37). A distância genética estimada pela média das diferenças entre as populações foi de 0,37.

Haplótipos das inserções Alu

A inferência de haplótipos é uma estratégia muito utilizada em substituição a análises familiares, que são muito mais dispendiosas em termos tanto de consumo quanto de tempo. Porém, os principais programas disponíveis para realizar essas inferências apresentam falhas. Considerando ainda que há uma distância relativamente grande entre as inserções *Alu* situadas no MHC de Classe I (1,6Mb de *AluMICB* até *AluHF*) analisadas no presente trabalho, buscou-se avaliar qual seria o melhor programa para ser utilizado com nossos dados. Para tanto foram realizadas análises familiares (trios) para a reconstrução dos haplótipos das inserções *Alu* aqui estudadas, que foi a mesma abordagem utilizada por Castelli *et al.* (2010) na reconstrução de haplótipos de *HLA-A*, *-B*, *-DRB1* e *-DQB1*.

O algoritmo PHASE, implementado no software PHASE v2.1, apresentou uma taxa de acerto de 83,33%, inferindo corretamente o haplótipo de 25 dos 30 indivíduos estudados. Já o algoritmo ELB, que apresentou as melhores inferências no trabalho de Castelli

et al. (2010), obteve uma taxa de 73,33% de acerto, inferindo corretamente os haplótipos de 22 dos 30 indivíduos estudados. Finalmente, o algoritmo EM, considerado até então a maneira mais robusta de se inferir haplótipos, mas que apresentou as piores inferências em Castelli *et al.* (2010) obteve a mesma taxa de acerto que o algoritmo ELB, errando exatamente os mesmos haplótipos dos mesmos indivíduos. Devido aos resultados obtidos no teste com trios, o algoritmo PHASE foi o escolhido para a inferência haplotípica dos marcadores selecionados no presente trabalho.

A tabela 08 mostra os haplótipos encontrados em Brasília e Kalunga para os cinco loci de marcadores *Alu*. Dos 32 haplótipos possíveis considerando apenas os marcadores *Alu*, foram inferidos 21 haplótipos para Brasília e 17 para Kalunga.

O haplótipo mais frequente encontrado em ambas populações (Brasília e Kalunga), também é o haplótipo mais frequente na maioria das populações que já foram estudadas para esses marcadores (Dunn *et al.*, 2003; Kulski & Dunn, 2005; Tian *et al.*, 2008; Yao *et al.*, 2009; García-Obregón *et al.*, 2011) e é o haplótipo ancestral, cujos *loci* se apresentam sem a inserção *Alu*. Essas populações compreendem indivíduos com ancestralidade européia (Australianos e Espanhóis), africana (Bantu, Kung e Sekele) e asiática (tailandeses, mongóis, e chineses de diferentes etnias). Nas populações chinesas, entretanto, o haplótipo mais frequente não foi o haplótipo ancestral 22222, mas o haplótipo 22211, que está presente em Brasília e Kalunga em apenas 5% dos indivíduos de cada população, aproximadamente.

O segundo haplótipo mais comum em Brasília é o terceiro haplótipo mais comum em Kalunga e apresenta inserção do elemento *Alu* somente no *locus* HJ. Esse mesmo haplótipo se encontra em frequência maior que 15% em sete das 11 populações que já foram estudadas para esses marcadores. Essa região do pseudo-gene *HLA-J* é considerada conservada, mas está muito próxima da região do gene *HLA-A*, de intensa variabilidade. A

persistência e a frequência expressiva dessa inserção na maioria das populações estudadas podem sugerir que o *locus* da inserção sofre um “efeito carona” na seleção de variabilidade do gene *HLA-A* (Meyer & Thomson, 2000).

Tabela 08. Listagem dos haplótipos (AluMICB:AluTF:AluHJ:AluHG:AluHF) encontrados nas duas populações estudadas, bem como suas frequências e desvio-padrão da estimativa de frequência. Os valores marcados indicam haplótipos com frequência maior ou igual a 10%.

Haplótipo	Brasília	Kalunga
2 2 2 2 2	0,2857 +- 0,0312	0,3667 +- 0,0442
2 2 2 2 1	0,0238 +- 0,0105	0,0167 +- 0,0117
2 2 2 1 2	0,0809 +- 0,0189	0,1750 +- 0,0348
2 2 2 1 1	0,0524 +- 0,0154	0,0500 +- 0,0200
2 2 1 2 2	0,1809 +- 0,0266	0,1083 +- 0,0285
2 2 1 2 1	0,0095 +- 0,0067	0,0083 +- 0,0083
2 2 1 1 2	n/o	0,0083 +- 0,0083
2 1 2 2 2	0,0714 +- 0,0178	0,0167 +- 0,0117
2 1 2 2 1	0,0238 +- 0,0105	0,0083 +- 0,0083
2 1 2 1 1	0,0048 +- 0,0048	n/o
2 1 1 2 2	0,0048 +- 0,0048	n/o
2 1 1 2 1	0,0048 +- 0,0048	n/o
2 1 1 1 2	0,0286 +- 0,0115	0,0333 +- 0,0165
1 2 2 2 2	0,0476 +- 0,0147	0,1000 +- 0,0275
1 2 2 2 1	0,0190 +- 0,0095	0,0083 +- 0,0083
1 2 2 1 2	n/o	0,0083 +- 0,0083
1 2 2 1 1	0,0429 +- 0,0140	0,0167 +- 0,0117
1 2 1 2 2	0,0095 +- 0,0067	0,0017 +- 0,0000
1 2 1 1 1	0,0048 +- 0,0048	n/o
1 1 2 2 2	0,0429 +- 0,0140	0,0583 +- 0,0215

Haplótipo	Brasília	Kalunga
1 1 2 2 1	0,0333 +- 0,0124	0,0083 +- 0,0083
1 1 2 1 2	0,0048 +- 0,0048	n/o
1 1 2 1 1	0,0238 +- 0,0105	0,0083 +- 0,0083

n/o: não observado; 1: inserção, 2: ausência de inserção.

Considerando os haplótipos observados utilizando os marcadores do tipo *Alu* foi possível identificar diferenças populacionais. Observou-se que as diferenças na distribuição dos haplótipos *Alu* nas populações de Brasília e Kalunga contrastaram com o encontrado para os haplótipos dos microssatélites TNF, o que sugere que as regiões de inserções *Alu* são mais conservadas do que as de microssatélites. A maior diferença encontrada entre as populações (1,68) foi menor do que a encontrada dentro de cada população (Brasília = 1,83; Kalunga = 1,48), justamente o oposto do que observamos para os haplótipos dos microssatélites TNF. Além disso, a distância entre as populações estimada com base na média de diferenças intra e inter-populacionais indicou grande similaridade, com distância estimada de 0,03, p-valor = 0,99.

Observou-se também que quanto maior o número de inserções na configuração de um haplótipo para os marcadores aqui estudados, menores são as frequências do haplótipo em questão nas populações estudadas. E ainda, o haplótipo 11111 não foi observado em nenhuma das populações que foram estudadas até agora. De certa maneira, esse resultado já era de se esperar para as populações com histórico mais jovens, como a asiática e a ameríndia. Como as populações africanas são geneticamente ancestrais, esperava-se que elas apresentassem maior variabilidade que as outras populações. No entanto, os dados presentes na literatura e gerados por este trabalho indicaram que as populações africanas foram também conservadoras com relação às inserções *Alu* estudadas aqui, mantendo, em sua maioria, os haplótipos ancestrais (García-Obregón *et al.*, 2010). Outra explicação possível para essa condição é que essas

inserções *Alu* pertencem às famílias *AluYa5* e *AluYb8* (Dunn *et al.*, 2002), que são extremamente recentes (Deininger & Batzer, 1999) e por isso ainda apresentam baixa frequência na população humana.

Microsatélites TNF e Inserções Alu

Ao todo, considerando o número de alelos por *locus* e o número de *loci* estudados, o número de combinações de haplótipos possíveis para o conjunto de marcadores selecionados é de 131.762. Levando em conta, agora, os alelos encontrados em cada população e o número de haplótipos que podem ser gerados a partir dos mesmos, foi estimado que a população de Brasília poderia apresentar 2861 e Kalunga 1959 haplótipos distintos. Após a inferência haplotípica, entretanto, foram inferidos 191 haplótipos para a população de Brasília e 116 para Kalunga. Destes, nove dos haplótipos gerados pelo PHASE apresentaram-se compartilhados pelas duas populações (Tabela 09). A Tabela contendo os haplótipos obtidos para cada população está apresentada nos Anexos 04 e 05.

Tabela 09. Haplótipos (TNFa:TNFb:TNFc:TNFd:TNFe:AluMICB:AluTF:AluHJ:AluHG:AluHF) compartilhados entre as populações de Brasília e Kalunga e a incidência dos mesmos em cada população.

Haplótipo	Brasília (n=210)	Kalunga (n=120)
h25 (9 5 1 2 3 2 1 2 1 2)	1	1
h34 (3 4 1 1 3 2 2 2 2 2)	1	1
h37 (6 3 1 2 3 2 2 2 1 1)	1	1
h38 (2 5 1 2 3 2 2 2 2 2)	1	1
h109 (3 2 2 4 3 2 2 2 2 2)	1	1
h111 (3 1 2 3 3 2 2 2 2 2)	2	1
h131 (7 1 1 2 3 2 2 2 2 2)	1	2
h148 (2 1 1 6 3 2 2 2 2 2)	1	1
h156 (3 4 1 1 1 2 2 2 2 2)	1	1

n: número de cromossomos

Como não há trabalhos que tenham estudado simultaneamente os dois conjuntos de marcadores do MHC de Classe III e de Classe I aqui analisados, não foi possível tecer comparações com dados de outras populações. Entretanto, cabe ressaltar que devido ao número amostral e o número de alelos encontrados para cada *locus*, o pequeno número de haplótipos em comum entre Brasília e Kalunga era esperado. E ainda, como já comentado, a diferença entre as populações já era esperada em decorrência dos trabalhos previamente publicados para essas duas populações com marcadores neutros, que mostram que as duas populações são, de fato, geneticamente muito diferentes (Barcelos *et al.*, 2006; Barcelos, 2006; Godinho, 2008; Gontijo, 2008; Arcanjo *et al.*, 2009; Dalton *et al.*, 2009; Oliveira *et al.*, 2009; Ribeiro *et al.*, 2009; Ribeiro, 2009; Silva, 2010; Trindade-Filho, 2010). A única exceção para essa tendência foi o trabalho de Toledo (2010), que ao estudar a região 3'UTR do gene *HLA-G* não verificou diferenças significativas entre as duas populações, além de ter obtido um teste significativo para o desvio da neutralidade nessa região.

Diferenciação populacional

Foram realizados dois testes de diferenciação populacional baseado em F_{ST} para o conjunto de marcadores *Alu* em Brasília, Kalunga e demais populações. O primeiro teste foi realizado utilizando as frequências alélicas para os *loci* em cada população (figura 12), e o segundo teste foi utilizando as frequências haplotípicas para cada população (figura 13).

Como podemos observar na figura 13, as populações que apresentaram maiores diferenças foram as populações de japoneses e mongóis quando comparados às populações africanas Bantu, Sekele e !Kung. A população de Brasília apresentou maior diferenciação populacional com relação a Bantu e !Kung, menor com relação a Kalunga, Australianos e Valência. A população australiana aqui mencionada e a população de Valência apresentam

ancestralidade europeia, assim como Brasília também apresenta alta ancestralidade europeia de acordo com estudo de marcadores uniparentais (91,6%; Barcelos, 2006) e marcadores informativos de ancestralidade (61%; Godinho, 2008). A população de Kalunga, entretanto, mostrou pequena diferenciação populacional com relação às populações europeias (Australianos e Valência), Brasília, Malaio-Chineses e Sekele, mas diferenciação moderada com relação às populações africanas !Kung e Bantu. Essa diferenciação maior com relação às populações africanas e menor com relação às populações de origem europeia podem refletir a intensa miscigenação que os ancestrais dos Kalunga já apresentavam antes mesmo da formação da comunidade. Estudo que estimaram a composição genética de Kalunga baseado em marcadores informativos de ancestralidade apontaram que a contribuição genética europeia para a formação dessa população varia de cerca de 10% (marcadores do cromossomo X; Ribeiro, 2009) até 60% (microsatélites do cromossomo Y; Ribeiro, 2009), passando por volta dos 30% quando considerados AIMs autossômicos (Pedrosa, 2006). Dessa forma, especula-se que a contribuição europeia na formação da população Kalunga possa ser o motivo da pequena diferenciação encontrada entre essa população e outras populações de ancestralidade europeia que foram estudadas para os marcadores apresentados neste trabalho.

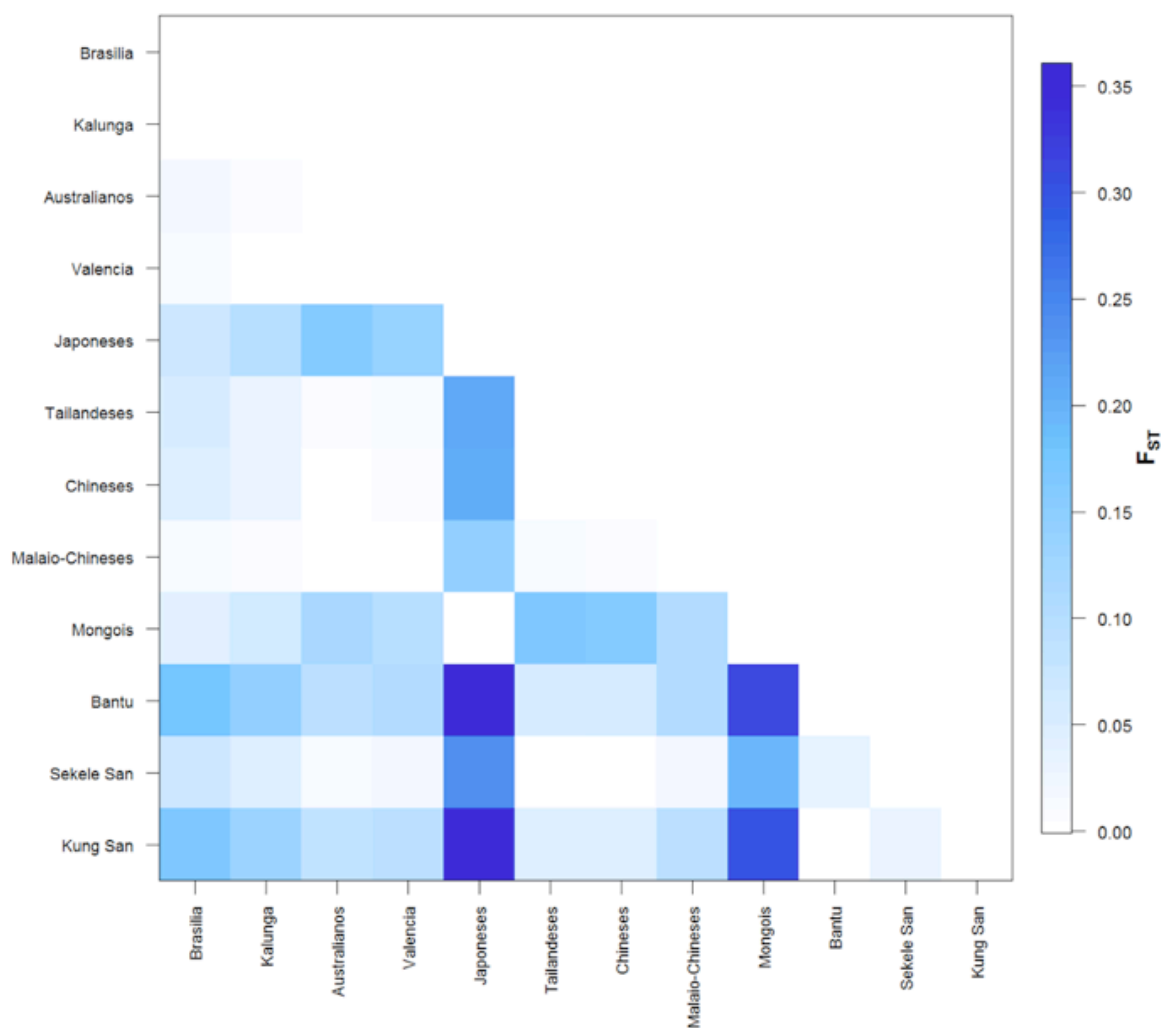


Figura 12. Matriz de diferenciação populacional baseado em F_{ST} para populações dos diversos continentes (Brasília, Kalunga, Australianos, Valência, Japoneses, Tailandeses, Chineses, Malaio-Chineses, Mongóis, Bantu, Sekele e !Kung), analisadas para o conjunto de marcadores *Alu* presentes neste trabalho.

Analisando a diferenciação populacional baseada nas frequências dos haplótipos estimados para as inserções *Alu*, Brasília e Kalunga apresentaram uma diferenciação um pouco maior em relação uma à outra (figura 13). Brasília apresentou maior diferenciação populacional com relação à população Sekele (africana) e menor diferenciação com relação à

Valência e Australianos. Esse resultado foi parcialmente compartilhado com Kalunga, que também apresentou maior diferenciação à Sekele, porém a menor diferenciação encontrada foi com relação aos Australianos. Entretanto, Kalunga apresentou maior diferenciação à Valência quando consideramos as frequências haplotípicas do que quando consideramos apenas as frequências alélicas (figura 12). A população de mongóis apresentou maior diferenciação a Sekele baseada nas frequências haplotípicas do que quando baseada em frequências alélicas.

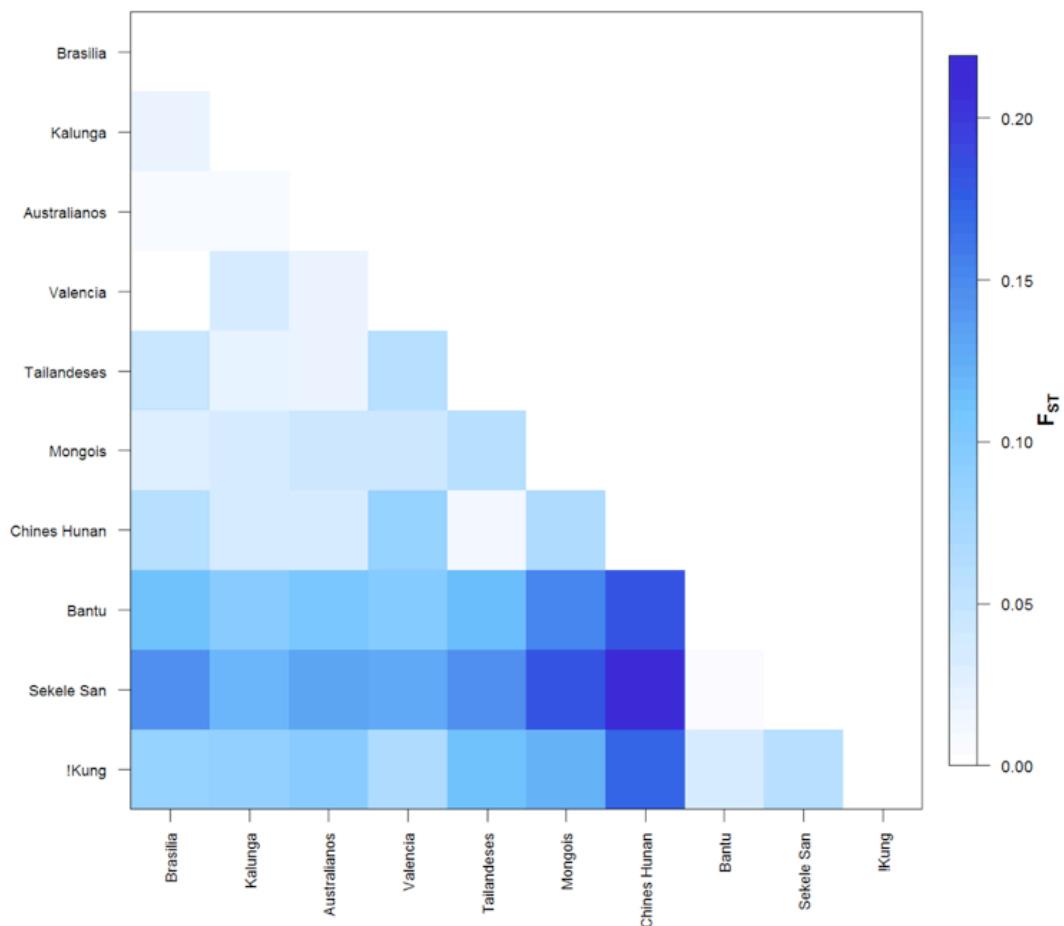


Figura 13. Diferenciação populacional baseada em F_{ST} com relação à frequência dos haplótipos encontrados em cada população, nas populações Brasília, Kalunga, Australianos, Valência, Tailandeses, Mongóis, Chineses Hunan, Bantu, Sekele e !Kung.

As diferenças médias intra e inter-populacionais calculadas para todas as populações baseadas nos haplótipos (figura 14) indicaram que as diferenças máximas observadas entre populações foram muito próximas às diferenças máximas observadas dentro de cada população. Brasília e Valência mostraram um grande número de diferenças intra-populacionais, mas a distância entre elas foi próxima de zero. A população mais distante de Brasília foi Sekele, apresentando entre 15% e 20% de distância dessa população. Kalunga, por sua vez, mostrou grande proximidade à população de australianos, uma vez que a distância de Nei estimada pela média de diferenças alélicas para essas duas populações foi próxima a zero. A população mais distante de Kalunga também foi Sekele, apresentando de 10% a 15% de distância dessa população. A menor média de diferenças intra-populacionais foi encontrada para a população Sekele, que também apresentou a menor média de diferenças inter-populacionais quando comparada à população Bantu.

Em resumo, o relacionamento genético entre as populações considerando esse conjunto de marcadores *Alu*, apesar de apresentar alguma coerência com a ancestralidade das populações, em linhas gerais não difere muito. Essa similaridade em termos de frequências alélicas e haplotípicas entre as populações pode ser resultado da inserção recente desses elementos móveis no genoma humano, ou de algum mecanismo evolutivo no sentido de conservar a sequência ancestral, sem a inserção *Alu*.

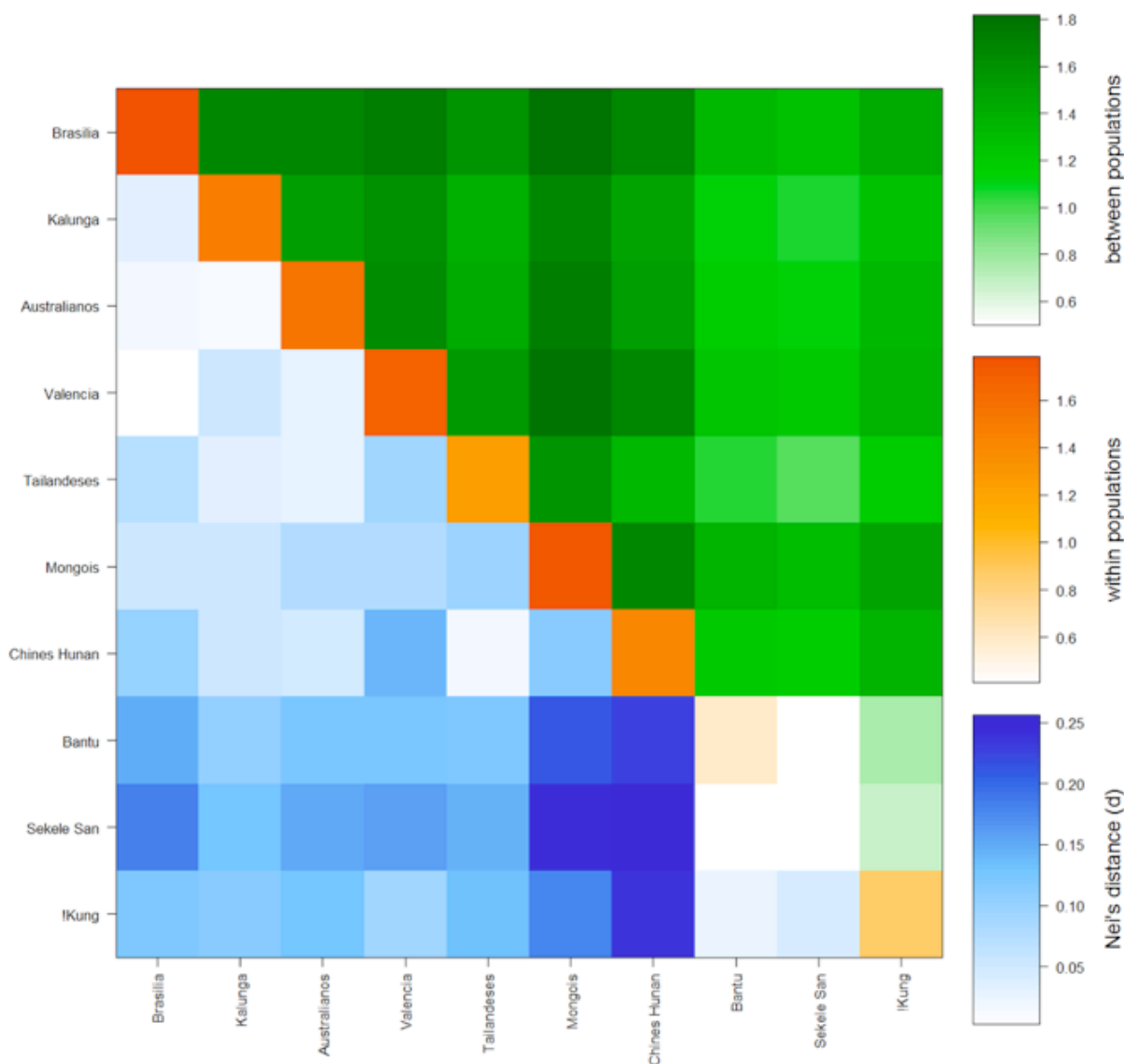


Figura 14. Média das diferenças intra (laranja) e inter-populacionais (verde) para cada par de populações analisadas, bem como a distância de Nei estimada pela média de diferenças encontradas entre cada par de populações (azul) para as populações Brasília, Kalunga, Australianos, Valência, Tailandeses, Mongóis, Chineses Hunan, Bantu, Sekele e !Kung.

De acordo com o índice de diferenciação populacional (F_{ST}) (tabela 10), há uma diferença estatisticamente significativa considerando os conjuntos de marcadores utilizados. Porém, considerando a classificação de F_{ST} de acordo com Wright (1978, apud Hartl & Clark, 2010), o resultado difere entre os marcadores. Os marcadores *Alu* indicaram pequena

diferenciação entre as populações. enquanto os marcadores microssatélites indicaram diferenciação populacional moderada. Em relação ao conjunto inteiro de marcadores, contudo, o índice indicou diferenciação populacional moderada, provavelmente porque o F_{ST} dos marcadores microssatélites elevou a média dos índices entre as populações.

Tabela 10. Índice de diferenciação populacional (F_{ST}) entre Brasília e Kalunga, calculado a partir dos *loci* estudados.

Conjunto de marcadores	F_{ST}	p-valor
Alu	0,01552	0,00000+-0,0000
STR	0,11348	0,00000+-0,0000
Todos os marcadores	0,07983	0,00000+-0,0000

Teste de neutralidade

O Teste de Neutralidade de Ewens-Watterson indicou que apenas os *loci* TNFa ($F = -1,4645$, p-valor: 0,0034), TNFb ($F = -1,6186$, p-valor: 0,0027) e TNFd ($F = -1,5236$, p-valor: 0,0072) não se comportam como regiões neutras do genoma nas duas populações estudadas. O Teste de Neutralidade de Ewens-Watterson aplicado aos haplótipos e aos marcadores do tipo inserção *Alu* não revelou desvios estatisticamente significativos do esperado em condições de neutralidade. Resultado semelhante foi encontrado em trabalho recente de García-Obregón *et al.* (2011), onde apenas o microssatélite TNFa mostrou desvio da neutralidade em um conjunto de 10 marcadores. Neste trabalho, os autores analisaram as cinco inserções *Alu* estudadas nesta dissertação de mestrado, e mais cinco microssatélites (somente o TNFa em comum com esta dissertação) em diversas populações europeias, africanas e asiáticas.

Conclusão

Os resultados aqui apresentados sugerem que as inserções *Alu* situadas no MHC de Classe I são mais conservadas que os microssatélites presentes nos genes *TNF*. Além disso, Brasília e Kalunga mostraram diferentes graus de diferenciação com relação as comparações dos dados obtidos com a análise das inserções *Alu*, dos microssatélites e dos haplótipos considerando todos os marcadores. Por outro lado, nesse trabalho não foram feitas inferências acerca do comportamento das regiões genômicas em si (MHC de classe I e MHC de classe III) pois o comportamento dos marcadores nem sempre reflete a região e sim, mais frequentemente, a classe de marcadores a qual pertencem.

Com este trabalho, pôde-se concluir que a dinâmica dos marcadores microssatélites e dos marcadores *Alu* nas duas populações estudadas são diferentes, porém mais similares quando se consideram as inserções *Alu*, e que as informações que podem ser obtidas desse conjunto de marcadores são complementares. Além disso, observou-se que três microssatélites *TNF* (*TNFa*, *TNFb*, *TNFd*) indicaram desvio da neutralidade assim como sugeriu que o *AluHG* esteja sofrendo seleção natural por efeito carona da seleção sofrida pelo gene *HLA-G*.

Para uma melhor compreensão dos possíveis mecanismos evolutivos que possam estar exercendo influência nas regiões do MHC de Classe I e de Classe III, sugere-se que a confirmação dessas conclusões seja feita a partir da análise de haplótipos de inserções *Alu* presentes no MHC de Classe III em contraste a haplótipos de microssatélites no MHC de Classe I.

Referências bibliográficas

- Arcanjo AC, Dalton GC, Godinho NMO, Gontijo CC, Barcelos RSS, Astolfi-Filho S, Pontes IM, Klautau-Guimarães MN, Oliveira SF. DRD2TaqI haplotypes in three urban Brazilian populations. *Forensic Science International: Genetics Supplement Series 2*:481-482, 2009.
- Barcelos RSS. Contribuição genética de duas populações urbanas da região Centro-Oeste brasileira estimada por marcadores uniparentais. *Tese de Doutorado*. Universidade de Brasília, 2006.
- Barcelos R, Ribeiro G, Silva Jr W, Abe-Sandes K, Godinho N, Marinho-Neto F, Gigonzac M, Klautau-Guimarães M, Oliveira S. Male contribution in the constitution of the Brazilian Centro-Oeste population estimated by Y-chromosome binary markers. *International Congress Series 1288*:228-230, Progress in Forensic Genetics 11,2006.
- Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselte TD, Hoppens CL, Deininger PL. Structure and variability of recently inserted *Alu* family members. *Nucleic Acids Research* 18: 6793-6798, 1990.
- Batzer MA, Deininger PL. A human-specific subfamily of *Alu* sequences. *Genomics* 9:481-487, 1991.
- Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer DW, Herrera RJ, Deininger PL. African origin of human-specific *Alu* insertions. *PNAS* 91: 12288-12292, 1994.
- Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJB, Deininger PL, Stoneking M. Genetic variation of recent *Alu* insertions in human populations. *Journal of Molecular Evolution* 42: 22-29, 1996.

- Callinan PA, Batzer MA. Retrotransposable Elements and Human Disease. *Genome and Disease. Genome Dyn. Basel* 1:104-115, 2006.
- Castelli, EC. Polimorfismos dos genes HLA-G e de citocinas em pacientes com histórico e carcinoma de células transicionais da bexiga urinária. *Tese de Doutorado*. USP, 2008.
- Castelli EC, Mendes-Junior CT, Veiga-Castelli LC, Pereira NF, Petzl-Erler ML, Donadi EA. Evaluation of computational methods for the reconstruction of HLA haplotypes. *Tissue Antigens* 79:459-466, 2010.
- Castelli EC, Mendes-Junior CT, Veiga-Castelli LC, Roger M, Moreau P, Donadi EA. A comprehensive study of polymorphic sites along the *HLA-G* gene: implication for gene regulation and evolution. *Mol Biol Evol* 28(11):3069-3086, 2011.
- Castro MJ, Morales P, Martinez-Laso J, Allende L, Rojo-Amigo R, Gonzalez-Hevilla M. Evolution of MHC-G in humans and primates based on three new 3'UTR polymorphisms. *Hum Immunol* 61(11): 1157-63, 2000.
- Comissão Pró-Índio de São Paulo. <http://www.cpisp.org.br/htm/leis/page.aspx?LeiID=141>, acessado em 18 de fevereiro de 2012.
- Cordeaux R, Hedges DJ, Herke SW, Batzer MA. Estimating the retrotransposition rate of human *Alu* elements. *Gene* 373:134-137, 2006.
- Cordeaux R, Srikanta D, Lee J, Stoneking M, Batzer MA. In search of polymorphic *Alu* insertions with restricted geographic distributions. *Genomics* 90:154-158, 2007.
- Dalton GC, Maia FAS, Mendes CRBO, Pak HI, de Paula KAA, Trindade-Filho A, Oliveira SF. Genetic profile of Federal District of Brazil based on 18 STR autosomal *loci*. *Forensic Science International: Genetics Supplement Series* 2:373-375, 2009.
- Deghaide NHS, Rodrigues MLV, Castelli EC, Mendes-Junior CT, Figueiredo JFC, Donadi EA. Tumor necrosis factor region polymorphisms are associated with AIDS and with cytomegalovirus retinitis. *AIDS* 23:1641-1647, 2009.

- Deininger PL, Batzer MA. Alu Repeats and Human Disease. *Molecular Genetics and Metabolism* 67, 183–193, 1999.
- Derkx HHF, Bruin CV, Jongneel LP, de Waal LP, Brinkman BMN, Verweij CL, Houwing-Duistermaat JJ, Rosendal FR and van De-Venter SJH. Familial differences in endotoxin-induced TNF release in whole blood mononuclear cells in vitro: relationship to TNF gene polymorphism. *J Endotoxin Res* 2:19–25, 1995.
- Dunn DS, Naruse T, Inoko H, Kulski JK. 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. *J Mol Evol* 55:718-726, 2002.
- Dunn DS, Inoko H, Kulski JK. Dimorphic Alu element located between the TFIID and CDSN genes within the major histocompatibility complex. *Electrophoresis* 24, 2740–2748, 2003.
- Edwards A, Civitello A, Hammond HA, Caskey CT. DNA Typing and Genetic Mapping with Trimeric and Tetrameric Tandem Repeats. *Am. J. Hum. Genet* 49:746-756, 1991.
- Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564-567, 2010.
- Foissac A, Crouau-Roy B, Faurè S, Thomsen M, Cambon-Thomsen A. Microsatellites in the HLA region: an overview. *Tissue Antigens* 49:197-214, 1997.
- Fundação Cultural Palmares - www.palmares.gov.br, acessado em 18 de fevereiro de 2012.
- Gallagher G, Eskdale J, Oh HH, Richards SD, Campbell DA, Field M. Polymorphisms in the TNF gene cluster and MHC serotypes in the West of Scotland. *Immunogenetics* 45:188-194, 1997.
- García-Obregón S, Alfonso-Sánchez MA, Gómez-Pérez L, Pérez-Miranda AM, Arroyo D, de Pancorbo MM, Peña JA. Microsatellites and Alu elements from the human MHC in

- Valência (Spain): analysis of genetic relationships and linkage disequilibrium. *International Journal of Immunogenetics* 38:483-491, 2011.
- Genographic Project. <https://genographic.nationalgeographic.com/genographic/index.html>, acessado em 18 de fevereiro de 2012.
- Godinho NMO. O impacto das migrações na constituição genética das populações Latino-Americanas. *Tese de Doutorado*. Universidade de Brasília, 2008.
- Gontijo CC. Composição Genética de Duas Populações Afro-derivadas Brasileiras Inferida a partir de Marcadores Informativos de Ancestralidade. *Dissertação de Mestrado*. Universidade de Brasília, 2008.
- Grubic Z, Stingl K, Zunec R, Car H, Cecuk-Jelicic E, Brkljadic-Kerhin V. Linkage disequilibria between human leucocyte antigen-B and closely linked microsatellites in the Croatian population. *Tissue Antigens* 69:86-94, 2006.
- Guo SW, Thompson EA. Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* 48(2):361-372, 1992.
- Hajeer AH, Hutchinson IV. TNF-alpha gene polymorphism: clinical and biological implications. *Microsc Res Tech*, 50(3):216-228, 2000.
- Hartl DL & Clark AG. Princípios de Genética de Populações. 4ª Edição. Ed Artmed, 2010.
- Horton R, Gibson R, Coghill P, Miretti M, Allcock RJ, Almeida J, Forbes S, Gilbert JGR, Halls K, Harrow JL, Hart E, Howe K, Jackson DK, Palmer S, Roberts AN, Sims S, Stewart CA, Traherne JA, Trevanion S, Wilming L, Rogers J, Jong PJ, Elliot JF, Sawcer S, Todd JA, Trowsdale J, Beck S. Variation analysis and gene annotation of eight MHC haplotypes: The MHC Haplotype Project. *Immunogenetics* 60:1-18, 2008.
- Houck CM, Rinehart FP, Schmid CW. A ubiquitous family of repeated DNA sequences in the human genome. *Journal of Molecular Biology* 132(2):289-306, 1979.

HUGO – Human Genome Organisation. <http://www.hugo-international.org/>, acessado em 18 de fevereiro de 2012.

International HapMap Project. (<http://hapmap.ncbi.nlm.nih.gov/>), acessado em 18 de fevereiro de 2012.

Janeway CA, Travers P, Walport M, Shlomchik MJ. *Imunobiologia*. 6ª Edição. Ed. Artmed, 2007.

Jongeneel CV, Briant L, Udalova IA, Sevin A, Nedospasov SA, Cambon-Thomsen A. Extensive genetic polymorphism in the human tumor necrosis factor region and relation to extended HLA haplotypes. *PNAS* 88: 9717-9721, 1991.

Kim HK, Han H, Choi HB, Pyo CW, Kim CC, Kim TG. Distribution of seven polymorphic markers and haplotypes within the human TNF gene cluster in Koreans. *Human Immunology* 61:1274-1280, 2000.

Klein J, Sato A. The HLA system. First of two parts. *N Engl J Med*. 343(10): 702-9, 2000.

Klein J, Sato A. The HLA system. Second of two parts. *N Engl J Med*. 343(11): 782-6, 2000.

de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genetics* 12:e1002384, 2011.

Koreth J, O’Leary JJ, McGee JOD. Microsatellites and PCR genomic analysis. *Journal of Pathology* 178:239-248, 1996.

Kriegs JO, Churakov G, Jurka J, Brosius J, Schimtz J. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *TRENDS in Genetics* 23(4):158-161, 2007.

Kulski JK, Martinez P, Longman-Jacobsen N, Wang W, Williamson J, Dawkins RL, Shiina T, Naruse T, Inoko H. The Association Between HLA-A Alleles and an Alu Dimorphism Near HLA-G. *J Mol Evol* 53:114–123, 2001.

Kulski JK, Dunn DS, Martinez JHP, Leelayuwat AVRC, Tay GK, Inoko AOH. Alu polymorphism within the MICB gene and association with HLA-B alleles. *Immunogenetics* 53:975–979, 2002.

- Kulski JK, Dunn DS. Polymorphic Alu insertions within the Major Histocompatibility Complex class I genomic region: a brief review. *Cytogenet Genome Res* 110:193–202, 2005.
- Lancaster A, Nelson MP, Meyer D, Thomson G, Single RM. PyPop: A software framework for population genomics: analyzing large-scale multi-locus genotype data. *Pac Symp Biocomput* 514-525, 2003.
- Malkki M, Single R, Carrington M, Thomson G, Petersdorf E. MHC microsatellite diversity and linkage disequilibrium among common HLA-A, HLA-B, DRB1 haplotypes: implications for unrelated donor hematopoietic transplantation and disease association studies. *Tissue Antigens* 66: 114–124, 2005.
- Marsh HP, Haldar NA, Bunce M, Marshall SE, le Monier K, Winsey SL, Christodoulos K, Cranston D, Welsh KI, Harris AL. Polymorphisms in tumour necrosis factor (TNF) are associated with risk of bladder cancer and grade of tumour at presentation. *British Journal of Cancer* 89:1096-1101, 2003.
- McDonnell GV, Kirk CW, Middleton D, Droogan AG, Hawkins SA, Patterson CC, Graham CA. Genetic association studies of tumor necrosis factor α and β and tumor necrosis factor receptor 1 and 2 polymorphisms across the clinical spectrum of multiple sclerosis. *Journal of Neurology* 246:1051-1058, 1999.
- Meyer D, Thomson G. How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics* 65:1-26, 2001.
- Miller, SA; Dykes, DD; Polesky, HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 16(3): 1215, 1988.
- Miretti MM, Walsh EC, Ke X, Delgado M, Griffiths M, Hunt S, Morrison J, Whittaker P, Lander ES, Cardon LR, Bentley DR, Rioux JD, Beck S, Deloukas P. A high-resolution

- linkage-disequilibrium map of the Human Major Histocompatibility Complex as first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet* 76:634-646, 2005.
- Nei M. *Molecular Evolutionary Genetics*. Columbia University Press, 1987.
- Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS* 76:5269-5273, 1979.
- Novick GE, Batzer MA, Deininger PL, Herrera RJ. The mobile genetic element “Alu” in the Human Genome. *BioScience* 46(1):32-41, 1996.
- Oliveira, LA. Diversidade Populacional de Microssatélites no Complexo De Histocompatibilidade Principal (MHC) em Populações Indígenas do Paraná e do Mato Grosso do Sul. *Dissertação de Mestrado*. Universidade Federal do Paraná, 2007.
- Oliveira SF, Godinho NM, Amorim CEG, Gontijo CC, Falcão-Alencar G, Barcelos RSS, Luizon, MR, Klautau-Guimarães MN, Simões AL, Mendes-Junior CT. STRs and AIMs informativeness for forensic purposes in an admixed Brazilian population. *Forensic Science International: Genetics Supplement Series* 2:475-476, 2009.
- Peña JA, Calderon R, Perez-Miranda A, Vidales C, Dugoujon JM, Carrion M, Crouau-Roy B. Microsatellite DNA markers from HLA region (D6S105, D6S265 and TNFa) in autochthonous Basques from Northern Navarre (Spain). *Annals of Human Biology* 29(2):176-191, 2002.
- Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *American Journal of Human Genetics* 68:198-207, 2001.
- Pociot F, Briant L, Jongeneel CV, Molvig J, Worsaae H, Abbal M, Thomsen M, Nerup J, Cambon-Thomsen A. Association of tumor necrosis factor (TNF) and class II major histocompatibility complex alleles with the secretion of TNF-alpha and TNF-beta by

- human mononuclear cells: a possible link to insulin-dependent diabetes mellitus. *Eur J Immunol* 23:224–231, 1993.
- Rassi DM, Wastowski IJ, Simões RT, Rodrigues S, Deghaide NNHS, Mendes-Júnior CT, Simões AL, Soares CP, Passos GAS, Donadi EA. (2006) TNFa-e Microsatellite, HLA-DRB1 and -DQB1 Alleles and Haplotypes in Brazilian Patients Presenting Recently Diagnosed Type 1 Diabetes Mellitus. *Ann. N.Y. Acad. Sci* 1079: 300–304, 2006.
- Ribeiro, GGBL. Análise genética de marcadores do tipo STR e Indel em cromossomos sexuais humanos em populações remanescentes de quilombos. *Tese de Doutorado*. Universidade de Brasília, 2009.
- Ribeiro GGBL, De Lima RR, Wiezel CEV, Ferreira LB, Sousa SMB, Rocha DMS, Canas MCT, Nardelli-Costa J, Klautau-Guimarães MN, Simões AL, Oliveira SF. Afro-derived Brazilian populations: Male genetic constitution estimated by Y-chromosomes STRs and YAP element polymorphisms. *American Journal of Human Biology* 21(3): 354-356, 2009.
- Rogers J. A straight LINE story. *Nature* 301:113-114, 1983.
- Salem AH, Ray DA, Batzer MA. Identity by descent and DNA sequence variation of human SINE and LINE elements. *Cytogenet Genome Res* 108:63–72, 2005.
- dos Santos IL. Análise dos Microssatélites D6S2874, D6S273, D6S2792, STR_MICA e HLABC-2CA da região MHC em uma amostra de afro-descendentes doadores voluntários de medula óssea. *Dissertação de Mestrado*. Universidade do Paraná, 81pp, 2009.
- Sens-Abuázar C, Santos PSC, Bicalho MG, Petzl-Erler ML, Sperandio-Roxo V. MHC microsatellites in a Southern Brazilian population. *Journal of Immunogenetics* 36:269-274, 2009.
- Shedlock AM, Okada N. SINE insertions: powerful tools for molecular systematics. *BioEssays* 22(2): 148-160, 2000.

- Shiina T, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations. *Tissue Antigens* 64:631-649, 2004.
- Silva EL. Marcadores moleculares associados à resistência/susceptibilidade ao HIV-1 no Distrito Federal (Brasil). *Dissertação de Mestrado*. Universidade de Brasília, 2010.
- Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462, 1995.
- Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989, 2001.
- Stewart C, Kural D, Strömberg MP, Walker JA, Konkel MK, Stütz AM, Urban AE, Grubert F, Lam HYK, Lee WP, Busby M, Indap AR, Garrison E, Huff C, Xing J, Snyder MP, Jorde LP, Batzer MA, Korbel JO, Marth GT, 1000 Genomes Project. A comprehensive map of Mobile Element Insertion polymorphisms in humans. *PLoS Genetics* 7(8):e1002236, 2011.
- Strachan T, Read AP. Human Molecular Genetics. Wiley-Liss, 2nd edition, 1999.
- Tan CY, Chong YS, Loganath A, Chan YH, Ravichandran J, Lee CG, Chong SS. Possible gene-gene interaction of KIR2DL4 with its cognate ligand HLA-G in modulating risk for preeclampsia. *Reprod Sci* 16:1135–1143, 2009.
- Tian W, Wang F, Cai JH, Li LX. 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* 72:559-567, 2008.
- Toledo RCP. Região 3'UTR do Gene HLA-G em populações humanas do Centro-Oeste. *Dissertação de Mestrado*. Universidade de Brasília, 61pp, 2010.
- Trindade-Filho A. Caracterização genética da população do Distrito Federal com base em marcadores STR do cromossomo X. *Dissertação de Mestrado*. Universidade de Brasília, 2010.

- Trowsdale J. The MHC, disease and selection. *Immunology Letters* 137:1-8, 2011.
- Turner DM, Grant SC, Lamb WR, Brenchley PE, Dyer PA, Sinnott PJ, Hutchinson IV. A genetic marker of high TNF-alpha production in heart transplant recipients. *Transplantation* 60:1113–1117, 1995.
- Udalova IA, Nedospasov SA, Webb GC, Chaplin DD, Turetskaya RL. Highly informative typing of the human TNF locus using six adjacent polymorphic markers. *Genomics* 1993; 16:180–186.
- Walsh EC, Mather KA, Schaffner SF, Farwell L, Daly MJ, Patterson N, Cullen M, Carrington M, Bugawan TL, Erlich H, Campbell J, Barrett J, Miller K, Thomson G, Lander ES, Rioux JD. An Integrated Haplotype Map of the Human Major Histocompatibility Complex. *Am. J. Hum. Genet.* 73:580–590, 2003.
- Wang J. Informativeness of genetic marker for pairwise relationship and relatedness inference. *Theoretical Population Biology* 70:300-321, 2006.
- Watterson G. The homozygosity test of neutrality. *Genetics* 88:405-417, 1978.
- Weber JL, Long C. Mutation of human short tandem repeats. *Human Molecular Genetics* 8(2):1123-1128, 1993.
- Whitherspoon DJ, Marchani EE, Watkins WS, Ostler CT, Wooding SP, Anders BA, Fowlkes JD, Boissinot S, Furano AV, Ray DA, Rogers AR, Batzer MA, Jorde LB. Human Population Genetic Structure and Diversity Inferred from Polymorphic L1 (LINE-1) and Alu Insertions. *Hum Hered.* 62:30–46, 2006.
- Yao Y, Shi L, Shi L, Lin K, Yu L, Sun H, Huang X, Tao Y, Yi W, Liu S, Chu J. The association between HLA-A, -B alleles and major histocompatibility complex class I polymorphic Alu insertions in four populations in China. *Tissue Antigens.* 73, 575 –581, 2009.

Zietkewicz E, Richer C, Makalowski W, Jurka J, Labuda D. A young Alu subfamily amplified independently in human and African great apes lineages. *Nucl Acid Res* 22:5608–5612, 1994.

Anexo 1 - Aprovações do Comitê de Ética em Pesquisa e do Conselho Nacional de Ética em Pesquisa para a utilização das amostras estudadas.

PROC

Pesquisador Responsável: Dr. Simone Pamella de Oliveira

Titulo: "Ancestralidade genética em populações indígenas brasileiras"

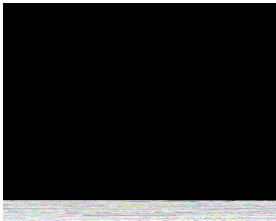
Artigo sobre o tema (item VII.13 da Resolução 196/96).

Brasília, 12 de

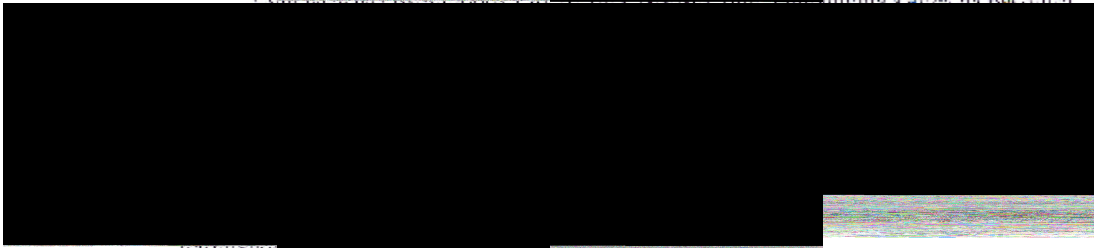
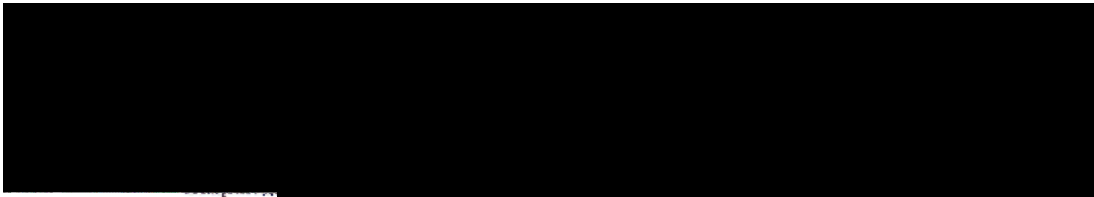
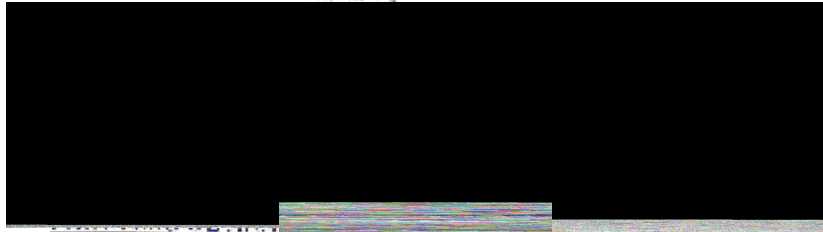
Comitê de Ética em Pesquisa

Brasília -

0.910.906/1 - Fone: (61) 3307-1199

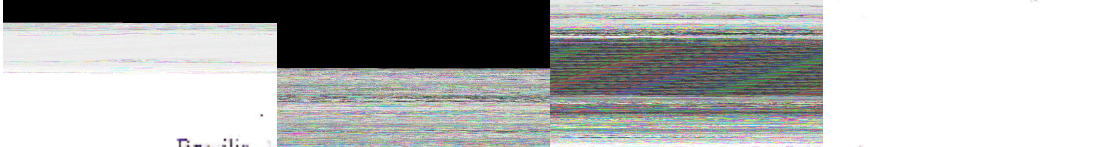
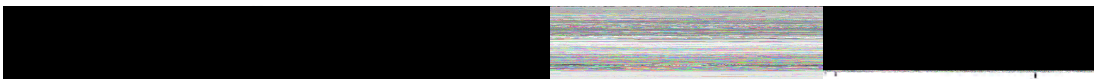


Brasília, 1



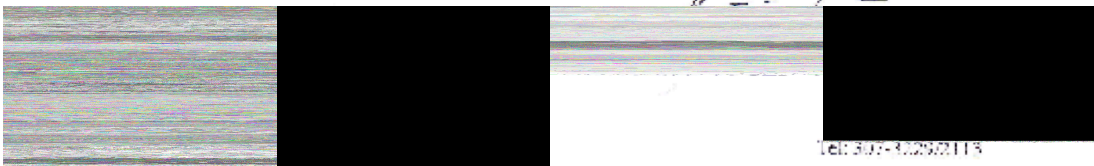
Brasília, 1

Brasília, 1



Brasília, 1

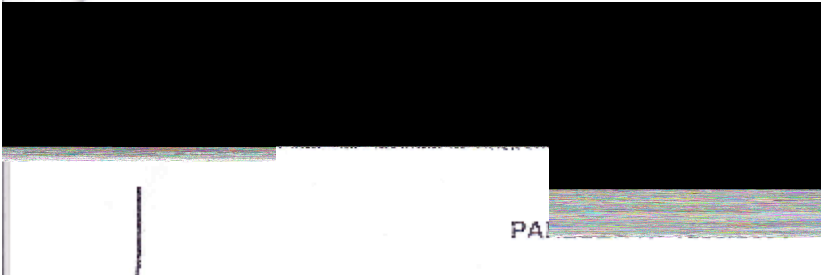
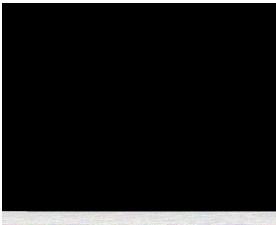
Handwritten signature



Telefone: 3029.2115



12/15



PA



...do Brasil, OND, DF

...do Brasil, OND, DF



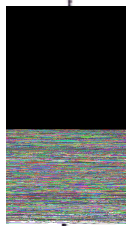
ca



de 2001



Coordenador da OND, OND



Anexo 02. Haplótipos dos microssatélites TNF encontrados em Brasília, bem como o número de cromossomos encontrados com cada haplótipo.

Haplótipo	TNFa	TNFb	TNFc	TNFd	TNFe	Nº de Cromossomos
h1	6	5	1	4	3	3
h2	7	1	1	4	1	1
h3	8	4	1	3	3	5
h4	10	4	1	3	3	2
h5	2	5	1	3	3	9
h6	6	4	1	4	3	5
h7	7	1	1	5	2	1
h8	10	4	1	4	3	2
h9	7	5	1	4	3	1
h10	8	4	1	4	3	3
h11	10	4	1	5	3	2
h12	10	4	2	4	3	2
h13	9	4	1	4	3	1
h14	11	4	2	4	3	1
h15	2	1	1	2	3	3
h16	8	4	1	3	1	2
h17	7	1	1	2	1	1
h18	8	3	1	4	3	2
h19	9	5	1	2	3	1
h20	5	1	1	2	3	2
h21	5	4	1	2	2	1
h22	6	3	1	2	3	2
h23	3	6	1	2	3	1
h24	4	4	1	3	3	11
h25	2	4	1	3	3	5
h26	3	4	1	1	3	1
h27	7	4	1	2	3	1
h28	5	3	1	1	3	1
h29	2	5	1	2	3	2
h30	3	3	1	2	3	3
h31	7	3	1	1	3	1
h32	8	3	1	5	1	1

Haplótipo	TNFa	TNFb	TNFc	TNFd	TNFe	Nº de Cromossomos
h33	10	2	1	4	1	1
h34	10	3	1	3	2	1
h35	2	5	1	4	3	3
h36	7	4	1	4	3	2
h37	5	1	1	5	2	1
h38	5	3	1	4	1	2
h39	2	1	1	5	3	1
h40	2	5	2	6	1	1
h41	3	4	1	2	1	1
h42	13	5	2	3	3	1
h43	14	6	2	3	3	1
h44	3	4	1	3	3	4
h45	5	3	1	3	3	2
h46	5	5	1	4	3	1
h47	4	5	1	3	3	2
h48	14	5	1	2	3	2
h49	1	5	1	3	3	2
h50	4	2	1	2	3	1
h51	4	5	1	4	3	5
h52	7	3	1	4	3	2
h53	3	4	1	4	3	2
h54	1	5	1	4	3	1
h55	2	4	2	1	3	1
h56	7	1	1	1	3	3
h57	7	4	1	3	3	2
h58	4	1	2	3	3	1
h59	4	5	1	4	1	2
h60	1	1	1	6	3	1
h61	3	3	2	1	3	1
h62	5	2	1	3	3	1
h63	5	4	1	2	3	1
h64	2	5	1	3	1	1
h65	13	1	1	4	3	1
h66	13	6	2	3	3	1
h67	14	1	1	4	3	1
h68	3	5	1	6	3	4
h69	4	3	1	4	3	10
h70	5	3	1	4	3	2
h71	4	1	1	3	3	2
h72	6	3	1	4	3	3

Haplótipo	TNFa	TNFb	TNFc	TNFd	TNFb	Nº de Cromossomos
h73	9	2	2	3	3	1
h74	7	3	1	2	3	1
h75	3	1	2	3	1	1
h76	2	2	1	1	3	1
h77	3	5	2	4	3	2
h78	3	2	2	4	3	1
h79	3	5	1	3	3	1
h80	3	1	2	3	3	3
h81	7	2	1	5	3	1
h82	7	1	1	2	3	2
h83	1	5	1	2	3	1
h84	4	1	1	4	3	1
h85	4	4	1	5	3	4
h86	3	3	1	4	3	3
h87	3	5	1	4	3	2
h88	4	3	1	2	2	1
h89	4	4	1	2	3	1
h90	3	5	2	7	2	1
h91	4	3	1	2	3	1
h92	13	1	1	5	1	1
h93	4	1	1	4	1	1
h94	4	5	1	5	3	4
h95	2	1	1	6	3	2
h96	9	4	1	5	3	1
h97	2	1	1	5	1	1
h98	3	7	1	4	3	1
h99	11	4	1	5	3	1
h100	11	5	1	6	3	1
h101	3	4	1	1	1	1
h102	4	4	1	2	1	1
h103	4	5	1	2	3	1
h104	7	2	2	2	3	1
h105	3	7	1	2	3	1
h106	3	4	1	2	2	1
h107	3	3	1	4	1	1
h108	2	3	1	3	3	1
h109	2	6	1	3	3	1
h110	2	4	1	4	3	1
h111	2	6	1	5	3	1
h112	6	5	1	3	3	1

Anexo 03. Haplótipos dos microssatélites TNF encontrados em Kalunga.

Haplótipos	TNFa	TNFb	TNFc	TNFd	TNFb	Nº de Cromossomos
h5	2	5	1	3	3	2
h19	9	5	1	2	3	1
h22	6	3	1	2	3	2
h26	3	4	1	1	3	1
h27	7	4	1	2	3	1
h29	2	5	1	2	3	1
h31	7	3	1	1	3	1
h56	7	1	1	1	3	2
h58	4	1	2	3	3	1
h59	4	5	1	4	1	1
h64	2	5	1	3	1	1
h67	14	1	1	4	3	1
h68	3	5	1	6	3	1
h74	7	3	1	2	3	3
h78	3	2	2	4	3	1
h80	3	1	2	3	3	1
h82	7	1	1	2	3	2
h95	2	1	1	6	3	2
h101	3	4	1	1	1	1
h104	7	2	2	2	3	1
h111	2	6	1	5	3	2
h113	6	5	1	1	3	1
h114	7	5	2	2	3	5
h115	5	7	1	2	3	1
h116	11	4	1	2	3	1
h117	3	5	1	2	3	1
h118	9	7	1	4	3	2
h119	7	5	2	6	1	2
h120	6	5	2	6	3	1
h121	10	3	2	5	3	1
h122	6	4	2	1	2	1
h123	9	5	1	1	2	1
h124	6	4	2	4	2	1
h125	14	1	2	2	1	2
h126	4	5	2	3	3	1
h127	11	4	2	2	3	1
h128	1	6	1	6	3	1
h129	7	3	1	3	1	2

Haplótipos	TNFa	TNFb	TNFc	TNFd	TNFb	Nº de Cromossomos
h130	6	7	2	3	3	3
h131	8	4	2	2	3	1
h132	2	6	2	2	3	1
h133	7	4	2	2	3	1
h134	14	7	2	3	3	1
h135	5	1	2	2	1	1
h136	6	5	2	1	3	1
h137	14	2	1	2	3	1
h138	1	3	1	2	3	1
h139	2	6	1	1	3	1
h140	7	6	1	2	3	1
h141	2	7	2	2	3	2
h142	3	2	2	1	3	3
h143	6	6	2	3	3	1
h144	14	2	2	6	2	1
h145	2	6	1	3	2	1
h146	2	4	1	1	1	1
h147	6	7	2	1	3	1
h148	7	4	2	6	3	1
h149	7	7	1	1	3	2
h150	14	5	2	2	3	2
h151	7	2	2	3	1	1
h152	2	6	1	1	1	1
h153	9	4	1	2	3	1
h154	2	3	2	1	2	1
h155	1	1	1	3	3	1
h156	3	4	2	1	3	1
h157	6	7	2	2	3	1
h158	14	6	2	2	3	1
h159	6	7	1	2	2	1
h160	2	5	2	2	3	2
h161	3	1	2	2	3	1
h162	3	6	2	2	3	1
h163	9	5	2	4	3	1
h164	1	6	2	5	3	1
h165	4	7	2	1	2	2
h166	2	3	2	4	3	1
h167	14	2	2	6	3	1
h168	1	7	2	3	3	1
h169	7	1	2	2	1	1
h170	14	1	1	3	1	1
h171	14	4	1	3	1	1
h172	3	5	2	6	3	1

Haplótipos	TNFa	TNFb	TNFc	TNFd	TNFb	Nº de Cromossomos
h173	7	7	1	2	3	1
h174	2	1	2	2	3	2
h175	2	4	2	2	3	1
h176	7	4	2	6	1	2
h177	2	1	2	3	1	2
h178	7	7	2	2	1	1

Haplótipos	TNFa	TNFb	TNFc	TNFd	TNFb	Nº de Cromossomos
h179	3	3	2	1	1	1
h180	2	5	2	4	3	1
h181	14	5	1	3	3	1
h182	14	7	1	6	3	1
h183	3	2	1	1	3	1

Anexo 04. Lista de haplótipos para todos os marcadores encontrados em Brasília.

Haplótipo	TNFa	TNFb	TNFe	TNFd	TNFe	AluMICB	AluTF	AluHJ	AluHG	AluHF	N.º Cromossomos
h1	6	5	1	4	3	1	2	1	2	2	1
h2	7	1	1	4	1	2	2	2	2	2	1
h3	8	4	1	3	3	2	2	2	2	2	2
h4	10	4	1	3	3	1	1	2	2	1	1
h5	2	5	1	3	3	1	1	1	2	2	1
h6	6	4	1	4	3	2	2	2	1	1	1
h7	6	1	1	5	2	2	1	2	2	2	1
h8	7	5	1	4	3	2	1	2	2	2	1
h9	8	4	1	3	3	1	2	2	2	1	1
h10	10	4	1	4	3	2	2	2	1	1	1
h11	7	5	1	4	3	2	2	2	2	2	1
h12	10	4	1	3	3	1	2	2	2	1	1
h13	2	5	1	3	3	2	2	1	2	2	3
h14	8	4	1	4	3	2	2	2	2	1	1
h15	10	4	1	5	3	1	2	1	2	2	1
h16	10	4	2	4	3	2	2	2	2	2	2
h17	9	4	1	4	3	2	2	1	2	2	1
h18	8	4	1	4	3	1	2	2	2	2	1
h19	11	4	2	4	3	2	1	2	1	1	1
h20	2	1	1	2	3	2	2	2	1	1	1
h21	8	4	1	3	1	2	2	2	1	1	1
h22	2	5	1	3	3	2	2	2	2	1	1
h23	7	1	1	2	1	2	2	1	2	2	1
h24	8	3	1	4	3	1	2	2	2	2	1
h25	9	5	1	2	3	2	1	2	1	2	1
h26	5	1	1	2	3	2	2	2	2	2	1
h27	5	4	1	2	2	2	1	2	2	2	1
h28	6	3	1	2	3	2	2	1	2	2	1
h29	10	4	1	4	3	2	2	2	1	2	1
h30	3	6	1	2	3	2	2	1	2	2	1
h31	4	4	1	3	3	2	2	1	2	2	3
h32	2	1	1	2	3	2	2	2	2	2	1
h33	2	4	1	3	3	2	2	2	1	2	1
h34	3	4	1	1	3	2	2	2	2	2	1
h35	7	4	1	2	3	2	1	2	1	1	1
h36	5	3	1	1	3	1	1	1	2	2	1
h37	6	3	1	2	3	2	2	2	1	1	1
h38	2	5	1	2	3	2	2	2	2	2	1
h39	2	5	1	3	3	2	2	2	1	2	1
h40	2	5	1	3	3	1	2	2	2	2	1
h41	3	3	1	2	3	2	2	2	1	2	1

Haplótipo	TNFa	TNFb	TNFc	TNFd	TNFe	AluMICB	AluTF	AluHJ	AluHG	AluHF	N.º Cromossomos
h42	7	3	1	1	3	1	2	2	2	2	1
h43	8	3	1	5	1	2	2	2	1	1	1
h44	10	2	1	4	1	1	1	2	2	2	1
h45	10	3	1	3	2	2	2	2	2	2	1
h46	2	5	1	4	3	2	2	2	2	2	1
h47	7	4	1	4	3	2	2	2	2	2	2
h48	5	1	1	5	2	2	2	1	2	2	1
h49	5	3	1	4	1	2	2	1	2	2	1
h50	2	1	1	5	3	2	2	1	2	2	1
h51	2	5	2	6	1	2	2	2	2	2	1
h52	3	4	1	2	1	2	2	1	2	2	1
h53	13	5	2	3	3	2	2	2	2	2	1
h54	14	6	2	3	3	2	2	2	2	2	1
h55	3	4	1	3	3	2	2	2	2	2	2
h56	5	3	1	3	3	2	2	1	2	2	1
h57	5	5	1	4	3	2	1	2	1	2	1
h58	4	5	1	3	3	2	2	2	2	2	1
h59	8	3	1	4	3	2	2	2	1	2	1
h60	14	5	1	3	3	2	2	2	2	2	1
h61	1	5	1	2	3	2	1	2	2	2	1
h62	4	2	1	2	3	2	2	1	2	2	1
h63	4	5	1	4	3	2	2	2	1	2	1
h64	2	4	1	3	3	2	2	2	2	2	1
h65	7	3	1	4	3	2	1	2	1	1	1
h66	3	4	1	4	3	2	1	2	2	1	1
h67	1	5	1	4	3	2	2	1	2	2	1
h68	2	4	2	1	3	2	1	1	2	1	1
h69	7	1	1	1	3	2	2	2	2	2	1
h70	7	4	1	3	3	2	1	2	2	1	1
h71	4	1	2	3	3	1	2	2	2	1	1
h72	4	5	1	4	1	1	1	2	1	1	1
h73	1	1	1	6	3	2	2	2	2	2	1
h74	3	3	2	1	3	2	1	2	2	2	1
h75	5	2	1	2	3	2	2	1	2	2	1
h76	5	4	1	3	3	2	1	2	2	1	1
h77	14	5	1	2	3	2	2	2	2	2	1
h78	8	4	1	3	3	2	1	2	1	1	1
h79	2	5	1	4	1	2	1	2	2	1	1
h80	13	1	1	4	3	2	2	2	2	2	1
h81	13	6	2	3	3	2	1	2	1	2	1
h82	14	1	1	4	3	2	2	2	2	2	1
h83	3	5	1	6	3	2	1	2	2	2	1
h84	4	3	1	4	3	2	2	2	2	2	2
h85	4	4	1	3	3	2	1	2	2	1	1

Haplótipo	TNFa	TNFb	TNFc	TNFd	TNFe	AluMICB	AluTF	AluHJ	AluHG	AluHF	N.º Cromossomos
h86	2	5	1	2	3	2	2	2	1	2	1
h87	2	5	1	3	3	2	1	2	1	1	1
h88	3	5	1	6	3	1	2	2	2	2	1
h89	5	3	1	3	3	2	1	2	1	1	1
h90	4	3	1	4	3	1	2	2	2	2	4
h91	8	4	1	4	3	2	1	2	1	1	1
h92	5	3	1	4	3	2	2	1	2	2	1
h93	8	4	1	3	3	2	1	2	1	2	1
h94	4	1	1	3	3	2	2	1	2	2	1
h95	4	3	1	4	3	2	2	1	2	2	2
h96	2	4	1	3	3	2	2	2	2	1	1
h97	4	4	1	3	3	2	2	2	1	1	1
h98	6	2	2	3	3	2	2	2	2	2	1
h99	9	3	1	4	3	2	1	2	1	2	1
h100	4	4	1	3	3	2	2	2	2	2	3
h101	7	3	1	2	3	1	1	2	2	1	1
h102	3	1	2	3	3	2	2	1	2	2	2
h103	3	4	1	4	1	2	2	2	2	1	1
h104	4	5	1	4	1	2	2	1	2	2	1
h105	2	2	1	1	3	2	2	2	2	2	1
h106	2	5	1	3	3	2	1	2	2	2	1
h107	1	5	1	3	3	2	2	1	2	2	1
h108	3	5	2	4	3	2	1	2	2	2	1
h109	3	2	2	4	3	2	2	2	2	2	1
h110	3	5	1	3	3	2	2	2	2	2	1
h111	3	1	2	3	3	2	2	2	2	2	2
h112	6	3	1	4	3	2	2	2	2	2	2
h113	7	2	1	5	3	2	2	2	1	1	1
h114	4	3	1	4	3	1	1	1	2	2	1
h115	7	1	1	2	3	2	1	2	1	2	1
h116	3	5	2	4	3	2	2	2	1	2	1
h117	1	5	1	2	3	2	2	1	2	2	1
h118	10	4	1	5	3	2	2	2	2	1	1
h119	4	1	1	4	3	2	2	1	2	2	1
h120	2	4	1	3	3	2	2	1	2	2	1
h121	2	1	1	2	3	1	2	1	2	2	1
h122	4	4	1	5	3	1	1	2	2	1	1
h123	4	5	1	3	3	2	1	1	1	2	1
h124	4	5	1	4	3	2	1	2	2	2	1
h125	4	4	1	5	3	1	1	2	2	2	1
h126	3	4	1	3	3	1	1	2	2	1	1
h127	7	3	1	4	3	2	2	2	2	1	1
h128	3	3	1	2	3	2	2	2	2	2	1
h129	3	5	1	4	3	2	2	2	2	2	1

Haplótipo	TNFa	TNFb	TNFc	TNFd	TNFe	AluMICB	AluTF	AluHJ	AluHG	AluHF	N.º Cromossomos
h174	4	4	1	3	3	1	2	2	2	2	1
h175	4	5	1	5	3	1	2	2	2	1	1
h176	4	5	1	4	3	1	2	2	2	2	1
h177	4	5	1	4	3	2	2	2	2	2	1
h178	3	3	1	4	1	1	2	2	2	1	1
h179	3	3	1	4	3	1	1	2	1	1	1
h180	3	3	1	4	3	1	2	1	2	2	1
h181	2	3	1	3	3	2	2	2	2	2	1
h182	2	6	1	3	3	2	2	2	2	2	1
h183	2	1	1	6	3	2	2	1	2	2	1
h184	2	4	1	4	3	2	1	2	1	1	1
h185	2	5	1	4	3	2	2	1	2	2	1
h186	2	6	1	5	3	2	2	2	2	2	1
h187	5	3	1	4	1	1	2	1	2	2	1
h188	6	4	1	4	3	1	2	1	1	1	1
h189	6	4	1	4	3	1	2	2	2	2	1
h190	6	5	1	3	3	2	2	2	1	1	1
h191	2	5	1	3	3	1	2	2	1	1	1

Anexo 05. Lista dos haplótipos para todos os marcadores encontrados na população de Kalunga.

Haplótipo	TNFa	TNFb	TNFc	TNFd	TNFe	AluMICB	AluTF	AluHJ	AluHG	AluHF	N.º Cromossomos
h25	9	5	1	2	3	2	1	2	1	2	1
h34	3	4	1	1	3	2	2	2	2	2	1
h37	6	3	1	2	3	2	2	2	1	1	1
h38	2	5	1	2	3	2	2	2	2	2	1
h109	3	2	2	4	3	2	2	2	2	2	1
h111	3	1	2	3	3	2	2	2	2	2	1
h131	7	1	1	2	3	2	2	2	2	2	2
h138	4	4	1	2	3	2	2	2	2	2	1
h156	3	4	1	1	1	2	2	2	2	2	1
h192	6	5	1	1	3	2	2	2	2	1	1
h193	7	5	2	2	3	2	2	2	1	1	1
h194	5	7	1	2	3	1	2	2	2	2	1
h195	11	4	1	2	3	2	2	2	1	1	1
h196	3	5	1	2	3	2	2	1	2	2	1
h197	9	7	1	4	3	2	2	2	1	2	1
h198	4	5	1	4	1	2	2	2	2	2	1
h199	7	5	2	6	1	2	1	2	2	1	1
h200	6	5	2	6	3	2	2	2	2	2	1
h201	10	3	2	5	3	2	2	2	2	2	1
h202	6	4	2	1	2	2	2	1	2	2	1
h203	9	5	1	1	2	2	1	1	1	2	1
h204	6	5	2	4	2	1	2	2	2	2	1
h205	7	4	2	6	1	2	1	2	1	1	1
h206	14	1	2	2	1	2	2	2	2	2	1
h207	7	4	1	2	3	2	2	2	2	2	1
h208	4	5	2	3	3	2	2	2	2	2	1
h209	11	4	2	2	3	2	2	2	1	2	1
h210	1	6	1	3	1	2	2	1	2	2	1
h211	7	3	1	6	3	2	2	2	2	2	1
h212	2	5	1	3	3	1	2	1	2	2	1
h213	6	7	2	3	3	2	2	2	1	2	1
h214	7	5	2	2	3	1	2	2	2	2	1
h215	8	4	2	2	3	2	2	2	1	2	1
h216	2	6	2	2	3	2	2	2	2	2	1
h217	7	4	2	2	3	2	2	2	2	2	1
h218	14	7	1	6	3	1	2	1	2	2	1
h219	2	1	2	3	3	2	2	2	2	2	1
h220	5	1	2	2	1	2	2	2	2	2	1
h221	6	5	2	1	3	2	2	2	1	1	1
h222	14	2	1	2	3	1	2	2	2	2	1

Haplótipo	TNFa	TNFb	TNFc	TNFd	TNFe	AluMICB	AluTF	AluHJ	AluHG	AluHF	N.º Cromossomos
h223	1	3	1	2	3	2	2	2	1	2	1
h224	2	6	1	1	3	1	2	1	2	2	1
h225	7	6	1	2	3	2	2	2	2	2	1
h226	2	7	2	2	3	1	2	2	2	2	1
h227	3	2	2	1	3	2	1	2	2	1	1
h228	4	1	2	3	3	1	2	1	2	2	1
h229	6	6	2	3	3	2	1	2	2	2	1
h230	2	5	1	3	1	2	2	2	2	2	1
h231	2	7	2	2	3	2	2	2	1	2	1
h232	14	2	2	6	2	2	2	1	2	2	1
h233	2	5	1	3	3	2	2	1	2	1	1
h234	2	6	1	3	2	2	2	2	2	1	1
h235	2	4	1	1	1	2	2	2	2	2	1
h236	6	7	2	1	3	2	1	2	1	1	1
h237	7	4	2	6	3	2	2	2	1	1	1
h238	7	3	1	2	3	1	2	2	2	2	1
h239	7	7	1	1	3	1	2	2	2	2	1
h240	7	1	1	1	3	2	2	1	2	2	1
h241	7	1	1	1	3	1	2	2	2	2	1
h242	14	5	2	2	3	2	2	1	2	2	1
h243	7	3	1	2	3	2	2	2	2	2	1
h244	7	2	2	3	1	2	2	1	2	2	1
h245	7	5	2	2	3	2	2	2	2	2	2
h246	14	1	1	4	1	2	2	2	2	2	1
h247	2	6	1	1	3	1	2	2	2	2	1
h248	9	4	1	2	3	2	2	2	2	2	1
h249	2	3	2	1	2	2	2	2	2	2	1
h250	7	3	1	2	3	2	2	2	1	2	1
h251	1	1	1	3	3	1	2	2	2	2	1
h252	6	3	1	2	3	2	2	2	1	2	1
h253	3	4	2	2	3	1	2	2	2	2	1
h254	6	7	2	1	3	2	1	2	1	2	1
h255	14	6	2	2	3	1	2	2	2	2	1
h256	6	7	1	2	2	2	1	2	1	2	1
h257	2	5	2	2	3	2	2	1	1	2	1
h258	3	2	2	1	3	2	1	1	1	2	1
h259	3	1	2	2	3	2	2	2	2	2	1
h260	3	6	2	2	3	2	2	2	1	2	1
h261	9	5	2	4	3	2	2	2	1	2	1
h262	9	7	1	4	3	2	1	1	2	2	1
h263	3	2	2	1	3	1	2	2	2	2	1
h264	7	5	2	2	3	2	2	2	1	2	1
h265	14	1	2	2	1	1	2	2	2	2	1

Haplótipo	TNFa	TNFb	TNFc	TNFd	TNFe	AluMICB	AluTF	AluHJ	AluHG	AluHF	N.º Cromossomos
h266	1	6	2	5	3	1	1	2	2	2	1
h267	4	7	2	1	2	2	2	2	2	2	1
h268	4	7	2	1	2	2	2	2	1	2	1
h269	2	3	2	4	3	2	2	2	2	2	1
h270	2	6	1	5	3	2	2	2	1	2	2
h271	14	2	2	6	3	2	2	2	2	2	1
h272	1	7	2	3	3	2	2	2	2	2	1
h273	7	1	2	2	1	1	2	2	2	2	1
h274	7	3	1	3	1	2	1	2	1	2	1
h275	3	5	1	6	3	1	1	2	2	2	1
h276	6	7	2	3	3	2	2	2	2	2	2
h277	14	1	1	3	1	2	2	1	2	2	1
h278	14	4	1	3	1	2	2	2	1	1	1
h279	3	5	2	6	3	1	2	2	2	1	1
h280	7	2	2	2	3	2	2	2	1	2	1
h281	7	7	1	1	3	2	2	2	2	2	1
h282	7	7	1	2	3	2	2	2	1	2	1
h283	2	1	2	2	3	2	2	2	2	2	1
h284	2	4	2	2	3	2	2	2	1	2	1
h285	7	4	2	6	1	2	1	2	2	2	1
h286	2	1	2	3	1	2	2	1	2	2	1
h287	2	1	2	2	3	1	2	2	2	2	1
h288	7	3	1	1	3	2	1	2	1	2	1
h289	7	4	2	6	1	2	2	2	2	2	1
h290	2	1	2	3	1	2	2	2	2	2	1
h291	2	5	2	2	3	2	2	2	2	2	1
h292	7	7	2	2	1	2	2	2	1	2	1
h293	3	3	2	1	1	2	2	2	2	2	1
h294	14	5	2	2	3	2	2	2	1	2	1
h295	2	5	2	4	3	2	2	2	1	1	1
h296	14	5	1	3	3	2	2	1	2	2	1
h297	14	7	1	6	3	1	2	2	2	2	1
h298	3	2	1	1	3	1	2	2	2	2	1