

**UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO  
DEPARTAMENTO DE GENÉTICA**

**DESEQUILÍBRIO DE LIGAÇÃO E BLOCOS DE HAPLÓTIPOS  
DETERMINADOS PELA ANÁLISE DE 250K SNPS EM TRÊS  
REMANESCENTES DE QUILOMBOS**

**EDILENE SANTOS DE ANDRADE**

**Ribeirão Preto  
2013**

**UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO  
DEPARTAMENTO DE GENÉTICA**

**DESEQUILÍBRIO DE LIGAÇÃO E BLOCOS DE HAPLÓTIPOS  
DETERMINADOS PELA ANÁLISE DE 250K SNPS EM TRÊS  
REMANESCENTES DE QUILOMBOS**

**EDILENE SANTOS DE ANDRADE**

Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, como requisito parcial para obtenção do título de Doutora em Ciências – Área de concentração: Genética.

Orientação: Prof. Dr. Celso Teixeira Mendes-Junior.

**Ribeirão Preto**

**2013**

**AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.**

### **FICHA CATALOGRÁFICA**

ANDRADE, Edilene Santos

Desequilíbrio de ligação e blocos de haplótipos determinados pela análise de 250K SNPs em três remanescentes de quilombos / Edilene Santos de Andrade; orientador Celso Teixeira Mendes-Junior. Ribeirão Preto, 2013.

147p.: 30cm

Tese (Doutorado – Programa de Pós-Graduação em Genética) - Faculdade de Medicina de Ribeirão – Universidade de São Paulo.

1. Desequilíbrio de ligação. 2. Blocos de haplótipos. 3. SNPs. 4. Remanescentes de quilombos. 5. História demográfica.

## FOLHA DE APROVAÇÃO

Edilene Santos de Andrade

Desequilíbrio de ligação e blocos de haplótipos determinados pela análise de 250K SNPs em três remanescentes de quilombos

Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, como requisito parcial para obtenção do título de Doutora em Ciências – Área de concentração: Genética.

Orientação: Prof. Dr. Celso Teixeira Mendes-Junior.

Aprovada em:

Banca examinadora

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_ Assinatura: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_ Assinatura: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_ Assinatura: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_ Assinatura: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_ Assinatura: \_\_\_\_\_



Ribeirão Preto, 02 de agosto de 2012

Ofício nº 2739/2012  
CEP/MGV

**Prezados Senhores,**

O trabalho intitulado “**PADRÕES DE DESEQUILÍBRIO DE LIGAÇÃO EM POPULAÇÕES REMANESCENTES DE QUILOMBOS DO ESTADO DO PIAUÍ**”, foi analisado pelo Comitê de Ética em Pesquisa, em sua 350ª Reunião Ordinária realizada em 30/07/2012, e enquadrado na categoria: **APROVADO**, de acordo com o Processo HCRP nº 4788/2012.

*Este Comitê segue integralmente a Conferência Internacional de Harmonização de Boas Práticas Clínicas (IGH-GCP), bem como a Resolução nº 196/96 CNS/MS.*

*Lembramos que devem ser apresentados a este CEP, o Relatório Parcial e o Relatório Final da pesquisa.*

Atenciosamente.

**DR<sup>ª</sup>. MARCIA GUIMARÃES VILLANOVA**  
**Coordenadora do Comitê de Ética em**  
**Pesquisa do HCRP e da FMRP-USP**

Ilustríssimos Senhores  
**EDILENE SANTOS DE ANDRADE**  
**PROF.DR. AGUINALDO LUIZ SIMÕES**  
Depto. de Genética – FMRP-USP

*Dedico este trabalho aos meus pais,  
Macrino e Maria Helena, pelo apoio  
incondicional!*

## **AGRADECIMENTOS**

A Deus, por ter renovado a minha fê e as minhas forças diante de todas as dificuldades que surgiram durante o desenvolvimento deste trabalho.

Aos meus pais, Macrino e Maria Helena, às minhas irmãs, Eliane e Etiene, e aos demais familiares que sempre me incentivaram e oraram por mim.

Ao Prof. Dr. Celso Teixeira Mendes-Junior pela orientação, pela oportunidade de desenvolver projetos na área de Genética Forense (minha grande paixão) e, sobretudo, pelo voto de confiança concedido quando aceitou me orientar e concordou com a continuidade deste trabalho.

Ao Prof. Dr. Aguinaldo Luiz Simões pela orientação no início deste trabalho, pelas muitas discussões que me permitiram adquirir um maior conhecimento dos fundamentos teóricos relativos a esta tese e por ter concordado com a continuidade deste trabalho, apesar de todas as dificuldades que surgiram ao longo do caminho.

Ao colega bioinformata Diego Martinez Salvanha, pelo desenvolvimento dos scripts que me permitiram analisar os dados, sem os quais a realização deste trabalho teria sido inviável.

Ao Prof. Dr. Ricardo Zorzetto Nicolliello Vêncio pela valiosa colaboração neste trabalho e por abrir as portas de seu grupo de pesquisa, o que despertou o meu interesse pela Bioinformática. Agradeço, sobretudo, pelo apoio e incentivo concedidos, não apenas durante o desenvolvimento deste trabalho, mas também em relação a projetos futuros.

Ao Prof. Dr. Henrique Krieger pela simpatia com a qual me recebeu em seu Laboratório para a realização da etapa experimental. Ao colega Lucas Pereira e, principalmente, ao Leandro Maza Garrido pela paciência e boa vontade durante o meu período de treinamento no sistema de genotipagem dos SNPs.

Aos professores membros da banca examinadora, pela disponibilidade em avaliar este trabalho e pelas valiosas contribuições.

Às amigas Ana Lúcia Pimentel e Maria do Carmo T. Canas, pelo inestimável auxílio laboratorial prestado, principalmente à Ana Lúcia por todo o tempo despendido e por sua persistência quando me auxiliou na quase impossível tarefa de extrair DNA de boa qualidade das amostras estudadas. Agradeço também as duas pela amizade, pelo apoio nos momentos difíceis, pelos muitos momentos de descontração e por estarem sempre dispostas a me ajudar, dentro e fora do Laboratório.

À amiga Edna Maria Pereira pelos muitos favores prestados e, principalmente, pelos momentos que rimos e choramos juntas, pelo ombro amigo e sábios conselhos que me ajudaram a superar os momentos de maior dificuldade.

À amiga e conterrânea Fernanda Carvalho pela paciência em passar quatro anos ouvindo as minhas lamentações e por ter sempre me incentivado a continuar. Agradeço pela amizade e por todo apoio, sobretudo durante a finalização da tese.

À Juliana Feres pela amizade, vários momentos de descontração e por ter sido em alguns momentos minha terapeuta, me ensinando a desenvolver paciência e a controlar minhas reações, que contribuiu muito para a minha formação geral.



A todos os amigos que fazem ou fizeram parte do Departamento de Genética: Juliana Massaro, Cláudia Wiesel, Natália, Cláudia Caixeta, Nádia, Lídia, Leonardo, Marcelo, Paulo, Rosana, Daniela, Ludmila, Marcela e Rômulo. Obrigada pelo sempre animado ambiente de trabalho, pela amizade e pelo auxílio que me ofereceram, laboratorial ou por meio de proveitosas conversas (acadêmicas e pessoais).

Às amigas e companheiras de futebol Patrícia e Tássia, por me hospedado em São Paulo durante o meu treinamento na genotipagem dos SNPs.

À amiga Adriana Vieira, pela revisão do texto e por todo apoio e incentivo, mesmo à distância.

Às amigas pernambucanas, Glória Raposo e Karina Alves, que sempre estiveram dispostas a ouvir meus problemas e me incentivaram a continuar.

Aos funcionários do Departamento de Genética, em especial às secretárias Susie e Sílvia, que sempre estiveram dispostas a me ajudar.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida para o desenvolvimento desta tese.

A todas as pessoas que de alguma forma contribuíram para a realização deste trabalho, MUITO OBRIGADA!

*“Pois existe a trajetória, e a trajetória não é apenas um modo de ir. A trajetória somos nós mesmos”.*

(Clarice Lispector)

## RESUMO

ANDRADE, E. S. **Desequilíbrio de ligação e blocos de haplótipos determinados pela análise de 250K SNPs em três remanescentes de quilombos.** 2013. 147p. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto.

A associação não aleatória entre alelos de diferentes loci caracteriza o que é chamado de desequilíbrio de ligação (DL) entre eles. A extensão do DL nas populações humanas pode ser influenciada por muitos fatores, tais como taxa de recombinação, características demográficas (idade, tamanho e taxa de crescimento) e fatores evolutivos (deriva genética, efeito fundador, gargalos populacionais, mutação, seleção e fluxo gênico). Portanto, o conhecimento dos padrões do DL fornecem dados que auxiliam na descrição dos eventos demográficos e evolutivos sofridos pelas populações. O objetivo deste estudo foi descrever os padrões de DL de quatro populações brasileiras e correlacioná-los com suas respectivas histórias demográficas, uma vez que estas populações experimentaram alguns dos eventos evolutivos que geram ou retardam o decréscimo do DL, como fundação por poucos indivíduos, miscigenação no momento da fundação e posterior isolamento. Foram analisadas amostras de três populações remanescentes de quilombos do Estado do Piauí, Gaucinha (GAU,  $n = 14$ ), Mimbó (MIB,  $n = 15$ ) e Sítio Velho (STV,  $n = 15$ ) e da população urbana de Teresina, Piauí (TES,  $n = 15$ ), além de sete amostras populacionais do projeto HapMap (CEU, CHB, JPT, ASW, LWK, MKK, YRI, todas com  $n = 15$ ). Foram genotipados mais de 250 mil SNPs (*Single Nucleotide Polymorphisms*) utilizando-se o *GeneChip® Human Mapping 250K Nsp I Array - Affymetrix®* nas amostras das quatro populações brasileiras. Os dados brutos das populações do HapMap para este *array* foram obtidos na página do projeto. Os genótipos para todas as amostras foram determinados pelo algoritmo CRLMM após comparação com o algoritmo BRLMM, e as análises de DL e determinação dos blocos de haplótipos

foram realizadas com o uso do programa Haploview. Considerando-se o número de blocos de haplótipos detectados em cada população estudada, padrão semelhante foi observado em todos os autossomos. Em geral, a população europeia (CEU) e as duas populações asiáticas (CHB e JPT) do HapMap apresentaram os maiores números de blocos, enquanto que os menores números foram observados nos quilombos GAU e MIB e na população TES. As populações africanas LWK, MKK e YRI e a população afro-americana ASW apresentaram os valores intermediários e a população afro-brasileira STV, apresentou um número de blocos apenas inferior a CEU, CHB e JPT. A grande contribuição africana nos quilombos GAU e MIB pode explicar o menor DL observado nestas comunidades. Por outro lado, o menor DL em TES se deve, provavelmente, à sua fundação, que envolveu um maior número de indivíduos e foi seguida por um rápido crescimento. A possível explicação para o maior DL observado em STV, em relação aos demais quilombos, consiste em sua peculiar história demográfica: esta comunidade experimentou uma miscigenação no momento de sua fundação, que foi seguida por um crescimento lento e pouca diferenciação. Assim, foi demonstrado como os eventos demográficos de cada população influenciam seus respectivos padrões de DL.

**Palavras-Chave:** desequilíbrio de ligação; blocos de haplótipos; SNPs; remanescentes de quilombos; história demográfica.

## ABSTRACT

ANDRADE, E. S. **Linkage disequilibrium and haplotype blocks determined by the analysis of 250K SNPs in three quilombo remnants communities.** 2013. 147p. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto.

The non-random association between alleles of different loci characterizes what is called linkage disequilibrium (LD) between them. The LD extent in human populations can be influenced by many factors, such as recombination rate, demographic features (age, size and growth rate) and evolutionary events (genetic drift, founder effects, population bottlenecks, mutation, selection and gene flow). Therefore, knowledge of the LD patterns provides data that assists in describing the evolutionary and demographic events experienced by populations. The aim of this study was to describe the LD patterns of four Brazilian populations and correlate these patterns with their respective demographic histories, since these populations have experienced some of the evolutionary events that produce or retard the LD decrease, such as foundation by few individuals, admixture at the founding moment and subsequent isolation. Samples from three quilombo remnants populations of the Piauí State, Gaucinha (GAU,  $n = 14$ ), Mimbó (MIB,  $n = 15$ ) and Sítio Velho (STV,  $n = 15$ ) and the urban population of Teresina, Piauí (TES,  $n = 15$ ), and seven population samples from the HapMap Project (CEU, CHB, JPT, ASW, LWK, MKK, YRI, all with  $n = 15$ ) were analyzed. More than 250 thousand SNPs (Single Nucleotide Polymorphisms) were genotyped using the GeneChip® Human Mapping 250K Nsp Array I - Affymetrix® in the samples of the four Brazilian populations. Raw data of the HapMap population samples for this array were obtained from the HapMap homepage. Genotypes for all samples were determined by CRLMM algorithm after comparison with the BRLMM algorithm. LD analyzes and determination of haplotype blocks were performed using the Haploview software. Considering the number

of haplotype blocks detected in each population, a consistent pattern was observed for all autosomes. The European population (CEU) and the two Asian populations (CHB and JPT) of the HapMap showed the highest numbers of blocks, while the lowest numbers were observed in the GAU and MIB quilombos and in the TES population. The African populations, LWK, MKK and YRI, and the African-American ASW exhibited intermediate values and the African-Brazilian population STV, presented a number of blocks smaller than that observed for CEU, CHB and JPT. The great African contribution in the GAU and MIB quilombos may explain the lower LD observed in these communities. On the other hand, the lower LD in TES is probably due to its foundation that involved a larger number of individuals and was followed by a fast growth. A possible explanation for the higher LD observed in STV, compared to other quilombos, consists in its particular demographic history: this community experienced admixture at the time of its foundation, which was followed by slow growth and low differentiation. Thus, it was shown how the demographic events of each population influence their respective LD patterns.

**Keywords:** linkage disequilibrium; haplotype blocks; SNPs; quilombo remnants; demographic history.

# SUMÁRIO

1. INTRODUÇÃO .....	16
1.1 Ligação, Haplótipos e Desequilíbrio de Ligação (DL) .....	16
1.2 Fatores demográficos e evolutivos que geram e/ou alteram o DL.....	17
1.3 Exemplos de análise populacional a partir do DL .....	19
2. HIPÓTESE .....	23
3. OBJETIVO .....	24
3.1 Objetivo Geral.....	24
3.2 Objetivos Específicos .....	24
4. METODOLOGIA .....	25
4.1 Aspectos éticos da pesquisa.....	25
4.2 Amostras populacionais.....	25
4.3 Extração do DNA.....	28
4.4 Genotipagem de SNPs com microarranjos de DNA .....	29
4.4.1 Microarranjos de DNA .....	29
4.4.2 Procedimentos laboratoriais.....	31
4.4.3 Análise dos dados e determinação dos genótipos .....	35
4.5 Análises de desequilíbrio de ligação e blocos de haplótipos .....	40
5. RESULTADOS .....	46
5.1 Comparação entre os algoritmos BRLMM e CRLMM.....	46
5.2 Influências do tamanho amostral e do número de SNPs analisados na estimativa do DL.....	49
5.3 Padrões genômicos de blocos de haplótipos e DL .....	53
5.4 Desequilíbrio de Ligação e Blocos de haplótipos no cromossomo 22.....	55
5.4.1 Análise descritiva .....	55
5.4.2 Compartilhamento de blocos de haplótipos.....	60
5.5 Desequilíbrio de Ligação e Blocos de haplótipos no cromossomo 6.....	65
5.5.1 Análise descritiva .....	65
5.5.1.2 Região do MHC.....	71
5.5.2 Compartilhamento de blocos de haplótipos.....	77
5.6 Análise comparativa dos blocos de haplótipos nos cromossomos 6 e 22 .....	85
6. DISCUSSÃO .....	90
6.1 Comparação entre os algoritmos BRLMM e CRLMM.....	90

6.2 Influências do tamanho amostral e do número de SNPs analisados na estimativa do DL.....	92
6.3 Padrões genômicos de blocos de haplótipos e DL .....	94
6.4 Compartilhamentos de blocos de haplótipos .....	102
6.5 Análise comparativa dos blocos de haplótipos nos cromossomos 6 e 22 .....	105
7. CONCLUSÕES .....	107
REFERÊNCIAS BIBLIOGRÁFICAS.....	109
MANUSCRITO .....	116



# 1. INTRODUÇÃO

## 1.1 Ligação, Haplótipos e Desequilíbrio de Ligação (DL)

A ligação pode ser definida como o fenômeno no qual alelos adjacentes situados em um mesmo segmento cromossômico são transmitidos juntos (em bloco) na meiose com uma frequência maior do que seria esperado pelo acaso. Esse conjunto de alelos ligados é denominado haplótipo.

Considerando-se dois locos ligados (A e B), cada um com dois alelos com frequências  $p_{A1}$ ,  $p_{A2}$ ,  $p_{B1}$ ,  $p_{B2}$ , respectivamente, espera-se, em uma situação de equilíbrio, que o haplótipo  $A1B1$  ocorra com frequência igual ao produto  $p_{A1}p_{B1}$ . As frequências esperadas dos outros haplótipos seriam estimadas de maneira semelhante (Hartl e Clark 1997).

O desequilíbrio de ligação (DL) é definido como a diferença entre as proporções haplotípicas observadas e aquelas esperadas se os alelos segregassem independentemente (Weiss e Clark 2002).

O coeficiente de desequilíbrio (D) ou desvio é uma medida do grau de associação não aleatória entre os alelos de dois locos. Este coeficiente é igual à diferença entre o produto das frequências dos haplótipos em acoplamento e o produto das frequências dos haplótipos em repulsão. A cada geração, devido aos eventos de recombinação, uma proporção de haplótipos em acoplamento se transforma em haplótipos em repulsão e vice-versa, reduzindo assim o DL a cada geração. O DL em uma geração qualquer pode ser calculado pela fórmula:

$$DL = (1 - r)^t DL_0$$

onde  $DL_0$  é o DL inicial,  $t$  é o número de gerações de cruzamento ao acaso (*panmixia*) e  $r$  é a taxa de recombinação. Observe-se que as frequências alélicas permanecem constantes, ao longo das gerações. O DL decairá, tendendo a zero, se nenhum outro processo além da

recombinação estiver agindo na população. Se a taxa de recombinação ( $r$ ) for pequena, o DL decairá lentamente. Quando  $r = 0,5$  (taxa de recombinação máxima, para loci em cromossomos diferentes) o decaimento é rápido, mas o equilíbrio só será atingido após algumas gerações. Adicionalmente, quando dois loci estão distantes entre si, a taxa de recombinação entre eles será mais alta em comparação à taxa dos loci muito próximos entre si (Hartl e Clark 1997).

Consequentemente, é esperado que populações jovens exibam uma maior extensão de DL (maior distância observada entre dois loci em desequilíbrio) do que as populações fundadas há mais tempo, uma vez que, para loci separados por longas distâncias, o número de gerações necessárias para atingir o equilíbrio (DL desfeito) é muito menor. Assim, espera-se um maior número de loci distantes em desequilíbrio nas populações mais jovens.

## **1.2 Fatores demográficos e evolutivos que geram e/ou alteram o DL**

Além da taxa de recombinação, a magnitude e o decaimento do DL podem ser alterados por diversos eventos como miscigenação, deriva genética, efeito fundador, gargalos populacionais, mutação e seleção. Outros fatores, tais como idade da população, tamanho e taxa de crescimento populacional, migração e endocruzamento também podem modificar os padrões de DL.

Eventos de miscigenação no momento da fundação de uma população resultam em um alto e extenso DL (Laan e Pääbo, 1997; Hall *et al.*, 2002; Service *et al.*, 2006; Xu *et al.*, 2007). A quantidade e a extensão do DL criado pela miscigenação dependem da proporção e do nível de diferenças nas frequências alélicas nas duas populações ancestrais, da contribuição a partir de cada população parental, além do tempo decorrido desde a miscigenação (Pfaff *et al.*, 2001).

Outro evento demográfico que influencia os padrões de DL é a deriva genética, que consiste em um dos mais importantes mecanismos

criadores de DL. Quanto menor o tamanho populacional, mais forte será a ação da deriva e, conseqüentemente, maior será o DL criado (Jorde 2000; Jorgensen *et al.*, 2002). Em populações pequenas e estáveis, a magnitude esperada do DL entre dois loci também pode variar significativamente em função da taxa de mutação, visto que a ocorrência de mutações, embora geralmente tenha baixa frequência, pode alterar as frequências alélicas e haplotípicas (Tenesa *et al.*, 2004).

Conforme uma população se expande, os efeitos da deriva genética e das mutações serão menores e, como consequência, a extensão do DL entre todos os pares de loci será diminuída de uma geração para outra. Um rápido crescimento limita o DL em decorrência do aumento da diversidade genética e do aparecimento de recombinantes. Espera-se então encontrar DL alto e extenso em populações isoladas que experimentaram um crescimento populacional recente (após um longo período de tamanho constante) e que foram fundadas por um pequeno número de indivíduos (Slatkin 1994; Kruglyak 1999).

Outro mecanismo criador de DL é a seleção natural. Sob evolução neutra, os novos alelos que surgem em uma população requerem muitas gerações para atingirem uma alta frequência. Durante este período, o DL ao redor destes novos alelos diminuirá em decorrência dos eventos de recombinação. Conseqüentemente, alelos comuns apresentarão DL de curta extensão. Por outro lado, a seleção positiva aumenta a frequência do alelo selecionado em relativamente poucas gerações, reduzindo os efeitos da recombinação, que de outra forma, iria desfazer o haplótipo que contém o alelo selecionado. Adicionalmente, os alelos contíguos e, conseqüentemente, o haplótipo correspondente terão suas frequências aumentadas, devido ao efeito carona, resultando em um extenso DL ao redor do alelo selecionado (Ahmad *et al.* 2003; Walsh *et al.*, 2003; Miretti *et al.*, 2005). Além disto, a seleção epistática para combinações de alelos em dois ou mais loci também pode influenciar os padrões de DL através dos haplótipos (Ahmad *et al.* 2003). Foi sugerido que na região do Complexo Principal

de Histocompatibilidade (MHC), os altos níveis de DL entre os lócus HLA separados por grandes distâncias físicas refletem os efeitos da seleção natural e dos eventos demográficos (Meyer *et al.*, 2006).

A extensão do DL em uma população é determinada, principalmente, pelo seu tamanho efetivo (influenciado fortemente pelo número de fundadores e a taxa de expansão da população) e o tempo decorrido desde a sua fundação. O tamanho efetivo estabelece o número de diferentes haplótipos presentes na população, sendo que em populações pequenas este número é limitado. O tempo, por sua vez, determina o número de oportunidades para que a recombinação entre os marcadores desfça o DL (Service *et al.*, 2006).

De modo geral, conclui-se que eventos que levam à redução da diversidade populacional e da heterozigose mimetizam a diminuição da taxa de recombinação, retardando o decaimento do DL ao longo das gerações. Assim, os padrões das associações alélicas em uma população, em um determinado período, resultam da interação complexa entre muitos fatores evolutivos.

### **1.3 Exemplos de análise populacional a partir do DL**

Uma vez que os fatores evolutivos citados anteriormente alteram o efeito da recombinação, populações submetidas a diferentes eventos histórico-demográficos diferem quanto à intensidade, à extensão e à velocidade do decréscimo do DL. Assim, analisando-se os padrões de DL em diferentes populações, pode-se realizar inferências sobre suas respectivas histórias demográficas.

A partir dos padrões de DL, é possível estimar o tamanho efetivo de uma população. Neste contexto, o tamanho efetivo da população islandesa foi estimado em cinco mil, a partir da determinação do DL em 12 regiões genômicas analisadas em 1.753 indivíduos, utilizando-se 179 microssatélites (Bataillon *et al.*, 2006). O tamanho efetivo populacional de populações estudadas no Projeto HapMap (catálogo das variações

genéticas comuns em humanos, acessível em <http://hapmap.ncbi.nlm.nih.gov/>), também foi determinado a partir do DL. Para a população de ancestralidade europeia (CEU, localizada em Utah, Estados Unidos), a japonesa (JPT) e a chinesa (CHB), o tamanho efetivo estimado foi de, aproximadamente, 3.100, enquanto que o tamanho estimado da população africana (YRI) foi de 7.500 (Tenesa *et al.*, 2007).

A partir dos mapas de DL gerados para os cromossomos 6, 21 e 22 (com dados de 24.940 SNPs - *Single Nucleotide Polymorphisms*), De La Vega e colaboradores (2005) determinaram o tempo efetivo de gargalo para quatro populações. O período estimado de duração do gargalo para as populações caucasiana (43.325 anos), chinesa (44.300 anos) e japonesa (41.039 anos) representam menos da metade do tempo presumido para o evento *out-of-Africa*, sendo consistente com o efeito cumulativo de gargalos subsequentes contribuindo para a criação do DL. O período estimado para a população afro-americana foi de 61.525 anos, refletindo uma história demográfica muito diferente, sendo caracterizada pela miscigenação.

Além da estimativa do tempo efetivo de gargalo, também é possível, a partir dos padrões do DL, realizar inferências relacionadas aos processos de expansão, a idade das populações e a ocorrência de eventos de miscigenação no momento da fundação. Uma análise comparativa da extensão do DL entre 15 STRs (*Short Tandem Repeats*) analisados em amostras da população isolada das Ilhas Faroe e das populações britânica e dinamarquesa, revelou uma maior extensão do DL na população isolada (3,8 Mb), quando comparada às demais (1,4 Mb na britânica e 1,2 Mb na dinamarquesa). Esta maior extensão do DL na população isolada foi atribuída ao seu antigo e lento crescimento, bem como a um maior efeito da deriva genética (Jorgensen *et al.*, 2002).

A determinação dos padrões do DL entre 2.486 SNPs do cromossomo 22 (34,2 Mb) de 11 populações isoladas de várias regiões do mundo (200

indivíduos de cada população) e da população CEU (Projeto HapMap), permitiu correlacionar o DL aos eventos demográficos ocorridos em cada população. O DL foi maior nas populações isoladas quando comparadas à população CEU (Service *et al.*, 2006). O perfil do DL de 771 SNPs do cromossomo 22 na população de Talana (101 indivíduos não relacionados), região da Sardenha, também foi comparado ao DL da população CEU. Talana apresentou níveis maiores e mais extensos de DL do que a população europeia, provavelmente devido a um evento fundador antigo, seguido por um período de tamanho populacional relativamente constante (Angius *et al.*, 2008).

A partir de uma análise em todo o genoma, utilizando-se 360 mil SNPs, foi estabelecido o DL de oito vilas isoladas da região de Ogliastra, Sardenha. O padrão do DL de cada vila foi comparado ao padrão de populações oriundas de regiões não isoladas da Sardenha e da população CEU. Como esperado, as cinco vilas menores e com maior isolamento apresentaram os maiores valores de DL, enquanto que as populações não isoladas da Sardenha e a do HapMap apresentaram valores de DL mais baixos quando comparadas a todas as oito vilas isoladas (Pistis *et al.*, 2009) .

No Brasil, a extensão do DL entre 11 STRs do cromossomo X (X-STRs) foi determinada para a população do Rio Grande do Sul (200 homens) e para um *pool* de amostras de 131 índios de 13 tribos da região amazônica. Na primeira população, o DL se estendeu até 45,5 Mb, provavelmente como resultado dos eventos de miscigenação ocorridos há pouco tempo atrás. O tempo decorrido não foi suficiente para que o DL fosse desfeito pela recombinação. A extensão do DL no *pool* de amostras indígenas foi de 142,2 Mb. Contudo, não foi possível distinguir se este DL extenso é decorrente de gargalos populacionais ou da subdivisão ocorrida durante os processos de formação das tribos (Leite *et al.*, 2009). Outro estudo realizado no Brasil descreveu os padrões de DL de outras três populações indígenas (132 cromossomos) e das populações do Rio Grande do Sul e da Costa Rica (124

cromossomos), a partir de 47 X-STRS. Como esperado, as populações ameríndias apresentaram uma menor diversidade genética e uma maior proporção de lócus em DL, quando comparadas as duas populações não isoladas. Foram detectados dois blocos de haplótipos no cromossomo X, ambos presentes apenas nos ameríndios (Amorim *et al.*, 2011). Resultados semelhantes foram obtidos pela análise de sete X-STRs em populações indígenas oriundas da Colômbia (cinco) e do Canadá (uma). Em tal estudo, os ameríndios apresentaram maior DL do que populações não isoladas, como as da Eurásia (Wang *et al.*, 2010).

Uma vez que as populações brasileiras remanescentes de quilombos experimentaram alguns dos eventos evolutivos descritos anteriormente, que geram ou retardam o decréscimo do DL, como fundação por poucos indivíduos, miscigenação no momento da fundação e posterior isolamento, buscou-se, por meio da análise de 250K SNPs, descrever os padrões de desequilíbrio de ligação nessas populações. A título de comparação, os padrões de DL também foram inferidos na população urbana de Teresina, capital do Estado do Piauí e em populações que experimentaram histórias demográficas diversas, analisadas pelo projeto HapMap.

## 2. HIPÓTESE

Fundamentado no conhecimento teórico relacionado ao Desequilíbrio de Ligação e à história demográfica de populações brasileiras remanescentes de quilombos, espera-se que os remanescentes apresentem um forte e extenso DL, devido à miscigenação (evento que cria DL) experimentada por essas populações no momento de sua fundação e ao seu posterior isolamento (o que teria permitido a manutenção dos altos níveis de DL criados pela miscigenação).

A população urbana, por sua vez, tem experimentado eventos de miscigenação por mais tempo (desde a sua fundação), além de não ter passado por nenhum período de isolamento. O tempo decorrido desde a fundação foi, provavelmente, suficiente para desfazer o DL criado entre a maioria dos loci. Assim, espera-se um DL menor e menos extenso nesta população.

Portanto, supõe-se que estes dois segmentos da população brasileira apresentem padrões diversos de DL em decorrência da estrutura genética que apresentam e de suas diferentes histórias demográficas.



### **3. OBJETIVO**

#### **3.1 Objetivo Geral**

O presente trabalho visa determinar e comparar os padrões de DL de três populações remanescentes de quilombos do Estado do Piauí (Gaucinha, Mimbó e Sítio Velho) e da população urbana de Teresina, capital do Estado do Piauí, além de sete populações do projeto HapMap (CEU, CHB, JPT, ASW, LWK, MKK, YRI), na tentativa de correlacionar os achados com a história demográfica dos remanescentes de quilombos.

#### **3.2 Objetivos Específicos**

- Genotipar 250 mil SNPs em amostra das quatro populações brasileiras;
- Comparar o desempenho dos algoritmos de determinação de genótipos BRLMM e CRLMM;
- Avaliar a influência do tamanho amostral na determinação dos padrões de desequilíbrio de ligação;
- Avaliar a quantidade de blocos de haplótipos nos diferentes cromossomos;
- Avaliar detalhadamente o padrão dos blocos de haplótipos nos cromossomo 6 e 22, e na região do MHC.

## 4. METODOLOGIA

### 4.1 Aspectos éticos da pesquisa

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (Processo HCRP nº 4788/2012).

As amostras analisadas neste estudo estão armazenadas em freezer a  $-20^{\circ}\text{C}$  no Laboratório de Genética Bioquímica do Departamento de Genética da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

### 4.2 Amostras populacionais

Foram analisadas 58 amostras provenientes de três comunidades afro-derivadas (remanescentes de quilombos), situadas no Estado do Piauí: Gaucinha (n = 14), Mimbó (n = 15) e Sítio Velho (n = 15) juntamente com uma amostra (n = 14) urbana de Teresina, a capital do Estado (Figura 1). As coletas foram realizadas em 1993 por professores da Universidade Federal do Piauí, liderados pela professora Dra. Zélia Arpini Sampaio, em projeto que incluía atendimento médico e alguns exames laboratoriais de assistência às comunidades estudadas.

#### Gaucinha (GAU):

Gaucinha está localizada a  $4^{\circ} 49'$  latitude Sul e  $42^{\circ} 10'$  longitude Oeste, distando 80 km da sede do município de Campo Maior e 180 km de Teresina. Na época em que foi realizada a visita à comunidade, esta contava com apenas 11 choupanas com localização esparsa, em decorrência da emigração dos moradores para a sede de Campo Maior.

#### Mimbó (MIB):

Mimbó localiza-se a  $6^{\circ} 14'$  latitude Sul e  $42^{\circ} 50'$  longitude Oeste e se encontra a 22 km da sede do município de Amarante e a 170 km de Teresina. Segundo moradores mais antigos, o povoado originou-se na

época da abolição da escravidão. Grupos de negros abandonaram as fazendas e migraram para a região em busca de um local para se fixarem e viverem sua liberdade.

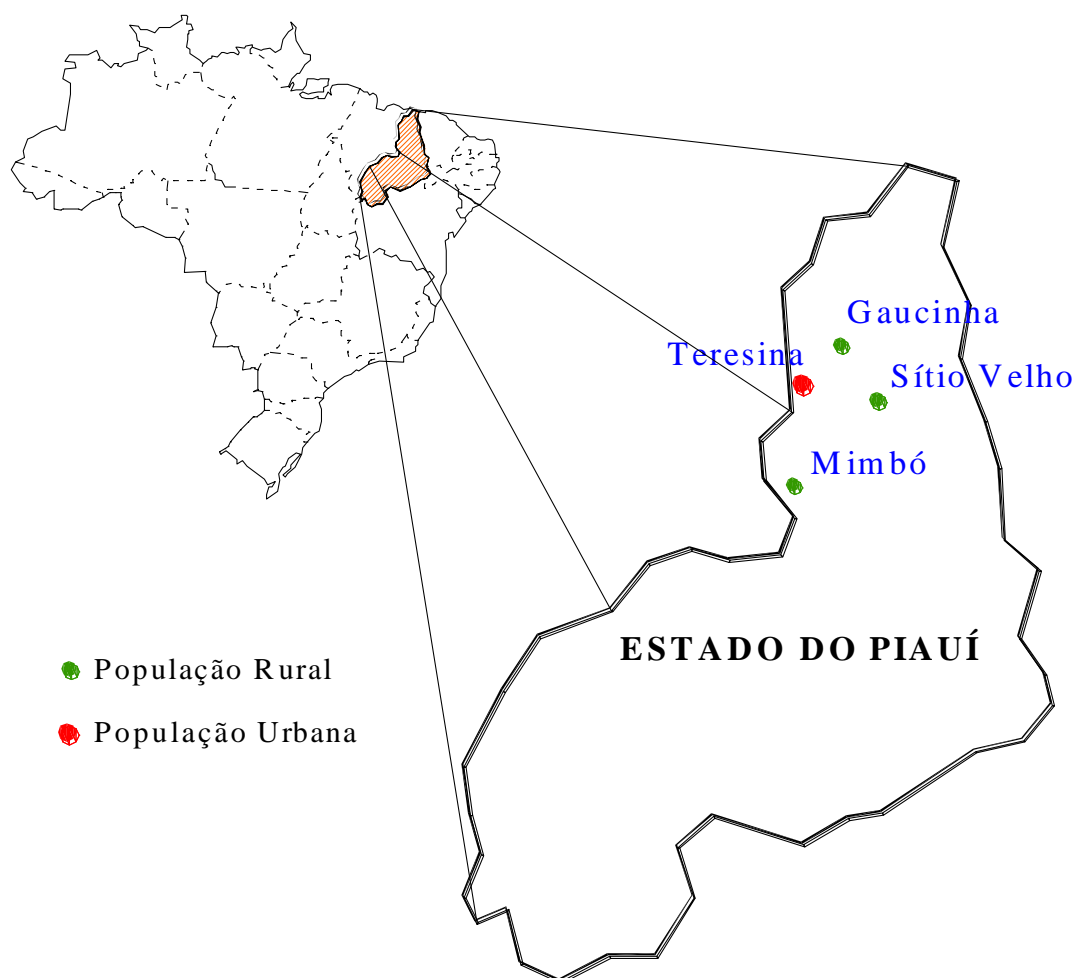


Figura 1. Localização geográfica das comunidades afro-derivadas do Estado do Piauí.  
Fonte: Wiezel, 2003.

### **Sítio Velho (STV):**

Sítio Velho localiza-se a 5° 20' latitude Sul e 41° 15' longitude Oeste, estando a 280 km de Teresina. De acordo com membros mais antigos, a comunidade surgiu no início do século XX nas margens de um riacho onde a caça era abundante, por aglomeração de africanos oriundos de

diferentes pontos de uma região relativamente ampla da fronteira entre os Estados do Piauí e Ceará.

### **Teresina (TES):**

A população de Teresina, por ser urbana, foi utilizada como população de referência para os três remanescentes de quilombos. A amostra é constituída de indivíduos nascidos no Estado do Piauí, selecionados entre os doadores de sangue do Centro de Hematologia e Hemoterapia do Estado do Piauí (HEMOPI).

### **Populações analisadas no projeto HapMap:**

Foram selecionadas sete populações estudadas pelo projeto HapMap (catálogo das variações genéticas comuns em humanos) para servirem também como referência para as comparações dos padrões de desequilíbrio de ligação entre as populações, por serem populações bem conhecidas e caracterizadas geneticamente. As populações do projeto HapMap incluídas neste estudo foram:

- 1) CEU: amostras de residentes dos Estados Unidos com ancestralidade do norte e oeste da Europa. As amostras foram coletadas em 1980 pelo Centro de Estudos do Polimorfismo Humano (CEPH). Assim, esta população foi usada como referência de uma população europeia;
- 2) CHB: amostras de chineses Han da cidade de Pequim;
- 3) JPT: amostras de japoneses da cidade de Tóquio;
- 4) YRI: amostras do grupo étnico Yoruba em Ibadan, Nigéria;
- 5) LWK: amostras do grupo étnico Luhya em Webuye, Quênia;
- 6) MKK: amostras do grupo étnico Maasai em Kinyawa, Quênia;
- 7) ASW: amostras de residentes da região sudoeste dos Estados Unidos que se auto-declararam afro-americanos.

A Tabela 1 apresenta o número de amostras disponíveis para estudo nas populações brasileiras e nas populações do projeto HapMap. Foram analisadas, a princípio, 15 amostras de cada uma das quatro populações brasileiras. Contudo, foram excluídas uma amostra de GAU e uma de TES por não produzirem resultados dentro dos padrões de qualidade exigidos pelo protocolo de genotipagem dos SNPs.

Tabela 1. Número de indivíduos das onze populações incluídas no estudo.

	Populações										
	GAU	MIB	STV	TES	CEU	CHB	JPT	YRI	LWK	MKK	ASW
Nº de Indivíduos disponíveis	15	15	15	15	165	84	86	167	90	171	83
Nº de Indivíduos analisados	14	15	15	14	15	15	15	15	15	15	15

GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Foram realizados testes para verificar se as análises de desequilíbrio de ligação sofrem influência das diferenças no tamanho amostral entre as populações brasileiras e do HapMap. Assim, de acordo com os resultados dos testes, optou-se pela padronização do tamanho amostral, analisando-se apenas 15 indivíduos de cada população do HapMap (Tabela 1).

Os genótipos gerados com o uso do *Human Mapping 250K Nsp Array* para as populações do HapMap foram obtidos na página do projeto (<http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/?N=D>). Assim, os procedimentos laboratoriais foram realizados utilizando-se apenas as amostras provenientes das quatro populações brasileiras (GAU, MIB, STV e TES).

### 4.3 Extração do DNA

O DNA genômico dos indivíduos selecionados para o estudo foi extraído a partir de amostras de sangue total utilizando-se o kit QIAamp DNA Investigator Kit (QIAGEN®, Germantown, MD, USA).

Este kit se baseia nas propriedades de ligação seletiva de colunas compostas por membranas de sílica. A partir de 100 $\mu$ L de sangue, seguiu-se o protocolo do fabricante. Primeiramente, cada amostra foi lisada, sob condições desnaturantes, com 10 $\mu$ L de proteinase K. Foram adicionados 50 $\mu$ L de etanol (96-100%) e, após incubação e centrifugação, o lisado foi transferido para as colunas de sílica. Este procedimento foi seguido por dois ciclos de lavagem com uso de reagentes específicos do kit. Foram adicionados, em seguida, 700 $\mu$ L de etanol (96-100%). Após dois ciclos de centrifugação, as amostras foram incubadas à temperatura ambiente. O DNA foi então eluído em água e sua concentração foi ajustada para 50 $\mu$ g/ $\mu$ L, após determinação da concentração inicial, utilizando-se o espectrofotômetro NanoDrop® (Nano Drop 2000c, Thermo Cientific, Wilmington, DE, USA). O espectrofotômetro foi também utilizado para determinar a pureza das amostras, com base na relação A260/A280. Ao final, as amostras de DNA foram armazenadas a -20°C até sua utilização.

#### **4.4 Genotipagem de SNPs com microarranjos de DNA**

##### **4.4.1 Microarranjos de DNA**

Os SNPs foram genotipados utilizando-se um sistema de microarranjos de DNA, o *GeneChip® Human Mapping 250K Nsp I Array - Affymetrix®*. Neste sistema, cada *array* inclui mais de 6,5 milhões de *spots*, cada um consistindo de mais de um milhão de cópias de uma sonda de oligonucleotídeos de 25 pb com uma sequência definida.

Para cada SNP, apenas dois alelos são considerados, uma vez que, para a maioria dos SNPs, apenas dois alelos são observados na natureza. Cada SNP é interrogado independentemente por seis ou dez quartetos de sondas, onde cada quarteto compreende duas sondas de complementação perfeita (PM, *perfect match*) para cada alelo do SNP e outras duas sondas de complementação imperfeita (MM, *mismatch*)



Existem ainda 50 SNPs em cada *array* que servem como controles internos. O *GeneChip® Genotyping Analysis Software* (GTYPE) (Affymetrix, Santa Clara, CA, USA) utiliza estes controles para realizar uma checagem cruzada dos genótipos a partir da mesma amostra em cada *array*, quando se usa concomitantemente os dois *arrays*: o 250K *Nsp I* e o 250K *Sty I*. O objetivo é assegurar que os dois *arrays* da mesma amostra não sejam confundidos com os demais, desde a preparação do DNA até a análise dos dados.

O *GeneChip® Human Mapping 250K Nsp Array* permite a análise de 262.244 SNPs humanos em um único experimento. Estes SNPs estão uniformemente distribuídos ao longo de todo o genoma, com uma distância média de 5,8 Kb entre eles e heterozigose média de 0,30, com base nas populações CEU, CHB, JPT e YRI do projeto HapMap. A frequência alélica mínima (MAF), ou seja, a frequência do alelo menos frequente é, em média, 22%. Anotações para cada SNP são fornecidas pelo GTYPE e pelo *NetAffx™ Analysis Center* (Affymetrix, Santa Clara, CA, USA). Estas anotações combinam dados de múltiplas fontes em uma única base de dados. As informações para cada SNP incluem a identificação do SNP (*rs number*) na base de dados dbSNP (NCBI), o gene mais próximo, a localização física e as frequências alélicas nas quatro populações citadas.

#### **4.4.2 Procedimentos laboratoriais**

Todos os procedimentos laboratoriais para a genotipagem dos SNPs foram realizados no Laboratório de Epidemiologia Genética do Departamento de Parasitologia do Instituto de Ciências Biomédicas da Universidade de São Paulo.

Os procedimentos seguiram estritamente as recomendações (incluindo quantidades e concentrações de reagentes) contidas no manual *GeneChip® 500K Assay Manual* (Affymetrix, Santa Clara, CA, USA).



O primeiro procedimento realizado foi a digestão do DNA genômico de cada indivíduo, em regiões específicas, utilizando-se a enzima de restrição *NspI* (New England Biolabs, Ipswich, MA, USA), visando a obtenção de fragmentos de DNA de menor complexidade para a amplificação e genotipagem de regiões específicas do genoma. Foram usados na reação 250ng de DNA de cada amostra. A digestão ocorreu por 2 horas a 37°C, sendo seguida pela inativação térmica da enzima (65°C por 20 minutos).

O segundo passo do protocolo foi a ligação dos adaptadores *NspI* (*Affymetrix GeneChip Mapping 250K Nsp Assay Kit*) nas extremidades das moléculas de DNA digeridas. Essa etapa foi realizada com o uso da enzima T4 DNA ligase (New England Biolabs, Ipswich, MA, USA) a 16°C por 3 horas.

Após a etapa de inativação térmica da T4 DNA ligase (70°C por 20 minutos), o DNA foi diluído quatro vezes com água (*Water molecular biology reagent* Sigma-Aldrich, St. Louis, MO, USA), sendo preparado para amplificação pela reação em cadeia da polimerase (PCR). Esta reação é otimizada com o objetivo de amplificar fragmentos de 250 a 1.100 pb. Os oligonucleotídeos iniciadores da reação estão também inclusos no *Affymetrix GeneChip Mapping 250K Nsp Assay Kit*. O DNA é amplificado com o uso do kit *TITANIUM™ DNA Amplification Kit* (Clontech, Mountain View, CA, USA). A amplificação de cada amostra é realizada em triplicata, visando à obtenção de uma maior quantidade de DNA amplificado ao final dessa etapa.

Após a amplificação do DNA, uma amostra de cada produto amplificado foi testada por meio de eletroforese em gel de agarose 2%. Nesta etapa, espera-se que os fragmentos de DNA amplificados apresentem um comprimento de 250 a 1.100 pb, confirmando o sucesso das reações de digestão, ligação e amplificação.

Os produtos amplificados foram então purificados utilizando-se o *DNA amplification Clean-Up Kit* (Clontech, Mountain View, CA, USA). Após

secagem, cada amostra foi eluída em 45µL de *Rb Buffer* e, em seguida, 2µL foram diluídos em 198µL de água. As amostras foram então quantificadas no NanoDrop 2000 (Thermo Scientific, Wilmington, DE, USA). Posteriormente, cada amostra foi diluída para uma concentração final de 2 µg/µl usando-se o tampão *RB Buffer* incluído no *DNA Amplification Clean-up Kit* (Clontech, Mountain View, CA, USA).

Cada amostra foi submetida à reação de fragmentação por meio da ação da enzima DNase I (0,25U) inclusa no *Affymetrix GeneChip Mapping 250K Nsp Assay Kit*. Na fragmentação, foram usados 45µL de cada amostra purificada na etapa anterior, com o objetivo de fragmentar 90 µg do DNA. A reação de fragmentação foi realizada a 37°C por 35 minutos. Em seguida, a enzima foi inativada a 95°C por 15 minutos. A eficiência da reação de fragmentação foi verificada por meio de eletroforese das amostras fragmentadas em gel de agarose de 4%, sendo validadas as amostras em que o DNA apresentasse tamanho inferior a 180 pb.

Após a etapa de fragmentação, as amostras foram marcadas com biotina (30 mM) (*GeneChip® DNA Labeling Reagent*) (Affymetrix, Santa Clara, CA, USA), por meio da ação da enzima TdT (*Terminal Deoxynucleotidyl Transferase*), que catalisa a transferência de um nucleotídeo biotinilado para as extremidades 3' do DNA. A reação de marcação ocorreu a 37°C por 4 horas.

O DNA marcado foi então hibridizado com as sondas presentes no *Genechip® Human mapping 250K Nsp Array* durante 16 a 18 horas a 50°C. A incubação foi realizada no *GeneChip® Hybridization Oven 645* (Affymetrix Services Lab, Santa Clara, CA, USA).

Os microarranjos foram lavados com os tampões *Wash A* (6X SSPE, 0,01% Tween 20) e *Wash B* (0,6X SSPE, 0,01% Tween 20) no *GeneChip® Fluidics Station 450* (Affymetrix, Santa Clara, CA, USA), visando a retirada de material em excesso e de sondas hibridizadas inespecificamente. A amplificação do sinal da marcação foi realizada por

meio de sucessivos ciclos de incubação com estreptavidina-ficoeritrina (1 mg/ml) (SAPE Molecular Probes/Life Technologies, Carlsbad, CA, USA) conjugada com o anticorpo anti-estreptavidina biotilado (0,5 mg/ml) (Vector Laboratories Inc., Burlingame, CA, USA) no *GeneChip Fluidics Station 450* (Affymetrix Services Lab, Santa Clara, CA, USA).

O último passo realizado foi a leitura do *chip* em um *scanner* (*GeneChip® Scanner 3000 7G* (Affymetrix Services Lab, Santa Clara, CA, USA)).

Fluxograma ilustrativo de todo o processo de análise laboratorial voltado para a genotipagem de SNPs pode ser observado nas Figuras 3 e 4.

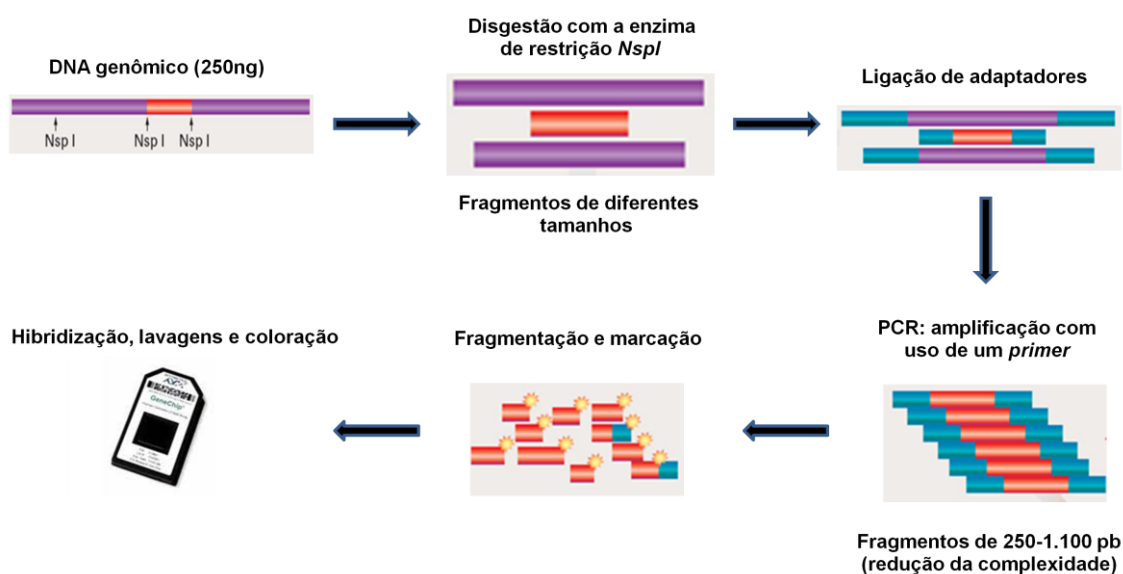


Figura 3. Representação esquemática dos procedimentos laboratoriais para análise de SNPs. O DNA genômico é submetido à digestão pela enzima de restrição *NspI*. Esta digestão produz fragmentos de tamanhos variados. Um adaptador comum é então ligado às extremidades destes fragmentos e é usado como *primer* para a reação de amplificação. A PCR ocorre sob condições controladas para que sejam gerados fragmentos de 250 a 1.100 pb, o que reduz a complexidade genômica. O produto da amplificação é fragmentado e marcado, resultando em uma amostra pronta para as etapas de hibridização, lavagens, coloração e *scanning*. Fonte: Andrade, 2013.

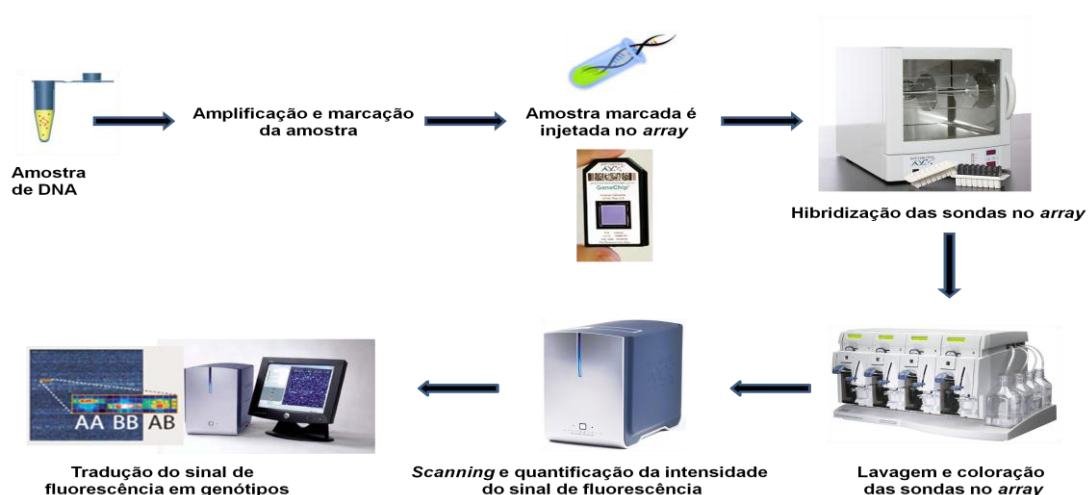


Figura 4. Fluxograma de um experimento com o sistema *GeneChip*. Após a obtenção do DNA, as reações de amplificação e marcação resultam em uma amostra marcada. Esta amostra é injetada no *array* e segue-se o processo de hibridização por 16 a 18 horas no forno de hibridização. Os ciclos de lavagem e coloração ocorrem na estação fluidica (quatro *arrays* simultaneamente). O *array* é então lido pelo *scanner* e, em seguida, a intensidade de fluorescência é traduzida em informações de genótipos com uso de diferentes programas. Fonte: Andrade, 2013.

#### 4.4.3 Análise dos dados e determinação dos genótipos

Após a leitura dos microarranjos no *GeneChip Scanner 3000 7G* (Affymetrix, Santa Clara, CA, USA), um arquivo de dados brutos (de extensão “.dat”) referentes à imagem lida é criado. A imagem dessa leitura pode ser observada com o programa de visualização *Affymetrix Gene Chip Viewer*, que permite também a verificação dos controles positivos do sistema para confirmar se as condições de hibridação do microarranjo foram adequadas.

O principal objetivo do pré-processamento dos *arrays* é normalizar as intensidades e prever os genótipos AA, AB ou BB com uma determinada medida de confiança. Estas previsões são definidas como *genotype calls*. As amostras que não alcançam um valor de confiança pré-definido (*cutoff*) para um determinado SNP, recebem *no call* para este SNP.

O arquivo “.dat” é processado pelo programa *Affymetrix GeneChip Command Console* (AGCC) gerando um arquivo de extensão “.cel” onde são armazenadas informações relativas a intensidade do sinal em cada pixel dos microarranjos. Estas informações incluem, para cada SNP, a intensidade do sinal, o desvio padrão da intensidade, o número de pixels utilizados no cálculo da intensidade e uma marcação dos pixels que apresentam valores acima do limite estabelecido pelo algoritmo. O arquivo “.cel” armazena também informações sobre o experimento relacionado ao microarranjo, inclusive sobre a qualidade da leitura dos dados, denominada *call rate*.

A *call rate*, representa a porcentagem de SNPs para os quais é possível atribuir um genótipo. Esta *call rate* é gerada para cada *array* pelo *GeneChip® Genotyping Analysis Software* (GTYPE), que usa um algoritmo automatizado de determinação de genótipos baseado em modelo dinâmico (DM) (Di *et al.*, 2002), fornecendo um *score* de confiança para cada genótipo individual. Uma *call rate* maior ou igual a 93% deve ser usada para determinar se uma amostra pode ou não ser utilizada. Contudo, o usuário pode ajustar esse *score* de acordo com as exigências das aplicações posteriores e considerando a qualidade das amostras de DNA (Matsuzaki *et al.*, 2004). Assim, os *chips* com *call rate* abaixo deste valor foram utilizados no presente estudo.

O algoritmo DM considera quatro diferentes modelos (nulo, AA, AB e BB) de intensidade das sondas para cada SNP e a determinação do genótipo para cada amostra é realizada com base na probabilidade de cada genótipo. Note-se que os genótipos são determinados diretamente a partir da intensidade de cada *array* separadamente. Este processo, contudo, gera um maior grau de erros de classificação para os genótipos heterozigotos (Carvalho *et al.*, 2007).

Os arquivos “.cel” são então analisados pelo programa *Genotyping Console* (GTC) da Affymetrix, que utiliza o algoritmo *Bayesian Robust Linear Model with Mahalanobis distance classifier* BRLMM (Affymetrix,

2006) para obter os genótipos dos indivíduos a partir dos dados de intensidade dos pixels nos microarranjos. O BRLMM usa os genótipos determinados pelo DM como um passo inicial para definir as regiões de cada tipo de genótipo. Estas regiões são então recalibradas por meio de uma abordagem Bayesiana, usando-se dados de múltiplos *arrays* de diferentes indivíduos para determinação dos genótipos (Rabbee e Speed 2005; Affymetrix, Inc 2006).

Este algoritmo enriquece significativamente os resultados do DM: ele melhora o desempenho geral (*call rate* e acurácia) e equaliza o desempenho dos genótipos homocigotos e heterocigotos. Os genótipos assim determinados, bem como informações tabuladas acerca dos indivíduos genotipados, ficam armazenados no programa GTC e podem ser exportados para arquivos de texto (".txt").

Uma vez que alguns chips apresentaram *call rate* inferior a 93%, buscou-se utilizar um método de determinação de genótipos que fosse capaz de extrair o máximo possível de informações dos dados brutos, visando, ao final, um aumento do número de SNPs que pudessem ter seus genótipos determinados com alta confiança e acurácia.

Assim, foram realizados testes com o objetivo de comparar o desempenho de dois algoritmos de determinação de genótipos: o já citado BRLMM (Affymetrix®) e o CRLMM (*Corrected Robust Linear Model with Maximum Likelihood Classification*) (Carvalho *et al.*, 2007; Lin *et al.*, 2008), uma vez que uma determinação precisa e confiável dos genótipos é um passo crucial nas estimativas dos haplótipos e nas análises do desequilíbrio de ligação, que são objetivos deste estudo.

O algoritmo CRLMM foi escolhido para comparação porque, a princípio, fornece uma determinação mais precisa dos genótipos e oferece uma significativa melhoria nas estimativas de acurácia para cada SNP de forma robusta (Lin *et al.*, 2008; de Andrade *et al.*, 2011), fazendo com que mais amostras e SNPs sejam aproveitados nas análises posteriores.

No CRLMM, para cada *array*, os efeitos da sequência da sonda e do comprimento do fragmento de DNA são estimados e removidos do *log* da intensidade. Em seguida, é usada uma medida de normalização contra uma amostra de referência para remover as variações entre os *arrays*. Para cada SNP, para cada um dos dois alelos e para cada uma das fitas de DNA é gerada uma medida sumária de todas as sondas PM (*perfect match*) usando-se um modelo linear. São calculadas também, para cada fita, a diferença do *log* entre os alelos A e B. Uma vez que os efeitos da sequência da sonda, do comprimento do fragmento e da intensidade total são dependentes do genótipo, um modelo de mistura é usado, que assume que cada genótipo desconhecido resulta em uma diferente distribuição Gaussiana.

Utilizando-se os dados do HapMap como treinamento, visto que nestes dados os genótipos são conhecidos para a maioria dos SNPs, a média e a variância são calculadas para cada SNP para as razões logarítmicas e são corrigidas para os efeitos estimados nos passos anteriores. Os genótipos das amostras do HapMap são baseados em um consenso a partir de diferentes tecnologias de genotipagem e, por isso, podem ser considerados como “padrão ouro”, servindo como base na determinação dos genótipos das amostras a serem analisadas pelo CRLMM. O modelo de mistura dos efeitos é então usado para obter as estimativas Bayesianas empíricas. Assim, para uma nova amostra e para cada SNP, o genótipo é predito como a maximização da probabilidade calculada, uma vez que a média e a variância são conhecidas. As razões de probabilidade são então usadas como medidas de incerteza (Carvalho *et al.*, 2007).

Uma implementação do CRLMM está disponível no pacote *oligo* do *software Bioconductor* (<http://www.bioconductor.org/>) (Gentleman *et al.*, 2004), um projeto de desenvolvimento de software aberto, contido no programa de estatística computacional R (<http://www.r-project.org/>). Assim, foram desenvolvidos *scripts* específicos no programa R para que, a partir da base de dados “pd.mapping250k.nsp”,

que inclui os dados brutos das amostras do HapMap, e utilizando o pacote *oligo*, o algoritmo CRLMM determinasse os genótipos para as amostras estudadas (as do HapMap e as brasileiras) utilizando os arquivos “.cel”, gerados pelo programa *Affymetrix Gene Chip Command Console* (AGCC). Além disso, foi utilizado o arquivo de anotação de SNPs do projeto HapMap (Mapping250K\_Nsp Annotations, CSV format, Release 31 (83 MB, 10/26/10), disponível em [http://www.affymetrix.com/Auth/analysis/downloads/na31/genotyping/Mapping250K\\_Nsp.na31.annot.csv.zip](http://www.affymetrix.com/Auth/analysis/downloads/na31/genotyping/Mapping250K_Nsp.na31.annot.csv.zip). Os genótipos determinados dessa forma pelo CRLMM foram então comparados com os genótipos gerados pelo BRLMM.

Fluxograma ilustrativo de todo o processo de análise computacional voltado para a determinação dos genótipos pode ser observado na Figura 5.



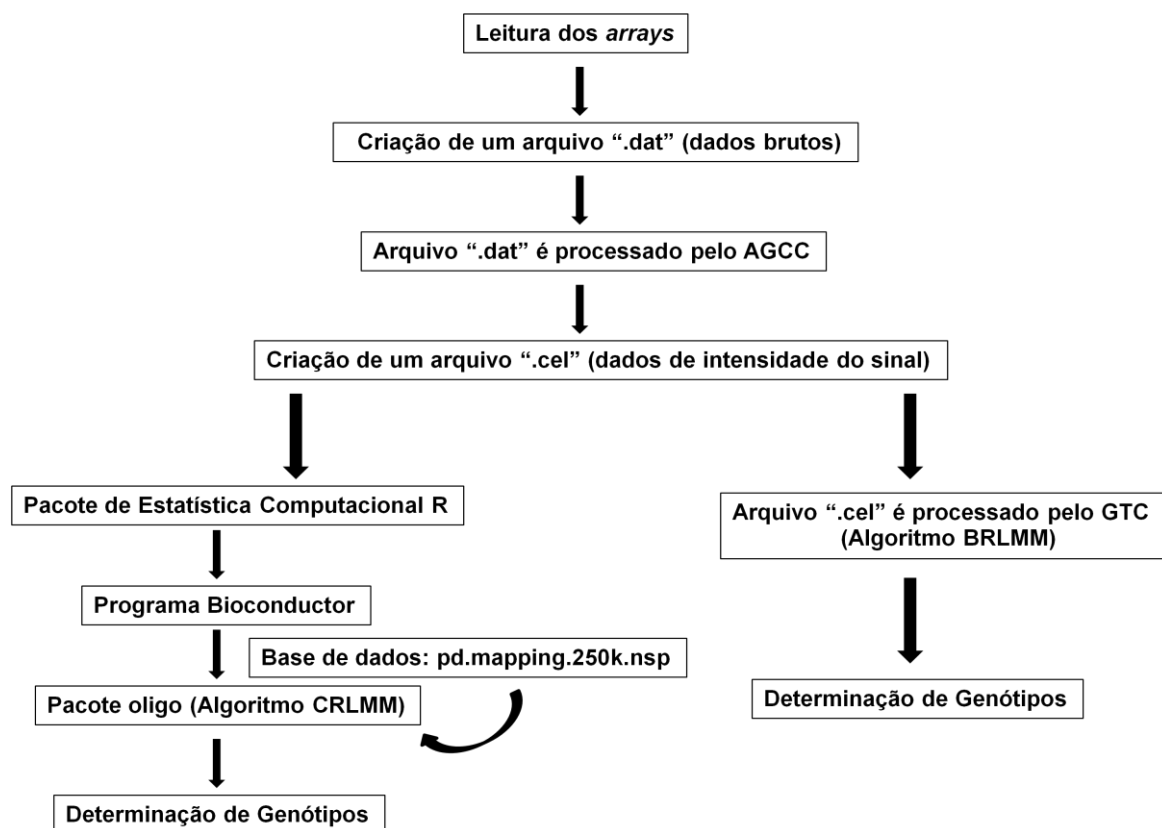


Figura 5. Fluxograma do processo de análise computacional dos dados. Após a leitura dos *arrays* pelo scanner, um arquivo de dados brutos (".dat") é criado. Este arquivo é processado pelo programa AGCC, gerando um arquivo contendo informações sobre a intensidade do sinal de fluorescência (".cel"). Este arquivo é processado pelo programa GTC, que utiliza o algoritmo BRLMM para determinar os genótipos. O arquivo ".cel" foi também analisado com uso do pacote *oligo*, do programa *Bioconductor*, contido no pacote R. Os genótipos foram então determinados pelo algoritmo CRLMM, implementado no pacote *oligo*, utilizando-se as informações contidas na base de dados *pd.mapping.250k.nsp*. Fonte: Andrade, 2013.

#### 4.5 Análises de desequilíbrio de ligação e blocos de haplótipos

A partir dos genótipos determinados pelo CRLMM com probabilidade  $\geq 0,90$ , foram realizadas as análises de desequilíbrio de ligação e blocos de haplótipos, utilizando-se o software Haploview (Barrett *et al.*, 2005; <http://www.broad.mit.edu/mpg/haploview>). Foram realizadas duas abordagens, considerando e não considerando os dados de SNPs cujos genótipos não puderam ser determinados (*no call*).

Inicialmente, foi realizada uma seleção para obtenção dos SNPs comuns a todas as 11 populações analisadas (as quatro brasileiras e as sete do HapMap), para que a comparação dos padrões de desequilíbrio de ligação e blocos de haplótipos fosse possível. Esta seleção foi realizada com o uso do programa R, utilizando-se *scripts* específicos. Outros *scripts* foram desenvolvidos para que os dados resultantes da seleção fossem formatados de acordo com um dos formatos de entrada do programa Haploview, o “HapMap format”.

Os haplótipos são então inferidos pelo Haploview por meio do algoritmo EM, utilizando o método de partição/ligação descrito em Qin e colaboradores 2002. Este método infere as fases dos haplótipos e gera estimativas de suas respectivas frequências populacionais com base em máxima verossimilhança, determinada a partir dos dados de genótipo de fase desconhecida.

Com o uso do Haploview, o Desequilíbrio de Ligação (DL) entre cada par de SNPs foi calculado a partir de três medidas: o *LOD (log of the likelihood odds ratio) score*, o coeficiente de desvio padronizado (*D'*) de Lewontin (1964) e a medida  $r^2$ .

O *LOD score* compara a probabilidade de se obter os dados testados se os dois loci estão verdadeiramente ligados e a probabilidade de se observar os mesmos dados por chance (acaso). O programa considera que  $LOD > 2$  indica um DL significativo.

O *D'* é calculado pelo Haploview da seguinte maneira: consideremos como exemplo dois genes (A e B) cada um com dois alelos: gene A com os alelos A e a e o gene B com os alelos B e b. Em qualquer geração, um cromossomo contendo os gametas AB pode ser submetido a eventos de recombinação com probabilidade =  $r$ , sendo  $r$  a frequência de recombinação. Por outro lado, o cromossomo pode não sofrer recombinação (probabilidade =  $1 - r$ ). Quando não ocorre recombinação, a frequência de AB é a mesma observada em gerações anteriores,  $P_{AB}$ . Entre os cromossomos que sofrem recombinação, a frequência de AB é

simplesmente o produto das frequências dos alelos  $A$  e  $B$  nas gerações prévias:  $p_{APB}$ . Assim, o  $D$  pode ser definido como a diferença entre  $P_{AB}$  (não recombinantes) e  $p_{APB}$  (recombinantes). O  $D$  é calculado da mesma forma para os demais gametas  $Ab$ ,  $aB$ ,  $ab$  e pode também ser definido como:  $D = P_{AB}P_{ab} - P_{Ab}P_{aB}$  (Hartl e Clark 1997). Uma vez que o valor de  $D$  é dependente das frequências alélicas, seus valores mínimo e máximo são assim definidos:  $D_{min} = \text{maior valor de } -p_{APB} \text{ e } -q_aq_b$ ;  $D_{max} = \text{maior valor de } p_{AQb} \text{ e } q_a p_B$ . Por esta razão, a magnitude do desequilíbrio de ligação é descrita por  $D'$ , que é definido como:  $D' = D/D_{max}$  (quando  $D$  for positivo) e  $D' = D/D_{min}$  (quando  $D$  for negativo) (Hartl e Clark 1997).

A medida  $r^2$ , por sua vez, é definida como:  $r^2 = (D')^2 / (p_A q_a p_B q_b)$ . A raiz quadrada de  $r^2$  representa o coeficiente de correlação no estado alélico entre os alelos no mesmo gameta (Hartl e Clark 1997).

Ambas as medidas,  $D'$  e  $r^2$ , são usadas para descrever a quantidade de desequilíbrio de ligação porque representam diferentes aspectos das associações gaméticas. Assim, quando  $D'$  é próximo de zero,  $r^2$  também se aproxima de zero. Contudo, à medida que  $D'$  aumenta,  $r^2$  pode apresentar qualquer valor entre zero e  $(D')^2$  (Hartl e Clark 1997).

Blocos de haplótipos foram gerados e analisados separadamente para cada cromossomo em cada população e foram posteriormente comparados entre todas as populações estudadas. Para tal análise, foi solicitado ao programa excluir os indivíduos que apresentassem mais de 50% de genótipos não determinados e forçar as análises para computar todos os pares de marcadores. Para cada SNP, foram utilizados os valores *default* do Haploview para os seguintes parâmetros: *cutoff* do valor de  $p$  do Equilíbrio de Hardy-Weinberg (0,0010), porcentagem mínima de genótipos (75%) e frequência alélica mínima (0,0010).

Os blocos de haplótipos foram definidos separadamente para cada população segundo Gabriel e colaboradores (2002). De acordo com o critério destes autores, um par de SNPs está em forte DL se  $D' > 0,98$ , o que é consistente com ausência de recombinação histórica, com limite

inferior do intervalo de confiança de 95% acima de 0,70. Em contrapartida, os pares de SNPs indicam forte evidência de recombinação histórica quando o limite superior do intervalo de confiança de  $D'$  é menor do que 0,9 (Gabriel *et al.*, 2002).

Deste modo, um bloco é definido como uma região sobre a qual uma proporção muito pequena (< 5%) de comparações entre pares de SNPs informativos mostra evidência de recombinação histórica. Dentro dos blocos, medidas independentes de DL entre pares de SNPs não declinam em função da distância entre os marcadores (Gabriel *et al.*, 2002). Este método de definição de blocos ignora SNPs com frequência alélica mínima (MAF) menor que 0,05. De modo geral, o método consiste em ordenar pelo tamanho a lista de todos os blocos possíveis; começando a partir do bloco maior, os demais blocos são adicionados enquanto não houver uma sobreposição com um bloco já definido (Gabriel *et al.*, 2002).

O Haploview permite representar o DL por meio de diferentes esquemas de cores. No mais tradicional deles (Figura 6), voltado para representação dos valores de  $D'$  e  $LOD$  score, branco indica  $LOD < 2$  e  $D' < 1$ ; azul indica  $LOD < 2$  e  $D' = 1$ ; tons de rosa/vermelho brilhante indicam  $LOD \geq 2$  e  $D' < 1$ ; e vermelho indica  $LOD \geq 2$  e  $D' = 1$ . Em um segundo esquema, voltado para a representação dos intervalos de confiança de  $D'$ , uma forte evidência de DL é representada pela cor cinza escuro enquanto que forte evidência de recombinação é representada por branco; cinza claro indica uma comparação não informativa. No terceiro esquema, branco indica  $r^2 = 0$ , tons de cinza indicam  $0 < r^2 < 1$  e preto indica  $r^2 = 1$ . O primeiro esquema foi selecionado para ilustrar os resultados obtidos no presente trabalho (Figura 6).

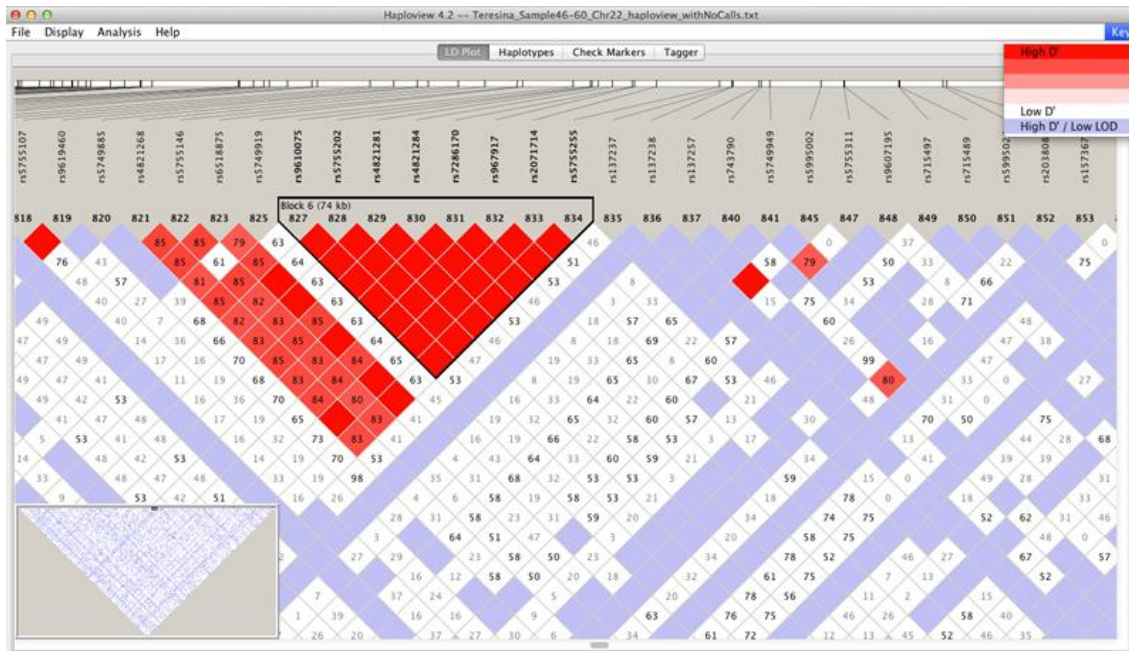


Figura 6. Esquema de cores utilizado pelo Haploview para representar o Desequilíbrio de Ligação em função de  $D'$  e  $LOD$  score.

Um valor de  $D'$  multialélico também é calculado pelo Haploview. Este consiste em uma medida de DL entre dois blocos de haplótipos e representa o nível de recombinação entre os dois blocos (Figura 7).

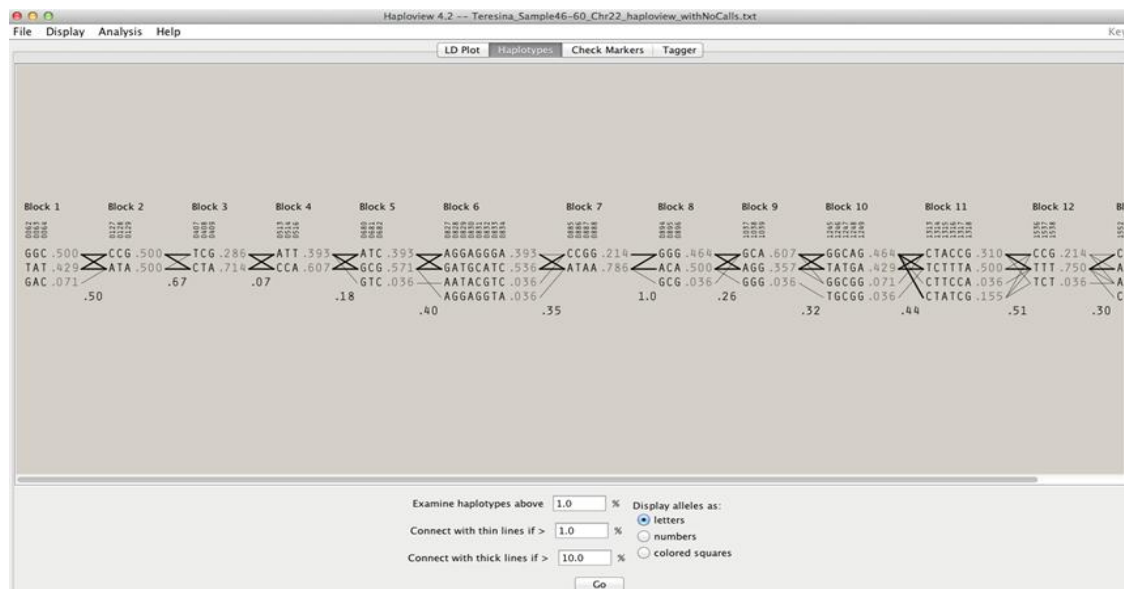


Figura 7. Frequências estimadas dos haplótipos inferidos em cada bloco e  $D'$  multialélico ilustrando o Desequilíbrio de Ligação entre cada bloco definido pelo programa Haploview.

Quanto mais próximo de zero, maior será a quantidade de recombinação histórica entre os blocos. O  $D'$  multialélico não é computado para os haplótipos raros, pois estes contribuem muito pouco para o valor geral.

## 5. RESULTADOS

### 5.1 Comparação entre os algoritmos BRLMM e CRLMM

Os resultados das comparações entre os algoritmos de determinação de genótipos, BRLMM e CRLMM, para todos os 262.244 SNPs contidos no *GeneChip 250 K Nsp* estão representados na Figura 8.

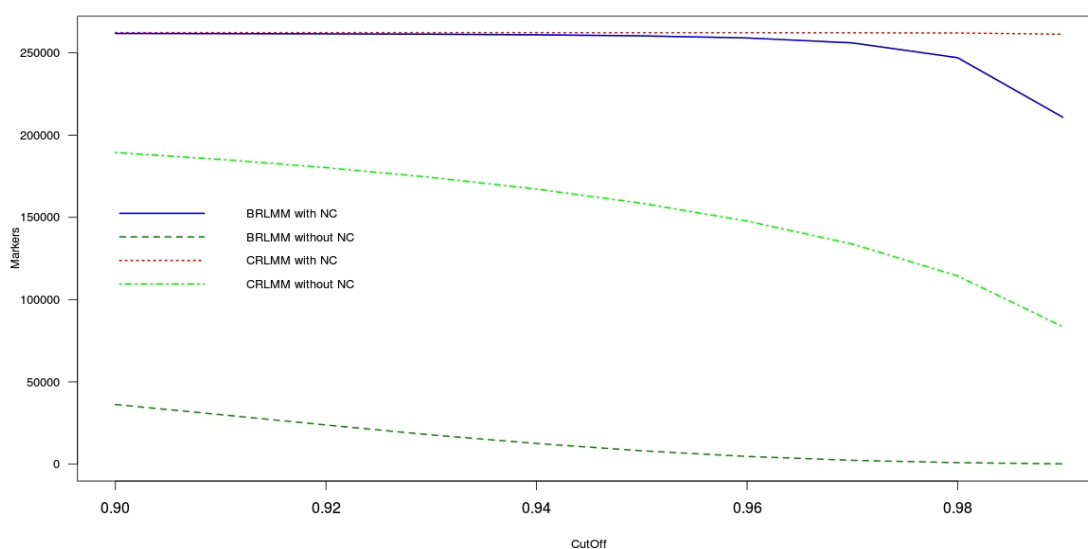


Figura 8. Resultados das comparações entre os algoritmos BRLMM e CRLMM. NC = *no call*. *With NC*: abordagem que considera todos os SNPs, inclusive aqueles cujos genótipos não foram determinados para todos os indivíduos. *Without NC*: abordagem que considera apenas os SNPs cujos genótipos foram determinados para todos os indivíduos.

O algoritmo CRLMM apresentou um melhor desempenho, sendo capaz de determinar os genótipos para praticamente todos os SNPs com grande acurácia (0.98), quando os SNPs cujos genótipos não foram determinados para todos os indivíduos foram incluídos nas análises.

Assim, este algoritmo foi escolhido para a determinação dos genótipos em todas as amostras estudadas, visando a uma menor perda de dados e a um maior poder de resolução na análise do DL.

A partir dos genótipos determinados pelo CRLMM (*cut off* = 0.90), com e sem *no call*, foram realizadas as análises de desequilíbrio de ligação e blocos de haplótipos, utilizando-se o *software* Haploview. Para tal análise, foram selecionados apenas os SNPs (com e sem *no call*) analisados em todas as 11 populações consideradas no presente estudo (as quatro brasileiras e as sete do HapMap), para que a comparação dos padrões de desequilíbrio de ligação e blocos de haplótipos fosse possível.

São apresentados, como exemplos, os resultados para os cromossomos 1 e 22 para a população remanescente de quilombo Gaucinha. Para o cromossomo 1, foram considerados 12.796 marcadores, sendo incluídos os SNPs cujos genótipos não foram determinados para todos os indivíduos (isto é, com *no call*). A partir destes marcadores, foram determinados 142 blocos de haplótipos. Quando foram incluídos nas análises apenas os SNPs cujos genótipos foram determinados para todos os indivíduos (sem *no call*) o número de marcadores foi reduzido para 10.005 e 117 blocos de haplótipos foram gerados pelo Haploview. Para o cromossomo 22, os números de blocos de haplótipos foram 14 (1.572 marcadores) e 12 (1.210 marcadores), seguindo a mesma descrição acima, respectivamente.

A partir destes resultados, observou-se que, ao considerar apenas os SNPs sem *no call* houve uma grande redução no número de SNPs incluídos nas análises (2.791 SNPs foram excluídos do cromossomo 1 e 362 SNPs do cromossomo 22) e, conseqüentemente, uma diminuição do número de blocos de haplótipos gerados. Um número muito reduzido de marcadores dificulta as análises de desequilíbrio de ligação e a determinação dos blocos de haplótipos. Assim, optou-se por utilizar a abordagem que considera inclusive os SNPs cujos genótipos não foram



determinados para todos os indivíduos (com *no call*), para maximizar o número de SNPs analisados.

A Tabela 2 apresenta o número total de SNPs contidos no *GeneChip 250 K Nsp* e o número de SNPs incluídos nas análises, para cada cromossomo. O número de SNPs analisados inclui os SNPs comuns a todas as populações, mesmo quando os seus respectivos genótipos não foram determinados para todos os indivíduos.

Tabela 2. Número de SNPs contidos no *GeneChip 250 K Nsp* e comuns a todas as populações estudadas.

Cromossomo	Nº de SNPs no <i>chip</i>	Nº de SNPs comuns (com <i>no call</i> )
1	19.855	12.796
2	22.204	14.603
3	18.358	12.240
4	19.053	12.254
5	17.152	11.423
6	17.104	11.603
7	13.932	9.327
8	14.834	9.701
9	11.942	7.747
10	14.263	9.188
11	13.287	8.677
12	13.046	8.448
13	11.072	7.167
14	8.175	5.348
15	7.002	4.590
16	7.010	4.626
17	4.844	3.178
18	8.145	5.282
19	2.691	1.765
20	5.838	3.840
21	3.936	2.619
22	2.499	1.572
X	5.714	3.540
Y	5	–
?	283	–
Total	262.244	167.994

Após a seleção para SNPs comuns, a redução do número total de SNPs (considerando-se todos os cromossomos) variou de 32,16% (cromossomo 6) a 38,05% (cromossomo X), com redução média de 34,69%.

## **5.2 Influências do tamanho amostral e do número de SNPs analisados na estimativa do DL**

Ao serem analisados todos os indivíduos disponíveis para cada população, conforme apresentado na Tabela 1, padrões bem distintos no que se refere ao número de blocos de haplótipos gerados pelo Haploview foram observados entre as populações brasileiras e as populações do projeto HapMap.

Para testar se esta variação se deve a grande diferença no número de indivíduos analisados, foram escolhidos aleatoriamente 15 indivíduos de cada uma das sete populações do HapMap para gerar uma amostragem uniforme, permitindo uma comparação de resultados mais confiável e informativa.

Quando o número de indivíduos foi fixado em 15, a diferença no número de blocos foi significativamente reduzida, passando a refletir apenas as diferenças nos padrões de desequilíbrio de ligação resultantes das diferentes histórias demográficas de cada população. Os resultados estão apresentados nas Figuras 9 e 10 para os cromossomos 6 e 22, respectivamente.

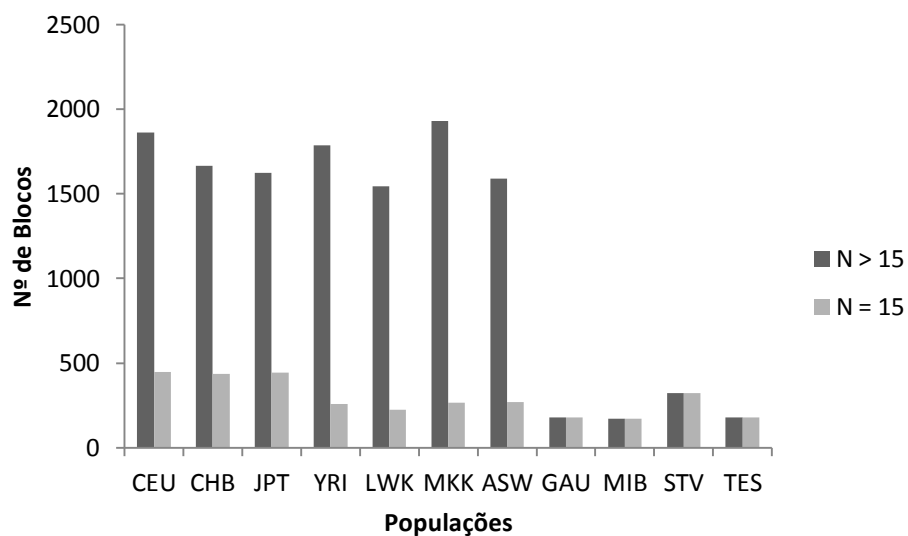


Figura 9. Número de blocos de haplótipos gerados pelo Haploview para o cromossomo 6 incluindo todos os indivíduos disponíveis e apenas 15 indivíduos escolhidos aleatoriamente de cada população.

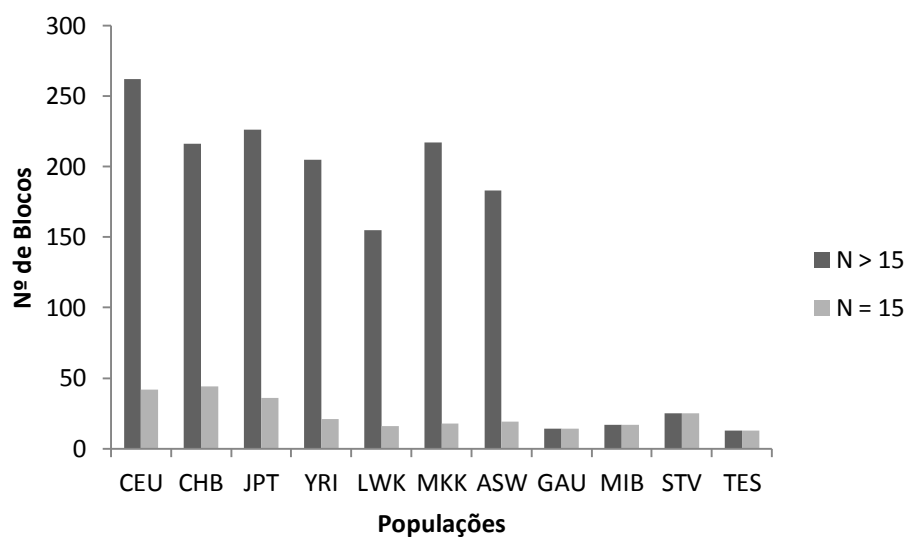


Figura 10. Número de blocos de haplótipos gerados pelo Haploview para o cromossomo 22 incluindo todos os indivíduos disponíveis e apenas 15 indivíduos escolhidos aleatoriamente de cada população.

O número de SNPs utilizados pelo Haploview para gerar os blocos de haplótipos para cada cromossomo foram menores do que os números apresentados na Tabela 2 (SNPs comuns a todas as populações), uma

vez que o Haploview exclui SNPs que falham em determinados testes, como o de Equilíbrio de Hardy-Weinberg. Consequentemente, o número real de SNPs analisados variou entre as populações. Considerando todos os indivíduos disponíveis, o número de SNPs excluídos foi extremamente pequeno ou nulo. Por outro lado, quando o número amostral foi reduzido para 15, muitos SNPs foram excluídos das análises (Figuras 11 e 12). Isto explica o número muito maior de blocos gerados para as populações do HapMap quando um número maior de indivíduos foi analisado.

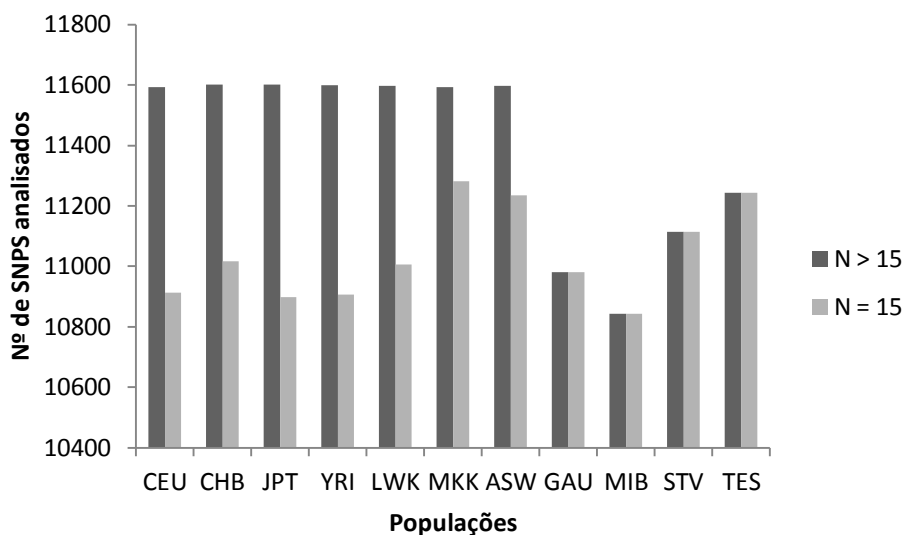


Figura 11. Número real de SNPs (dos 11.603) analisados pelo Haploview para gerar os blocos de haplótipos do cromossomo 6, considerando todos os indivíduos disponíveis e apenas 15 indivíduos escolhidos aleatoriamente de cada população.

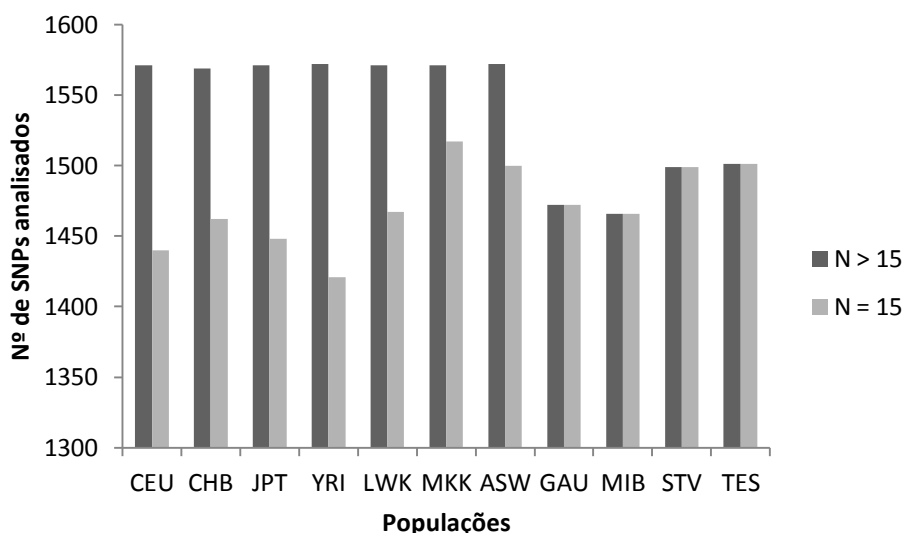


Figura 12. Número real de SNPs (dos 1.572) analisados pelo Haploview para gerar os blocos de haplótipos do cromossomo 22, considerando todos os indivíduos disponíveis e apenas 15 indivíduos escolhidos aleatoriamente de cada população.

Contudo, considerando-se apenas os dados dos 15 indivíduos, aparentemente não houve correlação entre o número de SNPs analisados e o número de blocos de haplótipos gerados, exceto para o cromossomo 6, em que a população que apresentou o menor número de SNPs incluídos nas análises (MIB) apresentou também o menor número de blocos. Assim, as diferenças no número de SNPs incluídos nas análises, quando o número de indivíduos foi ajustado para 15, parecem não ter afetado a determinação dos blocos de haplótipos.

Com base nestas observações, a partir deste ponto, serão apresentados e discutidos somente os resultados que consideraram apenas 15 indivíduos para cada população, exceto Gaucinha e Teresina, que possuem apenas 14 indivíduos amostrados.

### 5.3 Padrões genômicos de blocos de haplótipos e DL

A Tabela 3 apresenta o número de blocos de haplótipos gerados para as 11 populações e 22 autossomos analisados.

Tabela 3. Número de blocos de haplótipos gerados pelo Haploview para todas as populações analisadas.

Cromossomo	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
1	489	457	474	230	201	238	268	142	181	304	187
2	534	538	552	259	244	279	266	180	228	373	230
3	477	430	432	235	256	272	276	168	205	355	223
4	483	464	458	284	265	289	299	184	219	364	215
5	458	434	437	239	231	251	254	179	175	310	217
6	446	437	444	259	223	265	268	178	172	324	178
7	358	329	332	176	156	196	194	126	138	233	134
8	382	416	386	179	203	198	181	142	155	260	148
9	267	255	261	130	120	125	127	89	80	162	111
10	330	318	327	188	182	174	204	127	137	252	141
11	334	348	350	188	181	192	191	133	155	275	155
12	325	277	295	170	160	177	159	114	119	215	137
13	261	280	270	140	141	138	128	100	102	171	117
14	188	176	202	107	102	99	109	76	81	115	85
15	134	134	148	60	62	61	63	67	39	94	59
16	160	134	146	62	62	65	74	44	50	104	55
17	107	107	97	49	49	52	58	36	41	69	39
18	201	180	185	89	85	91	95	60	62	125	69
19	48	52	44	26	21	27	27	20	15	36	17
20	147	131	146	58	57	61	61	34	49	76	53
21	110	104	114	52	45	53	56	40	46	77	50
22	42	44	36	21	16	18	19	14	17	25	13
Total	6281	6045	6136	3201	3062	3321	3377	2253	2466	4319	2633

GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Para a maior parte dos cromossomos, a população CEU apresentou o maior número de blocos de haplótipos, seguida pela população CHB, que apresentou o maior número para os cromossomos 8, 13 19 e 22, enquanto que JPT (japoneses) apresentou mais blocos para os

cromossomos 11, 14, 15 e 21. No total, estas três populações apresentaram quantidades bastante superiores de blocos.

Por outro lado, a população GAU (Gaucinha) apresentou o menor número de blocos para a maioria dos cromossomos, exceto os cromossomos 5, 6, 9, 15 e 19, nos quais MIB (Mimbó) apresentou o menor número, e cromossomo 22, com a população TES (Teresina) apresentando o menor número. Estas três populações, em geral, apresentaram um número bem inferior de blocos (Tabela 3).

As populações africanas LWK, MKK e YRI e a população afro-americana ASW apresentaram os valores intermediários de número de blocos. A única exceção ao padrão geral foi observada no cromossomo 15, em que a população de quilombo GAU apresentou mais blocos que MIB, TES, ASW e as três populações africanas (Tabela 3).

É interessante ressaltar que a população remanescente de quilombo STV (Sítio Velho) apresentou, para todos os cromossomos, um número de blocos apenas inferior ao das populações CEU, CHB e JPT (Tabela 3). Isso se reflete no número total de blocos de haplótipos.

Devido a grande quantidade de blocos gerados para cada população, principalmente para os cromossomos maiores, os resultados serão apresentados e discutidos em maior detalhe apenas para os cromossomos 6 e 22. O cromossomo 22 foi escolhido por ser o menor e, conseqüentemente, ter números menores de blocos, o que facilita a visualização e o entendimento dos resultados. O cromossomo 6, por outro lado, foi escolhido por conter o Complexo Principal de Histocompatibilidade (MHC), uma região que pode influenciar significativamente os padrões de desequilíbrio de ligação ao seu redor em decorrência de fortes pressões seletivas (Ahmad et al., 2003; Walsh et al., 2003; Meyer *et al.*, 2006).

## **5.4 Desequilíbrio de Ligação e Blocos de haplótipos no cromossomo 22**

### **5.4.1 Análise descritiva**

Para o cromossomo 22, a população CHB apresentou o maior número de blocos (44), seguida por CEU com 42 e JPT com 36 (Tabela 3). Entre as populações africanas, o número de blocos variou de 16 (LWK) a 21 (YRI), enquanto que a população afro-americana ASW apresentou 19 blocos. A população com menor número de blocos foi TES (13), seguida por GAU com 14. Assim como observado para os demais cromossomos, a população de quilombo STV apresentou mais blocos (25) do que as demais populações brasileiras, as africanas e a afro-americana.

A extensão total do desequilíbrio de ligação foi determinada para cada população somando-se o comprimento de todos os blocos de haplótipos observados em cada uma delas (Figura 13). A maior extensão foi observada em JPT (868,00 Kb), seguida por CHB com 858,64 Kb e CEU com 819,24 Kb. A população ASW apresentou uma extensão maior (533,74 Kb) do que as três africanas (212,74 Kb a 366,24 Kb). A menor extensão foi observada em GAU (207,94 Kb), enquanto que em STV a extensão foi de 408,20 Kb.



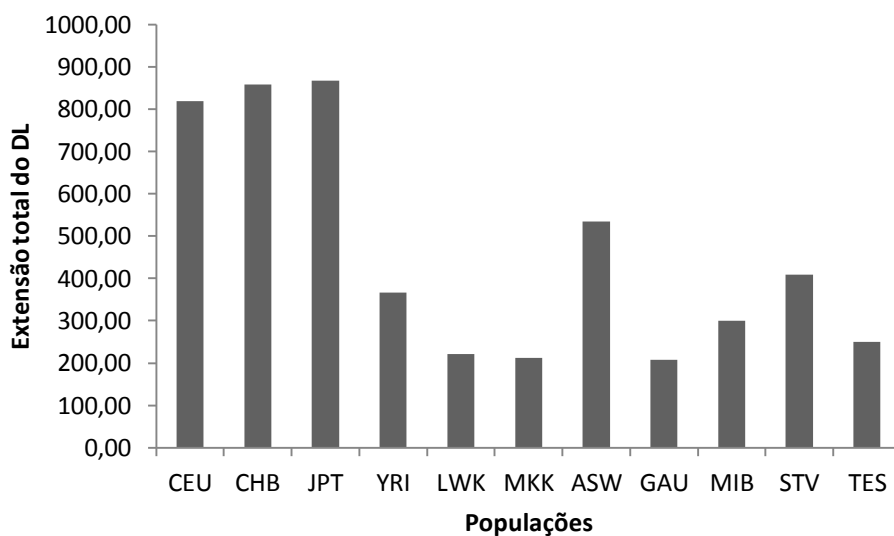


Figura 13. Extensão total do DL para o cromossomo 22, considerando os blocos de haplótipos identificados.

O tamanho médio dos blocos de desequilíbrio de ligação variou de 11,82 Kb em MKK a 28,09 Kb em ASW. Entre os quilombos, GAU apresentou o menor tamanho médio (14,85 Kb), enquanto que MIB apresentou blocos de maior tamanho, em média (17,64 Kb), superando STV (16,33 Kb), embora esta população tenha apresentado na grande maioria das vezes um número maior de blocos (Figura 14).

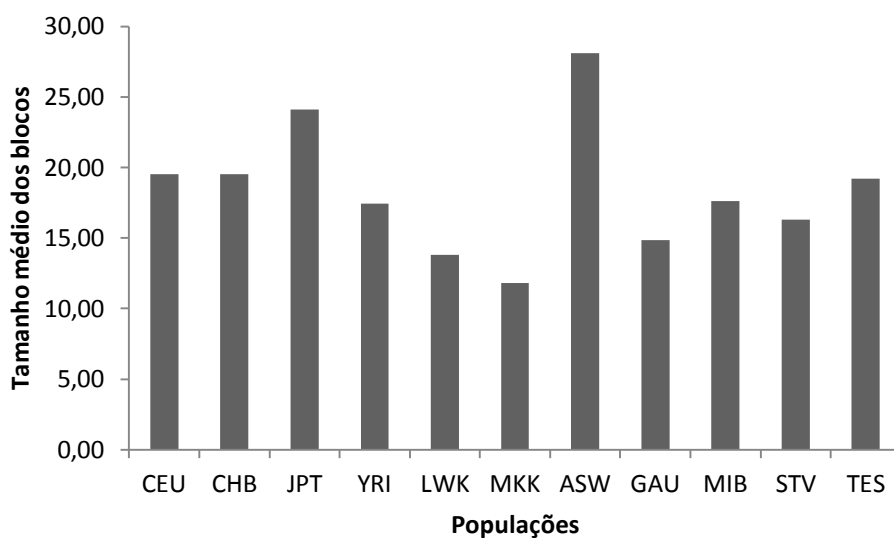


Figura 14. Tamanho médio dos blocos de desequilíbrio de ligação para o cromossomo 22.

Na população MKK, que apresentou o menor tamanho médio de blocos (11,82 Kb), o tamanho do maior bloco observado foi de 58,9 Kb (Figura 15). O bloco de maior tamanho entre todas as populações foi observado na população CEU (157,2 Kb), embora esta não tenha tido o maior tamanho médio de blocos. O bloco mais extenso observado nas três populações de quilombos, com 114 Kb, foi também o maior bloco da população ASW, enquanto que o maior bloco observado em TES foi de 74,7 Kb.

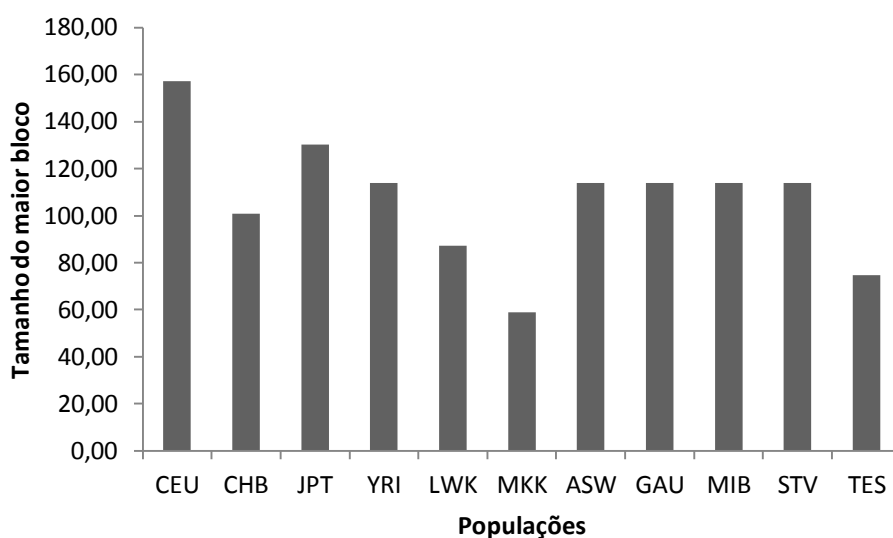


Figura 15. Tamanho do maior bloco de desequilíbrio de ligação para o cromossomo 22.

Os blocos de menor tamanho variaram de 0,04 Kb em CEU, CHB, YRI, MKK, ASW, GAU a 0,3 Kb em MIB (Figura 16).

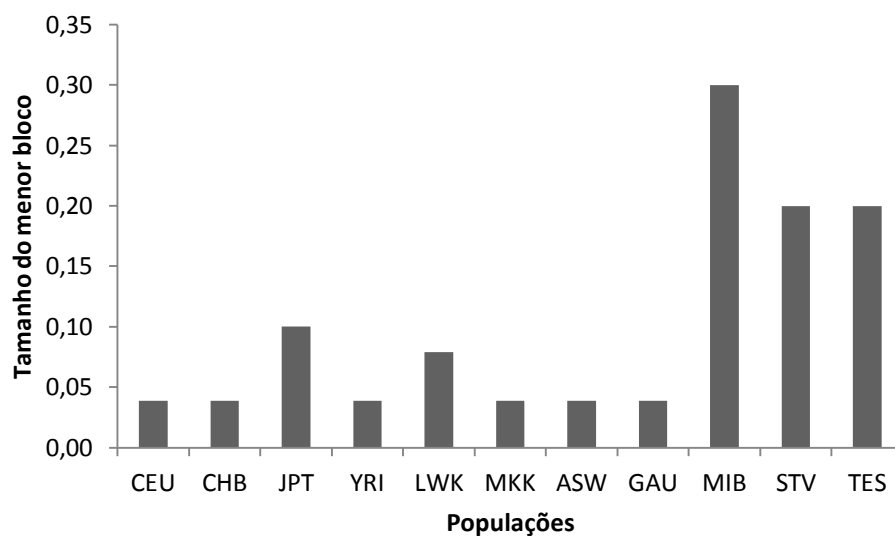


Figura 16. Tamanho do menor bloco de desequilíbrio de ligação para o cromossomo 22.

O número total de SNPs contidos em blocos de haplótipos variou de 52 em GAU a 205 em CHB. O número médio de SNPs por bloco variou de 3,71 em GAU e 3,75 em LWK a 5,0 em JPT. O número máximo de SNPs em cada bloco variou de 6,0 em GAU a 11 em CEU, CHB e JPT e 12 em YRI, enquanto que o número mínimo foi 2,0 em CEU e CHB e 3,0 nas demais populações. Blocos formados por apenas 3,0 SNPs foram os mais frequentes em todas as populações estudadas.

A Tabela 4 sumariza o panorama geral do padrão de DL e de blocos de haplótipos no cromossomo 22 nas diferentes populações consideradas.

Tabela 4. Panorama geral dos blocos de haplótipos no cromossomo 22.

Dados de DL	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
Nº de SNPs analisados	1440	1462	1448	1421	1467	1517	1500	1472	1466	1499	1501
Nº de blocos	42	44	36	21	16	18	19	14	17	25	13
Extensão total do DL (Kb)	819,24	858,64	868,00	366,24	221,08	212,74	533,74	207,94	299,90	408,20	249,80
Tamanho médio dos blocos (Kb)	19,51	19,51	24,11	17,44	13,82	11,82	28,09	14,85	17,64	16,33	19,22
Tamanho do maior bloco (Kb)	157,20	100,80	130,30	113,80	87,20	58,90	114,00	114,00	114,00	114,00	74,70
Tamanho do menor bloco (Kb)	0,04	0,04	0,10	0,04	0,08	0,04	0,04	0,04	0,30	0,20	0,20
Nº total de SNPs nos blocos	185	205	180	83	60	73	88	52	66	96	56
Nº médio de SNPs por bloco	4,40	4,66	5,00	3,95	3,75	4,06	4,63	3,71	3,88	3,84	4,31
Nº máximo de SNPs por bloco	11	11	11	12	7	10	12	6	9	7	9
Nº mínimo de SNPs por bloco	2	2	3	3	2	3	3	3	3	3	3

DL: Desequilíbrio de Ligação; GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

### 5.4.2 Compartilhamento de blocos de haplótipos

Muitos blocos de haplótipos foram compartilhados entre as populações estudadas. A Tabela 5 apresenta o número de blocos compartilhados entre todas as populações.

Tabela 5. Número de blocos de haplótipos do cromossomo 22 compartilhados entre as populações.

	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
CEU	–										
CHB	11	–									
JPT	9	19	–								
YRI	2	4	4	–							
LWK	4	4	3	7	–						
MKK	6	6	3	7	8	–					
ASW	5	4	3	9	6	7	–				
GAU	3	4	2	5	5	6	7	–			
MIB	4	4	2	6	6	8	8	6	–		
STV	5	8	7	6	8	7	8	7	8	–	
TES	3	2	0	0	2	3	0	0	3	1	–

GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

As duas populações asiáticas compartilharam um grande número de blocos (19), sendo o maior número de blocos compartilhados entre todas as comparações. A segunda população com mais blocos compartilhados com estas populações foi CEU: 11 blocos com CHB e nove com JPT. Ao contrário das asiáticas, as três populações africanas não apresentaram muitos blocos comuns. Dentre estas, a população que mais se assemelhou a população afro-americana (ASW) foi YRI, com nove blocos comuns.

Entre os quilombos, MIB e STV apresentaram mais blocos comuns (oito) e MIB foi o quilombo que mais compartilhou blocos com TES. A população GAU compartilhou sete blocos com ASW, enquanto que MIB

e STV compartilharam oito. Considerando-se as três populações africanas, GAU e MIB compartilharam mais blocos com MKK, enquanto que STV apresentou mais blocos comuns com LWK. A população MIB apresentou o mesmo número de blocos compartilhados com CEU e CHB (quatro), enquanto que GAU e STV compartilharam mais blocos com CHB do que com CEU e JPT. Entre todas as populações estudadas, TES foi a que apresentou o menor número de blocos compartilhados.

Em alguns casos, observou-se que o mesmo bloco, que anteriormente estava presente em algumas populações, foi reduzido ao longo do tempo devido aos eventos de recombinação. Na Tabela 6 estão apresentados alguns exemplos. Em geral, as populações que apresentaram os blocos mais extensos foram CEU e CHB. As populações africanas e as brasileiras, por outro lado, apresentaram os menores fragmentos de blocos.

Tabela 6. Blocos de haplótipos compartilhados do cromossomo 22 e que sofreram redução de tamanho em diferentes populações.

Bloco <sup>a</sup>	Tamanho <sup>b</sup> (SNPs)	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
A1	13,7 (4)											
A2	4,0 (3)											
B1	157,2 (7)											
B2	130,3 (6)											
B3	114,0 (6)											
B4	87,2 (5)											
C1	69,4 (4)											
C2	26,5 (3)											
D	3,2 (3)											
E	1,1 (3)											
F1	6,8 (6)											
F2	5,2 (3)											
G1	17,8 (7)											
G2	0,5 (4)											
H1	100,8 (11)											
H2	95,0 (10)											

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica; <sup>b</sup> Tamanho dos blocos em Kb. GAU: Gaucinha; MIB: Mimbó; STV: Sitio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 6. Continuação.

Bloco <sup>a</sup>	Tamanho <sup>b</sup> (SNPs)	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
I1	113,8 (12)				■			■				
I2	84,7 (10)		■							■		
I3	48,4 (8)						■					
I4	74,7 (8)											■
I5	12,9 (4)	■										
I6	37,3 (6)			■								
I7	12,8 (3)										■	
J1	4,1 (6)		■	■	■	■	■	■			■	
J2	3,8 (5)								■	■		
J3	3,6 (4)											■
K1	30,0 (6)		■	■				■			■	
K2	5,7 (3)						■			■		■
L1	60,5 (10)	■										
L2	23,0 (4)			■								
L3	14,0 (3)				■							
L4	38,7 (7)					■					■	
M1	104,5 (10)	■										
M2	67,9 (7)			■							■	
M3	36,7 (5)											■
M4	43,9 (6)		■									
N1	34,9 (6)	■	■			■	■			■		■
N2	0,6 (3)										■	
O1	58,9 (10)	■					■	■				
O2	16,0 (6)								■			
O3	51,8 (9)											■
O4	20,4 (4)			■								

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica; <sup>b</sup> Tamanho dos blocos em Kb. GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Em muitas comparações, observou-se que um bloco que visualmente parecia ser compartilhado entre as populações, isto é, que aparentemente apresentava o mesmo padrão de desequilíbrio de ligação no Haploview, não foi, em determinadas populações, definido de fato como um bloco de haplótipo.

Em TES, por exemplo, o bloco B3 (Tabela 6), embora tenha apresentando valores de desequilíbrio de ligação semelhantes aos das

três populações de quilombos, não foi definido como bloco pelo Haploview (Figura 17).

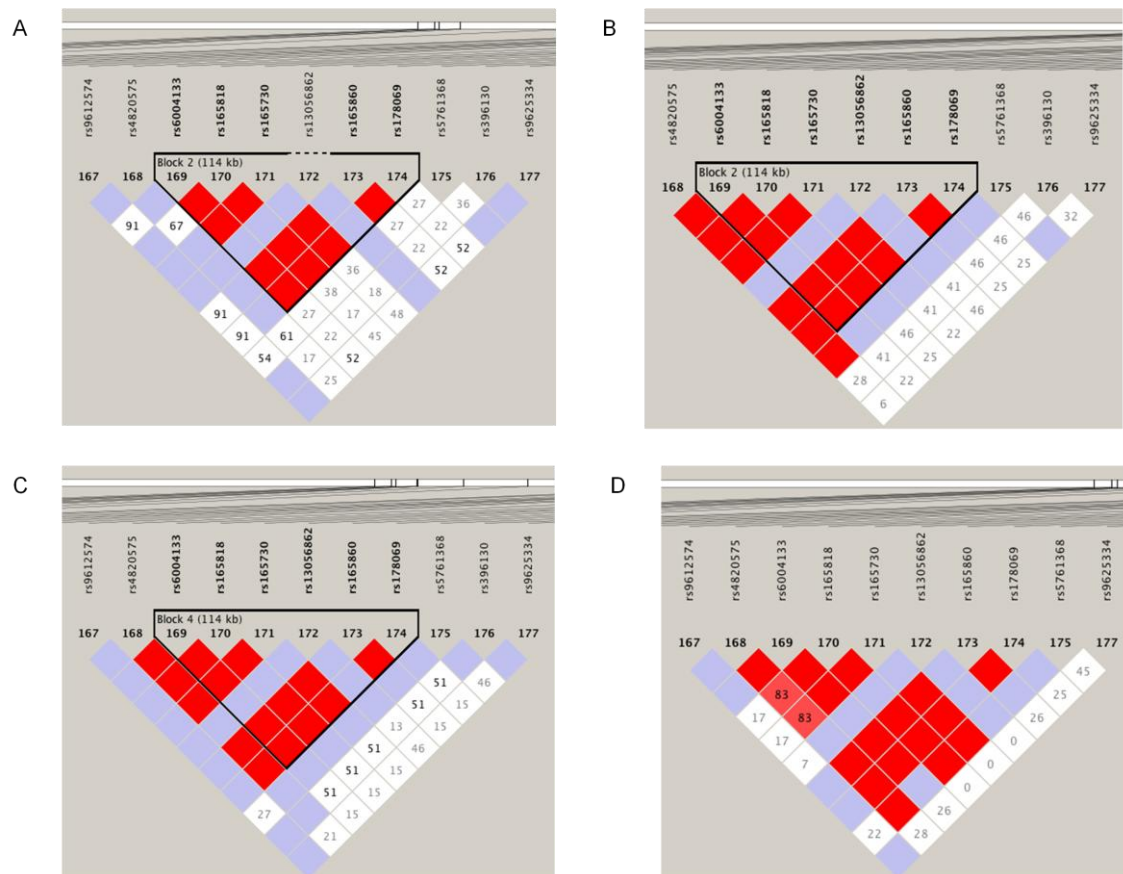


Figura 17. Diferentes definições do bloco B3 do cromossomo 22 pelo programa Haploview. A) GAU; B) MIB; C) STV; D) TES.

Alguns dos haplótipos compartilhados pelas populações de quilombos e pelas populações do HapMap, principalmente as africanas e a afro-americana, tiveram suas frequências consideradas e comparadas entre as diferentes populações (Tabela 7).



Tabela 7. Frequências populacionais dos haplótipos compartilhados do cromossomo 22.

Bloco <sup>a</sup>	Haplótipos	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
B1-B4	ATAGCC	0,467	–	0,567	0,500	0,733	–	0,600	0,607	0,667	0,567	–
	GCGGGT	0,300	–	0,300	0,200	0,167	–	0,367	0,357	0,267	0,267	–
	GCGTGT	0,200	–	0,033	0,133	0,100	–	–	–	0,067	0,167	–
	ACAGCC	–	–	–	–	–	–	0,033	0,018	–	–	–
	ACGGCC	–	–	0,067	–	–	–	–	0,018	–	–	–
	ATGTG–	–	–	0,033	–	–	–	–	–	–	–	–
C1-C2	TAT	0,367	0,533	0,600	0,333	0,233	0,333	0,400	0,393	0,200	0,400	–
	GGC	0,633	0,467	0,400	0,667	0,767	0,667	0,567	0,607	0,767	0,600	–
	TGC	–	–	–	–	–	–	0,033	–	0,033	–	–
	GAC	–	–	–	–	–	–	–	–	–	–	–
D	ATC	–	–	–	0,433	0,400	0,367	0,500	0,714	0,400	0,600	–
	GCA	–	–	–	0,567	0,600	0,633	0,500	0,286	0,600	0,400	–
E	TTT	0,400	0,433	0,400	0,233	0,500	0,233	0,500	–	0,067	0,100	–
	TAT	0,100	0,033	0,100	0,100	0,033	0,067	–	0,464	0,433	0,467	–
	CAG	0,500	0,533	0,500	0,567	0,467	0,700	0,500	0,536	0,500	0,433	–
G1-G2	CCAA	0,433	–	–	0,367	0,367	0,333	0,600	0,321	0,333	0,400	–
	GACT	0,567	–	–	0,633	0,633	0,667	0,400	0,679	0,667	0,600	–

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica. GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Em geral, considerando-se as frequências haplotípicas, o quilombo STV aparentou maior similaridade com a população CEU. O quilombo MIB, por sua vez, aparentou ser mais relacionados às populações africanas. GAU, contudo, não aparentou uma similaridade consistente com nenhuma das populações.

O desequilíbrio de ligação entre os blocos B e C, que distam 913,55 Kb, foi determinado para aquelas populações que não apresentaram outros blocos entre eles. Os valores foram os seguintes:  $D' = 0,41$  em CEU;  $D' = 0,33$  em JPT;  $D' = 0,22$  em LWK;  $D' = 0,14$  em ASW;  $D' = 0,65$  em GAU;  $D' = 0,32$  em MIB;  $D' = 0,13$  em STV e desequilíbrio total ( $D' = 1,0$ ) em YRI.

Entre os blocos E e G1 (distância de 863,91 Kb) os valores de desequilíbrio de ligação foram:  $D' = 0,21$  em YRI;  $D' = 0,14$  em ASW e LWK;  $D' = 0,30$  em GAU e desequilíbrio máximo ( $D' = 1,0$ ) em MKK.

## **5.5 Desequilíbrio de Ligação e Blocos de haplótipos no cromossomo 6**

### **5.5.1 Análise descritiva**

Para o cromossomo 6, a população com o maior número de blocos foi CEU (446), seguida por JPT com 444 e CHB com 437 (Tabela 3). Considerando-se as populações africanas, o número de blocos variou de 223 (LWK) a 265 (MKK), enquanto que a população afro-americana ASW apresentou 268 blocos. A população MIB apresentou o menor número de blocos (172), enquanto que GAU e TES apresentaram 178 blocos. Assim como observado para o cromossomo 22 e os demais, a população de quilombo STV apresentou mais blocos (324) do que as demais populações, exceto a europeia e as asiáticas (Tabela 3 e Figura 2).

A maior extensão total foi observada em JPT (19.456,80 Kb), seguida por CEU com 18.987,25 Kb e CHB com 17.021,00 Kb. A população ASW apresentou uma extensão maior (5.685,19 Kb) do que as três africanas (5.263,90 Kb a 5.353,00 Kb). A menor extensão foi observada em GAU (2.401,40 Kb), enquanto que em STV a extensão foi de 9.558,46 Kb, sendo maior do que as populações brasileiras, africanas e ASW (Figura 18).

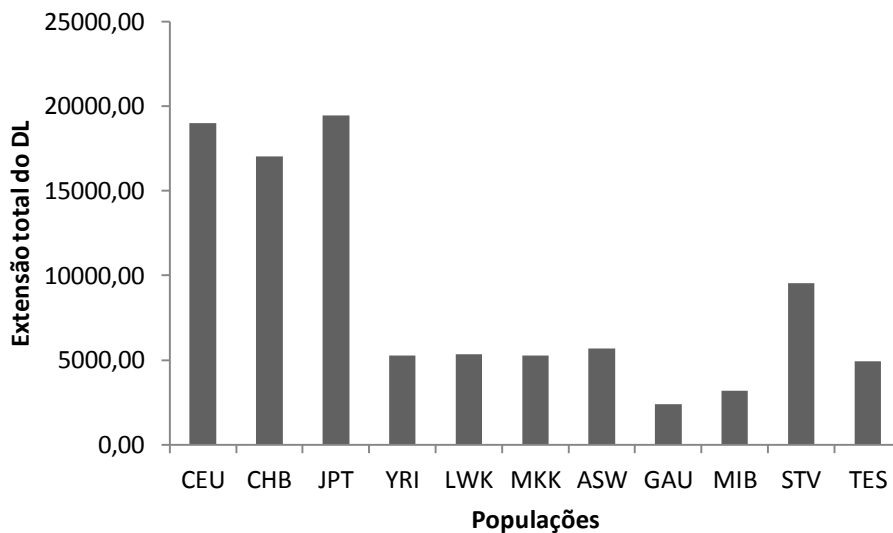


Figura 18. Extensão total do desequilíbrio de ligação para o cromossomo 6, considerando os blocos de haplótipos identificados.

O tamanho médio dos blocos de desequilíbrio de ligação variou de 13,49 Kb em GAU a 43,82 Kb em JPT. A população CEU apresentou um tamanho médio de 42,57 Kb, enquanto que em CHB a média foi de 38,95. Nas populações africanas, a menor média foi observada em MKK (19,86 Kb) e a maior em LWK (24,00 Kb). O tamanho médio dos blocos em ASW foi 21,21 Kb, bem próximo do observado nas três populações africanas. Entre os quilombos, STV apresentou a maior média (29,84 Kb), condizendo com o maior número de blocos e a maior extensão observados nesta população. A média em TES foi de 27,70 Kb (Figura 19).

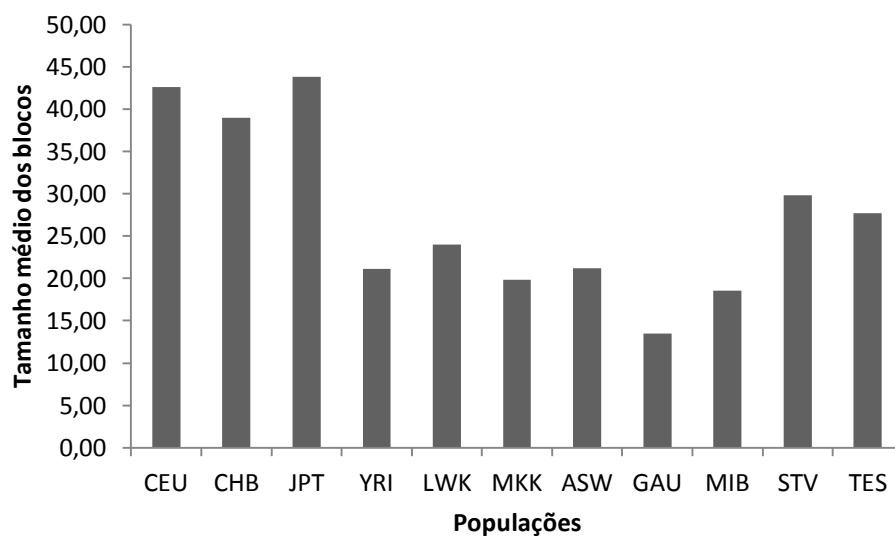


Figura 19. Tamanho médio dos blocos de desequilíbrio de ligação para o cromossomo 6.

O bloco de maior tamanho detectado no cromossomo 6, com 483,90 Kb, foi observado em CEU e STV. Estas populações, contudo, não apresentaram o maior tamanho médio de blocos. TES apresentou o segundo maior bloco, com 465,80 Kb, enquanto que o terceiro maior bloco foi observado em JPT (461,80 Kb) e o quarto em CHB (413,60 Kb). O bloco com maior extensão entre as populações africanas foi detectado em LWK (433,30 Kb). O bloco mais extenso em ASW tem comprimento de 332 Kb e o maior bloco observado em GAU foi de apenas 136,80 Kb, enquanto que em MIB o maior bloco mediu 225,60 Kb (Figura 20).

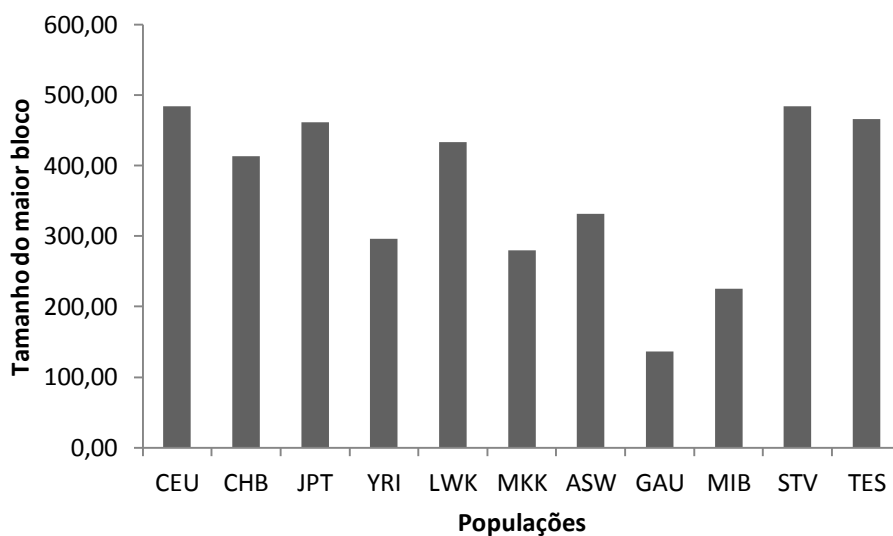


Figura 20. Tamanho do maior bloco de desequilíbrio de ligação para o cromossomo 6.

Os blocos de menor tamanho variaram de 0,01 Kb em CEU a 0,1 Kb nas três populações africanas, GAU e MIB (Figura 21).

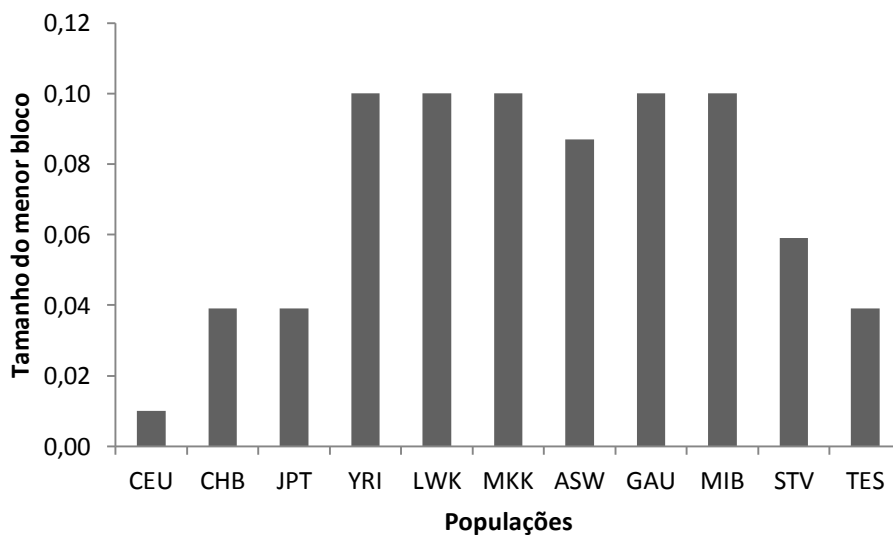


Figura 21. Tamanho do menor bloco de desequilíbrio de ligação para o cromossomo 6.

O número total de SNPs contidos em blocos de haplótipos variou de 741 em GAU a 2.687 em JPT. O número médio de SNPs por bloco variou de 4,16 em GAU a 6,05 em JPT. O número máximo de SNPs em cada bloco

variou de 14 em GAU a 45 em JPT, enquanto que o número mínimo foi 3,0 em MKK, GAU, MIB e TES e 2,0 nas demais populações. Blocos formados por apenas 3,0 SNPs foram observados com maior frequência em todas as populações estudadas.

A Tabela 8 sumariza o panorama geral do padrão de DL e de blocos de haplótipos no cromossomo 6 nas diferentes populações consideradas.

Tabela 8. Panorama geral dos blocos de haplótipos no cromossomo 6.

Dados de DL	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
Nº de SNPs analisados	10913	11017	10898	10907	11006	11282	11236	10981	10844	11115	11244
Nº de blocos	446	437	444	259	223	265	268	178	172	324	178
Extensão total do DL (Kb)	18987,25	17021,00	19456,80	5264,40	5353,00	5263,90	5685,19	2401,40	3195,80	9558,46	4931,13
Tamanho médio dos blocos (Kb)	42,57	38,95	43,82	21,14	24,00	19,86	21,21	13,49	18,58	29,84	27,70
Tamanho do maior bloco (Kb)	483,90	413,60	461,80	296,00	433,30	279,40	332,00	136,80	225,60	483,90	465,80
Tamanho do menor bloco (Kb)	0,01	0,04	0,04	0,10	0,10	0,10	0,09	0,10	0,10	0,06	0,04
Nº total de SNPs nos blocos	2679	2492	2687	1209	1091	1216	1290	741	780	1669	852
Nº médio de SNPs por bloco	6,01	5,70	6,05	4,88	4,89	4,59	4,81	4,16	4,53	5,15	4,79
Nº máximo de SNPs por bloco	32	28	45	18	21	19	17	14	17	23	27
Nº mínimo de SNPs por bloco	2	2	2	2	2	3	2	3	3	2	3

DL: Desequilíbrio de Ligação; GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

### 5.5.1.2 Região do MHC

Diferentemente do observado para todo o cromossomo 6, na região do MHC estendido (7.765,55 Kb), definida de acordo com Horton e colaboradores (2004), o número de blocos detectados foi muito semelhante na maioria das populações, variando de 20 a 22. Apenas TES e as populações de quilombos MIB e GAU apresentaram um número bem inferior de blocos: 7, 11 e 12, respectivamente (Figura 22).

Nesta região, a maior extensão total foi observada em JPT (1.162,90 Kb), seguida por CEU com 930,10 Kb e CHB com 779,60 Kb. A menor extensão foi observada em TES (66,70 Kb), enquanto que em STV a extensão foi de 699,30 Kb, sendo mais uma vez maior do que as demais populações brasileiras, africanas e ASW (Figura 23).

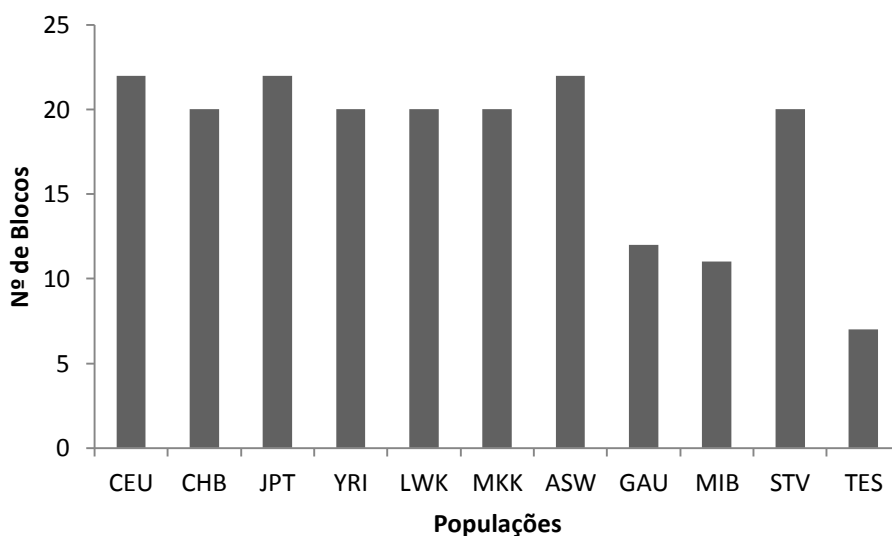


Figura 22. Número de blocos de haplótipos gerados pelo Haploview para a região do MHC no cromossomo 6.



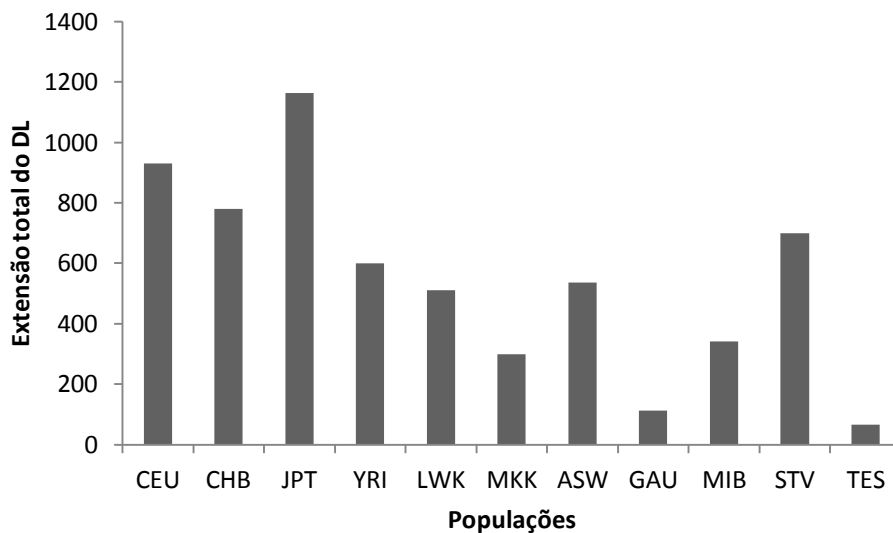


Figura 23. Extensão total do DL para a região do MHC no cromossomo 6, considerando os blocos de haplótipos identificados.

Considerando-se o tamanho médio dos blocos de desequilíbrio de ligação, os resultados foram bem similares aos descritos para todo o comprimento do cromossomo 6, variando de 9,46 Kb em GAU a 52,86 Kb em JPT. A população CEU apresentou um tamanho médio de 42,28 Kb, enquanto que em CHB a média foi de 38,98. Nas populações africanas, a menor média foi observada em MKK (14,99 Kb) e a maior em YRI (30,00 Kb). O tamanho médio dos blocos em ASW foi 24,43 Kb, próximo do observado nas três populações africanas. Entre os quilombos, STV apresentou a maior média (34,96 Kb), condizendo com o maior número de blocos e maior extensão observados nesta população, enquanto que a média em TES foi de 9,53 Kb (Figura 24).

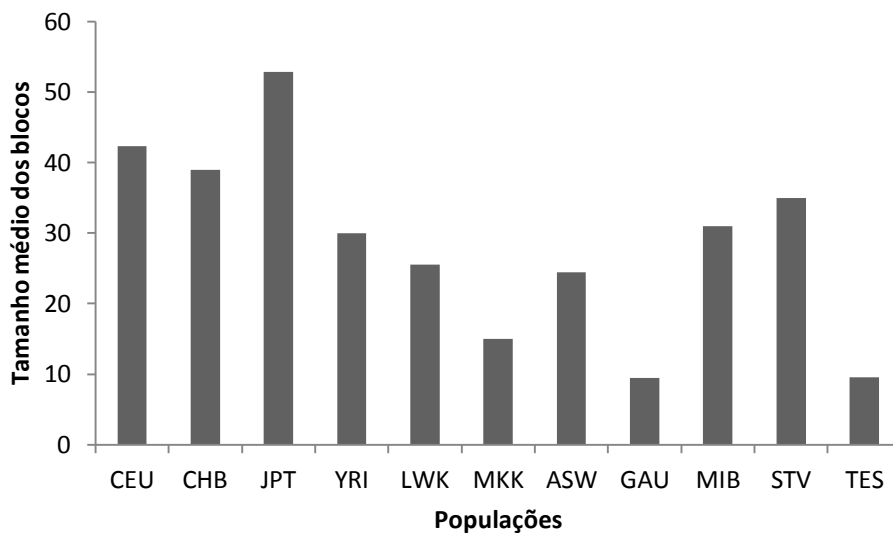


Figura 24. Tamanho médio dos blocos de desequilíbrio de ligação para a região do MHC no cromossomo 6.

O bloco de maior tamanho detectado na região do MHC estendido foi observado em CHB (231,50 Kb). O quilombo STV apresentou o segundo maior bloco, com 194,30 Kb, enquanto que o terceiro maior bloco foi observado em JPT (192,70 Kb) e o quarto em CEU (151,40 Kb). O bloco com maior extensão entre as populações africanas foi detectado em LWK (133,10 Kb). O bloco mais extenso em ASW tem comprimento de 140,10 Kb e o bloco de maior comprimento observado em GAU foi de apenas 25,70 Kb, enquanto que em MIB o maior bloco mediu 142,40 Kb (Figura 25).

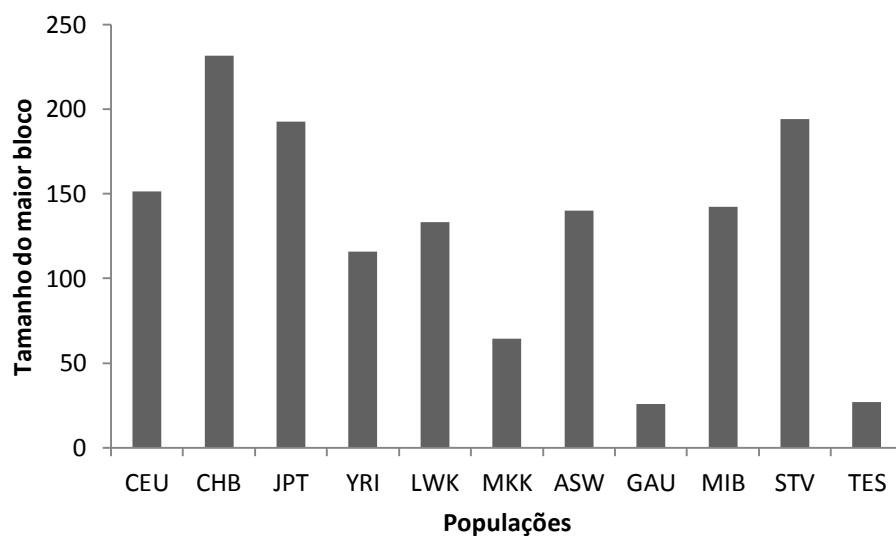


Figura 25. Tamanho do maior bloco de desequilíbrio de ligação para a região do MHC no cromossomo 6.

Os blocos de menor tamanho variaram de 0,10 Kb em JPT e LWK a 1,1 Kb em CHB (Figura 26).

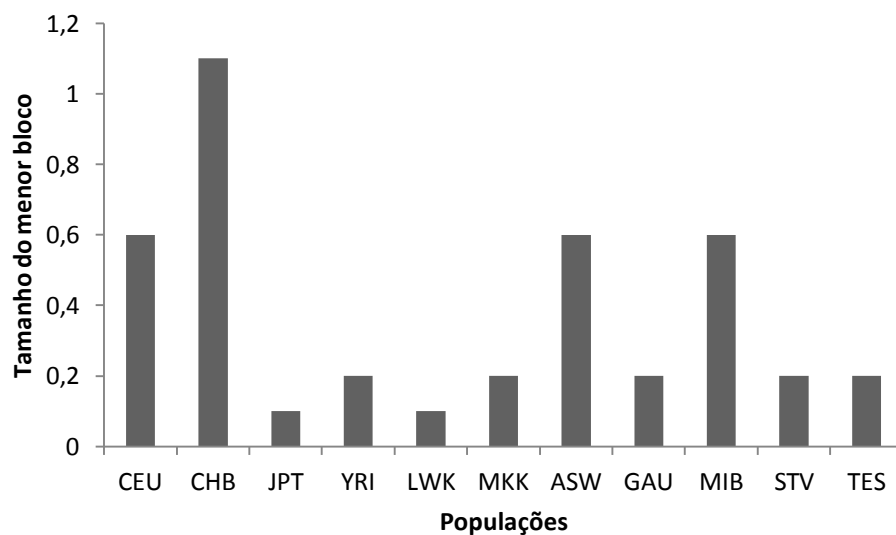


Figura 26. Tamanho do menor bloco de desequilíbrio de ligação para a região do MHC no cromossomo 6.

O número total de SNPs contidos em blocos de haplótipos variou de 30 em TES a 135 em JPT. O número médio de SNPs por bloco variou de 3,50 em GAU a 6,14 em JPT. O número máximo de SNPs em cada bloco variou de 5,0 em GAU a 16 em JPT, enquanto que o número mínimo foi 3,0 em todas as populações, exceto em JPT e ASW, que foi 2,0. Mais uma vez, blocos formados por apenas 3,0 SNPs foram observados com maior frequência em todas as populações estudadas, exceto TES com 4,0.

A Tabela 9 sumariza o panorama geral do padrão de DL e de blocos de haplótipos na região do MHC estendido nas diferentes populações consideradas.

Tabela 9. Panorama geral dos blocos de haplótipos na região do MHC no cromossomo 6.

Dados de DL	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
Nº de blocos	22	20	22	20	20	20	22	12	11	20	7
Extensão total do DL (Kb)	930,1	779,6	1162,9	600,1	511,9	299,8	537,4	113,6	341	699,3	66,7
Tamanho médio dos blocos (Kb)	42,28	38,98	52,86	30	25,56	14,99	24,43	9,46	31	34,96	9,53
Tamanho do maior bloco (Kb)	151,4	231,5	192,7	116	133,1	64,5	140,1	25,7	142,4	194,3	27
Tamanho do menor bloco (Kb)	0,6	1,1	0,1	0,2	0,1	0,2	0,6	0,2	0,6	0,2	0,2
Nº total de SNPs nos blocos	132	105	135	110	102	88	101	42	52	115	30
Nº médio de SNPs por bloco	6	5,25	6,14	5,5	5,1	4,4	4,59	3,5	4,73	5,75	4,28
Nº máximo de SNPs por bloco	12	14	16	11	12	11	11	5	8	16	8
Nº mínimo de SNPs por bloco	3	3	2	3	3	3	2	3	3	3	3

DL: Desequilíbrio de Ligação; GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

### 5.5.2 Compartilhamento de blocos de haplótipos

Assim como descrito para o cromossomo 22, também no cromossomo 6, alguns blocos de haplótipos foram compartilhados entre as populações estudadas (Tabela 10).

Tabela 10. Número de blocos de haplótipos do cromossomo 6 compartilhados entre as populações.

	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
CEU	–										
CHB	88	–									
JPT	84	172	–								
YRI	36	36	42	–							
LWK	48	46	51	96	–						
MKK	56	53	56	83	90	–					
ASW	47	50	49	91	79	86	–				
GAU	34	27	29	34	37	46	35	–			
MIB	28	31	30	43	47	54	51	38	–		
STV	79	76	78	73	70	91	67	49	52	–	
TES	37	30	31	42	35	41	37	31	30	51	–

GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

As duas populações asiáticas compartilharam o maior número de blocos (172), enquanto que CHB compartilhou 88 blocos com CEU e JPT compartilhou 84 blocos. O número de blocos compartilhados entre as populações africanas (83 a 96) foi bem inferior ao número de blocos compartilhados pelas populações asiáticas. A população africana que mais se assemelhou a população afro-americana (ASW) foi YRI, com 91 blocos comuns (Tabela 10).

Entre os quilombos, MIB e STV apresentaram o maior número de blocos comuns (52) e STV foi a população que compartilhou mais blocos com TES (51). A população GAU compartilhou 35 blocos com ASW, enquanto que MIB compartilhou 51 e STV 67 blocos. Em relação às três populações africanas, os três quilombos compartilharam mais blocos com MKK.

O quilombo STV apresentou um número semelhante de blocos compartilhados com CEU, CHB e JPT, enquanto que GAU compartilhou mais blocos com CEU e MIB com as duas populações asiáticas. A população TES apresentou o menor número de blocos compartilhados com as demais (Tabela 10).

A Tabela 11 apresenta o número de haplótipos compartilhados entre as populações estudadas, especificamente na região do MHC estendido.

Tabela 11. Número de blocos de haplótipos na região do MHC compartilhados entre as populações.

	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
CEU	–										
CHB	7	–									
JPT	9	9	–								
YRI	1	4	3	–							
LWK	4	7	8	9	–						
MKK	5	6	7	10	11	–					
ASW	4	8	7	10	10	11	–				
GAU	0	1	1	2	3	4	4	–			
MIB	2	4	3	7	6	4	4	3	–		
STV	5	8	7	7	9	10	6	2	3	–	
TES	4	3	4	3	4	4	2	2	1	3	–

GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

O padrão de compartilhamento de blocos na região do MHC estendido foi diferente do padrão observado nas demais regiões do cromossomo 6. O número de blocos compartilhados entre as três populações africanas (nove entre YRI e LWK e 10 entre YRI e MKK) foi ligeiramente maior do que o número de blocos compartilhados pelas populações asiáticas (nove) e entre CHB e CEU (sete) e JPT e CEU (nove). A população africana que mais se assemelhou a população afro-americana ASW foi MKK, com 11 blocos comuns (Tabela 11).

Entre os quilombos, MIB e STV compartilharam três blocos, mesmo número de blocos comuns entre MIB e GAU. Assim como observado em

todo o cromossomo 6, STV foi a população que compartilhou mais blocos com TES (três). As populações GAU e MIB compartilharam quatro blocos com ASW e STV seis blocos. Em relação às três populações africanas, GAU e STV compartilharam mais blocos com MKK, enquanto que MIB apresentou mais blocos comuns com YRI.

Diferentemente do descrito para o cromossomo 6, STV apresentou na região do MHC um número semelhante de blocos compartilhados com CHB e JPT, mas apresentou um número inferior de blocos comuns com CEU. O quilombo GAU, por sua vez, não compartilhou blocos com CEU. Outra diferença em relação a todo o comprimento do cromossomo 6 é que, no MHC, GAU foi a população que apresentou o menor número de blocos compartilhados com as demais. Como descrito para as demais regiões do cromossomo 6, o quilombo MIB compartilhou o maior número de blocos com as duas populações asiáticas, quando comparadas a CEU (Tabela 11).

A Tabela 12 apresenta os blocos compartilhados pelas populações estudadas na região do MHC estendido (7.765,55 Kb). Os blocos A1-A84 se sobrepõem por toda esta região e em alguns casos são observadas, entre os blocos, regiões de baixo desequilíbrio de ligação e/ou de  $D' = 0$  (que indica recombinação total). Assim, pode-se deduzir que anteriormente havia um único bloco de 7.465,88 Kb nesta região (formado por 536 SNPs) e que o mesmo foi sendo desfeito ao longo das gerações por eventos de recombinação, de acordo com a história evolutiva ocorrida em cada população.



Tabela 12. Blocos de haplótipos da região do MHC que sofreram redução de tamanho em diferentes populações.

Bloco <sup>a</sup>	Tamanho <sup>b</sup> (SNPs)	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
A1	62,0 (6)			■								
A2	12,6 (3)	■	■	■								
A3	15,2 (3)	■	■	■			■				■	
A4	1,30 (4)										■	
A5	37,5 (7)	■										
A6	1,4 (3)		■			■		■	■	■	■	■
A7	37,6 (7)			■								
A8	0,6 (4)						■	■	■	■	■	■
A9	137,9 (10)			■								
A10	88,4 (8)									■		
A11	14,9 (3)		■					■				
A12	34,5 (5)	■					■					
A13	116 (9)				■	■					■	
A14	25,7 (4)								■			
A15	81,5 (5)		■	■								
A16	116,6 (7)	■		■								
A17	1,1 (3)				■	■	■	■	■	■	■	■
A18	85,5 (6)		■									
A19	9,7 (2)							■				
A20	133,1 (12)					■						
A21	53,7 (9)		■									
A22	107,3 (9)				■			■				
A23	108,2 (12)										■	
A24	192,7 (16)			■								
A25	151,4 (12)	■										
A26	142,4 (7)									■		
A27	194,3 (8)										■	■
A28	6,0 (4)				■	■	■	■	■	■	■	■
A29	23,6 (3)									■		
A30	27,0 (4)	■	■	■		■	■				■	■
A31	14,4 (3)								■			
A32	75,3 (7)				■							
A33	140,1 (7)	■		■				■				
A34	231,5 (14)		■									
A35	34,3 (8)					■						
A36	19,3 (3)		■		■			■				
A37	0,8 (3)	■		■								■
A38	14,2 (3)			■	■	■	■	■		■		

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica; <sup>b</sup> Tamanho dos blocos em Kb. GAU: Gaucinha; MIB: Mimbó; STV: Sitio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 12. Continuação.

Bloco <sup>a</sup>	Tamanho <sup>b</sup> (SNPs)	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
A39	10,9 (7)											
A40	76,4 (10)											
A41	28,7 (9)											
A42	3,4 (5)											
A43	26,8 (4)											
A44	6,8 (3)											
A45	31,4 (8)											
A46	0,1 (3)											
A47	6,8 (4)											
A48	2,0 (3)											
A49	5,8 (3)											
A50	179,1 (16)											
A51	1,3 (4)											
A52	69,6 (7)											
A53	64,5 (11)											
A54	42,6 (5)											
A55	24,9 (3)											
A56	99,1 (9)											
A57	9,6 (4)											
A58	26,2 (4)											
A59	23,6 (3)											
A60	2,4 (4)											
A61	50,7 (6)											
A62	12,0 (3)											
A63	0,6 (6)											
A64	23,3 (8)											
A65	7,1 (5)											
A66	20,9 (5)											
A67	15,9 (5)											
A68	15,7 (4)											
A69	29,4 (4)											
A70	0,1 (2)											
A71	28,3 (8)											
A72	5,3 (3)											
A73	22,0 (8)											
A74	15,0 (10)											
A75	47,6 (5)											
A76	65,7 (16)											
A77	1,1 (4)											
A78	48,5 (11)											

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica; <sup>b</sup> Tamanho dos blocos em Kb. GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 12. Continuação.

Bloco <sup>a</sup>	Tamanho <sup>b</sup> (SNPs)	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
A79	22,2 (5)		■	■		■	■	■				
A80	0,2 (3)				■		■		■		■	■
A81	1,1 (3)	■	■	■		■	■	■		■	■	
A82	1,8 (3)						■	■				
A83	16,7 (4)				■	■	■	■			■	
A84	11,8 (3)								■			

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica; <sup>b</sup> Tamanho dos blocos em Kb. GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Em uma região distando 24.995,63 Kb do MHC estendido, foram detectados muitos blocos (B1-B45), grandes e pequenos, intercalados por pontos de recombinação em todas as 11 populações estudadas, indicando que um único grande bloco de haplótipo de 5.428,96 Kb (formado por 155 SNPs) estava presente nestas populações há muitas gerações (Tabela 13).

Tabela 13. Blocos de haplótipos em uma região do cromossomo 6 que sofreram redução de tamanho em diferentes populações.

Bloco <sup>a</sup>	Tamanho <sup>b</sup> (SNPs)	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
B1	209,9 (12)	■	■									
B2	161,6 (9)			■								
B3	166,6 (9)					■	■					
B4	13,5 (3)											■
B5	75,3 (5)				■							
B6	1,0 (3)			■					■			
B7	1,4 (4)		■									
B8	279,4 (8)						■					
B9	354,0 (19)		■									
B10	396,6 (21)					■						
B11	457,9 (32)	■										
B12	332,0 (17)							■				
B13	411,6 (45)			■								
B14	60,1 (13)				■					■		■
B15	31,4 (8)						■					
B16	72,2 (11)	■										
B17	156,5 (16)		■									
B18	66,3 (9)									■		
B19	116,1 (14)							■				
B20	88,0 (7)	■										
B21	160,2 (10)			■								
B22	137,2 (9)											■
B23	276,3 (28)			■								
B24	136,5 (9)				■							
B25	136,8 (11)					■			■			
B26	45,4 (4)									■		
B27	6,0 (4)							■			■	
B28	11,5 (4)		■									

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica; <sup>b</sup> Tamanho dos blocos em Kb. GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 13. Continuação.

Bloco <sup>a</sup>	Tamanho <sup>b</sup> (SNPs)	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
B29	111,3 (14)	■										
B30	11,7 (3)						■					
B31	90,8 (11)									■		
B32	40,1 (7)		■			■	■	■	■			
B33	35,6 (6)				■							
B34	54,5 (7)											■
B35	258,5 (11)											■
B36	262 (12)	■										
B37	47,7 (6)		■	■								
B38	153,9 (11)	■										
B39	40,1 (4)		■									
B40	66,6 (7)			■						■	■	
B41	19,0 (3)				■				■			
B42	45,4 (6)					■	■	■	■			
B43	147,8 (13)											■
B44	26,2 (3)		■									
B45	11,9 (3)	■		■		■	■	■				

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica; <sup>b</sup> Tamanho dos blocos em Kb. GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

As populações que apresentaram os blocos mais extensos foram CEU, CHB e JPT. Com exceção de LWK, as populações africanas apresentaram os menores fragmentos de blocos. Este padrão foi também observado nas populações brasileiras, com exceção de TES, que apresentou três fragmentos grandes de 137,20 Kb, 147,80 Kb e 258,50 Kb (Tabela 13).

As frequências dos haplótipos compartilhados pela maioria das populações estudadas foram comparadas e os resultados são apresentados na Tabela 14.

Tabela 14. Frequências populacionais dos haplótipos compartilhados do cromossomo 6.

Bloco <sup>a</sup> (Kb)	Haplótipos	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
C (4,4)	TAC	0,333	0,400	0,633	0,500	0,467	0,367	0,467	0,250	0,300	0,667	–
	CCT	0,667	0,600	0,367	0,500	0,533	0,633	0,533	0,750	0,700	0,333	–
D (1,0)	ATG	0,267	–	–	0,400	0,567	0,500	0,500	0,429	0,733	0,300	0,464
	GCA	0,733	–	–	0,600	0,433	0,500	0,500	0,571	0,267	0,700	0,536
E (7,0)	GAG	–	0,567	0,500	0,321	0,400	0,433	0,500	0,429	0,333	0,600	0,357
	AGA	–	0,433	0,500	0,679	0,600	0,567	0,500	0,571	0,667	0,400	0,643
F (23,2)	GGTT	0,417	0,500	0,400	0,633	0,533	0,500	0,500	0,393	0,267	0,400	0,500
	CCCC	0,500	0,500	0,500	0,367	0,467	0,500	0,500	0,464	0,633	0,567	0,393
	GGCT	–	–	–	–	–	–	–	0,036	0,100	–	–
	GGTC	–	–	–	–	–	–	–	–	–	0,033	–
	GCCC	–	–	–	–	–	–	–	0,036	–	–	–
	GCCT	–	–	–	–	–	–	–	0,071	–	–	–
	CGTT	–	–	0,100	–	–	–	–	–	–	–	0,036
	CGTC	–	–	–	–	–	–	–	–	–	–	0,036
	CGCC	–	–	–	–	–	–	–	–	–	–	0,036
	GCTT	0,083	–	–	–	–	–	–	–	–	–	–
	G (2,2)	GGCT	0,233	0,333	–	0,567	0,433	0,433	0,600	–	–	–
CCTC		0,767	0,600	–	0,433	0,533	0,567	0,367	–	–	–	–
GCTC		–	–	–	–	0,033	–	–	–	–	–	–
CGTC		–	–	–	–	–	–	0,033	0,393	0,500	0,533	0,714
GCCT		–	0,033	–	–	–	–	–	0,607	0,500	0,467	0,286
CCTT		–	0,033	–	–	–	–	–	–	–	–	–

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica. GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Em geral, o quilombo STV aparentou maior similaridade, em termos de frequência haplotípica, com as populações asiáticas. Os quilombos GAU e MIB, ao contrário, aparentaram ter maior similaridade com as populações africanas.

## 5.6 Análise comparativa dos blocos de haplótipos nos cromossomos 6 e 22

Quando os cromossomos 6 e 22 são comparados, observa-se um padrão semelhante dos blocos de haplótipos entre as diferentes populações.

Em ambos os cromossomos, a população europeia e as duas populações asiáticas apresentaram mais blocos que as demais

populações. A população afro-americana e as três africanas apresentaram números de blocos semelhantes, enquanto que os quilombos GAU e MIB e a população TES apresentaram sempre um número de blocos menor que as demais. Nos dois cromossomos, o quilombo STV apresentou mais blocos que as demais populações brasileiras, que as africanas e ASW, mas foi observado um número inferior ao de observado em CEU, CHB e JPT (Figura 27).

Contudo, no cromossomo 6, as populações com maior e menor número de blocos foram CEU e MIB, respectivamente, enquanto que no cromossomo 22 foram CHB e TES, nesta mesma ordem.

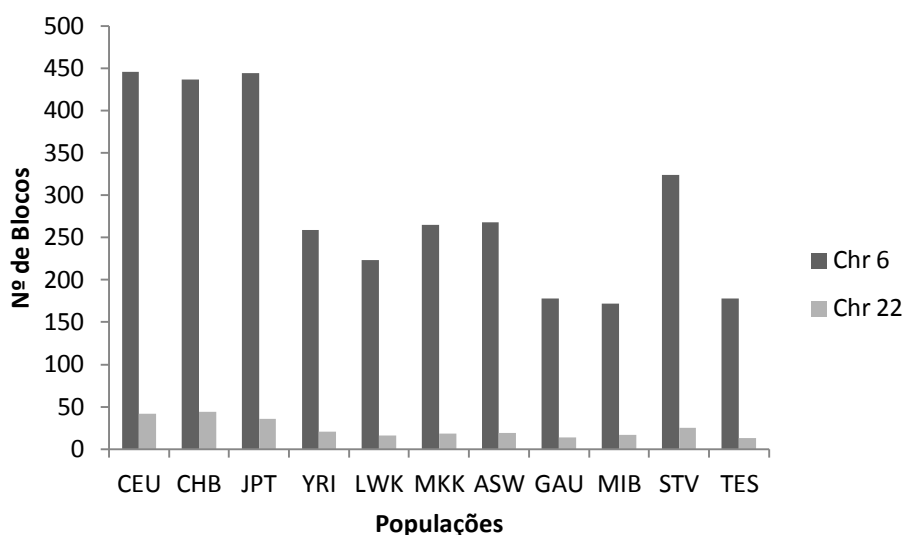


Figura 27. Número de blocos de haplótipos gerados pelo Haploview para os cromossomos 6 e 22 considerando 15 indivíduos de cada população.

A população com maior extensão do DL foi JPT e a que apresentou DL menos extenso foi GAU, em ambos os cromossomos (Figura 28). Para o cromossomo 22, entretanto, ASW apresentou DL mais extenso que as três populações africanas e as quatro brasileiras, enquanto que no cromossomo 6, entre estas mesmas populações, STV apresentou a maior extensão.

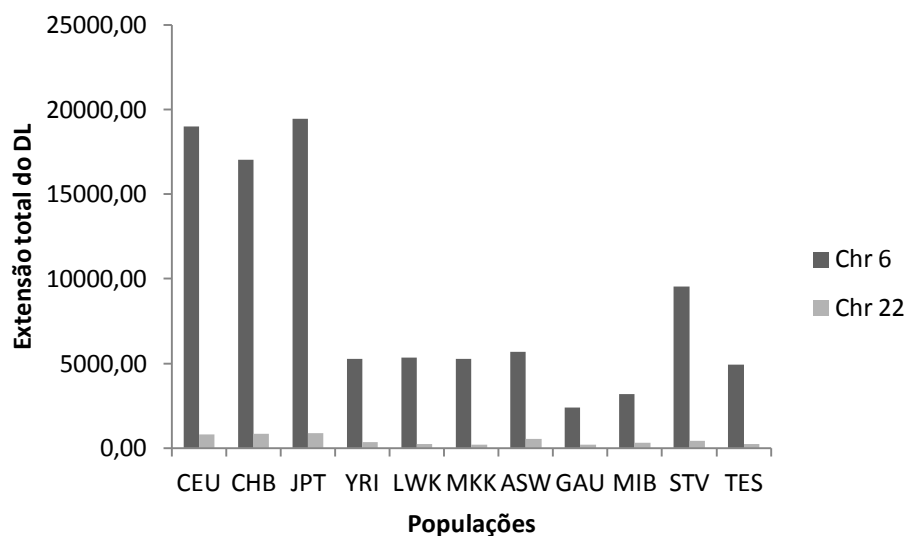


Figura 28. Extensão total do desequilíbrio de ligação para os cromossomos 6 e 22, considerando os blocos de haplótipos identificados em ambos.

No cromossomo 6, a população com maior tamanho médio de blocos foi JPT, mas para o cromossomo 22 foi ASW. Por outro lado, a população com menor tamanho médio de blocos foi GAU no cromossomo 6 e MKK no cromossomo 22 (Figura 29). Entre as populações brasileiras, STV apresentou um tamanho médio de blocos maior para o cromossomo 6, enquanto que para o 22, TES apresentou maior média.



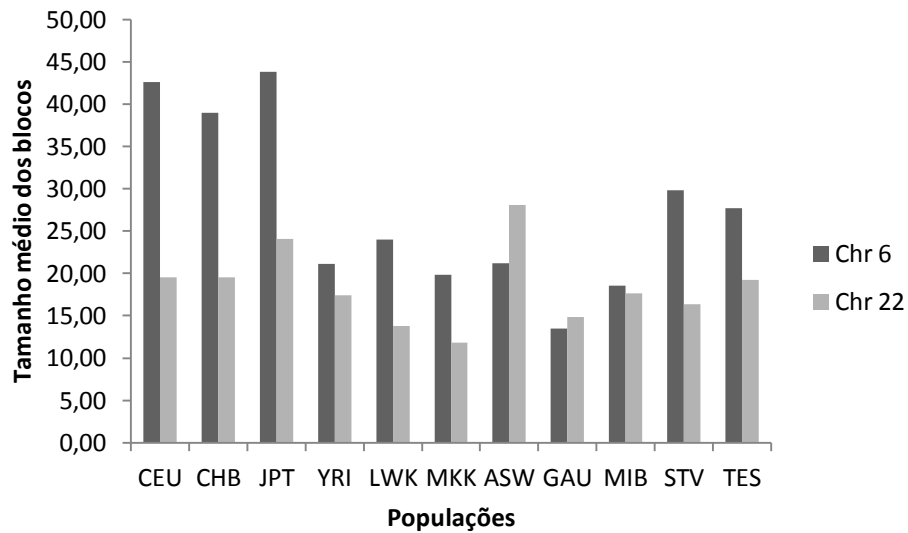


Figura 29. Tamanho médio dos blocos de desequilíbrio de ligação para os cromossomos 6 e 22.

CEU apresentou o bloco de maior tamanho entre todas as populações, tanto no cromossomo 6 quanto no 22. No cromossomo 6, STV compartilhou com CEU o maior bloco (Figura 30).

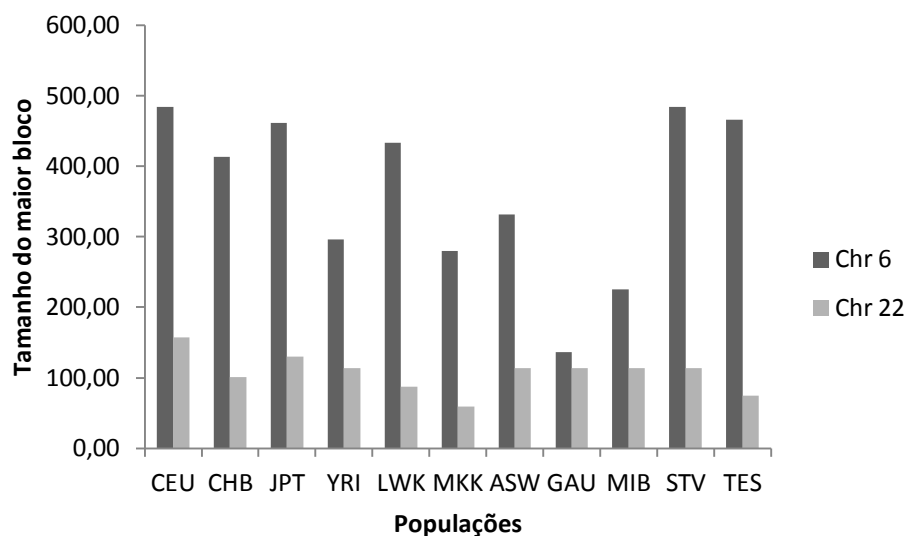


Figura 30. Tamanho do maior bloco de desequilíbrio de ligação para os cromossomos 6 e 22.

No cromossomo 6, o maior bloco observado em TES apresentou comprimento semelhante aos das populações CEU, CHB e JPT, sendo inclusive maior que o de CHB. Entretanto, no cromossomo 22, o maior bloco desta população foi bem menor do que os das três populações citadas.

O maior bloco observado nos quilombos GAU e MIB foi de tamanho inferior ao maior bloco das populações africanas e de ASW, no cromossomo 6. No cromossomo 22, o bloco de maior tamanho nos dois quilombos foi compartilhado com ASW e maior do que o maior bloco das três populações africanas.

De forma geral, os padrões foram bastante similares entre os dois cromossomos analisados. Além disso, considerando o número de blocos (Tabela 3), pode-se afirmar que os padrões foram consistentes entre as populações para todos os cromossomos, com as populações asiáticas e a europeia apresentando maior DL, enquanto que o menor DL foi observado nas populações brasileiras, com exceção de STV.

## 6. DISCUSSÃO

### 6.1 Comparação entre os algoritmos BRLMM e CRLMM

A determinação de genótipos com uma taxa de erro muito baixa, depende da capacidade dos algoritmos utilizados de extrair a maior quantidade possível de informações dos dados brutos, além da capacidade de quantificar a qualidade de cada SNP, fazendo com que SNPs com genótipos não confiáveis sejam eliminados das análises posteriores. Assim, o principal objetivo de um algoritmo de determinação de genótipos é aumentar o número de SNPs que podem ter seus genótipos determinados com grande acurácia.

O algoritmo CRLMM foi utilizado para a determinação dos genótipos em todas as amostras estudadas neste trabalho, por ter apresentado um melhor desempenho, quando comparado ao algoritmo BRLMM, resultando em uma maior acurácia (Figura 8), menor perda de dados e, conseqüentemente, um maior número de SNPs analisados, promovendo um maior poder de resolução nas análises do DL.

Em outro estudo, também foi observado que, com o uso do CRLMM, um menor número de amostras foi excluído das análises, quando os resultados foram comparados ao do algoritmo *Birdseed* ([http://www.affymetrix.com/products/software/specific/birdseed\\_algorithm.affx](http://www.affymetrix.com/products/software/specific/birdseed_algorithm.affx)), desenvolvido para o *Affymetrix SNP Array 6.0*, que consiste em uma evolução do BRLMM (de Andrade *et al.*, 2011).

Lin e colaboradores (2008) também observaram um melhor desempenho do CRLMM em comparação aos algoritmos *default* da Affymetrix (BRLMM e *Birdseed*). O CRLMM gerou genótipos mais precisos em diferentes conjuntos de dados, apresentando uma substancial melhoria na estimativa da acurácia (Lin *et al.*, 2008).

O melhor desempenho do CRLMM se deve, em grande parte, a utilização de dados oriundos do projeto HapMap como dados de

treinamento do algoritmo. Para a maioria dos SNPs no *chip* (64,06%), os genótipos para amostras do HapMap estão disponíveis. Estes genótipos são baseados em um consenso a partir de diferentes tecnologias de genotipagem e, por isso, podem ser considerados como “padrão ouro”, servindo como base na determinação dos genótipos das amostras a serem analisadas (Carvalho *et al.*, 2007).

Uma forma comum de analisar os algoritmos de genotipagem é a comparação de suas taxas de concordância com os genótipos do HapMap e suas taxas de *no call*. A partir destas comparações, o CRLMM proporcionou uma melhor separação dos *clusters* (AA, AB, BB), além de menores taxas de *no call* e maior concordância entre dados oriundos de diferentes laboratórios (Carvalho *et al.*, 2007). Mesmo quando uma classe de genótipos para um determinado SNP não é observado nas amostras do HapMap, o CRLMM consegue determinar este genótipo nas amostras oriundas de outros estudos. Os algoritmos RLMM (Rabbee e Speed, 2006) e BRLMM falharam em cenários como este (Carvalho *et al.*, 2007).

O melhor desempenho do CRLMM na análise de dados gerado pela metodologia da Affymetrix é esperado, uma vez que o algoritmo utiliza estes dados como treinamento. Contudo, o CRLMM superou o BRLMM em precisão mesmo quando dados outros que não da Affymetrix foram testados, além de ter apresentado melhor desempenho quando os dados foram estratificados em genótipos homocigotos e heterocigotos (Lin *et al.*, 2008).

Outra vantagem no uso do CRLMM é que ele facilita a identificação e exclusão de *chips* com dados de baixa qualidade, visto que estes *chips* podem distorcer os dados e provocar erros na determinação dos genótipos, mesmo para os *chips* de alta qualidade. Por outro lado, as medidas de confiança geradas pelo BRLMM não refletem precisamente a qualidade dos *chips*, fazendo com que este algoritmo seja inadequado na determinação da condição dos mesmos (Lin *et al.*, 2008).

Ao final das análises, o usuário deve ter uma medida de acurácia para identificar quais SNPs devem ser excluídos das análises posteriores. Quando se usa o BRLMM, o usuário é quem deve vincular as medidas de confiança geradas pelo algoritmo à acurácia. Contudo, esta vinculação não é direta, uma vez que as medidas parecem corresponder a diferentes graus de acurácia quando o genótipo é homocigoto ou heterocigoto (Lin *et al.*, 2008).

Lin e colaboradores (2008) observaram que as medidas de confiança geradas pelo CRLMM apresentavam limitações semelhantes, quanto à estimativa da acurácia, e propuseram um ajuste *ad-hoc* baseado em uma abordagem de treinamento, que está atualmente implementado no algoritmo. Assim, as medidas de confiança do CRLMM podem ser tratadas como sendo a acurácia predita para cada genótipo (Lin *et al.*, 2008), permitindo que, no presente estudo, apenas os SNPs com genótipos confiáveis pudessem ser incluídos nas análises posteriores de DL e blocos de haplótipos.

## **6.2 Influências do tamanho amostral e do número de SNPs analisados na estimativa do DL**

Foi sugerido que amostras pequenas (abaixo de 100 indivíduos) tendem a superestimar a quantidade de DL porque podem não incluir haplótipos recombinantes raros (Teare *et al.*, 2002; Tenesa *et al.*, 2004), além de promover uma subestimativa da taxa de decréscimo do DL com o aumento da distância física (Dunning *et al.*, 2000). Contudo, demonstrou-se que o DL ao redor dos alelos frequentes pode ser medido com amostras de pequeno tamanho, de 80 a 100 cromossomos (Reich *et al.*, 2001; Weiss e Clark 2002).

Stenzel e colaboradores (2004) estudaram os padrões de DL em diferentes populações (americanos com ancestralidade europeia, alemães, noruegueses, britânicos e afro-americanos) analisando amostras de 45 indivíduos e 320 a 920 SNPs. Os resultados

demonstraram que este tamanho amostral foi suficiente para descrever satisfatoriamente os padrões de DL.

Quando apenas um número ainda menor de indivíduos pode ser estudado, como no caso do presente estudo, a ausência de um maior tamanho amostral pode ser compensada pela genotipagem de um grande número de SNPs (Pritchard e Przeworski, 2001). Assim, pode-se concluir que o estudo de apenas 14 ou 15 indivíduos neste trabalho não prejudicou as análises do DL, uma vez que milhares de SNPs foram analisados.

O uso de haplótipos contendo muitos SNPs aumenta o poder para detectar o DL (Jorde, 2000). Apesar disto, verificou-se que quando as frequências alélicas dos SNPs são  $\geq 5\%$ , o número necessário de SNPs para detecção do DL não é modificado em função do tamanho amostral (Visscher, 2007).

No chip de genotipagem utilizado no presente estudo, a frequência alélica mínima (MAF), ou seja, a frequência do alelo menos frequente é maior que 5% (média de 22%), enquanto que a heterozigose média dos SNPs contidos no chip é de 30%, com base nas populações CEU, CHB, JPT e YRI do projeto HapMap. SNPs com MAF superior a 10-20% (heterozigose entre 20 e 30%, aproximadamente) são os mais indicados para estudos de DL, visto que marcadores mais informativos aumentam o poder de detecção do DL (Service *et al.*, 2006; Kauppi *et al.*, 2003) e fornecem estatísticas com menor variância (Weiss e Clark 2002; Abecasis *et al.*, 2001), tornando a determinação do DL mais realista (Ke *et al.*, 2004; Stenzel *et al.*, 2004).

Por outro lado, SNPs com MAF inferior a 10-20%, em geral, evolutivamente jovens, levam a valores superestimados do DL, principalmente em amostras pequenas (Hanchard *et al.*, 2006).

Deste modo, a fim de evitar a introdução de viés nas estimativas de DL, relacionado ao uso de populações com tamanhos amostrais bastante

variados, foram escolhidos aleatoriamente 15 indivíduos de cada uma das sete populações do HapMap para gerar uma amostragem uniforme. Este procedimento fez com que o número de blocos de haplótipos (Figuras 9 e 10) e de SNPs analisados (Figuras 11 e 12) fosse muito reduzido. Contudo, as diferenças no número de SNPs incluídos nas análises, quando o número de indivíduos foi ajustado para 15, parecem não ter afetado a determinação dos blocos de haplótipos, uma vez que, aparentemente, não houve correlação entre o número de SNPs analisados e o número de blocos de haplótipos gerados.

### **6.3 Padrões genômicos de blocos de haplótipos e DL**

A probabilidade de uma população apresentar uma determinada estrutura de blocos de DL está condicionada aos detalhes de sua história demográfica e, especialmente, da intensidade dos *hotspots* de recombinação (Stumpf e Goldstein, 2003).

Populações que foram submetidas aos mesmos eventos históricos apresentam alta similaridade quanto à extensão do DL (Shifman *et al.*, 2003). Um exemplo são as populações europeias, que compartilham uma longa história comum, com pequenas diferenças genéticas entre elas (Reich *et al.*, 2001). Em geral, populações da mesma região geográfica tendem a ser mais consistentes nos padrões de blocos do que populações oriundas de diferentes regiões (Gu *et al.*, 2007). Contudo, variação nos padrões de DL pode ser vista mesmo entre populações de mesma origem geográfica, como a observada entre duas populações de ancestralidade europeia (UK e CEPH) (Evans e Cardon, 2005).

Esta diferenciação é principalmente devido à história demográfica individual de cada população. Assim, determinados blocos de DL podem estar presentes em algumas populações e ausentes em outras, que possuem histórias demográficas diferentes (Stumpf e Goldstein, 2003). A variação também resulta dos efeitos combinados de fatores genéticos como deriva, mutação e recombinação (Gu *et al.*, 2007).

A variação na estrutura dos blocos de haplótipos entre as populações globais também reflete a considerável diversidade haplotípica entre elas (Sawyer *et al.*, 2005). Os haplótipos observados nas populações não africanas representam um subconjunto daqueles observados na África, indicando que, fora da África, houve acúmulo relativamente pequeno de novos haplótipos gerados por recombinação (Gu *et al.*, 2007). Nas populações africanas, por outro lado, os eventos de recombinação tem se acumulado e os blocos de haplótipos foram sendo desfeitos ao longo do tempo (De La Vega *et al.*, 2005). Consequentemente, as populações africanas apresentam DL menor e menos extenso quando comparadas às demais populações (Kruglyak, 1999; Frisse *et al.*, 2001; Reich *et al.*, 2001; Ke *et al.*, 2004; Bonnen *et al.*, 2006).

Este menor DL nas populações africanas pode ser atribuído ao maior tamanho populacional efetivo a longo prazo, devido a origem africana do homem moderno. Por outro lado, o alto DL mostrado pelas populações não africanas é explicado pelo tempo relativamente curto subsequente ao evento fundador associado à expansão *out of Africa*. Esta expansão está associada a eventos fundadores e gargalos populacionais que exerceram influência preponderante na redução do tamanho efetivo populacional e diversidade haplotípica, não fornecendo muitas oportunidades para a ocorrência de eventos de recombinação (Sawyer *et al.*, 2005).

Em concordância com estas descrições, no presente estudo, as três populações africanas analisadas apresentaram um número de blocos bem inferior ao observado na população europeia e nas duas populações asiáticas (Tabela 3).

Quanto às populações afro-americanas, seus blocos de haplótipos são consistentemente menores do que nas populações de euro-americanos, chineses e japoneses (De La Vega *et al.*, 2005). Foi demonstrado que as afro-americanas apresentam maior recombinação e menor DL do que populações de brancos (Shifman *et al.*, 2003; Ke *et al.*, 2004; Stenzel *et*



*al.*, 2004). Isto parece ser devido à história dessas populações, refletindo suas origens africanas (De La Vega *et al.*, 2005).

Os padrões de blocos de duas populações afro-brasileiras analisadas no presente estudo, GAU e MIB, estão de acordo com estas descrições. Os menores números de blocos e as menores extensões foram observados nestas populações remanescentes de quilombos (Tabelas 3, 4, 8 e 9).

Foi estimado, a partir de marcadores microssatélites, que a contribuição africana na população de MIB é de 52,90% (cromossomo Y) (Wiezel *et al.*, 2013), 81,76% (cromossomo X) (Wiezel, 2003) e 61,10% a 80,80% (autossômicos) (Arpini-Sampaio, 1998; Wiezel, 2003). Para GAU, a estimativa para o cromossomo Y foi de 59,80% (Wiezel *et al.*, 2013) e de 38,41% para os autossômicos (Wiezel, 2003). Assim, a maior contribuição africana nestas duas populações, explica o menor DL observado.

O valor mais baixo de contribuição africana observado em GAU para os marcadores autossômicos foi atribuído ao reduzido tamanho amostral, que teria gerado inconsistências nas estimativas de mistura étnica nesta população. Contudo, foi possível concluir que, nestas comunidades, o elemento fundador masculino foi de origem, predominantemente, africana e europeia. A contribuição ameríndia (10,37% em MIB e 17,17% em GAU, STRs autossômicos) provavelmente se restringiu a indivíduos fundadores do sexo feminino (Wiezel, 2003).

Mistura recente com caucasoides deixariam assinaturas distintas das observadas. Xu e colaboradores (2007) observaram que a extensão do DL, analisada a partir de 24.341 SNPs do cromossomo 21 (36,87 Mb), em uma população afro-americana (*Coriell Cell Repositories*) foi muito similar a da população africana do HapMap (YRI). Por outro lado, quando foram analisados apenas 344 SNPs com grandes diferenças em suas frequências alélicas entre CEU e YRI, a extensão do DL na

população afro-americana foi maior do que a observada nas populações europeia e africana (Xu *et al.*, 2007).

Outros dois estudos demonstraram que um DL aumentado em afro-americanos estava, provavelmente, correlacionado com o aumento das diferenças nas frequências alélicas dos marcadores entre as populações parentais. A partir do estudo de 200 indivíduos afro-americanos de Detroit, Estados Unidos, foi sugerido que uma miscigenação recente com caucasianos mantém um alto e extenso DL em escala genômica nos afro-americanos, sendo inclusive maior do que o DL observado em caucasianos (Rybicki *et al.*, 2002). Foi também descrito um maior DL em uma amostra composta por afro-americanos de diversas localidades dos Estados Unidos, quando comparada às populações parentais. Concluiu-se que o maior DL observado resultou de um contínuo fluxo gênico europeu na população afro-americana (Collins-Schramm *et al.*, 2003).

Estes resultados, entretanto, não podem ser generalizados para populações afro-americanas nas quais a contribuição caucasiana não é predominante, como é o caso das populações GAU (44,42% para os marcadores autossômicos e 31,00% para o cromossomo Y) e MIB (8,83% a 22,37% nos autossômicos e 45,10% no cromossomo Y) (Arpini-Sampaio, 1998; Wiezel, 2003; Wiezel *et al.*, 2013). Embora estas comunidades tenham recebido, no momento de sua fundação, uma considerável contribuição europeia masculina, elas se mantiveram isoladas após sua fundação, e os casamentos posteriores se deram entre indivíduos com herança predominantemente africana. Isto pode explicar os índices maiores de contribuição europeia observados apenas no cromossomo Y. Assim, estes quilombos apresentam padrões de DL mais similares aos apresentados pelas populações africanas, das quais receberam maior contribuição genética.

A população de Detroit é reconhecida por apresentar taxas maiores de miscigenação com caucasianos do que as populações do leste e sul dos

Estados Unidos (Rybicki *et al.*, 2002). Menor nível de contribuição europeia pode explicar o fato da população afro-americana do HapMap (ASW), analisada no presente estudo, ter apresentado, em geral, um número de blocos semelhante ao observado nas três populações africanas analisadas neste trabalho (Tabelas 3, 4, 8 e 9). As amostras da população ASW são oriundas de indivíduos residentes da região sudoeste dos Estados Unidos, que não apresenta alto grau de ancestralidade europeia. Assim, esta população se assemelha mais às populações africanas, quanto ao padrão de blocos.

Ao contrário das populações afro-americanas analisadas no presente estudo, a população urbana TES, apresenta altos níveis de ancestralidade europeia: 79,50%, 58,80% e 51,58%, estimados para os cromossomos Y (Wiezel *et al.*, 2013), X e autossômicos, respectivamente (Wiezel, 2003). Consequentemente, de acordo com o exposto, seria esperado que a mesma apresentasse um alto nível de DL e um número de blocos similar ao observado na população europeia analisada neste estudo. Contudo, esta população apresentou padrões de blocos semelhantes a GAU e MIB (Tabelas 3, 4, 8 e 9).

Uma possível explicação para esta observação é que, ao contrário das populações de quilombos, que representam isolados genéticos, TES teria sofrido sucessivos eventos de miscigenação ao longo de sua história, o que diminuiu o DL ao longo das gerações. Os níveis de DL aumentam durante as primeiras gerações, uma vez que os constantes eventos de miscigenação geram mais DL, impedindo o decréscimo do DL ocasionado pela recombinação entre os loci. Contudo, após algumas gerações, o DL começa a decair (Pfaff *et al.*, 2001).

O decréscimo do DL provavelmente se inicia quando a população miscigenada começa a se assemelhar a população parental a partir da qual recebeu contribuição genética contínua (Pfaff *et al.*, 2001). Aparentemente, foi o que aconteceu em Teresina. A população se tornou bastante semelhante à população europeia e, conseqüentemente, o seu

DL diminuiu. Todavia, a redução do DL em TES deveria ocorrer até o nível de DL da população parental europeia. Entretanto, os níveis de DL em TES são bem inferiores ao observado em CEU.

Esta acentuada redução do DL pode ser explicada pelo evento fundador ocorrido em TES. Comparada aos quilombos, a fundação de TES envolveu um maior número de indivíduos e foi seguida por um rápido crescimento. Foi demonstrado que valores intermediários de DL podem estar relacionados ao crescimento contínuo de uma população (Pistis *et al.*, 2009). As populações em expansão ou recentemente expandidas sofrem um rápido decréscimo do DL em um intervalo de tempo relativamente curto (Laan e Pääbo, 1997; Angius *et al.*, 2008). Um rápido crescimento aumenta a diversidade genética e a recombinação, enquanto que a deriva genética será mínima, resultando na diminuição do DL (Jorde, 2000; Pritchard e Przeworski 2001; Hall *et al.*, 2002; Aulchenko *et al.*, 2004).

Assim como os quilombos GAU e MIB, STV apresenta um alto percentual de contribuição africana: 60,20%, 96,39% e 48,09%, para os cromossomos Y (Wiezel *et al.*, 2013), X e autossômicos, respectivamente (Wiezel, 2003). Contudo, diferentemente dos outros dois quilombos, STV apresentou, para todos os cromossomos, um grande número de blocos, inferior apenas ao número observado nas populações CEU, CHB e JPT (Tabelas 3, 4, 8 e 9). Tendo ancestralidade predominantemente africana, os altos níveis de DL em STV não podem ser explicados por uma maior contribuição europeia, como sugerido por alguns autores (Rybicki *et al.*, 2002; Collins-Schramm *et al.*, 2003; Xu *et al.*, 2007).

Assim, a explicação para o maior DL observado em STV, em relação aos demais quilombos, consiste na história demográfica peculiar desta população.

Foi sugerido que três tipos de populações apresentam altos níveis de DL: 1) populações fundadas a partir da miscigenação de populações com diferentes frequências alélicas (Rybicki *et al.*, 2002; Collins-

Schramm *et al.*, 2003; Xu *et al.*, 2007); 2) populações pequenas estáveis (Service *et al.*, 2006; Angius *et al.*, 2008; Pistis *et al.*, 2009); 3) populações isoladas fundadas por um pequeno número de indivíduos, que experimentaram um rápido crescimento (Shifman *et al.*, 2003; Varilo *et al.*, 2003; De La Vega *et al.*, 2005). Estas características foram observadas na população STV, com exceção do rápido crescimento.

A miscigenação no momento da fundação de uma população gera alto e extenso DL (Laan e Pääbo, 1997; Hall *et al.*, 2002; Service *et al.*, 2006; Xu *et al.*, 2007). O quilombo STV experimentou um maior grau de miscigenação no momento da sua fundação, quando comparado a GAU e MIB. Embora também fundada por poucos indivíduos, o número de fundadores em STV foi maior do que nos demais quilombos, além de serem oriundos de áreas mais amplas, com uma população mais diversificada. GAU e MIB, por outro lado, foram fundados por um número muito pequeno de indivíduos, descendentes diretos de negros da região (Arpini-Sampaio, 1998).

Além disso, a partir da análise de STRs autossômicos, STV apresentou um maior índice de contribuição ameríndia (27,96%), quando comparado a GAU (17,17%) e MIB (10,37%) (Wiesel, 2003), o que evidencia o maior grau de miscigenação sofrido por este quilombo no momento de sua fundação. A maior contribuição ameríndia em STV pode ter contribuído para o maior DL observado neste quilombo, uma vez que as populações americanas nativas, incluindo as brasileiras (Leite *et al.*, 2009), são reconhecidas por apresentar altos níveis de DL e grandes blocos de haplótipos (Sawyer *et al.*, 2005; Conrad *et al.*, 2006; Gu *et al.*, 2007; Wang *et al.*, 2010; Amorim *et al.*, 2011).

Outra característica observada em STV que gera um alto DL é a estabilidade da população resultante do isolamento. Populações que experimentaram um longo período de tamanho constante no momento de sua fundação apresentam DL aumentado. Um crescimento lento no início (nas primeiras gerações) aumenta a perda de diversidade genética

durante o início da história de uma população isolada, resultando em um gargalo mais estreito, aumentando a extensão do DL nestas populações (Kruglyak, 1999; Jorgensen *et al.*, 2002).

Este cenário foi mais evidente em STV do que nos outros dois quilombos estudados, o que pode explicar o maior DL observado nesta população. Quando comparada a GAU e MIB, esta comunidade foi a que menos se diferenciou ao longo do tempo. STV apresentou o menor grau de variabilidade entre as gerações e a menor proporção de variabilidade genética atribuída à diferenciação da população ao longo do tempo (Arpini-Sampaio, 1998), além de ter apresentado a menor diversidade haplotípica, estimada por meio de marcadores autossômicos e do cromossomo Y (Wiesel, 2003).

Diferentemente do observado para todo o cromossomo 6, bem como para o cromossomo 22, na região do MHC estendido o número de blocos detectados foi muito semelhante entre a maioria das populações estudadas (Tabelas 4, 8 e 9). Uma similaridade nos padrões de DL nesta região foi também descrita para as populações africana e caucasiana (Hanchard *et al.*, 2006). Isto se deve ao fato desta região estar submetida a fortes pressões seletivas, fazendo com que os padrões de DL sejam conservados, mesmo entre as diversas populações globais (Ardlie *et al.*, 2002). A seleção epistática para combinações de alelos em dois ou mais loci pode influenciar os padrões de DL através dos haplótipos (Ahmad *et al.*, 2003). Walsh e colaboradores (2003) descreveram que três haplótipos estendidos desta região estavam associados com doenças autoimunes. Embora esta associação seja vista no presente, foi sugerido que, sob certas condições no passado, as variantes funcionais nestes haplótipos tenham sido benéficas, conferindo resistência a doenças, tendo sido, portanto, submetidas à seleção positiva.

No presente estudo, a região do MHC apresentou diversos blocos sobrepostos, muitos com extensão maior do que 100 Kb. A extensão do

DL nesta região foi similar à observada em todo o comprimento do cromossomo 22. Meyer e colaboradores (2006) descreveram altos níveis de DL na região do MHC, entre os loci HLA separados por grandes distâncias físicas, e sugeriram que o extenso DL resulta da seleção natural e dos eventos demográficos. Além da seleção natural, o DL mais extenso no MHC parece ser devido às taxas de recombinação reduzidas nesta região (Walsh *et al.*, 2003).

#### **6.4 Compartilhamentos de blocos de haplótipos**

A existência de haplótipos compartilhados por várias populações sugere que muito da estrutura dos haplótipos atuais é antiga e foi estabelecida antes da diversificação recente da humanidade. Como resultado, a estrutura de blocos de DL pode ser similar entre diversas populações (Kauppi *et al.*, 2003).

Em geral, os padrões dos blocos de haplótipos são muito similares entre as populações do leste da Ásia (que apresentam grande similaridade entre elas) e as populações da região ocidental da Eurásia (incluindo a Europa). Assim, pode-se considerar que as populações localizadas nas duas regiões compartilham uma única estrutura de DL (González-Neira *et al.*, 2004). Os padrões de blocos para as populações analisadas neste estudo estão de acordo com o descrito: as duas populações asiáticas compartilharam um grande número de blocos, mostrando grande similaridade, e ambas compartilharam muitos haplótipos com a população europeia CEU (Tabelas 5 e 10).

Por outro lado, neste estudo, as três populações africanas não apresentaram muitos blocos comuns, em relação ao número total de blocos detectados (Tabelas 5 e 10). Isto se deve a heterogeneidade genética observada entre as populações africanas. Nestas populações, há uma maior diversidade haplotípica e, conseqüentemente, é observado um menor compartilhamento de blocos e significativas diferenças nos padrões de DL (González-Neira *et al.*, 2004).

Entre os quilombos, MIB e STV apresentaram mais blocos comuns (Tabelas 5 e 10). Isto pode ser explicado pela menor distância genética entre estas populações quando comparadas com GAU, a despeito do maior DL observado em STV. Embora esta comunidade tenha experimentado diferentes eventos em sua fundação, que geraram altos níveis de DL, STV apresenta características comuns aos demais quilombos, como o isolamento e maior contribuição africana. A partir de dois microssatélites, a distância estimada foi de 4,10% entre MIB e STV; 6,12% entre GAU e MIB e de 6,31% entre GAU e STV (Arpini-Sampaio, 1998). Quando foram estudados oito locos proteicos, as distâncias foram de: 0,48%, 1,51% e 1,75%, respectivamente (Arpini-Sampaio, 1998). A partir dos locos de microssatélites, STV foi agrupado com TES, mas ambos próximos de MIB, enquanto que a partir dos locos proteicos, MIB e STV formaram um único agrupamento (Arpini-Sampaio, 1998). A partir de análises de microssatélites dos cromossomos Y e X, MIB e STV também foram colocados em um mesmo grupo, revelando a maior similaridade entre estas duas comunidades (Wiezel, 2003).

Os três quilombos compartilharam poucos blocos com TES (Tabelas 5 e 10), considerando-se o número total de blocos detectados em cada população. Isto se deve ao fato de TES apresentar maior contribuição europeia e os três quilombos apresentarem contribuição predominantemente africana. A partir da análise de microssatélites autossômicos e dos cromossomos X e Y, TES foi sempre colocada no mesmo grupo das populações europeias, enquanto que os três quilombos se agruparam sempre ao lado ou no mesmo grupo das populações africanas (Wiezel, 2003).

Embora apresentem contribuição majoritariamente africana, os três quilombos não compartilharam um grande número de blocos com as populações africanas analisadas, como seria esperado (Tabelas 5 e 10). Isto pode ser explicado pelo fato de que os haplótipos observados nas populações não africanas representam apenas um subconjunto daqueles observados na África, como já descrito (Gu *et al.*, 2007). De La



Vega e colaboradores (2005) observaram uma significativa proporção de haplótipos não compartilhados entre afro-americanos e africanos, embora os mapas de DL tenham apresentado formas similares.

Poucos haplótipos foram compartilhados pelas populações estudadas na região do MHC. Dentre os 84 blocos detectados nesta região, o maior número de blocos compartilhados foi 11 (entre LWK e MKK e entre MKK e ASW). Além disso, existe pouca evidência de haplótipos compartilhados por várias populações (Tabela 11). Isto sugere que esta região tem sido submetida a uma alta taxa de *turnover* de haplótipos e que as estruturas dos haplótipos atuais não foram estabelecidas no início da história humana. Contudo, observou-se que a estrutura dos blocos de DL permaneceu constante, quando foram comparadas três populações distintas (Reino Unido, Saami e Zimbabwe), sugerindo que os *hotspots* de recombinação têm se mantido suficientemente ativos para evitar a fusão dos blocos, mesmo em populações jovens (Kauppi *et al.*, 2003). Conseqüentemente, são observados nesta região, determinados blocos de DL que em algumas populações apresentam-se incompletos, mas que ainda incluem associações altamente significativas em toda a sua extensão (Kauppi *et al.*, 2003), como observado no presente estudo.

Considerando-se as frequências haplotípicas (Tabelas 7 e 14), é difícil relacioná-las diretamente ao DL. Em geral, os mesmos poucos haplótipos são os mais frequentes em quase todas as populações, mas suas frequências diferem consideravelmente, mesmo entre as populações não africanas. Assim, as diferenças no DL entre as populações africanas e as demais não estão correlacionadas com as diferenças nos haplótipos observados, mas sim com as diferenças nas frequências dos poucos haplótipos presentes (Sawyer *et al.*, 2005).

No presente estudo, os quilombos GAU e MIB aparentaram maior similaridade com as populações africanas, quanto às frequências haplotípicas (Tabelas 7 e 14). Isto está de acordo com a maior

contribuição africana na formação destas comunidades. O quilombo STV, por outro lado, aparentou maior similaridade com a população CEU (cromossomo 22) e com as populações asiáticas (cromossomo 6). Uma provável explicação seria a maior miscigenação ocorrida nesta população e a maior contribuição ameríndia, quando comparada aos demais quilombos.

### **6.5 Análise comparativa dos blocos de haplótipos nos cromossomos 6 e 22**

Alguns estudos têm descrito padrões de DL similares para diferentes regiões do genoma (Dunning *et al.*, 2000; Hall *et al.*, 2002), inclusive sugerindo a extrapolação de dados de um cromossomo para outras regiões (Service *et al.*, 2007).

Contudo, foi demonstrado que a distribuição do DL parece não ser uniforme dentro e entre os cromossomos (Stumpf e Goldstein 2003; De La Vega *et al.*, 2005). A variabilidade no DL observada entre as diferentes regiões genômicas pode ser explicada pela presença ou ausência de *hotspots* de recombinação em cada região (Shifman *et al.*, 2003).

A partir da análise de 24.940 SNPs, distribuídos por quase todo o comprimento dos cromossomos 6, 21 e 22, em quatro populações (afro-americanos, euro-americanos, chineses e japoneses), foi descrito que o cromossomo 6 apresenta uma taxa de decréscimo do DL 26-58% mais lenta do que os outros dois cromossomos. A extensão total do DL foi igual para os cromossomos 21 e 22 e inferior a do cromossomo 6. Isto foi atribuído ao maior tamanho do cromossomo 6 e ao fato deste apresentar taxas de recombinação menores e, conseqüentemente, diminuição mais lenta do DL, refletindo a ação da seleção natural (De La Vega *et al.*, 2005). No presente estudo, também foi observada uma extensão muito maior no cromossomo 6 em relação ao 22. Além disso, o cromossomo 22 inteiro (~51 Mb), mesmo contando com cerca de 1500 SNPs considerados na análise, possui extensão total de LD menor do

que a da região do MHC (~ 7,8 Mb) em 8 das 11 populações consideradas (Tabelas 4 e 9).

Foi descrito que os níveis de DL são muito variáveis em toda a extensão do cromossomo 6, desde blocos com forte DL interno até trechos de DL muito baixo. Áreas de alto DL sem uma estrutura clara de blocos e pares isolados de marcadores com forte e extenso DL também foram observados. Os *hotspots* de recombinação ocorrem entre os blocos de DL, nos quais a recombinação é suprimida e os marcadores estão em forte DL (Kauppi *et al.*, 2003; Stumpf e Goldstein 2003; Stenzel *et al.*, 2004; Xiang *et al.*, 2005). Padrões semelhantes foram observados no presente estudo.

Dentro da região do MHC, o maior bloco descrito se estendeu por 132 Kb na população de Cingapura (Xiang *et al.*, 2005). No presente estudo, o maior bloco detectado nesta região foi também observado em uma população de chineses, CHB (231,50 Kb). As populações CEU, JPT, LWK, ASW e os quilombos STV e MIB também apresentaram blocos maiores que 132 Kb. Estes grandes blocos podem ser explicados pela história evolutiva de cada população, com uma menor ocorrência de eventos de recombinação.

Em resumo, considerando-se o número de blocos de haplótipos detectados em cada população estudada, o padrão foi semelhante entre todos os cromossomos. Isto mostra que os padrões de DL são modelados principalmente pela história evolutiva de cada população, mais do que por fatores genômicos locais, como recombinação, mutação e seleção.

## 7. CONCLUSÕES

- O algoritmo CRLMM apresentou um melhor desempenho quando comparado ao BRLMM. Assim, o CRLMM foi utilizado para a determinação dos genótipos em todas as amostras estudadas.
- As populações CEU, CHB, JPT e STV apresentaram o maior número de blocos de haplótipos para todos os cromossomos, enquanto que as demais populações brasileiras apresentaram o menor número de blocos, inferior ao das populações africanas.
- Muitos blocos de haplótipos foram compartilhados entre as populações estudadas, sendo que as duas populações asiáticas compartilharam o maior número de blocos, refletindo a grande similaridade entre estas populações.
- As populações africanas apresentaram menor quantidade de blocos comuns, refletindo os muitos eventos de recombinação ocorridos nestas populações ao longo das gerações.
- O quilombo Sítio Velho compartilhou um número semelhante de blocos com CEU, CHB, JPT e as populações africanas, refletindo a maior mistura étnica ocorrida neste quilombo.
- Os quilombos GAU e MIB apresentaram uma maior contribuição africana, segundo os números de blocos observados nestas populações e o maior número de blocos compartilhados com as populações africanas.
- Muitos blocos, que anteriormente estavam presentes em algumas populações, foram reduzidos ao longo do tempo devido aos eventos de recombinação.

- As populações que apresentaram os blocos mais extensos foram CEU e CHB. As populações africanas e as brasileiras, por outro lado, apresentaram os menores fragmentos de blocos.
- Considerando-se as frequências haplotípicas, o quilombo STV aparentou possuir uma maior contribuição europeia, enquanto que GAU e MIB parecem ter uma maior contribuição africana.
- De forma geral, os padrões de desequilíbrio de ligação e blocos de haplótipos foram bastante similares entre os cromossomos 6 e 22.
- Na região do MHC, o padrão de blocos de haplótipos foi bem similar entre as populações, refletindo a ação da seleção natural nesta região genômica.

## REFERÊNCIAS BIBLIOGRÁFICAS

Abecasis GR, Noguchi E, Heinzmann A, Traherne JA, Bhattacharyya S, Leaves NI, Anderson GG, Zhang Y, Lench NJ, Carey A, Cardon LR, Moffatt MF, Cookson WOC (2001) Extent and distribution of linkage disequilibrium in three genomic regions. *Am J Hum Genet* 68(1):191-197.

Affymetrix (2006) BRLMM: an improved genotype calling method for the GeneChip Human Mapping 500K Array Set. Technical Report, White Paper. Santa Clara, CA: Affymetrix, Inc.

Affymetrix

([http://www.affymetrix.com/Auth/analysis/downloads/na31/genotyping/Mapping250K\\_Nsp.na31.annot.csv.zip](http://www.affymetrix.com/Auth/analysis/downloads/na31/genotyping/Mapping250K_Nsp.na31.annot.csv.zip)).

Affymetrix                      HapMap                      500K                      data  
([http://www.hapmap.org/downloads/raw\\_data/affy500k/](http://www.hapmap.org/downloads/raw_data/affy500k/))

Ahmad T, Neville M, Marshall SE, Armuzzi A, Mulcahy-Hawes K, Crawshaw J, Sato H, Ling KL, Barnardo M, Goldthorpe S, Walton R, Bunce M, Jewell DP, Welsh KI (2003) Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum Mol Genet* 12:647-656.

Amorim CE, Wang S, Marrero AR, Salzano FM, Ruiz-Linares A, Bortolini MC (2011) X-chromosomal genetic diversity and linkage disequilibrium patterns in Amerindians and non-Amerindian populations. *Am J Hum Biol* 23:299-304.

Angius A, Hyland FC, Persico I, Pirastu N, Woodage T, Pirastu M, De la Vega FM (2008) Patterns of linkage disequilibrium between SNPs in a Sardinian population isolate and the selection of markers for association studies. *Hum Hered* 65:9-22.

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299-309.

Arpini-Sampaio, Z (1998) Estrutura Genética de Três Isolados Afro-Brasileiros. Tese de Doutorado. 156 p. Departamento de Genética da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil.

Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, van Duijn CM (2004) Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* 12:527-534.

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263-265.

Bataillon T, Mailund T, Thorlacius S, Steingrimsdottir E, Rafnar T, Halldorsson MM, Calian V, Schierup MH (2006) The effective size of the Icelandic population and the prospects for LD mapping: inference from unphased microsatellite markers. *Eur J Hum Genet* 14:1044-1053.

Bioconductor (<http://www.bioconductor.org/>)

Birdseed

([http://www.affymetrix.com/products/software/specific/birdseed\\_algorithm.affx](http://www.affymetrix.com/products/software/specific/birdseed_algorithm.affx))

Bonnen PE, Pe'er I, Plenge RM, Salit J, Lowe JK, Shapero MH, Lifton RP, Breslow JL, Daly MJ, Reich DE, Jones KW, Stoffel M, Altshuler D, Friedman JM (2006) Evaluating potential for whole-genome studies in Kosrae, an isolated population in Micronesia. *Nat Genet* 38:214-217.

Carvalho B, Bengtsson H, Speed TP, and Irizarry RA (2007) Exploration, normalization, and genotype calls of high-density oligonucleotide SNP array data. *Biostatistics* 8(2):485-99.

Collins-Schramm HE, Chima B, Operario DJ, Criswell LA, Seldin MF (2003) Markers informative for ancestry demonstrate consistent megabase-length linkage disequilibrium in the African American population. *Hum Genet* 113:211-219.

Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 38:1251-1260.

de Andrade M, Atkinson EJ, Bamlet WR, Matsumoto ME, Maharjan S, Slager SL, Vachon CM, Cunningham JM, Kardia SLR (2011) Evaluating the influence of quality control decisions and software algorithms on SNP calling for the affymetrix 6.0 SNP array platform. *Hum Hered* 71(4):221-33.

De La Vega FM, Isaac H, Collins A, Scafe CR, Halldorsson BV, Su X, Lippert RA, Wang Y, Laig-Webster M, Koehler RT, Ziegler JS, Wogan LT, Stevens JF, Leinen KM, Olson SJ, Guegler KJ, You X, Xu LH, Hemken HG, Kalush F, Itakura M, Zheng Y, Thé G, O'Brien SJ, Clark AG, Istrail S, Hunkapiller MW, Spier EG, Gilbert DA (2005) The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. *Genome Res* 15:454-462.

Di X, Matsuzaki H, Webster TA, Hubbell E, Liu G, Dong S, Bartell D, Huang J, Chiles R, Yang G, Shen M, Kulp D, Kennedy GC, Mei R, Jones KW, Cawley S (2005). Dynamic model based algorithms for screening and genotyping over 100K SNPs on oligonucleotide microarrays. *Bioinformatics* 21, 1958–1963.

Dunning a M, Durocher F, Healey CS, Teare MD, McBride SE, Carlomagno F, Xu CF, Dawson E, Rhodes S, Ueda S, Lai E, Luben RN, Van Rensburg EJ, Mannermaa A, Kataja V, Rennart G, Dunham I, Purvis I, Easton D, Ponder BAJ (2000) The extent of linkage disequilibrium in four populations with distinct demographic histories. *Am J Hum Genet* 67(6):1544–54.

Evans DM, Cardon LR (2005) A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. *Am J Hum Genet* 76:681–687.

Frisse L, Hudson RR, Bartoszewicz a, Wall JD, Donfack J, and Di Rienzo A (2001) Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *American journal of human genetics* 69(4):831–43.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.

Gonzalez-Neira A, Calafell F, Navarro A, Lao O, Cann H, Comas D, Bertranpetit J (2004) Geographic stratification of linkage disequilibrium: a worldwide population study in a region of chromosome 22. *Hum Genomics* 1:399–409.

Gu S, Pakstis AJ, Li H, Speed WC, Kidd JR, Kidd KK (2007) Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. *Eur J Hum Genet* 15:302–312.

Hall D, Wijsman EM, Roos JL, Gogos JA, Karayiorgou M (2002) Extended intermarker linkage disequilibrium in the Afrikaners. *Genome Res* 12:956–961.



Hanchard N, Diakite M, Koch O, Keating B, Pinder M, Jallow M, Sisay-Joof F, Nijnik A, Wilson J, Udalova I, Kwiatkowski D, Rockett K (2006) Implications of inter-population linkage disequilibrium patterns on the approach to a disease association study in the human MHC class III. *Immunogenetics* 58:465-470.

HapMap Project  
(<http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/?N=D>)

Haploview (<http://www.broad.mit.edu/mpg/haploview>)

Hartl DL, Clark AG (1997) *Principles of population genetics*. Sinauer Associates Inc., Sunderland.

Horton R, Wilming L, Rand V, Lovering RC, Bruford E a, Khodiyar VK, Lush MJ, Povey S, Talbot Jr CC, Wright MW, Wain HM, Trowsdale J, Ziegler A, Beck S (2004) Gene map of the extended human MHC. *Nat Rev Genet* 5(12):889-99.

Jorde LB (2000) Linkage Disequilibrium and the Search for Complex Disease Genes. *Genome Res* 10(10):1435-1444.

Jorgensen TH, Degn B, Wang AG, Vang M, Gurling H, Kalsi G, McQuillin A, Kruse TA, Mors O, Ewald H (2002) Linkage disequilibrium and demographic history of the isolated population of the Faroe Islands. *Eur J Hum Genet* 10:381-387.

Kauppi L, Sajantila A, Jeffreys AJ (2003) Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum Mol Genet* 12:33-40.

Ke X, Hunt S, Tapper W, Lawrence R, Stavrides G, Ghori J, Whittaker P, Collins A, Morris AP, Bentley D, Cardon LR, Deloukas P (2004) The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum Mol Genet* 13:577-588.

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 22:139-144.

Laan M, Paabo S (1997) Demographic history and linkage disequilibrium in human populations. *Nat Genet* 17:435-438.

Leite FP, Santos SE, Rodriguez EM, Callegari-Jacques SM, Demarchi DA, Tsuneto LT, Petzl-Erler ML, Salzano FM, Hutz MH (2009) Linkage disequilibrium patterns and genetic structure of Amerindian and non-Amerindian Brazilian populations revealed by long-range X-STR markers. *Am J Phys Anthropol* 139:404-412.

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations; Heterotic models. *Genetics* 49: 49-67.

Lin S, Carvalho B, Cutler DJ, Arking DE, Chakravarti A, Irizarry RA (2008) Validation and extension of an empirical Bayes method for SNP calling on Affymetrix microarrays. *Genome Biology* 9:R63.

Matsuzaki H, Loi H, Dong S, Tsai Y-Y, Fang J, Law J, Di X, Liu W-M, Yang G, Liu G, Huang J, Kennedy GC, Ryder TB, Marcus GA, Walsh PS, Shriver MD, Puck JM, Jones KW, Mei R (2004) Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res* 14(3):414–25.1.

Meyer D, Single RM, Mack SJ, Erlich HA, Thomson G (2006) Signatures of demographic history and natural selection in the human major histocompatibility complex loci. *Genet* 173:2121-2142.

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 (2005) A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet* 76(4):634–46.

Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD (2001) Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am J Hum Genet* 68:198-207.

Pistis G, Piras I, Pirastu N, Persico I, Sassu A, Picciau A, Prodi D, Fraumene C, Mocchi E, Manias MT, Atzeni R, Cosso M, Pirastu M, Angius A (2009) High differentiation among eight villages in a secluded area of Sardinia revealed by genome-wide high density SNPs analysis. *PLoS One* 4:e4654.

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 69(1):1–14.

Qin ZS, Niu T, Liu JS (2002) Partition-Ligation-Expectation-Maximization Algorithm for Haplotype Inference with Single-Nucleotide Polymorphisms. *Am J Hum Genet* 71:1242–1247.

Rabbee N, Speed TP (2006) A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 22: 7–12.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. *Nat* 411:199–204.

Rybicki BA, Iyengar SK, Harris T, Liptak R, Elston RC, Sheffer R, Chen KM, Major M, Maliarik MJ, Iannuzzi MC (2002) The distribution of long range admixture linkage disequilibrium in an African-American population. *Hum Hered* 53:187-196.

Sawyer SL, Mukherjee N, Pakstis AJ, Feuk L, Kidd JR, Brookes AJ, Kidd KK (2005) Linkage disequilibrium patterns vary substantially among populations. *Eur J Hum Genet* 13:677-686.

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorius H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA, Heutink P, Aulchenko Y, Oostra B, van Duijn C, Jarvelin M-R, Varilo T, Peddle L, Rahman P, Piras G, Monne M, Murray S, Galver L, Peltonen L, Sabatti C, Collins A, Freimer N (2006) Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* 38:556-560.

Service S, Sabatti C, Freimer N (2007) Tag SNPs chosen from HapMap perform well in several population isolates. *Genet Epidemiol* 31:189-194.

Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A (2003) Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet* 12:771-776.

Slatkin M (1994) Linkage disequilibrium in growing and stable populations. *Genetics* 137:331-336.

Stenzel A, Lu T, Koch WA, Hampe J, Guenther SM, De La Vega FM, Krawczak M, Schreiber S (2004) Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. *Hum Genet* 114:377-385.

Stumpf MP, Goldstein DB (2003) Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium. *Curr Biol* 13:1-8.

The R Project for Statistical Computing (<http://www.r-project.org/>)

Teare MD, Dunning AM, Durocher F, Rennart G, Easton DF (2002) Sampling distribution of summary linkage disequilibrium measures. *Ann Hum Genet* 66:223-233.

Tenesa A, Wright AF, Knott SA, Carothers AD, Hayward C, Angius A, Persico I, Maestrale G, Hastie ND, Pirastu M, Visscher PM (2004) Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. *Hum Mol Genet* 13:25-33.

Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM (2007) Recent human effective population size estimated from linkage disequilibrium. *Genome Res* 17:520-526.

Varilo T, Paunio T, Parker A, Perola M, Meyer J, Terwilliger JD, Peltonen L (2003) The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum Mol Genet* 12:51-59.

Visscher PM (2007) Variation of estimates of SNP and haplotype diversity and linkage disequilibrium in samples from the same population due to experimental and evolutionary sample size. *Ann Hum Genet* 71:119-126.

Walsh EC, Mather K a, 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 (2003) An integrated haplotype map of the human major histocompatibility complex. *Am J Hum Genet* 73(3):580-90.

Wang S, Bedoya G, Labuda D, and Ruiz-Linares A (2010) Brief communication: patterns of linkage disequilibrium and haplotype diversity at Xq13 in six Native American populations. *American journal of physical anthropology* 142(3):476-80.

Weiss KM, Clark AG (2002) Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 18:19-24.

Wiesel CEV (2003) Diversidade genética e composição étnica em remanescentes de quilombos do estado do Piauí. 110 p. Tese de Doutorado. Departamento de Genética da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil.

Wiesel CEV, Luizon MR, Sousa SMB, Santos LMW, Muniz YCN, Mendes-Junior CT, Simões AL (2013) Y-Linked microsatellites in Amazonian Amerindians applied to ancestry estimates in Brazilian Afro-derived populations. *Am J Hum Biol* 25(3):313-317.

Xu S, Huang W, Wang H, He Y, Wang Y, Qian J, Xiong M, Jin L (2007) Dissecting linkage disequilibrium in African-American genomes: roles of markers and individuals. *Mol Biol Evol* 24:2049-2058.

Yu HX, Chia J-M, Bourque G, Wong MV, Chan SH, Ren EC (2005) A population-based LD map of the human chromosome 6p. *Immunogenetics* 57(8):559-65.

## **MANUSCRITO**

Requisito regimental da Pós-Graduação da Faculdade de Medicina de Ribeirão Preto – USP determina que seja encaminhado juntamente com a Tese ao menos um manuscrito derivado dos dados do trabalho. Assim, é apresentada a seguir a versão de um manuscrito que atende a exigência citada.

## **DESEQUILÍBRIO DE LIGAÇÃO E BLOCOS DE HAPLÓTIPOS DETERMINADOS PELA ANÁLISE DE 250K SNPS EM TRÊS REMANESCENTES DE QUILOMBOS**

Edilene Santos de Andrade<sup>1</sup>, Diego Martinez Salvanha<sup>2</sup>, Ricardo Zorzetto Nicolliello Vêncio<sup>2</sup>, Leandro Maza Garrido<sup>3</sup>, Henrique Krieger<sup>3</sup>, Aguinaldo Luiz Simões<sup>1</sup> e Celso Teixeira Mendes-Junior<sup>\*4</sup>.

<sup>1</sup> Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto - SP, Brasil.

<sup>2</sup> Departamento de Matemática e Computação, Faculdade de Filosofia, Ciências e Letras, Universidade de São Paulo, 14040-901, Ribeirão Preto - SP, Brasil.

<sup>3</sup> Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, 05508-000, São Paulo – SP, Brasil.

<sup>4</sup> Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto - SP, Brasil.

\* Corresponding author. Tel.: +55-16-3602-0417.

*E-mail address:* [ctmendes@ffclrp.usp.br](mailto:ctmendes@ffclrp.usp.br) (C. T. Mendes-Junior)

## RESUMO

A associação não aleatória entre alelos de diferentes loci caracteriza o que é chamado de desequilíbrio de ligação (DL) entre eles. Os padrões de DL nas populações humanas são influenciados pela taxa de recombinação e pelos eventos demográficos. Assim, o conhecimento dos padrões de DL auxilia na descrição dos eventos evolutivos sofridos pelas populações. A partir da análise de 250K SNPs (*Single Nucleotide Polymorphisms*), foram determinados os padrões de DL e blocos de haplótipos para três populações brasileiras remanescentes de quilombos e para uma população urbana. Os achados foram correlacionados com a história demográfica de cada população.

**Palavras-Chave:** desequilíbrio de ligação; blocos de haplótipos; SNPs; remanescentes de quilombos; história demográfica.

## INTRODUÇÃO

A ligação pode ser definida como o fenômeno no qual alelos adjacentes situados em um mesmo segmento cromossômico são transmitidos juntos (em bloco) na meiose com uma frequência maior do que seria esperado pelo acaso. Esse conjunto de alelos ligados é denominado haplótipo. O desequilíbrio de ligação (DL) é definido como a diferença entre as proporções haplotípicas observadas e aquelas esperadas se os alelos segregassem independentemente (Weiss e Clark 2002).

O coeficiente de desequilíbrio (D) ou desvio é uma medida do grau de associação não aleatória entre os alelos de dois loci. O DL decairá, tendendo a zero, se nenhum outro processo além da recombinação estiver agindo na população. Quando dois loci estão distantes entre si, a taxa de recombinação entre eles será mais alta em comparação à taxa dos loci muito próximos entre si. Consequentemente, é esperado que populações jovens exibam uma maior extensão de DL (maior distância observada entre dois loci em desequilíbrio) do que as populações fundadas há mais tempo, uma vez que, para loci separados por longas distâncias, o número de gerações necessárias para atingir o equilíbrio (DL desfeito) é muito menor em relação aos loci separados por curtas distâncias.

Além da taxa de recombinação, a magnitude e o decaimento do DL podem ser alterados por diversos eventos como miscigenação (Laan e Pääbo, 1997; Pfaff *et al.*, 2001; Hall *et al.*, 2002; Service *et al.*, 2006; Xu *et al.*, 2007), deriva genética (Jorde, 2000; Jorgensen *et al.*, 2002), efeito fundador (Service *et al.*, 2006), gargalos populacionais (Tenesa *et al.*, 2007), mutação (Tenesa *et al.*, 2004) e seleção natural (Ahmad *et al.*, 2003; Walsh *et al.*, 2003; Miretti *et al.*, 2005; Meyer *et al.*, 2006). Outros fatores, tais como idade da população, tamanho e taxa de crescimento populacional, migração e endocruzamento também podem modificar os padrões de DL (Slatkin, 1994; Kruglyak, 1999).

Populações submetidas a diferentes eventos histórico-demográficos diferem quanto à intensidade, à extensão e à velocidade do decréscimo do DL. Assim, analisando-se os padrões de DL em diferentes populações, pode-se realizar inferências sobre suas respectivas histórias demográficas (Jorgensen *et al.*, 2002; De La Vega *et al.*, 2005; Service *et al.*, 2006; Angius *et al.*, 2008; Pistis *et al.*, 2009).

Estudos dos padrões de DL em populações brasileiras demonstraram que as populações isoladas como as ameríndias apresentam maiores níveis e extensão do DL, quando comparadas às populações urbanas (Leite *et al.*, 2009; Amorim *et al.*, 2011). Resultados semelhantes foram obtidos em populações indígenas oriundas da Colômbia e do Canadá (Wang *et al.*, 2010).

Uma vez que as populações brasileiras remanescentes de quilombos experimentaram alguns dos eventos evolutivos que geram ou retardam o decréscimo do DL, como fundação por poucos indivíduos, miscigenação no momento da fundação e posterior isolamento, buscou-se, por meio da análise de 250K SNPs, descrever os padrões de desequilíbrio de ligação dessas populações. A título de comparação, os padrões de DL também foram inferidos na população urbana de Teresina, capital do Estado do Piauí e em populações que experimentaram histórias demográficas diversas, analisadas pelo projeto HapMap.

## **METODOLOGIA**

### **Aspectos éticos da pesquisa**

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (Processo HCRP nº 4788/2012).



## **Amostras populacionais**

Foram analisadas 58 amostras provenientes de três comunidades afro-derivadas (remanescentes de quilombos), situadas no Estado do Piauí: Gaucinha (GAU,  $n=14$ ), Mimbó (MIB,  $n=15$ ) e Sítio Velho (STV,  $n=15$ ) juntamente com uma amostra da população urbana de Teresina (TES,  $n=14$ ), a capital do Estado.

Buscando homogeneizar o tamanho amostral, foram selecionadas aleatoriamente 15 amostras de cada uma de sete populações estudadas pelo projeto HapMap: CEU (residentes dos Estados Unidos com ancestralidade do norte e oeste da Europa), CHB (chineses Han da cidade de Pequim), JPT (japoneses da cidade de Tóquio), YRI (grupo étnico Yoruba em Ibadan, Nigéria), LWK (grupo étnico Luhya em Webuye, Quênia), MKK (grupo étnico Maasai em Kinyawa, Quênia) e ASW (residentes da região sudoeste dos Estados Unidos que se auto-declararam afro-americanos). Estas amostras foram usadas como referência para as comparações dos padrões de desequilíbrio de ligação entre as populações, por serem populações bem conhecidas e caracterizadas geneticamente.

Os genótipos gerados com o uso do *Human Mapping 250K Nsp Array* para as populações do HapMap foram obtidos na página do projeto (<http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/?N=D>). Assim, os procedimentos laboratoriais foram realizados utilizando-se apenas as amostras provenientes das quatro populações brasileiras (GAU, MIB, STV e TES).

## **Extração do DNA**

O DNA genômico foi extraído a partir de amostras de 100 $\mu$ L de sangue total utilizando-se o kit QIAamp DNA Investigator Kit (QIAGEN®, Germantown, MD, USA), seguindo-se as instruções do fabricante. A determinação da concentração e pureza das amostras (relação

A260/A280) foi realizada utilizando-se o espectrofotômetro NanoDrop® (Nano Drop 2000c, Thermo Cientific, Wilmington, DE, USA). Ao final, as amostras de DNA foram armazenadas a -20°C até sua utilização.

### **Genotipagem de SNPs com microarranjos de DNA**

A partir de 250ng de DNA, os SNPs foram genotipados utilizando-se o *GeneChip® Human Mapping 250K Nsp I Array - Affymetrix®*, seguindo-se estritamente as recomendações contidas no manual *GeneChip® 500K Assay Manual* (Affymetrix, Santa Clara, CA, USA). A determinação dos genótipos foi realizada com o uso do algoritmo CRLMM (*Corrected Robust Linear Model with Maximum Likelihood Classification*) (Carvalho *et al.*, 2007; Lin *et al.*, 2008).

### **Análises de desequilíbrio de ligação e blocos de haplótipos**

Inicialmente, foi realizada uma seleção para obtenção dos SNPs comuns a todas as 11 populações analisadas (as quatro brasileiras e as sete do HapMap), para que a comparação dos padrões de desequilíbrio de ligação e blocos de haplótipos fosse possível.

A partir dos genótipos determinados pelo CRLMM com probabilidade  $\geq 0,90$  (incluindo os SNPs cujos genótipos não puderam ser determinados para todas as amostras), foram realizadas as análises de desequilíbrio de ligação e blocos de haplótipos, utilizando-se o software Haploview (Barrett *et al.*, 2005; <http://www.broad.mit.edu/mpg/haploview>).

Os haplótipos foram inferidos pelo Haploview por meio do algoritmo EM, utilizando o método de partição/ligação descrito em Qin e colaboradores 2002. O Desequilíbrio de Ligação (DL) entre cada par de SNPs foi calculado a partir de três medidas: o *LOD (log of the likelihood odds ratio) score*, o coeficiente de desvio padronizado ( $D'$ ) de Lewontin (1964) e a medida  $r^2$ . Os blocos de haplótipos foram definidos separadamente para cada população segundo Gabriel e colaboradores (2002).

## RESULTADOS

Após a seleção para SNPs comuns, muitos SNPs foram excluídos das análises de desequilíbrio de ligação e blocos de haplótipos. A redução do número total de SNPs (considerando-se todos os cromossomos) variou de 32,16% (cromossomo 6) a 38,05% (cromossomo X), com redução média de 34,69% (Tabela 1).

O número de blocos de haplótipos gerados para as 11 populações e 23 cromossomos analisados está apresentado na Tabela 2. Um padrão semelhante de blocos de haplótipos entre as diferentes populações foi observado para todos os cromossomos. As populações CEU, CHB e JPT apresentaram quantidades bastante superiores de blocos. Por outro lado, em geral, as populações brasileiras GAU, MIB e TES apresentaram um número bem inferior de blocos (Tabela 2). As três populações africanas (LWK, MKK e YRI) e a população afro-americana ASW apresentaram os valores intermediários de número de blocos. A única exceção ao padrão geral foi observada no cromossomo 15, em que a população de quilombo GAU apresentou mais blocos que MIB, TES, ASW e as três populações africanas (Tabela 2). A população remanescente de quilombo STV (Sítio Velho) apresentou, para todos os cromossomos, um número de blocos apenas inferior ao das populações CEU, CHB e JPT (Tabela 2). Isso se reflete no número total de blocos de haplótipos.

Foram realizadas comparações entre os cromossomos 6 e 22. O cromossomo 22 foi escolhido por ser o menor e, conseqüentemente, ter números menores de blocos, o que facilita a visualização e o entendimento dos resultados. O cromossomo 6, por outro lado, foi escolhido por conter o Complexo Principal de Histocompatibilidade (MHC), uma região que pode influenciar significativamente os padrões de desequilíbrio de ligação ao seu redor em decorrência de fortes pressões seletivas (Ahmad *et al.*, 2003; Walsh *et al.*, 2003; Meyer *et al.*, 2006).

Nos cromossomos 6 e 22, o número de blocos detectados em cada população seguiu o padrão descrito para todos os cromossomos (Tabela 2). Contudo, no cromossomo 6, as populações com maior e menor número de blocos foram CEU e MIB, respectivamente (Tabela 3), enquanto que no cromossomo 22 foram CHB e TES, nesta mesma ordem (Tabela 5). Diferentemente do observado para todo o cromossomo 6, na região do MHC estendido (7.765,55 Kb), definida de acordo com Horton e colaboradores (2004), o número de blocos detectados foi muito semelhante na maioria das populações, variando de 20 a 22. Apenas TES e as populações de quilombos MIB e GAU apresentaram um número bem inferior de blocos: sete, 11 e 12, respectivamente (Tabela 4).

A população com maior extensão do DL foi JPT e a que apresentou DL menos extenso foi GAU, em ambos os cromossomos (Tabelas 3 e 5). Para o cromossomo 22, entretanto, ASW apresentou DL mais extenso que as três populações africanas e as quatro brasileiras (Tabela 5), enquanto que no cromossomo 6, entre estas mesmas populações, STV apresentou a maior extensão (Tabela 3). Na região do MHC estendido, a maior extensão foi também observada em JPT, seguida por CEU e CHB. A menor extensão foi observada em TES, enquanto que em STV a extensão foi maior do que as demais populações brasileiras, africanas e ASW (Tabela 4).

No cromossomo 6, a população com maior tamanho médio de blocos foi JPT, mas para o cromossomo 22 foi ASW. Por outro lado, a população com menor tamanho médio de blocos foi GAU no cromossomo 6 e MKK no cromossomo 22. Entre as populações brasileiras, STV apresentou um tamanho médio de blocos maior para o cromossomo 6, enquanto que para o 22, TES apresentou maior média (Tabelas 3 e 5). No MHC, os resultados foram bem similares aos descritos para todo o comprimento do cromossomo 6, variando de 9,46 Kb em GAU a 52,86 Kb em JPT (Tabela 4).

CEU apresentou o bloco de maior tamanho entre todas as populações, tanto no cromossomo 6 quanto no 22, sendo que no cromossomo 6, STV compartilhou com CEU o maior bloco (Tabelas 3 e 5). O bloco mais extenso detectado na região do MHC estendido foi observado em CHB. O quilombo STV apresentou o segundo maior bloco, enquanto que o terceiro maior bloco foi observado em JPT (Tabela 4).

Muitos blocos de haplótipos foram compartilhados entre as populações estudadas. Em geral, o padrão de compartilhamento de blocos no cromossomo 6 foi bem semelhante ao observado no cromossomo 22. As duas populações asiáticas compartilharam o maior número de blocos, enquanto que o número de blocos compartilhados entre as populações africanas foi bem inferior. Entre os quilombos, para ambos os cromossomos, MIB e STV compartilharam mais blocos entre si e TES foi a população que apresentou o menor número de blocos compartilhados com as demais (Tabelas 6 e 8).

O padrão de compartilhamento de blocos na região do MHC estendido foi diferente do padrão observado nas demais regiões do cromossomo 6. O número de blocos compartilhados entre as três populações africanas foi ligeiramente maior do que o número de blocos compartilhados pelas populações asiáticas e entre CEU e as asiáticas. Entre os quilombos, MIB e STV compartilharam três blocos, mesmo número de blocos comuns entre MIB e GAU. Outra diferença em relação a todo o comprimento do cromossomo 6 é que, no MHC, GAU foi a população que apresentou o menor número de blocos compartilhados com as demais (Tabela 7).

Em alguns casos, observou-se que o mesmo bloco, que anteriormente estava presente em algumas populações, foi reduzido ao longo do tempo devido aos eventos de recombinação. Por exemplo, foram observados 84 blocos que se sobrepõem por toda a região do MHC estendido. Em alguns casos são observadas, entre os blocos, regiões de baixo desequilíbrio de ligação e/ou de  $D' = 0$  (que indica recombinação total). Assim, pode-se deduzir que anteriormente havia um único bloco de

7.465,88 Kb nesta região (formado por 536 SNPs) e que o mesmo foi sendo desfeito ao longo das gerações por eventos de recombinação, de acordo com a história evolutiva ocorrida em cada população.

O mesmo padrão foi observado em outra região do cromossomo 6, distando 24.995,63 Kb do MHC estendido. Foram detectados 45 blocos, grandes e pequenos, intercalados por pontos de recombinação em todas as 11 populações estudadas, indicando que um único grande bloco de haplótipo de 5.428,96 Kb (formado por 155 SNPs) estava presente nestas populações há muitas gerações. No cromossomo 22, também foram observados blocos sobrepostos com pontos de recombinação entre eles. Em geral, as populações que apresentaram os blocos mais extensos foram CEU, CHB e JPT. As populações africanas e as brasileiras, por outro lado, apresentaram os menores fragmentos de blocos.

As frequências dos haplótipos compartilhados pela maioria das populações estudadas foram comparadas e os resultados são apresentados nas Tabelas 9 e 10. Em geral, para o cromossomo 22, considerando-se as frequências haplotípicas, o quilombo STV aparentou maior similaridade com a população CEU. O quilombo MIB, por sua vez, aparentou ser mais relacionados às populações africanas. GAU, contudo, não aparentou uma similaridade consistente com nenhuma das populações. No cromossomo 6, em geral, o quilombo STV aparentou maior similaridade, com as populações asiáticas, enquanto que os quilombos GAU e MIB, aparentaram ter maior similaridade com as populações africanas.

## DISCUSSÃO

As populações africanas apresentam DL menor e menos extenso quando comparadas às demais populações (Kruglyak 1999; Frisse *et al.*, 2001; Reich *et al.*, 2001; Ke *et al.*, 2004; Bonnen *et al.*, 2006), uma vez que os eventos de recombinação tem se acumulado e os blocos de haplótipos foram sendo desfeitos ao longo do tempo (De La Vega *et al.*, 2005). Em concordância com estas descrições, no presente estudo, as três populações africanas analisadas apresentaram um número de blocos bem inferior ao observado na população europeia e nas duas populações asiáticas (Tabela 2).

Quanto às populações afro-americanas, foi demonstrado que apresentam maior recombinação e menor DL do que populações de brancos (Shifman *et al.*, 2003; Ke *et al.*, 2004; Stenzel *et al.*, 2004), provavelmente devido às suas origens africanas (De La Vega *et al.*, 2005). Os menores números de blocos e as menores extensões de DL foram observados em duas populações afro-brasileiras analisadas no presente estudo, GAU e MIB, o que está de acordo com estas descrições (Tabelas 2-5).

Foi estimado, a partir de marcadores microssatélites, que a contribuição africana na população de MIB é de 52,90% (cromossomo Y) (Wiesel *et al.*, 2013), 81,76% (cromossomo X) (Wiesel, 2003) e 61,10% a 80,80% (autossômicos) (Arpini-Sampaio, 1998; Wiesel, 2003). Para GAU, a estimativa para o cromossomo Y foi de 59,80% (Wiesel *et al.*, 2013) e de 38,41% para os autossômicos (Wiesel, 2003). Assim, a maior contribuição africana nestas duas populações, explica o menor DL observado. O valor mais baixo de contribuição africana observado em GAU para os marcadores autossômicos foi atribuído ao reduzido tamanho amostral, que teria gerado inconsistências nas estimativas de mistura étnica nesta população. Contudo, foi possível concluir que, nestas comunidades, o elemento fundador masculino foi de origem, predominantemente, africana e europeia. A contribuição ameríndia

(10,37% em MIB e 17,17% em GAU, STRs autossômicos) provavelmente se restringiu a indivíduos fundadores do sexo feminino (Wiezel, 2003).

Mistura recente com caucasoides deixariam assinaturas distintas das observadas. Foi sugerido que uma miscigenação recente com caucasianos mantém um alto e extenso DL em escala genômica nos afro-americanos, sendo inclusive maior do que o DL observado em caucasianos (Rybicki *et al.*, 2002; Collins-Schramm *et al.*, 2003; Xu *et al.*, 2007).

Estes resultados, entretanto, não podem ser generalizados para populações afro-americanas nas quais a contribuição caucasiana não é predominante, como é o caso das populações GAU (44,42% para os marcadores autossômicos e 31,00% para o cromossomo Y) e MIB (8,83% a 22,37% nos autossômicos e 45,10% no cromossomo Y) (Arpini-Sampaio, 1998; Wiezel, 2003; Wiezel *et al.*, 2013). Embora estas comunidades tenham recebido, no momento de sua fundação, uma considerável contribuição europeia masculina, elas se mantiveram isoladas após sua fundação, e os casamentos posteriores se deram entre indivíduos com herança predominantemente africana. Isto pode explicar os índices maiores de contribuição europeia observados apenas no cromossomo Y. Assim, estes quilombos apresentam padrões de DL mais similares aos apresentados pelas populações africanas, das quais receberam maior contribuição genética.

A população urbana TES, por outro lado, apresenta altos níveis de ancestralidade europeia: 79,50%, 58,80% e 51,58%, estimados para os cromossomos Y (Wiezel *et al.*, 2013), X e autossômicos, respectivamente (Wiezel, 2003). Consequentemente, de acordo com o exposto, seria esperado que a mesma apresentasse um alto nível de DL e um número de blocos similar ao observado na população europeia analisada neste estudo. Contudo, esta população apresentou padrões de blocos semelhantes à GAU e MIB (Tabelas 2-5).



Uma possível explicação para esta observação é que, ao contrário das populações de quilombos, que representam isolados genéticos, TES teria sofrido sucessivos eventos de miscigenação ao longo de sua história, o que diminuiu o DL ao longo das gerações (Pfaff *et al.*, 2001). A acentuada redução do DL em TES pode também ser explicada pelo seu evento fundador, que envolveu um maior número de indivíduos e foi seguido por um rápido crescimento, uma vez que populações em expansão ou recentemente expandidas sofrem um rápido decréscimo do DL em um intervalo de tempo relativamente curto (Laan e Pääbo, 1997; Angius *et al.*, 2008; Pistis *et al.*, 2009). Um rápido crescimento aumenta a diversidade genética e a recombinação, enquanto que a deriva genética será mínima, resultando na diminuição do DL (Jorde, 2000; Pritchard e Przeworski 2001; Hall *et al.*, 2002; Aulchenko *et al.*, 2004).

Assim como os quilombos GAU e MIB, o quilombo STV apresenta um alto percentual de contribuição africana: 60,20%, 96,39% e 48,09%, para os cromossomos Y (Witzel *et al.*, 2013), X e autossômicos, respectivamente (Witzel, 2003). Contudo, diferentemente dos outros dois quilombos, STV apresentou, para todos os cromossomos, um grande número de blocos, inferior apenas ao observado nas populações CEU, CHB e JPT (Tabela 2). Tendo ancestralidade predominantemente africana, os altos níveis de DL em STV não podem ser explicados por uma maior contribuição europeia, como sugerido por alguns autores (Rybicki *et al.*, 2002; Collins-Schramm *et al.*, 2003; Xu *et al.*, 2007). Assim, a explicação para o maior DL observado em STV, em relação aos demais quilombos, consiste na história demográfica peculiar desta população.

A miscigenação no momento da fundação de uma população gera alto e extenso DL (Laan e Pääbo, 1997; Pfaff *et al.*, 2001; Hall *et al.*, 2002; Service *et al.*, 2006; Xu *et al.*, 2007). O quilombo STV experimentou um maior grau de miscigenação no momento da sua fundação, quando comparado a GAU e MIB. Embora também fundada por poucos indivíduos, o número de fundadores em STV foi maior do que nos

demais quilombos, além de serem oriundos de áreas mais amplas, com uma população mais diversificada (Arpini-Sampaio, 1998).

Além disso, a partir da análise de STRs autossômicos, STV apresentou um maior índice de contribuição ameríndia (27,96%), quando comparado a GAU (17,17%) e MIB (10,37%) (Wiesel, 2003), o que evidencia o maior grau de miscigenação sofrido por este quilombo no momento de sua fundação. A maior contribuição ameríndia em STV pode ter contribuído para o maior DL observado neste quilombo, uma vez que as populações americanas nativas, incluindo as brasileiras (Leite *et al.*, 2009), são reconhecidas por apresentar altos níveis de DL e grandes blocos de haplótipos (Sawyer *et al.*, 2005; Conrad *et al.*, 2006; Gu *et al.*, 2007; Wang *et al.*, 2010; Amorim *et al.*, 2011).

Outra característica observada em STV que gera um alto DL é a estabilidade da população resultante do isolamento. Populações que experimentaram um longo período de tamanho constante no momento de sua fundação apresentam DL aumentado. Um crescimento lento no início (nas primeiras gerações) aumenta a perda de diversidade genética durante o início da história de uma população isolada, resultando em um gargalo mais estreito, aumentando a extensão do DL nestas populações (Kruglyak, 1999; Jorgensen *et al.*, 2002). Este cenário foi mais evidente em STV do que nos outros dois quilombos estudados. Quando comparada a GAU e MIB, esta comunidade foi a que menos se diferenciou ao longo do tempo (Arpini-Sampaio, 1998; Wiesel, 2003).

Diferentemente do observado para todo o cromossomo 6, bem como para o cromossomo 22, na região do MHC estendido o número de blocos detectados foi muito semelhante entre a maioria das populações estudadas (Tabela 4). Uma similaridade nos padrões de DL nesta região foi também descrita para as populações africana e caucasiana (Hanchard *et al.*, 2006). Isto se deve ao fato desta região estar submetida a fortes pressões seletivas, fazendo com que os padrões de DL sejam conservados, mesmo entre as diversas populações globais

(Ardlie *et al.*, 2002). A seleção epistática para combinações de alelos em dois ou mais locos pode influenciar os padrões de DL através dos haplótipos (Ahmad *et al.*, 2003). Meyer e colaboradores (2006) descreveram altos níveis de DL na região do MHC, entre os locos HLA separados por grandes distâncias físicas, e sugeriram que o extenso DL era resultante da seleção natural e dos eventos demográficos. Além da ação da seleção natural, o DL mais extenso no MHC parece ser devido às taxas de recombinação reduzidas nesta região (Walsh *et al.*, 2003).

A existência de haplótipos compartilhados por várias populações sugere que muito da estrutura dos haplótipos atuais é antiga e foi estabelecida antes da diversificação recente da humanidade. Em geral, os padrões dos blocos de haplótipos são muito similares entre as populações do leste da Ásia (que apresentam grande similaridade entre elas) e as populações da região ocidental da Eurásia (incluindo a Europa). Assim, pode-se considerar que as populações localizadas nas duas regiões compartilham uma única estrutura de DL (González-Neira *et al.*, 2004). Os padrões de blocos para as populações analisadas neste estudo estão de acordo com o descrito: as duas populações asiáticas compartilharam um grande número de blocos, mostrando grande similaridade, e ambas compartilharam muitos haplótipos com a população europeia CEU (Tabelas 6 e 8). Por outro lado, neste estudo, as três populações africanas não apresentaram muitos blocos comuns, em relação ao número total de blocos detectados (Tabelas 6 e 8). Isto se deve a maior diversidade haplotípica observada entre as populações (González-Neira *et al.*, 2004).

Entre os quilombos, MIB e STV apresentaram mais blocos comuns (Tabelas 6 e 8). Isto pode ser explicado pela menor distância genética entre estas populações quando comparadas com GAU (Arpini-Sampaio, 1998; Wiezel, 2003), a despeito do maior DL observado em STV. Os três quilombos compartilharam poucos blocos com TES (Tabelas 6 e 8), devido ao fato de TES apresentar maior contribuição europeia e os três quilombos apresentarem contribuição predominantemente africana.

Embora apresentem contribuição majoritariamente africana, os três quilombos não compartilharam um grande número de blocos com as populações africanas analisadas, como seria esperado (Tabelas 6 e 8). Isto pode ser explicado pelo fato de que os haplótipos observados nas populações não africanas representam apenas um subconjunto daqueles observados na África (Gu *et al.*, 2007). De La Vega e colaboradores (2005) observaram uma significativa proporção de haplótipos não compartilhados entre afro-americanos e africanos, embora os mapas de DL tenham apresentado formas similares.

Poucos haplótipos foram compartilhados pelas populações estudadas na região do MHC (Tabela 7). Nesta região, existe pouca evidência dentro dos blocos de DL para haplótipos compartilhados por várias populações. Isto sugere que esta região tem sido submetida a uma alta taxa de *turnover* de haplótipos e que as estruturas dos haplótipos atuais não foram estabelecidas no início da história humana (Kauppi *et al.*, 2003). Conseqüentemente, são observados nesta região, determinados blocos de DL que em algumas populações apresentam-se incompletos, mas que ainda incluem associações altamente significativas em toda a sua extensão (Kauppi *et al.*, 2003), como observado no presente estudo.

Considerando-se as frequências haplotípicas (Tabelas 9 e 10), é difícil relacioná-las diretamente ao DL. Em geral, os mesmos poucos haplótipos são os mais frequentes em quase todas as populações, mas suas frequências diferem consideravelmente, mesmo entre as populações não africanas (Sawyer *et al.*, 2005).

No presente estudo, os quilombos GAU e MIB aparentaram maior similaridade com as populações africanas, quanto às frequências haplotípicas (Tabelas 9 e 10). Isto está de acordo com a maior contribuição africana na formação destas comunidades. O quilombo STV, por outro lado, aparentou maior similaridade com a população CEU (cromossomo 22) e com as populações asiáticas (cromossomo 6). Uma provável explicação seria a maior miscigenação ocorrida nesta

população e a maior contribuição ameríndia, quando comparada aos demais quilombos.

A distribuição do DL parece não ser uniforme dentro e entre os cromossomos (Stumpf e Goldstein 2003; De La Vega *et al.*, 2005). A variabilidade no DL observada entre as diferentes regiões genômicas pode ser explicada pela presença ou ausência de *hotspots* de recombinação em cada região (Shifman *et al.*, 2003). Foi descrito que o cromossomo 6 apresenta uma taxa de decréscimo do DL 26-58% mais lenta do que os cromossomos 21 e 22. A extensão total do DL foi igual para estes cromossomos e inferior a do cromossomo 6. Isto foi atribuído ao maior tamanho do cromossomo 6 e ao fato deste apresentar taxas de recombinação menores e, conseqüentemente, diminuição mais lenta do DL, refletindo a ação da seleção natural (De La Vega *et al.*, 2005). No presente estudo, também foi observada uma extensão muito maior no cromossomo 6 em relação ao 22 (Tabelas 3 e 5). Além disso, o cromossomo 22 inteiro (~51 Mb), mesmo contando com cerca de 1.500 SNPs considerados na análise, possui extensão total de DL menor do que a da região do MHC (~7,8 Mb) em 8 das 11 populações consideradas.

Em resumo, considerando-se o número de blocos de haplótipos detectados em cada população estudada, o padrão foi semelhante entre todos os cromossomos. Isto mostra que os padrões de DL são modelados principalmente pela história evolutiva de cada população, mais do que por fatores genômicos locais, como recombinação, mutação e seleção.

## **AGRADECIMENTOS**

Este trabalho foi realizado com auxílio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e da Fundação de Apoio ao Ensino, Pesquisa e Assistência do HCFMRP-USP (FAEPA).

Agradecemos a Maria do Carmo Tomitão Canas e Ana Lúcia Pimentel pelo suporte técnico oferecido no processo de extração de DNA das amostras analisadas.

## REFERÊNCIAS BIBLIOGRÁFICAS

Ahmad T, Neville M, Marshall SE, Armuzzi A, Mulcahy-Hawes K, Crawshaw J, Sato H, Ling KL, Barnardo M, Goldthorpe S, Walton R, Bunce M, Jewell DP, Welsh KI (2003) Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum Mol Genet* 12:647-656.

Amorim CE, Wang S, Marrero AR, Salzano FM, Ruiz-Linares A, Bortolini MC (2011) X-chromosomal genetic diversity and linkage disequilibrium patterns in Amerindians and non-Amerindian populations. *Am J Hum Biol* 23:299-304.

Angius A, Hyland FC, Persico I, Pirastu N, Woodage T, Pirastu M, De la Vega FM (2008) Patterns of linkage disequilibrium between SNPs in a Sardinian population isolate and the selection of markers for association studies. *Hum Hered* 65:9-22.

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299-309.

Arpini-Sampaio, Z (1998) Estrutura Genética de Três Isolados Afro-Brasileiros. Tese de Doutorado. 156 p. Departamento de Genética da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil.

Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, van Duijn CM (2004) Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* 12:527-534.

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263-265.

Bonnen PE, Pe'er I, Plenge RM, Salit J, Lowe JK, Shapero MH, Lifton RP, Breslow JL, Daly MJ, Reich DE, Jones KW, Stoffel M, Altshuler D, Friedman JM (2006) Evaluating potential for whole-genome studies in Kosrae, an isolated population in Micronesia. *Nat Genet* 38:214-217.

Carvalho B, Bengtsson H, Speed TP, and Irizarry RA (2007) Exploration, normalization, and genotype calls of high-density oligonucleotide SNP array data. *Biostatistics* 8(2):485-99.

Collins-Schramm HE, Chima B, Operario DJ, Criswell LA, Seldin MF (2003) Markers informative for ancestry demonstrate consistent megabase-length linkage disequilibrium in the African American population. *Hum Genet* 113:211-219.

Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 38:1251-1260.

De La Vega FM, Isaac H, Collins A, Scafe CR, Halldorsson BV, Su X, Lippert RA, Wang Y, Laig-Webster M, Koehler RT, Ziegler JS, Wogan LT, Stevens JF, Leinen KM, Olson SJ, Guegler KJ, You X, Xu LH, Hemken HG, Kalush F, Itakura M, Zheng Y, Thé G, O'Brien SJ, Clark AG, Istrail S, Hunkapiller MW, Spier EG, Gilbert DA (2005) The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. *Genome Res* 15:454-462.

Frisse L, Hudson RR, Bartoszewicz a, Wall JD, Donfack J, and Di Rienzo A (2001) Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *American journal of human genetics* 69(4):831-43.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225-2229.

Gonzalez-Neira A, Calafell F, Navarro A, Lao O, Cann H, Comas D, Bertranpetit J (2004) Geographic stratification of linkage disequilibrium: a worldwide population study in a region of chromosome 22. *Hum Genomics* 1:399-409.

Gu S, Pakstis AJ, Li H, Speed WC, Kidd JR, Kidd KK (2007) Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. *Eur J Hum Genet* 15:302-312.

Hall D, Wijsman EM, Roos JL, Gogos JA, Karayiorgou M (2002) Extended intermarker linkage disequilibrium in the Afrikaners. *Genome Res* 12:956-961.

Hanchard N, Diakite M, Koch O, Keating B, Pinder M, Jallow M, Sisay-Joof F, Nijnik A, Wilson J, Udalova I, Kwiatkowski D, Rockett K (2006) Implications of inter-population linkage disequilibrium patterns on the approach to a disease association study in the human MHC class III. *Immunogenetics* 58:465-470.

HapMap Project  
(<http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/?N=D>)

Haploview (<http://www.broad.mit.edu/mpg/haploview>)



Horton R, Wilming L, Rand V, Lovering RC, Bruford E a, Khodiyar VK, Lush MJ, Povey S, Talbot Jr CC, Wright MW, Wain HM, Trowsdale J, Ziegler A, Beck S (2004) Gene map of the extended human MHC. *Nat Rev Genet* 5(12):889–99.

Jorde LB (2000) Linkage Disequilibrium and the Search for Complex Disease Genes. *Genome Res* 10(10):1435–1444.

Jorgensen TH, Degn B, Wang AG, Vang M, Gurling H, Kalsi G, McQuillin A, Kruse TA, Mors O, Ewald H (2002) Linkage disequilibrium and demographic history of the isolated population of the Faroe Islands. *Eur J Hum Genet* 10:381–387.

Kauppi L, Sajantila A, Jeffreys AJ (2003) Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum Mol Genet* 12:33–40.

Ke X, Hunt S, Tapper W, Lawrence R, Stavrides G, Ghori J, Whittaker P, Collins A, Morris AP, Bentley D, Cardon LR, Deloukas P (2004) The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum Mol Genet* 13:577–588.

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 22:139–144.

Laan M, Paabo S (1997) Demographic history and linkage disequilibrium in human populations. *Nat Genet* 17:435–438.

Leite FP, Santos SE, Rodriguez EM, Callegari-Jacques SM, Demarchi DA, Tsuneto LT, Petzl-Erler ML, Salzano FM, Hutz MH (2009) Linkage disequilibrium patterns and genetic structure of Amerindian and non-Amerindian Brazilian populations revealed by long-range X-STR markers. *Am J Phys Anthropol* 139:404–412.

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations; Heterotic models. *Genetics* 49: 49–67.

Lin S, Carvalho B, Cutler DJ, Arking DE, Chakravarti A, Irizarry RA (2008) Validation and extension of an empirical Bayes method for SNP calling on Affymetrix microarrays. *Genome Biology* 9:R63.

Meyer D, Single RM, Mack SJ, Erlich HA, Thomson G (2006) Signatures of demographic history and natural selection in the human major histocompatibility complex loci. *Genet* 173:2121–2142.

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 (2005) A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet* 76(4):634–46.

Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD (2001) Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am J Hum Genet* 68:198-207.

Pistis G, Piras I, Pirastu N, Persico I, Sassu A, Picciau A, Prodi D, Fraumene C, Mocchi E, Manias MT, Atzeni R, Cosso M, Pirastu M, Angius A (2009) High differentiation among eight villages in a secluded area of Sardinia revealed by genome-wide high density SNPs analysis. *PLoS One* 4:e4654.

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 69(1):1–14.

Qin ZS, Niu T, Liu JS (2002) Partition-Ligation-Expectation-Maximization Algorithm for Haplotype Inference with Single-Nucleotide Polymorphisms. *Am J Hum Genet* 71:1242–1247.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. *Nat* 411:199–204.

Rybicki BA, Iyengar SK, Harris T, Liptak R, Elston RC, Sheffer R, Chen KM, Major M, Maliarik MJ, Iannuzzi MC (2002) The distribution of long range admixture linkage disequilibrium in an African-American population. *Hum Hered* 53:187-196.

Sawyer SL, Mukherjee N, Pakstis AJ, Feuk L, Kidd JR, Brookes AJ, Kidd KK (2005) Linkage disequilibrium patterns vary substantially among populations. *Eur J Hum Genet* 13:677-686.

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorius H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA, Heutink P, Aulchenko Y, Oostra B, van Duijn C, Jarvelin M-R, Varilo T, Peddle L, Rahman P, Piras G, Monne M, Murray S, Galver L, Peltonen L, Sabatti C, Collins A, Freimer N (2006) Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* 38:556-560.

Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A (2003) Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet* 12:771-776.

Slatkin M (1994) Linkage disequilibrium in growing and stable populations. *Genetics* 137:331-336.

Stenzel A, Lu T, Koch WA, Hampe J, Guenther SM, De La Vega FM, Krawczak M, Schreiber S (2004) Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. *Hum Genet* 114:377-385.

Stumpf MP, Goldstein DB (2003) Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium. *Curr Biol* 13:1-8.

Tenesa A, Wright AF, Knott SA, Carothers AD, Hayward C, Angius A, Persico I, Maestrale G, Hastie ND, Pirastu M, Visscher PM (2004) Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. *Hum Mol Genet* 13:25-33.

Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM (2007) Recent human effective population size estimated from linkage disequilibrium. *Genome Res* 17:520-526.

Walsh EC, Mather K a, 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 (2003) An integrated haplotype map of the human major histocompatibility complex. *Am J Hum Genet* 73(3):580-90.

Wang S, Bedoya G, Labuda D, and Ruiz-Linares A (2010) Brief communication: patterns of linkage disequilibrium and haplotype diversity at Xq13 in six Native American populations. *American journal of physical anthropology* 142(3):476-80.

Weiss KM, Clark AG (2002) Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 18:19-24.

Wiesel CEV (2003) Diversidade genética e composição étnica em remanescentes de quilombos do estado do Piauí. 110 p. Tese de Doutorado. Departamento de Genética da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil.

Wiesel CEV, Luizon MR, Sousa SMB, Santos LMW, Muniz YCN, Mendes-Junior CT, Simões AL (2013) Y-Linked microsatellites in Amazonian Amerindians applied to ancestry estimates in Brazilian Afro-derived populations. *Am J Hum Biol* 25(3):313-317.

Xu S, Huang W, Wang H, He Y, Wang Y, Qian J, Xiong M, Jin L (2007) Dissecting linkage disequilibrium in African-American genomes: roles of markers and individuals. *Mol Biol Evol* 24:2049-2058.

## TABELAS

**Tabela 1.** Número de SNPs contidos no *GeneChip 250 K Nsp* e comuns a todas as populações estudadas.

**Tabela 2.** Número de blocos de haplótipos gerados pelo Haploview para todas as populações analisadas.

**Tabela 3.** Panorama geral dos blocos de haplótipos no cromossomo 6.

**Tabela 4.** Panorama geral dos blocos de haplótipos na região do MHC no cromossomo 6.

**Tabela 5.** Panorama geral dos blocos de haplótipos no cromossomo 22.

**Tabela 6.** Número de blocos de haplótipos do cromossomo 6 compartilhados entre as populações.

**Tabela 7.** Número de blocos de haplótipos na região do MHC compartilhados entre as populações.

**Tabela 8.** Número de blocos de haplótipos do cromossomo 22 compartilhados entre as populações.

**Tabela 9.** Frequências populacionais dos haplótipos compartilhados do cromossomo 6.

**Tabela 10.** Frequências populacionais dos haplótipos compartilhados do cromossomo 22.

Tabela 1. Número de SNPs contidos no *GeneChip 250 K Nsp* e comuns a todas as populações estudadas.

Cromossomo	Nº de SNPs no <i>chip</i>	Nº de SNPs comuns (com <i>no call</i> )
1	19.855	12.796
2	22.204	14.603
3	18.358	12.240
4	19.053	12.254
5	17.152	11.423
6	17.104	11.603
7	13.932	9.327
8	14.834	9.701
9	11.942	7.747
10	14.263	9.188
11	13.287	8.677
12	13.046	8.448
13	11.072	7.167
14	8.175	5.348
15	7.002	4.590
16	7.010	4.626
17	4.844	3.178
18	8.145	5.282
19	2.691	1.765
20	5.838	3.840
21	3.936	2.619
22	2.499	1.572
X	5.714	3.540
Y	5	–
?	283	–
Total	262.244	167.994

Tabela 2. Número de blocos de haplótipos gerados pelo Haploview para todas as populações analisadas.

Cromossomo	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
1	489	457	474	230	201	238	268	142	181	304	187
2	534	538	552	259	244	279	266	180	228	373	230
3	477	430	432	235	256	272	276	168	205	355	223
4	483	464	458	284	265	289	299	184	219	364	215
5	458	434	437	239	231	251	254	179	175	310	217
6	446	437	444	259	223	265	268	178	172	324	178
7	358	329	332	176	156	196	194	126	138	233	134
8	382	416	386	179	203	198	181	142	155	260	148
9	267	255	261	130	120	125	127	89	80	162	111
10	330	318	327	188	182	174	204	127	137	252	141
11	334	348	350	188	181	192	191	133	155	275	155
12	325	277	295	170	160	177	159	114	119	215	137
13	261	280	270	140	141	138	128	100	102	171	117
14	188	176	202	107	102	99	109	76	81	115	85
15	134	134	148	60	62	61	63	67	39	94	59
16	160	134	146	62	62	65	74	44	50	104	55
17	107	107	97	49	49	52	58	36	41	69	39
18	201	180	185	89	85	91	95	60	62	125	69
19	48	52	44	26	21	27	27	20	15	36	17
20	147	131	146	58	57	61	61	34	49	76	53
21	110	104	114	52	45	53	56	40	46	77	50
22	42	44	36	21	16	18	19	14	17	25	13
Total	6281	6045	6136	3201	3062	3321	3377	2253	2466	4319	2633

GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 3. Panorama geral dos blocos de haplótipos no cromossomo 6.

Dados de DL	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
Nº de blocos	446	437	444	259	223	265	268	178	172	324	178
Extensão total do DL (Kb)	18987,25	17021,00	19456,80	5264,40	5353,00	5263,90	5685,19	2401,40	3195,80	9558,46	4931,13
Tamanho médio dos blocos (Kb)	42,57	38,95	43,82	21,14	24,00	19,86	21,21	13,49	18,58	29,84	27,70
Tamanho do maior bloco (Kb)	483,90	413,60	461,80	296,00	433,30	279,40	332,00	136,80	225,60	483,90	465,80
Tamanho do menor bloco (Kb)	0,01	0,04	0,04	0,10	0,10	0,10	0,09	0,10	0,10	0,06	0,04
Nº total de SNPs nos blocos	2679	2492	2687	1209	1091	1216	1290	741	780	1669	852
Nº médio de SNPs por bloco	6,01	5,70	6,05	4,88	4,89	4,59	4,81	4,16	4,53	5,15	4,79
Nº máximo de SNPs por bloco	32	28	45	18	21	19	17	14	17	23	27
Nº mínimo de SNPs por bloco	2	2	2	2	2	3	2	3	3	2	3

DL: Desequilíbrio de Ligação; GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 4. Panorama geral dos blocos de haplótipos na região do MHC no cromossomo 6.

Dados de DL	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
Nº de blocos	22	20	22	20	20	20	22	12	11	20	7
Extensão total do DL (Kb)	930,1	779,6	1162,9	600,1	511,9	299,8	537,4	113,6	341	699,3	66,7
Tamanho médio dos blocos (Kb)	42,28	38,98	52,86	30	25,56	14,99	24,43	9,46	31	34,96	9,53
Tamanho do maior bloco (Kb)	151,4	231,5	192,7	116	133,1	64,5	140,1	25,7	142,4	194,3	27
Tamanho do menor bloco (Kb)	0,6	1,1	0,1	0,2	0,1	0,2	0,6	0,2	0,6	0,2	0,2
Nº total de SNPs nos blocos	132	105	135	110	102	88	101	42	52	115	30
Nº médio de SNPs por bloco	6	5,25	6,14	5,5	5,1	4,4	4,59	3,5	4,73	5,75	4,28
Nº máximo de SNPs por bloco	12	14	16	11	12	11	11	5	8	16	8
Nº mínimo de SNPs por bloco	3	3	2	3	3	3	2	3	3	3	3

DL: Desequilíbrio de Ligação; GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.



Tabela 5. Panorama geral dos blocos de haplótipos no cromossomo 22.

Dados de DL	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
Nº de blocos	42	44	36	21	16	18	19	14	17	25	13
Extensão total do DL (Kb)	819,24	858,64	868,00	366,24	221,08	212,74	533,74	207,94	299,90	408,20	249,80
Tamanho médio dos blocos (Kb)	19,51	19,51	24,11	17,44	13,82	11,82	28,09	14,85	17,64	16,33	19,22
Tamanho do maior bloco (Kb)	157,20	100,80	130,30	113,80	87,20	58,90	114,00	114,00	114,00	114,00	74,70
Tamanho do menor bloco (Kb)	0,04	0,04	0,10	0,04	0,08	0,04	0,04	0,04	0,30	0,20	0,20
Nº total de SNPs nos blocos	185	205	180	83	60	73	88	52	66	96	56
Nº médio de SNPs por bloco	4,40	4,66	5,00	3,95	3,75	4,06	4,63	3,71	3,88	3,84	4,31
Nº máximo de SNPs por bloco	11	11	11	12	7	10	12	6	9	7	9
Nº mínimo de SNPs por bloco	2	2	3	3	2	3	3	3	3	3	3

DL: Desequilíbrio de Ligação; GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 6. Número de blocos de haplótipos do cromossomo 6 compartilhados entre as populações.

	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
CEU	–										
CHB	88	–									
JPT	84	172	–								
YRI	36	36	42	–							
LWK	48	46	51	96	–						
MKK	56	53	56	83	90	–					
ASW	47	50	49	91	79	86	–				
GAU	34	27	29	34	37	46	35	–			
MIB	28	31	30	43	47	54	51	38	–		
STV	79	76	78	73	70	91	67	49	52	–	
TES	37	30	31	42	35	41	37	31	30	51	–

GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 7. Número de blocos de haplótipos na região do MHC compartilhados entre as populações.

	POPULAÇÕES										
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
CEU	–										
CHB	7	–									
JPT	9	9	–								
YRI	1	4	3	–							
LWK	4	7	8	9	–						
MKK	5	6	7	10	11	–					
ASW	4	8	7	10	10	11	–				
GAU	0	1	1	2	3	4	4	–			
MIB	2	4	3	7	6	4	4	3	–		
STV	5	8	7	7	9	10	6	2	3	–	
TES	4	3	4	3	4	4	2	2	1	3	–

GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 8. Número de blocos de haplótipos do cromossomo 22 compartilhados entre as populações.

	POPULAÇÕES											
	CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES	
CEU	–											
CHB	11	–										
JPT	9	19	–									
YRI	2	4	4	–								
LWK	4	4	3	7	–							
MKK	6	6	3	7	8	–						
ASW	5	4	3	9	6	7	–					
GAU	3	4	2	5	5	6	7	–				
MIB	4	4	2	6	6	8	8	6	–			
STV	5	8	7	6	8	7	8	7	8	–		
TES	3	2	0	0	2	3	0	0	3	1	–	

GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 9. Frequências populacionais dos haplótipos compartilhados do cromossomo 6.

Bloco <sup>a</sup> (Kb)	Haplótipos	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
C (4,4)	TAC	0,333	0,400	0,633	0,500	0,467	0,367	0,467	0,250	0,300	0,667	–
	CCT	0,667	0,600	0,367	0,500	0,533	0,633	0,533	0,750	0,700	0,333	–
D (1,0)	ATG	0,267	–	–	0,400	0,567	0,500	0,500	0,429	0,733	0,300	0,464
	GCA	0,733	–	–	0,600	0,433	0,500	0,500	0,571	0,267	0,700	0,536
E (7,0)	GAG	–	0,567	0,500	0,321	0,400	0,433	0,500	0,429	0,333	0,600	0,357
	AGA	–	0,433	0,500	0,679	0,600	0,567	0,500	0,571	0,667	0,400	0,643
F (23,2)	GGTT	0,417	0,500	0,400	0,633	0,533	0,500	0,500	0,393	0,267	0,400	0,500
	CCCC	0,500	0,500	0,500	0,367	0,467	0,500	0,500	0,464	0,633	0,567	0,393
	GGCT	–	–	–	–	–	–	–	0,036	0,100	–	–
	GGTC	–	–	–	–	–	–	–	–	–	0,033	–
	GCCC	–	–	–	–	–	–	–	0,036	–	–	–
	GCCT	–	–	–	–	–	–	–	0,071	–	–	–
	CGTT	–	–	0,100	–	–	–	–	–	–	–	0,036
	CGTC	–	–	–	–	–	–	–	–	–	–	0,036
	CGCC	–	–	–	–	–	–	–	–	–	–	0,036
	GCTT	0,083	–	–	–	–	–	–	–	–	–	–
G (2,2)	GGCT	0,233	0,333	–	0,567	0,433	0,433	0,600	–	–	–	–
	CCTC	0,767	0,600	–	0,433	0,533	0,567	0,367	–	–	–	–
	GCTC	–	–	–	–	0,033	–	–	–	–	–	–
	CGTC	–	–	–	–	–	–	0,033	0,393	0,500	0,533	0,714
	GCCT	–	0,033	–	–	–	–	–	0,607	0,500	0,467	0,286
	CCTT	–	0,033	–	–	–	–	–	–	–	–	–

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica. GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.

Tabela 10. Frequências populacionais dos haplótipos compartilhados do cromossomo 22.

Bloco <sup>a</sup>	Haplótipos	POPULAÇÕES										
		CEU	CHB	JPT	YRI	LWK	MKK	ASW	GAU	MIB	STV	TES
B1-B4	ATAGCC	0,467	–	0,567	0,500	0,733	–	0,600	0,607	0,667	0,567	–
	GCGGGT	0,300	–	0,300	0,200	0,167	–	0,367	0,357	0,267	0,267	–
	GCGTGT	0,200	–	0,033	0,133	0,100	–	–	–	0,067	0,167	–
	ACAGCC	–	–	–	–	–	–	0,033	0,018	–	–	–
	ACGGCC	–	–	0,067	–	–	–	–	0,018	–	–	–
	ATGTG–	–	–	0,033	–	–	–	–	–	–	–	–
C1-C2	TAT	0,367	0,533	0,600	0,333	0,233	0,333	0,400	0,393	0,200	0,400	–
	GGC	0,633	0,467	0,400	0,667	0,767	0,667	0,567	0,607	0,767	0,600	–
	TGC	–	–	–	–	–	–	0,033	–	0,033	–	–
	GAC	–	–	–	–	–	–	–	–	–	–	–
D	ATC	–	–	–	0,433	0,400	0,367	0,500	0,714	0,400	0,600	–
	GCA	–	–	–	0,567	0,600	0,633	0,500	0,286	0,600	0,400	–
E	TTT	0,400	0,433	0,400	0,233	0,500	0,233	0,500	–	0,067	0,100	–
	TAT	0,100	0,033	0,100	0,100	0,033	0,067	–	0,464	0,433	0,467	–
	CAG	0,500	0,533	0,500	0,567	0,467	0,700	0,500	0,536	0,500	0,433	–
G1-G2	CCAA	0,433	–	–	0,367	0,367	0,333	0,600	0,321	0,333	0,400	–
	GACT	0,567	–	–	0,633	0,633	0,667	0,400	0,679	0,667	0,600	–

<sup>a</sup> Os blocos estão ordenados segundo a posição cromossômica. GAU: Gaucinha; MIB: Mimbó; STV: Sítio Velho; TES: Teresina; CEU: população europeia; CHB: chineses; JPT: japoneses; YRI: Yoruba; LWK: Luhya; MKK: Maasai; ASW: afro-americanos.