UNIVERSIDADE DE SÃO PAULO

FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

DEPARTAMENTO DE GENÉTICA

LEONARDO ARDUINO MARANO

Estudo da diversidade dos genes *MC1R* e
SLC24A5 em populações globais:
avaliação de aspectos evolutivos e
ambientais

Ribeirão Preto
2015

LEONARDO ARDUINO MARANO

Estudo da diversidade dos genes *MC1R* e
SLC24A5 em populações globais:
avaliação de aspectos evolutivos e
ambientais

Tese apresentada ao
Departamento de Genética da
Faculdade de Medicina de
Ribeirão Preto da Universidade
de São Paulo para obtenção do
título de Doutor em Ciências

Área de Concentração:
Genética

Orientador:
Prof. Dr. Celso Teixeira Mendes
Junior

Ribeirão Preto
2015

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

Marano, Leonardo Arduino

Estudo da diversidade dos genes *MC1R* e *SLC24A5* em populações globais: avaliação de aspectos evolutivos e ambientais/ Leonardo Arduino Marano ; orientador Celso Teixeira Mendes Junior. – Ribeirão Preto, 2015.

157 f. : il. 30cm

Tese (Doutorado – Program de Pós-graduação em Genética)- Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, 2015.

1. *MC1R*. 2. *SLC24A5*. 3. Sequenciamento de nova-geração. 4. *1000 Genomes Project* 5. Genética de populações

Ficha de Aprovação

Nome: Marano, Leonardo Arduino

Título: Estudo da diversidade dos genes MC1R e SLC24A5 em populações globais: avaliação de aspectos evolutivos e ambientais

Tese apresentada ao Departamento de Genética da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Doutor em Ciências

Área de Concentração:
Genética

Aprovado em:

Banca Examinadora

Prof. Dr. _____	Instituição: _____
Julgamento: _____	Assinatura: _____

Prof. Dr. _____	Instituição: _____
Julgamento: _____	Assinatura: _____

Prof. Dr. _____	Instituição: _____
Julgamento: _____	Assinatura: _____

Prof. Dr. _____	Instituição: _____
Julgamento: _____	Assinatura: _____

Prof. Dr. _____	Instituição: _____
Julgamento: _____	Assinatura: _____

Dedicatória

*À todos meus amigos e familiares, pelo suporte e amor infinitos em
todos os momentos*

Agradecimentos

Inicialmente agradeço grandemente ao meu orientador **Prof. Dr. Celso Teixeira Mendes Junior**, por mais essa chance, pela disponibilidade, boa vontade, paciência e ensinamentos imensos oferecidos. Esse agradecimento vem do fundo do coração como forma de tentar dizer o quanto sua ajuda foi necessária para que as dificuldades surgidas durante esses 4 anos pudessem ser superadas. Obrigado mais uma vez por ter colaborado infinitamente em minha formação profissional;

A todo o pessoal da **Genética** e, mais recentemente, do nosso **Laboratório de Pesquisas Forenses e Genômicas** (uia!): Fernanda, Nathalia, Nádia, Edilene, Guilherme, Maria Luiza, Alison e Letícia;

A toda equipe do Departamento de Genética (secretárias, funcionários, coordenadores e chefes) por manterem o funcionamento de nossa casa e por toda ajuda ao longo desses anos;

Aos **gatos** (rs): Flávia, Natália, Rafa e Sil. Conversar com vocês me dá uma alegria sem tamanho, *mon ami!* Hahaha!

Ao pessoal da **41ª turma de Biologia da USP Ribeirão**, por ainda tentarmos o nosso máximo para manter contato, mesmo com todos tão ocupados (ainda bem!) e tão longe infelizmente. Parte especial do agradecimento vai às nossas duas mães, por estarem desempenhando tão bem seus papéis criando essas duas lindas que fazem nossa turma ser ainda mais unida (e babona...rs);

Ao pessoal do **LGMF** (Chris, Nana, Marcelo, Hemerson, Claudia, Mada, Carlos) por toda amizade e companheirismo nestes últimos anos;

Um agradecimento às minhas meninas da **Terra do Nunca** (sim, pra mim continua sendo a Terra do Nunca). Carol, Gi e Carlinha: vocês tiveram um papel vital nessa minha fase atual. A distância machuca demais quando lembro da nossa época juntos, mas saber que ainda posso contar com vocês me dá uma alegria sem tamanho. Amo vocês! S2

À minha **família**, por todo o amor que temos um pelo outro, por todo apoio que sempre me deram nessa minha caminhada como pesquisador e pelo orgulho que têm de mim. Agradecer vocês é mais que necessário, é uma forma de lembrá-los que penso em vocês todos os dias apesar da distância e que a vida fica mais leve quando sei que tenho uma família que me ama.

À **Deus**, por me oferecer estas oportunidades de crescimento (por mais que tenham vindo todas juntas rs) tanto profissional quanto pessoal. Agradeço todos os dias pelo caminho trilhado e por permitir que esse caminho esteja me tornando uma pessoa melhor.

*O conhecimento torna a alma jovem e diminui a
amargura da velhice. Colhe, pois, a sabedoria.
Armazena suavidade para o amanhã.*

Leonardo da Vinci

Resumo

MARANO, L. A. **Estudo da diversidade dos genes *MC1R* e *SLC24A5* em populações globais: avaliação de aspectos evolutivos e ambientais.** 2015. 157f. Tese (Doutorado) – Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, 2015.

Dentre os vários marcadores genéticos existentes, alguns SNPs (Single Nucleotide Polymorphisms) podem estar associados à determinação de uma série de características fenotípicas (cor de pele, olhos e cabelos, estatura, forma do rosto, espessura do fio de cabelo) tendo sua predição um grande valor nas investigações forenses. Dentre os principais genes conhecidos por controlarem a pigmentação humana, através da produção da melanina, o *SLC24A5* (*solute carrier family 24, member 5*) e o *MC1R* (*melanocortin 1-receptor*) apresentam um papel fundamental na melanogênese. O advento do sequenciamento de nova geração permitiu o processamento de várias regiões genômicas e indivíduos simultaneamente, aumentando a disponibilidade e precisão de dados de genomas completos, como alcançado em estudos como o *1000 Genomes Project*. Nossas análises preliminares demonstraram que os dados de Fase 3 do *1000 Genomes* são muito mais confiáveis do que as versões anteriores. Com esses dados, obtidos para 26 populações globais, foram realizadas análises populacionais (diversidade haplotípica, desequilíbrio de ligação, redes de haplótipo e de variância molecular) para os genes *MC1R* e *SLC24A5* a fim de se compreender melhor seus padrões de diversidade, correlacionando-os à prováveis eventos de seleção natural que tenham moldado sua história evolutiva, como por exemplo a intensidade da radiação UV nas diferentes regiões geográficas. Alguns padrões foram encontradas entre os grupos africano, europeu e asiático, de acordo com a história evolutiva já descrita para estes grupos. Apesar disso, foi observado um padrão claro de varredura seletiva para o *SLC24A5* nos resultados obtidos, como o alto desequilíbrio de ligação e baixa diversidade haplotípica. As análises da diversidade do *SLC24A5* associadas com as zonas de incidência UV, no entanto, não apresentaram uma correlação clara entre a intensidade da radiação UV e a diversidade haplotípica.

Palavras-chave: *MC1R*, *SLC24A5*, sequenciamento de nova geração, genética de populações

Abstract

MARANO, L. A. ***MC1R* and *SLC24A5* gene diversity among global populations: assessment of environmental and evolutionary aspects** . 2015. 157f. Tese (Doutorado) – Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, 2015.

Among the various existing genetic markers, some SNPs may be associated with the determination of a series of phenotypic characteristics (skin, eyes and hair color, height, face shape, hair thickness) and its prediction could have a great value in forensic investigations. Among the major genes known to control human pigmentation through melanin production, *SLC24A5* (solute carrier family 24, member 5) and *MC1R* (melanocortin-1 receptor) have a major role in melanogenesis. The advent of next-generation sequencing has enabled processing of several individuals and genomic regions simultaneously while increasing the availability and accuracy of whole genomes data, such as 1000 Genomes Project has achieved. Our preliminary analysis showed that Phase 3 data from the 1000 Genomes are far more reliable than previous versions. Using this data, obtained for 26 global populations, several analyzes were performed (haplotype diversity, linkage disequilibrium, haplotype networks and molecular variance) for *MC1R* and *SLC24A5* in order to better understand their diversity patterns, correlating them to natural selection events which may have shaped their evolutionary history, such as UV radiation intensity in different geographical regions. Some patterns were found between African, European and Asian groups, according to the evolutionary history already described for these groups. Nevertheless, a strong pattern of selective sweep was observed for *SLC24A5* in our data, such as high linkage disequilibrium and low haplotype diversity. Analysis of *SLC24A5* diversity associated to UV, however, did not show a clear correlation between the UV radiation intensity and haplotype diversity.

Keywords: *MC1R*, *SLC24A5*, next-generation sequencing, population genetics

SUMÁRIO

1-INTRODUÇÃO	11
1.1 Aspectos evolutivos da variação de pigmentação humana	12
1.2 Genética da pigmentação	14
1.3 Sequenciamento de nova geração	18
1.4 Bancos de dados de genomas completos	19
2-HIPÓTESE.....	22
3-OBJETIVOS	24
3.1 Geral.....	25
3.2 Específicos	25
4-MATERIAL E MÉTODOS.....	26
4.1 Amostras populacionais	27
4.2 Sequências dos genes <i>SLC24A5</i> e <i>MC1R</i> obtidas pelo <i>1000 Genomes Project</i>	28
4.3 Validação dos dados processados pelo <i>1000 Genomes Project</i>	29
4.4 Análises populacionais	34
4.5 Desequilíbrio de ligação.....	37
4.6 Rede de haplótipos.....	38
4.7 Correlação com a variação da radiação UV	39
5-RESULTADOS	42
5.1 Validação dos dados processados pelo <i>1000 Genomes Project</i>	43
5.2 Análises populacionais	46
5.3 Correlação com a variação da radiação UV	82
6-DISCUSSÃO	85
6.1 Validação dos dados processados pelo <i>1000 Genomes Project</i>	86
6.2 Aspectos populacionais e Evolutivos do <i>MC1R</i> e <i>SLC24A5</i>	89
6.3 Correlação com a variação da radiação UV	97
Considerações finais.....	99

7-CONCLUSÃO.....	101
8-REFERÊNCIAS BIBLIOGRÁFICAS	103
9-ANEXO	113
10-MANUSCRITO	139

1-Introdução

1.1 Aspectos evolutivos da variação de pigmentação humana

Muitas mutações localizadas em regiões gênicas codificantes ou regulatórias podem determinar substituições de aminoácidos, modificando as propriedades funcionais da proteína traduzida, ou alterar níveis de produção da mesma, acarretando na expressão de fenótipos distintos. Diversos estudos já avaliaram a existência de polimorfismos associados à cor da pele, cabelos e olhos (Frudakis et al., 2007; Graf et al., 2007; Myles et al., 2007; Stokowski et al., 2007; Sulem et al., 2007; Han et al., 2008; Kayser et al., 2008; Sturm et al., 2008; Branicki et al., 2009; Duffy et al., 2010; Edwards et al., 2010; Weeden et al., 2011), espessura de fios de cabelo (Fujimoto et al., 2009), formas faciais (Claes et al., 2014), estatura (Watson, 2000; Perola et al., 2001; Pulker et al., 2007; Fujimoto et al., 2008) e calvície (Nuwaihdy et al., 2014).

Os principais fatores determinantes desses fenótipos de pigmentação, como cor de cabelos, pele e olhos, são o conteúdo e a distribuição de melanina nos indivíduos (Rees, 2003). A melanina é uma substância sintetizada a partir da oxidação enzimática do aminoácido tirosina, sendo que sua produção ocorre dentro de organelas denominadas melanossomos em células especializadas, denominadas melanócitos. Esses melanossomos são posteriormente secretados para os queratinócitos localizados na junção entre a epiderme e a derme (Mcevoy et al., 2006). Já no caso dos melanócitos dos olhos, os melanossomos não são secretados, mas mantidos no citoplasma (Nordlund, 1998).

Variações nas cores de pele e cabelo, observadas em diferentes regiões geográficas, resultam da produção, distribuição e empacotamento de dois tipos distintos de melanina depositados na epiderme ou fios de cabelo: a eumelanina, que é um pigmento castanho/preto e a feomelanina que é um pigmento vermelho/amarelo (Gerstenblith et al., 2007). Em cabelos ruivos há uma predominância das quantidades de feomelanina comparadas às de eumelanina, enquanto nos cabelos escuros as quantidades de eumelanina prevalecem, e nos cabelos loiros há pouco de cada tipo (Rees, 2003). Um fato interessante é que, apesar da origem embrionária comum dos melanócitos da pele e do cabelo, os genes que afetam a pigmentação podem se expressar de maneira independente,

gerando combinações conhecidas de cabelos escuros e pele clara ou cabelos claros e pele escura (Sturm et al., 2001).

Sabe-se que, apesar de todas as colorações de olhos possuírem um número similar de melanócitos, os olhos azuis contêm um mínimo de pigmentos e poucos melanossomos; olhos verdes são resultado de níveis intermediários de pigmentos, intensidade de melanina e número de melanossomos, enquanto olhos castanhos são resultantes de níveis altos de melanina e grande quantidade de partículas melanossomais (Sturm e Frudakis, 2004). Entretanto, estudos acerca das proporções eumelanina/feomelanina que possam vir a determinar a cor dos olhos ainda são inconclusivos.

Um dos grandes problemas relacionados à biologia humana está justamente nas tentativas de se explicar a evolução da variação da pigmentação entre as diversas populações. Algumas teorias abordam a seleção natural e dimorfismo sexuais como possíveis explicações para essa diversidade (Aoki, 2002; Jablonski e Chaplin, 2014).

No entanto, explicação evolutiva mais aceita acerca dessa variabilidade na pigmentação se refere à proteção contra a radiação ultravioleta (UV), particularmente em áreas da pele com alta exposição ao sol. Os seres humanos diferem dos outros primatas e de muitos mamíferos por terem perdido a maior parte dos pelos do corpo. Essa perda dos pelos teria como vantagem evolutiva uma capacidade de transpirar de maneira mais eficiente, o que contribuiria no processo de regulação da temperatura corporal. Entretanto, essa ausência de pelos teria gerado a necessidade de outro tipo de proteção contra a radiação ultravioleta (UV) (Rees, 2003).

A melanina é capaz de absorver radiação eletromagnética no comprimento de onda UV, protegendo assim o DNA, as proteínas e outras macromoléculas dos efeitos nocivos desses raios (Rees, 2003). Outros estudos demonstram que a melanina teria também a capacidade de remover radicais livres, resultantes da interação da radiação UV com os lipídios das membranas e outros componentes celulares (Carletti et al., 2014). Outro aspecto importante da fotoproteção está no fato que os folatos (uma das vitaminas B) são destruídos pela radiação UV, sendo que os mesmos são de extrema importância no processo de divisão celular e formação embrionária, influenciando as taxas de sucesso reprodutivo (Moan et al., 2012).

Pensando nesse aspecto, a pele negra ofereceria maior proteção em ambientes caracterizados por elevada incidência de radiação UV (baixas latitudes e/ou áreas descobertas). Assim, a alta incidência de raios UV nessas regiões atuaria contra o surgimento de quaisquer mutações que gerassem fenótipos mais claros (Figura 1).

Figura 1 – Predição da intensidade de pigmentação da pele humana baseada em modelos de regressão utilizando a intensidade da radiação UV (retirado de Jablonski e Chaplin, 2000)

No entanto, com o início das migrações populacionais para fora da África (regiões de maior latitude), essa pressão seletiva negativa conferida pelos raios UV seria relaxada, tolerando eventuais mutações que gerassem fenótipos claros. Além disso, peles escuras poderiam diminuir a absorção adequada de luz em ambientes de incidência reduzida, enquanto peles claras facilitariam o processo de fotossíntese de vitamina D em regiões onde a incidência de radiação UV não é tão alta, praticamente levando à fixação deste fenótipo em altas latitudes (Soejima e Koda, 2007).

1.2 Genética da pigmentação

Estima-se que mais de 170 genes estejam envolvidos com a produção, distribuição e secreção da melanina através dos melanócitos, contribuindo com o processo da pigmentação (Branicki et al., 2009; Schneider et al., 2009). A enzima tirosinase, codificada pelo gene *TYR*, é a conversora da tirosina em dopaquinona,

precursora por sua vez da eumelanina e feomelanina. Mutações nesse gene podem levar à ausência da proteína funcional, impedindo totalmente a produção de melanina, condição chamada de albinismo oculocutâneo tipo 1. Mutações nulas no gene *OCA2* podem também resultar levar a outro tipo de albinismo, por reduzir a produção de eumelanina. Acredita-se que dois outros genes, *SLC24A5* e *SLC45A2*, codifiquem proteínas responsáveis pelo transporte transmembrana de sódio-cálcio e sódio-prótons nos melanócitos, respectivamente (Ito e Wakamatsu, 2010).

Um aspecto interessante da melanogênese consiste no fato de que o pH dos melanócitos pode ter uma grande influência nas vias de produção de melanina. Observou-se que melanócitos obtidos de peles claras possuem melanossomos de pH ácido, enquanto aqueles obtidos de peles escuras possuíam um pH próximo do neutro (Smith et al., 2004). Evidências recentes levam a acreditar que polimorfismos dos genes *OCA2* e *SLC24A5* possam estar associados ao controle do pH, levando à acidificação dos melanossomos (Ito e Wakamatsu, 2010).

Até o momento, o gene *MC1R* (*melanocortin-1 receptor* – MIM 155555) é o mais bem caracterizado dentre os genes descritos que influenciam a variação normal de pigmentação em humanos (Gerstenblith et al., 2007; Liu et al., 2013). Diversos estudos mostraram que populações do Norte da Europa apresentam o *MC1R* altamente polimórfico, com diversos alelos de perda de função com alta frequência (Smith et al., 1998), alguns deles associados à fenótipos de pele clara, sardas e cabelos ruivos. No entanto, outros genes, como o *SLC24A5*, também vêm sendo amplamente estudados, uma vez que as proteínas por eles codificadas aparentemente desempenham função relevante na síntese de melanina (Pulker et al., 2007; Soejima e Koda, 2007; Tully, 2007; Sturm, 2009; Liu et al., 2013).

O *MC1R* é um gene composto por apenas um éxon de 951 pares de bases, localizado no cromossomo 16 (16q24.3), responsável pela codificação de uma proteína transmembrana de 7 passos (317 aminoácidos), chamada de receptor de melanocortina 1. Esse receptor está localizado na superfície celular dos melanócitos e desempenha papel importante na via de produção da melanina. Observou-se que o gene *MC1R* regula a produção dos principais constituinte da pigmentação humana, a eumelanina e a feomelanina (Makova e Norton, 2005).

O receptor de melanocortina 1 pertence à família dos receptores associados à proteína G. Em mamíferos, esse receptor responde ao hormônio melanócito-estimulante (α -MSH) que, quando ligado, ativa a proteína G, que por sua vez ativa a

adenilato-ciclase a sintetizar AMP cíclico (cAMP), elevando seus níveis. Com os níveis de cAMP aumentados, a transcrição de uma série de genes é alterada, entre eles o gene TYR, aumentando os níveis de tirosinase, levando por fim à produção de eumelanina, de cor castanha/preta. No entanto, a ausência ou redução parcial dessa sinalização levaria à produção de feomelanina (Makova e Norton, 2005).

O gene *SLC24A5* (*solute carrier family 24, member 5* – MIM 609802) está localizado na região cromossômica 15q21.1, possui 21.421pb, 9 éxons e codifica a proteína NCKX5 de 500 aminoácidos, sendo tal proteína membro de um canal iônico intracelular de sódio/cálcio dependente de potássio (Sturm, 2006). Além de participar na regulação do pH dos melanossomos, recentemente foi proposto que o referido gene desempenha papel fundamental na regulação dos níveis de cálcio nos melanossomos, sendo a melanogênese e a maturação dos melanossomos altamente dependentes desses níveis em tais células (Norton et al., 2007).

Estudos adicionais revelaram que a proteína NCKX5 gerada pelo gene *SLC24A5* pode ser também encontrada na região *trans* do complexo de Golgi (TGN) (Ginger et al., 2008). Experimentos funcionais mostraram que a proteína contendo a variante 111Thr possui uma atividade de troca iônica significativamente reduzida quando comparada à ancestral (Ala111) (Ginger et al., 2008; Cheli et al., 2009). Tal atividade de troca reduzida da variante 111Thr seria a causadora de um processo de melanogênese mais reduzida, resultando em fenótipos mais claros (Wilson et al., 2013).

Dados obtidos por Stokowski et al. (2007) em um estudo de associação de varredura genômica sugerem que o polimorfismo rs1426654 (Ala111Thr) apresente o maior efeito na pigmentação da pele dentre todos os outros polimorfismos/genes identificados. O alelo 111Thr se apresentou correlacionado com a pigmentação clara em populações afro-americanas e afro-caribenhas miscigenadas (Sturm, 2006) e também na população brasileira (Cerqueira et al., 2014; De Araujo Lima et al., 2015).

Apesar de alguns estudos mostrarem que esse polimorfismo pode resultar em modificações no pH do melanossomo, levando a alterações nas vias de síntese da melanina, Ginger et al. (2008) afirmam que seria equivocado relacionar tais modificações somente a essa função de canal iônico transmembrana. Uma de suas hipóteses afirma que a redução na troca de cálcio na TGN (que é uma armazenadora de cálcio) poderia levar a acidificação da mesma, interferindo na produção da tirosinase e, por consequência, alterando a síntese de melanina. A

segunda hipótese afirma que a proteína NCKX5 poderia ter também algum papel ainda desconhecido no transporte e organização das proteínas componentes dos melanossomos.

Um estudo prévio desenvolvido por nosso grupo de trabalho com o gene *SLC24A5* demonstrou a existência de diversas associações significativas envolvendo alelos e genótipos do SNP Ala111Thr com características de pigmentação como a discriminação entre cabelos claros e escuros, olhos claros dos escuros, pele clara da pele escura, além da presença de sardas (Mendes-Junior et al., 2010). O alelo A (Thr) parece ter um efeito “clareador” da pele, parcialmente dominante sobre o alelo G (Ala), sendo que foi sugerido que o *SLC24A5* seria responsável por 25-38% da diferença entre os índices de melanina observados entre europeus e africanos (Lamason et al., 2005). Outros trabalhos feitos com amostras da população brasileira corroboraram nossos dados, demonstrando uma associação do alelo A com pele, olhos e cabelos claros (De Araujo Lima et al., 2015).

Em um experimento utilizando siRNAs para silenciamento do *SCL24A5*, foi verificado que o *MC1R* está entre os poucos genes de pigmentação que tiveram seus níveis de transcritos reduzidos, sugerindo sua interação com uma das principais vias conhecidas da melanogênese (Wilson et al., 2013). Essa co-dependência foi avaliada por outros trabalhos, que demonstraram que uma expressão aumentada do inibidor do *MC1R*, *ASIP* (*Agouti Signaling Protein*), está associada a uma regulação negativa do *SLC24A5*, enquanto um aumento do α -MSH está associado a uma elevação nos níveis de expressão de *SLC24A5*, evidenciando uma conexão entre esses dois genes (Nadeau et al., 2008; Le Pape et al., 2009). No entanto, devido à alta variabilidade de fenótipos em indivíduos contendo o alelo derivado A (111Thr), foi sugerido que o clareamento causado pela presença do 111Thr no *SLC24A5*, poderia ter um efeito tolerante para a atuação de outros genes envolvidos na pigmentação de olhos e cabelos em europeus (Lamason et al., 2005).

Um levantamento global feito com as frequências de diversos SNPs do *MC1R* revelou diferenças consideráveis entre as populações avaliadas (Brasil, EUA, sul da Europa, Norte da Europa, África e Ásia). Comparando-se os dados de frequência mundiais de 19 *loci* comuns entre as populações, observa-se que as populações do Brasil, do norte e sul da Europa e dos Estados Unidos apresentavam variação em entre 57,89% e 84,21% dos marcadores, enquanto as amostras populacionais da África (26,32%) e da Ásia (31,58%) apresentavam reduzida variabilidade,

demonstrando possíveis aspectos de seleção purificadora nessas duas situações (Marano, 2011).

O polimorfismo Ala111Thr (rs1426654) do *SLC24A5* tem seu alelo 111Thr predominante (frequência variando de 98,7-100%) nas populações de origem europeia, enquanto as populações africana, ameríndia e do leste da Ásia apresentavam o alelo ancestral Ala111 dominante (variando de 93-100%). Assim, o rs1426654 pode ser considerado um dos *loci* com maior diferenciação entre Europa e África avaliados pelo HapMap, sugerindo forte atuação da seleção natural (Smith et al., 2004; Lamason et al., 2005). Embora o alelo A (111Thr) deste *locus* seja potencialmente um grande determinante do “clareamento” dos europeus, tal marcador não pode ser utilizado para a explicação do “clareamento” dos asiáticos, uma vez que esta população compartilha o mesmo alelo ancestral G (111Ala) com os africanos, evidenciando a influência de outros marcadores e genes na melanogênese (Lamason et al., 2005).

Considerando-se ainda os valores de heterozigose na região flanqueadora do *SLC24A5*, observa-se uma queda considerável em indivíduos europeus quando comparados a indivíduos de outras populações, sendo esta uma evidência de uma varredura seletiva (*selective sweep*) nessa região. Tal varredura não teria ocorrido na população ancestral do leste asiático, uma vez que a alta heterozigose predomina nessa região (Lamason et al., 2005).

1.3 Sequenciamento de nova geração

O surgimento do sequenciamento de Sanger em 1977 e sua automação associada a eletroforese capilar proporcionou o sequenciamento de genomas completos das mais variadas espécies. No entanto, por conta de algumas limitações dessa técnica (alto custo, relativa lentidão e restrição do número de amostras/fragmentos sequenciados por vez, por exemplo), surgem dificuldades na obtenção de genomas completos em múltiplas amostras, quando estudos populacionais são o objetivo.

Com o surgimento de métodos de sequenciamento de nova geração (*Next Generation Sequencing*), baseados em leituras massivas em paralelo, vários

indivíduos/regiões puderam ser processados simultaneamente. Além disso, a precisão das técnicas de sequenciamento de nova geração permite maior confiabilidade dos dados, uma vez que um mesmo trecho é sequenciado diversas vezes e vários parâmetros foram desenvolvidos a fim de permitir avaliar detalhadamente a qualidade das sequências geradas (Nielsen et al., 2011). Diversas metodologias foram propostas neste contexto, incluindo, por exemplo, PCR em emulsão (454 – Roche, Ion Torrent – Life) e amplificação em fase sólida (Solexa – Illumina) (Metzker, 2010; Vigliar et al., 2015).

Duas aplicações de sequenciamento de nova geração podem ser destacadas: o sequenciamento de genoma completo (*Whole Genome Sequencing* - WGS) e o sequenciamento de exoma completo (*Whole Exome Sequencing* - WES). O WES trata-se do sequenciamento feito apenas nos trechos de DNA que codificam proteínas (em torno de 1-2% do genoma humano), conhecido como exoma. Trata-se de uma maneira mais prática, barata e eficaz do que o Sequenciamento de Genoma Completo (WGS), mas que ainda pode ser extremamente informativa (Tennessen et al., 2012).

Conforme os custos de sequenciamento caem, diversos projetos colaborativos voltados para sequenciamento de genomas completos vêm surgindo, como por exemplo, o *1000 Genomes Project* (Genomes Project et al., 2012). Tais projetos, utilizando-se de amostras populacionais distintas, proporcionam a descoberta de novas variantes raras, elucidando detalhes importantes sobre aspectos evolutivos e demográficos que moldaram o genoma humano (Matullo et al., 2013).

1.4 Bancos de dados de genomas completos

A primeira fase do *1000 Genomes Project* incluiu diversos centros de pesquisa trabalhando em conjunto, entre 2008 e 2010, sequenciando 1.092 genomas completos de 14 populações. Inicialmente, grupos de no máximo 100 indivíduos não aparentados, mas de populações relacionadas, eram sequenciados. Por conta do elevado custo e tamanho do genoma, os resultados consistiam principalmente de dados genômicos de baixa cobertura e exômicos de alta

cobertura, além de painéis de SNP *array* de alta densidade. Esses dados, após serem lidos e alinhados, eram então tratados com múltiplos algoritmos que identificavam variantes através da análise de diversos parâmetros (qualidade de mapeamento, qualidade da base, posição da base variante dentro do fragmento e coeficiente de endocruzamento). Cada genótipo obtido era categorizado em sítios bialélicos (0, 1 ou 2 cópias da variante) para cada amostra e em cada sítio. A partir desses dados, análise de desequilíbrio de ligação foi empregada para confirmação/inferência ou até mesmo correção das variantes obtidas dos dados não-exômicos de baixa cobertura (Genomes Project et al., 2012). No entanto, os dados obtidos apresentam diversas falhas, como a ausência de alguns polimorfismos devido às limitações dos métodos de *calling* de genótipos disponíveis na época. Além dessas falhas, o método usado para inferência de genótipos e haplótipos não possibilitava a diferenciação de SNPs tri-alélicos, considerando-os bi-alélicos, reduzindo a informatividade de diversos *loci* (Castelli et al., 2014). A última fase (Fase 3), liberada em outubro de 2014, englobava 2.504 genomas completos, apresentando dados multialelicos pela primeira vez.

Estudos como o *1000 Genomes Project* podem servir de base para inferência de variantes não detectadas, uma vez que diversas delas estão localizadas em blocos de desequilíbrio de ligação conhecidos (Howie et al., 2011). As variantes de baixa frequência podem sofrer imputação com precisão de 60-90%, mesmo em populações miscigenadas (Genomes Project et al., 2012). Apesar de muitas melhoras terem sido alcançadas nas etapas de mapeamento, imputação e *calling* de genótipos, as tecnologias atuais ainda apresentam taxas de erros de, no mínimo, 10^{-4} por nucleotídeo, levando a uma alta taxa de falsos positivos (em torno de 5%) quando a baixa cobertura é utilizada (Matullo et al., 2013).

Diversos estudos têm abordado este problema na análise de dados se utilizando de métodos *in silico* para a identificação e correção de qualquer viés que possa resultar em erros de mapeamento e/ou de *calling* de genótipos (Gilissen et al., 2012; Nagasaki et al., 2013; Castelli et al., 2014), além de destacar a importância de se validar os resultados obtidos por outras metodologias estabelecidas e validadas, tais como sequenciamento Sanger, para as regiões de interesse. Novos scripts e *pipelines* foram desenvolvidos para a detecção precisa de genótipos e haplótipo a partir de dados brutos do *1000 Genomes Project*, permitindo também se obter dados inéditos a partir de variantes que foram deixadas de fora pelo projeto.

Esses scripts foram validados em diferentes análises do gene *HLA-G* com os dados da Fase 1, demonstrando uma maior precisão no *calling* de genótipos para genes individuais ou regiões cromossômicas delimitadas (Castelli et al., 2014). No entanto, a precisão de *calling* de genótipos para os dados da Fase 3 ainda não foi avaliada até o momento.

Apesar dos erros existentes, os dados obtidos pelo sequenciamento de nova geração pelo *1000 Genomes Project* pode ser valiosa para estudos evolutivos e de população para genes específicos. Este conjunto de dados pode ser utilizado para esclarecer um dos principais problemas relacionados à biologia da evolução humana: a evolução da variação de pigmentação as diferentes populações. Como já abordado, os genes envolvidos no processo de melanogênese podem ser considerado um dos principais alvos da seleção natural, fato que pode ser verificado pela correlação entre a diversidade da pigmentação e a variação da incidência de radiação ultravioleta (UV), uma vez que a distribuição dos fenótipos está claramente associada à variação da latitude (Figura 1).

Alelos de muitos SNPs da região codificante do *MC1R* (Arg151Cys, Arg160Trp e Asp294His) já foram associados à cabelos ruivos, pele clara e sardas nos europeus (Harding et al., 2000; Sturm, 2006). Outros estudos também identificaram diversas variantes do *MC1R* relacionadas com susceptibilidade ao melanoma (Ghiorzo et al., 2012; Pellegrini et al., 2012). A compreensão da expressão e regulação deste receptor é essencial para que se compreenda seus efeitos sobre a melanogênese e o risco de melanoma (Swope et al., 2012). Como a seleção natural age principalmente moldando a diversidade de regiões que tenham impacto funcional sobre a expressão gênica, a compreensão dos padrões globais de diversidade genética do *MC1R* é o primeiro passo para entender a sua expressão e atividade de regulação.

Portanto, considerando essas limitações, foi utilizado um *pipeline* desenvolvido por Castelli (2014) para se obter dados de genótipos e haplótipos a partir dos dados brutos de sequenciamento disponíveis no site FTP do *1000 Genomes Project*, permitindo verificar sua variabilidade ao redor do mundo e obter dados de alguns sítios polimórficos falhos da Fase 3.

2-Hipótese

Os dados obtidos pelo *1000 Genomes Project*, quando analisados de forma criteriosa, podem trazer informações genéticas valiosas e precisas a respeito dos genes *SLC24A5* e *MC1R* em populações globais, os quais apresentam padrões de diversidade moldado por eventos demográficos.

Apesar da alta taxa de erros existentes nos bancos de dados de sequenciamento de nova geração, considerando-se a evolução e melhorias dos processos de obtenção e análise desse tipo de dados, avaliamos a hipótese que os dados gerados pela Fase 3 do *1000 Genomes Project* apresentam maior confiabilidade e precisão que os dados liberados em fases anteriores.

Considerando-se a análise de dois genes sabidamente influenciados pela seleção natural em diversas populações globais, avaliamos também a hipótese de maior diversidade presente nas populações africanas, por conta de sua origem mais antiga, mas também a hipótese dos demais grupos (europeu, leste da Ásia e sul da Ásia) apresentarem padrões de diversidade possivelmente moldados por processos de *sweep* seletivo.

Sabe-se que os diversos grupos étnicos distribuídos pelo globo apresentam características fenotípicas particulares em relação à pigmentação dos olhos, pele e cabelos. A variação geográfica com relação à pigmentação da pele faz sentido quando analisada em função da radiação UV: a pele escura oferece maior proteção em ambientes caracterizados por elevada incidência de radiação UV (baixas latitudes), enquanto que a pele clara evita que as pessoas apresentem deficiência de vitamina D em regiões onde a incidência de radiação UV não é tão alta (elevadas latitudes) (Soejima e Koda, 2007). Neste contexto, avaliamos ainda a hipótese de que a maior incidência UV está associada à padrões de diversidade específicos nas populações habitantes dessas regiões, enquanto a incidência menor de UV estaria associada à padrões distintos, tanto para o gene *SLC24A5* quanto para o *MC1R*.

3-Objetivos

3.1 Geral

- Avaliar os padrões globais de diversidade dos genes *SLC24A5* e *MC1R* através da utilização dos dados de 26 amostras populacionais analisadas pelo consórcio *1000 Genomes Project*, averiguando eventuais assinaturas de seleção natural associadas à pigmentação da pele e à intensidade da radiação UV da região geográfica que estas populações ocupam.

3.2 Específicos

- Avaliar a confiabilidade dos resultados de genótipos faseados obtidos em arquivos VCF do *1000Genomes Project* através da comparação com resultados obtidos pelo processamento dos dados brutos pré-alinhados aos genes *SLC24A5* e *MC1R*, por meio de scripts desenvolvidos por nosso grupo de pesquisa (Castelli et al., 2014), em um processo de co-validação ;
- Determinar as frequências de alelos, genótipos e haplótipos compostos pelos *loci* distribuídos ao longo dos genes *SLC24A5* e *MC1R*;
- Avaliar parâmetros de diversidade dos genes *SLC24A5* e *MC1R* nos grupos populacionais, tanto individualmente, como comparados entre si;
- Avaliar a intensidade do desequilíbrio de ligação entre os alelos encontrados ao longo dos genes *SLC24A5* e *MC1R*;
- Avaliar a dinâmica da correlação entre os haplótipos dos genes *SLC24A5* e *MC1R* por meio da análise das redes de haplótipos obtidas.
- Avaliar a correlação entre alelos/haplótipos dos *loci* dos genes *SLC24A5* e *MC1R* e as características ambientais de incidência UV para os grupos populacionais avaliados em função da região geográfica ocupada.

4-Material e Métodos

4.1 Amostras populacionais

Foram utilizados dados obtidos pelo *1000 Genomes Project Consortium* (Genomes Project et al., 2012), um consórcio internacional que, por meio de técnicas de sequenciamento de nova geração, realizou o sequenciamento completo dos genomas de 2504 indivíduos de 26 populações ao redor do globo, obtendo dados de mais de 55 milhões de variantes curtas (SNPs, InDels e mutações somáticas). Essas 26 populações foram subdivididas em 5 grupos étnicos principais em nossas análises (Tabela 1), de acordo com sua ancestralidade predominante.

Dessa maneira, o grupo África (AFR) incluiu os afro-caribenhos de Barbados (ACB), os Esan da Nigéria (ESN), os gambianos da Divisão Oeste, Gâmbia (GWD), os Luhya de Webuye, Quênia (LWK), os Mende de Serra Leoa (MSL) e os Yoruba de Ibadan, Nigéria (YRI).

O segundo grupo foi composto por amostras de populações miscigenadas das Américas (AMR) e contou com indivíduos afro-americanos do sudoeste dos EUA (ASW), colombianos de Medellín, Colômbia (CLM), indivíduos de ancestralidade mexicana residentes em Los Angeles (MXL), peruanos de Lima, Peru (PEL) e porto-riquenhos residentes em Porto Rico (PUR).

O grupo do leste da Ásia (EAS) compreendeu chineses Dai de Xishuangbanna, China (CDX), chineses Han de Pequim, China (CHB), chineses Han do sul da China (CHS), japoneses de Tóquio, Japão (JPT) e indivíduos de Kinh in Ho Chi Minh, Vietnã (KHV).

O grupo Europa (EUR) contou com residentes de Utah, Estados Unidos, com ancestrais do Norte da Europa e da Europa Ocidental (CEU), finlandeses residentes na Finlândia (FIN), britânicos da Inglaterra e Escócia (GBR), indivíduos de populações ibéricas da Espanha (IBS) e habitantes da Toscana, Itália (TSI).

O último grupo foi composto pelos indivíduos do sul da Ásia (SAS), sendo representados os Bengali de Bangladesh (BEB), indianos Gujarati residentes em Houston, Texas (GIH), indianos Telugu residentes no Reino Unido (ITU), os Punjabi de Lahore, Paquistão (PJL) e os Tâmis do Sri Lanka, residentes no Reino Unido (STU).

Tabela 1 – Composição dos grupos étnicos avaliados e respectivos tamanhos amostrais (*n*)

Grupo	Populações	<i>n</i>	Total
África (AFR)	ACB	96	1200
	ESN	99	
	GWD	113	
	LWK	99	
	MSL	85	
	YRI	108	
América (AMR)	ASW	61	816
	CLM	94	
	MXL	64	
	PEL	85	
	PUR	104	
Leste Ásia (EAS)	CDX	93	1008
	CHB	103	
	CHS	105	
	JPT	104	
	KHV	99	
Europa (EUR)	CEU	99	1006
	FIN	99	
	GBR	91	
	IBS	107	
	TSI	107	
Sul Ásia (SAS)	BEB	86	978
	GIH	103	
	ITU	102	
	PJL	96	
	STU	102	
Total			2504

4.2 Sequências dos genes *SLC24A5* e *MC1R* obtidas pelo 1000 Genomes Project

Os dados genotípicos, referentes às regiões 5' regulatória (5'URR), 5' não-traduzida (5'UTR), região codificadora (éxons e introns) e 3' não-traduzida (3'UTR) dos genes *SLC24A5* e *MC1R* foram obtidos em arquivos .VCF (*Variant Call Format*)

em fase, por meio do *browser* do *1000 Genomes Project* (<http://browser.1000genomes.org>). Tais dados foram completamente processados pelo consórcio.

O gene *SLC24A5* está localizado na região cromossômica 15q21.1, possui 21.422pb, 9 éxons e codifica a proteína NCKX5 de 500 aminoácidos. Sua região 5'UTR possui 74pb, a codificante (considerando os 9 éxons e introns) tem 21.307pb, e a 3'UTR tem 41pb. As coordenadas da região a ser avaliada foram determinadas com uma margem extra de 3000 pb *upstream* dos ponto de início do gene definido pelo *assembly* GRCh38.p2 (chr15:48.117.971 - 48.142.392), totalizando uma região de 24,422kb. Nesse intervalo foram observados 78 sítios de variação na região 5'URR, 3 sítios na região 5'UTR, 475 sítios na região codificadora (sendo 434 intrônicos) e nenhum na região 3'UTR. Desse total de *loci*, 122 foram excluídos da análise, um por ser monomórfico (rs192237146) e 121 por serem de ocorrência única.

O *MC1R* é um gene composto por 3.099pb, localizado na região cromossômica 16q24.3, responsável pela codificação do receptor de melanocortina 1 de 317 aminoácidos. Sua região 5'UTR possui 1.380pb, a codificante tem 951pb em um único éxon, e a 3'UTR tem 768pb. Os dados do gene *MC1R* seguiram os mesmos parâmetros acima, totalizando uma região de 6,1kb (chr16:89.914.878 - 89.920.977), sendo observado um total de 296 sítios de variação: 132 deles na região 5'URR, 50 na região 5'UTR, 87 na região codificante e 28 na região 3'UTR. Foram excluídas da análise 53 *loci* desse total, um por ser monomórfico (rs541340274) e 52 de ocorrência única.

Os dados genotípicos foram posteriormente organizados para que se obtivesse as sequências dos haplótipos de cada indivíduo separados por região de interesse (5'URR, 5'UTR, Codificadora e 3'UTR). Os haplótipos foram então listados e nomeados de acordo com o número e posição das mutações presentes nos mesmos.

4.3 Validação dos dados processados pelo 1000 Genomes Project

A fim de avaliar a confiabilidade dos dados obtidos nos arquivos .VCF, foram baixados os dados brutos de sequências pré-alinhadas ao gene *MC1R* (arquivos .SAM - *Sequence Alignment/Map*), referentes ao mesmo conjunto de indivíduos.

Uma vez que o processo de obtenção dos genomas faseados pelo *1000 Genomes Project* se utiliza de procedimentos de alinhamento, determinação de genótipos e imputação que podem enviesar a real variabilidade dos dados, a obtenção e tratamento desses dados brutos através de scripts (Figura 3) desenvolvidos por nosso grupo de pesquisa (Castelli et al., 2014) permite uma análise comparativa para identificação de erros e co-validação de ambos os métodos.

Essa análise se utilizou do ambiente virtual da NuvemUSP, o qual permitiu que os dados fossem baixados diretamente do servidor FTP do *1000 Genomes Project* (<ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/>), além de permitir que os scripts pudessem ser rodados em máquinas virtuais agrupadas com alta capacidade de processamento.

Inicialmente, usando o Samtools (Li et al., 2009), foram baixados os arquivos SAM (*sequence alignment/map*) contendo os dados para a região do gene *MC1R* (pré-mapeadas no chr16:89.981.286-89.987.385) diretamente a partir do servidor do *1000 Genomes* (<ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/>). O download foi realizado para todas as 2,537 amostras disponíveis na Fase III e incluiu tanto os dados de genoma de baixa cobertura como os de exoma quando disponíveis. Este processo gerou até dois arquivos SAM por indivíduo que foram, em seguida, fundidos em um único arquivo e convertidos em um arquivo BAM (*binary alignment/map*). Cada arquivo BAM foi então convertido em um arquivo de formato Fastq que recuperou todas as leituras previamente mapeadas para a região do *MC1R*, através da utilização do Bamtools (<https://github.com/pezmaster31/bamtools/>) e scripts Perl (desenvolvidos localmente) para remover *reads* duplicadas e classificar as *reads* como *paired* ou *unpaired*.

Tanto os arquivos Fastq *paired* ou *unpaired* foram mapeados num cromossomo 16 (hg19), em que apenas a região de *MC1R* estava disponível, tendo o restante sido substituído por "N" para preservar posições de nucleotídeos em relação ao hg19. Os arquivos BAM resultantes das *reads* recém mapeadas foram unidos usando o software Picard-tools (<http://broadinstitute.github.io/picard/>). O software Bamtools também foi usado para remover as *reads* mapeadas com qualidade de mapeamento (MP) baixas (MQ<40). O UnifiedGenotyper do GATK foi usado para inferir genótipos e um arquivo VCF foi gerado.

Dada a baixa cobertura dos dados 1000Genomes, alguns genótipos obtidos pelo GATK são incertos, principalmente em casos em que um genótipo homozigoto

é inferido a partir de um *locus* com baixa cobertura. Além disso, por conta da alta variabilidade do MC1R, é esperado que ocorram algumas *reads* mapeadas erroneamente, enviesando a inferência dos genótipos. Para contornar esse problema, o arquivo VCF gerado pelo GATK foi tratado com um script Perl desenvolvido localmente que aplicou algumas regras (descritas abaixo). Este script usa o número de diferentes *reads* detectadas para cada alelo numa dada posição (fornecida pelo GATK quando o arquivo VCF foi gerado):

- Homozigotos só são inferidos quando uma cobertura mínima de sete *reads* foi observada; caso contrário, é introduzido um *missing allele* neste genótipo. Este procedimento assegura ($p > 0,99$) que é um genótipo homozigoto resultante foi obtido por conta de falta de variação naquela posição e não porque um segundo alelo não foi eventualmente amostrado;

- Nos casos em que um dos alelos foi extremamente sub-representado (proporção de *reads* abaixo de 5%), o genótipo foi considerado como homozigoto para o alelo mais frequente. Este procedimento minimiza a influência de *reads* contendo erros de mapeamento e reduz o alto nível de erros de sequenciamento característicos dos dados de sequenciamento de nova geração. Esse tipo de correção foi aplicado apenas em situações caracterizadas por uma elevada cobertura (20 ou mais *reads* na posição avaliada).

- Para os genótipos em que um dos alelos estava levemente sub-representadas (com uma proporção de *reads* entre 5 e 20%), um *missing allele* foi introduzido no lugar deste alelo menos frequente. Este procedimento é muito útil nas situações de baixa cobertura (menos de 20 *reads* na posição avaliada), nas quais uma única *read* pode indicar a existência de um alelo alternativo. Tal *read* pode ser consequência de um erro de mapeamento (falso positivo) ou pode representar um genótipo heterozigoto desbalanceado. Assim, a definição deste genótipo (homo ou heterozigoto) será inferida durante um passo de imputação posterior.

- Genótipos nos quais a proporção de *reads* para o alelo menos frequente seja superior a 20% foram considerados como heterozigotos. Isto assegura que somente genótipos heterozigotos com alta qualidade são passados para o processo de imputação.

Algumas SNVs detectadas anteriormente (com baixa qualidade) foram consideradas para monomórficas, uma vez que o alelo alternativo foi removido ou considerado *missing*, sendo desconsideradas para análises posteriores. Ao usar o

pacote VCFtools (Danecek et al., 2011), foram removidos os SNVs que não apresentavam mais variação ou aqueles que foram representados apenas uma vez no conjunto de dados (*singletons*). Os *missing alleles* foram então imputados, assim como os haplótipos foram inferidos através da utilização do algoritmo fastPHASE (Scheet e Stephens, 2006).

Todos os conjuntos de dados foram subdivididos em categorias de acordo a região gênica, sendo gerados arquivos de análise para as regiões 5'URR, 5'UTR, codificante e 3'UTR isoladamente, bem como para o gene completo (incluindo 5'URR) e região total considerando-se somente a extensão do mRNA (excluindo-se a 5'URR).

A co-validação das metodologias para análise dos dados do *1000 Genomes Project* foi realizada pela comparação entre os dados finais gerados pelo consórcio e pelos scripts locais, contando com três passos principais de avaliação:

- a) diferenças quanto à determinação de genótipos em cada *locus* de cada indivíduo: nesta etapa foram comparados os genótipos obtidos *locus a locus* em todo os indivíduos. Foram verificadas inconsistências entre os dois bancos para as 445.712 comparações possíveis (178 loci x 2504 indivíduos), sendo observado se o genótipo obtido no VCF diferia daquele obtido pelo script (0|0, 0|1 ou 1|1).
- b) diferenças quanto às frequências alélicas por *locus* analisado: foram comparadas as frequências alélicas nos 178 *loci* analisados, afim de se verificar diferenças na frequência do alelo referência;
- c) diferenças quanto às frequências haplotípica das regiões consideradas: foram verificados os casos em que a reconstrução do par de haplótipos foi idêntica nos indivíduos amostrados. Foram ainda comparadas as frequências haplotípicas daqueles haplótipos comuns aos dois bancos, além de ser verificada a presença de outros haplótipos exclusivos para cada abordagem.

4.4 Análises populacionais

O programa Arlequin 3.5 (Excoffier e Lischer, 2010) foi utilizado para o cálculo de diversos parâmetros para avaliação da diversidade intra-populacional, como a aderência das frequências genotípicas e haplotípicas ao Equilíbrio de Hardy-Weinberg (EHW), diversidade nucleotídica (Tajima, 1983; Nei, 1987), e da diversidade inter-populacional, como Análise de Variância Molecular (AMOVA) (Excoffier et al., 1992), estimativas de F_{ST} entre pares de populações (Weir e Cockerham, 1984) e teste exato de diferenciação populacional baseado nas frequências haplotípicas (Raymond e Rousset, 1995), além do teste de neutralidade de Slatkin (Slatkin, 1994; 1996).

4.4.1 Diversidade Nucleotídica

Uma das maneiras de avaliar a diversidade dos genes pode ser feita através da análise dos sítios de segregação (os sítios de nucleotídeos que variam) em uma sequência de nucleotídeos. Essa diversidade nucleotídica (π) descreve a probabilidade que duas cópias de um mesmo nucleotídeo, retiradas de um conjunto de sequências, sejam diferentes entre si. Essa medida de diversidade é análoga à diversidade gênica de Nei. Ou seja, é calculada a média da diversidade gênica em razão dos L loci avaliados (Jobling et al., 2004).

Assim, a comparação da diversidade nucleotídica entre as regiões gênicas ou entre as populações analisadas pode trazer informações a respeito de processos evolutivos atuantes sobre cada uma delas. Para esta etapa foram utilizados não somente os haplótipos gerados a partir das variantes, mas sim a sequência completa dos haplótipos contendo todas as bases presentes no trecho em questão.

4.4.2 Equilíbrio de Hardy-Weinberg (EHW)

O teste exato do Equilíbrio de Hardy-Weinberg (EHW) foi realizado seguindo o procedimento descrito por Guo e Thompson (1992) se utilizando de um teste análogo ao teste exato de Fisher implementado pelo programa Arlequin 3.5. Como os dados empregados no presente estudo apresentavam fase gamética

conhecida, foi realizada tanto a análise de *locus a locus* quanto a análise do haplótipo todo.

4.4.3 AMOVA (Análise de Variância Molecular - *Analysis of Molecular Variance*)

A Análise de Variância Molecular (AMOVA) é utilizada para estimar índices de estruturas genéticas utilizando informações da composição alélica de um haplótipo, bem como suas frequências. Assim, os resultados obtidos permitem medir a diferenciação entre as populações analisadas, elucidando padrões de agrupamento/distanciamento entre elas.

A AMOVA aqui utilizada é similar a outros métodos baseados nas análises de variância de frequência gênica, no entanto, ela leva em consideração o número de mutações entre dois haplótipos. Ao serem definidos os grupos populacionais (África, Europa, leste da Ásia e sul da Ásia), também é definida a estrutura genética a ser testada. O grupo das amostras das Américas foi removido desta análise por não trazer nenhuma informatividade, já que se trata de uma população híbrida. A análise hierárquica da variância acaba particionando a variância total em componentes de covariância, devido a diferenças inter-individuais e/ou inter-populacionais.

As informações sobre a composição de cada haplótipo são inseridas em uma matriz e então são testadas as significâncias de todos os componentes de covariância associados em três níveis principais de estrutura genética: entre indivíduos (Φ_{ST}), entre populações (Φ_{SC}) e entre grupos de populações (Φ_{CT}) se utilizando de 1.000 permutações.

4.4.4 Comparações interpopulacionais (F_{ST})

O F_{ST} é um dos índices de fixação que mede o desvio das frequências dos heterozigotos observadas em relação às esperada sob EHW. Esse parâmetro mede a distribuição de variação genética entre as subpopulações, comparando a diversidade encontrada dentro de cada um delas em relação à diversidade encontrada na população total se utilizando das diferenças nas frequências alélicas entre as subpopulações. O F_{ST} é uma medida de diferenciação populacional, calculado utilizando o coeficiente de endogamia de um par de populações, se

utilizando da probabilidade de dois alelos aleatórios serem idênticos por descendência.

Foram calculados os F_{ST} para cada par de populações, utilizando-se como distância genética através da fórmula:

$$F_{ST} = \frac{V_p}{p(1-p)}$$

Onde V_p e p são, respectivamente, a variância e a média das frequências entre as duas populações analisadas (Jobling et al., 2004).

A significância das distâncias genéticas é testada pela permutação dos haplótipos ou indivíduos entre as populações. Foi utilizada a distância de Reynolds (1983), calculada através do F_{ST} linearizado para tempos curtos de divergência, com 1000 permutações e significância de 0,05.

4.4.5 Diferenciação populacional

O teste exato de diferenciação populacional avalia a hipótese de distribuição aleatória dos indivíduos entre pares de populações. É um teste análogo ao teste exato de Fisher feito sobre uma tabela 2 x 2, mas ampliado para uma tabela de 2 x (número de haplótipos). Foram selecionados o valor padrão de número de passos na cadeia de Markov (100.000). O número de memorização (o número de passos a serem realizados antes do início da comparação das probabilidades de tabela com a tabela observada) foi estabelecido em 10.000 e o nível de significância em 0,05.

4.4.6 Testes de Neutralidade

Foi utilizado o teste exato de neutralidade de seleção de Slatkin (Slatkin, 1994; 1996) baseado na teoria de amostragem de Ewens, o qual considera que as taxas de mutação e deriva da população estão em equilíbrio (Ewens, 1972). Neste teste, a probabilidade das amostras observadas é comparada à uma amostra aleatória neutra com o mesmo número indivíduos e alelos. A probabilidade de neutralidade seletiva da amostra é a proporção das amostras aleatórias, que sejam tão ou menos prováveis que as amostras observadas. O número de amostras simuladas (amostras aleatórias geradas para comparação com as amostras

observadas) foi estabelecido em 10.000. A definição de haplótipos para esse teste (parâmetro de grande importância, já que os testes são baseados no número de alelos nas amostras) foi inferida a partir da matriz de distância. Segundo a descrição do teste, os valores de probabilidade obtidos podem ser considerados significativos quando maiores que 0,95.

4.5 Desequilíbrio de ligação

Os dados haplotípicos foram formatados para serem inseridos no programa Haploview 4.2 (Barrett et al., 2005), o qual utiliza uma representação gráfica para visualização dos blocos haplotípicos presentes ao longo do intervalo analisado e das regiões com forte desequilíbrio de ligação. O Haploview permite a análise do nível de desequilíbrio entre cada par de *loci* do segmento de DNA em questão. Um dos valores que o programa leva em consideração é o coeficiente de desequilíbrio D' para os pares analisados, representado por uma legenda de cores indicando valores mais próximos de zero ou de um.

Nessa etapa, foram removidos os *loci* trialélicos, uma vez que o programa só aceita marcadores bialélicos. Além disso, foram também removidas as Inserções-Deleções (InDels), já que o *input* do Haploview só aceita dados obtidos de SNPs. Foram ainda gerados arquivos informando a distância entre cada um dos marcadores analisados para cada região gênica.

Foram considerados os blocos gerados por *default* pelo programa, sem alterações, representados pelo esquema padrão de cores. Os gráficos gerados se utilizam de quadrados com três cores principais para representação da intensidade de desequilíbrio entre os pares de *loci* através dos valores de D' e de LOD – *Log of the Likelihood Odds Ratio*. A cor branca representa pares com valor baixo de D' (próximos a 0), indicando ausência de desequilíbrio; os tons de vermelho representam pares com valor alto de D' (próximos a 1) e de LOD, indicando desequilíbrio elevado e significativo, enquanto a cor azul representa valores altos de D' , mas baixo LOD (menor que 2), representativo de um desequilíbrio elevado, mas não-significativo. Todos os valores de D' menores que 1 são representados dentro dos respectivos quadrados. Por fim, no alto do gráfico são representados os *loci*

utilizados na análise, identificados por seus respectivos *rs number* e localizados de acordo com sua distância relativa no trecho analisado.

4.6 Rede de haplótipos

Para cada região gênica, foram construídas *networks* (redes) de haplótipos em cada um dos grupo étnicos descritos anteriormente utilizando-se o programa Network 4.6 (Fluxus Engineering), através do algoritmo *median-joining* (Bandelt et al., 1999) relacionando os haplótipos obtidos do *1000 Genomes Project*. Nela, os haplótipos são representados por círculos enquanto as linhas que os unem representam as mutações que diferenciam uns dos outros. A frequência de cada haplótipo é representada proporcionalmente pelo tamanho dos círculos, enquanto as cores representam a frequência relativa do haplótipo entre as cinco populações parentais (Vermelho – África, Verde – Américas, Amarelo – leste da Ásia, Azul – Europa e Laranja – sul da Ásia). Esse tipo de recurso é muito útil na elaboração de inferências a respeito da evolução dos haplogrupos originados a partir de haplótipos ancestrais, além de permitir a visualização das relações existentes entre os haplótipos analisados.

Inicialmente foram gerados arquivos contendo os haplótipos alinhados de cada uma das regiões gênicas de todos indivíduos, contendo as variantes obtidas do *1000 Genomes Project*. Como parâmetros de análise, foram consideradas as nomenclaturas geradas pelo Arlequin para os haplótipos analisados. As variantes que diferenciam um haplótipo do outro são indicadas pelas linhas ligando os círculos e estão identificadas por seu respectivo *rs number*. Além disso, as InDels foram consideradas como tendo peso 2 quando comparadas às transições/transversões, de acordo com orientações do manual.

A rede foi calculada através do algoritmo *median-joining* e, para aqueles casos em que havia excesso de haplótipos de baixa frequência, a rede passou pelo processamento de máxima parcimônia (MP), o qual permite que a rede fique mais limpa e inteligível, removendo nós excessivos (Polzin, 2003). Foram então inseridas manualmente as frequências de cada haplótipo, relativas a cada uma das cinco populações, gerando as distribuições de cores observadas.

A partir da rede final obtida, os nós eram realinhados e organizados manualmente de modo a permitir uma melhor visualização e compreensão das distribuições dos haplótipos ao longo das populações.

4.7 Correlação com a variação da radiação UV

Foi ainda analisado se a variação da radiação UV (em função da latitude, principalmente) seria um fator que poderia afetar a distribuição das variações dos genes *MC1R* e *SLC24A5* entre as populações analisadas.

Para isto, foram usados dados utilizados por Jablonski e Chaplin (2000), no qual foram compiladas medições da dose mínima eritemal de UV (UVMED), que é a quantidade de radiação UV necessária para induzir uma vermelhidão mínima em peles claras, entre os anos de 1978 e 1993 obtidos pela NASA (*National Aeronautics and Space Administration*).

As 26 populações do *1000 Genomes Project* foram distribuídas em 3 grupos principais de acordo com as três faixas principais de intensidade da radiação UV: Baixa, Média e Alta intensidade (Figura 3). Foram removidas da análise as cinco populações do grupo Americano, além dos Afro-caribenhos em Barbados por serem originárias da miscigenação de populações parentais vindas de outros continentes, portanto, não representando nenhum vínculo evolutivo com o ambiente que ocupam atualmente. Foram utilizadas as coordenadas geográficas em que as amostras populacionais foram coletadas, considerando-se o ponto central da região ou sua respectiva capital.

Afim de se verificar possíveis correlações entre a variação das regiões gênicas do *MC1R* e *SLC24A5* e a posição geográfica ocupada pelas populações (a incidência UV, portanto), foi realizada uma AMOVA com a subdivisão das populações conforme a Tabela X. Os 2000 indivíduos das 20 populações foram distribuídos nos três grupos principais de acordo com as áreas estabelecidas na Figura 3.

Tabela 2 – Composição populacional dos grupos analisados em função da zona de incidência UV e respectivos tamanhos amostrais (*n*)

Grupo	Populações	<i>n</i>	Total
Baixa Incidência UV	GBR	91	289
	CEU	99	
	FIN	99	
Média Incidência UV	IBS	107	517
	TSI	107	
	PJL	96	
	CHB	103	
	JPT	104	
Alta Incidência UV	GWD	113	1194
	MSL	85	
	YRI	108	
	ESN	99	
	LWK	99	
	GIH	103	
	STU	102	
	ITU	102	
	BEB	86	
	CDX	93	
	KHV	99	
	CHS	105	
Total			2000

Figura 3 – Distribuição das 20 populações do 1000 Genomes Project de acordo com a região geográfica e zona de incidência UV ocupada. Estão indicadas no mapa a localização de cada uma das populações, além das regiões de alta, média e baixa incidência UV (adaptado de Jablonski e Chaplin (2000)).

5-Resultados

5.1 Validação dos dados processados pelo *1000 Genomes Project*

Após as etapas de conversão dos dados brutos do *1000 Genomes Project* para os haplótipos faseados do *MC1R* para todos os indivíduos, o processo resultou em dados completos para 201 *loci* em 2.537 indivíduos, os quais resultaram em 596 haplótipos distintos. Destes 201 *loci*, 24 pertenciam à 5'URR, 54 pertenciam à 5'UTR, 89 pertenciam à região codificante e 34 à 3'UTR. Como os dados obtidos diretamente do VCF do *1000 Genomes* tinham apenas 178 *loci* em comum (sendo 174 SNVs e 4 InDels) em 2.504 indivíduos, foram removidos os 23 *loci* (11 deles *singletons*) omitidos pelo VCF, além de 33 indivíduos que eram incompatíveis entre os dois conjuntos de dados, permitindo equivalência em comparações.

Após esta filtragem inicial, os genótipos de cada *locus* foram comparados entre os indivíduos, a fim de verificar se houveram incompatibilidades. Numa análise inicial, considerando-se o número total de *loci* (178) e indivíduos (2.504), 445.712 comparações eram possíveis entre genótipos das duas análises. Deste total, apenas 585 incompatibilidades foram observadas no *calling* de genótipos (0,13% do total).

Verificou-se que 87,58% da amostra (2.193 indivíduos) possuíam genótipos idênticos entre os dois conjuntos de dados para todos os *loci* analisados. Os 311 indivíduos restantes tinham pelo menos um *locus* com genótipos incompatíveis, sendo que esta taxa de erro chegou a atingir um máximo de 13 *loci* incompatíveis num único indivíduo (Figura 1). Ao ser analisado individualmente cada *locus*, observou-se que 102 deles eram compatíveis entre as duas abordagens para todos os indivíduos. Entre os 76 *loci* discordantes, observamos que alguns deles reuniam a maioria das discrepâncias: 13 deles eram responsáveis por quase 80% de todos os *mismatches* ao longo da análise e um deles (rs885479) foi responsável por 87 *mismatches* (14,87% do total) (Figura 2). Se considerarmos as regiões genéticas isoladamente, a 5'UTR foi a que apresentou mais *loci* contendo *mismatches*, totalizando 32 *loci* do 48 (66,67%) encontrados nesta região.

Figura 4 – Número de *mismatches* observados entre os dados gerados pelo nosso estudo e dados obtidos diretamente do VCF em relação aos 2.504 indivíduos analisados pelo *1000 Genomes Project*.

Figura 5 – Número de *mismatches* observados entre os dados gerados pelo nosso estudo e dados obtidos diretamente do VCF em relação aos 178 *loci* analisados pelo *1000 Genomes Project*.

Ao analisar as frequências alélicas para cada *locus* avaliado, notamos que, mesmo com os *mismatches* mencionados, não houveram diferenças significativas entre as duas abordagens. Mesmo o *locus* rs885479, que exibiu a maior quantidade de incompatibilidades, acabou apresentando uma diferença irrelevante de 1,22% para a frequência do alelo referência entre o VCF e os dados processados pelos *scripts*.

A fim de avaliar detalhadamente possíveis causas destes *mismatches*, foram verificados os arquivos de alinhamento BAM baixados diretamente do *browser* do *1000 Genomes Project* usando o software IGV para explorar manualmente os *mismatches* entre as duas abordagens. A partir das 585 diferenças genótípicas originais, foram avaliadas 40 *mismatches* aleatórios para o *locus* rs885479 e 60 *mismatches* escolhidos aleatoriamente entre os *loci* restantes (Tabela 3). A partir dos 40 *mismatches* avaliados para o *locus* rs885479, 26 deles tiveram sete *reads* ou menos, e foi automaticamente introduzido pelo *script* um *missing allele* em cada leitura, impedindo nossas comparações. Os outros 14 *mismatches* tinham menos de vinte *reads*, com 5 deles apresentando o alelo alternativo sub-representado (razão entre 5-20%), o que também resultou na introdução de um *missing allele* pelo *script*. Observou-se que 7 *mismatches* aparentemente ocorreram devido a algum erro de *calling* do genótipo em nosso *pipeline* (3 e 4 casos deveriam ter sido considerados heterozigotos e homozigotos, respectivamente) e outros 2 casos tinham *calling* de genótipo errado no VCF. A partir dos 60 *mismatches* aleatórios, observou-se que nosso *pipeline* introduziu *missing alleles* em 33 deles, seja por causa do baixo número de *reads* ou por causa da sub-representação do alelo alternativo, impedindo maiores comparações. Os outros 26 *mismatches* ocorreram devido a erros cometidos por nossos *scripts* (15 e 11 casos, deveria ter sido considerados como homozigotos e heterozigotos, respectivamente) e um dos casos o *calling* de genótipo foi realizado erroneamente pelo VCF.

Surpreendentemente, observou-se também que os 178 *loci* nos 2.504 indivíduos, que resultaram em 207 haplótipos distintos para os dados do VCF, acabaram resultando em 490 haplótipos diferentes para os dados gerados pelos nossos *scripts* (Tabela 1). Cabe notar que nem todos os 207 haplótipos obtidos a partir do VCF estavam entre aqueles obtidos pelos *scripts*. Desse total, 52 haplótipos não estavam entre os 490 inferidos pelo nosso *pipeline*. Comparando-se os pares de haplótipos obtidos para cada indivíduo, observou-se que, a partir dos 5.008

haplótipos gerados, 2.455 comparações de haplótipos eram idênticas e 2.553 possuíam algumas inconsistências. Este resultado é ainda mais surpreendente se for levado em conta o fato de que 87,58% da amostra (2.193 indivíduos) apresentaram dados de genótipos idênticos entre os dois conjuntos de dados para todos os *loci* analisados, sendo que apenas 1.147 deles possuíam reconstruções de haplótipos idênticas.

Tabela 3 – Comparação entre os dados do *1000 Genomes Project* obtidos pelas análises do nosso *pipeline* e pelo VCF.

<i>Mismatches</i>	<i>locus</i> rs885479	Sítios Aleatórios
<i>Missing alleles</i>	31	33
Erros do VCF	2	1
Erros do <i>Pipeline</i>	7	26

5.2 Análises populacionais

5.2.1 Equilíbrio de Hardy-Weinberg

Considerando-se a média de cada grupo populacional dos valores de probabilidade obtidos por meio de teste exato para aderência ao Equilíbrio de Hardy-Weinberg (EHW) nas regiões gênicas do *MC1R* e do *SLC24A5*, observou-se que em nenhuma das situações analisadas ocorreram desvios significativos do EHW ($p < 0,05$). Os valores de heterozigose média para AFR encontravam-se normalmente superiores em relação aos demais grupos populacionais em todas regiões analisadas do *MC1R* (Tabela 4), sendo que EAS constituiu no grupo populacional onde foram observados os maiores desvios entre observado e esperado (deficiência de heterozigotos) em todas as regiões gênicas. No que se refere ao *SLC24A5* (Tabela 5), a maior diversidade em AFR só não foi observada na região codificante (apenas éxons), enquanto EUR apresenta os menores valores de diversidade. É interessante observar que SAS apresenta normalmente baixos valores de

diversidade (inferiores apenas aos observados em EUR), em todas as situações, exceto na região codificante (somente éxons). Ao contrário do *MC1R*, não foram encontradas diferenças marcantes entre observado e esperado.

5.2.2 Testes de Neutralidade

No teste de neutralidade de Ewens-Watterson, a homozigose observada sob EHW (F_o) é comparada com a homozigose esperada, calculada por simulações sob a hipótese de neutralidade (F_s). Os valores de p obtidos consistem na probabilidade de se obter F_o maior ou igual a F_s . Segundo a descrição do teste, os valores de probabilidade obtidos podem ser considerados significativos quando inferiores a 0,05 ou quando superiores a 0,95, evidenciando a presença de assinaturas de seleção balanceadora ou direcional, respectivamente.

Foram observadas significâncias, para o *MC1R*, em praticamente todas as regiões gênicas, em praticamente todos os grupos populacionais, exceto nas regiões 5'UTR e 3'UTR (Tabela 4), sendo que os maiores valores de significância foram observados na região Total Estendida.

Para o *SLC24A5* foram observados testes significativos para todos os grupos, sendo os maiores valores de p observados principalmente nas regiões Codificante (éxons e introns) e Total (tanto estendida quanto mRNA), sendo observada a exceção na região 5'UTR, que não teve parte de seus testes de neutralidade calculado, por ser composta por apenas três sítios de variação que apresentam alelos raros (Tabela 5). EAS foi a única população a apresentar assinaturas de seleção em todas as regiões gênicas.

Tabela 4 – Heterozigose média observada, heterozigose média esperada, valores de probabilidade de aderência ao equilíbrio de Hardy-Weinberg obtidos por meio de teste exato, e testes de neutralidade de Ewens-Watterson para o gene **MC1R** em suas respectivas regiões gênicas, separadas pelos grupos populacionais analisados (os valores significativos estão marcados em negrito).

	<i>Ho</i>	<i>Hs</i>	<i>p</i> EHW	<i>Fo</i>	<i>Fs</i>	<i>p</i> Slatkin
5'URR						
AFR	0,9225	0,9050	0,5686	0,1034	0,0725	0,9722
AMR	0,8028	0,7896	0,4134	0,2155	0,1159	0,9974
EAS	0,7742	0,6956	0,3454	0,3078	0,1873	0,9864
EUR	0,7422	0,7342	0,4303	0,2379	0,1463	0,9547
SAS	0,8719	0,8571	0,4761	0,1473	0,1019	0,9851
5UTR						
AFR	0,8206	0,8142	0,4780	0,1912	0,2695	0,3790
AMR	0,6508	0,6479	0,4162	0,3563	0,3447	0,6615
EAS	0,5529	0,5080	0,3701	0,4945	0,5709	0,4937
EUR	0,6138	0,6005	0,5967	0,3659	0,3530	0,8021
SAS	0,6803	0,6829	0,4645	0,3206	0,2573	0,8827
Codificante						
AFR	0,6830	0,6639	0,5257	0,3518	0,2697	0,8530
AMR	0,6331	0,6412	0,4439	0,3630	0,2300	0,9757
EAS	0,5862	0,5474	0,4047	0,4553	0,3333	0,9610
EUR	0,6465	0,6679	0,3208	0,3259	0,2206	0,8126
SAS	0,4819	0,4927	0,3868	0,5098	0,2883	0,9375
3'UTR						
AFR	0,6911	0,6683	0,4068	0,3328	0,3964	0,5163
AMR	0,3317	0,3409	0,5630	0,6614	0,4442	0,7588
EAS	0,4325	0,3956	0,2604	0,6064	0,7813	0,3322
EUR	0,2797	0,2879	0,5282	0,6523	0,5240	0,7846
SAS	0,3780	0,3824	0,5501	0,6196	0,5226	0,5756
Total Estendida						
AFR	0,9715	0,9627	0,7999	0,0424	0,0326	0,9907
AMR	0,9270	0,9057	0,6539	0,1001	0,0516	0,9991
EAS	0,8826	0,8479	0,5969	0,1563	0,0637	1,0000
EUR	0,9559	0,9307	0,5931	0,0684	0,0486	0,9502
SAS	0,9420	0,9217	0,6459	0,0830	0,0454	0,9997
Total (mRNA)						
AFR	0,9208	0,9123	0,3360	0,0955	0,0890	0,8157
AMR	0,8844	0,8429	0,4248	0,1625	0,1128	0,9594
EAS	0,8034	0,7543	0,4568	0,2495	0,1294	0,9940
EUR	0,9280	0,8781	0,2030	0,1180	0,1042	0,7646
SAS	0,8854	0,8322	0,3287	0,1720	0,1168	0,9903

(*Ho* – Heterozigose Média Observada, *Hs* – Heterozigose Média Esperada, *p*EHW – probabilidade de aderência ao Equilíbrio de Hardy-Weinberg, *Fo* – *F* observado, *Fs* – *F* esperado, *p*Slatkin – probabilidade atribuída ao teste de neutralidade de Ewens-Watterson, sendo que valores superiores a 0,95 são considerados significantes, AFR – África, AMR – América, EAS – leste Ásia, EUR – Europa, SAS – sul Ásia)

Tabela 5 – Heterozigose média observada, heterozigose média esperada, valores de probabilidade de aderência ao equilíbrio de Hardy-Weinberg obtidos por meio de teste exato, e testes de neutralidade de Ewens-Watterson para o gene **SLC24A5** em suas respectivas regiões gênicas, separadas pelos grupos populacionais analisados (os valores significativos estão marcados em negrito).

	<i>Ho</i>	<i>Hs</i>	<i>p</i> EHW	<i>Fo</i>	<i>Fs</i>	<i>p</i> Slatkin
5'URR						
AFR	0,8351	0,8206	0,3582	0,1835	0,1947	0,6901
AMR	0,4283	0,4162	0,5274	0,5866	0,4169	0,7639
EAS	0,6394	0,6418	0,5287	0,3614	0,2407	0,9616
EUR	0,0317	0,0315	1,0000	0,9687	0,6324	0,9926
SAS	0,4104	0,4046	0,6581	0,5975	0,3505	0,9510
5'UTR						
AFR	0,0015	0,0015	1,0000	0,9912	0,8333	1,0000
AMR	0,0033	0,0033	1,0000	0,9837	0,8149	1,0000
EAS	0,0022	0,0022	1,0000	0,9893	0,8286	1,0000
EUR	0,0000	0,0000	1,0000	-	-	-
SAS	0,0000	0,0000	1,0000	-	-	-
Codificante (Éxons e Introns)						
AFR	0,9509	0,9484	0,7628	0,0564	0,0348	0,9961
AMR	0,7897	0,7723	0,6872	0,2327	0,0795	0,9999
EAS	0,9080	0,8963	0,5555	0,1082	0,0660	0,9994
EUR	0,3587	0,3625	0,5128	0,6393	0,1532	1,0000
SAS	0,6462	0,6544	0,7158	0,3490	0,0751	1,0000
Codificante (Somente Éxons)						
AFR	0,1725	0,1760	0,7522	0,8249	0,4577	0,9762
AMR	0,3970	0,4068	0,4594	0,5958	0,7059	0,5018
EAS	0,0631	0,0618	1,0000	0,9385	0,5647	0,9767
EUR	0,0103	0,0103	1,0000	0,9830	0,7883	0,9432
SAS	0,3676	0,3874	0,6577	0,6146	0,6594	0,6951
Total Estendida						
AFR	0,9691	0,9606	0,8293	0,0443	0,0298	0,9849
AMR	0,7959	0,7794	0,7366	0,2257	0,0730	0,9999
EAS	0,9241	0,9124	0,6538	0,0922	0,0510	0,9999
EUR	0,3823	0,3844	0,5957	0,6175	0,1336	1,0000
SAS	0,6737	0,6743	0,7589	0,3292	0,0671	1,0000
Total (mRNA)						
AFR	0,9509	0,9484	0,7706	0,0564	0,0348	0,9957
AMR	0,7897	0,7724	0,6961	0,2326	0,0789	0,9999
EAS	0,9080	0,8966	0,5452	0,1078	0,0653	0,9996
EUR	0,3587	0,3625	0,5079	0,6393	0,1529	1,0000
SAS	0,6462	0,6544	0,7030	0,3490	0,0752	1,0000

(*Ho* – Heterozigose Média Observada, *Hs* – Heterozigose Média Esperada, *p*EHW – probabilidade de aderência ao Equilíbrio de Hardy-Weinberg, *Fo* – *F* observado, *Fs* – *F* esperado, *p*Slatkin – probabilidade atribuída ao teste de neutralidade de Ewens-Watterson, sendo que valores superiores a 0,95 são considerados significantes, AFR – África, AMR – América, EAS – leste Ásia, EUR – Europa, SAS – sul Ásia)

5.2.3 Diversidade Nucleotídica

Os resultados de diversidade nucleotídica sumarizam as comparações par a par dos *loci* de cada região gênica entre todos os indivíduos de um mesmo grupo (número de *loci* multiplicado pelo número de indivíduos).

Foi observado para o *MC1R*, considerando-se a região total estendida, que os valores foram próximos entre si para todos os grupos. Considerando-se as regiões gênicas isoladamente, observou-se que os maiores valores de diversidade se concentravam na região 5'UTR (Tabela 9), com uma média de 0,2070%, sendo o maior valor obtido para os grupos africano e do sul da Ásia. As menores diversidades foram observadas para as regiões Codificante e 3'UTR, com valores médios de 0,0990% e 0,1023%, respectivamente. Cabe ressaltar, no entanto, que o grupo africano apresentou valores altos (0,1972%) para a região 3'UTR.

Para o *SLC24A5* foi verificado que os menores valores de diversidade foram observados nas regiões 5'UTR e Codificante (contendo somente éxons), com médias de 0,0019% e 0,0143% respectivamente (Tabela 10), enquanto as regiões Codificante (éxons e introns) e Total apresentaram os maiores valores (média de 0,0366% e 0,0362% respectivamente). A população europeia apresentou marcadamente os menores valores de diversidade ao longo do *SLC24A5*, com uma média de 0,0014% entre as regiões, enquanto o grupo do leste asiático também teve um valor de diversidade abaixo da média para a região exômica.

Tabela 9 – Valores médios de Diversidade Nucleotídica (em porcentagem) para as regiões gênicas do *MC1R*, separadas pelos grupos populacionais analisados (AFR – África, AMR – América, EAS – leste da Ásia, EUR – Europa, SAS – sul da Ásia)

	5'URR	5'UTR	Codificante	3'UTR	Total Estendida	Total (mRNA)
AFR	0,2313%	0,1481%	0,0940%	0,1972%	0,1867%	0,1436%
AMR	0,1947%	0,1938%	0,0908%	0,0774%	0,1636%	0,1333%
EAS	0,1887%	0,1314%	0,1449%	0,1037%	0,1582%	0,1288%
EUR	0,1855%	0,1995%	0,1033%	0,0582%	0,1598%	0,1350%
SAS	0,2350%	0,2190%	0,0617%	0,0750%	0,1842%	0,1351%
Média	0,2070%	0,1784%	0,0990%	0,1023%	0,1705%	0,1351%

Tabela 10 – Valores médios de Diversidade Nucleotídica (em porcentagem) para as regiões gênicas do **SLC24A5**, separadas pelos grupos populacionais analisados (AFR – África, AMR – América, EAS – leste da Ásia, EUR – Europa, SAS – sul da Ásia)

	5'URR	5'UTR	Codificante (Éxons e Introns)	Codificante (Somente Éxons)	Total Estendida	Total (mRNA)
AFR	0,0668%	0,0020%	0,0671%	0,0129%	0,0668%	0,0668%
AMR	0,0363%	0,0045%	0,0443%	0,0273%	0,0432%	0,0442%
EAS	0,0377%	0,0029%	0,0377%	0,0043%	0,0376%	0,0376%
EUR	0,0012%	0,0000%	0,0038%	0,0007%	0,0035%	0,0038%
SAS	0,0230%	0,0000%	0,0299%	0,0265%	0,0289%	0,0298%
Média	0,0330%	0,0019%	0,0366%	0,0143%	0,0360%	0,0364%

5.2.4 Rede de haplótipos

Foi observado um total de 91 haplótipos distintos na região 5'URR do gene *SLC24A5*, 04 haplótipos na região 5'UTR, 519 haplótipos da região codificadora (considerando-se éxons e introns), 41 haplótipos da região codificadora (considerando-se somente os éxons), 607 haplótipos da região total estendida e 522 haplótipos da região completa considerando-se somente o mRNA.

No caso do gene *MC1R* foram observados 227 haplótipos da região 5'URR, 56 da região 5'UTR, 101 da região codificadora, 31 da região 3'UTR, 452 haplótipos considerando-se a região total estendida e 190 da região total considerando-se somente o mRNA.

Na elaboração das redes de haplótipos só foram avaliados aqueles haplótipos que possuíam frequência maior ou igual a 0,5% dentro do grupo étnico em questão, uma vez que esse limiar excluiria a abundância de haplótipos raros obtidos, os quais dificultariam a visualização e entendimento da *network*. Uma segunda etapa de redução de haplótipos, por meio da exclusão daqueles com cinco ou menos ocorrências, permitiu (em algumas situações de diversidade haplotípica elevada) que se observasse melhor os haplótipos ancestrais de cada grupo étnico e sua relação com os demais. Como a região total (tanto estendida, quanto do mRNA) do *SLC24A5* apresentou um número acentuado de haplótipos, mesmo após essas duas

etapas de filtragem, a dinâmica de interação entre os mesmos se tornou prejudicada, gerando redes de difícil interpretação. Sendo assim, foi optado por serem mantidas apenas as redes de haplótipos para as demais regiões gênicas separadas.

Em praticamente todas as análises realizadas para as regiões do gene *SLC24A5*, as populações africanas apresentavam uma maior variabilidade, refletida por uma prevalência relativa no número de haplótipos diversos, em geral apresentando uma grade proporção de haplótipos raros, distintos entre si por poucos passos mutacionais (Figuras 6 a 9 – para valores detalhados de frequência ver Tabelas S1 – S12). Cabe ressaltar que a região 5'UTR deste gene apresentou pouquíssima variabilidade em todas as populações analisadas (presença de um único haplótipo predominante, com frequência próxima a 100%), uma vez que, além de ser extremamente curta (73 pb), apenas três mutações foram detectadas nesse trecho em nossas análises, gerando três haplótipos de distribuição restrita a um único grupo populacional.

Com exceção da região 5'UTR, as populações africanas foram as que apresentaram a maior diversidade no gene *MC1R*, evidenciada pelo maior número de haplótipos distintos (muitos deles de ocorrência única), assim como ocorreu para o gene *SLC24A5* (Figuras 10 a 15 – para valores detalhados de frequência ver Tabela S1). Para a região 5'UTR, as populações do sul da Ásia apresentaram um número de haplótipos maior (24 haplótipos distintos) até que as africanas (20 haplótipos distintos), sendo esta diferença causada principalmente pelo número aumentado de haplótipos raros exclusivos das populações SAS (17 haplótipos do total) (Figura 11, Tabelas S1 – S12).

O padrão de diversidade observado no *MC1R* foi semelhante ao do *SLC24A5*, com a prevalência de dois ou três haplótipos principais, compartilhados entre as populações, além do grupo africano continuar apresentando maior variabilidade de haplótipos, sendo a maioria deles raros, em grande parte das análises. Diferente, no entanto, do observado para o gene *SLC24A5*, as populações europeias não apresentaram uma variabilidade tão baixa, tendo valores muito próximos àqueles das populações asiáticas, americanas miscigenadas e até mesmo das africanas em algumas das regiões do *MC1R*.

Figura 6 - Rede de haplótipos gerada pelo programa Network 4.6 através do algoritmo *Median-Joining* contendo os haplótipos da região 5'URR do gene *SLC24A5* nas amostras do *1000Genomes Project*. Cada circunferência representa um haplótipo, com sua identificação à esquerda. O diâmetro de cada circunferência é proporcional à frequência do haplótipo na amostra populacional. As cores em cada circunferência representam a frequência relativa do haplótipo entre as cinco populações parentais. Os números em itálico representam os SNPs que diferenciam cada haplótipo.

Figura 7 - Rede de haplótipos gerada pelo programa Network 4.6 através do algoritmo *Median-Joining* contendo os haplótipos da região 5'UTR do gene *SLC24A5* nas amostras do *1000Genomes Project*. Cada circunferência representa um haplótipo, com sua identificação à esquerda. O diâmetro de cada circunferência é proporcional à frequência do haplótipo na amostra populacional. As cores em cada circunferência representam a frequência relativa do haplótipo entre as cinco populações parentais. Os números em itálico representam os SNPs que diferenciam cada haplótipo.

Figura 8 - Rede de haplótipos gerada pelo programa Network 4.6 através do algoritmo *Median-Joining* contendo os haplótipos da região Codificante (Éxons e Introns) do gene *SLC24A5* nas amostras do *1000Genomes Project*. Cada circunferência representa um haplótipo, com sua identificação à esquerda. O diâmetro de cada circunferência é proporcional à frequência do haplótipo na amostra populacional. As cores em cada circunferência representam a frequência relativa do haplótipo entre as cinco populações parentais.

Figura 10 - Rede de haplótipos gerada pelo programa Network 4.6 através do algoritmo *Median-Joining* contendo os haplótipos da região 5'URR do gene *MC1R* nas amostras do *1000Genomes Project*. Cada circunferência representa um haplótipo, com sua identificação à esquerda. O diâmetro de cada circunferência é proporcional à frequência do haplótipo na amostra populacional. As cores em cada circunferência representam a frequência relativa do haplótipo entre as cinco populações parentais. Os números em itálico representam os SNPs que diferenciam cada haplótipo.

Figura 12 - Rede de haplótipos gerada pelo programa Network 4.6 através do algoritmo *Median-Joining* contendo os haplótipos da região Codificante do gene *MC1R* nas amostras do *1000Genomes Project*. Cada circunferência representa um haplótipo, com sua identificação à esquerda. O diâmetro de cada circunferência é proporcional à frequência do haplótipo na amostra populacional. As cores em cada circunferência representam a frequência relativa do haplótipo entre as cinco populações parentais. Os números em itálico representam os SNPs que diferenciam cada haplótipo.

Figura 13 - Rede de haplótipos gerada pelo programa Network 4.6 através do algoritmo *Median-Joining* contendo os haplótipos da região 3'UTR do gene *MC1R* nas amostras do *1000Genomes Project*. Cada circunferência representa um haplótipo, com sua identificação à esquerda. O diâmetro de cada circunferência é proporcional à frequência do haplótipo na amostra populacional. As cores em cada circunferência representam a frequência relativa do haplótipo entre as cinco populações parentais. Os números em itálico representam os SNPs que diferenciam cada haplótipo.

Figura 14 - Rede de haplótipos gerada pelo programa Network 4.6 através do algoritmo *Median-Joining* contendo os haplótipos da região Total Estendida do gene *MC1R* nas amostras do *1000Genomes Project*. Cada circunferência representa um haplótipo, com sua identificação à esquerda. O diâmetro de cada circunferência é proporcional à frequência do haplótipo na amostra populacional. As cores em cada circunferência representam a frequência relativa do haplótipo entre as cinco populações parentais. Os números em itálico representam os SNPs que diferenciam cada haplótipo.

Figura 15 - Rede de haplótipos gerada pelo programa Network 4.6 através do algoritmo *Median-Joining* contendo os haplótipos da região Total (mRNA) do gene *MC1R* nas amostras do *1000Genomes Project*. Cada circunferência representa um haplótipo, com sua identificação à esquerda. O diâmetro de cada circunferência é proporcional à frequência do haplótipo na amostra populacional. As cores em cada circunferência representam a frequência relativa do haplótipo entre as cinco populações parentais. Os números em itálico representam os SNPs que diferenciam cada haplótipo.

5.2.5 Desequilíbrio de ligação

Considerando-se os SNPs inicialmente presentes nos dados analisados, observamos que grande parte deles foi eliminada das análises no Haploview por não satisfazerem os critérios de frequência alélica (0,1%) e probabilidade de aderência ao EHW (0,001) mínimos estabelecidos como padrão pelo software.

Considerando-se os *loci* analisados na região estendida do gene SLC24A5 e os critérios do Haploview mencionados acima, o grupo AFR ficou com 125 *loci* (19 na região 5'URR, nenhum na 5'UTR e 106 na codificante) agrupados em um único bloco de aproximadamente 19kb (Figura 16), o grupo AMR permaneceu com 156 *loci* (19 na região 5'URR, nenhum na 5'UTR e 137 na codificante) agrupado em um único bloco de aproximadamente 20kb (Figura 17); o grupo EAS continuou com 80 *loci* (16 na região 5'URR, nenhum na 5'UTR e 64 na codificante) agrupados em dois blocos, sendo um de 1kb englobando parte da região 5'URR e outro de 21kb englobando o fim da região 5'URR até quase a totalidade da codificante (Figura 18); o grupo EUR continuou com 23 *loci* (3 na região 5'URR, nenhum na 5'UTR e 20 na codificante) sem formar nenhum bloco em desequilíbrio (Figura 19) e finalmente o grupo SAS permaneceu com 140 *loci* (19 na região 5'URR, nenhum na 5'UTR e 121 na codificante) agrupado em um único bloco de aproximadamente 21kb (Figura 20).

Para a análise da região estendida do MC1R, utilizando-se os mesmos parâmetros, o grupo AFR ficou com 76 *loci* (32 na região 5'URR, 17 na 5'UTR, 19 na codificante e 8 na região 3'UTR) agrupados em 4 blocos (um bloco de aproximadamente 3kb, um bloco de aproximadamente 1kb e dois menores que 1kb) (Figura 21), o grupo AMR permaneceu com 104 *loci* (49 na região 5'URR, 18 na 5'UTR, 27 na codificante e 10 na região 3'UTR) agrupados em 2 blocos (um bloco de aproximadamente 3kb e um bloco menor que 1kb) (Figura 22), o grupo EAS continuou com 38 *loci* (18 na região 5'URR, 9 na 5'UTR, 9 na codificante e 2 na região 3'UTR) agrupados em um único bloco de aproximadamente 5kb (Figura 23), o grupo EUR continuou com 50 *loci* (23 na região 5'URR, 12 na 5'UTR, 13 na codificante e 2 na região 3'UTR) agrupados em um único bloco de aproximadamente 5kb (Figura 24) e finalmente o grupo SAS permaneceu com 98 *loci* (47 na região 5'URR, 23 na 5'UTR, 23 na codificante e 5 na região 3'UTR) agrupados em um único bloco de aproximadamente 4kb (Figura 25).

Figura 16 – Gráfico gerado pelo Haploview 4.2 mostrando o desequilíbrio de ligação entre pares de *loci* selecionados da região estendida do gene *SLC24A5* nas amostras do grupo africano do *1000Genomes Project*. A cor vermelha dos quadrados indica valores de D' próximos a 1 (valores menores que 1 estão indicados) e LOD maior ou igual a 2, branco indica valores de D' próximos a 1 e LOD menor que 2. Acima do gráfico estão representadas as regiões gênicas correspondentes a cada *locus*.

Figura 17 – Gráfico gerado pelo Haploview 4.2 mostrando o desequilíbrio de ligação entre pares de *loci* selecionados da região estendida do gene SLC24A5 nas amostras do grupo Americano do 1000Genomes Project. A cor vermelha dos quadrados indica valores de D' próximos a 1 (valores menores que 1 estão indicados) e LOD maior ou igual a 2, branco indica valores de D' próximos a 0 e azul indica valores de D' próximos a 1 e LOD menor que 2. Acima do gráfico estão representadas as regiões gênicas correspondentes a cada *locus*.

Figura 18 – Gráfico gerado pelo Haploview 4.2 mostrando o desequilíbrio de ligação entre pares de *loci* selecionados da região estendida do gene SLC24A5 nas amostras do grupo leste asiático do *1000Genomes Project*. A cor vermelha dos quadrados indica valores de D' próximos a 1 (valores menores que 1 estão indicados) e LOD maior ou igual a 2, branco indica valores de D' próximos de 0 e azul indica valores de D' próximos a 1 e LOD menor que 2. Acima do gráfico estão representadas as regiões gênicas correspondentes a cada *locus*.

Figura 19 – Gráfico gerado pelo Haploview 4.2 mostrando o desequilíbrio de ligação entre pares de *loci* selecionados da região estendida do gene SLC24A5 nas amostras do grupo europeu do 1000Genomes Project. A cor vermelha dos quadrados indica valores de D' próximos a 1 (valores menores que 1 estão indicados) e LOD maior ou igual a 2, branco indica valores de D' próximos de 0 e azul indica valores de D' próximos a 1 e LOD menor que 2. Acima do gráfico estão representadas as regiões gênicas correspondentes a cada *locus*.

Figura 20 – Gráfico gerado pelo Haploview 4.2 mostrando o desequilíbrio de ligação entre pares de *loci* selecionados da região estendida do gene SLC24A5 nas amostras do grupo sul asiático do 1000Genomes Project. A cor vermelha dos quadrados indica valores de D' próximos a 1 (valores menores que 1 estão indicados) e LOD maior ou igual a 2, branco indica valores de D' próximos a 0 e azul indica valores de D' próximos a 1 e LOD menor que 2. Acima do gráfico estão representadas as regiões gênicas correspondentes a cada *locus*.

Figura 21 – Gráfico gerado pelo Haploview 4.2 mostrando o desequilíbrio de ligação entre pares de *loci* selecionados da região estendida do gene *MC1R* nas amostras do grupo africano do 1000Genomes Project. A cor vermelha dos quadrados indica valores de D' próximos a 1 (valores menores que 1 estão indicados) e LOD maior ou igual a 2, branco indica valores de D' próximos a 0 e azul indica valores de D' próximos a 1 e LOD menor que 2. Acima do gráfico estão representadas as regiões gênicas correspondentes a cada *locus*.

Figura 22 – Gráfico gerado pelo Haploview 4.2 mostrando o desequilíbrio de ligação entre pares de *loci* selecionados da região estendida do gene *MC1R* nas amostras do grupo Americano do *1000Genomes Project*. A cor vermelha dos quadrados indica valores de D' próximos a 1 (valores menores que 1 estão indicados) e LOD maior ou igual a 2, branco indica valores de D' próximos de 0 e azul indica valores de D' próximos a 1 e LOD menor que 2. Acima do gráfico estão representadas as regiões gênicas correspondentes a cada *locus*.

Figura 23 – Gráfico gerado pelo Haploview 4.2 mostrando o desequilíbrio de ligação entre pares de *loci* selecionados da região estendida do gene *MC1R* nas amostras do grupo leste asiático do *1000Genomes Project*. A cor vermelha dos quadrados indica valores de D' próximos a 1 (valores menores que 1 estão indicados) e LOD maior ou igual a 2, branco indica valores de D' próximos de 0 e azul indica valores de D' próximos a 1 e LOD menor que 2. Acima do gráfico estão representadas as regiões gênicas correspondentes a cada *locus*..

Figura 24 – Gráfico gerado pelo Haploview 4.2 mostrando o desequilíbrio de ligação entre pares de *loci* selecionados da região estendida do gene *MC1R* nas amostras do grupo europeu do *1000Genomes Project*. A cor vermelha dos quadrados indica valores de D' próximos a 1 (valores menores que 1 estão indicados) e LOD maior ou igual a 2, branco indica valores de D' próximos a 0 e azul indica valores de D' próximos a 1 e LOD menor que 2. Acima do gráfico estão representadas as regiões gênicas correspondentes a cada *locus*.

Figura 25 – Gráfico gerado pelo Haploview 4.2 mostrando o desequilíbrio de ligação entre pares de *loci* selecionados da região estendida do gene *MC1R* nas amostras do grupo sul asiático do *1000Genomes Project*. A cor vermelha dos quadrados indica valores de D' próximos a 1 (valores menores que 1 estão indicados) e LOD maior ou igual a 2, branco indica valores de D' próximos a 0 e azul indica valores de D' próximos a 1 e LOD menor que 2. Acima do gráfico estão representadas as regiões gênicas correspondentes a cada *locus*.

5.2.6 AMOVA

Os valores obtidos para as análises de AMOVA considerando as regiões gênicas do *MC1R* e do *SLC24A5* mostraram, como esperado, que a maior proporção da variância ocorre entre indivíduos de uma mesma população (Φ_{ST}). No entanto, observou-se que a proporção da variância entre os grupos analisados (Φ_{CT}), também foi significativamente elevada (Tabela 6).

Considerando a variância dentro de populações (Φ_{ST}), os maiores valores observados para as análises foram, para o *MC1R*, aqueles referentes à região codificante do gene ($\Phi_{ST} = 0,27669$) e, para o *SLC24A5*, da região codificante contendo somente éxons do gene ($\Phi_{ST} = 0,74796$). No entanto, interessante, essa mesma região do *SLC24A5* apresentou valores altos de estatística inter-grupos (Φ_{CT}), consistindo na única situação em que a maior proporção da variância se localizasse entre os grupos (72,01% da variância).

Dentre todas análises realizadas, a única com valores de *P* não significativos foi referente à região 5'UTR do gene *SLC24A5*.

Tabela 6 - Análise de Variância Molecular (AMOVA) gerada à partir das diferentes regiões gênicas do *MC1R* e do *SLC24A5*, separadas em função da origem da variação, comparando-se as populações dos grupos África, Europa, leste da Ásia e sul da Ásia: entre grupos (Φ_{CT}), entre populações de um mesmo grupo (Φ_{SC}) ou dentro de populações (Φ_{ST}). Estão representados os valores das estatísticas Φ , a porcentagem da variância e a significância (P), com os valores significativos ($P < 0,05$) em negrito.

Gene	<i>MC1R</i>			<i>SLC24A5</i>		
	Estatística Φ	% Total	valor P	Estatística Φ	% Total	valor P
Região Gênica	5'URR			5'URR		
Entre os Grupos (Φ_{CT})	0,10640	10,64	0,0000	0,20043	20,04	0,0000
Entre as populações de um grupo (Φ_{SC})	0,01053	00,94	0,0000	0,01773	01,42	0,0000
Dentro das populações (Φ_{ST})	0,11581	88,42	0,0000	0,21460	78,54	0,0000
Região Gênica	5'UTR			5'UTR		
Entre os Grupos (Φ_{CT})	0,14665	14,66	0,0000	-0,00020	-0,02	0,83382
Entre as populações de um grupo (Φ_{SC})	0,01417	01,21	0,0000	-0,00008	-0,01	0,68524
Dentro das populações (Φ_{ST})	0,15874	84,13	0,0000	-0,00028	100,03	0,69599
Região Gênica	Codificante			Codificante (Éxons e Introns)		
Entre os Grupos (Φ_{CT})	0,26052	26,05	0,0000	0,19400	19,40	0,0000
Entre as populações de um grupo (Φ_{SC})	0,02187	01,62	0,0000	0,01298	01,05	0,0000
Dentro das populações (Φ_{ST})	0,27669	72,33	0,0000	0,20446	79,55	0,0000
Região Gênica	3'UTR			Codificante (Só Éxons)		
Entre os Grupos (Φ_{CT})	0,08285	08,28	0,0000	0,72013	72,01	0,0000
Entre as populações de um grupo (Φ_{SC})	0,02463	02,26	0,0000	0,09942	02,78	0,0000
Dentro das populações (Φ_{ST})	0,10544	89,46	0,0000	0,74796	25,20	0,0000
Região Gênica	Total Estendida			Total Estendida		
Entre os Grupos (Φ_{CT})	0,02756	02,76	0,0000	0,18089	18,09	0,0000
Entre as populações de um grupo (Φ_{SC})	0,01221	01,19	0,0000	0,01281	01,05	0,0000
Dentro das populações (Φ_{ST})	0,03944	96,06	0,0000	0,19138	80,86	0,0000
Região Gênica	Total (somente mRNA)			Total (somente mRNA)		
Entre os Grupos (Φ_{CT})	0,05204	05,20	0,0000	0,19395	19,39	0,0000
Entre as populações de um grupo (Φ_{SC})	0,02000	01,90	0,0000	0,01298	01,05	0,0000
Dentro das populações (Φ_{ST})	0,07099	92,90	0,0000	0,20441	79,56	0,0000

5.2.7 F_{ST} e Diferenciação Populacional

A princípio, valores de F_{ST} entre 0,00 e 0,05 indicam baixa diferenciação, 0,05 a 0,15 indicam diferenciação moderada e valores maiores que 0,15 indicam níveis altos de diferenciação (Hartl e Clark, 2007).

Observa-se que o padrão de F_{ST} dos genes *SLC24A5* e *MC1R* apresenta valores baixos com alta significância estatística, quando observa-se a região estendida desses genes (e na maioria das regiões individualizadas) para os diversos grupos populacionais. Alguns poucos valores mais elevados são mostrados quando as populações do leste asiático são comparadas com as africanas e europeias na região codificante do gene *MC1R* (Tabelas 7 e 8) e também quando o grupo africano é comparado ao europeu em todas as regiões do gene *SLC24A5*, principalmente na região codificante contendo somente éxons (Tabelas 9 e 10).

Os valores do teste exato de diferenciação populacional baseado em frequências haplotípicas foram significativos ($p < 0,05$) na maioria das comparações entre os grupos nas regiões avaliadas do *MC1R* e do *SLC24A5* (Tabelas 7 a 10).

Considerando-se as regiões gênicas do *MC1R* e do *SLC24A5* isoladamente, observa-se que os maiores valores de F_{ST} encontram-se nas comparações para a região codificante e codificante contendo somente éxons, respectivamente. Os valores de F_{ST} com maior significância para o *SLC24A5* ($p \leq 0,0023$) foram observados nas comparações da população europeia com os demais grupos (Tabelas 9 e 10). Para o *MC1R*, os valores significância foram altos para todas as comparações entre os grupos ($p \leq 0,0453$) (Tabelas 7 e 8).

No entanto, alguns resultados significativos foram encontrados mesmo para diferenciação entre populações de um mesmo grupo. Ao analisarmos a comparação entre as populações que compõe o grupo Americano, observamos que praticamente todos os valores de p do F_{ST} e do teste exato foram significativos, tanto para o *MC1R* quanto para o *SLC24A5*, o que é esperado em virtude de se tratarem de populações de padrões de ancestralidade altamente variáveis. Esse mesmo padrão de alto F_{ST} intragrupo foi observado quando as populações do sul da Ásia foram comparadas entre si, com valores próximos inclusive àqueles do grupo Americano para o *SLC24A5*.

Observamos também um padrão de alto F_{ST} significativo (inclusive para o teste exato) na comparação entre os dois grupos asiáticos (Sul e Leste), sendo que, para a região codificante do *MC1R*, esse foi o maior valor de F_{ST} obtido, da mesma maneira que a região codificante (contendo somente éxons) do *SLC24A5* apresentou valores altos na comparação destes grupos. Se considerarmos ainda as regiões isoladamente, observa-se que as regiões Codificante, Total (estendida) e Total (mRNA) do gene *MC1R* também apresentaram esse padrão de diferenciação intragrupo.

A região 5'UTR do *SLC24A5* foi a única em todas as análises a não apresentar nenhum valor de F_{ST} significativo, o que se justifica pela quase total ausência de variabilidade.

Tabela 7 – Valores médios de F_{ST} , percentagem de valores p significativos e valor médio de p (tanto para o F_{ST} quanto para o teste Exato) para as regiões 5'URR, 5'UTR e Codificante do gene *MC1R* nas populações avaliadas pelo 1000 Genomes Project (AFR – África, AMR- América, EAS – leste da Ásia, EUR – Europa, SAS – sul da Ásia)

	F_{ST} médio					F_{ST} médio					F_{ST} médio					F_{ST} médio				
	% p <0,05	p Médio	% p <0,05 ^a	p Médio ^a	p Médio ^a	% p <0,05	p Médio	% p <0,05 ^a	p Médio ^a	p Médio ^a	% p <0,05	p Médio	% p <0,05 ^a	p Médio ^a	p Médio ^a	% p <0,05	p Médio	% p <0,05 ^a	p Médio ^a	p Médio ^a
5'URR	AFR					AMR					EAS					EUR				
AFR	0,0054	60,0%	0,1006	73,3%	0,0741															
AMR	0,0693	93,3%	0,0140	96,7%	0,0208	0,0483	90,0%	0,0056	90,0%	0,0252	0,0369	80,0%	0,0525	80,0%	0,0570	0,0020	0,0%	0,2771	40,0%	0,2420
EAS	0,1094	100,0%	0,0000	100,0%	0,0000	0,1237	100,0%	0,0000	100,0%	0,0000	0,2184	100,0%	0,0000	100,0%	0,0000	0,0000	0,0%	0,2771	40,0%	0,2420
EUR	0,1374	100,0%	0,0000	100,0%	0,0000	0,0687	96,0%	0,0075	100,0%	0,0001	0,2184	100,0%	0,0000	100,0%	0,0000	0,0000	0,0%	0,2771	40,0%	0,2420
SAS	0,0563	100,0%	0,0000	100,0%	0,0000	0,0305	100,0%	0,0043	100,0%	0,0000	0,1269	100,0%	0,0000	100,0%	0,0000	0,0431	100,0%	0,0000	100,0%	0,0000
5'UTR	AFR					AMR					EAS					EUR				
AFR	0,0095	73,3%	0,1070	73,3%	0,0775															
AMR	0,1054	83,3%	0,0638	90,0%	0,0453	0,0896	100,0%	0,0015	100,0%	0,0002	0,0442	70,0%	0,0996	70,0%	0,0953	0,0094	40,0%	0,1647	40,0%	0,1509
EAS	0,1419	100,0%	0,0000	100,0%	0,0000	0,1504	100,0%	0,0002	100,0%	0,0000	0,2845	100,0%	0,0000	100,0%	0,0000	0,0000	40,0%	0,1647	40,0%	0,1509
EUR	0,1916	100,0%	0,0000	100,0%	0,0000	0,1140	92,0%	0,0159	96,0%	0,0030	0,1181	100,0%	0,0000	100,0%	0,0000	0,0000	40,0%	0,1647	40,0%	0,1509
SAS	0,1040	100,0%	0,0000	100,0%	0,0000	0,0498	92,0%	0,0104	100,0%	0,0000	0,1181	100,0%	0,0000	100,0%	0,0000	0,1030	100,0%	0,0000	100,0%	0,0000
Codificante	AFR					AMR					EAS					EUR				
AFR	0,0231	40,0%	0,1938	86,7%	0,0230															
AMR	0,1654	83,3%	0,0423	96,7%	0,0209	0,1129	90,0%	0,0060	90,0%	0,0082	0,0399	70,0%	0,0924	80,0%	0,0471	0,0182	70,0%	0,1297	80,0%	0,0925
EAS	0,3816	100,0%	0,0000	100,0%	0,0000	0,2582	100,0%	0,0000	100,0%	0,0000	0,3329	100,0%	0,0000	100,0%	0,0000	0,0659	100,0%	0,0000	100,0%	0,0000
EUR	0,1628	100,0%	0,0000	100,0%	0,0000	0,0934	96,0%	0,0059	100,0%	0,0002	0,4458	100,0%	0,0000	100,0%	0,0000	0,0659	100,0%	0,0000	100,0%	0,0000
SAS	0,1357	100,0%	0,0000	100,0%	0,0000	0,1325	100,0%	0,0000	100,0%	0,0000	0,4458	100,0%	0,0000	100,0%	0,0000	-0,0003	0,0%	0,4468	50,0%	0,2059

a – valores para o teste exato

Tabela 8 – Valores médios de F_{ST} , porcentagem de valores p significativos e valor médio de p (tanto para o teste Exato) para as regiões 3'UTR, Total Estendida e Total (mRNA) do gene *MC1R* nas populações avaliadas pelo *1000 Genomes Project* (AFR – África, AMR- América, EAS – leste da Ásia, EUR – Europa, SAS – sul da Ásia)

	F_{ST} médio					F_{ST} médio					F_{ST} médio					F_{ST} médio				
	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05
3'UTR	AFR					AMR					EAS					EUR				
AFR	0,0158	33,3%	0,2783	66,7%	0,1596															
AMR	0,1325	83,3%	0,0633	86,7%	0,0229	0,0652	80,0%	0,0471	70,0%	0,0728										
EAS	0,0956	100,0%	0,0000	100,0%	0,0000	0,1193	96,0%	0,0042	100,0%	0,0000	0,0658	70,0%	0,0812	70,0%	0,0802					
EUR	0,1519	100,0%	0,0000	100,0%	0,0000	0,0387	68,0%	0,1185	80,0%	0,0998	0,1027	96,0%	0,0090	100,0%	0,0004	0,0189	40,0%	0,1878	40,0%	0,2401
SAS	0,0976	100,0%	0,0000	100,0%	0,0000	0,0410	100,0%	0,0055	100,0%	0,0000	0,0784	100,0%	0,0020	100,0%	0,0000	0,0301	96,0%	0,0104	100,0%	0,0000
																-0,0024	0,0%	0,6642	0,0%	0,5664
Total Estendida	AFR					AMR					EAS					EUR				
AFR	0,0040	73,3%	0,0410	100,0%	0,0052															
AMR	0,0319	100,0%	0,0011	96,7%	0,0038	0,0182	90,0%	0,0129	90,0%	0,0096										
EAS	0,0688	100,0%	0,0000	100,0%	0,0000	0,0423	100,0%	0,0002	100,0%	0,0000	0,0204	90,0%	0,0484	100,0%	0,0035					
EUR	0,0183	93,3%	0,0427	93,3%	0,0215	0,0197	88,0%	0,0099	100,0%	0,0028	0,0584	100,0%	0,0001	100,0%	0,0000	0,0141	70,0%	0,1330	80,0%	0,0387
SAS	0,0223	96,7%	0,0067	96,7%	0,0033	0,0183	84,0%	0,0608	92,0%	0,0128	0,0507	100,0%	0,0000	100,0%	0,0000	0,0180	100,0%	0,0007	100,0%	0,0000
																0,0124	70,0%	0,1710	70,0%	0,1857
Total mRNA	AFR					AMR					EAS					EUR				
AFR	0,0062	73,3%	0,0317	100,0%	0,0010															
AMR	0,0493	100,0%	0,0028	93,3%	0,0189	0,0349	100,0%	0,0033	90,0%	0,0074										
EAS	0,1187	100,0%	0,0000	100,0%	0,0000	0,0738	100,0%	0,0001	100,0%	0,0000	0,0256	90,0%	0,0396	100,0%	0,0034					
EUR	0,0338	93,3%	0,0304	96,7%	0,0172	0,0319	96,0%	0,0106	96,0%	0,0035	0,0996	100,0%	0,0000	100,0%	0,0000	0,0256	80,0%	0,0850	80,0%	0,0680
SAS	0,0401	96,7%	0,0032	96,7%	0,0018	0,0339	80,0%	0,0632	96,0%	0,0107	0,0912	100,0%	0,0000	100,0%	0,0000	0,0389	100,0%	0,0001	100,0%	0,0000
																0,0259	70,0%	0,1440	70,0%	0,1418

a – valores para o teste exato

Tabela 9 – Valores médios de F_{ST} , percentagem de valores p significativos e valor médio de p (tanto para o teste Exato) para as regiões 5'UTR, 5'UTR e Codificante (Éxons e Introns) do gene *SLC24A5* nas populações avaliadas pelo *1000 Genomes Project* (AFR – África, AMR- América, EAS – leste da Ásia, EUR – Europa, SAS – sul da Ásia)

5'UTR	F_{ST} médio					F_{ST} médio					F_{ST} médio					F_{ST} médio				
	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05
5'UTR	AFR					AMR					EAS					EUR				
	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05
AFR	0,0058	33,3%	0,2518	53,3%	0,1670															
AMR	0,1687	93,3%	0,0130	90,0%	0,0300	0,0680	70,0%	0,0616	70,0%	0,0937										
EAS	0,0842	100,0%	0,0000	100,0%	0,0000	0,1512	100,0%	0,0000	100,0%	0,0000	0,0033	20,0%	0,3048	30,0%	0,2402					
EUR	0,3838	100,0%	0,0000	100,0%	0,0000	0,1681	100,0%	0,0000	100,0%	0,0000	0,3675	100,0%	0,0000	100,0%	0,0000	-0,0021	0,0%	0,7359	0,0%	0,6203
SAS	0,1734	100,0%	0,0000	100,0%	0,0000	0,0728	100,0%	0,0009	100,0%	0,0000	0,1318	100,0%	0,0016	100,0%	0,0000	0,1530	80,0%	0,0372	84,0%	0,0273
																0,0776	80,0%	0,0372	100,0%	0,0087
5'UTR	AFR					AMR					EAS					EUR				
	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05
AFR	-0,0002	0,0%	0,9990	0,0%	1,0000															
AMR	0,0004	0,0%	0,8968	0,0%	0,8834	0,0011	0,0%	0,7681	0,0%	0,7665										
EAS	-0,0001	0,0%	0,9306	0,0%	0,9206	0,0005	0,0%	0,8581	0,0%	0,8436	0,0002	0,0%	0,7944	0,0%	0,7903					
EUR	-0,0001	0,0%	0,9990	0,0%	1,0000	0,0009	0,0%	0,8732	0,0%	0,8756	0,0001	0,0%	0,9155	0,0%	0,9159	0,0000	0,0%	0,9990	0,0%	1,0000
SAS	-0,0001	0,0%	0,9990	0,0%	1,0000	0,0008	0,0%	0,8752	0,0%	0,8769	0,0001	0,0%	0,9162	0,0%	0,9169	0,0000	0,0%	0,9990	0,0%	1,0000
																0,0000	0,0%	0,9990	0,0%	1,0000
Codificante (Éxon e Intron)	AFR					AMR					EAS					EUR				
	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05	% p <0,05
AFR	0,0043	60,0%	0,1022	66,7%	0,1126															
AMR	0,0967	86,7%	0,0405	90,0%	0,0280	0,0692	90,0%	0,0485	90,0%	0,0142										
EAS	0,0423	100,0%	0,0000	100,0%	0,0000	0,1219	100,0%	0,0000	100,0%	0,0000	-0,0001	0,0%	0,4820	40,0%	0,2362					
EUR	0,3164	100,0%	0,0000	100,0%	0,0000	0,1650	100,0%	0,0002	100,0%	0,0000	0,3668	100,0%	0,0000	100,0%	0,0000	0,0063	30,0%	0,1368	70,0%	0,0534
SAS	0,1684	100,0%	0,0000	100,0%	0,0000	0,0786	100,0%	0,0034	100,0%	0,0000	0,2035	100,0%	0,0000	100,0%	0,0000	0,0754	88,0%	0,0480	92,0%	0,0100
																0,0522	80,0%	0,0251	90,0%	0,0123

a – valores para o teste exato

Tabela 10 – Valores médios de F_{ST} , porcentagem de valores p significativos e valor médio de p (tanto para o teste Exato) para as regiões Codificante (somente éxons), Total Estendida e Total (mRNA) do gene SLC24A5 nas populações avaliadas pelo 1000 Genomes Project (AFR – África, AMR- América, EAS – leste da Ásia, EUR – Europa, SAS – sul da Ásia)

Codificante (Só Éxons)	AFR					AMR					EAS					EUR					SAS				
	F_{ST} médio	% p <0,05	p Médio	% p <0,05 ^a	p Médio ^a	F_{ST} médio	% p <0,05	p Médio	% p <0,05 ^a	p Médio ^a	F_{ST} médio	% p <0,05	p Médio	% p <0,05 ^a	p Médio ^a	F_{ST} médio	% p <0,05	p Médio	% p <0,05 ^a	p Médio ^a	F_{ST} médio	% p <0,05	p Médio	% p <0,05 ^a	p Médio ^a
AFR	0,0165	53,3%	0,2059	60,0%	0,1399																				
AMR	0,3611	96,7%	0,0038	96,7%	0,0071	0,2074	80,0%	0,0313	80,0%	0,0351															
EAS	0,0326	73,3%	0,1266	76,7%	0,0647	0,4609	100,0%	0,0000	100,0%	0,0000	0,0041	40,0%	0,2653	20,0%	0,2723										
EUR	0,9003	100,0%	0,0000	100,0%	0,0000	0,5125	100,0%	0,0000	100,0%	0,0000	0,9633	100,0%	0,0000	100,0%	0,0000	0,0012	0,0%	0,6008	10,0%	0,3511					
SAS	0,5449	100,0%	0,0000	100,0%	0,0000	0,1896	80,0%	0,0907	80,0%	0,1050	0,6366	100,0%	0,0000	100,0%	0,0000	0,3035	100,0%	0,0023	100,0%	0,0017	0,1532	90,0%	0,0286	90,0%	0,0205
Total Estendida																									
AFR	0,0037	60,0%	0,0930	73,3%	0,1321																				
AMR	0,0910	90,0%	0,0404	90,0%	0,0362	0,0651	90,0%	0,0565	90,0%	0,0183															
EAS	0,0373	100,0%	0,0000	100,0%	0,0000	0,1147	100,0%	0,0000	100,0%	0,0000	0,0002	0,0%	0,4229	60,0%	0,2098										
EUR	0,2996	100,0%	0,0000	100,0%	0,0000	0,1556	100,0%	0,0001	100,0%	0,0000	0,3477	100,0%	0,0000	100,0%	0,0000	0,0056	30,0%	0,1366	80,0%	0,0663					
SAS	0,1559	100,0%	0,0000	100,0%	0,0000	0,0742	100,0%	0,0021	100,0%	0,0000	0,1888	100,0%	0,0000	100,0%	0,0000	0,0742	88,0%	0,0510	92,0%	0,0096	0,0527	80,0%	0,0498	90,0%	0,0129
Total mRNA																									
AFR	0,0043	60,0%	0,0991	66,7%	0,1153																				
AMR	0,0967	90,0%	0,0415	93,3%	0,0296	0,0691	90,0%	0,0533	90,0%	0,0172															
EAS	0,0422	100,0%	0,0000	100,0%	0,0000	0,1217	100,0%	0,0000	100,0%	0,0000	-0,0001	0,0%	0,4888	40,0%	0,2212										
EUR	0,3164	100,0%	0,0000	100,0%	0,0000	0,1649	100,0%	0,0000	100,0%	0,0000	0,3666	100,0%	0,0000	100,0%	0,0000	0,0063	30,0%	0,1291	80,0%	0,0591					
SAS	0,1684	100,0%	0,0000	100,0%	0,0000	0,0785	100,0%	0,0033	100,0%	0,0000	0,2034	100,0%	0,0000	100,0%	0,0000	0,0754	88,0%	0,0502	96,0%	0,0103	0,0522	80,0%	0,0242	90,0%	0,0115

a – valores para o teste exato

5. 3 Correlação com a variação da radiação UV

Os valores obtidos para as análises de AMOVA considerando as regiões gênicas do *MC1R* e do *SLC24A5* em função da incidência UV, com base nos padrões demonstrado na Figura 3, mostraram, como os dados de AMOVA para os grupos continentais, que a maior parte da variância ocorre entre indivíduos de uma mesma população (Φ_{ST}), tanto na comparação entre as três zonas principais de UV (baixa, média e alta, Tabela 11), quanto na comparação entre as zonas extremas (baixa e alta incidência, Tabela 12). Dentre todas análises realizadas, as únicas com valor de *P* não significativo foram para a região 5'UTR do gene *SLC24A5* e para várias das comparações entre os grupos para o gene *MC1R*, nas duas abordagens comparativas (Tabelas 11 e 12).

No entanto, a forma de agrupamento em três zonas principais é claramente ineficiente, uma vez que apresenta maior variância entre populações de mesmo grupo do que entre grupos em todas as regiões de ambos os genes (Tabela 11) enquanto a comparação entre as zonas extremas apresentou-se ligeiramente mais eficiente, ao menos para o gene *MC1R* (Tabela 12). A diferença principal entre as duas abordagens utilizadas se refletiu em um aumento na diversidade entre as regiões de incidência (Φ_{CT}), quando as zonas de baixa e alta incidência UV eram comparadas entre si (Tabelas 11 e 12).

Tabela 11 - Análise de Variância Molecular (AMOVA) gerada à partir das diferentes regiões gênicas do *MC1R* e do *SLC24A5*, separadas em função da origem da variação, comparando-se as populações dos grupos Alta, Média e Baixa incidência UV: entre grupos (Φ_{CT}), entre populações de um mesmo grupo (Φ_{SC}) ou dentro de populações (Φ_{ST}). Estão representados os valores das estatísticas Φ , a porcentagem da variância e a significância (P), com os valores significativos ($P < 0,05$) em negrito.

Gene	<i>MC1R</i>			<i>SLC24A5</i>		
	Estatística Φ	% Total	valor P	Estatística Φ	% Total	valor P
Região Gênica	5'URR			5'URR		
Entre os Grupos (Φ_{CT})	0,0309	3,09	0,0293	0,1016	10,16	0,0088
Entre as populações de um grupo (Φ_{SC})	0,0815	7,9	0,0000	0,1310	11,77	0,0000
Dentro das populações (Φ_{ST})	0,1099	89,01	0,0000	0,2193	78,07	0,0000
Região Gênica	5'UTR			5'UTR		
Entre os Grupos (Φ_{CT})	0,054	5,41	0,0088	-0,0005	-0,04	0,7439
Entre as populações de um grupo (Φ_{SC})	0,107	10,11	0,0000	0,0000	0,00	0,6999
Dentro das populações (Φ_{ST})	0,155	84,48	0,0000	-0,0004	100,04	0,6383
Região Gênica	Codificante			Codificante (Éxons e Introns)		
Entre os Grupos (Φ_{CT})	0,0045	0,45	0,3568	0,0697	6,97	0,0284
Entre as populações de um grupo (Φ_{SC})	0,2388	23,77	0,0000	0,1391	12,94	0,0000
Dentro das populações (Φ_{ST})	0,2422	75,78	0,0000	0,1991	80,09	0,0000
Região Gênica	3'UTR			Codificante (Só Éxons)		
Entre os Grupos (Φ_{CT})	0,0469	4,69	0,0078	0,3330	33,30	0,0196
Entre as populações de um grupo (Φ_{SC})	0,0655	6,24	0,0000	0,6275	41,85	0,0000
Dentro das populações (Φ_{ST})	0,1093	89,07	0,0000	0,7516	24,84	0,0000
Região Gênica	Total Estendida			Total Estendida		
Entre os Grupos (Φ_{CT})	0,0019	0,19	0,2757	0,0663	6,63	0,0332
Entre as populações de um grupo (Φ_{SC})	0,0339	3,38	0,0000	0,1289	12,04	0,0000
Dentro das populações (Φ_{ST})	0,0357	96,43	0,0000	0,1867	81,33	0,0000
Região Gênica	Total (somente mRNA)			Total (somente mRNA)		
Entre os Grupos (Φ_{CT})	0,0038	0,38	0,2825	0,0697	6,97	0,0372
Entre as populações de um grupo (Φ_{SC})	0,0606	6,04	0,0000	0,1390	12,94	0,0000
Dentro das populações (Φ_{ST})	0,0642	93,58	0,0000	0,1991	80,10	0,0000

Tabela 12 - Análise de Variância Molecular (AMOVA) gerada à partir das diferentes regiões gênicas do *MC1R* e do *SLC24A5*, separadas em função da origem da variação, comparando-se as populações dos grupos Alta e Baixa incidência UV: entre grupos (Φ_{CT}), entre populações de um mesmo grupo (Φ_{SC}) ou dentro de populações (Φ_{ST}). Estão representados os valores das estatísticas Φ , a porcentagem da variância e a significância (P), com os valores significativos ($P < 0,05$) em negrito.

Gene	<i>MC1R</i>			<i>SLC24A5</i>		
	Estatística Φ	% Total	valor P	Estatística Φ	% Total	valor P
Região Gênica	5'UTR			5'UTR		
Entre os Grupos (Φ_{CT})	0,0732	7,32	0,0088	0,1895	18,95	0,0039
Entre as populações de um grupo (Φ_{SC})	0,0645	5,98	0,0000	0,1066	8,64	0,0000
Dentro das populações (Φ_{ST})	0,1330	86,70	0,0000	0,2759	72,41	0,0000
Região Gênica	5'UTR			5'UTR		
Entre os Grupos (Φ_{CT})	0,1145	11,45	0,0020	-0,0007	-0,07	0,8016
Entre as populações de um grupo (Φ_{SC})	0,0829	7,34	0,0000	-0,0001	-0,01	0,6686
Dentro das populações (Φ_{ST})	0,1879	81,21	0,0000	-0,0008	100,08	0,7556
Região Gênica	Codificante			Codificante (Éxons e Introns)		
Entre os Grupos (Φ_{CT})	0,0170	1,70	0,2121	0,1590	15,90	0,0127
Entre as populações de um grupo (Φ_{SC})	0,2137	21,00	0,0000	0,1031	8,67	0,0000
Dentro das populações (Φ_{ST})	0,2270	77,30	0,0000	0,2457	75,43	0,0000
Região Gênica	3'UTR			Codificante (Só Éxons)		
Entre os Grupos (Φ_{CT})	0,0640	6,40	0,0196	0,5957	59,56	0,0049
Entre as populações de um grupo (Φ_{SC})	0,0748	7,00	0,0000	0,5002	20,22	0,0000
Dentro das populações (Φ_{ST})	0,1340	86,60	0,0000	0,7979	20,21	0,0000
Região Gênica	Total Estendida			Total Estendida		
Entre os Grupos (Φ_{CT})	0,0014	0,14	0,3001	0,1522	15,22	0,0088
Entre as populações de um grupo (Φ_{SC})	0,0273	2,72	0,0000	0,0957	8,11	0,0000
Dentro das populações (Φ_{ST})	0,0286	97,14	0,0000	0,2334	76,66	0,0000
Região Gênica	Total (somente mRNA)			Total (somente mRNA)		
Entre os Grupos (Φ_{CT})	0,0033	0,33	0,2590	0,1589	15,89	0,0088
Entre as populações de um grupo (Φ_{SC})	0,0490	4,89	0,0000	0,1030	8,67	0,0000
Dentro das populações (Φ_{ST})	0,0522	94,78	0,0000	0,2456	75,44	0,0000

6-Discussão

6.1 Validação dos dados processados pelo 1000 Genomes Project

A maior parte dos 585 *mismatches* avaliados entre os conjuntos de dados foi ocasionada pelo elevado nível de *missing alleles* (cerca de 64%, Tabela 3). Esta alta taxa de incertezas foi causada pela baixa cobertura (abaixo de 7 ou 20 *reads*) observada entre os sítios polimórficos analisados, que levaram aos passos de imputação incluídos em nosso *pipeline*. Assim, essa alta taxa de discordância entre o arquivo VCF e os genótipos obtidos por nossos *scripts* é, provavelmente, um reflexo das diferenças relativas aos métodos de imputação utilizados em cada abordagem. No entanto, nesse momento não é possível afirmar empiricamente qual dos métodos de imputação é mais propenso a erros em tais situações de baixa cobertura extremas.

A imputação de genótipos pode ser extremamente útil, permitindo que dados faltantes obtidos de chips de baixa densidade possam ser preenchidos, trazendo redução dos custos (Jimenez-Montero et al., 2013), combinando conjuntos de dados gerados por diferentes plataformas (Khankhanian et al., 2015), ou mesmo prever características fenotípicas complexas (Felipe et al., 2014). Provou-se que genótipos imputados apresentam precisão semelhante àqueles obtidos a partir de painéis de SNPs de alta densidade, em casos em que uma população de referência adequada é usada (Weigel et al., 2010). No entanto, os benefícios da imputação dependem fortemente de sua precisão, que está diretamente relacionada à estrutura populacional das amostras utilizadas e à composição do gene alvo avaliado (Felipe et al., 2014).

Embora a baixa cobertura seja um problema recorrente nos sequenciamentos de nova geração para genomas completos, outra questão importante é o *read mapping bias*, já que regiões altamente polimórficas irão gerar *reads* compostas de muitos alelos alternativos. Algumas destas *reads* altamente diversas não serão capazes de se alinhar a nenhum local no genoma referência e serão descartadas durante a análise. Portanto, seria interessante dispor de diversos haplótipos de determinado alvo para servirem como referência em passos múltiplos de alinhamento, a fim de melhorar o *calling* de genótipos para estas regiões polimórficas (Brandt et al., 2015).

A maioria dos estudos que usam imputação de genótipos para dados de sequenciamento de genoma completo consideram somente modelos de predição lineares, como abordagens de *ridge regression*, *Bayesian LASSO* (*least absolute selection and shrinkage operator*) ou *GBLUP* (*genomic best linear unbiased prediction*), que são mais adequados para modelos de genética aditiva, mas não são tão precisos na detecção de efeitos não-aditivos, como dominância e epistasia (Weigel et al., 2010; Felipe et al., 2014).

Mesmo modelos diferentes de imputação podem apresentar viés em alguns casos específicos, como a alta heterogeneidade resultando em subestimação dos alelos derivados; a presença do alelo alternativo em janelas genéticas pequenas não abrangidas por um número suficiente de SNPs para que a imputação seja eficaz e alelos alternativos com frequências raras. Embora o processo de imputação possa gerar conjuntos de dados com maior poder estatístico e novas associações em estudos populacionais ou de caso-controle, esse viés intrínseco poderia gerar a deficiência de dados ou erros do tipo I, que poderiam contradizer os resultados destes bancos de dados, como a perda ou subestimativa de associações genotípicas (Khankhanian et al., 2015).

Portanto, deve ser realizada uma avaliação precisa da estrutura genética dos dados, a fim de escolher o método de imputação mais adequado e confiável para os genes ou regiões genômicas avaliados. Além da escolha adequada do método de imputação, um método de validação de regressão deve ser usado, a fim de examinar os resultados, em busca de artefatos e assegurar a sua validade (Hoffmann et al., 2015) ou então, outra metodologia validada deveria ser utilizada, tal como sequenciamento Sanger.

Finalmente, foi verificado que 33% dos *mismatches* ocorreram por conta de falhas do pipeline, que acabaram fazendo *calling* de genótipos errados, como a detecção de homozigotos como sendo heterozigotos e vice versa. Tais erros deverão ser analisados para que se avalie quais passos ou regras dos *scripts* estejam sendo aplicados de forma errada, resultando em tais discrepâncias.

Dados de sequenciamento de nova geração devem ser tratados com cuidado, para que informações mais precisas possam ser obtidas deles. Em nossa comparação, foi observada uma taxa muito baixa de incompatibilidades para o *calling* de genótipo do *MC1R* (0,13% do total). Como só foram encontrados marcadores bialélicos para a região do *MC1R*, seriam necessários outros estudos

semelhantes, que incluíssem genes e regiões com marcadores multialélicos. Apesar disso, nossa comparação sugere que tanto o arquivo VCF quanto o *pipeline* descrito podem ser usados para obtenção de *calling* de genótipos confiáveis e precisos em estudos de genes ou regiões cromossômicas da Fase 3 do *1000 Genomes Project*. Um problema que deve ser melhor avaliado está na discordância significativa nos procedimentos de inferência de haplótipos, que pode ter ocorrido tanto por uma subestimação dos haplótipos no arquivo VCF quanto uma superestimação produzida pelo fastPHASE. O maior problema do arquivo VCF seria a aparente exclusão de 11,44% dos sítios variantes, que poderiam eventualmente representar mutações raras associadas a novos fenótipos ou doenças.

Embora a baixa cobertura possa ser um problema em relação a este tipo de análise, os métodos de imputação podem ser úteis para preencher lacunas de dados, desde que um painel de referência preciso seja escolhido. Um aumento na cobertura também poderia melhorar a precisão de *calling* de genótipos, mas o desenvolvimento de novos métodos para a identificação e correção de viés pode ser a única solução em casos que apenas os dados de baixa cobertura estão disponíveis. Portanto, apesar das diferenças causadas pelo métodos de imputação e pelos erros gerados pelo *pipeline*, as análises mostram que os dados da Fase 3 são muito mais confiáveis que aqueles obtidos anteriormente pelo *1000 Genomes*, baseando-se nos passos de alinhamento, avaliação de mapeamento e *calling* de genótipos aqui explorados.

6.2 Aspectos populacionais e Evolutivos do *MC1R* e *SLC24A5*

6.2.1 Equilíbrio de Hardy-Weinberg

Todos os valores de heterozigose observada e esperada obtidos para os genes *MC1R* e *SLC24A5* foram muito próximos entre si para todos os grupos populacionais (Tabelas 4 e 5), demonstrando não haver, aparentemente, nenhum fator contribuindo para o desvio do equilíbrio. Contudo, seria esperado que tais genes, sabidamente influenciados pela seleção natural, apresentassem certos desvios do EHW, considerando-se a influência que a incidência UV teria desempenhado em sua evolução.

No entanto, esse aparente equilíbrio não significa que as populações e os genes em questão estejam totalmente livres da ação de fatores evolutivos. Diversas premissas do princípio de equilíbrio podem estar sendo quebradas simultaneamente, fazendo com que os fatores de desvio estejam anulando uns aos outros (Guo e Thompson, 1992). Dentre os possíveis desvios concomitantes, podemos mencionar alguns como a ocorrência de mistura étnica, acasalamentos preferenciais, deriva genética, seleção, endogamia, entre outros. Considerando-se os grupos populacionais analisados, podemos destacar facilmente a presença de acasalamentos preferenciais e endogamia em populações do sul da Ásia (discutidos mais a frente), além de mistura étnica nas populações das Américas, por conta de sua contribuição tri-híbrida (europeia, indígena e africana).

Foi observado para o *SLC24A5* que os valores de heterozigose do grupo EUR foram drasticamente menores que os outros grupos em todas situações, enquanto o grupo AFR apresentou os maiores valores em praticamente todas análises. Esse padrão de baixa heterozigose reflete a homogeneidade característica das populações europeias, enquanto as populações africanas mantêm altos valores de heterozigose, reflexo da maior diversidade de haplótipos (Figuras 6 a 9).

6.2.2 AMOVA

Os valores de AMOVA obtidos foram significativos para quase todas as análises das regiões gênicas do *MC1R* e do *SLC24A5*, mostrando que a maior proporção da diversidade haplotípica foi observada entre os indivíduos de uma mesma população (Φ_{ST}). A única exceção se encontra na região 5'UTR do *SLC24A5*, na qual todos os valores de P foram extremamente altos, o que pode ser explicado pelo fato desta ser uma região extremamente curta contendo apenas 3 *loci* variantes, resultando em valores de estatística Φ negativos.

Um estudo feito com estimativas de AMOVA em 52 populações das Américas, Oceania, Ásia, Europa e África em 377 marcadores microssatélites verificou que o padrão esperado para maioria das comparações entre populações é de se encontrar a maior parte da variação entre os indivíduos de uma mesma população. Outras abordagens, considerando-se o agrupamento destas populações mediante fatores epidemiológicos, refletiram esse mesmo padrão observado anteriormente de maior parte da variação estar dentro das populações (Rosenberg et al., 2002).

Apesar dos dados aqui obtidos confirmarem o esperado para a maioria das regiões gênicas, ou seja, a maior parte da diferenciação está dentro das populações que compõe um grupo, observou-se que os valores obtidos nas análises da região codificante contendo somente os éxons do *SLC24A5* apresentavam um padrão distinto, com a contribuição da diversidade entre os grupos (72,01%) superando a contribuição da diversidade intra-populacional (25,2%). Apesar dos resultados mostrarem, nesse caso, que a maior parte da diversidade está entre os grupos, ao observarmos os valores absolutos obtidos pelas estatística Φ , nota-se que o valor de Φ_{ST} também é alto (0,7496), como o esperado.

Esse valor inesperado de Φ_{CT} nesta região do gene *SLC24A5* pode ser resultante de alguns fatores seletivos atuantes neste trecho do genoma. Se observarmos a diversidade para esta região, veremos que há pouca variabilidade na região codificante, representada pelo baixo número de haplótipos (Figura 9). Foram gerados apenas 41 haplótipos para a região codificante (sem introns), sendo que 28 deles eram de ocorrência única (Tabelas S1 – S12). Entre os 13 haplótipos restantes, pode-se observar um padrão de separação entre o haplótipo 1, prevalente na África e no leste da Ásia, e o haplótipo 2, prevalente na Europa e no sul da Ásia

(Figura 9). Portanto, o predomínio de um haplótipo principal distinto em cada um desses dois conjuntos (haplótipo 1 para AFR e EAS; haplótipo 2 para EUR e SAS), seria a causa principal dos padrões de diferenciação entre grupos observados nos resultados da AMOVA (Tabela 6).

6.2.3 Rede de Haplótipos

A visualização das frequências haplotípicas e sua distribuição ao longo das populações através das redes de haplótipos permite notar que, para o gene *SLC24A5*, as populações europeias possuem menor número de haplótipos e diversidade menores quando comparada às africanas, por exemplo. Enquanto observa-se a ocorrência de um número reduzido de haplótipos exclusivamente europeus, nota-se que os grupos asiático e africano possuem uma maior diversidade, representada por uma maior quantidade de haplótipos de baixa frequência para diferentes regiões do gene (Figuras 1-4 e Tabelas S1 - S12).

Ao analisarmos a região codificante total (éxons e introns) do *SLC24A5*, observou-se que o haplótipo 02 (mais frequente em AFR), o haplótipo 88 (mais frequente em EAS) e o haplótipo 12 (mais frequente em EUR, SAS e AMR) diferem entre si por poucos passos mutacionais, sugerindo que o haplótipo 02 tenha sido o ancestral para o surgimento dos outros dois variantes na Europa e Ásia. Ao considerar-se somente os éxons da região codificante, vemos que os haplótipos mais frequentes (haplótipo 01 na África e leste da Ásia e haplótipo 02 na Europa, sul da Ásia e América) diferem entre si por uma única mutação (rs1462554). Essa observação somada à baixa variabilidade (apenas 41 haplótipos distintos, sendo 28 deles de ocorrência única) demonstra ser uma região altamente conservada.

Para o *MC1R*, a população africana novamente apresenta uma maior diversidade juntamente com a população Americana. No entanto, as populações europeias, diferente do que ocorreu no gene *SLC24A5*, passam a exibir uma diversidade um pouco maior, apresentando mais haplótipos de frequência intermediária, mas no geral possuem uma diversidade reduzida quando comparadas ao grupo africano, que continua apresentando um número elevado de haplótipos de baixa frequência (Figuras 5-10).

Já na região 5'UTR do *MC1R*, as populações do sul da Ásia apresentaram uma maior diversidade que as populações africanas, diversidade esta gerada pelo alto número de haplótipos raros exclusivos (17 haplótipos no total). Esse padrão de alta diversidade no grupo SAS é observado novamente na rede da região Total Estendida.

Ao observar-se a disposição dos haplótipos do gene *SLC24A5* e *MC1R*, nota-se a presença de diversos haplótipos pouco frequentes com prevalência daqueles de origem africana formando um grupo bem delimitado enquanto o restante dos haplótipos tem frequências mais altas com contribuições mais bem distribuídas entre os cinco grupos étnicos (Figuras 1-10). Esse arranjo evidencia a baixa variabilidade gênica das populações europeias e asiáticas, um provável resultado do efeito gargalo que essas populações sofreram com o processo de saída da África, somado ainda à possível atuação da seleção positiva nas regiões de menor incidência solar e consequente dificuldade na fixação de vitamina D (Hudjashov et al., 2013). Se considerarmos que, após o efeito gargalo, o haplótipo mais frequente possuía uma vantagem seletiva em relação aos outros, por exemplo, seria esperado que sua frequência aumentasse, tornando-se um dos haplótipos prevalentes nesse grupo, como observado nos dados.

6.2.4 F_{ST} e Diferenciação populacional

O Índice de Fixação (F_{ST}) é a redução média da heterozigose em determinada subpopulação (quando comparada à população total) devido à deriva genética entre todas subpopulações analisadas, sendo portanto um parâmetro de avaliação da diferenciação genética entre as mesmas (Hartl e Clark, 2007). Assim sendo, valores baixos de F_{ST} refletem a inexistência de diferenças entre os grupos avaliados.

A interpretação do F_{ST} , no entanto, depende de alguns aspectos como taxas de migração, tempo de separação das populações e eventos fundadores que as mesmas tenham sofrido. Assim, alguns aspectos evolutivos das populações analisadas podem ser elucidados, como divergências populacionais ocorridas, estruturação populacional e miscigenação. A predominância de valores baixos de F_{ST} é esperada nos casos de regiões genômicas sujeitas à seleção balanceadora ou direcional. Se observarmos o padrão obtido nos testes de neutralidade (homozigose

observada maior que a esperada), nota-se um indicativo da atuação da seleção natural direcional, tanto no *MC1R* quanto no *SLC24A5*.

Podemos considerar tal padrão como esperado, uma vez que o *MC1R* e o *SLC24A5* são genes que tiveram sua trajetória evolutiva potencialmente influenciada pela maior ou menor incidência de luminosidade solar em determinada região geográfica habitada por cada população (diretamente relacionados à outros fatores, como a biodisponibilidade de vitamina D e ácido fólico, além de proteção contra danos da radiação UV). Cabe ressaltar que esses não são os únicos fatores atuantes sobre os genes *MC1R* e *SLC24A5*, uma vez que cada gene sofre diversas pressões seletivas distintas (algumas delas inclusive sem qualquer relação com a radiação UV), refletidas em algumas diferenças nos padrões do teste de neutralidade entre ambos. Além do mais, muitos genes relacionados à variação de pigmentação têm uma vasta gama de funções relacionada a outros fenótipos que não sejam somente os de cor de olhos, pelo ou cabelos (Hider et al., 2013).

Observando-se as comparações na região codificante do *MC1R* e do *SLC24A5*, nota-se que os maiores valores de F_{ST} encontram-se nas comparações das regiões codificante e codificante contendo somente éxons, respectivamente (Tabelas 7 e 8). Esses valores de diferenciação moderada a alta poderiam ser reflexo de assinaturas de seleção positiva mais intensa na região codificante dos dois genes. O *MC1R* e o *SLC24A5* já foram descritos em diversos estudos de varredura do genoma para assinaturas de seleção (Mcevoy et al., 2006; Canfield et al., 2013; Hider et al., 2013), demonstrando que a seleção positiva teria moldado consideravelmente seus padrões de variação.

Os valores de F_{ST} observados para a região codificante (contendo somente éxons) do gene *SLC24A5* foram extremamente elevados e significativos na comparação das populações europeias com o restante, chegando a diferenciação de 0,9633 (Tabela 8). Pela definição da análise de F_{ST} , valores tão próximos de 1,0 seriam evidência de uma possível fixação de alelos, trazendo indícios de que as populações que compõe o grupo europeu teriam algum genótipo/haplótipo fixados. Podemos confirmar esses dados com a observação que, para a região codificante (contendo somente éxons) do *SLC24A5*, as populações europeias apresentam um único haplótipo predominante, o haplótipo 02 (Figura 9), que tem frequência de 99,7% nesse grupo (Tabelas S1 - S12).

Outro padrão similar de diferenciação elevada e extremamente significativa (tanto para o F_{ST} quanto para o teste exato) é observado ainda na região codificante do *SLC24A5* quando as populações do sul da Ásia são comparadas com as do leste da Ásia (0,6366) e com as da África (0,5449) (Tabela 7). Nota-se nesse caso que a diferenciação deve-se provavelmente por conta da altíssima frequência do haplótipo 01 no grupo africano (90,66%) e no grupo leste asiático (96,82%), enquanto esse haplótipo tem frequência de 30,8% no grupo do sul da Ásia (Figura 9 e Tabelas S1 - S12).

Podemos também analisar essa alta diferenciação na comparação entre os dois grupos asiáticos (Sul e Leste) ao ponderar que, apesar de ambos estarem localizados em um mesmo continente, suas composições populacionais diferem grandemente. Ao observarmos os valores de F_{ST} obtidos das comparação entre as populações componentes do grupo SAS para o gene *SLC245*, nota-se que esse grupo possui valores de F_{ST} relativamente altos, muito próximos inclusive daqueles do grupo AMR, sabidamente composto por populações heterogêneas. Enfatiza-se também nesse caso o padrão de alta diversidade observado na dinâmica das redes de haplótipos do gene *MC1R* para o grupo SAS (Figuras 11 e 14). Esses resultados coincidentes poderiam ser o reflexo do grupo SAS ser composto por diversas populações consideradas isoladas.

De fato, um estudo feito com populações do sul da Ásia mostrou a presença de subgrupos no subcontinente Indiano, principalmente por conta do sistema de castas, no qual casamentos preferenciais endogâmicos são estimulados, além da presença de grupos etno-linguísticos distintos no norte e sul (Moorjani et al., 2013). Análises de componente principal ainda mostraram que populações do norte da Índia foram consideradas mais heterogêneas do que os grupos do sul do país (Wong et al., 2014).

O grupo AMR apresentou valores significativos de diferenciação populacional quando comparados a outros grupos e, inclusive quando suas populações são comparadas entre si. Tal observação provavelmente se deve ao fato deste grupo ser altamente miscigenado, tanto pelo agrupamento de populações distintas que o compõe, quanto pelas proporções das contribuições ancestrais que cada uma destas populações apresenta.

Esse padrão de diferenciação, no entanto, não se apresenta tão forte quando o grupo AMR é comparado ao AFR, já que, por conta da população ASW (indivíduos

afro-americanos do sudoeste dos EUA) ter sido considerada pertencente ao grupo Americano. A porcentagem de valores significativos é reduzida nessas comparações entre AMR e AFR, uma vez que a alta contribuição africana dessa população acaba tornando esses dois grupos mais similares entre si do que o esperado. Considerando-se o histórico de colonização das Américas, pode-se ainda considerar que as demais populações do grupo Americano também tenham alta contribuição africana, por conta do grande fluxo gênico ocorrido durante o período de tráfico de escravos.

Norton e colaboradores (2007) obtiveram valores de F_{ST} para o *SLC24A5* (rs1426654) para populações africanas, asiáticas, Ameríndias e europeias com um padrão que confirma nossos dados obtidos: a população europeia apresenta altos valores de F_{ST} (variando de 0,389 a 0,957) ao ser comparada com os demais grupos.

6.2.5 Desequilíbrio de ligação

Sabe-se que a estrutura dos blocos de desequilíbrio de ligação está relacionada a diversos fatores da história evolutiva da população em questão, resultantes da combinação de efeitos como deriva, mutação e recombinação gênica (Gu et al., 2007). Assim, populações que tenham passado por eventos evolutivos semelhantes ou populações mais próximas geograficamente possuirão maiores semelhanças no padrão de desequilíbrio do que populações originárias de regiões/situações distintas. Porém, outros fatores evolutivos como mutação, seleção natural, deriva e recombinação, específicos de cada população, acabarão influenciando o desequilíbrio de ligação de maneira particular a cada uma delas (Ardlie et al., 2002).

Foi observado um padrão de alto desequilíbrio de ligação para o *SLC24A5*, com poucos blocos de tamanho longo para praticamente todos os grupos populacionais. Esse padrão similar de desequilíbrio foi visto entre as amostras dos grupos AFR, AMR, EAS e SAS, que apresentaram um bloco semelhante de 20kb em média, englobando as regiões 5'URR e praticamente toda a região codificante do gene *SLC24A5*. No entanto, não foi observada a formação de nenhum bloco em desequilíbrio para as populações do grupo.

Sabe-se que o processo de recombinação gera diversidade de haplótipos, fazendo com que não haja apenas um haplótipo dominante em determinada população. No entanto, em casos de varredura seletiva a seleção positiva para determinado alelo resultará também na seleção dos alelos fisicamente próximos, diminuindo a possibilidade de recombinação. Assim, o padrão de alto desequilíbrio observado no *SLC24A5*, mesmo nas populações africanas, pode ser considerado mais uma forte evidência de sweep seletivo recente.

Para o *MC1R*, no entanto, foi observado um padrão distinto, com blocos menores e em maior número. Foi observado um compartilhamento de um bloco de 3 – 4 kb que englobava parte das regiões 5'URR e 5'UTR entre os grupos AFR, AMR e SAS. Um outro bloco de tamanho aproximado de 1kb, englobando parte das regiões 5'UTR e codificante, foi compartilhado entre os grupos AFR e AMR, enquanto os grupos EUR e EAS apresentaram um mesmo padrão, com um bloco único de aproximadamente 5kb englobando o gene todo.

Sabe-se que o baixo desequilíbrio de ligação nas populações africanas é esperado, já que são populações mais antigas, além de possuírem um maior tamanho populacional efetivo, se considerarmos a teoria mais aceita de origem do homem na África. Esse acúmulo de eventos de recombinação ao longo do tempo seria o responsável pela desestruturação dos blocos, sendo refletindo no menor desequilíbrio, com blocos mais curtos, quando comparados à populações mais recentes evolutivamente (Mcevoy et al., 2011). Esse padrão pode ser observado para o grupo AFR no gene *MC1R*, que apresentou 4 blocos de tamanho variando entre menos de 1kb e 3kb (Figura 21).

Foi observado nas demais populações (EUR, EAS e SAS) para o *MC1R* a presença de um único bloco variando entre 4 – 5kb, demonstrando o alto desequilíbrio nesses grupos (Figuras 23 a 25). Populações relativamente mais recentes que as africanas tendem a apresentar valores de desequilíbrio mais baixos, já que um dos fatores dominantes nessas situações seria o efeito gargalo sofrido pelas mesmas no processo de saída da África para a Eurásia, responsável pela diminuição drástica no tamanho populacional e consequente diversidade que levou, portanto, à menores possibilidades de recombinação e surgimento de blocos maiores de desequilíbrio (Sawyer et al., 2005).

As populações do grupo AMR, no entanto, apresentaram dois blocos (um de 3kb e outro menor que 1kb), representando um padrão intermediário entre os

demais grupos (Figura 22). Apesar da alta contribuição africana em algumas populações do grupo AMR, sabe-se que a miscigenação pode elevar o padrão de desequilíbrio, por conta da recombinação entre os blocos existentes nas populações parentais (Gu et al., 2007).

6.3 Correlação com a variação da radiação UV

Foram realizadas dois tipos de AMOVA avaliando a correlação entre as zonas de incidência UV e a diversidade dos genes *MC1R* e *SLC24A5* nas populações do *1000 Genomes Project*. Em uma primeira análise, foram analisadas as três faixas principais de incidência UV entre si, enquanto na segunda abordagem somente as zonas extremas (alta e baixa) de incidência UV foram comparadas.

Para as duas abordagens, os valores de AMOVA obtidos foram significativos para a maior parte todas as análises das regiões gênicas do *MC1R* e do *SLC24A5*, mostrando que a maior parte da diversidade haplotípica foi observada entre os indivíduos de uma mesma população (Φ_{ST}), sem qualquer correlação com a localização destas populações.

Para a análise entre as três zonas UV, a região codificante (contendo somente éxons) do *SLC24A5* apresentou um alto valor de diversidade quando considerou-se as populações dentro de uma mesma zona de incidência UV (Φ_{SC}), demonstrando que a maior diversidade neste caso está contida dentro de cada faixa UV.

Somente na análise considerando as duas zonas extremas, foi observado que o valor de contribuição de Φ_{CT} foi o mais alto para a região codificante (somente éxons) do *SLC24A5*, revelando que a maior parte da diversidade estava na comparação entre as faixas de alta e baixa incidência UV.

O fato de nossa segunda abordagem para a AMOVA (considerando as regiões extremas) ter apresentado maiores valores na comparação entre grupos (Φ_{CT}), pode ser reflexo da maior discrepância ambiental entre os mesmos, potencialmente resultando em pressões seletivas distintas para essas duas zonas de incidência UV e, portanto, em maiores discrepâncias genéticas.

No entanto, ao compararmos os dados desta análise com a AMOVA realizada com os grupos populacionais originais (Tabela 6), observamos que, apesar da região codificante do *SLC24A5* parecer influenciada pela incidência UV na comparação entre as zonas extremas, esse mesmo padrão foi observado nessas análises prévias.

Nossos dados demonstraram que as populações europeias têm baixa variabilidade nesta região do *SLC24A5*. Quando o grupo europeu é dividido em função da incidência UV (GBR, CEU e FIN no grupo de baixa UV; IBS e TSI no grupo de média UV) observa-se que a diversidade dessas populações é levemente distinta. Enquanto as populações europeias de média intensidade UV apresenta 2 haplótipos, o grupo de baixa intensidade UV apresenta 4 haplótipos distintos. Essa relativa diversidade aumentada nas populações da zona de baixa incidência UV poderia ser explicada em função das hipóteses de *fitness* adaptativo da síntese de vitamina D em locais de baixa intensidade solar (Jablonski e Chaplin, 2010).

O grupo formado pelas populações em zona de alta incidência UV apresentou maior diversidade. Obviamente, parte dessa alta diversidade se dá por conta das populações do continente africano, mais antigas que todas as restantes, mas nota-se que as populações asiáticas que fazem parte do grupo de alta UV (BEB, GIH, ITU, STU, CDX, CHS, KHV) também têm uma diversidade relativamente alta, apresentando 13 haplótipos distintos, enquanto as populações africanas possuem 19 haplótipos. Assim, esse padrão nas populações asiáticas também poderia ser influenciado por pressões seletivas exercidas pela alta incidência UV, já que tratam-se de populações com diversidade quase tão alta quanto as africanas, no entanto, bem mais recentes.

Considerações finais

Os resultados aqui apresentados demonstraram alguns aspectos interessantes sobre a evolução de dois genes associados à variação da pigmentação humana.

Os resultados de AMOVA demonstraram que a maior parte da diversidade da região codificante do gene *SLC24A5* está na comparação entre os 4 grupos principais (AFR, EUR, EAS e SAS), com outras evidências apontando para um padrão independente para o grupo europeu, como sua heterozigose extremamente reduzida, sua baixa diversidade nucleotídica e de haplótipos, com a presença de um haplótipo único em sua *network*, além dos valores de diferenciação populacional mostrarem que suas populações são extremamente diferentes das restantes e homogêneas entre si. Além disso, o padrão de alto desequilíbrio em conjunto com os demais resultados corroboram o processo de varredura seletiva já descrito nestas populações para a região codificante do *SLC24A5*. Esse processo de *sweep* pode ter como origem a alta pressão seletiva sofrida pelas populações na saída da África.

É interessante observar que a única diferença entre os haplótipos 1 e 2 (Figura 9) é o polimorfismo Ala111Thr, extensamente estudado para o gene *SLC24A5*. Um trabalho de Beleza et al. (2013) demonstra que os alelos específicos de populações europeias e asiáticas do gene *SLC24A5* são oriundos de *sweeps* seletivos muito mais recentes (11-19 mil anos atrás) do que aqueles que geraram os alelos de outros genes de pigmentação (30 mil anos atrás), corroborando a composição da rede haplotípica pouco diversa nesta região quando comparada às outras. Acredita-se que tal padrão tenha sido ainda influenciado pelo aumento no tamanho populacional do grupo europeu, o qual teria diminuído a deriva e favorecido o aumento de frequência de mutações vantajosas.

Essa estimativa do *sweep* seletivo foi confirmada por Canfield e colaboradores (2013), que obtiveram uma data em torno de 12,4 mil anos para o polimorfismo Ala111Thr. Curiosamente, esses períodos estimados caem justamente dentro do último máximo glacial, período em que houve maior sazonalidade da incidência de luz (ou seja, períodos mais longos de alta e baixa luminosidade alternados). Presume-se que durante esse período houveram mudanças culturais nas populações humanas, como uso de roupas mais protetoras e busca de abrigos, que acabaram levando à uma menor exposição à radiação UV, e por consequência,

à maiores chances de deficiência de vitamina D, sendo estes períodos conhecidos como “invernos da vitamina D”. O aumento populacional após este período poderia ser outro fator importante relacionado ao início deste processo de varredura seletiva (Jablonski e Chaplin, 2010). Apesar desse padrão evidente do *sweep* seletivo, ainda não foi esclarecido onde esse alelo variante teria surgido, mas Canfield e colaboradores (2013) afirmam que esse alelo teria surgido fora da África e posteriormente ao seu surgimento teria sido reintroduzido no continente.

Outros aspectos sobre a evolução desses genes evidenciam que alguns polimorfismos do *MC1R* presentes em alta frequência nas populações do leste asiático estão em baixa frequência (ou mesmo ausentes) nas populações europeias e africanas, sugerindo a ocorrência de evolução convergente para a despigmentação na Europa e Ásia (Hider et al., 2013). Além disso, o fato do *sweep* seletivo ocorrido no *SLC24A5* ter ocorrido bem depois da divergência entre os europeus e asiáticos aumenta as chances de evolução convergente, ao diminuir o intervalo de tempo para que as variantes vantajosas se espalhassem pela Eurásia (Beleza et al., 2013). Assim, poderia se supor que a alta intensidade UV nessas regiões geográficas teria moldado uma forte pressão seletiva a ponto de causar a evolução convergente para estes genes de pigmentação.

7-Conclusão

O presente estudo comprovou que os dados resultantes da Fase 3 do *1000 Genomes Project* apresentam maiores confiabilidade, informatividade e precisão quando comparados aos dados obtidos pelo *pipeline* local e de dados de fases anteriores, demonstrando que os dados liberados na última versão podem ser usados para inferências populacionais.

Também foi confirmado através dos padrões de frequência haplotípica, desequilíbrio de ligação, dinâmica e diversidade da rede de haplótipos do *SLC24A5* as hipóteses existentes de *sweep* seletivo recente para este gene. Os padrões de baixa diversidade haplotípica associada ao alto nível de desequilíbrio de ligação aqui apresentados são fortes evidências da ocorrência de varredura seletiva.

Os altos valores de F_{ST} dos grupos americano (AMR) e do sul da Ásia (SAS) demonstram uma alta diversidade entre as populações que os compõe. Tais valores são potencialmente o reflexo da alta heterogeneidade destes grupos, possivelmente decorrente da miscigenação, no caso das populações americanas e do isolamento entre grupos, já descrito para diversas populações do sul da Ásia.

Finalmente, as análises realizadas não evidenciaram uma correlação significativa entre a diversidade dos genes *MC1R* e *SLC24A5* e a distribuição da incidência UV global. Apesar da hipótese de pressão seletiva gerada pela radiação UV ter sido demonstrada por outros trabalhos, o teste aqui realizado não foi suficiente para evidenciar uma provável correlação entre os fatores ambientais e a diversidade desses dois genes.

8-Referências Bibliográficas

AOKI, K. Sexual selection as a cause of human skin colour variation: Darwin's hypothesis revisited. **Ann Hum Biol**, v. 29, n. 6, p. 589-608, Nov-Dec 2002.

ARDLIE, K. G.;KRUGLYAK, L.;SEIELSTAD, M. Patterns of linkage disequilibrium in the human genome. **Nat Rev Genet**, v. 3, n. 4, p. 299-309, Apr 2002.

BANDELT, H. J.;FORSTER, P.;ROHL, A. Median-joining networks for inferring intraspecific phylogenies. **Mol Biol Evol**, v. 16, n. 1, p. 37-48, Jan 1999.

BARRETT, J. C.;FRY, B.;MALLER, J.;DALY, M. J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, v. 21, n. 2, p. 263-5, Jan 15 2005.

BELEZA, S.;SANTOS, A. M.;MCEVOY, B.;ALVES, I.;MARTINHO, C.;CAMERON, E.;SHRIVER, M. D.;PARRA, E. J.;ROCHA, J. The timing of pigmentation lightening in Europeans. **Mol Biol Evol**, v. 30, n. 1, p. 24-35, Jan 2013.

BRANDT, D. Y.;AGUIAR, V. R.;BITARELLO, B. D.;NUNES, K.;GOUDET, J.;MEYER, D. Mapping Bias Overestimates Reference Allele Frequencies at the HLA Genes in the 1000 Genomes Project Phase I Data. **G3**, v. 5, n. 5, p. 931-41, May 2015.

BRANICKI, W.;BRUDNIK, U.;WOJAS-PELC, A. Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. **Ann Hum Genet**, v. 73, n. 2, p. 160-70, Mar 2009.

CANFIELD, V. A.;BERG, A.;PECKINS, S.;WENTZEL, S. M.;ANG, K. C.;OPPENHEIMER, S.;CHENG, K. C. Molecular phylogeography of a human autosomal skin color locus under natural selection. **G3**, v. 3, n. 11, p. 2059-67, Nov 2013.

CARLETTI, G.;NERVO, G.;CATTIVELLI, L. Flavonoids and Melanins: a common strategy across two kingdoms. **Int J Biol Sci**, v. 10, n. 10, p. 1159-70, 2014.

CASTELLI, E. C.;RAMALHO, J.;PORTO, I. O.;LIMA, T. H.;FELICIO, L. P.;SABBAGH, A.;DONADI, E. A.;MENDES-JUNIOR, C. T. Insights into HLA-G Genetics Provided by Worldwide Haplotype Diversity. **Frontiers in immunology**, v. 5, p. 476, 2014.

CERQUEIRA, C. C.;HUNEMEIER, T.;GOMEZ-VALDES, J.;RAMALLO, V.;VOLASKO-KRAUSE, C. D.;BARBOSA, A. A.;VARGAS-PINILLA, P.;DORNELLES, R. C.;LONGO, D.;ROTHHAMMER, F.;BEDOYA, G.;CANIZALES-QUINTEROS, S.;ACUNA-ALONZO, V.;GALLO, C.;POLETTI, G.;GONZALEZ-JOSE, R.;SALZANO, F. M.;CALLEGARI-JACQUES, S. M.;SCHULER-FACCINI, L.;RUIZ-LINARES, A.;CATIRA BORTOLINI, M.;FOR, C. Implications of the admixture process in skin color molecular assessment. **PLoS One**, v. 9, n. 5, p. e96886, 2014.

CHELI, Y.;LUCIANI, F.;KHALED, M.;BEURET, L.;BILLE, K.;GOUNON, P.;ORTONNE, J. P.;BERTOLOTTO, C.;BALLOTTI, R. α MSH and Cyclic AMP

elevating agents control melanosome pH through a protein kinase A-independent mechanism. **J Biol Chem**, v. 284, n. 28, p. 18699-706, Jul 10 2009.

CLAES, P.;HILL, H.;SHRIVER, M. D. Toward DNA-based facial composites: preliminary results and validation. **Forensic Sci Int Genet**, v. 13, p. 208-16, Nov 2014.

DANECEK, P.;AUTON, A.;ABECASIS, G.;ALBERS, C. A.;BANKS, E.;DEPRISTO, M. A.;HANDSAKER, R. E.;LUNTER, G.;MARTH, G. T.;SHERRY, S. T.;MCVEAN, G.;DURBIN, R.;GENOMES PROJECT ANALYSIS, G. The variant call format and VCFtools. **Bioinformatics**, v. 27, n. 15, p. 2156-8, Aug 1 2011.

DE ARAUJO LIMA, F.;DE TOLEDO GONCALVES, F.;FRIDMAN, C. SLC24A5 and ASIP as phenotypic predictors in Brazilian population for forensic purposes. **Leg Med (Tokyo)**, v. 17, n. 4, p. 261-6, Jul 2015.

DUFFY, D. L.;ZHAO, Z. Z.;STURM, R. A.;HAYWARD, N. K.;MARTIN, N. G.;MONTGOMERY, G. W. Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. **J Invest Dermatol**, v. 130, n. 2, p. 520-8, Feb 2010.

EDWARDS, M.;BIGHAM, A.;TAN, J.;LI, S.;GOZDZIK, A.;ROSS, K.;JIN, L.;PARRA, E. J. Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. **PLoS Genet**, v. 6, n. 3, p. e1000867, Mar 2010.

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

EXCOFFIER, L.;LISCHER, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. **Mol Ecol Resour**, v. 10, n. 3, p. 564-7, May 2010.

EXCOFFIER, L.;SMOUSE, P. E.;QUATTRO, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. **Genetics**, v. 131, n. 2, p. 479-91, Jun 1992.

FELIPE, V. P.;OKUT, H.;GIANOLA, D.;SILVA, M. A.;ROSA, G. J. Effect of genotype imputation on genome-enabled prediction of complex traits: an empirical study with mice data. **BMC Genet**, v. 15, p. 149, 2014.

FRUDAKIS, T.;TERRAVAINEN, T.;THOMAS, M. Multilocus OCA2 genotypes specify human iris colors. **Hum Genet**, v. 122, n. 3-4, p. 311-26, Nov 2007.

FUJIMOTO, A.;KIMURA, R.;OHASHI, J.;OMI, K.;YULIWULANDARI, R.;BATUBARA, L.;MUSTOFA, M. S.;SAMAKKARN, U.;SETTHEETHAM-ISHIDA, W.;ISHIDA, T.;MORISHITA, Y.;FURUSAWA, T.;NAKAZAWA, M.;OHTSUKA, R.;TOKUNAGA, K. A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. **Hum Mol Genet**, v. 17, n. 6, p. 835-43, Mar 15 2008.

FUJIMOTO, A.;NISHIDA, N.;KIMURA, R.;MIYAGAWA, T.;YULIWULANDARI, R.;BATUBARA, L.;MUSTOFA, M. S.;SAMAKKARN, U.;SETTHEETHAM-ISHIDA, W.;ISHIDA, T.;MORISHITA, Y.;TSUNODA, T.;TOKUNAGA, K.;OHASHI, J. FGFR2 is associated with hair thickness in Asian populations. **J Hum Genet**, v. 54, n. 8, p. 461-5, Aug 2009.

GENOMES PROJECT, C.;ABECASIS, G. R.;AUTON, A.;BROOKS, L. D.;DEPRISTO, M. A.;DURBIN, R. M.;HANDSAKER, R. E.;KANG, H. M.;MARTH, G. T.;MCVEAN, G. A. An integrated map of genetic variation from 1,092 human genomes. **Nature**, v. 491, n. 7422, p. 56-65, Nov 1 2012.

GERSTENBLITH, M. R.;GOLDSTEIN, A. M.;FARGNOLI, M. C.;PERIS, K.;LANDI, M. T. Comprehensive evaluation of allele frequency differences of MC1R variants across populations. **Hum Mutat**, v. 28, n. 5, p. 495-505, May 2007.

GHIORZO, P.;BONELLI, L.;PASTORINO, L.;BRUNO, W.;BARILE, M.;ANDREOTTI, V.;NASTI, S.;BATTISTUZZI, L.;GROSSO, M.;BIANCHI-SCARRA, G.;QUEIROLO, P. MC1R variation and melanoma risk in relation to host/clinical and environmental factors in CDKN2A positive and negative melanoma patients. **Exp Dermatol**, v. 21, n. 9, p. 718-20, Sep 2012.

GILISSEN, C.;HOISCHEN, A.;BRUNNER, H. G.;VELTMAN, J. A. Disease gene identification strategies for exome sequencing. **Eur J Hum Genet**, v. 20, n. 5, p. 490-7, May 2012.

GINGER, R. S.;ASKEW, S. E.;OGBORNE, R. M.;WILSON, S.;FERDINANDO, D.;DADD, T.;SMITH, A. M.;KAZI, S.;SZERENCSEI, R. T.;WINKFEIN, R. J.;SCHNETKAMP, P. P.;GREEN, M. R. SLC24A5 encodes a trans-Golgi network protein with potassium-dependent sodium-calcium exchange activity that regulates human epidermal melanogenesis. **J Biol Chem**, v. 283, n. 9, p. 5486-95, Feb 29 2008.

GRAF, J.;VOISEY, J.;HUGHES, I.;VAN DAAL, A. Promoter polymorphisms in the MATP (SLC45A2) gene are associated with normal human skin color variation. **Hum Mutat**, v. 28, n. 7, p. 710-7, Jul 2007.

GU, S.;PAKSTIS, A. J.;LI, H.;SPEED, W. C.;KIDD, J. R.;KIDD, K. K. Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. **Eur J Hum Genet**, v. 15, n. 3, p. 302-12, Mar 2007.

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

HAN, J.;KRAFT, P.;NAN, H.;GUO, Q.;CHEN, C.;QURESHI, A.;HANKINSON, S. E.;HU, F. B.;DUFFY, D. L.;ZHAO, Z. Z.;MARTIN, N. G.;MONTGOMERY, G. W.;HAYWARD, N. K.;THOMAS, G.;HOOVER, R. N.;CHANOCK, S.;HUNTER, D. J. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. **PLoS Genet**, v. 4, n. 5, p. e1000074, May 2008.

HARDING, R. M.;HEALY, E.;RAY, A. J.;ELLIS, N. S.;FLANAGAN, N.;TODD, C.;DIXON, C.;SAJANTILA, A.;JACKSON, I. J.;BIRCH-MACHIN, M. A.;REES, J. L. Evidence for variable selective pressures at MC1R. **Am J Hum Genet**, v. 66, n. 4, p. 1351-61, Apr 2000.

HARTL, D. L.;CLARK, A. G. **Principles of population genetics**. 4th. Sunderland, Mass.: Sinauer Associates, 2007. xv, 652 p. p.

HIDER, J. L.;GITTELMAN, R. M.;SHAH, T.;EDWARDS, M.;ROSENBLOOM, A.;AKEY, J. M.;PARRA, E. J. Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. **BMC Evol Biol**, v. 13, p. 150, 2013.

HOFFMANN, T. J.;SAKODA, L. C.;SHEN, L.;JORGENSEN, E.;HABEL, L. A.;LIU, J.;KVALE, M. N.;ASGARI, M. M.;BANDA, Y.;CORLEY, D.;KUSHI, L. H.;QUESENBERRY, C. P., JR.;SCHAEFER, C.;VAN DEN EEDEN, S. K.;RISCH, N.;WITTE, J. S. Imputation of the rare HOXB13 G84E mutation and cancer risk in a large population-based cohort. **PLoS Genet**, v. 11, n. 1, p. e1004930, Jan 2015.

HOWIE, B.;MARCHINI, J.;STEPHENS, M. Genotype imputation with thousands of genomes. **G3**, v. 1, n. 6, p. 457-70, Nov 2011.

HUDJASHOV, G.;VILLEMS, R.;KIVISILD, T. Global patterns of diversity and selection in human tyrosinase gene. **PLoS One**, v. 8, n. 9, p. e74307, 2013.

ITO, S.;WAKAMATSU, K. Human hair melanins: what we have learned and have not learned from mouse coat color pigmentation. **Pigment Cell Melanoma Res**, Sep 14 2010.

JABLONSKI, N. G.;CHAPLIN, G. The evolution of human skin coloration. **J Hum Evol**, v. 39, n. 1, p. 57-106, Jul 2000.

_____. Colloquium paper: human skin pigmentation as an adaptation to UV radiation. **Proc Natl Acad Sci U S A**, v. 107 Suppl 2, p. 8962-8, May 11 2010.

_____. The evolution of skin pigmentation and hair texture in people of African ancestry. **Dermatol Clin**, v. 32, n. 2, p. 113-21, Apr 2014.

JIMENEZ-MONTERO, J. A.;GIANOLA, D.;WEIGEL, K.;ALENDA, R.;GONZALEZ-RECIO, O. Assets of imputation to ultra-high density for productive and functional traits. **J Dairy Sci**, v. 96, n. 9, p. 6047-58, Sep 2013.

JOBLING, M. A.;HURLES, M.;TYLER-SMITH, C. **Human evolutionary genetics : origins, peoples & disease**. New York: Garland Science, 2004. xx, 523 p. p.

KAYSER, M.;LIU, F.;JANSSENS, A. C.;RIVADENEIRA, F.;LAO, O.;VAN DUIJN, K.;VERMEULEN, M.;ARP, P.;JHAMAI, M. M.;VAN IJCKEN, W. F.;DEN DUNNEN, J. T.;HEATH, S.;ZELENKA, D.;DESPRIET, D. D.;KLAVER, C. C.;VINGERLING, J. R.;DE JONG, P. T.;HOFMAN, A.;AULCHENKO, Y. S.;UITTERLINDEN, A. G.;OOSTRA, B. A.;VAN DUIJN, C. M. Three genome-wide association studies and a

linkage analysis identify HERC2 as a human iris color gene. **Am J Hum Genet**, v. 82, n. 2, p. 411-23, Feb 2008.

KHANKHANIAN, P.;DIN, L.;CAILLIER, S. J.;GOURRAUD, P. A.;BARANZINI, S. E. SNP imputation bias reduces effect size determination. **Frontiers in genetics**, v. 6, p. 30, 2015.

LAMASON, R. L.;MOHIDEEN, M. A.;MEST, J. R.;WONG, A. C.;NORTON, H. L.;AROS, M. C.;JURYNEC, M. J.;MAO, X.;HUMPHREVILLE, V. R.;HUMBERT, J. E.;SINHA, S.;MOORE, J. L.;JAGADEESWARAN, P.;ZHAO, W.;NING, G.;MAKALOWSKA, I.;MCKEIGUE, P. M.;O'DONNELL, D.;KITTLES, R.;PARRA, E. J.;MANGINI, N. J.;GRUNWALD, D. J.;SHRIVER, M. D.;CANFIELD, V. A.;CHENG, K. C. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. **Science**, v. 310, n. 5755, p. 1782-6, Dec 16 2005.

LE PAPE, E.;PASSERON, T.;GIUBELLINO, A.;VALENCIA, J. C.;WOLBER, R.;HEARING, V. J. Microarray analysis sheds light on the dedifferentiating role of agouti signal protein in murine melanocytes via the Mc1r. **Proc Natl Acad Sci U S A**, v. 106, n. 6, p. 1802-7, Feb 10 2009.

LI, H.;HANDSAKER, B.;WYSOKER, A.;FENNELL, T.;RUAN, J.;HOMER, N.;MARTH, G.;ABECASIS, G.;DURBIN, R.;GENOME PROJECT DATA PROCESSING, S. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-9, Aug 15 2009.

LIU, F.;WEN, B.;KAYSER, M. Colorful DNA polymorphisms in humans. **Semin Cell Dev Biol**, v. 24, n. 6-7, p. 562-75, Jun-Jul 2013.

MAKOVA, K.;NORTON, H. Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. **Peptides**, v. 26, n. 10, p. 1901-8, Oct 2005.

MARANO, L. A. **Análise de polimorfismos do gene MC1R associados a fenótipos humanos de pigmentação na população brasileira**. 2011. 130 f. (Dissertação de Mestrado) - Departamento de Genética - FMRP, Universidade de São Paulo

MATULLO, G.;DI GAETANO, C.;GUARRERA, S. Next generation sequencing and rare genetic variants: from human population studies to medical genetics. **Environ Mol Mutagen**, v. 54, n. 7, p. 518-32, Aug 2013.

MCEVOY, B.;BELEZA, S.;SHRIVER, M. D. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. **Hum Mol Genet**, v. 15 Spec No 2, p. R176-81, Oct 15 2006.

MCEVOY, B. P.;POWELL, J. E.;GODDARD, M. E.;VISSCHER, P. M. Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. **Genome Res**, v. 21, n. 6, p. 821-9, Jun 2011.

MENDES-JUNIOR, C. T.;FRACASSO, N. C. A.;ZANÃO, L. R.;ANDRADE, C. C. F.;SILVA, M. S.;MARANO, L. A.;WIEZEL, C. E. V.;DONADI, E. A.;SIMÕES, A. L.

Alleles from SNPs at the SLC24A5 and SLC45A2 genes are associated with the presence of freckles and eye, hair and skin pigmentation in Brazil., 60th ASHG Annual Meeting, 2010. Washington, DC. 2 - 6 November, 2010.

METZKER, M. L. Sequencing technologies - the next generation. **Nat Rev Genet**, v. 11, n. 1, p. 31-46, Jan 2010.

MOAN, J.;NIELSEN, K. P.;JUZENIENE, A. Immediate pigment darkening: its evolutionary roles may include protection against folate photosensitization. **FASEB J**, v. 26, n. 3, p. 971-5, Mar 2012.

MOORJANI, P.;THANGARAJ, K.;PATTERSON, N.;LIPSON, M.;LOH, P. R.;GOVINDARAJ, P.;BERGER, B.;REICH, D.;SINGH, L. Genetic evidence for recent population mixture in India. **Am J Hum Genet**, v. 93, n. 3, p. 422-38, Sep 5 2013.

MYLES, S.;SOMEL, M.;TANG, K.;KELSO, J.;STONEKING, M. Identifying genes underlying skin pigmentation differences among human populations. **Hum Genet**, v. 120, n. 5, p. 613-21, Jan 2007.

NADEAU, N. J.;MINVIELLE, F.;ITO, S.;INOUE-MURAYAMA, M.;GOURICHON, D.;FOLLETT, S. A.;BURKE, T.;MUNDY, N. I. Characterization of Japanese quail yellow as a genomic deletion upstream of the avian homolog of the mammalian ASIP (agouti) gene. **Genetics**, v. 178, n. 2, p. 777-86, Feb 2008.

NAGASAKI, H.;MOCHIZUKI, T.;KODAMA, Y.;SARUHASHI, S.;MORIZAKI, S.;SUGAWARA, H.;OHYANAGI, H.;KURATA, N.;OKUBO, K.;TAKAGI, T.;KAMINUMA, E.;NAKAMURA, Y. DDBJ read annotation pipeline: a cloud computing-based pipeline for high-throughput analysis of next-generation sequencing data. **DNA research : an international journal for rapid publication of reports on genes and genomes**, v. 20, n. 4, p. 383-90, Aug 2013.

NEI, M. **Molecular evolutionary genetics**. New York: Columbia University Press, 1987. x, 512 p. p.

NIELSEN, R.;PAUL, J. S.;ALBRECHTSEN, A.;SONG, Y. S. Genotype and SNP calling from next-generation sequencing data. **Nat Rev Genet**, v. 12, n. 6, p. 443-51, Jun 2011.

NORDLUND, L. A. Trends in smoking habits and lung cancer in Sweden. **Eur J Cancer Prev**, v. 7, n. 2, p. 109-16, Apr 1998.

NORTON, H. L.;KITTELS, R. A.;PARRA, E.;MCKEIGUE, P.;MAO, X.;CHENG, K.;CANFIELD, V. A.;BRADLEY, D. G.;MCEVOY, B.;SHRIVER, M. D. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. **Mol Biol Evol**, v. 24, n. 3, p. 710-22, Mar 2007.

NUWAIHYD, R.;REDLER, S.;HEILMANN, S.;DRICHEL, D.;WOLF, S.;BIRCH, P.;DOBSON, K.;LUTZ, G.;GIEHL, K. A.;KRUSE, R.;TAZI-AHNINI, R.;HANNEKEN, S.;BOHM, M.;MIESEL, A.;FISCHER, T.;WOLFF, H.;BECKER, T.;GARCIA-BARTELS, N.;BLUME-PEYTAVI, U.;NOTHEN, M. M.;MESSENGER, A. G.;BETZ, R.

C. Investigation of four novel male androgenetic alopecia susceptibility loci: no association with female pattern hair loss. **Arch Dermatol Res**, v. 306, n. 4, p. 413-8, May 2014.

PELLEGRINI, C.;FARGNOLI, M. C.;SUPPA, M.;PERIS, K. MC1R variants predisposing to concomitant primary cutaneous melanoma in a monozygotic twin pair. **BMC Med Genet**, v. 13, p. 81, 2012.

PEROLA, M.;OHMAN, M.;HIEKKALINNA, T.;LEPPAVUORI, J.;PAJUKANTA, P.;WESSMAN, M.;KOSKENVUO, M.;PALOTIE, A.;LANGE, K.;KAPRIO, J.;PELTONEN, L. Quantitative-trait-locus analysis of body-mass index and of stature, by combined analysis of genome scans of five Finnish study groups. **Am J Hum Genet**, v. 69, n. 1, p. 117-23, Jul 2001.

PULKER, H.;LAREU, M. V.;PHILLIPS, C.;CARRACEDO, A. Finding genes that underlie physical traits of forensic interest using genetic tools. **Forensic Sci Int-Gen**, v. 1, n. 2, p. 100-104, Jun 2007.

RAYMOND, M.;ROUSSET, F. An exact test for population differentiation. **Evolution**, v. 49, p. 1280-1283, 1995.

REES, J. L. Genetics of hair and skin color. **Annu Rev Genet**, v. 37, p. 67-90, 2003.

REYNOLDS, J.;WEIR, B. S.;COCKERHAM, C. C. Estimation of the coancestry coefficient: basis for a short-term genetic distance. **Genetics**, v. 105, n. 3, p. 767-79, Nov 1983.

ROSENBERG, N. A.;PRITCHARD, J. K.;WEBER, J. L.;CANN, H. M.;KIDD, K. K.;ZHIVOTOVSKY, L. A.;FELDMAN, M. W. Genetic structure of human populations. **Science**, v. 298, n. 5602, p. 2381-5, Dec 20 2002.

SAWYER, S. L.;MUKHERJEE, N.;PAKSTIS, A. J.;FEUK, L.;KIDD, J. R.;BROOKES, A. J.;KIDD, K. K. Linkage disequilibrium patterns vary substantially among populations. **Eur J Hum Genet**, v. 13, n. 5, p. 677-86, May 2005.

SCHEET, P.;STEPHENS, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. **Am J Hum Genet**, v. 78, n. 4, p. 629-44, Apr 2006.

SCHNEIDER, M. R.;SCHMIDT-ULLRICH, R.;PAUS, R. The hair follicle as a dynamic miniorgan. **Curr Biol**, v. 19, n. 3, p. R132-42, Feb 10 2009.

SLATKIN, M. An exact test for neutrality based on the Ewens sampling distribution. **Genet Res**, v. 64, n. 1, p. 71-4, Aug 1994.

_____. A correction to the exact test based on the Ewens sampling distribution. **Genet Res**, v. 68, n. 3, p. 259-60, Dec 1996.

SMITH, D. R.;SPAULDING, D. T.;GLENN, H. M.;FULLER, B. B. The relationship between Na(+)/H(+) exchanger expression and tyrosinase activity in human melanocytes. **Exp Cell Res**, v. 298, n. 2, p. 521-34, Aug 15 2004.

SMITH, R.;HEALY, E.;SIDDIQUI, S.;FLANAGAN, N.;STEIJLEN, P. M.;ROSDAHL, I.;JACQUES, J. P.;ROGERS, S.;TURNER, R.;JACKSON, I. J.;BIRCH-MACHIN, M. A.;REES, J. L. Melanocortin 1 receptor variants in an Irish population. **J Invest Dermatol**, v. 111, n. 1, p. 119-22, Jul 1998.

SOEJIMA, M.;KODA, Y. Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. **Int J Legal Med**, v. 121, n. 1, p. 36-9, Jan 2007.

STOKOWSKI, R. P.;PANT, P. V.;DADD, T.;FEREDAY, A.;HINDS, D. A.;JARMAN, C.;FILSELL, W.;GINGER, R. S.;GREEN, M. R.;VAN DER OUDERAA, F. J.;COX, D. R. A genomewide association study of skin pigmentation in a South Asian population. **Am J Hum Genet**, v. 81, n. 6, p. 1119-32, Dec 2007.

STURM, R. A. A golden age of human pigmentation genetics. **Trends Genet**, v. 22, n. 9, p. 464-8, Sep 2006.

_____. Molecular genetics of human pigmentation diversity. **Hum Mol Genet**, v. 18, n. R1, p. R9-17, Apr 15 2009.

STURM, R. A.;DUFFY, D. L.;ZHAO, Z. Z.;LEITE, F. P.;STARK, M. S.;HAYWARD, N. K.;MARTIN, N. G.;MONTGOMERY, G. W. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. **Am J Hum Genet**, v. 82, n. 2, p. 424-31, Feb 2008.

STURM, R. A.;FRUDAKIS, T. N. Eye colour: portals into pigmentation genes and ancestry. **Trends Genet**, v. 20, n. 8, p. 327-32, Aug 2004.

STURM, R. A.;TEASDALE, R. D.;BOX, N. F. Human pigmentation genes: identification, structure and consequences of polymorphic variation. **Gene**, v. 277, n. 1-2, p. 49-62, Oct 17 2001.

SULEM, P.;GUDBJARTSSON, D. F.;STACEY, S. N.;HELGASON, A.;RAFNAR, T.;MAGNUSSON, K. P.;MANOLESCU, A.;KARASON, A.;PALSSON, A.;THORLEIFSSON, G.;JAKOBSDOTTIR, M.;STEINBERG, S.;PALSSON, S.;JONASSON, F.;SIGURGEIRSSON, B.;THORISDOTTIR, K.;RAGNARSSON, R.;BENEDIKTSDOTTIR, K. R.;ABEN, K. K.;KIEMENEY, L. A.;OLAFSSON, J. H.;GULCHER, J.;KONG, A.;THORSTEINSDOTTIR, U.;STEFANSSON, K. Genetic determinants of hair, eye and skin pigmentation in Europeans. **Nat Genet**, v. 39, n. 12, p. 1443-52, Dec 2007.

SWOPE, V. B.;JAMESON, J. A.;MCFARLAND, K. L.;SUPP, D. M.;MILLER, W. E.;MCGRAW, D. W.;PATEL, M. A.;NIX, M. A.;MILLHAUSER, G. L.;BABCOCK, G. F.;ABDEL-MALEK, Z. A. Defining MC1R regulation in human melanocytes by its agonist alpha-melanocortin and antagonists agouti signaling protein and beta-defensin 3. **J Invest Dermatol**, v. 132, n. 9, p. 2255-62, Sep 2012.

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

TENNESSEN, J. A.;BIGHAM, A. W.;O'CONNOR, T. D.;FU, W.;KENNY, E. E.;GRAVEL, S.;MCGEE, S.;DO, R.;LIU, X.;JUN, G.;KANG, H. M.;JORDAN, D.;LEAL, S. M.;GABRIEL, S.;RIEDER, M. J.;ABECASIS, G.;ALTSHULER, D.;NICKERSON, D. A.;BOERWINKLE, E.;SUNYAEV, S.;BUSTAMANTE, C. D.;BAMSHAD, M. J.;AKEY, J. M.;BROAD, G. O.;SEATTLE, G. O.;PROJECT, N. E. S. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. **Science**, v. 337, n. 6090, p. 64-9, Jul 6 2012.

TULLY, G. Genotype versus phenotype: human pigmentation. **Forensic Sci Int Genet**, v. 1, n. 2, p. 105-10, Jun 2007.

VIGLIAR, E.;MALAPELLE, U.;DE LUCA, C.;BELLEVICINE, C.;TRONCONE, G. Challenges and opportunities of next-generation sequencing: a cytopathologist's perspective. **Cytopathology : official journal of the British Society for Clinical Cytology**, v. 26, n. 5, p. 271-83, Oct 2015.

WATSON, A. New tools. A new breed of high-tech detectives. **Science**, v. 289, n. 5481, p. 850-4, Aug 11 2000.

WEEDEN, T.;STEFANO, J.;DUAN, S.;EDLING, A.;HOU, L.;CHUANG, W. L.;PERRICONE, M. A.;PAN, C.;DZURIS, J. L. A retro-inverso alpha-melanocyte stimulating hormone analog with MC1R-binding selectivity. **J Pept Sci**, v. 17, n. 1, p. 47-55, Jan 2011.

WEIGEL, K. A.;DE LOS CAMPOS, G.;VAZQUEZ, A. I.;ROSA, G. J.;GIANOLA, D.;VAN TASSELL, C. P. Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. **J Dairy Sci**, v. 93, n. 11, p. 5423-35, Nov 2010.

WEIR, B. S.;COCKERHAM, C. C. Estimating F-Statistics for the Analysis of Population Structure. **Evolution**, v. 38, n. 6, p. 1358-1370, november 1984 1984.

WILSON, S.;GINGER, R. S.;DADD, T.;GUNN, D.;LIM, F. L.;SAWICKA, M.;SANDEL, M.;SCHNETKAMP, P. P.;GREEN, M. R. NCKX5, a natural regulator of human skin colour variation, regulates the expression of key pigment genes MC1R and alpha-MSH and alters cholesterol homeostasis in normal human melanocytes. **Adv Exp Med Biol**, v. 961, p. 95-107, 2013.

WONG, L. P.;LAI, J. K.;SAW, W. Y.;ONG, R. T.;CHENG, A. Y.;PILLAI, N. E.;LIU, X.;XU, W.;CHEN, P.;FOO, J. N.;TAN, L. W.;KOO, S. H.;SOONG, R.;WENK, M. R.;LIM, W. Y.;KHOR, C. C.;LITTLE, P.;CHIA, K. S.;TEO, Y. Y. Insights into the genetic structure and diversity of 38 South Asian Indians from deep whole-genome sequencing. **PLoS Genet**, v. 10, n. 5, p. e1004377, May 2014.

Tabela S1 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região 5'URR do gene *MC1R* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	-	1,0000	-	-	-	1	Hapl 115	-	1,0000	-	-	-	1
Hapl 002	0,0879	0,2043	0,3531	0,1655	0,1892	1263	Hapl 116	-	1,0000	-	-	-	1
Hapl 003	-	0,8000	-	-	0,2000	5	Hapl 117	-	1,0000	-	-	-	1
Hapl 004	-	0,7813	-	0,0313	0,1875	32	Hapl 118	-	1,0000	-	-	-	1
Hapl 005	0,0049	0,9064	-	-	0,0887	203	Hapl 119	-	1,0000	-	-	-	4
Hapl 006	0,3644	0,2055	0,0064	0,2309	0,1928	472	Hapl 120	-	1,0000	-	-	-	4
Hapl 007	0,1319	0,3077	-	0,2967	0,2637	91	Hapl 121	1,0000	-	-	-	-	2
Hapl 008	-	1,0000	-	-	-	1	Hapl 122	1,0000	-	-	-	-	6
Hapl 009	-	0,8400	-	-	0,1600	50	Hapl 123	1,0000	-	-	-	-	3
Hapl 010	0,0731	0,2009	0,4749	0,0685	0,1826	219	Hapl 124	1,0000	-	-	-	-	1
Hapl 011	-	0,7143	-	-	0,2857	7	Hapl 125	-	-	-	1,0000	-	4
Hapl 012	-	0,9286	-	-	0,0714	14	Hapl 126	-	-	-	0,5000	0,5000	2
Hapl 013	0,4570	0,0254	0,0342	0,2549	0,2285	1024	Hapl 127	-	-	-	1,0000	-	2
Hapl 014	-	0,9143	-	-	0,0857	35	Hapl 128	-	-	-	0,8333	0,1667	6
Hapl 015	-	1,0000	-	-	-	1	Hapl 129	-	-	-	1,0000	-	4
Hapl 016	-	0,9298	-	-	0,0702	57	Hapl 130	-	-	-	1,0000	-	1
Hapl 017	-	0,8778	-	-	0,1222	90	Hapl 131	-	-	-	1,0000	-	1
Hapl 018	-	0,8000	-	-	0,2000	15	Hapl 132	-	-	-	1,0000	-	1
Hapl 019	0,1273	0,0727	0,0909	0,6545	0,0545	55	Hapl 133	-	-	-	1,0000	-	7
Hapl 020	-	0,9524	-	-	0,0476	21	Hapl 134	-	1,0000	-	-	-	1
Hapl 021	-	1,0000	-	-	-	1	Hapl 135	-	1,0000	-	-	-	4
Hapl 022	-	1,0000	-	-	-	1	Hapl 136	-	1,0000	-	-	-	1
Hapl 023	0,1720	0,0108	0,7366	0,0430	0,0376	372	Hapl 137	-	1,0000	-	-	-	1
Hapl 024	0,0833	0,6667	0,0833	0,0833	0,0833	12	Hapl 138	-	1,0000	-	-	-	1
Hapl 025	-	0,5714	-	0,2857	0,1429	7	Hapl 139	-	1,0000	-	-	-	1
Hapl 026	-	0,8571	-	-	0,1429	7	Hapl 140	-	1,0000	-	-	-	1
Hapl 027	-	1,0000	-	-	-	3	Hapl 141	-	1,0000	-	-	-	1
Hapl 028	-	0,9333	-	-	0,0667	15	Hapl 142	-	0,6667	0,3333	-	-	3
Hapl 029	-	0,8000	-	-	0,2000	10	Hapl 143	-	1,0000	-	-	-	1
Hapl 030	-	1,0000	-	-	-	1	Hapl 144	-	1,0000	-	-	-	1
Hapl 031	-	0,8000	-	-	0,2000	5	Hapl 145	-	1,0000	-	-	-	1
Hapl 032	-	0,9091	-	-	0,0909	33	Hapl 146	-	1,0000	-	-	-	1
Hapl 033	-	1,0000	-	-	-	2	Hapl 147	1,0000	-	-	-	-	1
Hapl 034	0,7119	0,0169	-	0,0169	0,2542	59	Hapl 148	1,0000	-	-	-	-	1
Hapl 035	-	0,8889	-	-	0,1111	9	Hapl 149	-	-	-	1,0000	-	1
Hapl 036	0,2400	0,5600	-	0,0800	0,1200	25	Hapl 150	-	-	-	1,0000	-	1
Hapl 037	-	1,0000	-	-	-	3	Hapl 151	-	-	-	1,0000	-	1
Hapl 038	-	0,4113	-	0,5161	0,0726	124	Hapl 152	-	-	-	1,0000	-	1
Hapl 039	-	0,6667	-	0,3333	-	3	Hapl 153	-	-	-	1,0000	-	2
Hapl 040	0,0189	0,3019	-	0,6415	0,0377	53	Hapl 154	-	-	-	1,0000	-	1
Hapl 041	-	0,7000	-	-	0,3000	10	Hapl 155	-	-	-	1,0000	-	1
Hapl 042	-	1,0000	-	-	-	1	Hapl 156	-	-	-	1,0000	-	1
Hapl 043	-	1,0000	-	-	-	2	Hapl 157	-	-	-	1,0000	-	2
Hapl 044	-	-	-	-	1,0000	1	Hapl 158	-	-	-	1,0000	-	1
Hapl 045	0,6667	-	-	-	0,3333	3	Hapl 159	-	-	-	1,0000	-	2
Hapl 046	-	-	-	-	1,0000	1	Hapl 160	-	-	-	1,0000	-	1
Hapl 047	-	-	0,5000	-	0,5000	4	Hapl 161	-	-	1,0000	-	-	1
Hapl 048	-	0,8000	-	-	0,2000	5	Hapl 162	-	-	1,0000	-	-	1
Hapl 049	-	-	-	-	1,0000	1	Hapl 163	-	-	1,0000	-	-	1
Hapl 050	-	-	-	-	1,0000	1	Hapl 164	-	-	1,0000	-	-	1
Hapl 051	-	0,6667	-	-	0,3333	3	Hapl 165	-	-	1,0000	-	-	1
Hapl 052	-	-	-	-	1,0000	1	Hapl 166	-	-	1,0000	-	-	1
Hapl 053	-	-	-	-	1,0000	1	Hapl 167	-	-	1,0000	-	-	1
Hapl 054	-	-	-	-	1,0000	1	Hapl 168	-	-	1,0000	-	-	1
Hapl 055	-	-	-	-	1,0000	1	Hapl 169	-	-	1,0000	-	-	1
Hapl 056	-	-	-	-	1,0000	1	Hapl 170	-	-	1,0000	-	-	1
Hapl 057	0,3529	0,0588	-	0,1176	0,4706	17	Hapl 171	-	-	1,0000	-	-	1
Hapl 058	-	0,6667	-	-	0,3333	6	Hapl 172	-	-	1,0000	-	-	1
Hapl 059	-	-	-	-	1,0000	1	Hapl 173	-	1,0000	-	-	-	1
Hapl 060	-	-	-	1,0000	-	2	Hapl 174	-	1,0000	-	-	-	1
Hapl 061	0,0147	0,0147	0,0147	0,9559	-	68	Hapl 175	-	1,0000	-	-	-	1
Hapl 062	-	-	-	1,0000	-	3	Hapl 176	-	1,0000	-	-	-	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 063	0,1316	-	-	0,8684	-	38	Hapl 177	-	1,0000	-	-	-	4
Hapl 064	-	-	-	1,0000	-	1	Hapl 178	-	1,0000	-	-	-	3
Hapl 065	-	-	-	1,0000	-	6	Hapl 179	-	1,0000	-	-	-	1
Hapl 066	-	-	-	1,0000	-	2	Hapl 180	-	1,0000	-	-	-	1
Hapl 067	-	-	-	1,0000	-	2	Hapl 181	-	1,0000	-	-	-	1
Hapl 068	-	-	-	1,0000	-	1	Hapl 182	-	0,5000	-	-	0,5000	2
Hapl 069	-	-	-	1,0000	-	2	Hapl 183	-	1,0000	-	-	-	1
Hapl 070	-	-	-	1,0000	-	1	Hapl 184	-	1,0000	-	-	-	1
Hapl 071	0,3636	-	-	0,4545	0,1818	11	Hapl 185	-	1,0000	-	-	-	1
Hapl 072	-	-	0,6667	0,3333	-	3	Hapl 186	-	1,0000	-	-	-	1
Hapl 073	-	-	-	1,0000	-	1	Hapl 187	-	1,0000	-	-	-	1
Hapl 074	0,5000	-	-	0,5000	-	2	Hapl 188	-	1,0000	-	-	-	1
Hapl 075	-	-	-	1,0000	-	1	Hapl 189	-	1,0000	-	-	-	1
Hapl 076	-	-	0,5000	0,1667	0,3333	6	Hapl 190	-	1,0000	-	-	-	2
Hapl 077	-	-	1,0000	-	-	2	Hapl 191	-	1,0000	-	-	-	1
Hapl 078	-	-	1,0000	-	-	30	Hapl 192	-	1,0000	-	-	-	1
Hapl 079	-	0,1111	0,4444	0,2222	0,2222	9	Hapl 193	-	1,0000	-	-	-	1
Hapl 080	0,2075	-	0,7170	0,0377	0,0377	53	Hapl 194	-	-	-	-	1,0000	1
Hapl 081	-	-	1,0000	-	-	1	Hapl 195	-	-	-	-	1,0000	1
Hapl 082	-	-	1,0000	-	-	4	Hapl 196	-	-	-	-	1,0000	1
Hapl 083	-	-	1,0000	-	-	6	Hapl 197	-	-	-	-	1,0000	1
Hapl 084	-	-	1,0000	-	-	1	Hapl 198	-	-	-	-	1,0000	1
Hapl 085	-	-	1,0000	-	-	4	Hapl 199	-	-	-	1,0000	-	1
Hapl 086	-	-	1,0000	-	-	4	Hapl 200	-	-	-	1,0000	-	1
Hapl 087	1,0000	-	-	-	-	1	Hapl 201	-	-	-	1,0000	-	2
Hapl 088	0,7143	-	-	-	0,2857	14	Hapl 202	-	-	-	1,0000	-	1
Hapl 089	0,7805	-	-	0,0732	0,1463	41	Hapl 203	-	-	-	1,0000	-	1
Hapl 090	1,0000	-	-	-	-	1	Hapl 204	-	-	-	1,0000	-	2
Hapl 091	0,1875	-	0,0625	0,0625	0,6875	16	Hapl 205	-	-	-	1,0000	-	1
Hapl 092	1,0000	-	-	-	-	1	Hapl 206	-	-	-	1,0000	-	1
Hapl 093	1,0000	-	-	-	-	2	Hapl 207	-	-	-	-	1,0000	1
Hapl 094	1,0000	-	-	-	-	1	Hapl 208	-	-	-	-	1,0000	1
Hapl 095	1,0000	-	-	-	-	1	Hapl 209	-	-	-	1,0000	-	1
Hapl 096	-	-	1,0000	-	-	1	Hapl 210	-	-	-	1,0000	-	1
Hapl 097	-	-	1,0000	-	-	2	Hapl 211	-	-	-	1,0000	-	1
Hapl 098	-	-	1,0000	-	-	2	Hapl 212	-	-	-	1,0000	-	1
Hapl 099	-	-	1,0000	-	-	4	Hapl 213	-	-	-	1,0000	-	1
Hapl 100	-	-	1,0000	-	-	1	Hapl 214	-	-	-	1,0000	-	1
Hapl 101	0,1667	-	0,8333	-	-	6	Hapl 215	1,0000	-	-	-	-	2
Hapl 102	-	-	1,0000	-	-	1	Hapl 216	1,0000	-	-	-	-	1
Hapl 103	-	-	0,5000	-	0,5000	2	Hapl 217	1,0000	-	-	-	-	1
Hapl 104	-	-	1,0000	-	-	1	Hapl 218	-	1,0000	-	-	-	1
Hapl 105	-	-	1,0000	-	-	2	Hapl 219	-	1,0000	-	-	-	1
Hapl 106	-	-	1,0000	-	-	1	Hapl 220	-	1,0000	-	-	-	1
Hapl 107	-	-	1,0000	-	-	1	Hapl 221	-	1,0000	-	-	-	1
Hapl 108	-	-	1,0000	-	-	1	Hapl 222	-	1,0000	-	-	-	1
Hapl 109	0,5000	-	0,1667	0,1667	0,1667	6	Hapl 223	-	1,0000	-	-	-	1
Hapl 110	0,2000	-	-	0,6000	0,2000	5	Hapl 224	-	1,0000	-	-	-	1
Hapl 111	-	-	-	-	1,0000	1	Hapl 225	-	1,0000	-	-	-	1
Hapl 112	-	-	-	-	1,0000	1	Hapl 226	-	1,0000	-	-	-	1
Hapl 113	-	-	-	-	1,0000	1	Hapl 227	-	1,0000	-	-	-	1
Hapl 114	-	1,0000	-	-	-	1							

Tabela S2 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região 5'UTR do gene *MC1R* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	-	0,8633	-	-	0,1367	139	Hapl 029	-	0,8333	-	-	0,1667	6
Hapl 002	-	0,9259	-	-	0,0741	27	Hapl 030	1,0000	-	-	-	-	3
Hapl 003	0,0086	0,6567	-	0,2146	0,1202	233	Hapl 031	1,0000	-	-	-	-	2
Hapl 004	0,1254	0,1723	0,3035	0,2251	0,1738	2066	Hapl 032	-	-	-	1,0000	-	2
Hapl 005	0,1884	0,5870	-	0,0193	0,2053	414	Hapl 033	-	-	-	1,0000	-	4
Hapl 006	-	1,0000	-	-	-	1	Hapl 034	-	-	-	0,8333	0,1667	6
Hapl 007	0,1139	0,3015	0,4824	0,0318	0,0704	597	Hapl 035	-	-	-	1,0000	-	5
Hapl 008	0,4641	0,0384	0,0645	0,2247	0,2083	1224	Hapl 036	1,0000	-	-	-	-	1
Hapl 009	-	0,9655	-	-	0,0345	29	Hapl 037	1,0000	-	-	-	-	1
Hapl 010	0,1277	0,1064	0,0532	0,6809	0,0319	94	Hapl 038	-	-	-	1,0000	-	1
Hapl 011	-	1,0000	-	-	-	16	Hapl 039	-	-	1,0000	-	-	2
Hapl 012	-	1,0000	-	-	-	2	Hapl 040	-	-	1,0000	-	-	1
Hapl 013	-	1,0000	-	-	-	1	Hapl 041	-	0,6667	-	-	0,3333	9
Hapl 014	-	1,0000	-	-	-	1	Hapl 042	-	1,0000	-	-	-	1
Hapl 015	-	0,6667	-	-	0,3333	3	Hapl 043	-	0,6667	-	-	0,3333	3
Hapl 016	-	-	-	1,0000	-	1	Hapl 044	-	-	-	1,0000	-	1
Hapl 017	-	-	-	1,0000	-	1	Hapl 045	-	-	-	1,0000	-	1
Hapl 018	-	-	-	1,0000	-	59	Hapl 046	-	-	-	1,0000	-	1
Hapl 019	-	-	-	1,0000	-	2	Hapl 047	-	-	-	1,0000	-	1
Hapl 020	-	-	-	1,0000	-	6	Hapl 048	-	-	-	-	1,0000	1
Hapl 021	-	-	-	1,0000	-	3	Hapl 049	-	-	-	1,0000	-	1
Hapl 022	-	-	1,0000	-	-	3	Hapl 050	-	-	-	1,0000	-	2
Hapl 023	-	-	1,0000	-	-	1	Hapl 051	-	-	-	1,0000	-	1
Hapl 024	-	-	1,0000	-	-	1	Hapl 052	1,0000	-	-	-	-	3
Hapl 025	0,7143	-	-	-	0,2857	7	Hapl 053	1,0000	-	-	-	-	1
Hapl 026	-	-	1,0000	-	-	1	Hapl 054	1,0000	-	-	-	-	1
Hapl 027	-	-	-	-	1,0000	12	Hapl 055	1,0000	-	-	-	-	1
Hapl 028	-	1,0000	-	-	-	1	Hapl 056	1,0000	-	-	-	-	1

Tabela S3 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região Codificante do gene *MC1R* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	0,2635	0,2075	0,0227	0,3243	0,1819	2072	Hapl 052	-	-	-	1,0000	-	1
Hapl 002	0,0188	0,6558	0,0063	0,2286	0,0905	796	Hapl 053	-	-	-	1,0000	-	1
Hapl 003	0,0755	0,0032	0,6451	0,0410	0,2352	927	Hapl 054	-	-	-	1,0000	-	1
Hapl 004	-	1,0000	-	-	-	2	Hapl 055	-	1,0000	-	-	-	1
Hapl 005	-	0,8214	-	-	0,1786	28	Hapl 056	-	1,0000	-	-	-	1
Hapl 006	-	1,0000	-	-	-	1	Hapl 057	0,5000	0,5000	-	-	-	2
Hapl 007	-	0,8727	-	-	0,1273	55	Hapl 058	-	1,0000	-	-	-	1
Hapl 008	-	0,8723	-	-	0,1277	94	Hapl 059	-	1,0000	-	-	-	1
Hapl 009	-	1,0000	-	-	-	1	Hapl 060	1,0000	-	-	-	-	1
Hapl 010	-	0,9474	-	-	0,0526	19	Hapl 061	1,0000	-	-	-	-	1
Hapl 011	0,6384	0,0339	-	0,0395	0,2881	177	Hapl 062	-	-	-	1,0000	-	4
Hapl 012	0,1756	0,0102	0,7252	0,0458	0,0433	393	Hapl 063	-	-	-	1,0000	-	1
Hapl 013	0,7742	0,0215	0,0108	0,0538	0,1398	93	Hapl 064	-	-	-	1,0000	-	3
Hapl 014	-	1,0000	-	-	-	11	Hapl 065	-	-	-	1,0000	-	1
Hapl 015	0,7500	0,2500	-	-	-	4	Hapl 066	-	-	-	1,0000	-	2
Hapl 016	0,7692	0,0769	-	0,0769	0,0769	13	Hapl 067	-	-	-	1,0000	-	1
Hapl 017	-	1,0000	-	-	-	1	Hapl 068	-	-	1,0000	-	-	1
Hapl 018	0,4706	0,0588	-	-	0,4706	17	Hapl 069	-	-	1,0000	-	-	2
Hapl 019	-	1,0000	-	-	-	1	Hapl 070	-	-	1,0000	-	-	1
Hapl 020	0,8592	0,0141	-	0,0423	0,0845	71	Hapl 071	-	-	1,0000	-	-	1
Hapl 021	0,6429	0,0714	-	-	0,2857	14	Hapl 072	-	-	1,0000	-	-	1
Hapl 022	-	-	-	-	1,0000	1	Hapl 073	-	-	1,0000	-	-	1
Hapl 023	-	0,6667	-	-	0,3333	3	Hapl 074	-	-	1,0000	-	-	1
Hapl 024	-	0,8571	-	-	0,1429	7	Hapl 075	-	1,0000	-	-	-	1
Hapl 025	0,0667	0,0333	-	0,9000	-	30	Hapl 076	-	1,0000	-	-	-	1
Hapl 026	-	-	-	1,0000	-	3	Hapl 077	-	1,0000	-	-	-	1
Hapl 027	-	-	1,0000	-	-	42	Hapl 078	-	1,0000	-	-	-	1
Hapl 028	-	-	1,0000	-	-	4	Hapl 079	-	0,7500	-	-	0,2500	4
Hapl 029	-	-	1,0000	-	-	3	Hapl 080	-	-	-	-	1,0000	1
Hapl 030	-	-	1,0000	-	-	1	Hapl 081	-	-	-	-	1,0000	1
Hapl 031	-	-	1,0000	-	-	1	Hapl 082	-	-	-	-	1,0000	4
Hapl 032	-	-	1,0000	-	-	1	Hapl 083	-	-	-	1,0000	-	1
Hapl 033	-	-	1,0000	-	-	4	Hapl 084	-	-	-	-	1,0000	1
Hapl 034	0,5000	-	-	0,1250	0,3750	16	Hapl 085	-	-	-	-	1,0000	1
Hapl 035	1,0000	-	-	-	-	1	Hapl 086	-	-	-	-	1,0000	1
Hapl 036	0,7500	-	-	-	0,2500	4	Hapl 087	-	0,5000	-	-	0,5000	2
Hapl 037	-	-	1,0000	-	-	1	Hapl 088	-	-	-	-	1,0000	1
Hapl 038	-	-	1,0000	-	-	1	Hapl 089	-	-	-	1,0000	-	1
Hapl 039	-	-	1,0000	-	-	1	Hapl 090	-	-	-	1,0000	-	1
Hapl 040	-	-	1,0000	-	-	1	Hapl 091	1,0000	-	-	-	-	1
Hapl 041	-	-	1,0000	-	-	2	Hapl 092	1,0000	-	-	-	-	1
Hapl 042	-	-	1,0000	-	-	1	Hapl 093	1,0000	-	-	-	-	1
Hapl 043	-	-	1,0000	-	-	1	Hapl 094	-	1,0000	-	-	-	1
Hapl 044	-	-	-	-	1,0000	1	Hapl 095	-	1,0000	-	-	-	2
Hapl 045	-	-	-	-	1,0000	1	Hapl 096	-	1,0000	-	-	-	1
Hapl 046	-	1,0000	-	-	-	5	Hapl 097	-	1,0000	-	-	-	2
Hapl 047	-	1,0000	-	-	-	1	Hapl 098	-	1,0000	-	-	-	1
Hapl 048	-	1,0000	-	-	-	2	Hapl 099	-	1,0000	-	-	-	1
Hapl 049	0,6667	-	-	0,3333	-	6	Hapl 100	-	1,0000	-	-	-	1
Hapl 050	1,0000	-	-	-	-	1	Hapl 101	-	1,0000	-	-	-	1
Hapl 051	1,0000	-	-	-	-	5							

Tabela S4 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região 3'UTR do gene *MC1R* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	0,2360	0,1662	0,2008	0,2129	0,1840	3555	Hapl 017	-	1,0000	-	-	-	2
Hapl 002	0,0083	0,4959	-	0,4008	0,0950	242	Hapl 018	1,0000	-	-	-	-	1
Hapl 003	0,1151	0,3210	0,3734	0,1228	0,0678	782	Hapl 019	-	-	-	1,0000	-	1
Hapl 004	-	0,8863	-	-	0,1137	211	Hapl 020	-	-	1,0000	-	-	1
Hapl 005	-	1,0000	-	-	-	2	Hapl 021	-	-	1,0000	-	-	1
Hapl 006	0,4857	0,0786	-	0,0571	0,3786	140	Hapl 022	-	1,0000	-	-	-	1
Hapl 007	-	0,9000	-	-	0,1000	20	Hapl 023	-	1,0000	-	-	-	1
Hapl 008	-	1,0000	-	-	-	2	Hapl 024	-	-	-	-	1,0000	1
Hapl 009	-	1,0000	-	-	-	1	Hapl 025	-	-	-	-	1,0000	1
Hapl 010	-	0,9000	-	-	0,1000	10	Hapl 026	-	-	-	-	1,0000	1
Hapl 011	-	0,6667	-	-	0,3333	3	Hapl 027	-	-	-	-	1,0000	2
Hapl 012	-	-	-	1,0000	-	18	Hapl 028	1,0000	-	-	-	-	1
Hapl 013	-	1,0000	-	-	-	1	Hapl 029	1,0000	-	-	-	-	1
Hapl 014	1,0000	-	-	-	-	2	Hapl 030	1,0000	-	-	-	-	1
Hapl 015	1,0000	-	-	-	-	1	Hapl 031	-	1,0000	-	-	-	1
Hapl 016	-	-	-	1,0000	-	1							

Tabela S5 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região Total (mRNA) do gene *MC1R* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	0,1984	0,5079	0,1270	0,1111	0,0556	126	Hapl 096	1,0000	-	-	-	-	2
Hapl 002	0,1118	0,0499	0,4680	0,1759	0,1944	921	Hapl 097	1,0000	-	-	-	-	1
Hapl 003	0,1323	0,4815	0,1005	0,1852	0,1005	189	Hapl 098	1,0000	-	-	-	-	1
Hapl 004	0,1614	0,2756	0,1025	0,3133	0,1472	849	Hapl 099	1,0000	-	-	-	-	1
Hapl 005	0,2871	0,1566	0,1056	0,2301	0,2206	843	Hapl 100	1,0000	-	-	-	-	1
Hapl 006	-	1,0000	-	-	-	1	Hapl 101	1,0000	-	-	-	-	1
Hapl 007	0,3782	0,2773	0,0084	0,0504	0,2857	119	Hapl 102	1,0000	-	-	-	-	1
Hapl 008	0,2222	0,4444	0,1111	0,1111	0,1111	9	Hapl 103	-	0,5000	-	0,5000	-	2
Hapl 009	-	1,0000	-	-	-	1	Hapl 104	-	-	-	1,0000	-	2
Hapl 010	0,2326	0,4884	0,0465	0,1163	0,1163	43	Hapl 105	-	-	-	1,0000	-	1
Hapl 011	0,3869	0,2738	0,0595	0,0536	0,2262	168	Hapl 106	-	-	-	1,0000	-	1
Hapl 012	0,1687	0,5582	0,1245	0,1004	0,0482	249	Hapl 107	-	-	-	1,0000	-	1
Hapl 013	0,1221	0,6183	0,0763	0,0992	0,0840	131	Hapl 108	-	0,2000	0,2000	0,6000	-	5
Hapl 014	0,1071	0,4286	0,1786	0,0714	0,2143	28	Hapl 109	-	1,0000	-	-	-	1
Hapl 015	0,0526	0,6316	0,0526	0,1579	0,1053	19	Hapl 110	-	1,0000	-	-	-	1
Hapl 016	0,1828	0,5269	0,1075	0,0968	0,0860	93	Hapl 111	-	1,0000	-	-	-	1
Hapl 017	-	0,7143	0,1429	-	0,1429	7	Hapl 112	1,0000	-	-	-	-	1
Hapl 018	0,1319	0,2088	0,0989	0,4835	0,0769	91	Hapl 113	-	-	0,2500	0,7500	-	4
Hapl 019	0,3846	0,1648	0,0659	0,0440	0,3407	91	Hapl 114	-	-	-	1,0000	-	1
Hapl 020	0,2941	0,2353	-	0,0588	0,4118	17	Hapl 115	-	-	-	1,0000	-	1
Hapl 021	0,1881	0,0773	0,4562	0,1392	0,1392	388	Hapl 116	-	-	-	1,0000	-	1
Hapl 022	0,1000	0,5000	0,3000	0,1000	-	10	Hapl 117	-	-	-	1,0000	-	3
Hapl 023	0,2500	0,7500	-	-	-	4	Hapl 118	-	-	-	1,0000	-	2
Hapl 024	0,0714	0,7857	0,0357	0,0714	0,0357	28	Hapl 119	-	-	-	1,0000	-	1
Hapl 025	0,2500	0,5833	-	-	0,1667	12	Hapl 120	-	-	-	1,0000	-	1
Hapl 026	0,4000	0,6000	-	-	-	5	Hapl 121	-	-	1,0000	-	-	1
Hapl 027	0,6286	0,0714	0,0429	0,0286	0,2286	70	Hapl 122	-	0,2500	0,2500	-	0,5000	4
Hapl 028	0,2143	0,3571	0,4286	-	-	14	Hapl 123	-	-	0,2727	-	0,7273	11
Hapl 029	-	1,0000	-	-	-	2	Hapl 124	-	-	1,0000	-	-	1
Hapl 030	-	1,0000	-	-	-	1	Hapl 125	-	-	1,0000	-	-	1
Hapl 031	-	1,0000	-	-	-	1	Hapl 126	-	-	1,0000	-	-	2
Hapl 032	0,4615	0,2308	0,0769	-	0,2308	13	Hapl 127	-	-	1,0000	-	-	1
Hapl 033	-	-	-	-	1,0000	1	Hapl 128	-	-	1,0000	-	-	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 034	-	-	-	-	1,0000	1	Hapl 129	-	-	1,0000	-	-	1
Hapl 035	-	-	-	0,5000	0,5000	4	Hapl 130	-	-	1,0000	-	-	1
Hapl 036	-	-	0,3333	0,3333	0,3333	3	Hapl 131	-	-	0,6667	0,3333	-	3
Hapl 037	0,1111	0,7778	-	0,0556	0,0556	18	Hapl 132	-	-	1,0000	-	-	1
Hapl 038	-	0,2778	0,1111	0,5556	0,0556	18	Hapl 133	-	-	1,0000	-	-	1
Hapl 039	-	0,2703	0,0270	0,6216	0,0811	37	Hapl 134	-	-	1,0000	-	-	1
Hapl 040	0,5000	-	-	0,1667	0,3333	6	Hapl 135	-	-	1,0000	-	-	1
Hapl 041	0,3704	0,3333	0,0741	0,1852	0,0370	27	Hapl 136	-	-	1,0000	-	-	1
Hapl 042	0,1034	0,0345	0,0690	0,7586	0,0345	29	Hapl 137	-	-	1,0000	-	-	1
Hapl 043	-	0,5000	-	0,5000	-	6	Hapl 138	-	1,0000	-	-	-	1
Hapl 044	-	-	-	1,0000	-	1	Hapl 139	-	1,0000	-	-	-	1
Hapl 045	-	-	0,5000	0,5000	-	2	Hapl 140	-	0,6667	-	-	0,3333	3
Hapl 046	-	-	-	1,0000	-	6	Hapl 141	-	1,0000	-	-	-	1
Hapl 047	-	-	0,3333	0,3333	0,3333	6	Hapl 142	-	1,0000	-	-	-	1
Hapl 048	-	0,3333	-	0,6667	-	3	Hapl 143	-	1,0000	-	-	-	1
Hapl 049	-	-	-	1,0000	-	2	Hapl 144	-	1,0000	-	-	-	2
Hapl 050	0,1667	-	0,5952	0,1667	0,0714	42	Hapl 145	-	1,0000	-	-	-	1
Hapl 051	0,2500	-	0,5000	-	0,2500	4	Hapl 146	-	1,0000	-	-	-	1
Hapl 052	-	-	0,6667	0,3333	-	3	Hapl 147	-	0,5000	-	-	0,5000	2
Hapl 053	0,3333	-	0,6667	-	-	3	Hapl 148	-	-	-	-	1,0000	2
Hapl 054	-	-	1,0000	-	-	1	Hapl 149	-	-	-	-	1,0000	1
Hapl 055	-	-	1,0000	-	-	1	Hapl 150	-	-	-	-	1,0000	1
Hapl 056	0,4444	0,3333	0,2222	-	-	9	Hapl 151	-	-	-	-	1,0000	1
Hapl 057	0,4667	0,2000	-	0,0667	0,2667	15	Hapl 152	-	-	-	-	1,0000	1
Hapl 058	1,0000	-	-	-	-	1	Hapl 153	-	-	-	0,5000	0,5000	2
Hapl 059	0,7500	0,2500	-	-	-	4	Hapl 154	-	-	-	-	1,0000	1
Hapl 060	1,0000	-	-	-	-	1	Hapl 155	-	-	-	-	1,0000	1
Hapl 061	1,0000	-	-	-	-	1	Hapl 156	-	-	-	-	1,0000	1
Hapl 062	1,0000	-	-	-	-	1	Hapl 157	-	-	-	-	1,0000	1
Hapl 063	1,0000	-	-	-	-	1	Hapl 158	-	-	-	1,0000	-	1
Hapl 064	-	-	1,0000	-	-	1	Hapl 159	-	-	-	1,0000	-	1
Hapl 065	0,3333	-	0,3333	0,3333	-	3	Hapl 160	-	-	-	1,0000	-	1
Hapl 066	-	-	1,0000	-	-	1	Hapl 161	-	-	-	1,0000	-	1
Hapl 067	0,1000	0,6000	0,3000	-	-	10	Hapl 162	-	-	-	-	1,0000	1
Hapl 068	-	0,5000	0,5000	-	-	2	Hapl 163	-	-	-	-	1,0000	1
Hapl 069	-	0,3333	0,6667	-	-	3	Hapl 164	-	-	-	-	1,0000	1
Hapl 070	-	-	1,0000	-	-	1	Hapl 165	-	-	-	1,0000	-	1
Hapl 071	-	-	1,0000	-	-	1	Hapl 166	-	-	-	1,0000	-	1
Hapl 072	-	-	1,0000	-	-	1	Hapl 167	-	-	-	1,0000	-	1
Hapl 073	-	-	1,0000	-	-	2	Hapl 168	1,0000	-	-	-	-	1
Hapl 074	-	-	1,0000	-	-	1	Hapl 169	1,0000	-	-	-	-	3
Hapl 075	-	-	0,3333	0,6667	-	3	Hapl 170	1,0000	-	-	-	-	1
Hapl 076	-	-	1,0000	-	-	1	Hapl 171	1,0000	-	-	-	-	1
Hapl 077	-	-	-	-	1,0000	1	Hapl 172	1,0000	-	-	-	-	1
Hapl 078	-	-	-	-	1,0000	1	Hapl 173	1,0000	-	-	-	-	1
Hapl 079	-	-	-	-	1,0000	1	Hapl 174	1,0000	-	-	-	-	1
Hapl 080	-	-	-	-	1,0000	1	Hapl 175	1,0000	-	-	-	-	1
Hapl 081	0,4286	0,3571	-	-	0,2143	14	Hapl 176	1,0000	-	-	-	-	1
Hapl 082	-	-	-	-	1,0000	1	Hapl 177	1,0000	-	-	-	-	1
Hapl 083	-	1,0000	-	-	-	1	Hapl 178	1,0000	-	-	-	-	1
Hapl 084	-	1,0000	-	-	-	1	Hapl 179	1,0000	-	-	-	-	1
Hapl 085	-	0,6000	0,4000	-	-	5	Hapl 180	1,0000	-	-	-	-	1
Hapl 086	-	0,6000	-	0,2000	0,2000	5	Hapl 181	-	1,0000	-	-	-	1
Hapl 087	-	1,0000	-	-	-	2	Hapl 182	-	1,0000	-	-	-	2
Hapl 088	-	1,0000	-	-	-	1	Hapl 183	-	1,0000	-	-	-	1
Hapl 089	-	1,0000	-	-	-	1	Hapl 184	-	1,0000	-	-	-	1
Hapl 090	1,0000	-	-	-	-	1	Hapl 185	-	1,0000	-	-	-	1
Hapl 091	1,0000	-	-	-	-	1	Hapl 186	-	1,0000	-	-	-	1
Hapl 092	1,0000	-	-	-	-	1	Hapl 187	-	1,0000	-	-	-	1
Hapl 093	1,0000	-	-	-	-	3	Hapl 188	-	1,0000	-	-	-	1
Hapl 094	0,3333	-	-	-	0,6667	3	Hapl 189	-	1,0000	-	-	-	1
Hapl 095	1,0000	-	-	-	-	1	Hapl 190	-	1,0000	-	-	-	1

Tabela S6 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região Total Estendida do gene *MC1R* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	-	1,0000	-	-	-	1	Hapl 227	1,0000	-	-	-	-	3
Hapl 002	0,1075	0,0493	0,4642	0,1612	0,2179	670	Hapl 228	0,3333	-	-	-	0,6667	3
Hapl 003	0,2000	0,8000	-	-	-	5	Hapl 229	1,0000	-	-	-	-	1
Hapl 004	-	0,2667	0,1333	0,0667	0,5333	15	Hapl 230	1,0000	-	-	-	-	1
Hapl 005	0,1290	0,6452	0,0645	0,0323	0,1290	31	Hapl 231	1,0000	-	-	-	-	1
Hapl 006	0,1602	0,2641	0,1039	0,2814	0,1905	231	Hapl 232	1,0000	-	-	-	-	2
Hapl 007	0,2143	0,5408	0,1020	0,1122	0,0306	98	Hapl 233	1,0000	-	-	-	-	2
Hapl 008	0,2592	0,1501	0,1147	0,2521	0,2238	706	Hapl 234	1,0000	-	-	-	-	1
Hapl 009	-	1,0000	-	-	-	1	Hapl 235	1,0000	-	-	-	-	1
Hapl 010	0,3889	0,2778	0,0093	0,0556	0,2685	108	Hapl 236	1,0000	-	-	-	-	1
Hapl 011	0,1628	0,5581	0,0698	0,1163	0,0930	43	Hapl 237	1,0000	-	-	-	-	1
Hapl 012	0,1667	0,6667	-	-	0,1667	6	Hapl 238	1,0000	-	-	-	-	1
Hapl 013	-	1,0000	-	-	-	1	Hapl 239	0,2000	0,2000	0,4000	0,1000	0,1000	10
Hapl 014	0,2063	0,2857	0,1429	0,1905	0,1746	63	Hapl 240	1,0000	-	-	-	-	1
Hapl 015	0,1563	0,5313	0,0625	0,1250	0,1250	32	Hapl 241	1,0000	-	-	-	-	1
Hapl 016	0,3765	0,2840	0,0617	0,0556	0,2222	162	Hapl 242	0,6667	-	-	-	0,3333	3
Hapl 017	-	1,0000	-	-	-	1	Hapl 243	-	-	-	1,0000	-	4
Hapl 018	0,2500	0,2500	-	-	0,5000	8	Hapl 244	-	-	-	1,0000	-	1
Hapl 019	0,1835	0,2785	0,1171	0,3196	0,1013	316	Hapl 245	-	-	-	1,0000	-	1
Hapl 020	0,4615	0,2308	0,1538	0,0769	0,0769	13	Hapl 246	-	-	-	1,0000	-	1
Hapl 021	0,1719	0,5469	0,1354	0,1094	0,0365	192	Hapl 247	-	-	-	1,0000	-	1
Hapl 022	0,0941	0,6118	0,1059	0,0941	0,0941	85	Hapl 248	-	-	-	1,0000	-	1
Hapl 023	0,1000	0,4500	0,2000	0,1000	0,1500	20	Hapl 249	-	0,2500	0,2500	0,5000	-	4
Hapl 024	0,1667	0,8333	-	-	-	12	Hapl 250	-	-	-	1,0000	-	1
Hapl 025	0,5610	0,1707	0,0488	0,0488	0,1707	41	Hapl 251	-	-	-	1,0000	-	1
Hapl 026	0,0526	0,6316	0,0526	0,1579	0,1053	19	Hapl 252	-	-	-	1,0000	-	1
Hapl 027	0,1754	0,4912	0,1053	0,1053	0,1228	57	Hapl 253	-	0,2500	0,2500	0,5000	-	4
Hapl 028	-	1,0000	-	-	-	2	Hapl 254	-	-	-	1,0000	-	1
Hapl 029	-	1,0000	-	-	-	1	Hapl 255	-	-	-	1,0000	-	1
Hapl 030	0,2308	0,6538	-	0,0769	0,0385	26	Hapl 256	-	-	-	1,0000	-	1
Hapl 031	0,3333	0,5000	-	0,1667	-	6	Hapl 257	-	0,1429	-	0,7143	0,1429	7
Hapl 032	0,3820	0,1685	0,0562	0,0449	0,3483	89	Hapl 258	-	-	-	1,0000	-	1
Hapl 033	0,6667	0,3333	-	-	-	6	Hapl 259	-	1,0000	-	-	-	1
Hapl 034	0,3125	0,2500	-	0,0625	0,3750	16	Hapl 260	-	1,0000	-	-	-	1
Hapl 035	0,1951	0,0687	0,4643	0,1346	0,1374	364	Hapl 261	-	1,0000	-	-	-	2
Hapl 036	0,1419	0,0581	0,4645	0,2000	0,1355	155	Hapl 262	-	0,1667	0,3333	0,3333	0,1667	6
Hapl 037	0,1000	0,5000	0,3000	0,1000	-	10	Hapl 263	-	1,0000	-	-	-	1
Hapl 038	-	0,6667	-	0,1667	0,1667	6	Hapl 264	-	1,0000	-	-	-	1
Hapl 039	0,3333	0,2000	0,0667	0,1333	0,2667	15	Hapl 265	-	1,0000	-	-	-	1
Hapl 040	0,2500	0,7500	-	-	-	4	Hapl 266	-	0,3333	-	0,6667	-	3
Hapl 041	0,0476	0,8571	0,0476	-	0,0476	21	Hapl 267	-	1,0000	-	-	-	1
Hapl 042	0,0909	0,6667	0,0909	0,1212	0,0303	33	Hapl 268	-	1,0000	-	-	-	1
Hapl 043	0,1538	0,4615	0,1538	0,1538	0,0769	13	Hapl 269	-	1,0000	-	-	-	1
Hapl 044	0,3333	0,5556	-	-	0,1111	9	Hapl 270	-	1,0000	-	-	-	1
Hapl 045	0,4750	0,2750	0,0250	0,0750	0,1500	40	Hapl 271	-	1,0000	-	-	-	1
Hapl 046	0,1429	0,7143	-	-	0,1429	7	Hapl 272	-	1,0000	-	-	-	1
Hapl 047	0,4000	0,6000	-	-	-	5	Hapl 273	-	1,0000	-	-	-	1
Hapl 048	0,1667	0,5417	0,1667	0,0833	0,0417	24	Hapl 274	-	1,0000	-	-	-	1
Hapl 049	0,6286	0,0714	0,0429	0,0286	0,2286	70	Hapl 275	-	1,0000	-	-	-	1
Hapl 050	0,1951	0,2683	0,1463	0,3415	0,0488	41	Hapl 276	1,0000	-	-	-	-	1
Hapl 051	0,1667	0,5000	0,1667	0,1667	-	6	Hapl 277	1,0000	-	-	-	-	1
Hapl 052	0,2500	0,3333	0,4167	-	-	12	Hapl 278	-	-	-	1,0000	-	1
Hapl 053	-	0,6667	0,3333	-	-	3	Hapl 279	-	-	0,2500	0,7500	-	4
Hapl 054	0,1067	0,4000	0,1200	0,2800	0,0933	75	Hapl 280	-	-	-	1,0000	-	1
Hapl 055	-	1,0000	-	-	-	2	Hapl 281	-	-	-	1,0000	-	1
Hapl 056	-	1,0000	-	-	-	1	Hapl 282	-	-	-	1,0000	-	1
Hapl 057	-	1,0000	-	-	-	1	Hapl 283	-	-	-	1,0000	-	1
Hapl 058	0,1071	0,3571	0,1786	0,2500	0,1071	28	Hapl 284	-	-	-	1,0000	-	2
Hapl 059	0,4615	0,2308	0,0769	-	0,2308	13	Hapl 285	-	-	-	1,0000	-	1
Hapl 060	-	0,5000	-	0,5000	-	2	Hapl 286	-	-	-	1,0000	-	1
Hapl 061	-	-	-	-	1,0000	1	Hapl 287	-	-	-	1,0000	-	2
Hapl 062	-	-	-	-	1,0000	1	Hapl 288	-	-	-	1,0000	-	2

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 063	0,3333	-	-	-	0,6667	3	Hapl 289	-	-	-	1,0000	-	2
Hapl 064	0,1045	0,1045	0,0597	0,6119	0,1194	67	Hapl 290	-	-	-	1,0000	-	2
Hapl 065	-	-	-	-	1,0000	1	Hapl 291	-	-	-	1,0000	-	1
Hapl 066	-	-	-	-	1,0000	1	Hapl 292	-	-	-	1,0000	-	1
Hapl 067	-	-	-	-	1,0000	1	Hapl 293	-	-	-	1,0000	-	2
Hapl 068	-	-	-	-	1,0000	1	Hapl 294	-	-	-	1,0000	-	1
Hapl 069	-	-	-	-	1,0000	2	Hapl 295	-	-	-	1,0000	-	1
Hapl 070	-	-	0,3333	0,3333	0,3333	3	Hapl 296	-	-	0,5000	0,5000	-	2
Hapl 071	-	-	-	-	1,0000	1	Hapl 297	-	-	-	1,0000	-	1
Hapl 072	-	0,6000	0,2000	-	0,2000	5	Hapl 298	-	-	-	1,0000	-	1
Hapl 073	0,2222	0,3889	-	0,1111	0,2778	18	Hapl 299	-	-	-	1,0000	-	1
Hapl 074	-	0,5000	-	-	0,5000	2	Hapl 300	-	-	1,0000	-	-	1
Hapl 075	-	-	-	-	1,0000	1	Hapl 301	-	-	1,0000	-	-	1
Hapl 076	-	-	-	-	1,0000	1	Hapl 302	-	-	0,2727	-	0,7273	11
Hapl 077	0,1111	0,1389	0,0278	0,5833	0,1389	36	Hapl 303	-	-	1,0000	-	-	1
Hapl 078	0,2000	-	-	0,4000	0,4000	5	Hapl 304	-	-	1,0000	-	-	1
Hapl 079	-	-	-	-	1,0000	1	Hapl 305	-	-	0,3333	0,3333	0,3333	3
Hapl 080	-	-	-	-	1,0000	1	Hapl 306	-	-	1,0000	-	-	2
Hapl 081	-	-	-	-	1,0000	1	Hapl 307	-	-	1,0000	-	-	1
Hapl 082	-	0,3333	-	-	0,6667	3	Hapl 308	-	-	1,0000	-	-	1
Hapl 083	-	-	-	0,5000	0,5000	2	Hapl 309	-	-	1,0000	-	-	1
Hapl 084	0,1667	0,3333	0,1667	-	0,3333	6	Hapl 310	-	-	1,0000	-	-	1
Hapl 085	-	0,4615	-	0,4615	0,0769	13	Hapl 311	-	-	0,5000	-	0,5000	2
Hapl 086	0,5000	-	-	0,1667	0,3333	6	Hapl 312	-	-	1,0000	-	-	1
Hapl 087	0,3462	0,3462	0,0769	0,1923	0,0385	26	Hapl 313	-	-	1,0000	-	-	1
Hapl 088	0,0741	0,0370	0,0741	0,7778	0,0370	27	Hapl 314	-	-	1,0000	-	-	1
Hapl 089	-	0,5000	-	0,5000	-	6	Hapl 315	-	-	1,0000	-	-	1
Hapl 090	-	-	-	1,0000	-	3	Hapl 316	-	-	1,0000	-	-	1
Hapl 091	-	-	-	1,0000	-	1	Hapl 317	-	-	1,0000	-	-	1
Hapl 092	-	-	-	1,0000	-	1	Hapl 318	-	-	1,0000	-	-	2
Hapl 093	-	-	-	1,0000	-	1	Hapl 319	-	-	1,0000	-	-	1
Hapl 094	-	-	-	1,0000	-	2	Hapl 320	-	-	1,0000	-	-	1
Hapl 095	-	-	0,5000	0,5000	-	2	Hapl 321	-	-	1,0000	-	-	1
Hapl 096	-	-	-	1,0000	-	6	Hapl 322	-	-	1,0000	-	-	1
Hapl 097	-	-	0,5000	0,5000	-	2	Hapl 323	-	-	1,0000	-	-	1
Hapl 098	-	0,3333	-	0,6667	-	3	Hapl 324	-	-	1,0000	-	-	1
Hapl 099	-	-	-	1,0000	-	1	Hapl 325	-	-	1,0000	-	-	1
Hapl 100	-	-	-	1,0000	-	1	Hapl 326	-	-	1,0000	-	-	1
Hapl 101	-	0,1429	0,1429	0,4286	0,2857	7	Hapl 327	-	-	1,0000	-	-	1
Hapl 102	0,1579	-	0,6053	0,1842	0,0526	38	Hapl 328	-	-	1,0000	-	-	1
Hapl 103	0,5000	-	-	0,5000	-	2	Hapl 329	-	-	1,0000	-	-	1
Hapl 104	-	-	0,6333	0,3667	-	30	Hapl 330	-	1,0000	-	-	-	1
Hapl 105	0,5000	-	-	0,5000	-	2	Hapl 331	-	0,3333	-	-	0,6667	3
Hapl 106	-	-	0,5000	0,5000	-	2	Hapl 332	-	1,0000	-	-	-	1
Hapl 107	0,3333	0,3333	-	0,3333	-	3	Hapl 333	-	1,0000	-	-	-	2
Hapl 108	-	-	0,5000	0,5000	-	4	Hapl 334	-	1,0000	-	-	-	1
Hapl 109	-	-	-	1,0000	-	1	Hapl 335	-	1,0000	-	-	-	1
Hapl 110	-	0,1500	0,0500	0,7000	0,1000	20	Hapl 336	-	1,0000	-	-	-	3
Hapl 111	-	-	0,5000	0,5000	-	2	Hapl 337	-	1,0000	-	-	-	1
Hapl 112	0,3333	-	0,6667	-	-	3	Hapl 338	-	1,0000	-	-	-	1
Hapl 113	0,2000	0,2000	0,2000	-	0,4000	5	Hapl 339	-	1,0000	-	-	-	1
Hapl 114	-	-	1,0000	-	-	1	Hapl 340	-	1,0000	-	-	-	1
Hapl 115	-	-	1,0000	-	-	1	Hapl 341	-	1,0000	-	-	-	1
Hapl 116	0,2000	0,2000	0,2000	-	0,4000	5	Hapl 342	-	1,0000	-	-	-	1
Hapl 117	-	0,5000	0,2500	0,2500	-	4	Hapl 343	-	0,6667	-	-	0,3333	3
Hapl 118	0,2500	0,5000	0,2500	-	-	4	Hapl 344	-	1,0000	-	-	-	1
Hapl 119	0,5000	-	0,5000	-	-	2	Hapl 345	-	1,0000	-	-	-	1
Hapl 120	-	-	0,7500	0,2500	-	4	Hapl 346	-	1,0000	-	-	-	1
Hapl 121	-	-	1,0000	-	-	1	Hapl 347	-	1,0000	-	-	-	2
Hapl 122	-	-	1,0000	-	-	1	Hapl 348	-	1,0000	-	-	-	1
Hapl 123	0,6667	0,3333	-	-	-	3	Hapl 349	-	1,0000	-	-	-	1
Hapl 124	1,0000	-	-	-	-	1	Hapl 350	-	1,0000	-	-	-	1
Hapl 125	0,2500	0,7500	-	-	-	4	Hapl 351	-	1,0000	-	-	-	2
Hapl 126	1,0000	-	-	-	-	1	Hapl 352	-	1,0000	-	-	-	1
Hapl 127	0,5000	0,5000	-	-	-	4	Hapl 353	-	1,0000	-	-	-	1
Hapl 128	0,1667	0,8333	-	-	-	6	Hapl 354	-	1,0000	-	-	-	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 129	0,3846	0,2308	-	0,0769	0,3077	13	Hapl 355	-	1,0000	-	-	-	1
Hapl 130	1,0000	-	-	-	-	1	Hapl 356	-	1,0000	-	-	-	1
Hapl 131	0,6250	0,1250	0,1250	-	0,1250	8	Hapl 357	-	1,0000	-	-	-	1
Hapl 132	0,1250	0,3750	0,1250	-	0,3750	8	Hapl 358	-	0,5000	-	-	0,5000	2
Hapl 133	0,7500	0,2500	-	-	-	4	Hapl 359	-	-	-	-	1,0000	2
Hapl 134	0,3333	0,3333	0,3333	-	-	6	Hapl 360	-	-	-	-	1,0000	1
Hapl 135	1,0000	-	-	-	-	1	Hapl 361	-	-	-	-	1,0000	1
Hapl 136	0,5000	-	-	-	0,5000	2	Hapl 362	-	-	-	-	1,0000	1
Hapl 137	1,0000	-	-	-	-	1	Hapl 363	-	-	-	-	1,0000	1
Hapl 138	1,0000	-	-	-	-	1	Hapl 364	-	-	-	-	1,0000	1
Hapl 139	1,0000	-	-	-	-	1	Hapl 365	-	-	-	-	1,0000	1
Hapl 140	1,0000	-	-	-	-	1	Hapl 366	-	-	-	0,5000	0,5000	2
Hapl 141	1,0000	-	-	-	-	1	Hapl 367	-	-	-	-	1,0000	1
Hapl 142	1,0000	-	-	-	-	1	Hapl 368	-	-	-	-	1,0000	1
Hapl 143	1,0000	-	-	-	-	1	Hapl 369	-	-	-	-	1,0000	1
Hapl 144	1,0000	-	-	-	-	1	Hapl 370	-	-	-	-	1,0000	1
Hapl 145	1,0000	-	-	-	-	1	Hapl 371	-	-	-	-	1,0000	1
Hapl 146	1,0000	-	-	-	-	1	Hapl 372	-	-	-	-	1,0000	1
Hapl 147	1,0000	-	-	-	-	1	Hapl 373	-	-	-	-	1,0000	1
Hapl 148	1,0000	-	-	-	-	1	Hapl 374	-	-	-	-	1,0000	1
Hapl 149	0,1429	0,5714	-	0,2857	-	7	Hapl 375	-	-	-	-	1,0000	1
Hapl 150	0,3333	0,3333	0,3333	-	-	3	Hapl 376	-	-	-	-	1,0000	1
Hapl 151	-	-	1,0000	-	-	1	Hapl 377	-	-	-	-	1,0000	1
Hapl 152	-	-	0,5000	0,5000	-	4	Hapl 378	-	-	-	1,0000	-	1
Hapl 153	0,3333	-	0,3333	0,3333	-	3	Hapl 379	-	-	-	1,0000	-	1
Hapl 154	-	-	0,5000	0,5000	-	2	Hapl 380	-	-	-	1,0000	-	1
Hapl 155	-	-	1,0000	-	-	1	Hapl 381	-	-	-	0,5000	0,5000	2
Hapl 156	0,3333	-	0,3333	-	0,3333	3	Hapl 382	-	-	-	1,0000	-	1
Hapl 157	-	0,5000	0,5000	-	-	4	Hapl 383	-	-	-	1,0000	-	1
Hapl 158	0,2000	0,6000	0,2000	-	-	5	Hapl 384	-	-	-	1,0000	-	1
Hapl 159	-	0,5000	0,5000	-	-	2	Hapl 385	-	-	-	1,0000	-	1
Hapl 160	-	0,3333	0,6667	-	-	3	Hapl 386	-	-	-	1,0000	-	1
Hapl 161	-	-	0,5000	-	0,5000	2	Hapl 387	-	-	-	1,0000	-	1
Hapl 162	-	-	1,0000	-	-	1	Hapl 388	-	-	-	1,0000	-	1
Hapl 163	-	-	1,0000	-	-	1	Hapl 389	-	-	-	1,0000	-	1
Hapl 164	-	-	1,0000	-	-	1	Hapl 390	-	-	-	1,0000	-	1
Hapl 165	-	-	1,0000	-	-	1	Hapl 391	-	-	-	-	1,0000	1
Hapl 166	-	-	1,0000	-	-	1	Hapl 392	-	-	-	-	1,0000	1
Hapl 167	-	-	1,0000	-	-	1	Hapl 393	-	-	-	-	1,0000	1
Hapl 168	-	-	1,0000	-	-	1	Hapl 394	-	-	-	-	1,0000	1
Hapl 169	-	-	1,0000	-	-	2	Hapl 395	-	-	-	-	1,0000	1
Hapl 170	-	-	1,0000	-	-	1	Hapl 396	-	-	-	-	1,0000	1
Hapl 171	-	-	1,0000	-	-	1	Hapl 397	-	-	-	-	1,0000	1
Hapl 172	-	-	0,6667	-	0,3333	3	Hapl 398	-	-	-	-	1,0000	1
Hapl 173	-	-	0,3333	0,6667	-	3	Hapl 399	-	-	-	-	1,0000	1
Hapl 174	-	-	1,0000	-	-	3	Hapl 400	-	-	-	1,0000	-	1
Hapl 175	-	-	1,0000	-	-	1	Hapl 401	-	-	-	1,0000	-	1
Hapl 176	-	-	0,5000	-	0,5000	2	Hapl 402	-	-	-	1,0000	-	1
Hapl 177	-	0,3077	0,0769	0,6154	-	13	Hapl 403	-	-	-	1,0000	-	1
Hapl 178	-	-	1,0000	-	-	1	Hapl 404	-	-	-	1,0000	-	1
Hapl 179	-	-	1,0000	-	-	6	Hapl 405	-	-	-	1,0000	-	1
Hapl 180	-	-	1,0000	-	-	1	Hapl 406	-	-	-	1,0000	-	1
Hapl 181	-	-	1,0000	-	-	1	Hapl 407	-	-	-	1,0000	-	1
Hapl 182	-	-	1,0000	-	-	1	Hapl 408	-	-	-	1,0000	-	1
Hapl 183	-	-	1,0000	-	-	1	Hapl 409	-	-	-	1,0000	-	1
Hapl 184	-	-	-	-	1,0000	1	Hapl 410	-	-	-	1,0000	-	1
Hapl 185	0,2000	0,2000	-	0,4000	0,2000	5	Hapl 411	-	-	-	1,0000	-	1
Hapl 186	-	-	-	-	1,0000	1	Hapl 412	-	-	-	1,0000	-	1
Hapl 187	-	-	-	-	1,0000	1	Hapl 413	-	-	-	1,0000	-	1
Hapl 188	-	-	-	-	1,0000	1	Hapl 414	-	-	-	1,0000	-	1
Hapl 189	-	-	0,5000	-	0,5000	2	Hapl 415	-	-	-	1,0000	-	1
Hapl 190	-	-	-	-	1,0000	1	Hapl 416	-	-	-	1,0000	-	1
Hapl 191	-	-	-	-	1,0000	1	Hapl 417	1,0000	-	-	-	-	1
Hapl 192	-	-	-	-	1,0000	1	Hapl 418	1,0000	-	-	-	-	1
Hapl 193	-	-	-	-	1,0000	1	Hapl 419	1,0000	-	-	-	-	1
Hapl 194	-	-	-	-	1,0000	1	Hapl 420	1,0000	-	-	-	-	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 195	-	0,5000	-	-	0,5000	2	Hapl 421	1,0000	-	-	-	-	2
Hapl 196	-	-	-	-	1,0000	1	Hapl 422	1,0000	-	-	-	-	1
Hapl 197	-	-	-	-	1,0000	1	Hapl 423	1,0000	-	-	-	-	1
Hapl 198	-	1,0000	-	-	-	1	Hapl 424	1,0000	-	-	-	-	1
Hapl 199	-	1,0000	-	-	-	1	Hapl 425	1,0000	-	-	-	-	1
Hapl 200	-	1,0000	-	-	-	1	Hapl 426	1,0000	-	-	-	-	1
Hapl 201	-	1,0000	-	-	-	1	Hapl 427	1,0000	-	-	-	-	1
Hapl 202	-	1,0000	-	-	-	1	Hapl 428	1,0000	-	-	-	-	1
Hapl 203	-	1,0000	-	-	-	1	Hapl 429	1,0000	-	-	-	-	1
Hapl 204	0,1429	0,8571	-	-	-	7	Hapl 430	1,0000	-	-	-	-	1
Hapl 205	0,1429	0,4286	0,1429	0,2857	-	7	Hapl 431	1,0000	-	-	-	-	1
Hapl 206	-	1,0000	-	-	-	1	Hapl 432	1,0000	-	-	-	-	1
Hapl 207	-	0,5000	0,5000	-	-	4	Hapl 433	1,0000	-	-	-	-	1
Hapl 208	-	1,0000	-	-	-	1	Hapl 434	-	1,0000	-	-	-	1
Hapl 209	-	1,0000	-	-	-	1	Hapl 435	-	1,0000	-	-	-	1
Hapl 210	0,2222	0,4444	-	0,2222	0,1111	9	Hapl 436	-	1,0000	-	-	-	1
Hapl 211	0,5455	0,2727	-	-	0,1818	11	Hapl 437	-	1,0000	-	-	-	2
Hapl 212	-	0,6000	-	0,2000	0,2000	5	Hapl 438	-	1,0000	-	-	-	1
Hapl 213	-	1,0000	-	-	-	1	Hapl 439	-	1,0000	-	-	-	1
Hapl 214	-	1,0000	-	-	-	1	Hapl 440	-	1,0000	-	-	-	1
Hapl 215	-	1,0000	-	-	-	1	Hapl 441	-	1,0000	-	-	-	1
Hapl 216	-	1,0000	-	-	-	1	Hapl 442	-	1,0000	-	-	-	1
Hapl 217	-	1,0000	-	-	-	1	Hapl 443	-	1,0000	-	-	-	1
Hapl 218	1,0000	-	-	-	-	2	Hapl 444	-	1,0000	-	-	-	1
Hapl 219	0,6000	0,4000	-	-	-	5	Hapl 445	-	1,0000	-	-	-	1
Hapl 220	1,0000	-	-	-	-	1	Hapl 446	-	1,0000	-	-	-	1
Hapl 221	1,0000	-	-	-	-	1	Hapl 447	-	1,0000	-	-	-	1
Hapl 222	1,0000	-	-	-	-	1	Hapl 448	-	1,0000	-	-	-	1
Hapl 223	1,0000	-	-	-	-	1	Hapl 449	-	1,0000	-	-	-	1
Hapl 224	1,0000	-	-	-	-	1	Hapl 450	-	1,0000	-	-	-	1
Hapl 225	1,0000	-	-	-	-	1	Hapl 451	-	1,0000	-	-	-	1
Hapl 226	1,0000	-	-	-	-	1	Hapl 452	-	1,0000	-	-	-	1

Tabela S7 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região 5'URR do gene *SLC24A5* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	-	0,2434	0,5007	0,1826	0,0733	723	Hapl 047	-	-	-	-	1,0000	1
Hapl 002	0,3106	0,1174	0,1500	0,2287	0,1933	3187	Hapl 048	-	0,5000	-	-	0,5000	2
Hapl 003	0,0071	0,0643	0,3571	0,0786	0,4929	140	Hapl 049	-	1,0000	-	-	-	1
Hapl 004	-	0,8820	-	-	0,1180	178	Hapl 050	-	1,0000	-	-	-	1
Hapl 005	-	0,7113	0,0069	0,1787	0,1031	291	Hapl 051	-	1,0000	-	-	-	1
Hapl 006	-	0,9060	-	-	0,0940	117	Hapl 052	-	1,0000	-	-	-	1
Hapl 007	-	0,9722	-	-	0,0278	36	Hapl 053	-	1,0000	-	-	-	4
Hapl 008	-	0,9091	-	-	0,0909	33	Hapl 054	1,0000	-	-	-	-	1
Hapl 009	-	1,0000	-	-	-	12	Hapl 055	1,0000	-	-	-	-	4
Hapl 010	-	1,0000	-	-	-	1	Hapl 056	-	-	-	1,0000	-	2
Hapl 011	-	0,9500	-	-	0,0500	20	Hapl 057	-	1,0000	-	-	-	1
Hapl 012	-	1,0000	-	-	-	1	Hapl 058	-	1,0000	-	-	-	2
Hapl 013	-	1,0000	-	-	-	1	Hapl 059	-	1,0000	-	-	-	1
Hapl 014	-	0,9600	-	-	0,0400	25	Hapl 060	-	1,0000	-	-	-	2
Hapl 015	-	0,8889	-	-	0,1111	9	Hapl 061	-	1,0000	-	-	-	1
Hapl 016	-	1,0000	-	-	-	1	Hapl 062	-	1,0000	-	-	-	3
Hapl 017	-	0,7500	-	-	0,2500	4	Hapl 063	-	1,0000	-	-	-	1
Hapl 018	-	-	-	-	1,0000	1	Hapl 064	-	1,0000	-	-	-	1
Hapl 019	-	0,0833	0,5833	0,2917	0,0417	24	Hapl 065	-	1,0000	-	-	-	1
Hapl 020	-	-	0,5135	0,4865	-	37	Hapl 066	1,0000	-	-	-	-	1
Hapl 021	0,2500	-	0,5000	0,2500	-	4	Hapl 067	-	-	-	1,0000	-	2
Hapl 022	0,7500	-	-	0,2500	-	4	Hapl 068	-	-	-	1,0000	-	1
Hapl 023	-	-	0,5000	0,5000	-	2	Hapl 069	-	-	1,0000	-	-	1
Hapl 024	-	-	-	1,0000	-	3	Hapl 070	-	-	1,0000	-	-	1
Hapl 025	-	-	-	1,0000	-	1	Hapl 071	-	-	1,0000	-	-	1
Hapl 026	-	-	-	1,0000	-	1	Hapl 072	-	-	1,0000	-	-	2
Hapl 027	-	-	-	1,0000	-	6	Hapl 073	-	-	1,0000	-	-	1
Hapl 028	-	-	-	1,0000	-	1	Hapl 074	-	-	1,0000	-	-	1
Hapl 029	-	-	1,0000	-	-	2	Hapl 075	-	1,0000	-	-	-	3
Hapl 030	-	-	1,0000	-	-	22	Hapl 076	-	1,0000	-	-	-	1
Hapl 031	-	-	1,0000	-	-	15	Hapl 077	-	1,0000	-	-	-	2
Hapl 032	-	-	1,0000	-	-	4	Hapl 078	-	1,0000	-	-	-	2
Hapl 033	-	-	1,0000	-	-	11	Hapl 079	-	1,0000	-	-	-	1
Hapl 034	-	-	1,0000	-	-	7	Hapl 080	-	1,0000	-	-	-	1
Hapl 035	-	-	1,0000	-	-	2	Hapl 081	-	1,0000	-	-	-	1
Hapl 036	0,6667	-	-	-	0,3333	3	Hapl 082	-	-	-	-	1,0000	1
Hapl 037	1,0000	-	-	-	-	2	Hapl 083	-	-	-	1,0000	-	2
Hapl 038	-	-	1,0000	-	-	1	Hapl 084	-	-	-	1,0000	-	2
Hapl 039	-	-	1,0000	-	-	1	Hapl 085	-	-	-	-	1,0000	1
Hapl 040	-	-	1,0000	-	-	2	Hapl 086	-	-	-	1,0000	-	1
Hapl 041	-	-	1,0000	-	-	1	Hapl 087	-	-	-	1,0000	-	1
Hapl 042	-	-	1,0000	-	-	2	Hapl 088	-	-	-	1,0000	-	2
Hapl 043	-	-	1,0000	-	-	1	Hapl 089	-	-	-	1,0000	-	1
Hapl 044	-	-	1,0000	-	-	1	Hapl 090	1,0000	-	-	-	-	1
Hapl 045	-	-	1,0000	-	-	1	Hapl 091	-	1,0000	-	-	-	1
Hapl 046	-	-	-	-	1,0000	1							

Tabela S8 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região 5'UTR do gene *SLC24A5* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	0,2010	0,2396	0,2012	0,1954	0,1628	5005
Hapl 002	-	-	-	-	1,0000	1

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 003	-	-	1,0000	-	-	1
Hapl 004	-	1,0000	-	-	-	1

Tabela S9 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região Codificante (Éxons e Introns) do gene *SLC24A5* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	-	0,2949	0,4290	0,1662	0,1099	373
Hapl 002	-	0,3641	0,3871	0,0230	0,2258	434
Hapl 003	-	0,0551	0,3937	0,0630	0,4882	127
Hapl 004	-	0,8889	-	-	0,1111	27
Hapl 005	-	1,0000	-	-	-	15
Hapl 006	-	0,9007	-	-	0,0993	151
Hapl 007	-	1,0000	-	-	-	2
Hapl 008	-	0,8837	-	-	0,1163	43
Hapl 009	-	0,7667	-	-	0,2333	30
Hapl 010	-	0,9643	-	-	0,0357	28
Hapl 011	-	0,8636	-	-	0,1364	22
Hapl 012	0,4552	0,0352	0,0045	0,3110	0,1941	1762
Hapl 013	-	0,8909	-	-	0,1091	55
Hapl 014	-	0,5000	-	-	0,5000	2
Hapl 015	-	1,0000	-	-	-	17
Hapl 016	-	0,9615	-	-	0,0385	26
Hapl 017	-	1,0000	-	-	-	2
Hapl 018	-	0,7500	-	-	0,2500	12
Hapl 019	-	0,8333	-	-	0,1667	42
Hapl 020	-	1,0000	-	-	-	14
Hapl 021	-	0,9000	-	-	0,1000	10
Hapl 022	-	1,0000	-	-	-	4
Hapl 023	-	1,0000	-	-	-	2
Hapl 024	-	0,9000	-	-	0,1000	10
Hapl 025	-	1,0000	-	-	-	1
Hapl 026	-	0,8000	-	-	0,2000	10
Hapl 027	-	0,9545	-	-	0,0455	22
Hapl 028	-	1,0000	-	-	-	7
Hapl 029	-	1,0000	-	-	-	2
Hapl 030	-	0,5000	0,5000	-	-	6
Hapl 031	-	1,0000	-	-	-	3
Hapl 032	-	1,0000	-	-	-	2
Hapl 033	-	0,9091	-	-	0,0909	11
Hapl 034	-	0,6000	-	-	0,4000	5
Hapl 035	-	1,0000	-	-	-	1
Hapl 036	-	0,9091	-	-	0,0909	11
Hapl 037	-	1,0000	-	-	-	1
Hapl 038	-	0,8333	-	-	0,1667	12
Hapl 039	-	1,0000	-	-	-	1
Hapl 040	-	1,0000	-	-	-	1
Hapl 041	-	0,8750	-	-	0,1250	8
Hapl 042	-	1,0000	-	-	-	2
Hapl 043	-	1,0000	-	-	-	1
Hapl 044	-	1,0000	-	-	-	4
Hapl 045	0,5246	0,0164	0,0164	0,2295	0,2131	61
Hapl 046	-	1,0000	-	-	-	5
Hapl 047	-	1,0000	-	-	-	5
Hapl 048	-	1,0000	-	-	-	13

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 261	1,0000	-	-	-	-	1
Hapl 262	1,0000	-	-	-	-	2
Hapl 263	1,0000	-	-	-	-	1
Hapl 264	1,0000	-	-	-	-	1
Hapl 265	0,6667	0,3333	-	-	-	3
Hapl 266	1,0000	-	-	-	-	1
Hapl 267	-	-	-	1,0000	-	1
Hapl 268	-	-	-	1,0000	-	1
Hapl 269	-	-	-	1,0000	-	1
Hapl 270	-	-	-	1,0000	-	1
Hapl 271	-	-	-	1,0000	-	2
Hapl 272	-	-	-	1,0000	-	4
Hapl 273	-	-	-	1,0000	-	1
Hapl 274	-	-	-	1,0000	-	1
Hapl 275	-	-	-	1,0000	-	2
Hapl 276	-	-	-	1,0000	-	1
Hapl 277	-	-	-	1,0000	-	1
Hapl 278	-	-	-	1,0000	-	1
Hapl 279	-	-	-	1,0000	-	1
Hapl 280	-	-	-	1,0000	-	1
Hapl 281	-	1,0000	-	-	-	1
Hapl 282	-	1,0000	-	-	-	2
Hapl 283	-	1,0000	-	-	-	1
Hapl 284	-	1,0000	-	-	-	1
Hapl 285	-	1,0000	-	-	-	2
Hapl 286	-	0,7500	-	-	0,2500	4
Hapl 287	-	1,0000	-	-	-	1
Hapl 288	-	1,0000	-	-	-	1
Hapl 289	-	1,0000	-	-	-	1
Hapl 290	-	1,0000	-	-	-	2
Hapl 291	-	1,0000	-	-	-	5
Hapl 292	-	1,0000	-	-	-	5
Hapl 293	-	1,0000	-	-	-	3
Hapl 294	-	0,7500	-	-	0,2500	4
Hapl 295	-	1,0000	-	-	-	1
Hapl 296	-	1,0000	-	-	-	3
Hapl 297	-	0,6667	-	-	0,3333	3
Hapl 298	-	1,0000	-	-	-	1
Hapl 299	-	1,0000	-	-	-	2
Hapl 300	-	1,0000	-	-	-	1
Hapl 301	-	1,0000	-	-	-	1
Hapl 302	-	1,0000	-	-	-	3
Hapl 303	-	1,0000	-	-	-	1
Hapl 304	-	1,0000	-	-	-	1
Hapl 305	-	1,0000	-	-	-	1
Hapl 306	-	1,0000	-	-	-	1
Hapl 307	-	1,0000	-	-	-	3
Hapl 308	-	1,0000	-	-	-	1

Continuação													
Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 049	-	1,0000	-	-	-	1	Hapl 309	-	1,0000	-	-	-	3
Hapl 050	-	0,8500	-	-	0,1500	20	Hapl 310	-	1,0000	-	-	-	1
Hapl 051	-	1,0000	-	-	-	1	Hapl 311	-	1,0000	-	-	-	1
Hapl 052	-	0,8750	-	-	0,1250	16	Hapl 312	-	1,0000	-	-	-	1
Hapl 053	-	0,8571	-	-	0,1429	7	Hapl 313	-	1,0000	-	-	-	1
Hapl 054	-	0,8846	-	-	0,1154	26	Hapl 314	-	1,0000	-	-	-	1
Hapl 055	-	0,9167	-	-	0,0833	12	Hapl 315	-	1,0000	-	-	-	1
Hapl 056	-	1,0000	-	-	-	6	Hapl 316	1,0000	-	-	-	-	1
Hapl 057	-	1,0000	-	-	-	1	Hapl 317	1,0000	-	-	-	-	1
Hapl 058	-	1,0000	-	-	-	1	Hapl 318	1,0000	-	-	-	-	2
Hapl 059	0,4943	0,0230	0,0115	0,3333	0,1379	87	Hapl 319	0,6667	-	-	0,3333	-	6
Hapl 060	-	1,0000	-	-	-	1	Hapl 320	0,5000	-	-	-	0,5000	2
Hapl 061	-	1,0000	-	-	-	1	Hapl 321	1,0000	-	-	-	-	1
Hapl 062	-	1,0000	-	-	-	1	Hapl 322	1,0000	-	-	-	-	1
Hapl 063	-	0,6667	-	-	0,3333	3	Hapl 323	0,2500	-	-	0,7500	-	4
Hapl 064	-	0,6667	-	-	0,3333	3	Hapl 324	0,6667	-	-	-	0,3333	3
Hapl 065	-	1,0000	-	-	-	2	Hapl 325	1,0000	-	-	-	-	1
Hapl 066	-	0,8571	-	-	0,1429	7	Hapl 326	-	-	-	1,0000	-	1
Hapl 067	-	-	-	-	1,0000	1	Hapl 327	-	-	-	1,0000	-	2
Hapl 068	-	0,8750	-	-	0,1250	8	Hapl 328	-	-	-	1,0000	-	2
Hapl 069	-	-	-	-	1,0000	1	Hapl 329	-	-	-	1,0000	-	2
Hapl 070	-	-	-	-	1,0000	1	Hapl 330	-	-	-	1,0000	-	1
Hapl 071	-	-	-	-	1,0000	1	Hapl 331	-	-	-	1,0000	-	1
Hapl 072	-	-	-	-	1,0000	1	Hapl 332	-	-	-	1,0000	-	5
Hapl 073	-	-	-	-	1,0000	1	Hapl 333	-	-	-	1,0000	-	1
Hapl 074	-	-	0,5000	-	0,5000	2	Hapl 334	-	-	-	1,0000	-	2
Hapl 075	-	-	-	-	1,0000	1	Hapl 335	-	-	-	1,0000	-	1
Hapl 076	-	-	-	-	1,0000	1	Hapl 336	-	-	-	1,0000	-	2
Hapl 077	-	0,6667	-	-	0,3333	3	Hapl 337	-	-	-	1,0000	-	1
Hapl 078	0,6667	-	-	0,1111	0,2222	9	Hapl 338	-	-	-	1,0000	-	1
Hapl 079	-	-	-	-	1,0000	1	Hapl 339	-	-	-	1,0000	-	1
Hapl 080	-	-	-	-	1,0000	2	Hapl 340	-	-	-	1,0000	-	2
Hapl 081	-	-	-	-	1,0000	1	Hapl 341	-	-	1,0000	-	-	2
Hapl 082	0,7143	-	-	-	0,2857	7	Hapl 342	-	-	1,0000	-	-	1
Hapl 083	-	0,5000	-	-	0,5000	2	Hapl 343	-	-	1,0000	-	-	1
Hapl 084	-	-	-	-	1,0000	1	Hapl 344	-	-	1,0000	-	-	1
Hapl 085	-	-	-	-	1,0000	2	Hapl 345	-	-	1,0000	-	-	1
Hapl 086	-	-	-	-	1,0000	1	Hapl 346	-	-	1,0000	-	-	1
Hapl 087	-	-	-	-	1,0000	1	Hapl 347	-	-	1,0000	-	-	4
Hapl 088	0,0073	-	0,6850	0,1282	0,1795	273	Hapl 348	-	-	1,0000	-	-	1
Hapl 089	-	-	-	-	1,0000	1	Hapl 349	-	-	1,0000	-	-	1
Hapl 090	-	0,5000	-	-	0,5000	2	Hapl 350	-	-	1,0000	-	-	1
Hapl 091	-	-	-	-	1,0000	1	Hapl 351	-	-	1,0000	-	-	1
Hapl 092	0,6316	-	-	0,1053	0,2632	19	Hapl 352	-	-	1,0000	-	-	1
Hapl 093	-	-	-	-	1,0000	1	Hapl 353	-	-	1,0000	-	-	1
Hapl 094	-	-	-	-	1,0000	1	Hapl 354	-	-	1,0000	-	-	2
Hapl 095	-	-	0,4375	0,5625	-	32	Hapl 355	-	-	1,0000	-	-	3
Hapl 096	-	-	-	1,0000	-	2	Hapl 356	-	-	1,0000	-	-	1
Hapl 097	-	-	-	1,0000	-	2	Hapl 357	-	-	1,0000	-	-	1
Hapl 098	0,6000	-	-	0,2000	0,2000	5	Hapl 358	-	-	1,0000	-	-	1
Hapl 099	-	-	0,5000	0,5000	-	34	Hapl 359	-	-	1,0000	-	-	1
Hapl 100	-	0,0235	0,7059	0,2706	-	85	Hapl 360	-	-	1,0000	-	-	1
Hapl 101	-	0,0286	-	0,9714	-	35	Hapl 361	-	-	1,0000	-	-	1
Hapl 102	-	-	-	1,0000	-	5	Hapl 362	-	-	1,0000	-	-	1
Hapl 103	-	-	0,0714	0,9286	-	14	Hapl 363	-	-	1,0000	-	-	1
Hapl 104	-	-	-	1,0000	-	4	Hapl 364	-	-	1,0000	-	-	1
Hapl 105	-	-	-	1,0000	-	4	Hapl 365	-	-	1,0000	-	-	1
Hapl 106	-	-	-	1,0000	-	5	Hapl 366	-	-	1,0000	-	-	1
Hapl 107	-	-	-	1,0000	-	1	Hapl 367	-	-	0,5000	-	0,5000	2
Hapl 108	-	-	-	1,0000	-	1	Hapl 368	-	-	1,0000	-	-	1
Hapl 109	-	-	-	1,0000	-	1	Hapl 369	-	-	1,0000	-	-	1
Hapl 110	-	-	-	1,0000	-	2	Hapl 370	-	-	1,0000	-	-	1
Hapl 111	-	-	-	1,0000	-	1	Hapl 371	-	-	1,0000	-	-	1
Hapl 112	-	-	-	1,0000	-	2	Hapl 372	-	-	1,0000	-	-	1
Hapl 113	0,8667	-	-	0,0667	0,0667	15	Hapl 373	-	-	1,0000	-	-	1
Hapl 114	-	-	-	1,0000	-	1	Hapl 374	-	-	1,0000	-	-	1

Continuação													
Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 115	-	-	-	1,0000	-	1	Hapl 375	-	-	1,0000	-	-	1
Hapl 116	0,5000	-	-	0,5000	-	2	Hapl 376	-	1,0000	-	-	-	1
Hapl 117	-	-	-	1,0000	-	7	Hapl 377	-	1,0000	-	-	-	1
Hapl 118	-	-	-	1,0000	-	1	Hapl 378	-	1,0000	-	-	-	1
Hapl 119	-	-	-	1,0000	-	1	Hapl 379	-	1,0000	-	-	-	2
Hapl 120	-	-	-	1,0000	-	1	Hapl 380	-	1,0000	-	-	-	1
Hapl 121	-	-	-	1,0000	-	1	Hapl 381	-	1,0000	-	-	-	1
Hapl 122	-	-	-	1,0000	-	1	Hapl 382	-	1,0000	-	-	-	1
Hapl 123	-	-	-	1,0000	-	1	Hapl 383	-	1,0000	-	-	-	1
Hapl 124	0,5455	-	-	0,3636	0,0909	11	Hapl 384	-	1,0000	-	-	-	1
Hapl 125	-	-	-	1,0000	-	1	Hapl 385	-	1,0000	-	-	-	1
Hapl 126	-	-	0,8571	0,1429	-	7	Hapl 386	-	1,0000	-	-	-	1
Hapl 127	-	-	-	0,9000	0,1000	10	Hapl 387	-	1,0000	-	-	-	1
Hapl 128	-	-	-	1,0000	-	1	Hapl 388	-	1,0000	-	-	-	1
Hapl 129	-	-	-	1,0000	-	1	Hapl 389	-	1,0000	-	-	-	1
Hapl 130	0,5000	-	-	0,5000	-	2	Hapl 390	-	1,0000	-	-	-	1
Hapl 131	-	-	1,0000	-	-	9	Hapl 391	-	1,0000	-	-	-	3
Hapl 132	-	-	1,0000	-	-	1	Hapl 392	-	1,0000	-	-	-	1
Hapl 133	-	-	1,0000	-	-	97	Hapl 393	-	1,0000	-	-	-	1
Hapl 134	-	-	1,0000	-	-	5	Hapl 394	-	1,0000	-	-	-	1
Hapl 135	-	-	1,0000	-	-	1	Hapl 395	-	1,0000	-	-	-	1
Hapl 136	-	-	1,0000	-	-	13	Hapl 396	-	1,0000	-	-	-	1
Hapl 137	-	-	1,0000	-	-	2	Hapl 397	-	1,0000	-	-	-	2
Hapl 138	-	-	1,0000	-	-	4	Hapl 398	-	1,0000	-	-	-	1
Hapl 139	-	-	1,0000	-	-	6	Hapl 399	-	1,0000	-	-	-	1
Hapl 140	-	-	1,0000	-	-	1	Hapl 400	-	1,0000	-	-	-	1
Hapl 141	-	-	1,0000	-	-	2	Hapl 401	-	1,0000	-	-	-	1
Hapl 142	-	-	1,0000	-	-	2	Hapl 402	-	1,0000	-	-	-	1
Hapl 143	-	-	1,0000	-	-	2	Hapl 403	-	1,0000	-	-	-	1
Hapl 144	-	-	1,0000	-	-	1	Hapl 404	-	1,0000	-	-	-	1
Hapl 145	-	-	1,0000	-	-	1	Hapl 405	-	1,0000	-	-	-	1
Hapl 146	-	-	1,0000	-	-	9	Hapl 406	-	1,0000	-	-	-	1
Hapl 147	-	-	0,6667	-	0,3333	3	Hapl 407	-	1,0000	-	-	-	1
Hapl 148	-	-	1,0000	-	-	1	Hapl 408	-	1,0000	-	-	-	1
Hapl 149	-	-	1,0000	-	-	2	Hapl 409	-	1,0000	-	-	-	1
Hapl 150	-	0,1250	0,6250	-	0,2500	16	Hapl 410	-	1,0000	-	-	-	1
Hapl 151	-	-	1,0000	-	-	1	Hapl 411	-	1,0000	-	-	-	1
Hapl 152	-	-	1,0000	-	-	1	Hapl 412	-	1,0000	-	-	-	1
Hapl 153	-	-	1,0000	-	-	21	Hapl 413	-	1,0000	-	-	-	2
Hapl 154	-	-	0,6471	-	0,3529	17	Hapl 414	-	1,0000	-	-	-	1
Hapl 155	-	-	1,0000	-	-	4	Hapl 415	-	1,0000	-	-	-	2
Hapl 156	-	-	1,0000	-	-	4	Hapl 416	-	1,0000	-	-	-	1
Hapl 157	-	-	1,0000	-	-	9	Hapl 417	-	1,0000	-	-	-	1
Hapl 158	-	-	1,0000	-	-	1	Hapl 418	-	1,0000	-	-	-	1
Hapl 159	-	-	1,0000	-	-	1	Hapl 419	-	1,0000	-	-	-	1
Hapl 160	0,5000	-	-	-	0,5000	2	Hapl 420	-	1,0000	-	-	-	1
Hapl 161	0,8000	-	-	0,1333	0,0667	15	Hapl 421	-	1,0000	-	-	-	1
Hapl 162	1,0000	-	-	-	-	1	Hapl 422	-	1,0000	-	-	-	1
Hapl 163	0,6667	-	-	0,3333	-	3	Hapl 423	-	1,0000	-	-	-	1
Hapl 164	1,0000	-	-	-	-	1	Hapl 424	-	1,0000	-	-	-	1
Hapl 165	1,0000	-	-	-	-	2	Hapl 425	-	1,0000	-	-	-	1
Hapl 166	1,0000	-	-	-	-	1	Hapl 426	-	1,0000	-	-	-	1
Hapl 167	1,0000	-	-	-	-	1	Hapl 427	-	1,0000	-	-	-	1
Hapl 168	0,2857	-	-	0,4286	0,2857	7	Hapl 428	-	-	-	-	1,0000	1
Hapl 169	1,0000	-	-	-	-	1	Hapl 429	-	-	-	-	1,0000	1
Hapl 170	1,0000	-	-	-	-	1	Hapl 430	-	-	-	-	1,0000	1
Hapl 171	1,0000	-	-	-	-	1	Hapl 431	-	-	-	-	1,0000	1
Hapl 172	1,0000	-	-	-	-	1	Hapl 432	-	-	-	-	1,0000	1
Hapl 173	1,0000	-	-	-	-	1	Hapl 433	-	-	-	-	1,0000	1
Hapl 174	-	-	1,0000	-	-	1	Hapl 434	-	-	-	-	1,0000	1
Hapl 175	-	-	1,0000	-	-	7	Hapl 435	-	-	-	-	1,0000	1
Hapl 176	-	-	1,0000	-	-	1	Hapl 436	-	-	-	-	1,0000	1
Hapl 177	-	-	1,0000	-	-	1	Hapl 437	-	0,5000	-	-	0,5000	2
Hapl 178	-	-	1,0000	-	-	1	Hapl 438	-	-	-	-	1,0000	1
Hapl 179	-	-	1,0000	-	-	2	Hapl 439	-	-	-	-	1,0000	1
Hapl 180	-	-	1,0000	-	-	1	Hapl 440	-	-	-	-	1,0000	1

Continuação													
Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 181	-	-	1,0000	-	-	2	Hapl 441	-	-	-	-	1,0000	1
Hapl 182	-	-	1,0000	-	-	5	Hapl 442	-	-	-	-	1,0000	1
Hapl 183	-	-	1,0000	-	-	2	Hapl 443	-	-	-	-	1,0000	2
Hapl 184	-	-	1,0000	-	-	2	Hapl 444	-	-	-	-	1,0000	1
Hapl 185	-	-	1,0000	-	-	4	Hapl 445	-	-	-	-	1,0000	1
Hapl 186	-	-	0,7500	-	0,2500	4	Hapl 446	-	-	-	-	1,0000	1
Hapl 187	-	-	1,0000	-	-	1	Hapl 447	-	-	-	-	1,0000	1
Hapl 188	-	-	1,0000	-	-	1	Hapl 448	-	-	-	-	1,0000	1
Hapl 189	-	-	1,0000	-	-	1	Hapl 449	-	-	-	-	1,0000	1
Hapl 190	-	-	1,0000	-	-	2	Hapl 450	-	-	-	1,0000	-	1
Hapl 191	-	-	1,0000	-	-	1	Hapl 451	-	-	-	1,0000	-	1
Hapl 192	-	0,5000	0,5000	-	-	2	Hapl 452	-	-	-	1,0000	-	2
Hapl 193	-	-	1,0000	-	-	2	Hapl 453	-	-	-	1,0000	-	1
Hapl 194	-	-	1,0000	-	-	1	Hapl 454	-	-	-	1,0000	-	2
Hapl 195	-	-	1,0000	-	-	1	Hapl 455	-	-	-	1,0000	-	1
Hapl 196	-	-	1,0000	-	-	1	Hapl 456	-	-	-	1,0000	-	1
Hapl 197	-	-	1,0000	-	-	1	Hapl 457	-	-	-	1,0000	-	1
Hapl 198	-	-	1,0000	-	-	1	Hapl 458	-	-	-	1,0000	-	2
Hapl 199	-	-	1,0000	-	-	1	Hapl 459	-	-	-	1,0000	-	1
Hapl 200	-	-	1,0000	-	-	1	Hapl 460	-	-	-	1,0000	-	1
Hapl 201	-	-	1,0000	-	-	1	Hapl 461	-	-	-	1,0000	-	1
Hapl 202	-	-	1,0000	-	-	1	Hapl 462	-	-	-	1,0000	-	1
Hapl 203	-	-	1,0000	-	-	1	Hapl 463	-	-	-	-	1,0000	4
Hapl 204	-	-	1,0000	-	-	1	Hapl 464	-	-	-	-	1,0000	1
Hapl 205	-	-	1,0000	-	-	1	Hapl 465	-	-	-	-	1,0000	1
Hapl 206	-	-	1,0000	-	-	1	Hapl 466	-	-	-	-	1,0000	1
Hapl 207	-	-	1,0000	-	-	1	Hapl 467	-	-	-	-	1,0000	1
Hapl 208	-	-	1,0000	-	-	1	Hapl 468	-	-	-	-	1,0000	2
Hapl 209	-	-	1,0000	-	-	2	Hapl 469	-	-	-	-	1,0000	1
Hapl 210	-	-	0,7500	-	0,2500	4	Hapl 470	-	-	-	-	1,0000	1
Hapl 211	-	-	1,0000	-	-	1	Hapl 471	-	-	-	-	1,0000	1
Hapl 212	-	-	1,0000	-	-	1	Hapl 472	-	-	-	-	1,0000	2
Hapl 213	-	-	1,0000	-	-	1	Hapl 473	-	-	-	-	1,0000	1
Hapl 214	-	-	1,0000	-	-	1	Hapl 474	-	-	-	-	1,0000	1
Hapl 215	-	-	-	-	1,0000	2	Hapl 475	-	-	-	-	1,0000	1
Hapl 216	-	-	-	-	1,0000	1	Hapl 476	-	-	-	-	1,0000	1
Hapl 217	-	-	-	-	1,0000	2	Hapl 477	0,5000	-	-	-	0,5000	2
Hapl 218	-	-	-	-	1,0000	1	Hapl 478	-	-	-	1,0000	-	1
Hapl 219	-	-	-	-	1,0000	2	Hapl 479	-	-	-	1,0000	-	1
Hapl 220	-	-	-	-	1,0000	1	Hapl 480	-	-	-	1,0000	-	1
Hapl 221	-	-	-	-	1,0000	3	Hapl 481	-	-	-	1,0000	-	2
Hapl 222	-	-	-	-	1,0000	1	Hapl 482	-	-	-	1,0000	-	2
Hapl 223	-	-	-	-	1,0000	1	Hapl 483	-	-	-	1,0000	-	1
Hapl 224	-	-	-	-	1,0000	1	Hapl 484	-	-	-	1,0000	-	1
Hapl 225	-	-	-	-	1,0000	1	Hapl 485	-	-	-	1,0000	-	1
Hapl 226	-	1,0000	-	-	-	2	Hapl 486	-	-	-	1,0000	-	1
Hapl 227	-	1,0000	-	-	-	5	Hapl 487	-	-	-	1,0000	-	1
Hapl 228	-	1,0000	-	-	-	2	Hapl 488	-	-	-	1,0000	-	2
Hapl 229	-	1,0000	-	-	-	1	Hapl 489	-	-	-	1,0000	-	1
Hapl 230	-	1,0000	-	-	-	1	Hapl 490	-	-	-	1,0000	-	1
Hapl 231	-	1,0000	-	-	-	1	Hapl 491	-	-	-	1,0000	-	1
Hapl 232	-	1,0000	-	-	-	1	Hapl 492	-	-	-	1,0000	-	1
Hapl 233	-	1,0000	-	-	-	3	Hapl 493	-	-	-	1,0000	-	2
Hapl 234	-	1,0000	-	-	-	1	Hapl 494	-	-	-	1,0000	-	1
Hapl 235	-	1,0000	-	-	-	1	Hapl 495	-	-	-	1,0000	-	1
Hapl 236	-	1,0000	-	-	-	1	Hapl 496	-	-	-	1,0000	-	1
Hapl 237	-	1,0000	-	-	-	1	Hapl 497	-	-	-	1,0000	-	1
Hapl 238	-	1,0000	-	-	-	1	Hapl 498	-	-	-	1,0000	-	1
Hapl 239	-	1,0000	-	-	-	1	Hapl 499	-	-	-	1,0000	-	1
Hapl 240	-	1,0000	-	-	-	3	Hapl 500	1,0000	-	-	-	-	1
Hapl 241	-	1,0000	-	-	-	1	Hapl 501	1,0000	-	-	-	-	1
Hapl 242	-	1,0000	-	-	-	1	Hapl 502	1,0000	-	-	-	-	1
Hapl 243	-	1,0000	-	-	-	1	Hapl 503	1,0000	-	-	-	-	1
Hapl 244	-	1,0000	-	-	-	2	Hapl 504	1,0000	-	-	-	-	1
Hapl 245	-	1,0000	-	-	-	1	Hapl 505	1,0000	-	-	-	-	1
Hapl 246	1,0000	-	-	-	-	1	Hapl 506	1,0000	-	-	-	-	1

Continuação													
Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 247	0,6667	-	-	-	0,3333	3	Hapl 507	1,0000	-	-	-	-	1
Hapl 248	0,7500	-	-	0,2500	-	4	Hapl 508	-	1,0000	-	-	-	1
Hapl 249	1,0000	-	-	-	-	1	Hapl 509	-	1,0000	-	-	-	1
Hapl 250	1,0000	-	-	-	-	1	Hapl 510	-	1,0000	-	-	-	1
Hapl 251	1,0000	-	-	-	-	1	Hapl 511	-	1,0000	-	-	-	1
Hapl 252	1,0000	-	-	-	-	1	Hapl 512	-	1,0000	-	-	-	1
Hapl 253	1,0000	-	-	-	-	1	Hapl 513	-	1,0000	-	-	-	1
Hapl 254	1,0000	-	-	-	-	1	Hapl 514	-	1,0000	-	-	-	1
Hapl 255	1,0000	-	-	-	-	1	Hapl 515	-	1,0000	-	-	-	2
Hapl 256	1,0000	-	-	-	-	1	Hapl 516	-	1,0000	-	-	-	1
Hapl 257	1,0000	-	-	-	-	1	Hapl 517	-	1,0000	-	-	-	1
Hapl 258	1,0000	-	-	-	-	3	Hapl 518	-	1,0000	-	-	-	1
Hapl 259	1,0000	-	-	-	-	1	Hapl 519	-	1,0000	-	-	-	1
Hapl 260	1,0000	-	-	-	-	1							

Tabela S10 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região Codificante (somente Éxons) do gene *SLC24A5* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	0,0011	0,3958	0,3550	0,1099	0,1382	2749	Hapl 022	-	1,0000	-	-	-	1
Hapl 002	0,4581	0,0343	0,0050	0,3053	0,1973	2185	Hapl 023	1,0000	-	-	-	-	1
Hapl 003	-	1,0000	-	-	-	1	Hapl 024	1,0000	-	-	-	-	1
Hapl 004	-	0,8571	-	-	0,1429	14	Hapl 025	-	1,0000	-	-	-	2
Hapl 005	-	1,0000	-	-	-	1	Hapl 026	-	1,0000	-	-	-	1
Hapl 006	-	1,0000	-	-	-	2	Hapl 027	-	1,0000	-	-	-	1
Hapl 007	-	1,0000	-	-	-	1	Hapl 028	-	-	-	1,0000	-	2
Hapl 008	-	-	-	-	1,0000	1	Hapl 029	-	-	-	1,0000	-	1
Hapl 009	-	-	-	1,0000	-	1	Hapl 030	-	-	-	1,0000	-	2
Hapl 010	-	-	-	1,0000	-	1	Hapl 031	-	-	1,0000	-	-	1
Hapl 011	-	-	0,8750	0,1250	-	8	Hapl 032	-	-	1,0000	-	-	1
Hapl 012	-	-	1,0000	-	-	4	Hapl 033	-	-	1,0000	-	-	1
Hapl 013	-	-	1,0000	-	-	2	Hapl 034	-	-	1,0000	-	-	1
Hapl 014	-	-	1,0000	-	-	1	Hapl 035	-	1,0000	-	-	-	3
Hapl 015	-	-	1,0000	-	-	1	Hapl 036	-	1,0000	-	-	-	1
Hapl 016	-	-	1,0000	-	-	1	Hapl 037	-	1,0000	-	-	-	1
Hapl 017	-	-	1,0000	-	-	1	Hapl 038	-	1,0000	-	-	-	1
Hapl 018	-	-	-	-	1,0000	1	Hapl 039	-	1,0000	-	-	-	1
Hapl 019	-	1,0000	-	-	-	2	Hapl 040	-	-	-	-	1,0000	1
Hapl 020	-	1,0000	-	-	-	5	Hapl 041	-	-	-	1,0000	-	1
Hapl 021	-	1,0000	-	-	-	1							

Tabela S11 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região Total (mRNA) do gene *SLC24A5* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	-	0,2957	0,4274	0,1667	0,1102	372	Hapl 262	1,0000	-	-	-	-	1
Hapl 002	-	0,3641	0,3871	0,0230	0,2258	434	Hapl 263	1,0000	-	-	-	-	1
Hapl 003	-	0,0476	0,3968	0,0635	0,4921	126	Hapl 264	1,0000	-	-	-	-	2
Hapl 004	-	0,8889	-	-	0,1111	27	Hapl 265	1,0000	-	-	-	-	1
Hapl 005	-	1,0000	-	-	-	15	Hapl 266	1,0000	-	-	-	-	1
Hapl 006	-	0,9007	-	-	0,0993	151	Hapl 267	0,6667	0,3333	-	-	-	3
Hapl 007	-	1,0000	-	-	-	2	Hapl 268	1,0000	-	-	-	-	1
Hapl 008	-	0,8837	-	-	0,1163	43	Hapl 269	-	-	-	1,0000	-	1
Hapl 009	-	0,7667	-	-	0,2333	30	Hapl 270	-	-	-	1,0000	-	1
Hapl 010	-	0,9643	-	-	0,0357	28	Hapl 271	-	-	-	1,0000	-	1
Hapl 011	-	0,8636	-	-	0,1364	22	Hapl 272	-	-	-	1,0000	-	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 012	0,4552	0,0352	0,0045	0,3110	0,1941	1762	Hapl 273	-	-	-	1,0000	-	2
Hapl 013	-	0,9074	-	-	0,0926	54	Hapl 274	-	-	-	1,0000	-	4
Hapl 014	-	0,5000	-	-	0,5000	2	Hapl 275	-	-	-	1,0000	-	1
Hapl 015	-	1,0000	-	-	-	17	Hapl 276	-	-	-	1,0000	-	1
Hapl 016	-	0,9615	-	-	0,0385	26	Hapl 277	-	-	-	1,0000	-	2
Hapl 017	-	1,0000	-	-	-	2	Hapl 278	-	-	-	1,0000	-	1
Hapl 018	-	0,7500	-	-	0,2500	12	Hapl 279	-	-	-	1,0000	-	1
Hapl 019	-	0,8333	-	-	0,1667	42	Hapl 280	-	-	-	1,0000	-	1
Hapl 020	-	1,0000	-	-	-	14	Hapl 281	-	-	-	1,0000	-	1
Hapl 021	-	0,9000	-	-	0,1000	10	Hapl 282	-	-	-	1,0000	-	1
Hapl 022	-	1,0000	-	-	-	4	Hapl 283	-	1,0000	-	-	-	1
Hapl 023	-	1,0000	-	-	-	2	Hapl 284	-	1,0000	-	-	-	2
Hapl 024	-	0,9000	-	-	0,1000	10	Hapl 285	-	1,0000	-	-	-	1
Hapl 025	-	1,0000	-	-	-	1	Hapl 286	-	1,0000	-	-	-	1
Hapl 026	-	0,8000	-	-	0,2000	10	Hapl 287	-	1,0000	-	-	-	2
Hapl 027	-	0,9545	-	-	0,0455	22	Hapl 288	-	0,7500	-	-	0,2500	4
Hapl 028	-	1,0000	-	-	-	7	Hapl 289	-	1,0000	-	-	-	1
Hapl 029	-	1,0000	-	-	-	2	Hapl 290	-	1,0000	-	-	-	1
Hapl 030	-	0,5000	0,5000	-	-	6	Hapl 291	-	1,0000	-	-	-	1
Hapl 031	-	1,0000	-	-	-	3	Hapl 292	-	1,0000	-	-	-	2
Hapl 032	-	1,0000	-	-	-	2	Hapl 293	-	1,0000	-	-	-	5
Hapl 033	-	0,9091	-	-	0,0909	11	Hapl 294	-	1,0000	-	-	-	5
Hapl 034	-	0,6000	-	-	0,4000	5	Hapl 295	-	1,0000	-	-	-	3
Hapl 035	-	1,0000	-	-	-	1	Hapl 296	-	0,7500	-	-	0,2500	4
Hapl 036	-	0,9091	-	-	0,0909	11	Hapl 297	-	1,0000	-	-	-	1
Hapl 037	-	1,0000	-	-	-	1	Hapl 298	-	1,0000	-	-	-	3
Hapl 038	-	0,8333	-	-	0,1667	12	Hapl 299	-	0,6667	-	-	0,3333	3
Hapl 039	-	1,0000	-	-	-	1	Hapl 300	-	1,0000	-	-	-	1
Hapl 040	-	1,0000	-	-	-	1	Hapl 301	-	1,0000	-	-	-	2
Hapl 041	-	0,8750	-	-	0,1250	8	Hapl 302	-	1,0000	-	-	-	1
Hapl 042	-	1,0000	-	-	-	2	Hapl 303	-	1,0000	-	-	-	1
Hapl 043	-	1,0000	-	-	-	1	Hapl 304	-	1,0000	-	-	-	1
Hapl 044	-	1,0000	-	-	-	4	Hapl 305	-	1,0000	-	-	-	3
Hapl 045	0,5246	0,0164	0,0164	0,2295	0,2131	61	Hapl 306	-	1,0000	-	-	-	1
Hapl 046	-	1,0000	-	-	-	5	Hapl 307	-	1,0000	-	-	-	1
Hapl 047	-	1,0000	-	-	-	5	Hapl 308	-	1,0000	-	-	-	1
Hapl 048	-	1,0000	-	-	-	13	Hapl 309	-	1,0000	-	-	-	1
Hapl 049	-	1,0000	-	-	-	1	Hapl 310	-	1,0000	-	-	-	3
Hapl 050	-	0,8500	-	-	0,1500	20	Hapl 311	-	1,0000	-	-	-	1
Hapl 051	-	1,0000	-	-	-	1	Hapl 312	-	1,0000	-	-	-	3
Hapl 052	-	0,8750	-	-	0,1250	16	Hapl 313	-	1,0000	-	-	-	1
Hapl 053	-	0,8571	-	-	0,1429	7	Hapl 314	-	1,0000	-	-	-	1
Hapl 054	-	0,8846	-	-	0,1154	26	Hapl 315	-	1,0000	-	-	-	1
Hapl 055	-	0,9167	-	-	0,0833	12	Hapl 316	-	1,0000	-	-	-	1
Hapl 056	-	1,0000	-	-	-	6	Hapl 317	-	1,0000	-	-	-	1
Hapl 057	-	1,0000	-	-	-	1	Hapl 318	-	1,0000	-	-	-	1
Hapl 058	-	1,0000	-	-	-	1	Hapl 319	1,0000	-	-	-	-	1
Hapl 059	0,4943	0,0230	0,0115	0,3333	0,1379	87	Hapl 320	1,0000	-	-	-	-	1
Hapl 060	-	1,0000	-	-	-	1	Hapl 321	1,0000	-	-	-	-	2
Hapl 061	-	1,0000	-	-	-	1	Hapl 322	0,6667	-	-	0,3333	-	6
Hapl 062	-	1,0000	-	-	-	1	Hapl 323	0,5000	-	-	-	0,5000	2
Hapl 063	-	0,6667	-	-	0,3333	3	Hapl 324	1,0000	-	-	-	-	1
Hapl 064	-	0,6667	-	-	0,3333	3	Hapl 325	1,0000	-	-	-	-	1
Hapl 065	-	1,0000	-	-	-	2	Hapl 326	0,2500	-	-	0,7500	-	4
Hapl 066	-	0,8571	-	-	0,1429	7	Hapl 327	0,6667	-	-	-	0,3333	3
Hapl 067	-	-	-	-	1,0000	1	Hapl 328	1,0000	-	-	-	-	1
Hapl 068	-	0,8750	-	-	0,1250	8	Hapl 329	-	-	-	1,0000	-	1
Hapl 069	-	-	-	-	1,0000	1	Hapl 330	-	-	-	1,0000	-	2
Hapl 070	-	-	-	-	1,0000	1	Hapl 331	-	-	-	1,0000	-	2
Hapl 071	-	-	-	-	1,0000	1	Hapl 332	-	-	-	1,0000	-	2
Hapl 072	-	-	-	-	1,0000	1	Hapl 333	-	-	-	1,0000	-	1
Hapl 073	-	-	-	-	1,0000	1	Hapl 334	-	-	-	1,0000	-	1
Hapl 074	-	-	0,5000	-	0,5000	2	Hapl 335	-	-	-	1,0000	-	5
Hapl 075	-	-	-	-	1,0000	1	Hapl 336	-	-	-	1,0000	-	1
Hapl 076	-	-	-	-	1,0000	1	Hapl 337	-	-	-	1,0000	-	2
Hapl 077	-	0,6667	-	-	0,3333	3	Hapl 338	-	-	-	1,0000	-	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 078	0,6667	-	-	0,1111	0,2222	9	Hapl 339	-	-	-	1,0000	-	2
Hapl 079	-	-	-	-	1,0000	1	Hapl 340	-	-	-	1,0000	-	1
Hapl 080	-	-	-	-	1,0000	2	Hapl 341	-	-	-	1,0000	-	1
Hapl 081	-	-	-	-	1,0000	1	Hapl 342	-	-	-	1,0000	-	1
Hapl 082	0,7143	-	-	-	0,2857	7	Hapl 343	-	-	-	1,0000	-	2
Hapl 083	-	0,5000	-	-	0,5000	2	Hapl 344	-	-	1,0000	-	-	2
Hapl 084	-	-	-	-	1,0000	1	Hapl 345	-	-	1,0000	-	-	1
Hapl 085	-	-	-	-	1,0000	2	Hapl 346	-	-	1,0000	-	-	1
Hapl 086	-	-	-	-	1,0000	1	Hapl 347	-	-	1,0000	-	-	1
Hapl 087	-	-	-	-	1,0000	1	Hapl 348	-	-	1,0000	-	-	1
Hapl 088	0,0073	-	0,6850	0,1282	0,1795	273	Hapl 349	-	-	1,0000	-	-	1
Hapl 089	-	-	-	-	1,0000	1	Hapl 350	-	-	1,0000	-	-	4
Hapl 090	-	0,5000	-	-	0,5000	2	Hapl 351	-	-	1,0000	-	-	1
Hapl 091	-	-	-	-	1,0000	1	Hapl 352	-	-	1,0000	-	-	1
Hapl 092	0,6316	-	-	0,1053	0,2632	19	Hapl 353	-	-	1,0000	-	-	1
Hapl 093	-	-	-	-	1,0000	1	Hapl 354	-	-	1,0000	-	-	1
Hapl 094	-	-	-	-	1,0000	1	Hapl 355	-	-	1,0000	-	-	1
Hapl 095	-	-	-	-	1,0000	1	Hapl 356	-	-	1,0000	-	-	1
Hapl 096	-	-	0,4375	0,5625	-	32	Hapl 357	-	-	1,0000	-	-	2
Hapl 097	-	-	-	1,0000	-	2	Hapl 358	-	-	1,0000	-	-	3
Hapl 098	-	-	-	1,0000	-	2	Hapl 359	-	-	1,0000	-	-	1
Hapl 099	0,6000	-	-	0,2000	0,2000	5	Hapl 360	-	-	1,0000	-	-	1
Hapl 100	-	-	0,5000	0,5000	-	34	Hapl 361	-	-	1,0000	-	-	1
Hapl 101	-	0,0235	0,7059	0,2706	-	85	Hapl 362	-	-	1,0000	-	-	1
Hapl 102	-	0,0286	-	0,9714	-	35	Hapl 363	-	-	1,0000	-	-	1
Hapl 103	-	-	-	1,0000	-	5	Hapl 364	-	-	1,0000	-	-	1
Hapl 104	-	-	0,0714	0,9286	-	14	Hapl 365	-	-	1,0000	-	-	1
Hapl 105	-	-	-	1,0000	-	4	Hapl 366	-	-	1,0000	-	-	1
Hapl 106	-	-	-	1,0000	-	4	Hapl 367	-	-	1,0000	-	-	1
Hapl 107	-	-	-	1,0000	-	5	Hapl 368	-	-	1,0000	-	-	1
Hapl 108	-	-	-	1,0000	-	1	Hapl 369	-	-	1,0000	-	-	1
Hapl 109	-	-	-	1,0000	-	1	Hapl 370	-	-	0,5000	-	0,5000	2
Hapl 110	-	-	-	1,0000	-	1	Hapl 371	-	-	1,0000	-	-	1
Hapl 111	-	-	-	1,0000	-	2	Hapl 372	-	-	1,0000	-	-	1
Hapl 112	-	-	-	1,0000	-	1	Hapl 373	-	-	1,0000	-	-	1
Hapl 113	-	-	-	1,0000	-	2	Hapl 374	-	-	1,0000	-	-	1
Hapl 114	0,8667	-	-	0,0667	0,0667	15	Hapl 375	-	-	1,0000	-	-	1
Hapl 115	-	-	-	1,0000	-	1	Hapl 376	-	-	1,0000	-	-	1
Hapl 116	-	-	-	1,0000	-	1	Hapl 377	-	-	1,0000	-	-	1
Hapl 117	0,5000	-	-	0,5000	-	2	Hapl 378	-	-	1,0000	-	-	1
Hapl 118	-	-	-	1,0000	-	7	Hapl 379	-	1,0000	-	-	-	1
Hapl 119	-	-	-	1,0000	-	1	Hapl 380	-	1,0000	-	-	-	1
Hapl 120	-	-	-	1,0000	-	1	Hapl 381	-	1,0000	-	-	-	1
Hapl 121	-	-	-	1,0000	-	1	Hapl 382	-	1,0000	-	-	-	2
Hapl 122	-	-	-	1,0000	-	1	Hapl 383	-	1,0000	-	-	-	1
Hapl 123	-	-	-	1,0000	-	1	Hapl 384	-	1,0000	-	-	-	1
Hapl 124	-	-	-	1,0000	-	1	Hapl 385	-	1,0000	-	-	-	1
Hapl 125	0,5455	-	-	0,3636	0,0909	11	Hapl 386	-	1,0000	-	-	-	1
Hapl 126	-	-	-	1,0000	-	1	Hapl 387	-	1,0000	-	-	-	1
Hapl 127	-	-	0,8571	0,1429	-	7	Hapl 388	-	1,0000	-	-	-	1
Hapl 128	-	-	-	0,9000	0,1000	10	Hapl 389	-	1,0000	-	-	-	1
Hapl 129	-	-	-	1,0000	-	1	Hapl 390	-	1,0000	-	-	-	1
Hapl 130	-	-	-	1,0000	-	1	Hapl 391	-	1,0000	-	-	-	1
Hapl 131	0,5000	-	-	0,5000	-	2	Hapl 392	-	1,0000	-	-	-	1
Hapl 132	-	-	1,0000	-	-	9	Hapl 393	-	1,0000	-	-	-	1
Hapl 133	-	-	1,0000	-	-	1	Hapl 394	-	1,0000	-	-	-	3
Hapl 134	-	-	1,0000	-	-	97	Hapl 395	-	1,0000	-	-	-	1
Hapl 135	-	-	1,0000	-	-	5	Hapl 396	-	1,0000	-	-	-	1
Hapl 136	-	-	1,0000	-	-	1	Hapl 397	-	1,0000	-	-	-	1
Hapl 137	-	-	1,0000	-	-	13	Hapl 398	-	1,0000	-	-	-	1
Hapl 138	-	-	1,0000	-	-	2	Hapl 399	-	1,0000	-	-	-	1
Hapl 139	-	-	1,0000	-	-	4	Hapl 400	-	1,0000	-	-	-	2
Hapl 140	-	-	1,0000	-	-	6	Hapl 401	-	1,0000	-	-	-	1
Hapl 141	-	-	1,0000	-	-	1	Hapl 402	-	1,0000	-	-	-	1
Hapl 142	-	-	1,0000	-	-	2	Hapl 403	-	1,0000	-	-	-	1
Hapl 143	-	-	1,0000	-	-	2	Hapl 404	-	1,0000	-	-	-	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 144	-	-	1,0000	-	-	2	Hapl 405	-	1,0000	-	-	-	1
Hapl 145	-	-	1,0000	-	-	1	Hapl 406	-	1,0000	-	-	-	1
Hapl 146	-	-	1,0000	-	-	1	Hapl 407	-	1,0000	-	-	-	1
Hapl 147	-	-	1,0000	-	-	9	Hapl 408	-	1,0000	-	-	-	1
Hapl 148	-	-	0,6667	-	0,3333	3	Hapl 409	-	1,0000	-	-	-	1
Hapl 149	-	-	1,0000	-	-	1	Hapl 410	-	1,0000	-	-	-	1
Hapl 150	-	-	1,0000	-	-	2	Hapl 411	-	1,0000	-	-	-	1
Hapl 151	-	0,1250	0,6250	-	0,2500	16	Hapl 412	-	1,0000	-	-	-	1
Hapl 152	-	-	1,0000	-	-	1	Hapl 413	-	1,0000	-	-	-	1
Hapl 153	-	-	1,0000	-	-	1	Hapl 414	-	1,0000	-	-	-	1
Hapl 154	-	-	1,0000	-	-	21	Hapl 415	-	1,0000	-	-	-	1
Hapl 155	-	-	0,6471	-	0,3529	17	Hapl 416	-	1,0000	-	-	-	2
Hapl 156	-	-	1,0000	-	-	4	Hapl 417	-	1,0000	-	-	-	1
Hapl 157	-	-	1,0000	-	-	1	Hapl 418	-	1,0000	-	-	-	2
Hapl 158	-	-	1,0000	-	-	4	Hapl 419	-	1,0000	-	-	-	1
Hapl 159	-	-	1,0000	-	-	9	Hapl 420	-	1,0000	-	-	-	1
Hapl 160	-	-	1,0000	-	-	1	Hapl 421	-	1,0000	-	-	-	1
Hapl 161	-	-	1,0000	-	-	1	Hapl 422	-	1,0000	-	-	-	1
Hapl 162	0,5000	-	-	-	0,5000	2	Hapl 423	-	1,0000	-	-	-	1
Hapl 163	0,8000	-	-	0,1333	0,0667	15	Hapl 424	-	1,0000	-	-	-	1
Hapl 164	1,0000	-	-	-	-	1	Hapl 425	-	1,0000	-	-	-	1
Hapl 165	0,6667	-	-	0,3333	-	3	Hapl 426	-	1,0000	-	-	-	1
Hapl 166	1,0000	-	-	-	-	1	Hapl 427	-	1,0000	-	-	-	1
Hapl 167	1,0000	-	-	-	-	2	Hapl 428	-	1,0000	-	-	-	1
Hapl 168	1,0000	-	-	-	-	1	Hapl 429	-	1,0000	-	-	-	1
Hapl 169	1,0000	-	-	-	-	1	Hapl 430	-	1,0000	-	-	-	1
Hapl 170	0,2857	-	-	0,4286	0,2857	7	Hapl 431	-	-	-	-	1,0000	1
Hapl 171	1,0000	-	-	-	-	1	Hapl 432	-	-	-	-	1,0000	1
Hapl 172	1,0000	-	-	-	-	1	Hapl 433	-	-	-	-	1,0000	1
Hapl 173	1,0000	-	-	-	-	1	Hapl 434	-	-	-	-	1,0000	1
Hapl 174	1,0000	-	-	-	-	1	Hapl 435	-	-	-	-	1,0000	1
Hapl 175	1,0000	-	-	-	-	1	Hapl 436	-	-	-	-	1,0000	1
Hapl 176	-	-	1,0000	-	-	1	Hapl 437	-	-	-	-	1,0000	1
Hapl 177	-	-	1,0000	-	-	7	Hapl 438	-	-	-	-	1,0000	1
Hapl 178	-	-	1,0000	-	-	1	Hapl 439	-	-	-	-	1,0000	1
Hapl 179	-	-	1,0000	-	-	1	Hapl 440	-	0,5000	-	-	0,5000	2
Hapl 180	-	-	1,0000	-	-	1	Hapl 441	-	-	-	-	1,0000	1
Hapl 181	-	-	1,0000	-	-	2	Hapl 442	-	-	-	-	1,0000	1
Hapl 182	-	-	1,0000	-	-	1	Hapl 443	-	-	-	-	1,0000	1
Hapl 183	-	-	1,0000	-	-	2	Hapl 444	-	-	-	-	1,0000	1
Hapl 184	-	-	1,0000	-	-	5	Hapl 445	-	-	-	-	1,0000	1
Hapl 185	-	-	1,0000	-	-	2	Hapl 446	-	-	-	-	1,0000	2
Hapl 186	-	-	1,0000	-	-	2	Hapl 447	-	-	-	-	1,0000	1
Hapl 187	-	-	1,0000	-	-	4	Hapl 448	-	-	-	-	1,0000	1
Hapl 188	-	-	0,7500	-	0,2500	4	Hapl 449	-	-	-	-	1,0000	1
Hapl 189	-	-	1,0000	-	-	1	Hapl 450	-	-	-	-	1,0000	1
Hapl 190	-	-	1,0000	-	-	1	Hapl 451	-	-	-	-	1,0000	1
Hapl 191	-	-	1,0000	-	-	1	Hapl 452	-	-	-	-	1,0000	1
Hapl 192	-	-	1,0000	-	-	2	Hapl 453	-	-	-	1,0000	-	1
Hapl 193	-	-	1,0000	-	-	1	Hapl 454	-	-	-	1,0000	-	1
Hapl 194	-	0,5000	0,5000	-	-	2	Hapl 455	-	-	-	1,0000	-	2
Hapl 195	-	-	1,0000	-	-	2	Hapl 456	-	-	-	1,0000	-	1
Hapl 196	-	-	1,0000	-	-	1	Hapl 457	-	-	-	1,0000	-	2
Hapl 197	-	-	1,0000	-	-	1	Hapl 458	-	-	-	1,0000	-	1
Hapl 198	-	-	1,0000	-	-	1	Hapl 459	-	-	-	1,0000	-	1
Hapl 199	-	-	1,0000	-	-	1	Hapl 460	-	-	-	1,0000	-	1
Hapl 200	-	-	1,0000	-	-	1	Hapl 461	-	-	-	1,0000	-	2
Hapl 201	-	-	1,0000	-	-	1	Hapl 462	-	-	-	1,0000	-	1
Hapl 202	-	-	1,0000	-	-	1	Hapl 463	-	-	-	1,0000	-	1
Hapl 203	-	-	1,0000	-	-	1	Hapl 464	-	-	-	1,0000	-	1
Hapl 204	-	-	1,0000	-	-	1	Hapl 465	-	-	-	1,0000	-	1
Hapl 205	-	-	1,0000	-	-	1	Hapl 466	-	-	-	-	1,0000	4
Hapl 206	-	-	1,0000	-	-	1	Hapl 467	-	-	-	-	1,0000	1
Hapl 207	-	-	1,0000	-	-	1	Hapl 468	-	-	-	-	1,0000	1
Hapl 208	-	-	1,0000	-	-	1	Hapl 469	-	-	-	-	1,0000	1
Hapl 209	-	-	1,0000	-	-	1	Hapl 470	-	-	-	-	1,0000	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 210	-	-	1,0000	-	-	1	Hapl 471	-	-	-	-	1,0000	2
Hapl 211	-	-	1,0000	-	-	2	Hapl 472	-	-	-	-	1,0000	1
Hapl 212	-	-	0,7500	-	0,2500	4	Hapl 473	-	-	-	-	1,0000	1
Hapl 213	-	-	1,0000	-	-	1	Hapl 474	-	-	-	-	1,0000	1
Hapl 214	-	-	1,0000	-	-	1	Hapl 475	-	-	-	-	1,0000	2
Hapl 215	-	-	1,0000	-	-	1	Hapl 476	-	-	-	-	1,0000	1
Hapl 216	-	-	1,0000	-	-	1	Hapl 477	-	-	-	-	1,0000	1
Hapl 217	-	-	-	-	1,0000	2	Hapl 478	-	-	-	-	1,0000	1
Hapl 218	-	-	-	-	1,0000	1	Hapl 479	-	-	-	-	1,0000	1
Hapl 219	-	-	-	-	1,0000	2	Hapl 480	0,5000	-	-	-	0,5000	2
Hapl 220	-	-	-	-	1,0000	1	Hapl 481	-	-	-	1,0000	-	1
Hapl 221	-	-	-	-	1,0000	2	Hapl 482	-	-	-	1,0000	-	1
Hapl 222	-	-	-	-	1,0000	1	Hapl 483	-	-	-	1,0000	-	1
Hapl 223	-	-	-	-	1,0000	3	Hapl 484	-	-	-	1,0000	-	2
Hapl 224	-	-	-	-	1,0000	1	Hapl 485	-	-	-	1,0000	-	2
Hapl 225	-	-	-	-	1,0000	1	Hapl 486	-	-	-	1,0000	-	1
Hapl 226	-	-	-	-	1,0000	1	Hapl 487	-	-	-	1,0000	-	1
Hapl 227	-	-	-	-	1,0000	1	Hapl 488	-	-	-	1,0000	-	1
Hapl 228	-	1,0000	-	-	-	2	Hapl 489	-	-	-	1,0000	-	1
Hapl 229	-	1,0000	-	-	-	5	Hapl 490	-	-	-	1,0000	-	1
Hapl 230	-	1,0000	-	-	-	2	Hapl 491	-	-	-	1,0000	-	2
Hapl 231	-	1,0000	-	-	-	1	Hapl 492	-	-	-	1,0000	-	1
Hapl 232	-	1,0000	-	-	-	1	Hapl 493	-	-	-	1,0000	-	1
Hapl 233	-	1,0000	-	-	-	1	Hapl 494	-	-	-	1,0000	-	1
Hapl 234	-	1,0000	-	-	-	1	Hapl 495	-	-	-	1,0000	-	1
Hapl 235	-	1,0000	-	-	-	3	Hapl 496	-	-	-	1,0000	-	2
Hapl 236	-	1,0000	-	-	-	1	Hapl 497	-	-	-	1,0000	-	1
Hapl 237	-	1,0000	-	-	-	1	Hapl 498	-	-	-	1,0000	-	1
Hapl 238	-	1,0000	-	-	-	1	Hapl 499	-	-	-	1,0000	-	1
Hapl 239	-	1,0000	-	-	-	1	Hapl 500	-	-	-	1,0000	-	1
Hapl 240	-	1,0000	-	-	-	1	Hapl 501	-	-	-	1,0000	-	1
Hapl 241	-	1,0000	-	-	-	1	Hapl 502	-	-	-	1,0000	-	1
Hapl 242	-	1,0000	-	-	-	3	Hapl 503	1,0000	-	-	-	-	1
Hapl 243	-	1,0000	-	-	-	1	Hapl 504	1,0000	-	-	-	-	1
Hapl 244	-	1,0000	-	-	-	1	Hapl 505	1,0000	-	-	-	-	1
Hapl 245	-	1,0000	-	-	-	1	Hapl 506	1,0000	-	-	-	-	1
Hapl 246	-	1,0000	-	-	-	2	Hapl 507	1,0000	-	-	-	-	1
Hapl 247	-	1,0000	-	-	-	1	Hapl 508	1,0000	-	-	-	-	1
Hapl 248	1,0000	-	-	-	-	1	Hapl 509	1,0000	-	-	-	-	1
Hapl 249	0,6667	-	-	-	0,3333	3	Hapl 510	1,0000	-	-	-	-	1
Hapl 250	0,7500	-	-	0,2500	-	4	Hapl 511	-	1,0000	-	-	-	1
Hapl 251	1,0000	-	-	-	-	1	Hapl 512	-	1,0000	-	-	-	1
Hapl 252	1,0000	-	-	-	-	1	Hapl 513	-	1,0000	-	-	-	1
Hapl 253	1,0000	-	-	-	-	1	Hapl 514	-	1,0000	-	-	-	1
Hapl 254	1,0000	-	-	-	-	1	Hapl 515	-	1,0000	-	-	-	1
Hapl 255	1,0000	-	-	-	-	1	Hapl 516	-	1,0000	-	-	-	1
Hapl 256	1,0000	-	-	-	-	1	Hapl 517	-	1,0000	-	-	-	1
Hapl 257	1,0000	-	-	-	-	1	Hapl 518	-	1,0000	-	-	-	2
Hapl 258	1,0000	-	-	-	-	1	Hapl 519	-	1,0000	-	-	-	1
Hapl 259	1,0000	-	-	-	-	1	Hapl 520	-	1,0000	-	-	-	1
Hapl 260	1,0000	-	-	-	-	3	Hapl 521	-	1,0000	-	-	-	1
Hapl 261	1,0000	-	-	-	-	1	Hapl 522	-	1,0000	-	-	-	1

Tabela S12 – Frequências haplotípicas relativas para cada um dos grupos populacionais na região Total (Estendida) do gene *SLC24A5* e seu respectivo tamanho amostral (AFR – África, AMR – América, EAS – Leste da Ásia, EUR – Europa, SAS – Sul da Ásia)

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 001	-	0,2360	0,4382	0,1985	0,1273	267	Hapl 305	1,0000	-	-	-	-	1
Hapl 002	-	0,3656	0,3797	0,0236	0,2311	424	Hapl 306	1,0000	-	-	-	-	1
Hapl 003	-	0,0513	0,3590	0,0684	0,5214	117	Hapl 307	1,0000	-	-	-	-	1
Hapl 004	-	0,8696	-	-	0,1304	23	Hapl 308	1,0000	-	-	-	-	2
Hapl 005	-	1,0000	-	-	-	15	Hapl 309	1,0000	-	-	-	-	1
Hapl 006	-	0,8785	-	-	0,1215	107	Hapl 310	1,0000	-	-	-	-	1
Hapl 007	-	1,0000	-	-	-	2	Hapl 311	0,6667	0,3333	-	-	-	3
Hapl 008	-	0,8750	-	-	0,1250	40	Hapl 312	1,0000	-	-	-	-	1
Hapl 009	-	0,7500	-	-	0,2500	28	Hapl 313	-	-	-	1,0000	-	1
Hapl 010	-	0,9643	-	-	0,0357	28	Hapl 314	-	-	-	1,0000	-	1
Hapl 011	-	0,8636	-	-	0,1364	22	Hapl 315	-	-	-	1,0000	-	1
Hapl 012	0,4565	0,0353	0,0046	0,3076	0,1958	1726	Hapl 316	-	-	-	1,0000	-	1
Hapl 013	-	0,9245	-	-	0,0755	53	Hapl 317	-	-	-	1,0000	-	2
Hapl 014	-	0,5000	-	-	0,5000	2	Hapl 318	-	-	-	1,0000	-	4
Hapl 015	-	1,0000	-	-	-	16	Hapl 319	-	-	-	1,0000	-	1
Hapl 016	-	0,9615	-	-	0,0385	26	Hapl 320	-	-	-	1,0000	-	1
Hapl 017	-	1,0000	-	-	-	2	Hapl 321	-	-	-	1,0000	-	2
Hapl 018	-	0,7500	-	-	0,2500	12	Hapl 322	-	-	-	1,0000	-	2
Hapl 019	-	0,8333	-	-	0,1667	42	Hapl 323	-	-	-	1,0000	-	1
Hapl 020	-	1,0000	-	-	-	12	Hapl 324	-	-	-	1,0000	-	1
Hapl 021	-	0,8889	-	-	0,1111	9	Hapl 325	-	-	-	1,0000	-	1
Hapl 022	-	1,0000	-	-	-	4	Hapl 326	-	-	-	1,0000	-	1
Hapl 023	-	0,9149	-	-	0,0851	47	Hapl 327	-	-	-	1,0000	-	1
Hapl 024	-	1,0000	-	-	-	2	Hapl 328	-	1,0000	-	-	-	1
Hapl 025	-	1,0000	-	-	-	1	Hapl 329	-	1,0000	-	-	-	1
Hapl 026	-	0,9000	-	-	0,1000	10	Hapl 330	-	1,0000	-	-	-	2
Hapl 027	-	1,0000	-	-	-	1	Hapl 331	-	1,0000	-	-	-	1
Hapl 028	-	0,8000	-	-	0,2000	10	Hapl 332	-	1,0000	-	-	-	1
Hapl 029	-	0,9545	-	-	0,0455	22	Hapl 333	-	1,0000	-	-	-	1
Hapl 030	-	1,0000	-	-	-	7	Hapl 334	-	1,0000	-	-	-	2
Hapl 031	-	1,0000	-	-	-	1	Hapl 335	-	0,7500	-	-	0,2500	4
Hapl 032	-	0,5000	0,5000	-	-	6	Hapl 336	-	1,0000	-	-	-	1
Hapl 033	-	1,0000	-	-	-	2	Hapl 337	-	1,0000	-	-	-	3
Hapl 034	-	0,9286	-	-	0,0714	14	Hapl 338	-	1,0000	-	-	-	1
Hapl 035	-	1,0000	-	-	-	2	Hapl 339	-	1,0000	-	-	-	1
Hapl 036	-	0,8889	-	-	0,1111	9	Hapl 340	-	1,0000	-	-	-	2
Hapl 037	-	0,6000	-	-	0,4000	5	Hapl 341	-	1,0000	-	-	-	5
Hapl 038	-	1,0000	-	-	-	1	Hapl 342	-	1,0000	-	-	-	5
Hapl 039	-	1,0000	-	-	-	1	Hapl 343	-	1,0000	-	-	-	3
Hapl 040	-	0,9000	-	-	0,1000	10	Hapl 344	-	0,7500	-	-	0,2500	4
Hapl 041	-	1,0000	-	-	-	1	Hapl 345	-	1,0000	-	-	-	1
Hapl 042	-	0,8333	-	-	0,1667	12	Hapl 346	-	1,0000	-	-	-	3
Hapl 043	-	1,0000	-	-	-	1	Hapl 347	-	0,6667	-	-	0,3333	3
Hapl 044	-	1,0000	-	-	-	1	Hapl 348	-	1,0000	-	-	-	1
Hapl 045	-	0,8750	-	-	0,1250	8	Hapl 349	-	1,0000	-	-	-	2
Hapl 046	-	1,0000	-	-	-	2	Hapl 350	-	1,0000	-	-	-	2
Hapl 047	-	1,0000	-	-	-	1	Hapl 351	-	1,0000	-	-	-	1
Hapl 048	-	1,0000	-	-	-	4	Hapl 352	-	1,0000	-	-	-	1
Hapl 049	0,5167	0,0167	0,0167	0,2333	0,2167	60	Hapl 353	-	1,0000	-	-	-	1
Hapl 050	-	1,0000	-	-	-	5	Hapl 354	-	1,0000	-	-	-	3
Hapl 051	-	1,0000	-	-	-	5	Hapl 355	-	1,0000	-	-	-	1
Hapl 052	-	1,0000	-	-	-	13	Hapl 356	-	1,0000	-	-	-	1
Hapl 053	-	0,9600	-	-	0,0400	25	Hapl 357	-	1,0000	-	-	-	1
Hapl 054	-	1,0000	-	-	-	2	Hapl 358	-	1,0000	-	-	-	1
Hapl 055	-	1,0000	-	-	-	1	Hapl 359	-	1,0000	-	-	-	1
Hapl 056	-	0,8500	-	-	0,1500	20	Hapl 360	-	1,0000	-	-	-	1
Hapl 057	-	1,0000	-	-	-	1	Hapl 361	-	1,0000	-	-	-	3
Hapl 058	-	0,8667	-	-	0,1333	15	Hapl 362	-	1,0000	-	-	-	1
Hapl 059	-	0,8571	-	-	0,1429	7	Hapl 363	-	1,0000	-	-	-	3
Hapl 060	-	0,8846	-	-	0,1154	26	Hapl 364	-	1,0000	-	-	-	1
Hapl 061	-	0,9091	-	-	0,0909	11	Hapl 365	-	1,0000	-	-	-	1
Hapl 062	-	1,0000	-	-	-	1	Hapl 366	-	1,0000	-	-	-	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 063	-	1,0000	-	-	-	6	Hapl 367	-	1,0000	-	-	-	1
Hapl 064	-	1,0000	-	-	-	1	Hapl 368	-	1,0000	-	-	-	1
Hapl 065	-	1,0000	-	-	-	1	Hapl 369	-	1,0000	-	-	-	1
Hapl 066	0,5000	0,0233	0,0116	0,3256	0,1395	86	Hapl 370	-	1,0000	-	-	-	1
Hapl 067	-	1,0000	-	-	-	1	Hapl 371	-	1,0000	-	-	-	1
Hapl 068	-	1,0000	-	-	-	1	Hapl 372	1,0000	-	-	-	-	1
Hapl 069	-	1,0000	-	-	-	1	Hapl 373	1,0000	-	-	-	-	1
Hapl 070	-	0,6667	-	-	0,3333	3	Hapl 374	1,0000	-	-	-	-	2
Hapl 071	-	0,6667	-	-	0,3333	3	Hapl 375	1,0000	-	-	-	-	1
Hapl 072	-	1,0000	-	-	-	2	Hapl 376	0,6667	-	-	0,3333	-	6
Hapl 073	-	0,8571	-	-	0,1429	7	Hapl 377	1,0000	-	-	-	-	3
Hapl 074	-	-	-	-	1,0000	1	Hapl 378	0,5000	-	-	-	0,5000	2
Hapl 075	-	0,7500	-	-	0,2500	4	Hapl 379	1,0000	-	-	-	-	1
Hapl 076	-	-	-	-	1,0000	1	Hapl 380	1,0000	-	-	-	-	1
Hapl 077	-	-	-	-	1,0000	1	Hapl 381	0,2500	-	-	0,7500	-	4
Hapl 078	-	-	-	-	1,0000	1	Hapl 382	0,6667	-	-	-	0,3333	3
Hapl 079	-	-	-	-	1,0000	1	Hapl 383	1,0000	-	-	-	-	1
Hapl 080	-	-	-	-	1,0000	1	Hapl 384	-	-	-	1,0000	-	1
Hapl 081	-	-	0,5000	-	0,5000	2	Hapl 385	-	-	-	1,0000	-	2
Hapl 082	-	-	-	-	1,0000	1	Hapl 386	-	-	-	1,0000	-	2
Hapl 083	-	-	-	-	1,0000	1	Hapl 387	-	-	-	1,0000	-	2
Hapl 084	-	-	-	-	1,0000	1	Hapl 388	-	-	-	1,0000	-	1
Hapl 085	-	0,6667	-	-	0,3333	3	Hapl 389	-	-	-	1,0000	-	1
Hapl 086	0,6667	-	-	0,1111	0,2222	9	Hapl 390	-	-	-	1,0000	-	1
Hapl 087	-	-	-	-	1,0000	1	Hapl 391	-	-	-	1,0000	-	1
Hapl 088	-	-	-	-	1,0000	2	Hapl 392	-	-	-	1,0000	-	5
Hapl 089	-	-	-	-	1,0000	1	Hapl 393	-	-	-	1,0000	-	1
Hapl 090	0,7143	-	-	-	0,2857	7	Hapl 394	-	-	-	1,0000	-	2
Hapl 091	-	0,5000	-	-	0,5000	2	Hapl 395	-	-	-	1,0000	-	1
Hapl 092	-	-	-	-	1,0000	1	Hapl 396	-	-	-	1,0000	-	2
Hapl 093	-	-	-	-	1,0000	2	Hapl 397	-	-	-	1,0000	-	1
Hapl 094	-	-	-	-	1,0000	1	Hapl 398	-	-	-	1,0000	-	1
Hapl 095	-	-	-	-	1,0000	1	Hapl 399	-	-	-	1,0000	-	1
Hapl 096	0,0074	-	0,6838	0,1287	0,1801	272	Hapl 400	-	-	-	1,0000	-	2
Hapl 097	-	-	-	-	1,0000	1	Hapl 401	-	-	-	1,0000	-	1
Hapl 098	-	0,5000	-	-	0,5000	2	Hapl 402	-	-	1,0000	-	-	1
Hapl 099	-	-	-	-	1,0000	1	Hapl 403	-	-	1,0000	-	-	1
Hapl 100	0,6316	-	-	0,1053	0,2632	19	Hapl 404	-	-	1,0000	-	-	1
Hapl 101	-	-	-	-	1,0000	1	Hapl 405	-	-	1,0000	-	-	1
Hapl 102	-	-	-	-	1,0000	1	Hapl 406	-	-	1,0000	-	-	1
Hapl 103	-	-	-	-	1,0000	1	Hapl 407	-	-	1,0000	-	-	1
Hapl 104	-	0,1667	0,4167	0,3333	0,0833	12	Hapl 408	-	-	1,0000	-	-	1
Hapl 105	-	-	0,3793	0,6207	-	29	Hapl 409	-	-	1,0000	-	-	4
Hapl 106	-	-	-	1,0000	-	2	Hapl 410	-	-	1,0000	-	-	1
Hapl 107	-	-	-	1,0000	-	2	Hapl 411	-	-	1,0000	-	-	1
Hapl 108	0,6000	-	-	0,2000	0,2000	5	Hapl 412	-	-	1,0000	-	-	1
Hapl 109	-	-	0,4848	0,5152	-	33	Hapl 413	-	-	1,0000	-	-	1
Hapl 110	-	0,0235	0,7059	0,2706	-	85	Hapl 414	-	-	1,0000	-	-	1
Hapl 111	-	0,0286	-	0,9714	-	35	Hapl 415	-	-	1,0000	-	-	1
Hapl 112	-	-	-	1,0000	-	5	Hapl 416	-	-	1,0000	-	-	2
Hapl 113	0,5000	-	-	0,5000	-	2	Hapl 417	-	-	1,0000	-	-	3
Hapl 114	-	-	0,0769	0,9231	-	13	Hapl 418	-	-	1,0000	-	-	1
Hapl 115	-	-	-	1,0000	-	4	Hapl 419	-	-	1,0000	-	-	1
Hapl 116	-	-	-	1,0000	-	4	Hapl 420	-	-	1,0000	-	-	1
Hapl 117	-	-	-	1,0000	-	5	Hapl 421	-	-	1,0000	-	-	1
Hapl 118	-	-	-	1,0000	-	1	Hapl 422	-	-	1,0000	-	-	1
Hapl 119	0,7500	-	-	0,2500	-	4	Hapl 423	-	-	1,0000	-	-	1
Hapl 120	-	-	-	1,0000	-	1	Hapl 424	-	-	1,0000	-	-	1
Hapl 121	-	-	-	1,0000	-	1	Hapl 425	-	-	1,0000	-	-	1
Hapl 122	-	-	-	1,0000	-	2	Hapl 426	-	-	1,0000	-	-	1
Hapl 123	-	-	-	1,0000	-	1	Hapl 427	-	-	1,0000	-	-	1
Hapl 124	-	-	-	1,0000	-	3	Hapl 428	-	-	1,0000	-	-	1
Hapl 125	-	-	-	1,0000	-	2	Hapl 429	-	-	1,0000	-	-	1
Hapl 126	0,8667	-	-	0,0667	0,0667	15	Hapl 430	-	-	1,0000	-	-	1
Hapl 127	-	-	-	1,0000	-	1	Hapl 431	-	-	1,0000	-	-	1
Hapl 128	-	-	-	1,0000	-	1	Hapl 432	-	-	0,5000	-	0,5000	2

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 129	0,5000	-	-	0,5000	-	2	Hapl 433	-	-	1,0000	-	-	1
Hapl 130	-	-	-	1,0000	-	7	Hapl 434	-	-	1,0000	-	-	1
Hapl 131	-	-	-	1,0000	-	1	Hapl 435	-	-	1,0000	-	-	1
Hapl 132	-	-	-	1,0000	-	1	Hapl 436	-	-	1,0000	-	-	1
Hapl 133	-	-	-	1,0000	-	5	Hapl 437	-	-	1,0000	-	-	1
Hapl 134	-	-	-	1,0000	-	1	Hapl 438	-	-	1,0000	-	-	1
Hapl 135	-	-	-	1,0000	-	1	Hapl 439	-	-	1,0000	-	-	1
Hapl 136	-	-	-	1,0000	-	1	Hapl 440	-	-	1,0000	-	-	1
Hapl 137	-	-	-	1,0000	-	1	Hapl 441	-	1,0000	-	-	-	3
Hapl 138	-	-	-	1,0000	-	1	Hapl 442	-	1,0000	-	-	-	1
Hapl 139	-	-	-	1,0000	-	1	Hapl 443	-	1,0000	-	-	-	1
Hapl 140	0,6000	-	-	0,3000	0,1000	10	Hapl 444	-	1,0000	-	-	-	1
Hapl 141	-	-	-	1,0000	-	1	Hapl 445	-	1,0000	-	-	-	1
Hapl 142	-	-	0,8571	0,1429	-	7	Hapl 446	-	1,0000	-	-	-	1
Hapl 143	-	-	-	0,9000	0,1000	10	Hapl 447	-	1,0000	-	-	-	1
Hapl 144	-	-	-	1,0000	-	1	Hapl 448	-	1,0000	-	-	-	1
Hapl 145	-	-	-	1,0000	-	1	Hapl 449	-	1,0000	-	-	-	1
Hapl 146	0,5000	-	-	0,5000	-	2	Hapl 450	-	1,0000	-	-	-	1
Hapl 147	-	-	1,0000	-	-	9	Hapl 451	-	1,0000	-	-	-	1
Hapl 148	-	-	1,0000	-	-	1	Hapl 452	-	1,0000	-	-	-	1
Hapl 149	-	-	1,0000	-	-	87	Hapl 453	-	1,0000	-	-	-	1
Hapl 150	-	-	1,0000	-	-	5	Hapl 454	-	1,0000	-	-	-	1
Hapl 151	-	-	1,0000	-	-	1	Hapl 455	-	1,0000	-	-	-	1
Hapl 152	-	-	1,0000	-	-	13	Hapl 456	-	1,0000	-	-	-	1
Hapl 153	-	-	1,0000	-	-	2	Hapl 457	-	1,0000	-	-	-	1
Hapl 154	-	-	1,0000	-	-	19	Hapl 458	-	1,0000	-	-	-	1
Hapl 155	-	-	1,0000	-	-	15	Hapl 459	-	1,0000	-	-	-	3
Hapl 156	-	-	1,0000	-	-	4	Hapl 460	-	1,0000	-	-	-	1
Hapl 157	-	-	1,0000	-	-	8	Hapl 461	-	1,0000	-	-	-	1
Hapl 158	-	-	1,0000	-	-	4	Hapl 462	-	1,0000	-	-	-	1
Hapl 159	-	-	1,0000	-	-	6	Hapl 463	-	1,0000	-	-	-	1
Hapl 160	-	-	1,0000	-	-	6	Hapl 464	-	1,0000	-	-	-	1
Hapl 161	-	-	1,0000	-	-	1	Hapl 465	-	1,0000	-	-	-	1
Hapl 162	-	-	1,0000	-	-	2	Hapl 466	-	1,0000	-	-	-	1
Hapl 163	-	-	1,0000	-	-	2	Hapl 467	-	1,0000	-	-	-	1
Hapl 164	-	-	1,0000	-	-	2	Hapl 468	-	1,0000	-	-	-	2
Hapl 165	-	-	1,0000	-	-	1	Hapl 469	-	1,0000	-	-	-	1
Hapl 166	-	-	1,0000	-	-	1	Hapl 470	-	1,0000	-	-	-	1
Hapl 167	-	-	1,0000	-	-	9	Hapl 471	-	1,0000	-	-	-	1
Hapl 168	-	-	0,6667	-	0,3333	3	Hapl 472	-	1,0000	-	-	-	1
Hapl 169	-	-	1,0000	-	-	1	Hapl 473	-	1,0000	-	-	-	1
Hapl 170	-	-	1,0000	-	-	2	Hapl 474	-	1,0000	-	-	-	1
Hapl 171	-	0,1250	0,6250	-	0,2500	16	Hapl 475	-	1,0000	-	-	-	1
Hapl 172	-	-	1,0000	-	-	1	Hapl 476	-	1,0000	-	-	-	1
Hapl 173	-	-	1,0000	-	-	1	Hapl 477	-	1,0000	-	-	-	1
Hapl 174	-	-	1,0000	-	-	21	Hapl 478	-	1,0000	-	-	-	1
Hapl 175	-	-	0,6000	-	0,4000	15	Hapl 479	-	1,0000	-	-	-	1
Hapl 176	-	-	1,0000	-	-	4	Hapl 480	-	1,0000	-	-	-	1
Hapl 177	-	-	1,0000	-	-	1	Hapl 481	-	1,0000	-	-	-	1
Hapl 178	-	-	1,0000	-	-	1	Hapl 482	-	1,0000	-	-	-	1
Hapl 179	-	-	1,0000	-	-	4	Hapl 483	-	1,0000	-	-	-	1
Hapl 180	-	-	1,0000	-	-	8	Hapl 484	-	1,0000	-	-	-	1
Hapl 181	-	-	1,0000	-	-	1	Hapl 485	-	1,0000	-	-	-	2
Hapl 182	-	-	1,0000	-	-	1	Hapl 486	-	1,0000	-	-	-	1
Hapl 183	-	-	1,0000	-	-	1	Hapl 487	-	1,0000	-	-	-	2
Hapl 184	0,6667	-	-	-	0,3333	3	Hapl 488	-	1,0000	-	-	-	1
Hapl 185	0,5000	-	-	-	0,5000	2	Hapl 489	-	1,0000	-	-	-	1
Hapl 186	0,8000	-	-	0,1333	0,0667	15	Hapl 490	-	1,0000	-	-	-	1
Hapl 187	1,0000	-	-	-	-	2	Hapl 491	-	1,0000	-	-	-	1
Hapl 188	1,0000	-	-	-	-	1	Hapl 492	-	1,0000	-	-	-	1
Hapl 189	0,6667	-	-	0,3333	-	3	Hapl 493	-	1,0000	-	-	-	1
Hapl 190	1,0000	-	-	-	-	1	Hapl 494	-	1,0000	-	-	-	1
Hapl 191	1,0000	-	-	-	-	2	Hapl 495	-	1,0000	-	-	-	1
Hapl 192	1,0000	-	-	-	-	1	Hapl 496	-	1,0000	-	-	-	1
Hapl 193	1,0000	-	-	-	-	1	Hapl 497	-	1,0000	-	-	-	1
Hapl 194	0,2857	-	-	0,4286	0,2857	7	Hapl 498	-	1,0000	-	-	-	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 195	1,0000	-	-	-	-	1	Hapl 499	-	1,0000	-	-	-	1
Hapl 196	1,0000	-	-	-	-	1	Hapl 500	-	1,0000	-	-	-	1
Hapl 197	1,0000	-	-	-	-	1	Hapl 501	-	1,0000	-	-	-	1
Hapl 198	1,0000	-	-	-	-	1	Hapl 502	-	1,0000	-	-	-	1
Hapl 199	1,0000	-	-	-	-	1	Hapl 503	-	-	-	-	1,0000	1
Hapl 200	-	-	1,0000	-	-	1	Hapl 504	-	-	-	-	1,0000	1
Hapl 201	-	-	1,0000	-	-	7	Hapl 505	-	-	-	-	1,0000	1
Hapl 202	-	-	1,0000	-	-	1	Hapl 506	-	-	-	-	1,0000	1
Hapl 203	-	-	1,0000	-	-	1	Hapl 507	-	-	-	-	1,0000	1
Hapl 204	-	-	1,0000	-	-	1	Hapl 508	-	-	-	-	1,0000	1
Hapl 205	-	-	1,0000	-	-	1	Hapl 509	-	-	-	-	1,0000	1
Hapl 206	-	-	1,0000	-	-	2	Hapl 510	-	-	-	-	1,0000	1
Hapl 207	-	-	1,0000	-	-	1	Hapl 511	-	-	-	-	1,0000	1
Hapl 208	-	-	1,0000	-	-	2	Hapl 512	-	0,5000	-	-	0,5000	2
Hapl 209	-	-	1,0000	-	-	1	Hapl 513	-	-	-	-	1,0000	1
Hapl 210	-	-	1,0000	-	-	1	Hapl 514	-	-	-	-	1,0000	1
Hapl 211	-	-	1,0000	-	-	2	Hapl 515	-	-	-	-	1,0000	1
Hapl 212	-	-	1,0000	-	-	5	Hapl 516	-	-	-	-	1,0000	1
Hapl 213	-	-	1,0000	-	-	2	Hapl 517	-	-	-	-	1,0000	1
Hapl 214	-	-	1,0000	-	-	2	Hapl 518	-	-	-	-	1,0000	1
Hapl 215	-	-	1,0000	-	-	2	Hapl 519	-	-	-	-	1,0000	1
Hapl 216	-	-	1,0000	-	-	4	Hapl 520	-	-	-	-	1,0000	2
Hapl 217	-	-	0,7500	-	0,2500	4	Hapl 521	-	-	-	-	1,0000	1
Hapl 218	-	-	1,0000	-	-	1	Hapl 522	-	-	-	-	1,0000	1
Hapl 219	-	-	1,0000	-	-	1	Hapl 523	-	-	-	-	1,0000	1
Hapl 220	-	-	1,0000	-	-	1	Hapl 524	-	-	-	-	1,0000	1
Hapl 221	-	-	1,0000	-	-	1	Hapl 525	-	-	-	-	1,0000	1
Hapl 222	-	-	1,0000	-	-	2	Hapl 526	-	-	-	-	1,0000	1
Hapl 223	-	-	1,0000	-	-	1	Hapl 527	-	-	-	1,0000	-	1
Hapl 224	-	0,5000	0,5000	-	-	2	Hapl 528	-	-	-	1,0000	-	1
Hapl 225	-	-	1,0000	-	-	2	Hapl 529	-	-	-	1,0000	-	2
Hapl 226	-	-	1,0000	-	-	1	Hapl 530	-	-	-	1,0000	-	2
Hapl 227	-	-	1,0000	-	-	1	Hapl 531	-	-	-	1,0000	-	1
Hapl 228	-	-	1,0000	-	-	1	Hapl 532	-	-	-	1,0000	-	2
Hapl 229	-	-	1,0000	-	-	1	Hapl 533	-	-	-	1,0000	-	1
Hapl 230	-	-	1,0000	-	-	1	Hapl 534	-	-	-	1,0000	-	1
Hapl 231	-	-	1,0000	-	-	1	Hapl 535	-	-	-	1,0000	-	2
Hapl 232	-	-	1,0000	-	-	1	Hapl 536	-	-	-	1,0000	-	1
Hapl 233	-	-	1,0000	-	-	1	Hapl 537	-	-	-	1,0000	-	2
Hapl 234	-	-	1,0000	-	-	1	Hapl 538	-	-	-	1,0000	-	1
Hapl 235	-	-	1,0000	-	-	1	Hapl 539	-	-	-	1,0000	-	1
Hapl 236	-	-	1,0000	-	-	1	Hapl 540	-	-	-	1,0000	-	1
Hapl 237	-	-	1,0000	-	-	1	Hapl 541	-	-	-	1,0000	-	1
Hapl 238	-	-	1,0000	-	-	1	Hapl 542	-	-	-	-	1,0000	4
Hapl 239	-	-	1,0000	-	-	1	Hapl 543	-	-	-	-	1,0000	1
Hapl 240	-	-	1,0000	-	-	1	Hapl 544	-	-	-	-	1,0000	1
Hapl 241	-	-	1,0000	-	-	1	Hapl 545	-	-	-	-	1,0000	1
Hapl 242	-	-	1,0000	-	-	1	Hapl 546	-	-	-	-	1,0000	1
Hapl 243	-	-	1,0000	-	-	1	Hapl 547	-	-	-	-	1,0000	1
Hapl 244	-	-	1,0000	-	-	1	Hapl 548	-	-	-	-	1,0000	2
Hapl 245	-	-	1,0000	-	-	2	Hapl 549	-	-	-	-	1,0000	1
Hapl 246	-	-	0,7500	-	0,2500	4	Hapl 550	-	-	-	-	1,0000	1
Hapl 247	-	-	1,0000	-	-	1	Hapl 551	-	-	-	-	1,0000	1
Hapl 248	-	-	1,0000	-	-	1	Hapl 552	-	-	-	-	1,0000	2
Hapl 249	-	-	1,0000	-	-	1	Hapl 553	-	-	-	-	1,0000	1
Hapl 250	-	-	1,0000	-	-	1	Hapl 554	-	-	-	-	1,0000	1
Hapl 251	-	-	1,0000	-	-	1	Hapl 555	-	-	-	-	1,0000	1
Hapl 252	-	-	-	-	1,0000	2	Hapl 556	-	-	-	-	1,0000	1
Hapl 253	-	-	-	-	1,0000	1	Hapl 557	0,5000	-	-	-	0,5000	2
Hapl 254	-	-	-	-	1,0000	1	Hapl 558	-	-	-	1,0000	-	1
Hapl 255	-	-	-	-	1,0000	2	Hapl 559	-	-	-	1,0000	-	1
Hapl 256	-	-	-	-	1,0000	1	Hapl 560	-	-	-	1,0000	-	1
Hapl 257	-	-	-	-	1,0000	2	Hapl 561	-	-	-	1,0000	-	1
Hapl 258	-	-	-	-	1,0000	1	Hapl 562	-	-	-	1,0000	-	2
Hapl 259	-	-	-	-	1,0000	3	Hapl 563	-	-	-	1,0000	-	2
Hapl 260	-	-	-	-	1,0000	1	Hapl 564	-	-	-	1,0000	-	1

Continuação

Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL	Haplótipo	EUR	AFR	EAS	SAS	AMR	n TOTAL
Hapl 261	-	-	-	-	1,0000	1	Hapl 565	-	-	-	1,0000	-	1
Hapl 262	-	-	-	-	1,0000	1	Hapl 566	-	-	-	1,0000	-	1
Hapl 263	-	-	-	-	1,0000	1	Hapl 567	-	-	-	1,0000	-	1
Hapl 264	-	-	-	-	1,0000	1	Hapl 568	-	-	-	1,0000	-	1
Hapl 265	-	0,5000	-	-	0,5000	2	Hapl 569	-	-	-	1,0000	-	1
Hapl 266	-	1,0000	-	-	-	2	Hapl 570	-	-	-	1,0000	-	1
Hapl 267	-	1,0000	-	-	-	5	Hapl 571	-	-	-	1,0000	-	2
Hapl 268	-	1,0000	-	-	-	2	Hapl 572	-	-	-	1,0000	-	1
Hapl 269	-	1,0000	-	-	-	1	Hapl 573	-	-	-	1,0000	-	1
Hapl 270	-	1,0000	-	-	-	1	Hapl 574	-	-	-	1,0000	-	1
Hapl 271	-	1,0000	-	-	-	1	Hapl 575	-	-	-	1,0000	-	1
Hapl 272	-	1,0000	-	-	-	1	Hapl 576	-	-	-	1,0000	-	1
Hapl 273	-	1,0000	-	-	-	3	Hapl 577	-	-	-	1,0000	-	2
Hapl 274	-	1,0000	-	-	-	1	Hapl 578	-	-	-	1,0000	-	1
Hapl 275	-	1,0000	-	-	-	1	Hapl 579	-	-	-	1,0000	-	1
Hapl 276	-	1,0000	-	-	-	1	Hapl 580	-	-	-	1,0000	-	1
Hapl 277	-	1,0000	-	-	-	1	Hapl 581	-	-	-	1,0000	-	1
Hapl 278	-	1,0000	-	-	-	1	Hapl 582	-	-	-	1,0000	-	1
Hapl 279	-	1,0000	-	-	-	1	Hapl 583	-	-	-	1,0000	-	1
Hapl 280	-	1,0000	-	-	-	1	Hapl 584	1,0000	-	-	-	-	1
Hapl 281	-	1,0000	-	-	-	1	Hapl 585	1,0000	-	-	-	-	1
Hapl 282	-	1,0000	-	-	-	1	Hapl 586	1,0000	-	-	-	-	1
Hapl 283	-	1,0000	-	-	-	1	Hapl 587	1,0000	-	-	-	-	1
Hapl 284	-	1,0000	-	-	-	3	Hapl 588	1,0000	-	-	-	-	1
Hapl 285	-	1,0000	-	-	-	1	Hapl 589	1,0000	-	-	-	-	1
Hapl 286	-	1,0000	-	-	-	1	Hapl 590	1,0000	-	-	-	-	1
Hapl 287	-	1,0000	-	-	-	1	Hapl 591	1,0000	-	-	-	-	1
Hapl 288	-	1,0000	-	-	-	2	Hapl 592	1,0000	-	-	-	-	1
Hapl 289	-	1,0000	-	-	-	1	Hapl 593	-	1,0000	-	-	-	1
Hapl 290	1,0000	-	-	-	-	1	Hapl 594	-	1,0000	-	-	-	1
Hapl 291	1,0000	-	-	-	-	1	Hapl 595	-	1,0000	-	-	-	1
Hapl 292	0,6667	-	-	-	0,3333	3	Hapl 596	-	1,0000	-	-	-	1
Hapl 293	0,7500	-	-	0,2500	-	4	Hapl 597	-	1,0000	-	-	-	1
Hapl 294	1,0000	-	-	-	-	1	Hapl 598	-	1,0000	-	-	-	1
Hapl 295	1,0000	-	-	-	-	1	Hapl 599	-	1,0000	-	-	-	1
Hapl 296	1,0000	-	-	-	-	1	Hapl 600	-	1,0000	-	-	-	1
Hapl 297	1,0000	-	-	-	-	1	Hapl 601	-	1,0000	-	-	-	1
Hapl 298	1,0000	-	-	-	-	1	Hapl 602	-	1,0000	-	-	-	1
Hapl 299	1,0000	-	-	-	-	1	Hapl 603	-	1,0000	-	-	-	2
Hapl 300	1,0000	-	-	-	-	1	Hapl 604	-	1,0000	-	-	-	1
Hapl 301	1,0000	-	-	-	-	1	Hapl 605	-	1,0000	-	-	-	1
Hapl 302	1,0000	-	-	-	-	1	Hapl 606	-	1,0000	-	-	-	1
Hapl 303	1,0000	-	-	-	-	1	Hapl 607	-	1,0000	-	-	-	1
Hapl 304	1,0000	-	-	-	-	3							

1. Title Page

Research Article

Evaluation of *MC1R* high-throughput nucleotide sequencing data generated by the 1000 Genomes Project: a comparison between available VCF and raw data sets

Leonardo Arduino Marano ¹, Erick C. Castelli ², Celso Teixeira Mendes-Junior ^{3,*}

¹ Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto-SP, Brazil;

² Departamento de Patologia, Faculdade de Medicina de Botucatu, Universidade Estadual Paulista - Unesp, Botucatu - SP, Brazil;

³ Departamento de Química, Laboratório de Pesquisas Forenses e Genômicas, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto-SP, Brazil.

Work developed in Ribeirão Preto, São Paulo, Brazil

Running Title

Next-gen sequencing data evaluation

Keywords

Melanocortin-1 Receptor; The 1000 Genomes Project; Next-generation sequencing

* Corresponding author:

C. T. Mendes-Junior, Ph.D.

Departamento de Química

Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto

Universidade de São Paulo

Avenida dos Bandeirantes, 3900

14040-901, Ribeirão Preto, SP, Brazil.

Tel.: + 55 16 3602-0417; fax: + 55 16 3602-4838

E-mail: ctmendes@ffclrp.usp.br

2. Abstract

The advent of next-generation sequencing allowed several genomic regions / individuals to be processed simultaneously, increasing the availability and accuracy of whole-genome data. However, these new approaches may present some errors and bias due to alignment, genotype calling and imputation methods. Despite these flaws, data obtained by next-generation sequencing can be valuable for population and evolutionary studies for specific genes, such as those related to one of the main topics in human evolutionary biology: the evolution of pigmentation among populations. One of the most studied genes involved in pigmentation variation, melanocortin-1 receptor (*MC1R*), has already been suggested to have effects on melanogenesis and melanoma risk, thus being one of the best models to comprehend natural selection acting on pigmentation. We employed a locally developed pipeline to obtain genotype and haplotype data from the raw sequencing data provided by 1000 Genomes FTP site for *MC1R* and compared it to Phase 3 VCF in order to evaluate its quality and eventually discover any polymorphic sites that may be overlooked. Our analysis demonstrated that data from Phase III is far more reliable than previous releases, considering the steps for alignment, mapping quality evaluation and genotype calling here explored.

3. Text

a. Introduction

The advent of Sanger sequencing in 1977 and its capillary electrophoresis automation provided the complete genome sequencing of various species. However, due to some limitations of this technique (high cost, time consuming and low number of samples/fragments sequenced at a time), some difficulties may arise when complete genomes for multiple samples are the goal, especially in population studies.

The rise of next-generation sequencing (NGS) methods allowed several individuals/regions to be processed simultaneously with high accuracy, since the same region is sequenced several times and various parameters have been developed to enable quality assessment of the generated sequences (Nielsen et al., 2011). Several methods have been proposed in this context, involving, for example, emulsion PCR (e.g., 454 – Roche, Ion Torrent – Life) and solid-phase amplification (e.g., Solexa – Illumina) (Metzker, 2010; Vigliar et al., 2015). As the sequencing costs decline, several collaborative projects for whole-genome sequencing have emerged, such as the 1000 Genomes Project (Genomes Project et al., 2012). Such projects, using distinct population samples, allow the discovery of new variants, elucidating important details about evolutionary and demographic aspects that shaped the human genome (Matullo et al., 2013).

The Phase I of the 1000 Genomes Project included several research centers working together, from 2008 to 2010, sequencing 1,092 complete genomes from 14 populations. Due to high cost and genome size, these first results consisted mainly of low-coverage whole-genome sequencing, high-coverage exome sequencing and high-density SNP Array panels. After data treatment and analysis, each genotype obtained was set into biallelic categories for each sample and each site. From these data, linkage disequilibrium analysis was used to

confirm/infer or even correct the variants obtained from low-coverage non-exome data, in an imputation process. The retrieved data thus present several flaws and bias, such as lack of some polymorphisms due to limitations of the methods of genotype calling available at the time. In addition, the method used to infer genotypes and haplotypes did not allow the differentiation of triallelic SNPs, considering them to be biallelic, decreasing informativity of various *loci* (Castelli et al., 2014). The last phase (Phase 3), released on October 2014, encompassed 2,504 complete genomes, providing multiallelic data for the first time. Studies such as the 1000 Genomes Project can serve as basis for inference of undetected variants, since there are several rare variants located within known linkage disequilibrium blocks (Howie et al., 2011). Low frequency variants may have 60-90% precision on imputation, even in admixed populations (Genomes Project et al., 2012). Although many improvements were achieved on mapping, imputation and genotype calling, the current NGS technologies still have error rates of, at least, 10^{-4} per nucleotide, leading to a high rate of false positives (around 5%) when low-coverage is employed (Matullo et al., 2013).

Several studies have addressed this data analysis issue by using *in silico* methods for identification and correction for any bias that may result in mismapping and/or genotype calling errors (Gilissen et al., 2012; Nagasaki et al., 2013; Castelli et al., 2014), in addition to highlight the importance of validating the obtained results using other established and validated methodologies, such as Sanger sequencing, for regions of interest. New scripts and pipelines have been developed for accurate genotype and haplotype detection from raw data from the 1000 Genomes Project, also enabling to retrieve unreleased data from variants overlooked by the Consortium. Such scripts have been validated in different analyzes of *HLA-G* gene concerning Phase 1 data, demonstrating a higher accuracy in genotype calling on studies of individual genes or delimited chromosome regions (Castelli et al., 2014). Genotype calling accuracy of Phase 3 data has not been approached so far.

Despite the existing errors, data obtained by next-generation sequencing by the 1000 Genomes Project can be valuable for population and evolutionary studies for specific genes. This data set can be used to address one of the major problems related to human evolutionary biology: to explain the evolution of pigmentation variation between different populations. Genes involved in melanogenesis process can be considered one of the major targets of natural selection, which can be verified by the correlation between pigmentation diversity to the ultraviolet (UV) incidence variation, since phenotypes distribution are clearly associated to the latitude variation. Some theories address natural selection and sexual dimorphism as possible explanations for this diversity (Aoki, 2002; Jablonski e Chaplin, 2014). The most accepted evolutionary explanation concerning this pigmentation diversity refers to protection against UV radiation, particularly in skin areas highly exposed to sun. One of the differences between human beings from other primates and many mammals is that humans have lost most of the body hair. This fur loss would have an evolutionary advantage, such as the capacity to exude more efficiently, which would help in body temperature regulation. However, the absence of this protective layer might have created a need of another type of protection against ultraviolet (UV) radiation (Rees, 2003). Dark skin pigmentation has high levels of melanin, which can act as protection against UV damage (Jablonski e Chaplin, 2014).

Alleles from many *MC1R* coding SNPs (Arg151Cys, Arg160Trp and Asp294His) have already been associated to red hair, fair skin and freckling in Europeans (Harding et al., 2000; Sturm, 2006). Other studies have also identified several *MC1R* coding variants related to melanoma susceptibility (Ghiorzo et al., 2012; Pellegrini et al., 2012). Understanding the expression and activity regulation of this receptor is essential to comprehend its effects on melanogenesis and melanoma risk (Swope et al., 2012). Since natural selection acts primarily shaping global genetic diversity of regions with functional impact on the expression of genes

and their products, understanding worldwide *MC1R* genetic diversity patterns is the first step in order to understand its expression and activity regulation.

Therefore, considering the limitations here exposed, we employed a locally developed pipeline to obtain genotype and haplotype data from the raw sequencing data provided by 1000 Genomes FTP site and comparing it to Phase 3 VCF in order to evaluate its quality and eventually discover any polymorphic sites that may be overlooked by the VCF.

b. Results

After the conversion steps from the raw data from the 1000 Genomes Project to the phased haplotypes of *MC1R* for all individuals, the process resulted in complete data for 201 *loci* for 2,537 individuals, determining 596 distinct haplotypes. Of these 201 *loci*, 24 belonged to the 5'URR (Upstream Regulatory Region), 54 belonged to the 5'UTR (Untranslated Region), 89 belonged to the CDS (Coding Sequence) and 34 to the 3'UTR. As the data obtained directly from the 1000 Genomes Project VCF had 178 *loci* (174 SNVs and 4 InDels) in common for the same 2,504 individuals, we removed the 23 *loci* (11 singletons) and 33 individuals that were incompatible between the two datasets, allowing equivalence in comparisons.

After this initial filtering, genotypes for each *locus* were compared among individuals in order to verify if there were incompatibilities. Conducting an initial analysis, considering the total number of *loci* (178) and individuals (2,504), there were 445,712 possible genotype comparisons between the two analysis. From this total, only 585 incompatibilities were observed on genotype calling (0.13% of total).

It was found that 87.58% of the sample (2,193 individuals) had identical genotype data between the two datasets for all analyzed *loci*. The remaining 311 subjects had at least

one *locus* with incompatible genotypes, this error rate reaching a maximum of 13 incompatible *loci* in a single individual (Figure 1). By analyzing each *locus* individually, it was observed that 102 of them showed compatibility between the two approaches for all individuals. Among the 76 discordant *loci*, we observed that some of them condensed most of the discrepancies: 13 of them being responsible for almost 80% of all mismatches throughout the study and one of them (rs885479) responsible for 87 mismatches (14.87% of total) (Figure 2). If we consider the genetic regions isolated, the 5'UTR was the one with more *loci* containing mismatches, totaling 32 *loci* from the 48 encompassed by the region (66.67%).

Figure 1 – Number of mismatches observed between data generated by our pipeline and data obtained directly from the VCF concerning 2,504 individuals analyzed by the 1000 Genomes Project

Figure 2 – Number of mismatches observed between data generated by our pipeline and data obtained directly from the VCF concerning 178 *loci* analyzed by the 1000 Genomes Project

When analyzing the allele frequencies for each evaluated *locus*, we noted that even with the mentioned mismatches, there were no significant differences between the two approaches. Even *locus* rs885479, containing the largest amount of mismatches, ended presenting an irrelevant difference of 1.22% for the reference allele frequency between the VCF and data processed by the scripts.

From these results, we checked the BAM files alignment using IGV software to manually explore those mismatches between the two approaches. From the original 585 genotype differences we evaluated 40 random mismatches for rs885479 *locus* and 60 mismatches randomly chosen among the remaining *loci* (Table 1). From the 40 mismatches evaluated for *locus* rs885479, 26 of them had seven reads or less, and were automatically considered by the script with missing alleles, turning the data comparison uninformative. The other 14 mismatches had less than twenty reads, with 5 of them presenting an alternate allele underrepresented (ratio between 5-20%), which also resulted in the introduction of missing alleles. It was observed that 7 mismatches apparently occurred due to some genotype calling error in our pipeline (3 and 4 cases should have been considered heterozygous and homozygous, respectively) and 2 cases had wrong genotype calling included in the VCF. From the 60 random mismatches, it was observed that the pipeline introduced missing alleles in 33 of them, either because of low read number or underrepresentation of the alternative allele, turning the data comparison uninformative. Other 26 mismatches were due to errors made by our script (15 and 11 cases should have been regarded as homozygous and heterozygous, respectively) and 1 of the cases the genotype calling was erroneously made by the VCF.

Surprisingly, it was also observed that the 178 *loci* on the 2,504 individuals, which resulted in 207 distinct haplotype for the VCF data, ended up resulting in 490 different haplotypes on the data generated by our scripts (Table 1). It should be noted that not all 207

haplotypes obtained from VCF were among those obtained by the scripts. From the total, 52 haplotypes were not among the 490 inferred by our pipeline. Comparing the haplotype pairs obtained for each individual, we observed that, from the 5,008 haplotypes generated, 2,455 haplotype comparisons were identical and 2,553 included some inconsistencies. This result is even more astonishing if it is taken into account the fact that 87.58% of the sample (2,193 individuals) had identical genotype data between the two datasets for all analyzed *loci*: only 1,147 (52.3%) of them had identical haplotype reconstructions.

Table 1 – Comparison between data from the 1000 Genomes Project obtained from our pipeline analysis and from the VCF

Mismatches	<i>locus rs885479</i>	Random sites
Missing alleles	31%	33%
VCF error	2%	1%
Pipeline error	7%	26%

c. Discussion

Most part of the 585 evaluated mismatches among the datasets were due to a high level of missing alleles (approximately 64%, Table 1). This high rate of missing alleles was caused by the low coverage (below 7 or 20 reads) observed among the analyzed polymorphic sites, leading to the imputation steps included in our pipeline. Thus, this high inconsistency ratio between the VCF file and the genotypes obtained by our scripts is probably a reflex of differences concerning the imputation methods used in each approach. At the moment, it is not possible to empirically state which imputation method is more error prone in such extremely low coverage situations.

Genotype imputation can be extremely useful by allowing missing data from low-density chips to be filled, reducing costs (Jimenez-Montero et al., 2013), merging datasets

with non-overlapping genotypes generated by different platforms (Khankhanian et al., 2015), or even predicting complex traits (Felipe et al., 2014). Imputed genotypes have already been proven to have similar accuracy as those obtained from high-density SNP panels, in cases where an appropriate reference population is used (Weigel et al., 2010). However, imputation benefits rely deeply on its accuracy, directly related to population structure of the samples used and the composition of the gene target evaluated (Felipe et al., 2014).

Although low coverage is a recurrent problem regarding next-generation sequencing of whole genomes, another important issue is the read mapping bias, as highly polymorphic regions will generate reads consisting of many alternate alleles. Some of these greatly diverse reads will not be able to align to a single place in the reference genome and will be discarded during analysis. Therefore, it would be interesting to have multiple haplotypes from a given target as reference for several alignment steps, in order to improve genotype calling for these polymorphic regions (Brandt et al., 2015).

Most studies that use genotype imputation for whole-genome sequencing data regard only linear prediction models such as ridge regression, Bayesian LASSO (least absolute selection and shrinkage operator) or GBLUP (genomic best linear unbiased prediction) approaches, that are more suitable for additive genetic models, but not so accurate to detect non-additive effects such as dominance and epistasis (Weigel et al., 2010; Felipe et al., 2014).

But even different imputation models can present bias in specific cases, such as high heterogeneity resulting in underestimation of derived alleles; presence of the alternative allele in small genetic windows not spanned by enough SNPs to make imputation effective; rare frequency of the alternative allele. Although the imputation process can lead to more powerful datasets and new association in case-control or population studies, the inherent bias can lead to data deficiency or type I error that could counteract the results from these datasets, such as missing or underestimating genotype associations (Khankhanian et al., 2015).

Therefore, there must be an accurate evaluation regarding the genetic structure of the data, in order to choose the most fitting and reliable imputation method for the investigated genes or genomic regions. Besides choosing the most suitable imputation method for the data presented, a validation regression method should be used in order to examine the results, searching for artifacts and assure its validity (Hoffmann et al., 2015) or other established and validated methodology should be used, such as Sanger sequencing.

Finally, it was verified that 33% of the mismatches were due to pipeline flaws that made wrong genotype calls, such as detection of homozygous as heterozygous and vice versa. These inaccuracies shall be investigated to evaluate which steps or rules of our scripts are being misapplied resulting in such discrepancies.

In conclusion, data from next-generation sequencing should be treated carefully, in order to obtain the most accurate information from it. A very low incompatibility rate on *MC1R* genotype calling (0.13% of total) was observed in the present comparison. Since no multiallelic marker was retrieved for the *MC1R* region, similar studies including multiallelic markers must be done. Despite that, the present comparison suggests that either the VCF file or the presently described pipeline could be used in order to obtain reliable and accurate genotype calling on studies of individual genes or delimited chromosome regions from 1000 Genomes Phase 3. A negative issue that needs to be further investigated was the significant disagreement on the haplotype inference procedure, which may be due to an underestimation of haplotypes on the VCF file or to an overestimation produced by fastPHASE. Although low coverage can be a big issue regarding these kind of analysis, imputation methods can be useful to fill missing data gaps, as long as a accurate reference panel is chosen. Higher coverage could also improve the precision of genotyping calling, but the development of new methods for identification and correction for bias could be the only answer in cases where only low-coverage data is available. Thus, apart from differences caused by diverse

imputation methods, our analysis demonstrated that data from Phase III is far more reliable than previous releases, considering the steps for alignment, mapping quality evaluation and genotype calling here explored.

d. Methods

This study used two different datasets maintained by the 1000Genomes consortium (Genomes Project et al., 2012).

The first dataset consisted on a VCF file obtained at the official website (<http://browser.1000genomes.org/>) from Phase 3 Release, including 2,504 samples from 26 populations. The *MC1R* regions evaluated here (chr16:89981286-89987385) encompassed 1kb from its 5' Upstream Regulatory Region (5'URR), as well as its single exon composed of a 5' Untranslated Region (5'UTR), Coding Sequence (CDS) and 3' Untranslated Region (3'UTR).

The second dataset consisted on SAM files obtained directly from the 1000 Genomes server (<ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp>). The methodology used in this article was previously validated by our group (Castelli et al., 2014). First, by using Samtools (Li et al., 2009) subroutine view, we downloaded slices of the SAM (sequence alignment/map) files containing the 1000Genomes data for the *MC1R* gene region mentioned above. The download was performed for all of the 2,537 samples available on Phase 3 and included data from both low-coverage whole-genome and high-coverage exome sequencing when available. This process generated up to two SAM files per individual, which were merged as a single file and converted into a BAM (binary alignment/map) file. Each BAM file was then converted into a Fastq format file retrieving all reads that were previously mapped to the *MC1R* region, by use of Bamtools (<https://github.com/pezmaster31/bamtools/>) and Perl scripts (locally developed) to filter out duplicated reads and to classify the reads as paired or unpaired.

Both paired and unpaired Fastq files were re-mapped to a masked chromosome 16 (hg19), in which only the *MC1R* region was available and the rest of the chromosome was masked with “N” to preserve nucleotide positions regarding hg19. The resulting BAM files from the re-mapped reads, from both paired-end and unpaired sequences, were joined using Picard-tools (<http://picard.sourceforge.net/index.shtml>). The Bamtools software was also used to remove reads mapped with low mapping quality (MQ) scores (MQ<40). The GATK routine UnifiedGenotyper was used to infer genotypes and a VCF file (variant call format) was generated.

Given the low coverage nature of the 1000 Genomes data, some genotypes called by GATK are far uncertain, mainly in situations in which a homozygous genotype is inferred when that position presents low depth coverage. In addition, given the polymorphic nature of *MC1R* some level of mismapped reads is expected and might bias genotype inference. To circumvent this issue, the VCF file generated by GATK was treated with a locally developed Perl script that applied the rules described below. This script uses the number of different reads detected for each allele at a given position (provided by GATK when the VCF file was generated).

- Homozygosity was only inferred when a minimal coverage of seven reads was achieved; otherwise, a missing allele was introduced in this genotype. This procedure assures ($p>0.99$) that a homozygous genotype is called because of lack of variance at that position and not because the second allele was not sampled.

- Genotypes, in which one allele was extremely underrepresented (proportion of reads under 5%), were considered as homozygous for the most represented allele. This procedure minimizes the influence of mismapped reads and the high level of sequencing errors that characterizes next-generation sequencing data, and such correction was applied only in

situations characterized by high depth of coverage (20 or more reads available for the evaluated position).

- For genotypes in which one allele was mildly underrepresented (with a proportion of reads between 5 and 20%), a missing allele was introduced representing this underrepresented allele. This procedure is particularly helpful in situations characterized by low depth of coverage (less than 20 reads available for the evaluated position), in which a single read may indicate the existence of an alternative allele, such read may be a mismatched read (false positive variant) or may represent a true unbalanced heterozygous genotype (true positive variant). Therefore, the definitive status of this kind of genotype (homozygous or heterozygous) was inferred during a final imputation step.

- Genotypes in which the proportion of reads for the less represented allele was higher than 20% were considered to be heterozygous. This procedure assures that only high-quality heterozygous genotypes are passed forward to the imputation procedure.

By using the VCFtools package (Danecek et al., 2011), we removed SNVs that were no longer variable or that were represented just once in the dataset (i.e., singletons). The missing alleles were imputed and *MCIR* haplotypes were inferred by using fastPHASE (Scheet e Stephens, 2006).

4. Acknowledgements

This study was supported by CNPq/Brazil (Conselho Nacional de Desenvolvimento Científico e Tecnológico) Grant #448242/2014-1, FAPESP/Brazil (Fundação de Amparo à Pesquisa do Estado de São Paulo) Grant #2013/15447-0. C.T.M.J. (309572/2014-2) and E.C.C. (304471/2013-5) are supported by Research fellowships from CNPq/Brazil. L.A.M. was supported by a PhD. fellowship from CAPES/Brazil.

5. References

- Aoki, K. (2002). Sexual selection as a cause of human skin colour variation: Darwin's hypothesis revisited. *Ann Hum Biol* 29, 589-608.
- Brandt, D. Y., Aguiar, V. R., Bitarello, B. D., Nunes, K., Goudet, J., and Meyer, D. (2015). Mapping Bias Overestimates Reference Allele Frequencies at the HLA Genes in the 1000 Genomes Project Phase I Data. *G3* 5, 931-41.
- Castelli, E. C., Ramalho, J., Porto, I. O., Lima, T. H., Felicio, L. P., Sabbagh, A., Donadi, E. A., and Mendes-Junior, C. T. (2014). Insights into HLA-G Genetics Provided by Worldwide Haplotype Diversity. *Frontiers in immunology* 5, 476.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156-8.
- Felipe, V. P., Okut, H., Gianola, D., Silva, M. A., and Rosa, G. J. (2014). Effect of genotype imputation on genome-enabled prediction of complex traits: an empirical study with mice data. *BMC Genet* 15, 149.
- Genomes Project, C., Abecasis, G. R., Auton, A., Brooks, L. D., Depristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T., and Mcvean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56-65.
- Ghiorzo, P., Bonelli, L., Pastorino, L., Bruno, W., Barile, M., Andreotti, V., Nasti, S., Battistuzzi, L., Grosso, M., Bianchi-Scarra, G., et al. (2012). MC1R variation and melanoma risk in relation to host/clinical and environmental factors in CDKN2A positive and negative melanoma patients. *Exp Dermatol* 21, 718-20.
- Gilissen, C., Hoischen, A., Brunner, H. G., and Veltman, J. A. (2012). Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 20, 490-7.
- Harding, R. M., Healy, E., Ray, A. J., Ellis, N. S., Flanagan, N., Todd, C., Dixon, C., Sajantila, A., Jackson, I. J., Birch-Machin, M. A., et al. (2000). Evidence for variable selective pressures at MC1R. *Am J Hum Genet* 66, 1351-61.
- Hoffmann, T. J., Sakoda, L. C., Shen, L., Jorgenson, E., Habel, L. A., Liu, J., Kvale, M. N., Asgari, M. M., Banda, Y., Corley, D., et al. (2015). Imputation of the rare HOXB13 G84E mutation and cancer risk in a large population-based cohort. *PLoS Genet* 11, e1004930.

- Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. *G3* 1, 457-70.
- Jablonski, N. G., and Chaplin, G. (2014). The evolution of skin pigmentation and hair texture in people of African ancestry. *Dermatol Clin* 32, 113-21.
- Jimenez-Montero, J. A., Gianola, D., Weigel, K., Alenda, R., and Gonzalez-Recio, O. (2013). Assets of imputation to ultra-high density for productive and functional traits. *J Dairy Sci* 96, 6047-58.
- Khankhanian, P., Din, L., Caillier, S. J., Gourraud, P. A., and Baranzini, S. E. (2015). SNP imputation bias reduces effect size determination. *Frontiers in genetics* 6, 30.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-9.
- Matullo, G., Di Gaetano, C., and Guarrera, S. (2013). Next generation sequencing and rare genetic variants: from human population studies to medical genetics. *Environ Mol Mutagen* 54, 518-32.
- Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet* 11, 31-46.
- Nagasaki, H., Mochizuki, T., Kodama, Y., Saruhashi, S., Morizaki, S., Sugawara, H., Ohyanagi, H., Kurata, N., Okubo, K., Takagi, T., et al. (2013). DDBJ read annotation pipeline: a cloud computing-based pipeline for high-throughput analysis of next-generation sequencing data. *DNA research : an international journal for rapid publication of reports on genes and genomes* 20, 383-90.
- Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet* 12, 443-51.
- Pellegrini, C., Fargnoli, M. C., Suppa, M., and Peris, K. (2012). MC1R variants predisposing to concomitant primary cutaneous melanoma in a monozygotic twin pair. *BMC Med Genet* 13, 81.
- Rees, J. L. (2003). Genetics of hair and skin color. *Annu Rev Genet* 37, 67-90.
- Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* 78, 629-44.
- Sturm, R. A. (2006). A golden age of human pigmentation genetics. *Trends Genet* 22, 464-8.
- Swope, V. B., Jameson, J. A., Mcfarland, K. L., Supp, D. M., Miller, W. E., McGraw, D. W., Patel, M. A., Nix, M. A., Millhauser, G. L., Babcock, G. F., et al. (2012). Defining

- MC1R regulation in human melanocytes by its agonist alpha-melanocortin and antagonists agouti signaling protein and beta-defensin 3. *J Invest Dermatol* 132, 2255-62.
- Vigliar, E., Malapelle, U., De Luca, C., Bellevicine, C., and Troncone, G. (2015). Challenges and opportunities of next-generation sequencing: a cytopathologist's perspective. *Cytopathology : official journal of the British Society for Clinical Cytology* 26, 271-83.
- Weigel, K. A., De Los Campos, G., Vazquez, A. I., Rosa, G. J., Gianola, D., and Van Tassell, C. P. (2010). Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. *J Dairy Sci* 93, 5423-35.