

Universidade de Brasília

Instituto de Ciências Biológicas

Programa de Pós-Graduação em Biologia Animal

**Na saúde e na doença: variabilidade genética humana
estimada por marcadores genéticos neutros e em genes**

Ana Carolina Arcanjo Silva

Brasília – DF

2016

Na saúde e na doença: variabilidade genética humana estimada por marcadores genéticos neutros e em genes

Tese apresentada ao Programa de Pós-graduação em Biologia Animal como requisito parcial para a obtenção do título de Doutor em Biologia Animal.

Ana Carolina Arcanjo Silva

Orientadora: Dra. Silviene Fabiana de Oliveira

Brasília – DF

2016

Trabalho desenvolvido no Laboratório de Genética Humana da Universidade de Brasília – UnB e no *Batzer Laboratory of Comparative Genomics* da *Louisiana State University System* – LSU, com suporte financeiro da Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa do Distrito Federal (FAPDF), com período de doutorado-sanduíche no exterior (PDSE).

Aos meus, dedico.

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

Autor ignorado

Agradecimentos

Às agências financiadoras que permitiram que esse trabalho fosse desenvolvido: Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelas bolsas de Doutorado e de Doutorado-Sanduíche, FAPDF pelos auxílios de participação em eventos internacionais.

À banca examinadora deste trabalho, pela disponibilidade, boa vontade e pelas contribuições para tornar essa tese ainda melhor; à Rafaela Toledo e à Carolina Carvalho Gontijo que me auxiliaram nas genotipagens dos marcadores *Alu*. Obrigada por me ajudarem a garantir os resultados aqui apresentados.

Ao pessoal do Batzer Laboratory of Comparative Genomics, minha casa na Louisiana State University. Em especial, ao Dr. Mark Batzer por ter me aceitado em seu laboratório para o Doutorado-Sanduíche, à Ms. Jerilyn Walker por todo o apoio e suporte, e aos meus colegas Jasmine Brown (Ms Sassy Brown), Cody Steely e Vallmer Jordan por terem me mostrado os caminhos das pedras da pesquisa por aquelas bandas.

À minha querida orientadora Silviene Fabiana de Oliveira, que se tornou amiga, mestre, exemplo, professora. Finalmente concluímos quase dez anos de trabalho juntas, e eu só tenho a agradecer por você ter feito de mim uma cientista, sempre respeitando minhas limitações e me incentivando a ir mais longe.

Aos colegas do Laboratório de Genética Humana da Universidade de Brasília, que me ajudaram a chegar até aqui. Em especial à Rafaela Toledo, que foi quem abriu as portas do laboratório para mim. Se hoje sou uma boa cientista, com certeza isso é devido aos seus gritos de “Arcanjo, deixa de ser mané”! Não vou citar outros nomes para não fazer injustiça, mas saibam que todos vocês fazem parte da minha jornada acadêmica. Desde quando eu entrei como IC, até agora em que saio Doutora em Biologia Animal.

Aos meus pais, Miguel Arcanjo e Lucimar Barbosa, por todo o amor, carinho, compreensão, sem contar o apoio emocional e financeiro. Sem os incentivos de vocês, eu jamais chegaria aqui.

Aos meus amados irmãos, Augusto, Gustavo e Ana Paula, por me fazerem sentir tão especial e capaz de conquistar essa batalha que foi o doutorado. Cada vez que vocês olham para mim com orgulho, minha força para continuar cresce mais. Muito obrigada.

Aos meus queridos pais franco-germano-americanos. Martin e Isabelle McConnell, Olivier e Marie-Antoinette Moretéau, muito obrigada por terem feito de mim parte das suas famílias quando eu me encontrava tão solitária no Doutorado-Sanduíche. Minha gratidão por vocês estarem presentes na minha história não cabe em palavras.

Aos meus irmãos de coração, Remy e Julia McConnell, vocês fizeram e fazem a minha vida mais leve e feliz. Me sinto muito honrada por vocês terem me adotado como Big Sis. A saudade aperta.

Às minhas sobrinhas maravilhosas, Amanda e Isadora, que me estimulam a melhorar a cada dia, para que eu seja sempre um exemplo a ser seguido por elas.

Aos meus queridos amigos, em especial Thaís Moysés (minha fiel escudeira), Gabriel Rechden, Ana Paula Gabattelli (comadre, amiga, querida!), Lucas Venancio, Juliana Lis, por todos os momentos de desabafo, de compartilhamento de dramas e por me ajudarem a achar soluções para os meus problemas. Seus lindos!

Finalmente, gostaria de agradecer à minha amada ursinha, Tatianne Monteiro. Sem o seu apoio nesses últimos três anos, e sem a confiança de que sobreviveríamos ao Sanduíche, eu não teria conseguido terminar todo esse trabalho. Obrigada por me cuidar, obrigada por me apresentar à música eletrônica e por me amar todos os dias. Mesmo quando eu não vou dormir porque tenho que escrever a tese madrugada a dentro.

Resumo

O estado da arte do conhecimento sobre as relações entre variabilidade genética e a dinâmica das populações humanas é reflexo de anos de descrições de frequências alélicas e genotípicas. Os padrões de variação de marcadores genéticos em regiões codificadoras e não-codificadoras não são necessariamente os mesmos, uma vez que mutações em regiões codificadoras podem ter impactos significativos na sobrevivência de um indivíduo. Esse trabalho tem por principal objetivo explorar a variabilidade genética humana sob duas perspectivas. Na saúde refere-se à variabilidade genética acessada por marcadores genéticos neutros, que caracterizam as relações de variabilidade entre os grupos populacionais de uma forma não-enviesada. Para isso, utilizou-se um painel de 100 inserções *Alu* em um conjunto de 1125 amostras de populações mundiais. O painel evidenciou grande desequilíbrio de ligação nas populações brasileiras e na população de Utah, fruto da miscigenação recente (250 anos) durante a formação dessas populações. A miscigenação detectada pelo painel distorce as relações entre os grupos populacionais, uma vez que Kalunga forma um grupo à parte dentro do grupo africano (devido à miscigenação europeia) e Brasília se assemelha mais ao grupo do Oriente Médio do que ao europeu (devido à miscigenação africana). Além disso, o painel evidenciou que os grupos populacionais não podem ser considerados homogêneos, apesar de pelo menos uma população de cada grupo não apresentar diferenciação populacional às outras populações do seu grupo. Na doença, a variabilidade genética de genes e regiões gênicas foram avaliadas usando como exemplo a susceptibilidade de indivíduos a formas severas da infecção pelo vírus da influenza H1N1, a partir de uma busca da variabilidade genética global acessada por meio do 1000 Genomes Project. A variabilidade genética para marcadores associados à susceptibilidade a formas severas da infecção pelo vírus da influenza é pequena, ocorrendo principalmente em regiões intergênicas e que, a priori, não afetam a expressão dos genes. No entanto, mutações que já haviam apresentado associação a quadros severos da infecção alcançaram maiores frequências nas populações que tiveram maiores taxas de mortalidade durante a pandemia de H1N1/2009, que estão também relacionadas à ancestralidade compartilhada dessas populações. Portanto, os marcadores genéticos neutros foram suficientes para a diferenciação de grupos de amostras quanto à ancestralidade geográfica, onde as populações miscigenadas tenderam a se agrupar a populações que representassem o grupo com maior contribuição genética daquela população. Já os marcadores genéticos associados à susceptibilidade à infecção pelo vírus da influenza, permitiram identificar poucos grupos, normalmente também congruentes com a ancestralidade genética das amostras. Dessa forma, entender a ancestralidade genética de um grupo populacional facilitaria entender a susceptibilidade a doenças infecciosas, isto é, as duas classes de marcadores se complementam na busca de um melhor entendimento da transição entre a saúde e a doença.

Palavras-chave: genética de populações; 1000 Genomes Project; vírus da influenza; inserção *Alu*; ancestralidade genética; composição genética; diferenciação populacional

Abstract

The state of the art of the knowledge about the relationship between genetic variability and the dynamics of human populations is a result of years of describing allelic and genotypic frequencies. The patterns of variation for genetic markers in coding and non-coding regions are not the same, since mutations in coding regions might have significant impact in one's ability to survive. The main goal for this work is to explore the human genetic variability under two different perspectives. In health, it refers to the genetic variability that is accessed by neutral genetic markers, which describe the variability relationships among population groups in an unbiased way. To achieve that, a panel of 100 *Alu* insertions was used in a set of 1125 worldwide samples, including 160 samples from two admixed Brazilian populations, Brasília and Kalunga. The insertion panel showed a tenfold increase in the number of linked *loci* to each *locus* in Brasilia, Kalunga and Utah when compared to African, European, Asian and Indian populations. Which is mainly due to the recent admixture of different populations in the making of those populations (250 years). Kalunga clusters with the African populations, but still shows differentiation due to being admixed, Brasilia clusters with both the European and the Middle-Eastern groups, being the last one more genetically similar to Brasilia than the first due to an African genetic component. The population groups cannot be considered homogeneous, despite at least one population of each group not showing population differentiation to the other populations of its group. In health, the genetic variability of genes and coding regions were evaluated using as a model the susceptibility of an individual to severe forms of H1N1 influenza virus infection, by searching a global genetic variability that exists in the 1000 Genome Project database. The genetic variability of genetic markers that are associated to susceptibility to severe forms of influenza infection is small, occurring mainly in intergenic regions that, a priori, do not affect gene expression. However, mutations that have been associated to severe forms of infection have reached larger frequencies in populations that had larger mortality rates during the H1N1/2009 pandemic. Therefore, the neutral genetic markers were sufficient to the clustering of samples related to their geographic region. An exception to that were the admixed Brazilian populations that clustered with the groups that had contributed most to their genetic composition. The genetic markers related to susceptibility to influenza virus infection, on the other hand, could tell apart few groups, usually coincidental to their genetic ancestry. In this way, understanding the genetic ancestry of a population group would make it easier to understand the susceptibility to infectious diseases, that is to say that both classes of genetic markers complete each other in the search for a better understanding of the transition between health and sickness.

Keywords: population genetics; 1000 Genomes Project; influenza virus; *Alu* insertion; genetic ancestry; genetic profile; population differentiation

Sumário

Introdução geral	1
Capítulo I – Contextualização do Brasil no cenário genético mundial: uma história de variabilidade genética contada por 100 inserções <i>Alu</i>	8
I.1. Introdução	9
I.1.1. <i>Povoamento do Brasil</i>	10
I.2. Material e Métodos	14
I.2.1. <i>Material biológico e genótipos cedidos</i>	14
I.2.1.1. <i>África</i>	18
I.2.1.2. <i>Ásia</i>	19
I.2.1.3. <i>Europa</i>	20
I.2.1.4. <i>Índia</i>	21
I.2.1.5. <i>Oriente Médio</i>	22
I.2.2. <i>Amplificação de genoma total</i>	23
I.2.4. <i>Genotipagem dos marcadores</i>	27
I.2.5. <i>Análises estatísticas</i>	30
I.2.5.1. <i>Índices-padrão intrapopulacionais</i>	30
I.2.5.2. <i>Índices-padrão interpopulacionais</i>	31
I.3. Resultados e Discussão	34
I.3.1. <i>Parâmetros intrapopulacionais</i>	34
I.3.1.1. <i>Equilíbrio de Hardy-Weinberg</i>	34
I.3.1.2. <i>Diversidade gênica</i>	35
I.3.1.3. <i>Desequilíbrio de ligação</i>	40
I.3.2. <i>Parâmetros interpopulacionais</i>	41
I.3.3. <i>Estimativa de composição genética das populações miscigenadas brasileiras</i>	49
I.3.4. <i>Caracterização do painel de 100 inserções <i>Alu</i> nas populações mundiais</i>	52
I.3.4.1. <i>África</i>	52
I.3.4.2. <i>Ásia</i>	57
I.3.4.3. <i>Europa</i>	60
I.3.4.4. <i>Índia</i>	64
I.3.4.5. <i>Oriente Médio</i>	70
I.3.5. <i>Teste de Neutralidade</i>	73
1.4. Conclusões	75
Capítulo II – A variabilidade genética humana e sua influência na susceptibilidade à infecção severa pelo vírus da Influenza A	78
II.1. Introdução	79

II.2. A ligação do vírus à célula	80
II.3. As propriedades protetoras do muco contra o vírus da influenza A	86
II.4. Dentro da célula hospedeira: resposta imune inata à infecção pelo vírus da influenza A	89
II.4.1. Receptores Toll-Like (TLRs)	92
II.4.2. Receptores RIG-I-like (RLRs)	93
II.4.3. Receptores NOD-like (NLRs)	94
II.4.4. Interferon-induced proteins with tetratricopeptide repeats (IFITs)	95
II.4.5. Escape viral da imunidade inata	95
II.4.6. Mutações relacionadas à infecção pelo vírus da influenza	97
II.5. Espalhando a infecção: a resposta imune adaptativa do hospedeiro ao vírus da influenza A	100
II.5.1. Resposta imune humoral ao vírus da influenza A H1N1	101
II.5.2.1. Resposta imune celular ao vírus da influenza A H1N1 – Células T CD4+	104
II.5.2.2. Resposta imune celular ao vírus da influenza A H1N1 – Células T CD8+	104
II.5.3. A variabilidade genética na resposta imune adaptativa	106
II.6. Considerações finais e conclusão	109
Considerações finais	113
Referências	115
Anexo I.1 Pareceres dos comitês de ética para a utilização das amostras Kalunga e Brasília.	133
	134
Anexo I.2 Protocolo para Amplificação de Genoma Total com o kit GenomiPhi®	136
Anexo I.3 Purificação dos genomas amplificados.	137
Anexo I.4 Tabela contendo os nomes das inserções <i>Alu</i> , o cromossomo em que se encontra, os iniciadores 5' (forward) e 3' (reverse), os tamanhos dos fragmentos amplificados pela PCR (presença de inserção e ausência de inserção, respectivamente), a temperatura de anelamento utilizada no programa da MyCycler™ e a concentração de cloreto de magnésio (MgCl ₂) exigida para cada inserção <i>Alu</i> .	138
Anexo II.1 Artigo "Role of the host genetic variability in the influenza A vírus susceptibility"	145
Apêndice I.1 Amostras selecionadas para amplificação de genoma total devido à pequena quantidade de DNA disponível para caracterização dos 100 <i>Alus</i> .	146
Apêndice I.2. P-valor para diferenciação populacional entre as populações ameríndias e asiáticas para cada inserção <i>Alu</i> cujos dados estavam disponíveis. As células em cinza mostram diferenciação populacional significativa.	149

Índice de Tabelas

Tabela 1 Lista de repositórios públicos de variabilidade genética humana, incluindo tipo de repositório, endereço eletrônico e referência.....	3
Tabela I. 1 Relação das amostras utilizadas neste trabalho, com respectiva origem geográfica e ancestralidade. As amostras que não são brasileiras tiveram os genótipos gentilmente cedidos por Mark A. Batzer, PhD, e Jerilyn A. Walker, MSc (Louisiana State University, Baton Rouge, LA, EUA).....	16
Tabela I. 2 Reagentes utilizados para as reações de PCR e suas concentrações de estoque e final por tubo de reação de PCR.....	27
Tabela I. 3 Programa geral do termociclador utilizado para amplificação dos marcadores selecionados.	28
Tabela I. 4 Estimativas de Diversidade gênica, heterozigose observada e esperada para as populações, baseadas nas frequências de 100 inserções Alu, em contraste com o estimado por Watkins e colaboradores (2003).	37
Tabela I. 5 Estimativa de contribuição genética para as populações miscigenadas brasileiras Brasília e Kalunga através de marcadores genéticos autossômicos.....	50
Tabela I. 6 Pares de populações africanas que não apresentaram diferenciação populacional significativa, com os respectivos valores que corroboram o teste de hipótese de não-diferenciação.	54
Tabela I. 7 Pares de populações asiáticas que não apresentaram diferenciação populacional significativa.....	59
Tabela I. 8 Índice de diferenciação populacional para os pares de populações europeias que não apresentaram diferenciação significativa.	62
Tabela I. 9 Índice de diferenciação populacional para os pares de populações indianas que não apresentam diferenciação significativa. As populações 1 são apenas aquelas que não apresentaram diferenciação genética a mais de oito populações.	66
Tabela I. 10 Índice de diferenciação populacional e significância do teste (p-Valor) para todos os pares de populações do grupo Oriente-Médio. Os valores que indicam diferenciação estão marcados com um asterisco (*).	72
Tabela II. 1 Lista de variantes para sete genes envolvidos na ligação do vírus da influenza à célula (dados retirados do banco de dados do 1000 Genomes Project).	83
Tabela II. 2 Genes do hospedeiro envolvidos nos mecanismos de escape viral. DNAJC3 (codifica o inibidor P58IPK); EIFAK2 (codifica PKR); TRIM25 (codifica a proteína que contém um motivo-tripartido); VISA (codifica a proteína MAVS); (dados retirados do banco de dados do 1000 Genomes Project)..	96
Tabela II. 3 Lista das variantes encontradas nos genes envolvidos na resposta imune inata à infecção pelo vírus da influenza (dados retirados do banco de dados do 1000 Genomes Project).	100
Tabela II. 4a,b Variabilidade genética dos principais loci gênicos de genes HLA de classe I clássicos (dados retirados do bando de dados 1000 Genomes Project)	107

Índice de Figuras

Figura 1. Visão esquemática da evolução da biodiversidade humana a partir de frequências alélicas de marcadores do tipo SNP. Os pontos coloridos representam genótipos diferentes, os gráficos de pizza da letra E representam as frequências alélicas em cinco regiões ao final do processo. Datas aproximadas para os cinco panoramas: (a, b) > 60.000 anos atrás; (c) 60.000 anos atrás; (d) 40.000 anos atrás; (e) 30.000 anos atrás. Retirado de (Barbujani & Colonna, 2010).....	4
Figura I.2 Localização geográfica das populações africanas cujos genótipos foram gentilmente cedidos pelo Dr. Mark A. Batzer. Figura adaptada de Watkins e colaboradores (2003).	19
Figura I.3 Localização geográfica das amostras asiáticas para as quais havia informação de local de coleta. Adaptado de Watkins e colaboradores (2003).	20
Figura I.4 Localização geográfica das amostras do grupo europeu. Apesar de Utah se localizar nos Estados Unidos da América, a ancestralidade deste grupo representa a população do Norte Europeu, e, por esse motivo, Utah aparece localizada no norte da Europa. Adaptado de Watkins e colaboradores (2003).	21
Figura I.5 Localização geográfica das amostras indianas utilizadas neste trabalho. Os pontos em vermelho mostram as tribos indianas, e os pontos pretos evidenciam as castas e a população paquistanesa. Adaptado de Watkins e colaboradores (2003).	22
Figura I.6 Localização geográfica das amostras do Oriente-Médio. Adaptado de Watkins e colaboradores (2003).	23
Figura I.7 Esquema de funcionamento do kit de amplificação do genoma total. 1) os iniciadores hexâmetros se ligam às fitas de DNA da amostra que se pretende amplificar; 2) Inicia a polimerização pela enzima DNA Polimerase Phi29; 3) A polimerização se dá de um iniciador até o próximo ponto. 4) Novas fitas de DNA se soltam parcialmente do DNA molde, de forma a manter a nova fita em constante síntese. 5) Os iniciadores se ligam às novas fitas de DNA, recém-sintetizadas; 6) Novas fitas são sintetizadas a partir delas. Figura adaptada da bula do kit (Amersham Biosciences ®).....	24
Figura I.8 Localização cromossômica das 100 inserções <i>Alu</i> analisadas neste trabalho. À esquerda do cromossomo estão indicadas as coordenadas cromossômicas aproximadas de cada <i>locus</i> e à direita de cada cromossomo estão indicadas as respectivas inserções <i>Alu</i> que se encontram nessas coordenadas.....	26
Figura I.9 Gel de eletroforese em agarose a 2% corado com brometo de etídio. Neste gel pode-se observar a genotipagem de algumas amostras para o marcador <i>Alu048</i> (Yb8NBC120). Todos os géis foram feitos da mesma forma, utilizando-se o marcador de peso molecular 100bp (bandas que se iniciam em 100pb e são incrementadas com esse mesmo intervalo, número 1 da figura). Em seguida, o controle negativo (mix da PCR sem amostra, 2), o controle ancestral (chimpanzé, 3), o controle de amplificação em humanos (célula <i>HeLa</i> , 4) e as amostras em seguida. Podemos observar claramente que os indivíduos 4, 5, 6, 10, 11, 12, 13, 14, 15, 18, 19, 20, 23, 24, 29, 30, 32, 33, 35, 37, 38, 39, 41, 42, 43 e 49 são homozigotos para a ausência de inserção <i>Alu</i> (apresentam apenas uma banda de aproximadamente 150pb); os indivíduos 9, 22, 31 e 48 são heterozigotos (apresentam uma banda de 150pb e uma banda de 450pb), os indivíduos 16, 17, 21, 25, 26, 27, 28, 34, 36, 40, 44, 45, 46 e 47 são	

homozigotos para a inserção <i>Alu</i> (apresentam apenas uma banda de 450pb) e que os indivíduos 7 e 8 não apresentaram amplificação do marcador, tendo sido repetidos em uma PCR posterior.	29
Figura I.10 Índice de diferenciação populacional (F_{ST}) entre os pares de populações subsaarianas.	53
Figura I.11 Distância de Nei entre os pares de população (em azul, diagonal inferior), número médio de diferenças dentro de cada população (em laranja, diagonal) e o número médio de diferenças entre cada par de populações (em verde, diagonal superior).....	55
Figura I.12 Representação gráfica da análise de componentes principais da variação genética dentro do grupo africano. O eixo X representa o componente principal 1 (que captura a maior variação da amostra) e o eixo Y representa o componente principal 2 (que captura a segunda maior variação dentro da amostra).	57
Figura I.13 Índice de diferenciação populacional baseado em F_{ST} entre pares de populações asiáticas, estimado através do painel de 100 inserções <i>Alu</i>	58
Figura I.14 Índice de diferenciação populacional entre as populações do grupo europeu, baseado em 100 inserções <i>Alu</i>	61
Figura I.15 Análise de componente principal para o grupo de populações europeias. O eixo X representa o componente principal 1 (captura a maior variação) e o eixo Y representa o componente principal 2 (captura a segunda maior variação).	64
Figura I.16 Índice de diferenciação populacional para as subpopulações do grupo indiano, estimado a partir de 100 inserções <i>Alu</i>	65
Figura I.17 Análise de discriminantes mostrando os primeiros dois componentes principais de variação dentro do grupo indiano. O eixo X representa o componente principal 1 (maior variação) e o eixo Y representa o componente principal 2 (segunda maior variação).....	70
Figura I.18 Índice de diferenciação populacional para os pares de populações do grupo do Oriente-Médio, estimado por 100 inserções <i>Alu</i>	71
Figura I.19 Análise discriminante de componente principal para o grupo de populações do Oriente Médio, baseado em 100 inserções <i>Alu</i> . O eixo X representa a variação explicada pelo primeiro componente principal (maior variação) enquanto o eixo Y representa a variação explicada pelo segundo componente principal (segunda maior variação).	73
Figura I.20 Inserções <i>Alu</i> que estão sob seleção para todos os grupos de populações analisadas. As linhas vermelhas indicam os limites de 1% de significância para a relação F_{ST} /heterozigose. Apenas marcadores que estejam sobre ou excedam os limites mínimo e máximo do intervalo são considerados sob seleção.	74

Figura II. 1 Ciclo de Vida do Vírus da influenza. Durante o primeiro passo da infecção pelo vírus da influenza, a glicoproteína viral HA se liga a receptores na superfície da célula hospedeira. O vírus entra na célula por meio de endocitose. O baixo pH dos endossomos causa uma alteração na conformação na HA, levando à fusão do vírus ao endossomo e a consequente liberação dos complexos virais de ribonucleoproteínas (vRNPs) no citoplasma. As vRNPs são transportadas para o núcleo, onde a transcrição do RNAm e a replicação/montagem de novos vRNPs ocorrem. A tradução do RNAm viral acontece tanto no citoplasma quanto no Retículo Endoplasmático (RE). Seguindo as modificações Pós-

Traducionais (PTM) de determinadas proteínas virais, os componentes estruturais da partícula viral são transportados para a membrana plasmática junto com novos vRNPs, onde a montagem de novas partículas virais e o brotamento do vírus ocorrem. 86

Figura II. 2 Escape da resposta adaptativa. Durante a infecção pelo vírus da influenza, o RNA viral é detectado e várias vias são ativadas. Isso resulta na produção de interferons tipo-I, citocinas pró-inflamatórias e indução de um estado antiviral da célula. As proteínas do vírus influenza interagem com essa cascata. NS1 se liga ao RNA viral para mascarar a detecção por PRRs. Ela bloqueia a ativação de RIG-I se ligando a TRIM25 e PKR, inibindo sua fosforilação, que ocorre quando PKR se liga ao RNA viral. NP interage com o complexo P58IPK/hsp40, que então libera P58IPK e inibe a fosforilação de PKR. M2, por outro lado, restaura o complexo P58IPK/hsp40, inibindo a síntese proteica e induz a célula à apoptose, facilitando a liberação de partículas virais. PB1-F2 se liga à proteína sinalizadora antiviral mitocondrial (MAVS), inibindo essa proteína e a produção de interferon tipo-I. PB2 também se liga a MAVS e inibe sua função. Além disso, PB2 se liga ao estimulador de promotor interferon 1 (IPS-1), resultando na inativação do promotor que levaria à produção de interferon tipo-I. Setas pontilhadas representam vias e linhas sólidas e setas representam interações diretas. 90

Figura II. 3 Vias de Apresentação antigênica por HLAs e ativação da resposta imune adaptativa. 102

Introdução geral

Em julho de 2010, os renomados geneticistas Guido Barbujani e Vincenzo Colonna pareciam estupefatos com a nova possibilidade de se comparar o genoma completo de treze indivíduos (de continentes diferentes) ao genoma referência humano e os desdobramentos que a novidade poderia trazer para o melhor conhecimento da variabilidade genética humana (Barbujani & Colonna, 2010). Isso porque o estudo da variabilidade genética humana era baseado exclusivamente em fenótipos observáveis, tais como o sistema ABO, proteínas eritrocitárias e isoenzimas (Markert, 1968; Lewontin, 1972) até final dos anos 70. Na década de 70, iniciou-se um movimento tecnológico que permitiu avanços na área da biologia molecular com o início do uso das enzimas de restrição e a posterior descrição de polimorfismos de fragmentos de restrição, os RFLPs.

Os primeiros trabalhos a utilizarem diretamente a avaliação do material genético ocorreram no início da década de 80 com a descrição de RFLPs nos genes da família da hemoglobina (Gill, et al., 1985; Nakatsuji, et al., 1985). A partir do desenvolvimento da técnica de sequenciamento por terminação de cadeia por Sanger, Nicklen e Coulson (1977), foi possível descobrir novos marcadores genéticos e os descrever em diversas populações humanas. Entretanto, devido a dificuldades desse tipo de análise (custo, tempo, equipamento e pessoal), poucos marcadores genéticos eram utilizados em um mesmo estudo e, frequentemente, em um número pequeno de amostras. O desenvolvimento de tecnologias de automação do processo de eletroforese, conhecido como eletroforese capilar (Zagursky & McCormick, 1990, *apud* Dovichi, 1997), permitiu a mudança no ritmo de produção de dados que culminou no sequenciamento completo do genoma humano e a subsequente montagem do genoma referência em 2004 (International Human Genome Sequencing Consortium, 2004).

Pouco mais de três meses depois da revisão publicada por Barbujani e Colonna, o consórcio internacional *The 1000 Genomes Project Consortium* publicou um artigo que comparou os genomas completos de 188 indivíduos ao genoma

referência e entre si, e um total de 1024 indivíduos considerando sequenciamento total e exoma total (The 1000 Genomes Project Consortium, 2010). Em 2015, o mesmo consórcio publicou um mapa global da variabilidade genética humana onde, entre genomas e exomas completos por sequenciamento de alto desempenho (NGS – *next generation sequencing*) e *array-CGH*, exploraram um total de 2.500 indivíduos distribuídos pelos cinco continentes (The 1000 Genomes Project Consortium, 2015a).

Isso é, a velocidade com a qual o conhecimento científico vem sendo gerado, em especial com relação ao melhor entendimento do genoma da nossa espécie, leva a crer, novamente, que estejamos próximos de ter respostas finais aos questionamentos relativos às diferenciações populacionais da nossa espécie, assim como à dinâmica das populações. Isso vem ao encontro do que era demandado há décadas: quanto maior o número de marcadores genéticos maior a probabilidade de entender as relações evolutivas entre as populações e, com isso, traçar hipóteses cada vez mais consistentes sobre origens e rotas de povoamento (Salzano & Callegari-Jacques, 1988; Saitou, 1995), além da miscigenação ocorrida mais intensamente com as grandes navegações pós-colombianas.

O estado da arte do conhecimento sobre as relações entre variabilidade genética e a dinâmica das populações humanas é reflexo de anos de descrições de frequências alélicas e genotípicas. Este panorama passou por diversas fases, o que incluiu a busca por marcadores genéticos população-específicos (*population-specific alleles*; Neel, 1973) e marcadores informativos de ancestralidade (AIMs – *ancestry informative markers*, Bonilla, et al., 2004; Shriver, et al., 1997; Parra, et al., 1998), passando por marcadores associados a doenças (Piel, et al., 2010) e identificação de indivíduos (Gill, et al., 1985; Budowle, et al., 1999) até o aperfeiçoamento de técnicas para a detecção de milhares de variantes por estudo/experimento (Schoumans, et al., 2004; The 1000 Genomes Project Consortium, 2015a; The 1000 Genomes Project Consortium, 2015b; Grotenhuis, et al., 2016).

O resultado desses esforços para descrever a variabilidade humana resultou num grande volume de dados que necessitaram catalogação, organização e curadoria, o que demandou iniciativas internacionais para desenvolver bancos públicos contendo essa variabilidade (tabela 1). Assim surgiram repositórios públicos

de dados genéticos, alguns específicos para um determinado tipo de variante (p.ex. dbSNP, dbRIP, dbGaP, dbVar), alguns específicos para a finalidade do estudo (p.ex. ClinVar, ClinicalTrials, dbMHC, Genes and Disease) e outros que acabam aliando as duas facetas (p.ex. HapMap – desativado recentemente, 1000 Genomes Project).

Tabela 1 Lista de repositórios públicos de variabilidade genética humana, incluindo tipo de repositório, endereço eletrônico e referência.

Repositório	Finalidade/Tipo de marcador	Website	Referência
dbSNP	Polimorfismos de nucleotídeo único	www.ncbi.nlm.nih.gov/SNP	(Sherry, et al., 2001)
dbRIP	Polimorfismo de inserção de retrotransposon	dbrip.org/	(Wang, et al., 2006)
dbGaP	Genótipos e fenótipos	www.ncbi.nlm.nih.gov/gap	(Mailman, et al., 2007)
dbVar	Variantes estruturais	www.ncbi.nlm.nih.gov/dbvar	(Kitts, et al., 2014)
ClinVar	Variantes com importância clínica	www.ncbi.nlm.nih.gov/clinvar	(Landrum, et al., 2014)
dbMHC	Dados referentes ao MHC	www.ncbi.nlm.nih.gov/gv/mhc	(Gourraud, et al., 2006)
COSMIC	Mutações somáticas em câncer	cancer.sanger.ac.uk/cosmic	(Bamford, et al., 2004)
Decipher	Variantes estruturais com importância clínica	decipher.sanger.ac.uk/	(Firth, et al., 2009)
BGMUT	Variantes relacionadas a eritrócitos	www.ncbi.nlm.nih.gov/gv/rbc	(Patnaik, et al., 2012)
HapMap	Variantes comuns	Descontinuado	(The International HapMap Consortium, 2003)
1000 Genomes Project	Todos os tipos de variante	www.1000genomes.org/home	(The 1000 Genomes Project Consortium, 2010)

Os muitos anos de descrição de frequências alélicas permitiram, então, comparar grupos populacionais de distintas posições geográficas e padrões de composição genética foram sendo descritos. Por exemplo, estudos do *Centre d'Étude*

du *Polymorphisme Humain* (CEPH) com 783 marcadores microssatélites neutros (Ramachandran, et al., 2005; Liu, et al., 2006) mostraram que há uma forte relação entre a diversidade genética de uma população e a distância em que essa população se encontra de um ponto arbitrário na África, onde supostamente surgiu o humano moderno (*Homo sapiens*). Mais do que isso, esses estudos evidenciaram que o índice de diferenciação populacional (F_{ST}) era corretamente estimado pelas distâncias entre as populações e que a diversidade genética de uma população tende a cair fortemente quanto mais a população se distancia do ponto de origem africano (Figura 1e).

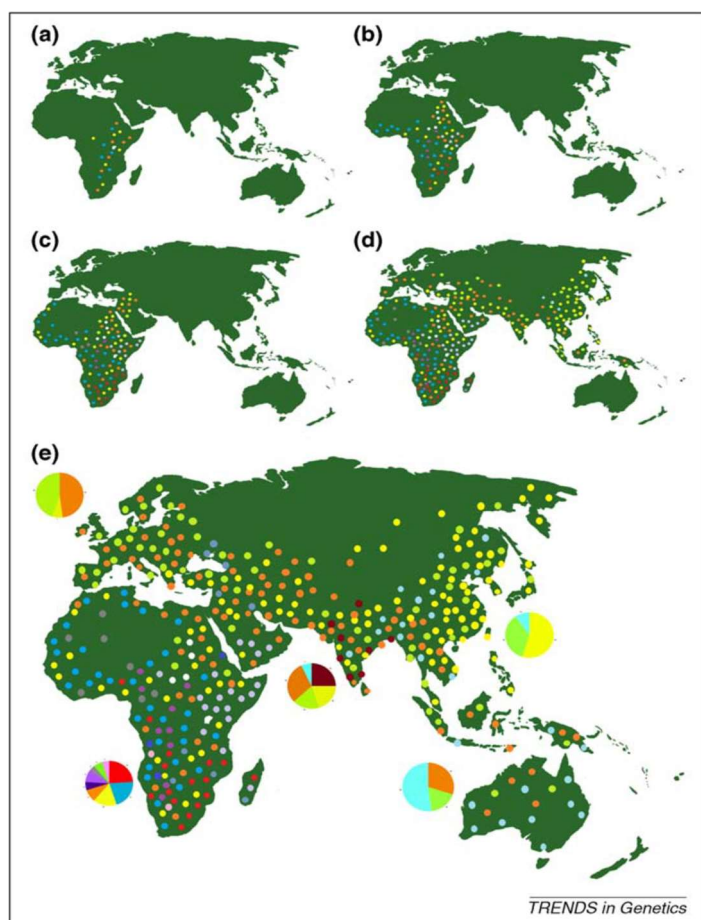


Figura 1. Visão esquemática da evolução da biodiversidade humana a partir de frequências alélicas de marcadores do tipo SNP. Os pontos coloridos representam genótipos diferentes, os gráficos de pizza da letra E representam as frequências alélicas em cinco regiões ao final do processo. Dados aproximados para os cinco panoramas: (a, b) > 60.000 anos atrás; (c) 60.000 anos atrás; (d) 40.000 anos atrás; (e) 30.000 anos atrás. Retirado de Barbujani & Colonna (2010).

Diversos trabalhos com marcadores autossômicos, DNAm e cromossomo Y chegaram ao mesmo resultado: populações africanas apresentam maior diversidade genética do que populações não-africanas (*The International HapMap Consortium*, 2005; Garrigan, et al., 2007; Jakobsson, et al., 2008; Li, et al., 2008; *The 1000 Genomes Project Consortium*, 2010; *The 1000 Genomes Project Consortium*,

2015), o que potencialmente corrobora a origem africana do humano moderno. Isso porque sua dispersão pelos continentes foi permeada por diversos gargalos de garrafa seletivos e efeito fundador, fazendo com que a diversidade genética fosse perdida ao longo dessa expansão (Ramachandran, et al., 2005).

Os padrões de variação de marcadores genéticos em regiões codificadoras e não-codificadoras não são necessariamente os mesmos, uma vez que mutações em regiões codificadoras podem ter impactos significativos na sobrevivência de um indivíduo. Regiões codificadoras – i.e., em genes, cujo exemplo de variabilidade será discutido no capítulo 2 deste trabalho – sofrem um monitoramento mais rigoroso, sendo que mutações nessas regiões tendem a desaparecer antes de alcançar frequência significativa na população. Essa pressão seletiva tende a diminuir a variabilidade genética encontrada nas regiões codificadoras, a não ser que a nova mutação tenha valor adaptativo, conferindo proteção ao indivíduo portador que poderá se reproduzir e aumentar a frequência dessa mutação na população.

Os marcadores neutros, por sua vez, se diversificam principalmente através de processos demográficos e deriva genética, uma vez que a seleção dessas regiões é frequentemente mais afrouxada (Kirk & Freeland, 2011). Dado esse cenário, uma mutação numa região neutra não-codificadora do genoma tem maiores chances, em comparação com regiões codificadoras, de aumentar de frequência em gerações futuras, já que possivelmente não terá impacto na capacidade de sobrevivência ou de reprodução do indivíduo (Tishkoff & Verrelli, 2003). Isso implica que a variabilidade genética relacionada a marcadores genéticos neutros é maior do que a de marcadores genéticos que estejam em regiões do genoma sob seleção.

Assim, esse trabalho tem por principal objetivo explorar a variabilidade genética humana sob duas perspectivas. O primeiro enfoque, na saúde, refere-se à variabilidade genética acessada por marcadores genéticos neutros, que *a priori* não têm impacto no *fitness* dos indivíduos, e que caracterizam as relações de variabilidade entre os grupos populacionais de uma forma não-enviesada, permitindo, portanto, que a ancestralidade genética de populações miscigenadas seja acessada, e que grupos com diferentes ancestralidades genéticas sejam distinguidos. Nesse caso, o modelo de marcadores neutros selecionado é um painel de 100 inserções *Alu*. No segundo

enfoque, na doença, a variabilidade genética de genes e regiões gênicas foram avaliadas usando como exemplo a susceptibilidade de indivíduos a formas severas da infecção pelo vírus da influenza H1N1. Levou-se em conta tanto a variabilidade genética global para determinados genes que, possivelmente, estão associados aos mecanismos celulares de resposta à infecção, quanto variantes específicas que já foram associadas a mau ou bom prognóstico em face à infecção pelo vírus da influenza.

No primeiro capítulo, buscou-se caracterizar as relações de variabilidade genética entre populações humanas de diversos continentes, introduzindo no cenário internacional grupos populacionais normalmente ignorados devido à sua alta miscigenação. Assim, duas populações miscigenadas brasileiras foram utilizadas como modelo de população miscigenada e acrescentadas a um painel de 50 outras populações não miscigenadas. Parâmetros clássicos de descrição de variabilidade foram estimados para 100 marcadores do tipo inserção *Alu*, que são bi alélicos, a fim de melhor compreender as relações genéticas entre populações miscigenadas e não-miscigenadas. Além disso, buscou-se estimar a composição/ancestralidade genética de populações miscigenadas a partir desse painel. Neste caso, o principal indicativo da variabilidade genética entre populações é a estimativa das frequências de cada um dos alelos para cada um dos 100 marcadores utilizados.

No segundo capítulo, buscou-se relacionar a variabilidade genética de marcadores do tipo SNP encontrada em diferentes populações humanas a dados epidemiológicos gerados após o surto pandêmico de influenza H1N1 em 2009. Neste caso, o objetivo foi buscar compreender como se distribui a variabilidade de regiões gênicas que estão, direta ou indiretamente, envolvidas na resposta imune a patógenos. Como modelo de patógeno, foi escolhido o vírus da influenza humana devido ao seu amplo espectro de prognósticos e impacto na saúde pública mundial todos os anos (Dawood, et al., 2012; Simonsen, et al., 2013; WHO, 2014). Neste contexto, a estimativa de variabilidade genética para regiões codificadoras está relacionada, principalmente, à diversidade alélica nas diferentes regiões dos genes como íntrons, exons, regiões regulatórias, e que sofrem diferentes pressões seletivas devido ao seu impacto no fenótipo do indivíduo, como será detalhado. Para finalizar, a distribuição global das frequências alélicas e genotípicas de marcadores genéticos

que já foram previamente associados a bom ou mau prognóstico da infecção pelo vírus da influenza serve como uma forma de acessar a variabilidade genética global frente a marcadores que não são neutros. A comparação entre as diferenças de frequências desses marcadores não neutros com as frequências dos marcadores neutros *Alu* pode esclarecer como a variabilidade genética é distribuída no genoma (regiões neutras x não-neutras), num âmbito global.

Capítulo I – Contextualização do Brasil no cenário genético mundial: uma história de variabilidade genética contada por 100 inserções *Alu*

I.1. Introdução

A história de povoamento humano das Américas foi permeada por sucessivos eventos demográficos, que incluíram episódios de mistura genética entre povos de três distintos continentes com perfil genético diferenciado entre eles. Inicialmente, o continente foi povoado por grupos humanos advindos do leste asiático que se dispersaram pelo continente americano a partir da entrada na América do Norte pelo Estreito de Bering (Reich, et al., 2012). Segundo a história reconstruída a partir de registros fósseis, a entrada na América do Norte se iniciou há aproximadamente 24 mil anos, tendo a dispersão finalizado por volta de 12 mil anos atrás, com o estabelecimento das populações nativas sul-americanas (Dillehay, 2003).

Entre 12 mil e 1.500 anos, entretanto, o continente americano se manteve praticamente isolado, uma vez que o Estreito de Bering, passagem da Ásia para a América, degelou ao final da glaciação impedindo o fluxo genético entre as populações doadoras e as populações americanas. Assim sendo, pode-se afirmar que a deriva genética (associada a diversos processos demográficos de expansão e contração de populações) foi o principal fator que moldou a composição genética das populações nativas americanas durante esse período (Raghavan, et al., 2014).

Entretanto, há 1500 anos o perfil genético das populações americanas começou a mudar. Com a chegada dos europeus, boa parte das populações que estavam estabelecidas no território americano foram dizimadas ou se amalgamaram com as que aqui chegavam. Durante o desenvolvimento e colonização dos territórios americanos, populações africanas subsaarianas foram trazidas em larga escala para servirem de mão-de-obra nas grandes lavouras que aqui foram estabelecidas. Esses três principais eventos migratórios (o povoamento das américas, a colonização europeia e o tráfico negreiro) foram fundamentais para o estabelecimento do que hoje conhecemos como perfil genético das populações americanas.

Essa mistura é extremamente recente se comparada a eventos migratórios e de dispersão em outros continentes. Além disso, a contribuição de cada população parental foi distinta quando se comparam as várias regiões americanas, sendo que

essa distribuição está relacionada tanto a posição geográfica dos distintos países, quanto a políticas internas dos mesmos (Godinho, et al., 2009). Como produto desse processo, hoje observa-se que determinadas regiões apresentam uma maior proporção de contribuição parental africana (tais como o Caribe), enquanto outras apresentam alta frequência de contribuição europeia, o que inclui espanhóis e italianos (América Central e do Sul), europeus de língua inglesa (América do Norte) e portugueses que chegaram ao Brasil (Godinho, 2008; Montinaro, et al., 2015).

1.1.1. Povoamento do Brasil

O Brasil é um exemplo clássico de país cuja população é derivada de eventos generalizados de miscigenação. Essa população é essencialmente tri híbrida, isso é, a formação contou, principalmente, com a contribuição genética de três grupos parentais: africanos, ameríndios e europeus. A intensa miscigenação desses três grupos resultou numa população altamente heterogênea e que apresenta grande variabilidade genética para diversos marcadores já estudados (Parra, et al., 2003; Godinho, et al., 2009).

A história de formação do Brasil se inicia com as grandes navegações, que trouxeram um grande número de migrantes europeus (colonização) e africanos (tráfico negreiro) e promoveram um encolhimento significativo das populações ameríndias que aqui viviam. Toda essa miscigenação pode ser capturada na população de Brasília (Tabela 1.2), já que sua formação, em 1960, foi marcada pela miscigenação recente de populações já estabelecidas de outras partes do país. Por isso, essa população pode ser considerada reflexo genético da população brasileira (Barcelos, 2006).

Por outro lado, no Brasil existem diversas comunidades, cidades e mesmo bairros que não estiveram sujeitos a tal grau de miscigenação. Dentre estas estão os quilombos, encontrados em todas as regiões do país, em especial na área rural. Os escravos africanos e afro derivados, quando conseguiam fugir das fazendas ou eram abandonados pelos seus senhores, organizavam-se em comunidades em locais de

difícil acesso. Estas comunidades, os *quilombos*, são grupos étnico-raciais, segundo critérios de auto atribuição, com trajetória histórica própria, dotados de relações territoriais específicas, com presunção de ancestralidade negra relacionada com a resistência à opressão histórica sofrida (Ministério da Cultura, 2004).

Os quilombos são de especial interesse na reconstrução histórica da população brasileira. Apesar de serem formados por indivíduos já miscigenados (em especial com europeu, euro-descendentes e ameríndios), essas comunidades conseguiram se manter semi-isoladas e configuram um cenário de pequena miscigenação recente se comparado à população brasileira como um todo (Pedrosa, 2006), tendo sido apontados como repositório genético de alelos ameríndios (Gontijo, et al., 2014). O maior quilombo brasileiro em extensão territorial é Kalunga, ocupando uma área de 258.000 hectares nos municípios de Cavalcante, Teresina e Monte Alegre, no estado de Goiás.

A composição genética de Brasília e Kalunga já foi estimada para diversos tipos de marcadores genéticos, tanto neutros quanto não-neutros: AIMs, STRs, inserções *Alu*, cromossomo Y, DNAMt (Barcelos, 2006; Pedrosa, 2006; Godinho, 2008; Ribeiro, 2009). Nestes trabalhos, foi possível observar como as populações parentais contribuíram para a composição genética atual de Brasília e Kalunga (figura 1). A caracterização desses marcadores neutros nas duas populações levou à conclusão que são populações geneticamente diferentes, ainda que tenham as mesmas populações formadoras em comum, o que vai de encontro com o que foi

encontrado para marcadores genéticos não neutros, que sinalizaram pequena diferenciação populacional entre Brasília e Kalunga (Toledo, 2011; Arcanjo, 2012).

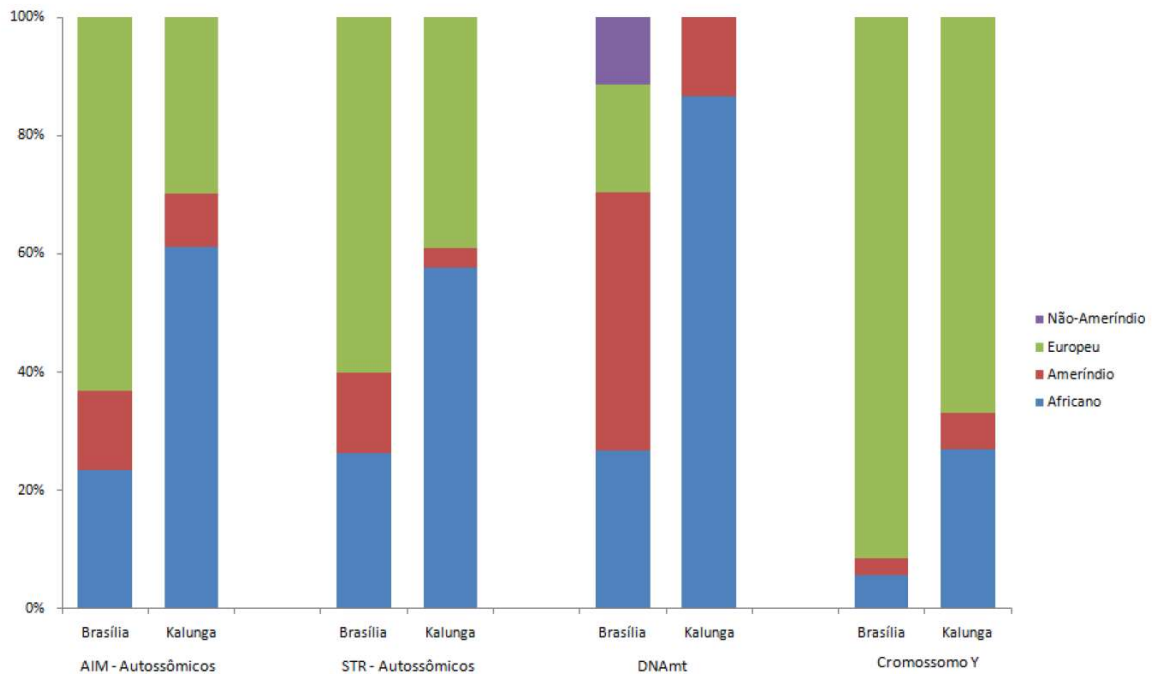


Figura I.1 Estimativa de contribuição genética das populações parentais para a composição genética de Brasília e Kalunga. Retirado de Toledo (2011).

O tipo de marcador genético utilizado pode trazer informações distintas sobre as estimativas de variabilidade, composição genética e história evolutiva de uma população (Wang, 2006; Kirk & Freeland, 2011). Portanto, é necessário avaliar o tipo e herança do marcador genético a ser utilizado, assim como entender suas limitações, para que não sejam tiradas conclusões fora do seu alcance. Tendo em mente a busca por variabilidade genética estável, diferenciação populacional compatível com a história evolutiva das populações e a acuidade na localização geográfica de amostras, marcadores do tipo *retrotransposon* surgem como ferramentas ímpares no estudo da variabilidade humana.

Elementos transponíveis são sequências de DNA que podem mudar sua localização e criar rearranjos inovadores no genoma, sendo agentes de variação genética. Dentre as várias famílias de elementos transponíveis destaca-se a inserção *Alu*, elemento SINE (*short interspersed nuclear element*) mais abundante do genoma

humano, alcançando mais de 1,1 milhão de cópias (Kriegs, et al., 2007; The 1000 Genomes Project Consortium, 2015b). As inserções *Alu* são sequências diméricas de aproximadamente 300pb, derivadas do RNA sinalizador 7SL (Ullu & Tschudi, 1984), sendo que cada inserção nova é um evento mutacional estável, único e unidirecional (Shedlock & Okada, 2000; Salem, et al., 2005). Essas sequências apresentam baixa taxa de inserção, são idênticas por descendência e seu estado ancestral é a ausência da inserção em determinado *locus* (Batzer & Deininger, 1991; Batzer, et al., 1994; Cordeaux, et al., 2006).

Diversos trabalhos têm demonstrado que um painel de 100 inserções *Alu*, quando utilizadas em conjunto, pode indicar ancestralidade geográfica de uma amostra (Bamshad, et al., 2001; Bamshad, et al., 2003; Watkins, et al., 2001; Watkins, et al., 2003). Contudo, as populações estudadas para esse conjunto de marcadores não são miscigenadas como as populações brasileiras aqui descritas. No único trabalho em que esse painel foi utilizado para amostras miscigenadas, foi observado que o sucesso de determinação de origem geográfica é muito menor do que em amostras não miscigenadas, porém apontando a ancestralidade do indivíduo (Ray, et al., 2005).

Populações miscigenadas apresentam uma constituição genética que representa a contribuição proporcional de cada população ancestral. Em decorrência disso, o genoma dessas populações são quebra-cabeças de sequências oriundas das diversas populações parentais. Ainda, quanto mais recente for a miscigenação, maior é o desequilíbrio de ligação, isto é, a distribuição não aleatória dos alelos na população. Como consequência, populações miscigenadas tendem a interferir na análise de relacionamento e/ou distância genética de populações não miscigenadas. As populações miscigenadas podem, potencialmente, levar duas populações parentais a apresentarem semelhanças/diferenças que não são condizentes com uma análise sem essas populações. Isso é, a inclusão de populações miscigenadas pode levar a distorções em árvores de relacionamento sugerindo similaridades não reais (Aistle & Balding, 2009; Hunley, et al., 2015).

Sabe-se que marcadores genéticos não se comportam da mesma forma em populações não miscigenadas e miscigenadas e que, infelizmente, a maioria dos

estudos de genética de populações tende a evitar o estudo destas últimas. Portanto, a caracterização desse painel de 100 inserções *Alu* em duas populações miscigenadas se mostrou especialmente relevante, pois pode acessar a eficiência desse conjunto de marcadores na determinação geográfica de amostras biológicas.

A pergunta que norteia o desenrolar desse capítulo, portanto, é se o painel de 100 inserções *Alu* desenvolvido por Batzer e colaboradores (Patente US20070003944 A1) permite inferir origem geográfica e ancestralidade genética em populações miscigenadas. Desta forma, o principal objetivo deste capítulo foi traçar o perfil genético das populações miscigenadas brasileiras a partir de um grande número de marcadores genéticos conservadores, configurado no painel de 100 inserções *Alu*.

Para alcançar esse objetivo geral, os seguintes objetivos específicos foram propostos:

- Caracterização do painel de 100 inserções *Alu* quanto a parâmetros descritivos (frequências alélicas e genotípicas, aderência ao EHW, desequilíbrio de ligação, heterozigose e estruturação populacional, sinais de seleção) em populações mundiais;
- Caracterização das populações brasileiras miscigenadas quanto ao painel de 100 inserções *Alu*;
- Contextualização das populações miscigenadas no cenário genético mundial estimado para populações distribuídas na Europa, África e Ásia;
- Identificação de grupos genéticos e estimativas de mistura genética para as populações miscigenadas.

I.2. Material e Métodos

I.2.1. Material biológico e genótipos cedidos

Para o desenvolvimento do estudo, foram selecionadas 102 amostras de DNA de habitantes da comunidade Kalunga (Goiás) e 60 de Brasília (Distrito Federal) (Figura I.1). Essas amostras fazem parte do banco de amostras do Laboratório de Genética Humana e tiveram aprovação do Comitê de Ética para serem utilizadas neste trabalho CEP 151/07 (Kalunga); CEP 028/2001 e CONEP 1259/2001 (Brasília) (Anexo I.1). O material biológico coletado foi sangue venoso periférico, que foi processado e de onde foi extraído DNA. Utilizou-se o kit comercial Illustra® da Amersham, para Kalunga e o protocolo *salting-out* para as amostras do Distrito Federal.



Figura I.2 Localização geográfica das populações Brasília e Kalunga dentro do território brasileiro.

Além das amostras de Brasília e Kalunga, outras 973 amostras de diferentes origens geográficas foram também acrescentadas a este trabalho, em

forma de genótipos previamente obtidos pela equipe do Batzer Laboratório de Genômica Comparativa (LA, USA). Os genótipos das populações da África, Europa, Ásia, Índia e Oriente-Médio, geradas em trabalhos anteriores (Bamshad, et al., 2003; Watkins, et al., 2003; Ray, et al., 2005; Witherspoon, et al., 2006; Ray, et al., 2007), foram gentilmente cedidas pelo Dr. Mark A. Batzer, da Louisiana State University. A relação de populações e o número de amostras que compõem o número final de indivíduos estudados neste trabalho está na tabela I.1.

Tabela I. 1 Relação das amostras utilizadas neste trabalho, com respectiva origem geográfica e ancestralidade. As amostras que não são brasileiras tiveram os genótipos gentilmente cedidos por Mark A. Batzer, PhD, e Jerilyn A. Walker, MSc (Louisiana State University, Baton Rouge, LA, EUA).

População	Origem	Ancestralidade	Número de amostras	Referência
Brasília	Brasil	Mista	60	Presente estudo
Kalunga	Brasil	Mista	102	Presente estudo
Total Mista			162	
Kenya	Quênia	Africana	20	(Ray, et al., 2005)
Nigeria	Nigéria	Africana	10	(Ray, et al., 2005)
Alur	Congo	Africana	10	(Watkins, et al., 2003)
Biaka Pygmy	África	Africana	5	(Watkins, et al., 2003)
Hema	Congo	Africana	18	(Watkins, et al., 2003)
Mbuti Pygmy	África	Africana	4	(Watkins, et al., 2003)
Nande	Congo	Africana	19	(Watkins, et al., 2003)
Nguni-Xhosa	África	Africana	14	(Watkins, et al., 2003)
!Kung	África	Africana	15	(Stoneking, et al., 1997)
Congo Pygmy	Congo	Africana	33	(Watkins, et al., 2003)
Sotho-Tswana	África	Africana	26	(Watkins, et al., 2003)
Tsonga	África	Africana	7	(Watkins, et al., 2003)
Africa	África	Africana	4	(Watkins, et al., 2003)
Total Africana			185	

População	Origem	Ancestralidade	Número de amostras	Referência
Cambodia	Cambója	Asiática	12	(Watkins, et al., 2003)
Chinese	China	Asiática	23	(Watkins, et al., 2003; Ray, et al., 2005)
Japanese	Japão	Asiática	30	(Watkins, et al., 2003; Ray, et al., 2005)
Malasian	Malásia	Asiática	6	(Watkins, et al., 2003)
Mixed Asian	Ásia	Asiática	14	(Ray, et al., 2005)
South East Asia	Sudeste Asiático	Asiática	10	(Ray, et al., 2005)
Taiwan	Taiwan	Asiática	23	(Ray, et al., 2005)
Vietnamese	Vietnã	Asiática	9	(Watkins, et al., 2003)
Total Asiática			127	
Basque	Espanha	Européia	10	(Ray, et al., 2005)
Europe	Polônia	Européia	10	(Bamshad, et al., 2001)
France	França	Européia	20	(Watkins, et al., 2003)
Iberia	Espanha	Européia	10	(Ray, et al., 2005)
Kuopio-Turku	Finlândia	Européia	20	(Watkins, et al., 2003)
Russian	Rússia	Européia	19	(Ray, et al., 2005)
Utah	Estados Unidos	Européia	68	(Bamshad, et al., 2003)
Total Européia			157	
Brahmin	Índia	Indiana	18	(Bamshad, et al., 2001)
Dora	Índia	Indiana	25	(Watkins, et al., 2001)
Gond	Índia	Indiana	20	(Watkins, et al., 2001)
India	Índia	Indiana	9	(Watkins, et al., 2003)
Indo Pakistan	Paquistão	Indiana	9	(Watkins, et al., 2003)
Irula	Índia	Indiana	33	(Watkins, et al., 2001)
Kapu	Índia	Indiana	9	(Bamshad, et al., 2001)

População	Origem	Ancestralidade	Número de amostras	Referência
Kashatriyas	Índia	Indiana	11	(Bamshad, et al., 2001)
Madiga	Índia	Indiana	28	(Bamshad, et al., 2001)
Mala	Índia	Indiana	26	(Bamshad, et al., 2001)
Niyogi	Índia	Indiana	13	(Bamshad, et al., 2001)
Pakistan	Paquistão	Indiana	19	(Watkins, et al., 2003)
Relli	Índia	Indiana	18	(Bamshad, et al., 2001)
Santal	Índia	Indiana	16	(Watkins, et al., 2001)
Telaga	Índia	Indiana	27	(Bamshad, et al., 2001)
Turpu	Índia	Indiana	19	(Bamshad, et al., 2001)
Vydika	Índia	Indiana	22	(Bamshad, et al., 2001)
Vysya	Índia	Indiana	10	(Bamshad, et al., 2001)
Yadava	Índia	Indiana	53	(Bamshad, et al., 2001)
Total Indiana			394	
Bahrain	Bahrein	Oriente-Médio	20	(Ray, et al., 2005)
Egypt	Egito	Oriente-Médio	20	(Ray, et al., 2005)
Jordan	Jordânia	Oriente-Médio	20	(Ray, et al., 2005)
Syria	Síria	Oriente-Médio	20	(Ray, et al., 2005)
UAE	Emirados Árabes Unidos	Oriente-Médio	20	(Ray, et al., 2005)
Total Oriente-Médio			100	
Total Amostras			1125	

1.2.1.1. África

O grupo étnico África, para esse trabalho e outros que já foram publicados para as 100 inserções *Alu*, é composto por 13 subpopulações. Inicialmente, considerou-se agrupar as subpopulações de acordo com grupo linguístico, etnia

reconhecida e/ou proximidade geográfica. No entanto, um estudo mais aprofundado da história da África mostrou que isso poderia ser um grande equívoco e tornar artificiais as relações entre as subpopulações (Vansina & Obenga, 2011). Desta forma, optou-se por manter a classificação das populações de acordo com local de origem. A localização geográfica das populações africanas pode ser observada na figura 1.2.

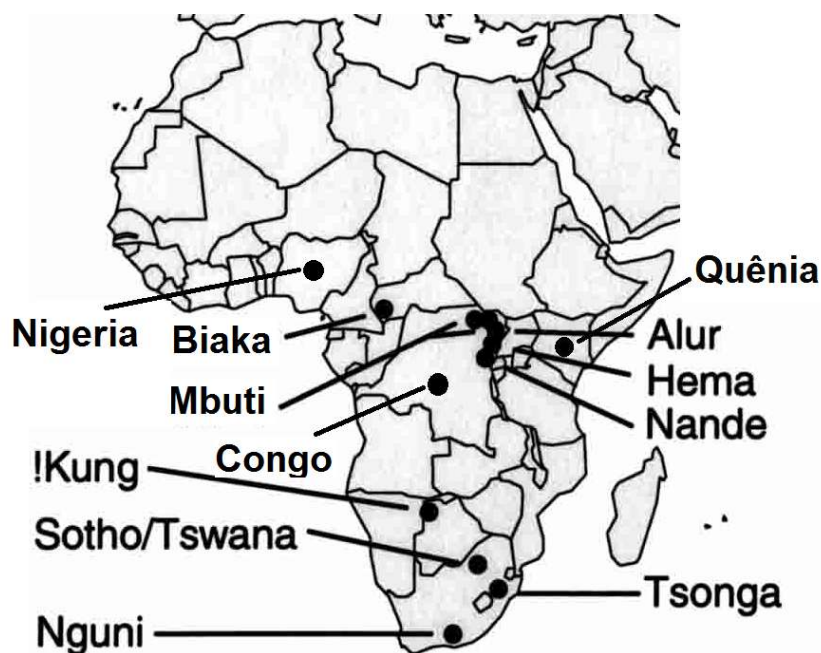


Figura 1.2 Localização geográfica das populações africanas cujos genótipos foram gentilmente cedidos pelo Dr. Mark A. Batzer. Figura adaptada de Watkins e colaboradores (2003).

1.2.1.2. Ásia

O grupo étnico Ásia é composto por oito subpopulações, sendo elas Leste da Ásia, China, Japão, Sudeste da Ásia, Cambódia, Malásia, Taiwan, e um grupo de amostras sem origem geográfica definida, que foi chamada de Asiáticos de Ancestralidade Mista. A figura 1.3 mostra a localização geográfica das populações asiáticas cujas origens foram informadas.

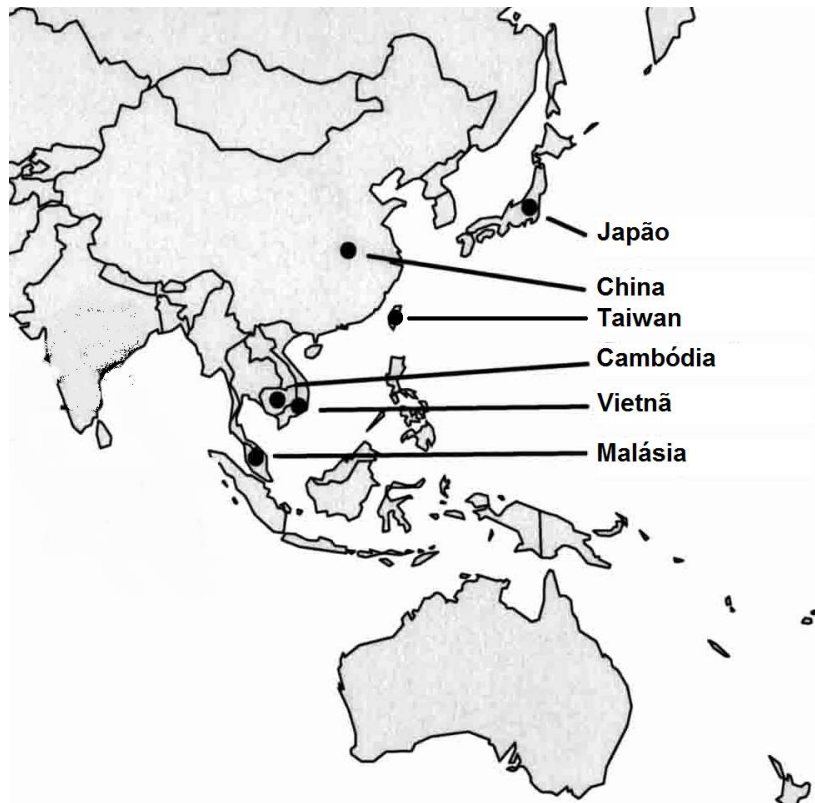


Figura I.3 Localização geográfica das amostras asiáticas para as quais havia informação de local de coleta. Adaptado de Watkins e colaboradores (2003).

I.2.1.3. Europa

O grupo europeu incluído neste trabalho conta com seis subpopulações: França, Polônia, Finlândia, Ibéria, País Basco, Utah e Rússia (Figura I.4). Ainda que a população de Utah seja americana, seu histórico de formação e estudos conduzidos por grandes consórcios de variação genética humana (como o extinto HapMap, o 1000 Genomes Project e o CEPH) determinaram que esse grupo de amostras são representativos de ancestralidade do norte europeu (Dausset, et al., 1990; *The*

International HapMap Consortium, 2005; The 1000 Genomes Project Consortium, 2010).

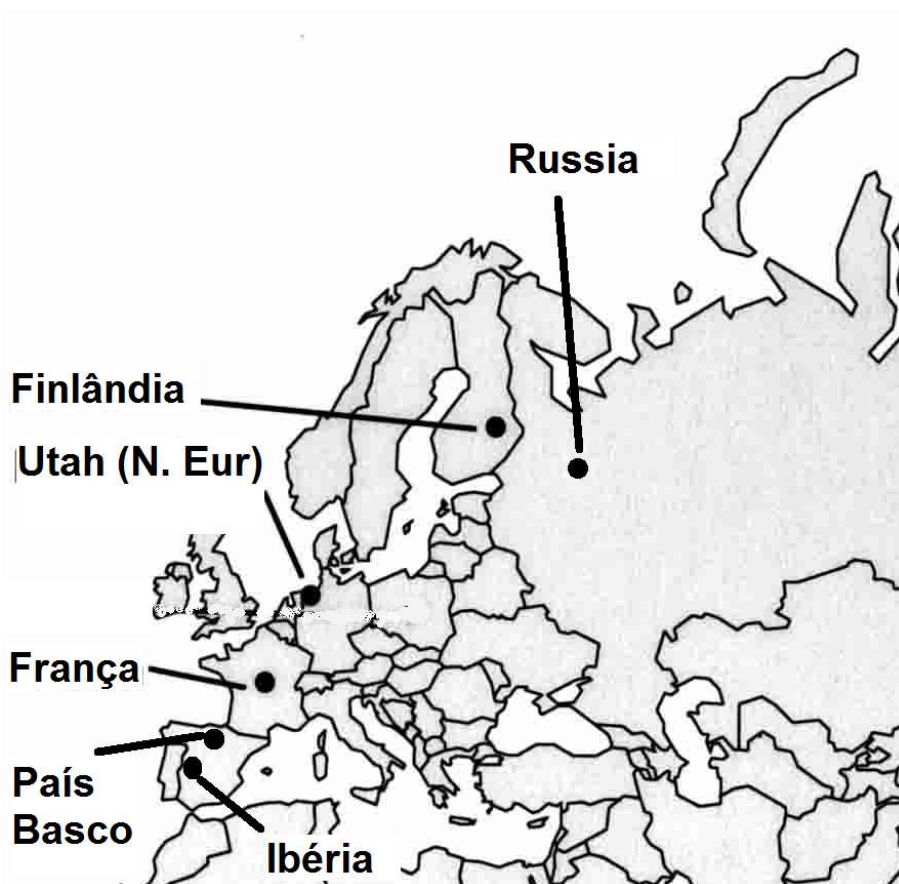


Figura I.4 Localização geográfica das amostras do grupo europeu. Apesar de Utah se localizar nos Estados Unidos da América, a ancestralidade deste grupo representa a população do Norte Europeu, e, por esse motivo, Utah aparece localizada no norte da Europa. Adaptado de Watkins e colaboradores (2003).

I.2.1.4. Índia

O grupo étnico indiano é, talvez, o mais complexo e mais numeroso dos grupos étnicos já estudados para as 100 inserções *Alu* que fazem parte deste trabalho. Não somente por ter o maior número de subpopulações, mas, principalmente, pela subestruturação populacional que é observada neste país (Kashyap, et al., 2006). Essa subestruturação é resultado de endogamia e de casamentos preferenciais praticados de acordo com a cultura local, que é baseada em castas, além de várias ondas migratórias de países colonizadores e comerciantes que acabaram por deixar

seu legado linguístico e genético na composição da população atual (Barbujani & Colonna, 2010).

O campo amostral do grupo étnico indiano é composto por 19 subpopulações, sendo eles: Brahmin, Nyiogi, Kashatryias, Vydika, Vysya (castas altas); Telaga, Turpu, Kapu, Yadava (castas médias); Relli, Madiga, Mala (castas baixas, intocáveis); Dora, Gond, Irula, Santal (tribos isoladas); India, Indo-Paquistaneses e Paquistaneses (indivíduos sem casta definida). As descrições de casta, origem e afinidade estão descritas em Bamshad e colaboradores (2001) e Watkins e colaboradores (2001). A localização geográfica das populações está explicitada na figura I.5.

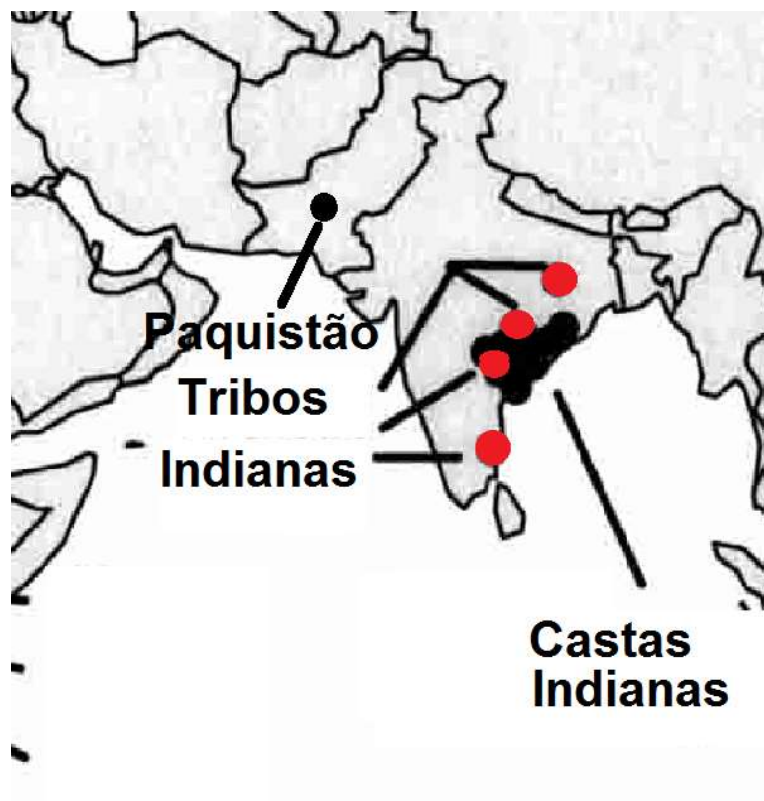


Figura I.5 Localização geográfica das amostras indianas utilizadas neste trabalho. Os pontos em vermelho mostram as tribos indianas, e os pontos pretos evidenciam as castas e a população paquistanesa. Adaptado de Watkins e colaboradores (2003).

1.2.1.5. Oriente Médio

O grupo de amostras do Oriente Médio é composto por cinco subpopulações: Jordânia, Bahrein, Emirados Árabes Unidos, Síria e Egito. As amostras do Oriente-Médio também são amostras inéditas deste trabalho, visto que ainda não foram publicadas em jornais científicos. A localização geográfica das amostras está ilustrada na figura 1.6.

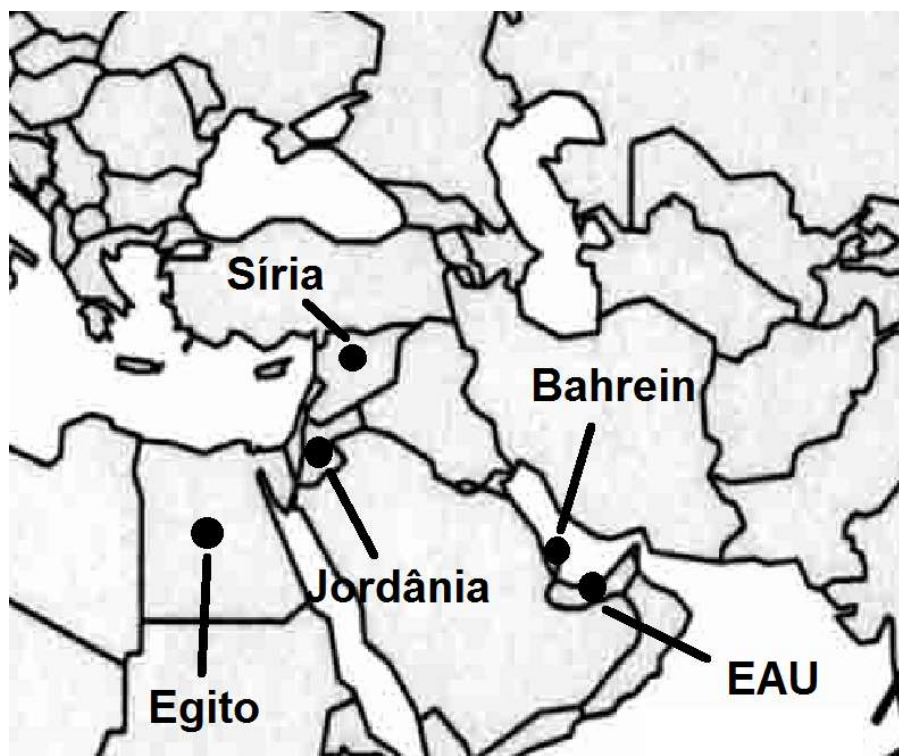


Figura 1.6 Localização geográfica das amostras do Oriente-Médio. Adaptado de Watkins e colaboradores (2003).

1.2.2. Amplificação de genoma total

Considerando o número de marcadores estudados e a quantidade de DNA necessária por amostra para cada marcador, determinou-se que seria necessário fazer amplificação de genoma total para as amostras com concentração igual ou menor que 6µg de DNA (Apêndice I.1). Para tanto, as amostras selecionadas foram submetidas ao kit GenomiPhi V2 da Amersham Biosciences®, com adaptações às instruções do fabricante (Anexo I.2). O kit GenomiPhi é baseado na técnica de *Isothermal Strand Displacement*, diferente do protocolo de amplificação de genoma total utilizando-se PCR. As vantagens do kit em questão estão relacionadas ao tipo

de polimerase utilizada (Phi29, altamente eficiente e com sistema rigoroso de correção de erros), a iniciadores curtos inespecíficos (hexâmetros) e ao sistema de deslocamento de fita nova, que permite que várias polimerases transcrevam continuamente uma mesma fita de DNA deslocando a fita nova do molde enquanto escrevem a nova fita, isso tudo mantendo a temperatura da reação constante, a 30°C (Figura I.7). Após o protocolo, os genomas totais amplificados foram purificados para retirar quaisquer resquícios de proteínas e/ou reagentes do kit, usando protocolo padrão do Laboratório de Genômica Comparativa da *Louisiana State University* (Anexo I.3).

I.2.3. Seleção dos marcadores

Nesse trabalho foi utilizado um painel de 100 inserções *Alu* selecionadas pela equipe do Dr. Mark A. Batzer (da *Louisiana State University*) para acessar a ancestralidade genômica de indivíduos e populações e, inclusive, deram origem a uma patente de material para determinação de origem geográfica de amostras desconhecidas (Patente US20070003944 A1). Esses marcadores já foram utilizados

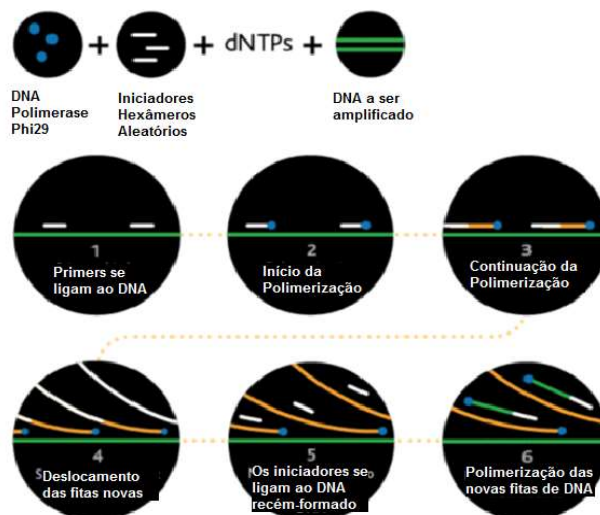


Figura I.7 Esquema de funcionamento do kit de amplificação do genoma total. 1) os iniciadores hexâmetros se ligam às fitas de DNA da amostra que se pretende amplificar; 2). Inicia a polimerização pela enzima DNA Polimerase Phi29; 3) A polimerização se dá de um iniciador até o próximo ponto. 4) Novas fitas de DNA se soltam parcialmente do DNA molde, de forma a manter a nova fita em constante síntese. 5) Os iniciadores se ligam às novas fitas de DNA, recém-sintetizadas; 6) Novas fitas são sintetizadas a partir delas. Figura adaptada da bula do kit (Amersham Biosciences ®).

em diversos trabalhos do grupo, seja uma seleção parcial ou total do painel (Bamshad, et al., 2003; Watkins, et al., 2003; Ray, et al., 2005). Foi mostrado que esse painel de marcadores é suficiente para determinar origem geográfica de amostras (Ray, et al., 2007), diferenciar populações e mostrar tendências de agrupamento por ancestralidade (Bamshad, et al., 2003). Também se mostraram úteis na avaliação da estruturação genética de populações humanas (Watkins, et al., 2003; Witherspoon, et al., 2006), e para fins forenses (Ray, et al., 2005). A tabela contendo os iniciadores e o nome para cada marcador está contida no Anexo I.4.

O painel utilizado é composto de 100 inserções *Alu* autossômicas distribuídas ao longo de todo o genoma humano. A escolha da inserções visou observar a dinâmica de variabilidade desses marcadores nas populações de uma forma global, diminuindo a influência de desvios de equilíbrio devido à ligação entre marcadores.

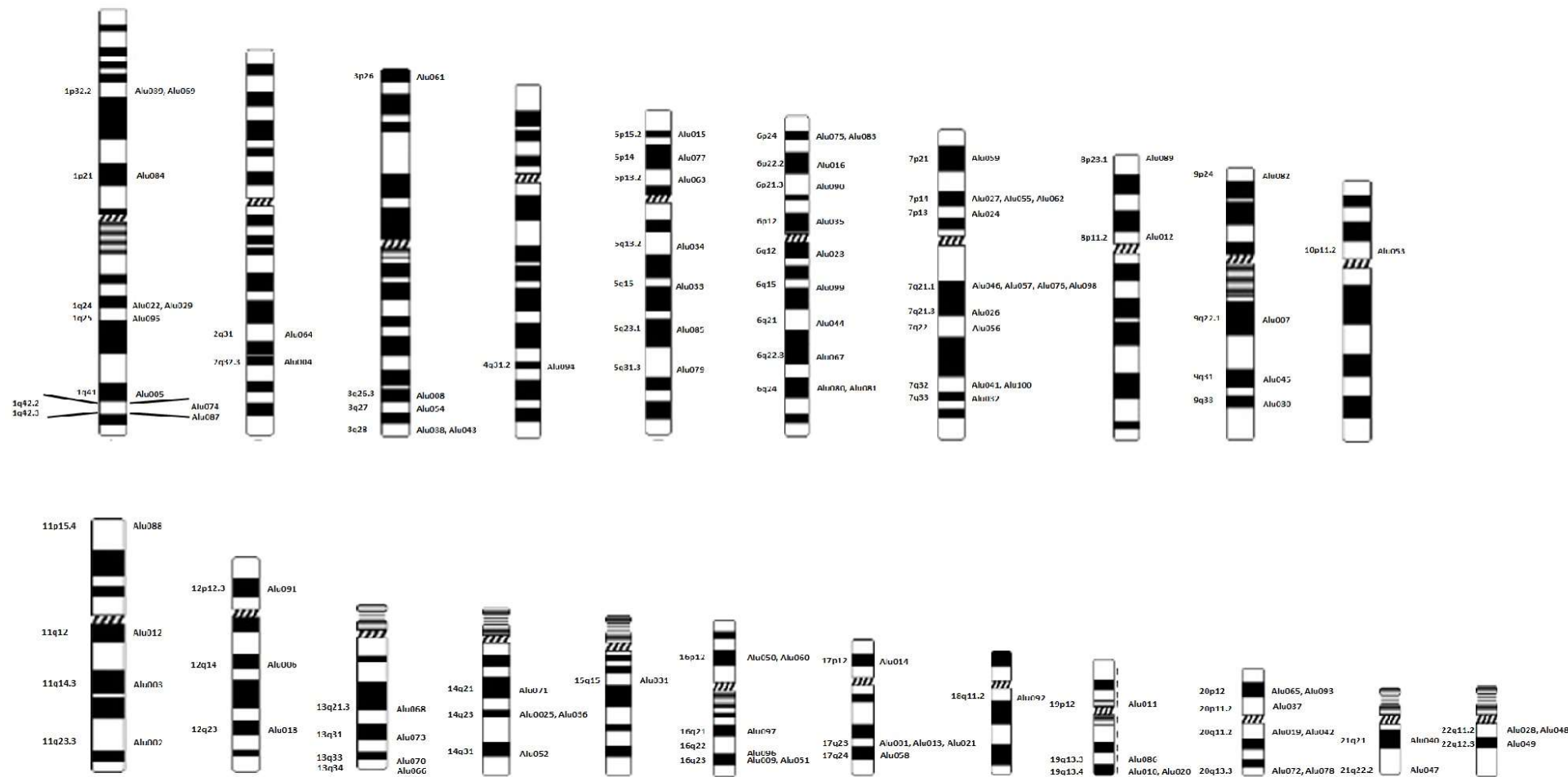


Figura 1.8 Localização cromossômica das 100 inserções *Alu* analisadas neste trabalho. À esquerda do cromossomo estão indicadas as coordenadas cromossômicas aproximadas de cada *locus* e à direita de cada cromossomo estão indicadas as respectivas inserções *Alu* que se encontram nessas coordenadas.

1.2.4. Genotipagem dos marcadores

A genotipagem das 100 inserções *Alu* foi feita pelo método tradicional de amplificação de fragmento por PCR (tabela I.2) em reações de volume final 25µl. Em seguida, foi feita a visualização do amplificado em gel de agarose a 2% e corado com brometo de etídio (figura I.9). Os processos de PCR e eletroforese são descritos abaixo.

Tabela I. 2 Reagentes utilizados para as reações de PCR e suas concentrações de estoque e final por tubo de reação de PCR.

Reagente	□ estoque	□ por reação
Tampão (50mM KCl; 10mM TrisHCl; pH8,4)	10X	1X
dNTP	100mM cada	0,2 mM cada
Primer	100µM cada	200nM cada
Rediload™ (ThermoFisher® – Waltham, MA)	10X	1X
MgCl ₂ (Invitrogen – Waltham, MA)	50mM	Varia de acordo com Anexo I.4
Taq Polymerase (Invitrogen – Waltham, MA)	5U/µl	5U
DNA	5ng/µl	25ng

As condições de PCR para cada marcador foram diferentes, especialmente com relação à temperatura de anelamento dos iniciadores e à concentração de cloreto de magnésio (MgCl₂) na reação (anexo I.4). Todas as reações foram feitas em termocicladores modelo MyCycler da BioRad v.3.0.1 (Hercules, CA), seguindo o programa mostrado na tabela I.3:

Tabela I. 3 Programa geral do termociclador utilizado para amplificação dos marcadores selecionados.

Passo	Ciclo	Temperatura	Tempo	
1	Denaturação	95°C	5 mins	
2	Denaturação	94°C	30 segs	Repetir 30X
3	Anelamento dos iniciadores	Diferente para cada marcador (Anexo I.4)	30 segs	
4	Extensão do fragmento	72°C	30 segs	
5	Extensão final	72°C	10 mins	
6	Inativação da enzima	4°C	Infinito	

À PCR seguiu-se o processo de eletroforese em gel de agarose. Para tanto, 10g de agarose eram diluídos em 500ml de tampão TBE 0,5X (Tris-Ácido Bórico-EDTA, Amresco® - Solon, OH), para que o gel ficasse à concentração de 2% de agarose. Em seguida, a solução era aquecida em microondas por 4 minutos, para que a agarose dissolvesse por completo na solução. À essa solução eram adicionados 15µl de Brometo de Etídio e o gel era vertido no recipiente próprio para solidificação. Após solidificado, as amostras e o marcador de peso molecular (100pb da Invitrogen® - Waltham, MA) eram colocadas nos poços e o gel era imerso em tampão TBE 0,5X. Os eletrodos eram ligados aos respectivos polos da cuba de eletroforese e a corrida acontecia por 40min a 200V, com amperagem e potência variáveis. A visualização dos fragmentos foi feita através de luz UV e documentada através do sistema fotodocumentador BioRad ChemiDoc XRS (Hercules, CA).

Como procedimento padrão, a genotipagem dos alelos para cada marcador em cada indivíduo se dá pela altura das bandas de DNA no gel (figura I.9). Fragmentos menores tendem a migrar para o polo positivo mais rápido do que fragmentos maiores. Como a diferença entre alelos para as inserções *Alu* é, aproximadamente, 300 pares de base (Jagadeeswaran, et al., 1981), a visualização dos alelos se dá de forma fácil e inequívoca seguindo o protocolo acima descrito, como está exemplificado na figura

I.9. Os tamanhos dos fragmentos com inserção e sem inserção *Alu* para cada marcador estão listados no Anexo I.4.

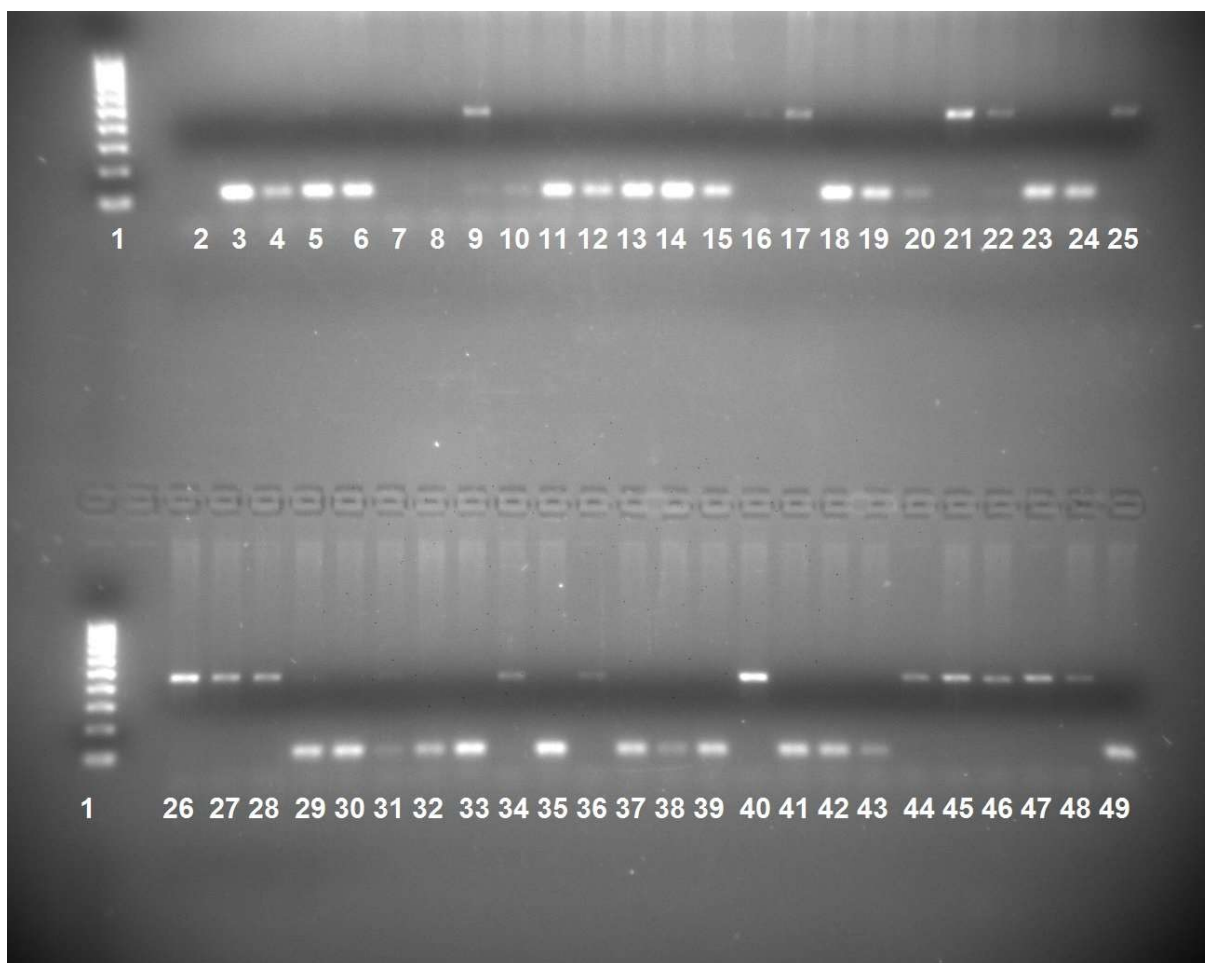


Figura I.9 Gel de eletroforese em agarose a 2% corado com brometo de etídio. Neste gel pode-se observar a genotipagem de algumas amostras para o marcador *Alu048* (Yb8NBC120). Todos os géis foram feitos da mesma forma, utilizando-se o marcador de peso molecular 100bp (bandas que se iniciam em 100pb e são incrementadas com esse mesmo intervalo, número 1 da figura). Em seguida, o controle negativo (mix da PCR sem amostra, 2), o controle ancestral (chimpanzé, 3), o controle de amplificação em humanos (célula *HeLa*, 4) e as amostras em seguida. Podemos observar claramente que os indivíduos 4, 5, 6, 10, 11, 12, 13, 14, 15, 18, 19, 20, 23, 24, 29, 30, 32, 33, 35, 37, 38, 39, 41, 42, 43 e 49 são homocigotos para a ausência de inserção *Alu* (apresentam apenas uma banda de aproximadamente 150pb); os indivíduos 9, 22, 31 e 48 são heterocigotos (apresentam uma banda de 150pb e uma banda de 450pb), os indivíduos 16, 17, 21, 25, 26, 27, 28, 34, 36, 40, 44, 45, 46 e 47 são homocigotos para a inserção *Alu* (apresentam apenas uma banda de 450pb) e que os indivíduos 7 e 8 não apresentaram amplificação do marcador, tendo sido repetidos em uma PCR posterior.

Foram observadas diferenças no sucesso de amplificação das 100 inserções *Alu* entre Brasília e Kalunga. Isso se deve, possivelmente, à qualidade do DNA extraído, já que foram conduzidos procedimentos diferentes para a extração de DNA do sangue das amostras de Brasília (*salting out*) e Kalunga (kit comercial). No entanto, não houve prejuízo das análises estatísticas, já que todas as inserções *Alu* apresentaram mais de 90% de sucesso de amplificação nessas amostras.

1.2.5. Análises estatísticas

As análises estatísticas foram feitas em diversos níveis, iniciando pela descrição dos parâmetros intra-populacionais em cada população brasileira, e terminando com análises inter-populacionais mais complexas em um número de indivíduos superior a 1000, incluindo-se populações não miscigenadas da África, Europa, Ásia, Índia e Oriente-Médio. Cabe salientar que é a primeira vez que essas 100 inserções são caracterizadas, em conjunto, em amostras miscigenadas da América do Sul.

1.2.5.1. Índices-padrão intrapopulacionais

As frequências alélicas para cada um dos marcadores foi obtida através de contagem simples dos alelos presentes em cada população (Brasília e Kalunga). Após isso, a estimativa foi feita utilizando o software Arlequin 3.5 (Excoffier & Lischer, 2010).

A partir das frequências alélicas obtidas para cada marcador em Brasília e Kalunga, foi testada a aderência ao Equilíbrio de Hardy-Weinberg (Guo & Thomson, 1992) e a probabilidade de desequilíbrio de ligação entre os marcadores (Slatkin & Excoffier, 1996). Não foi feita inferência haplotípica porque o conjunto de marcadores se encontra distribuído pelos 22 cromossomos autossômicos.

Foram também estimadas a diversidade molecular de cada população e as diferenças par-a-par entre elas, por meio de estimativa simples de frequência dos

alelos diferentes presentes em cada população. Encerrando os índices-padrão intrapopulacionais, foi feita estimativa de heterozigose em cada população. Todas as estimativas acima citadas foram realizadas utilizando-se a configuração padrão do software Arlequin 3.5, conforme instruções do manual (Excoffier & Lischer, 2010).

1.2.5.2. Índices-padrão interpopulacionais

A partir dos dados gerados no item anterior, foi possível estimar o índice de diferenciação populacional (F_{ST}) entre as populações estudadas (Brasília e Kalunga) e entre as populações para as quais haviam dados disponíveis (tabela I.1). O F_{ST} foi estimado através do método de Slatkin (1991), e sua significância foi estimada segundo Excoffier e colaboradores (1992), permutando os genótipos individuais entre as populações. O teste de F_{ST} é baseado nas diferenças par-a-par, comparando diretamente as frequências alélicas entre os pares de populações (Weir & Cockerham, 1984). O p-valor é a medida-limite da probabilidade que eventos ao acaso resultem na mesma situação que os dados observados. E quando abaixo do valor estipulado (0,05 para este trabalho), indica que a hipótese de igualdade entre as populações (H_0 , hipótese nula) é rejeitada, o que significa que as diferenças encontradas entre essas populações não se devem ao acaso, mas a algum fator evolutivo que esteja atuando nessas populações. A partir da linearização do F_{ST} , como descrito por Slatkin (1995), foi possível estimar outros parâmetros interpopulacionais, tais como a distância de Reynolds (Reynolds, et al., 1983), o número absoluto de migrantes entre as populações – M (Excoffier & Lischer, 2010) e a distância de Nei (Nei & Li, 1979).

Também foram feitas Análise de Componente Principal (PCA), através do pacote *Adegenet* (Jombart, 2008). A PCA é uma análise bastante informativa, que mostra se as diferenças genéticas observadas entre os grupos podem ser decompostas em vetores em um plano cartesiano. A visualização dos resultados dessas análises foi feita através do pacote *scatterplot3d* (Ligges & Mächler, 2003) e de outras funções do pacote básico do R.

Para analisar as relações entre os grupos de populações, e a existência de estrutura populacional entre as amostras estudadas, foi utilizado o software Structure 2.3.4 (Pritchard, et al., 2000). A inferência de estrutura populacional é feita a partir de um método de agrupamento baseado em modelos, utilizando-se dados de genótipos de marcadores genéticos não ligados, e pode ser usado para identificar populações em um conjunto de amostras, atribuir indivíduos à populações, identificar migrantes e indivíduos miscigenados (Pritchard, et al., 2000). O modelo funciona a partir da premissa de que existem K populações que são, cada uma, caracterizadas por um perfil de frequências alélicas diferentes. A partir do estabelecimento desse número de populações (que pode tanto ser determinado *a priori*, quanto inferido através do próprio programa), as amostras, ou indivíduos, são probabilisticamente atribuídos a uma das populações (ou mais de uma, caso as genótipos de um indivíduo indiquem que ele é miscigenado).

Para iniciar as estimativas, é necessário que o pesquisador determine alguns parâmetros (ou premissas). Os autores frisam que essas configurações não podem ser determinadas por meio de extrapolações teóricas: ou já se sabe exatamente quais parâmetros se quer usar, de acordo com as perguntas a serem respondidas; ou é necessário que várias corridas sejam feitas, com várias configurações diferentes, para se estipular empiricamente os melhores parâmetros (Pritchard, et al., 2010). Os parâmetros escolhidos foram determinados a partir da revisão feita por Porras-Hurtado e colaboradores (2013). As configurações, então, foram *burn-in* de 10.000 iterações, com 100.000 replicações da corrida, três réplicas para cada número de K (variando de 3 a 7), modelo de mistura (ADMIXTURE) com utilização de informação prévia para as amostras (LOCPRIOR com POPFLAG exceto para Brasília e Kalunga). Foram testados diversos números de K para que se pudesse determinar o melhor número de grupos, já que o número de amostras utilizadas neste trabalho foi maior do que em Ray e colaboradores (2005).

Para fazer a estimativa da composição genética de qualquer população miscigenada, é necessário que se conheçam as frequências alélicas das populações doadoras. Ou seja, é necessário que se conheça o perfil genético das populações que contribuíram geneticamente para a formação da população analisada para o conjunto de marcadores que se está estudando.

No caso das populações brasileiras, as populações doadoras são, conhecidamente, africana, europeia e ameríndia. Contudo, os dados para ameríndios utilizando o painel de 100 inserções *Alu* ainda não foram gerados, apesar de Battilana e colaboradores (2006) terem publicado sete dos marcadores que constam no painel. Dessa forma, optou-se por utilizar o grupo de populações asiáticas como substituta para a doadora ameríndia, assim como proposto por Montinaro e colaboradores (2015).

Para verificar se a dissimilaridade entre populações ameríndias e as populações asiáticas seriam grandes o suficiente para não permitir tal substituição, foram feitos testes de diferenciação populacional para cada um dos sete loci publicados por Battilana e colaboradores (2006). Os testes mostraram grande similaridade entre as populações ameríndias e asiáticas para os marcadores HS4.32, HS4.65, HS4.75, APO e PV92. Para esses marcadores, as populações ameríndias apresentaram diferenciação significativa principalmente de Vietnã, China, Japão e Camboja (apêndice I.2). Por isso, as populações asiáticas foram consideradas substitutas das populações ameríndias para a estimativa de composição genética das populações miscigenadas brasileiras.

Para realizar a caracterização das relações genéticas entre os grupos populacionais mundiais, o campo amostral (número total de amostras trabalhadas neste trabalho) foi dividido em vários subgrupos em determinados momentos das análises. O objetivo principal desta prática foi entender melhor as relações entre as populações descritas para as 100 inserções *Alu* estudadas neste trabalho. Espera-se, com isso, que as relações evidenciadas pelas análises interpopulacionais possam servir de guia para futuros trabalhos que tratem destes grupos populacionais.

I.3. Resultados e Discussão

Neste trabalho, os resultados e a discussão serão apresentados concomitantemente, buscando manter a fluidez e a conexão entre os diversos tópicos apresentados. Inicia-se pela caracterização das populações brasileiras individualmente, comparando as estimativas quando possível, para somente depois contextualizar essas duas populações no cenário global de amostras. Em seguida, serão apresentadas as estimativas dentro dos grupos continentais para explicitar as relações de variabilidade existentes em cada um deles. Entende-se que assim o leitor pode ter uma visão gradativa de como a variabilidade se distribui: inicia-se na população, passamos pelo grupo populacional, terminando na análise global da variabilidade.

I.3.1. Parâmetros intrapopulacionais

I.3.1.1. Equilíbrio de Hardy-Weinberg

Kalunga apresentou 12% dos marcadores fora do Equilíbrio de Hardy-Weinberg em uma primeira análise (marcadores 6, 11, 15, 48, 50, 58, 60, 78, 85, 86, 88, 96 e 97) e Brasília 13% (marcadores 9, 23, 33, 34, 48, 51, 53, 58, 84, 85, 88, 96 e 98). Porém, após a Correção de Bonferroni, apenas os marcadores 58 e 96 em Brasília se mantiveram fora do equilíbrio. Como é esperado que até 5% dos marcadores possam estar fora do equilíbrio de Hardy-Weinberg em ambas as populações, podemos considerar que esse desvio está dentro do esperado e, portanto, não é significativo.

A primeira hipótese que justificaria o desequilíbrio de Hardy-Weinberg nesses marcadores em Brasília e Kalunga é o erro de genotipagem. Para contornar esse problema, os géis foram genotipados por três pesquisadores diferentes. Após a comparação das genotipagens, apenas 2,3% apresentaram discordância de

genotipagem (32 em 1358), o que não influenciaria o desequilíbrio encontrado. Dessa forma, estabelece-se que a não-aderência desses marcadores nas populações de Brasília e Kalunga ao Equilíbrio de Hardy-Weinberg está possivelmente associado a outros fatores, tais como a recente miscigenação sofrida por essas populações (250 anos para Kalunga e 60 anos para Brasília).

Battilana e colaboradores (2006), estudando sete dos 100 marcadores contantes no painel, reportaram a não aderência ao Equilíbrio de Hardy-Weiberg para alguns marcadores desse painel em populações ameríndias da América do Sul (Aché, Gavião, Guarani, Kaingang, Quechua QT, Surui e Zocó).

1.3.1.2. Diversidade gênica

Brasília se mostrou monomórfica para quatro das 100 inserções *Alu* (21, 26, 55 e 91), enquanto Kalunga se mostrou monomórfica para seis (14, 17, 21, 26, 44 e 91). Apesar de apresentar mais *loci* monomórficos, Kalunga apresentou maior diversidade gênica entre os *loci* ($0,344781 \pm 0,166779$) do que Brasília ($0,301744 \pm 0,146863$). Considerando que o cálculo da diversidade gênica é uma função do número de cópias de genes disponíveis em cada população (Librado & Rozas, 2009) e que o número de cópias gênicas em Kalunga (204) é quase o dobro do de Brasília (120 para Brasília), postulou-se, *a priori*, que essa diferença entre as populações fosse decorrente dessa questão amostral. Porém, essa proposta não foi confirmada visto que a simulação das estimativas com o número de amostras mais próximo (60 Brasília, 65 Kalunga), a diversidade de Kalunga ainda se mostrou maior ($0,343906 \pm 0,166796$). Isso indica, então, que a maior diversidade gênica encontrada em Kalunga é função das características de formação da população.

A distribuição da diversidade gênica entre os diversos grupos analisados não se difere substancialmente quando se utiliza as 100 inserções *Alu*. Apesar disso, é digno de nota que as populações africanas apresentam maior diversidade gênica, mostrando as maiores heterozigoses observadas dentre as populações estudadas por Watkins e colaboradores (2003) (tabela I.4). Essa observação foi confirmada pelas

estimativas de diversidade gênica para o grupo africano no presente trabalho (tabela I.4).

Kalunga apresentou diversidade gênica mais próxima à encontrada para o grupo africano do que os grupos europeu e asiático. Essa semelhança decorre da constituição genética dessa população, produto direto de sua história. Kalunga é um quilombo, caracterizado por alta contribuição genética africana, como estimado por AIMS autossômicos (~60%), STRs autossômicos (~60%) e DNAm (~80%) (Toledo, 2011). Por outro lado, a população de Brasília apresentou uma diversidade gênica próxima do estimado para a Europa. Estimativas de ancestralidade genética na formação dessa população evidenciam uma maior contribuição europeia (61%) do que africana (23%) (Godinho, 2008). Dessa forma, podemos afirmar que a diversidade gênica encontrada para os 100 marcadores *Alu* nas duas populações brasileiras corrobora as estimativas anteriores de ancestralidade.

Tabela I. 4 Estimativas de Diversidade gênica, heterozigose observada e esperada para as populações, baseadas nas frequências de 100 inserções Alu, em contraste com o estimado por Watkins e colaboradores (2003).

População	Presente Estudo				Watkins et al (2003)	
	N	Loci	Div. Gênica	Het. Obs.	Het. Esp.	Div. Gênica
Brasília	60	100	0,302	0,300	0,337	-
Kalunga	102	100	0,345	0,336	0,374	-
África	(7-33)	(24-99)	(0,301-0,367)	(0,345-0,500)	(0,341-0,409)	(0,322-0,354)
Alur	10	83	0,357	0,401	0,397	0,3544
Nande	19	92	0,340	0,398	0,382	0,3393
Hema	18	94	0,345	0,395	0,373	0,3503
Congo	33	97	0,301	0,370	0,341	-
Biaka-Mbuti	10	79	0,334	0,365	0,385	0,3221
!Kung San	14	24	0,253	0,371	0,385	0,3390
Northern Sotho	11	95	0,332	0,365	0,379	-
Sotho-Tswana	15	78	0,322	0,387	0,378	0,3411
Zulu-Xhosa	14	80	0,337	0,356	0,374	-
Quênia	20	99	0,316	0,345	0,367	-
Nigéria	10	98	0,315	0,373	0,388	-
Tsonga	7	69	0,367	0,500	0,409	0,3510
Ásia	(6-28)	(53-100)	(0,274-0,323)	(0,356-0,432)	(0,372-0,412)	(0,2947-0,3526)
Vietnã	9	88	0,305	0,412	0,390	0,2965
Sem origem definida	14	98	0,289	0,378	0,379	-
Camboja	12	99	0,288	0,409	0,378	0,2947
Malásia	6	100	0,326	0,432	0,412	0,3526

População	Presente Estudo				Watkins et al (2003)	
	N	Loci	Div. Gênica	Het. Obs.	Het. Esp.	Div. Gênica
China	23	99	0,322	0,385	0,392	0,3178
Japão	28	53	0,274	0,385	0,388	0,3064
Sudeste Asiático	10	91	0,323	0,386	0,405	-
Taiwan	23	100	0,292	0,356	0,372	-
Europa	(10-68)	(70-98)	(0,244-0,308)	(0,324-0,380)	(0,354-0,0,377)	(0,2927-0,3009)
França	20	94	0,308	0,357	0,354	0,3009
Polonês	10	70	0,275	0,324	0,354	0,2798
Finlândia	20	84	0,289	0,373	0,361	0,2927
Ibéria	10	98	0,305	0,380	0,377	-
País Basco	10	97	0,284	0,361	0,370	-
Rússia	19	98	0,301	0,350	0,359	-
Utah	68	76	0,244	0,334	0,337	0,2964
Índia	(9-53)	(14-99)	(0,162-0,324)	(0,335-0,402)	(0,353-0,399)	(0,2993-0,3220)
Brahmin	18	98	0,315	0,361	0,381	0,3128
Dora	27	99	0,301	0,385	0,363	0,3050
Gond	22	93	0,290	0,335	0,361	0,3029
India	9	82	0,300	0,362	0,394	-
Indo-Paquistaneses	9	99	0,309	0,369	0,392	-
Irula	34	97	0,291	0,362	0,353	0,3068
Kapu	9	88	0,324	0,401	0,392	0,3117
Kashatryias	11	98	0,305	0,399	0,384	0,3031
Madiga	29	99	0,309	0,365	0,361	0,3103

População	Presente Estudo				Watkins et al (2003)	
	N	<i>Loci</i>	Div. Gênica	Het. Obs.	Het. Esp.	Div. Gênica
Mala	26	98	0,310	0,365	0,375	0,3113
Nyogi	14	96	0,318	0,384	0,380	-
Paquistão	20	88	0,293	0,371	0,376	-
Relli	19	98	0,317	0,402	0,398	0,3220
Santal	16	14	0,162	0,369	0,367	0,3007
Telaga	28	99	0,305	0,356	0,366	-
Turpu	18	97	0,297	0,369	0,377	-
Vydika	22	98	0,295	0,342	0,358	-
Vysya	10	17	0,183	0,388	0,399	0,2993
Yadava	53	97	0,294	0,370	0,364	0,3127
Oriente-Médio	20	(45-95)	(0,292-0,318)	(0,320-0,333)	(0,347-0,362)	-
Jordânia	20	94	0,298	0,320	0,347	-
Bahrein	20	93	0,318	0,333	0,359	-
EAU	20	81	0,297	0,321	0,362	-
Síria	20	95	0,301	0,327	0,353	-
Egito	20	45	0,292	0,324	0,355	-

N = número de indivíduos amostrados; *Loci* = número de *loci* com menos de 10% de amostras sem genótipo; d.p. = desvio-padrão associado a cada estimativa.

1.3.1.3. Desequilíbrio de ligação

Determinadas bandas cromossômicas apresentam uma maior concentração de inserções *Alu* selecionadas para o painel. Esse é o caso, por exemplo, da banda 7q21.1, que contém quatro inserções – 046, 057, 075 e 098. Análise de desequilíbrio de ligação entre esses marcadores em populações com baixa miscigenação (Europa, África e Ásia) mostrou que essas inserções não apresentam desequilíbrio de ligação entre elas. Inclusive, nas populações africanas, o marcador 098 não apresentou desequilíbrio de ligação com absolutamente nenhum dos outros 99 marcadores presentes no painel de 100 inserções *Alu*.

Em Brasília, todos os marcadores genéticos, com exceção de um, apresentaram-se em desequilíbrio de ligação com ao menos um dos outros. Em Kalunga, por sua vez, seis marcadores não apresentaram desequilíbrio de ligação com nenhum outro marcador. Se comparamos com populações mais antigas ou mesmo isoladas, observamos que o número de marcadores em equilíbrio de ligação é muito maior do que em populações novas e miscigenadas, como é o caso do Brasil. Por exemplo, na população do País Basco foram encontrados 30 *loci* que não estavam em desequilíbrio de ligação com nenhum outro marcador dos 100 avaliados. Já Brasília e Kalunga apresentaram, para cada marcador genético, uma média de 49,48 e 27,36 *loci* em desequilíbrio, respectivamente. A média para as outras populações ficou entre 2,08 *loci* em desequilíbrio por marcador, para os Indo-Paquistaneses, e 8,36, para os poloneses. Exceção a esse quadro foi observada na população de Utah ser considerada uma população representativa do norte da Europa, a recente formação dessa população (por volta de 500 anos) pode ter causado a miscigenação de diferentes populações do norte europeu, gerando um grande desequilíbrio de ligação, uma média de 35,1 *loci* em desequilíbrio por marcador. Pfaff e colaboradores (2001) demonstraram que o processo de miscigenação cria desequilíbrio de ligação entre marcadores genéticos, tanto os efetivamente ligados quanto os que não são ligados. Ao comparar o desequilíbrio de ligação estimado para as 100 inserções *Alu* das populações miscigenadas Brasília e Kalunga a populações menos miscigenadas, como os grupos africano e europeu, observou-se o mesmo cenário encontrado por

Pfaff e colaboradores (2001). Além disso, a estimativa de desequilíbrio de ligação para a população de Utah também sugere miscigenação, um resultado bastante interessante, e que já foi encontrado indiretamente (Wang, et al., 2010), se considerarmos que a maior referência de banco de dados de variabilidade genética humana (The 1000 Genomes Project Consortium, 2010) considera a população de Utah como não miscigenada. Isso pode acarretar em erros metodológicos para muitos trabalhos que partem da premissa que esse grupo de amostras é não-miscigenado.

1.3.2. Parâmetros interpopulacionais

O teste exato de diferenciação entre Brasília e Kalunga indicou que estas não são significativamente diferentes com relação às 100 inserções *Alu*, ainda que o F_{ST} tenha indicado diferenciação moderada ($F_{ST}=0,05933$, $p=0,00000$). Brasília apresentou diferenciação genética com as outras populações mundiais aqui analisadas, com exceção do País Basco ($F_{st} = 0,00953$; $p = 0,12613 \pm 0,27842$), Síria ($F_{st} = 0,00171$; $p = 0,40541 \pm 0,0389$) e Egito ($F_{st} = -0,01858$; $p = 0,99099 \pm 0,0030$). A figura 1.10 explícita a relação de similaridade entre populações europeias/oriente-médio e Brasília e entre as populações africanas e Kalunga.

As regiões mais claras da figura 1.10 explicitam menores valores de diferenciação populacional dentro dos grupos de populações africanas, asiáticas, europeias, indianas e do oriente-médio (da parte superior da diagonal, para baixo, respectivamente). Pode-se observar que a diferenciação populacional entre Brasília e as populações do grupo do Oriente-Médio é bastante pequena, praticamente como se aquela população pertencesse a esse grupo (figura 1.10). Também fica explícita a similaridade entre as populações do Oriente-Médio (e Brasília) às populações do grupo europeu, entre Kalunga e as populações africanas, e a grande diferenciação populacional entre as populações africanas e todos os outros grupos populacionais (regiões mais escuras da figura).

Matrix of pairwise F_{ST}

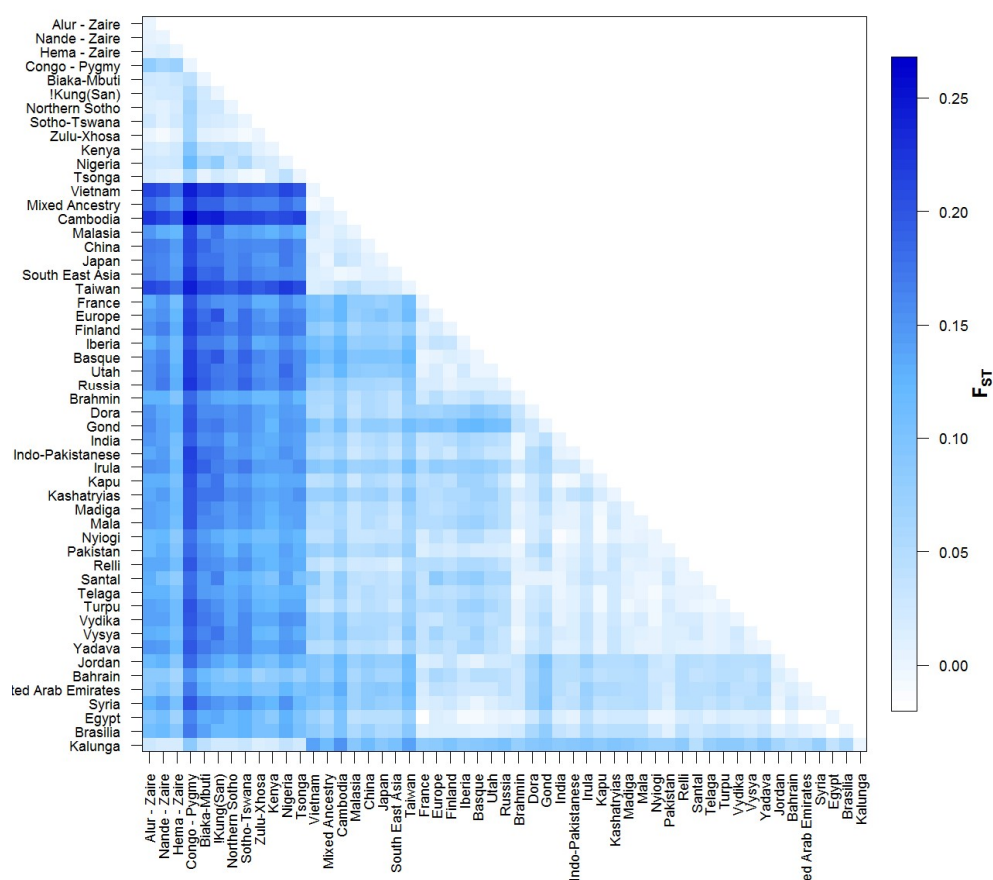


Figura I.10 Índice de diferenciação populacional (par a par) entre as populações analisadas neste trabalho, baseado nas diferenças de frequências alélicas das 100 inserções *Alu*.

A análise de componentes principais incluindo todas as amostras aqui analisadas, definiu quatro grupos de populações (figura I.11). As coordenadas plotadas para esse gráfico configuram a média de coordenadas por grupo, uma vez que o número de indivíduos tornou a plotagem das coordenadas individuais pouco informativa. O primeiro componente principal (eixo x) divide as populações africanas de não africanas, enquanto que o segundo componente principal (eixo y) divide as demais populações em dois grupos: Europa+Oriente Médio e Ásia (incluindo Índia), sendo possível observar uma continuidade populacional entre os dois últimos grupos.

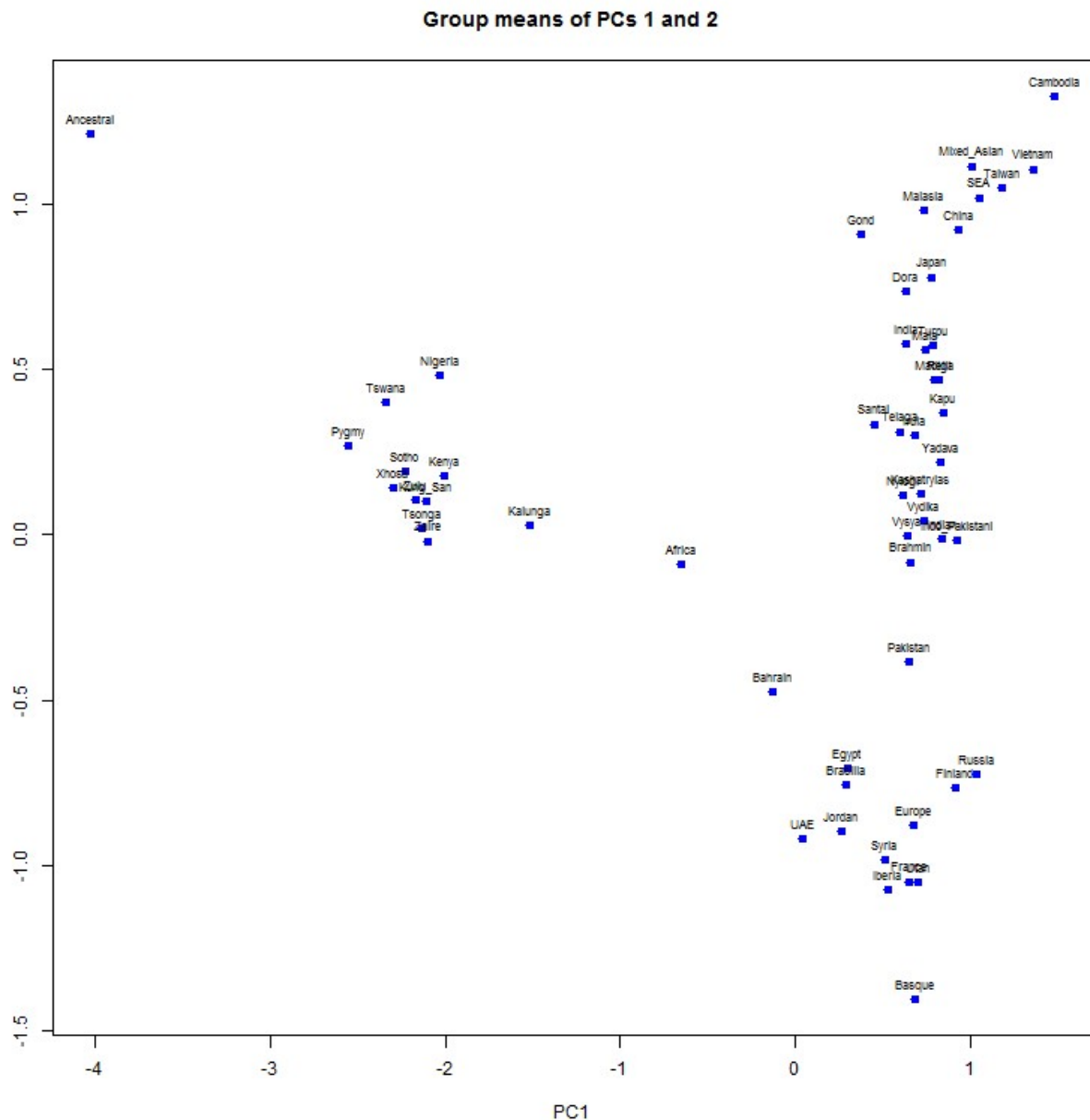


Figura I.11 Análise de componentes principais para o conjunto de populações utilizando dados de 100 inserções *Alu*.

Explorando melhor a variação dos grupos populacionais, buscou-se detalhar também o terceiro componente principal de variação. A figura I.12 mostra a análise de componentes principais com as coordenadas dos três primeiros componentes principais. Observa-se que o terceiro componente principal é capaz de diferenciar os grupos asiáticos e indianos, mas não é capaz de diferenciar satisfatoriamente os grupos europeus e do Oriente Médio. Brasília e Kalunga se posicionam em grupos distintos no gráfico de componentes principais, a primeira no

grupo que contém populações europeias e Oriente Médio, enquanto a segunda se localiza próximo a populações africanas subsaarianas, o que deve ser reflexo do histórico de formação e composição genética dessas populações. De acordo com análise de outros marcadores genéticos, Brasília tem uma alta ancestralidade européia, enquanto que Kalunga, alta ancestralidade africana (Oliveira, et al., 2004; Godinho, 2008).

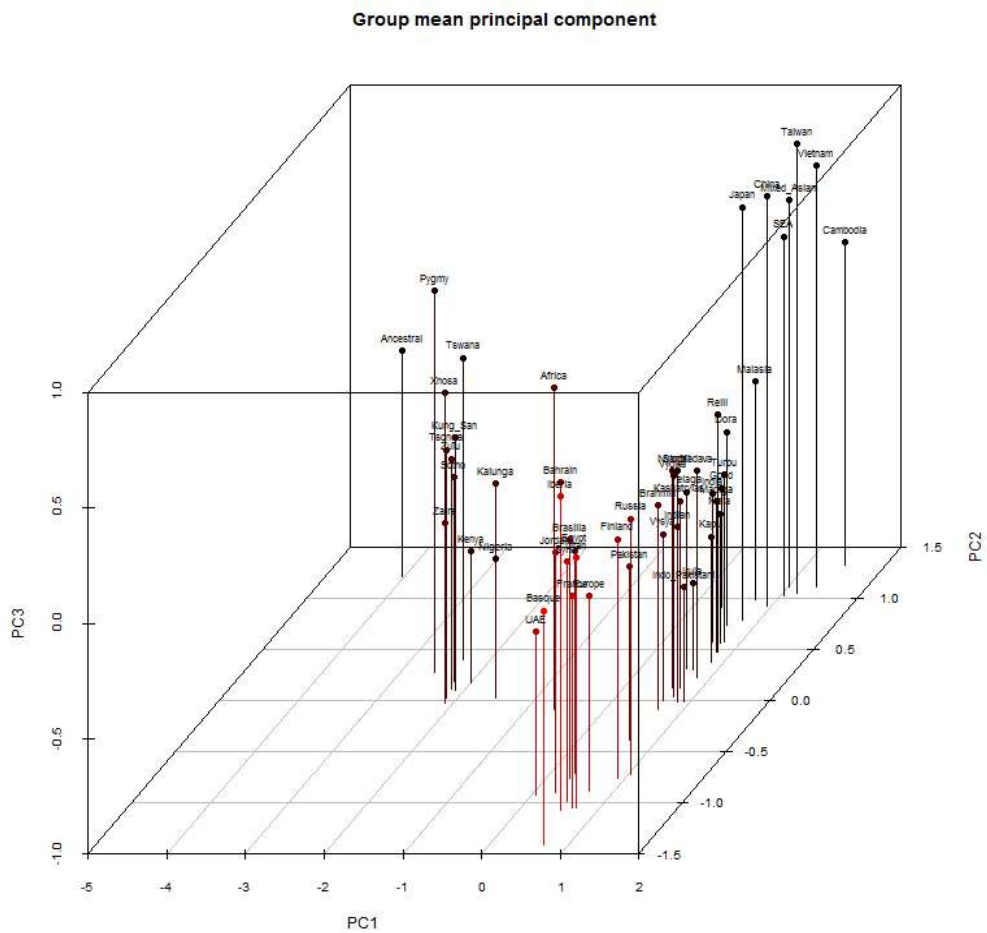


Figura I.12 Representação em 3D dos três primeiros componentes principais de variação genética para as populações, baseada em 100 inserções *Alu*.

Para detalhar melhor a relação entre as populações brasileiras, Brasília e Kalunga, e as populações com as quais apresentaram maior similaridade (figura I.10), foram feitas análises de componentes principais (PCA) adicionando Brasília a um grupo Europa/Oriente Médio e Kalunga ao grupo africano. Pôde-se observar, pelos componentes principais 1 e 2 (que capturam a maior variação dentro do grupo) para o grupo Brasília/Europa-Oriente Médio, que o grupo é dividido em Bahrein e todas as outras populações, não sendo possível discriminar a maior parte das populações (Figura I.13). A similaridade genética entre as populações da Europa e do Oriente Médio não é espantosa, uma vez que 65% dos haplótipos de DNAm_t existentes hoje na Europa foram introduzidos no continente em tempos históricos recentes, tais como a Romanização, a conquista árabe da Península Ibérica e Sicília e do período de tráfico negreiro (Cerezo, et al., 2012).

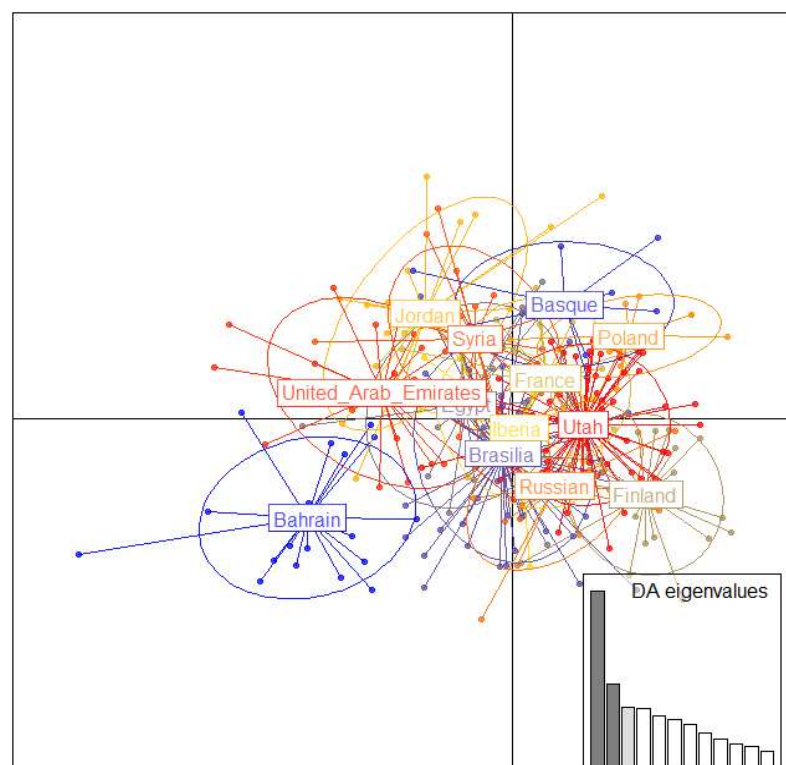


Figura I.13 Análise de componentes principais para o grupo Brasília/Europa/Oriente Médio. O eixo x representa a variação captada pelo primeiro componente principal, e o eixo y representa a variação captada pelo segundo componente principal. O box à direita evidencia os componentes principais, as barras hachuradas representam os componentes principais plotados e quanto de variação eles capturam.

Uma vez que a análise de componente principal não foi eficiente para resolver as relações genéticas entre Brasília e as populações do Oriente Médio e da Europa, fez-se necessário utilizar outra estratégia para resolver essas relações. A utilização de um algoritmo de clusterização de amostras do software *Structure* identificou três grupos aos quais as amostras Brasília/Oriente Médio/Europa poderiam ser designadas de acordo com as composições genéticas individuais (Figura I.14). Conhecendo a composição genética da população de Brasília por marcadores informativos de ancestralidade (Godinho, 2008), pode-se identificar que a cor azul é relativa à ancestralidade europeia, a cor verde é relativa à ancestralidade africana e a cor vermelha é relativa à ancestralidade ameríndia. Por mais que pareça inusitado que exista um componente ameríndio nas populações do Oriente-Médio e Europa, a própria análise de componente principal sugere que o painel de inserções Alu aqui estudadas é eficiente para diferenciar africanos de não-africanos. Por esse motivo, o componente africano é mais facilmente detectado pelo conjunto de marcadores, quando presentes na população em questão. Sendo assim, a clusterização realizada indica que a população de Brasília é geneticamente mais parecida com as populações do Oriente Médio do que com as populações da Europa, justamente pela presença do componente genético africano que inexistente nesse último grupo.

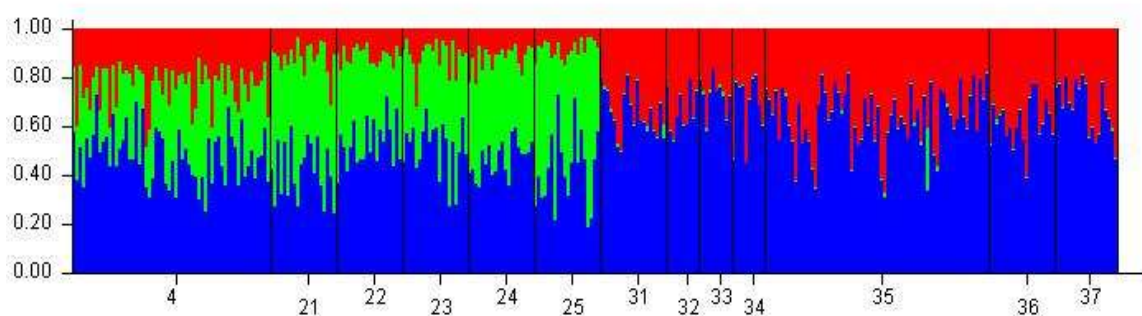


Figura I.14 Clusterização das amostras de Brasília (4), Oriente Médio (21-25) e Europa (31-37) de acordo com as frequências das 100 inserções Alu. Os grupos determinados pelo Structure são as cores azul, verde e vermelha.

A análise de componente principal do grupo África/Kalunga, por sua vez, mostrou três grupos definidos (Figura I.15). O primeiro componente principal diferencia o grupo que contém as populações pigmeias do Congo, com limites bem definidos e nenhuma sobreposição aos outros dois grupos de populações, um

segundo grupo contendo todas as outras populações africanas, e um terceiro grupo composto apenas pelas amostras de Kalunga. Esse aparente isolamento da população de Kalunga deve estar relacionado à miscigenação da população brasileira, que tem componentes europeus e ameríndios na sua composição genética, enquanto que as populações africanas não apresentam miscigenação com outros grupos continentais. Já o isolamento das tribos pigmeias do Congo ao restante das populações africanas é estimado em 60.000 anos (Patin, et al., 2009), e por isso o painel de inserções *Alu* parece ter sido capaz de identificar divergência populacional antiga.

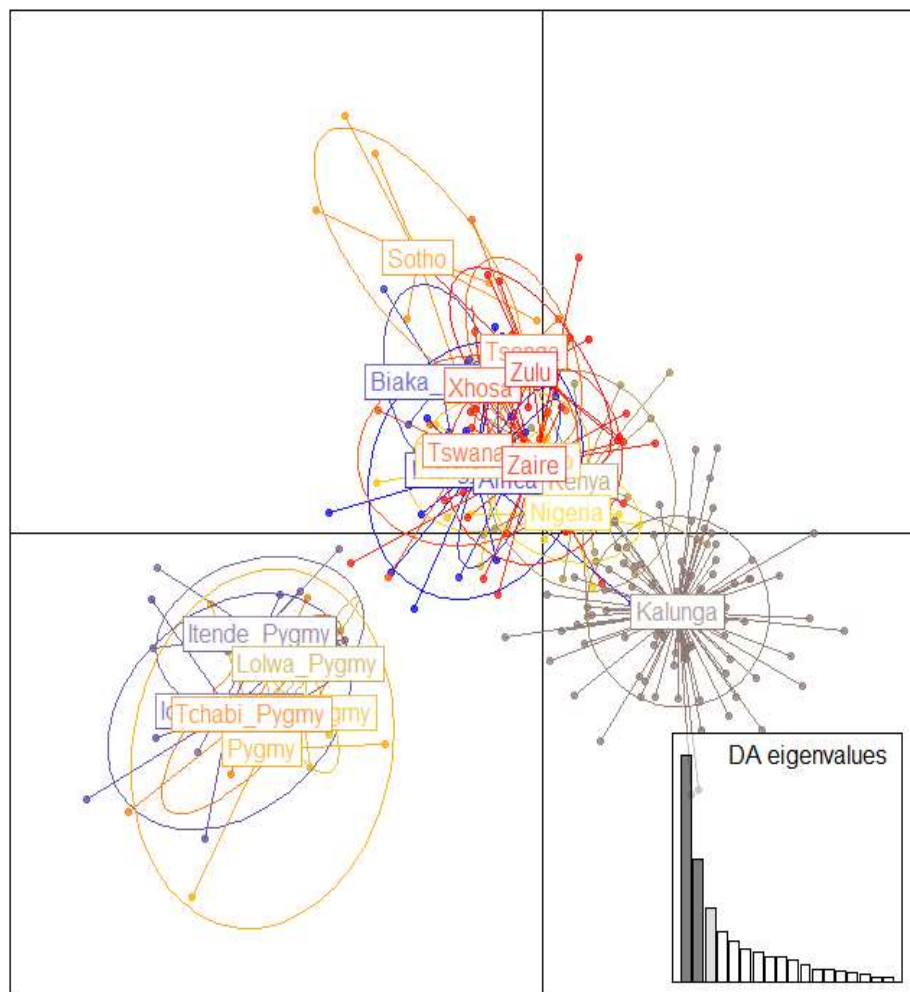


Figura I.15 Análise de componente principal para o grupo África/Kalunga. O eixo x representa a variação captada pelo primeiro componente principal, o eixo y representa a variação captada pelo segundo componente principal. O box à direita evidencia os componentes principais, as barras hachuradas representam os componentes principais plotados e quanto de variação eles capturam.

A análise de clusterização para o grupo de populações África/Europa/Kalunga evidenciou o motivo do isolamento desta última ao grupo africano apresentado na figura I.15. Como pode ser observado na figura I.16, a análise de clusterização identificou três grupos aos quais as amostras África/Europa/Kalunga poderiam ser designadas. Kalunga apresenta dois componentes em sua composição, um de origem africana (verde) e um de origem europeia (vermelho). O componente em azul, muito frequente nas tribos pigmeias (número 16), provavelmente representam um componente africano mais antigo (>60.000 anos), uma vez que pode servir para diferenciar bem as populações pigmeias das outras populações africanas. Estes achados estão em concordância com estimativas de idade de divergência das populações pigmeias (Quintana-Murci, et al., 2007; Patin, et al., 2009; Verdu, et al., 2009).

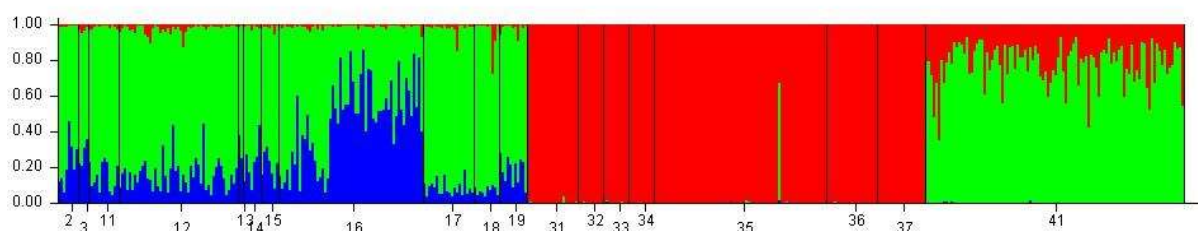


Figura I.16 Agrupamento das amostras de Kalunga (41), África (2-19) e Europa (31-37) de acordo com as frequências das 100 inserções *Alu*. Os grupos determinados pelo *Structure* são as cores azul, verde e vermelha.

Contextualizando as amostras brasileiras junto às suas populações doadoras, a figura I.17 mostra os quatro componentes genéticos que foram melhor estimados para o conjunto de amostras. Pode-se observar um componente africano exclusivo do grupo africano (em amarelo), como visto anteriormente; um componente africano ancestral, que está presente em todas as populações (ainda que em proporções mínimas, vermelho) e é o principal componente para a formação da população de Kalunga; um componente asiático que é mais frequente em Brasília do que em Kalunga (azul) e um componente europeu, que é o principal componente encontrado na população de Brasília.

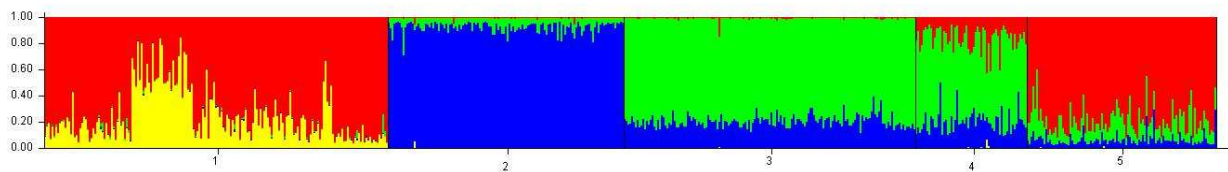


Figura I.17 Clusterização das populações miscigenadas brasileiras junto às suas populações doadoras estimada pelo painel de 100 inserções *Alu* para $K=4$ (melhor estimativa para o grupo amostral segundo o STRUCTURE). 1) africanos, 2) asiáticos, 3) europeus, 4) Brasília e 5) Kalunga.

I.3.3. Estimativa de composição genética das populações miscigenadas brasileiras

A partir das análises realizadas pelo software Structure, foi possível fazer a estimativa de composição genética das populações miscigenadas brasileiras baseada no painel de 100 inserções *Alu*. Tais estimativas foram feitas a partir da média da constituição genética dos indivíduos pertencentes a cada uma das populações. Dessa forma, foram obtidas estimativas congruentes com as inferidas em trabalhos anteriores (Pedrosa, 2006; Godinho, 2008), como pode ser observado na tabela I.5. As estimativas obtidas a partir do painel em Kalunga foram intermediárias às obtidas por marcadores autossômicos do tipo AIMs e STRs, enquanto que a estimativa de contribuição europeia para Brasília foi maior pelo painel do que a partir dos outros tipos de marcadores. O grande número de marcadores utilizados neste trabalho (100, se comparados aos 13 marcadores AIMs e 14 marcadores STRs) pode explicar essas diferenças, além de aumentar a confiança na robustez das estimativas.

Tabela I. 5 Estimativa de contribuição genética para as populações miscigenadas brasileiras Brasília e Kalunga através de marcadores genéticos autossômicos.

	Brasília			Kalunga		
	Africano	Europeu	Ameríndio	Africano	Europeu	Ameríndio
100	0,1900	0,7698	0,0040	0,7308	0,2519	0,0174
Alus	(0,1387)	(0,1653)	(0,1026)	(0,1663)	(0,1569)	(0,3465)
STRs	0,2620	0,6030	0,1360	0,8873	0,0037	0,0075
	(0,0050)	(0,0080)	(0,0070)	(0,0036)	(0,0036)	(0,0013)
AIMs	0,2290	0,6140	0,157	0,6105	0,2976	0,0918
	(0,0080)	(0,0080)	(0,0070)	(0,0147)	(0,0068)	(0,0016)

Considerando, então, as relações genéticas entre as populações dentro de cada grupo geográfico (Ásia, Europa, África) e as relações genéticas entre as populações miscigenadas brasileiras e os grupos de populações, pode-se observar que os grupos geográficos não são geneticamente homogêneos. Mesmo utilizando um painel de marcadores conservadores, foi possível identificar diferenças significativas entre as populações de um mesmo grupo geográfico, mostrando que não podemos considerar qualquer população como representativa da variabilidade/composição genética de um grupo geográfico. Isso implica um certo cuidado no momento de escolher as amostras parentais/doadoras para a estimativa de composição genética de populações miscigenadas.

Um bom exemplo disso são os diversas estimativas de composição genética geradas para cada indivíduo dependendo de como consideramos o grupo de amostras mundiais. A figura I.18 mostra o gráfico de agrupamento das amostras utilizadas de duas formas: em A, as populações parentais foram mantidas com suas populações originais, e o STRUCTURE é quem foi o responsável pela estimativa de agrupamento das amostras. Em B, as populações foram agrupadas por continente (África, Europa e Ásia).

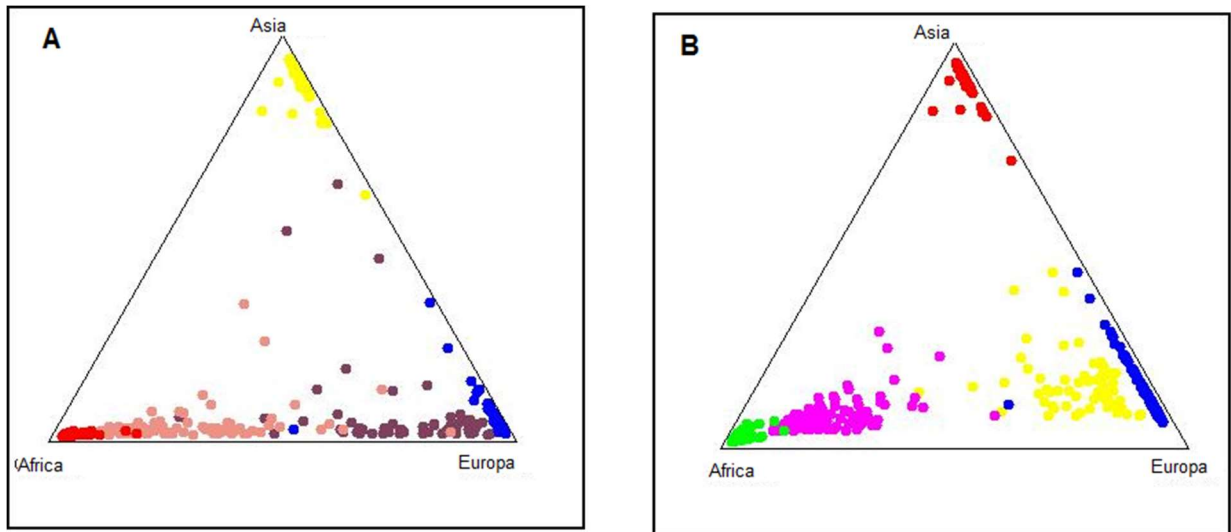


Figura I.18 Comparação entre duas abordagens de agrupamento para o mesmo conjunto de amostras. Cada ponto representa um indivíduo e os vértices representam os grupos parentais. Esquema de cores: A) vermelho – África, Azul – Europa; Amarelo – Ásia; Rosa – Kalunga; Roxo – Brasília. B) Verde – África; Azul – Europa; Vermelho – Ásia; Rosa – Kalunga; Amarelo – Brasília.

Assim como pode ser observado na figura I.16, a figura I.19 mostra que o grupo africano é composto de dois grupos de populações: um com grande componente genético azul, que possivelmente refere-se a um componente genético antigo e exclusivo da África subsaariana, e um com grande componente genético rosa. Este segundo, sim, acabou sendo dispersado pelos continentes e pode ser encontrado nas populações do Oriente Médio e do Brasil. Além disso, o painel de 100 inserções *Alu* foi capaz de identificar um componente genético bastante antigo e ubíquo, que pode ser encontrado (ainda que em baixa proporção) em absolutamente todas as amostras que aqui foram analisadas (Figura I.19, vermelho).

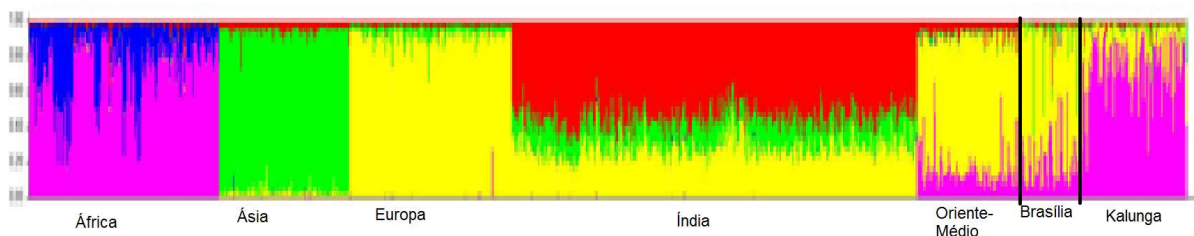


Figura I.19. Identificação de grupos populacionais baseado em perfis genéticos estimados pelo painel de 100 inserções *Alu* para K=5 (número de grupos estimado pelo STRUCTURE).

Em um estudo mais detalhado das relações genéticas entre populações mundiais, Montinaro e colaboradores (2015) utilizaram-se de 1414 indivíduos divididos em 42 populações mundiais para estimar a composição genética e a diferenciação entre grupos. Os autores observaram uma subestruturação ainda maior do que a encontrada aqui para os grupos populacionais, tendo as populações africanas sido divididas em nada menos do que 33 grupos com perfis genéticos diferentes. As populações europeias foram divididas regionalmente (Norte, Sudeste, Sudoeste e Nordeste da Europa – 10 grupos), tendo indivíduos do País Basco e da Sardenha formado dois grupos exclusivos e não pertencentes aos grupos regionais.

Portanto, considerando as estimativas de diferenciação populacional dentro dos grupos geográficos aqui apresentadas e os grande número de grupos encontrados dentro de cada região por Montinaro e colaboradores (2015), pode-se afirmar que os grupos geográficos não são geneticamente homogêneos, como será detalhado a seguir. Por esse motivo, a escolha da população doadora para a estimativa da composição genética de uma população miscigenada ou mesmo a escolha de populações-controle em estudos de genética médica e clínica deve ser feita com cautela e de forma consciente.

1.3.4. Caracterização do painel de 100 inserções Alu nas populações mundiais

1.3.4.1. África

Para testar se a diferenciação populacional entre as subpopulações africanas era significativa, conforme hipótese que incitou a análise segmentada dos grandes continentes, utilizaram-se os mesmos parâmetros que foram aplicados a Brasília e Kalunga e ao conjunto total de populações. Observou-se, então, que a maior diferenciação genética entre os subgrupos africanos foi entre as tribos pigmeias do Congo (ex-Zaire) e a população nigeriana ($F_{ST}=0,11488$, p -valor = 0,00000, figura I.20), considerada uma diferenciação populacional alta, de acordo com o critério

estabelecido por (Hartl & Clark, 2010). A maior parte dos pares de populações do grupo Africano, no entanto, apresentam diferenciação populacional moderada, tornando este grupo o mais estruturado dos grupos populacionais aqui estudados, como está descrito nas próximas seções.

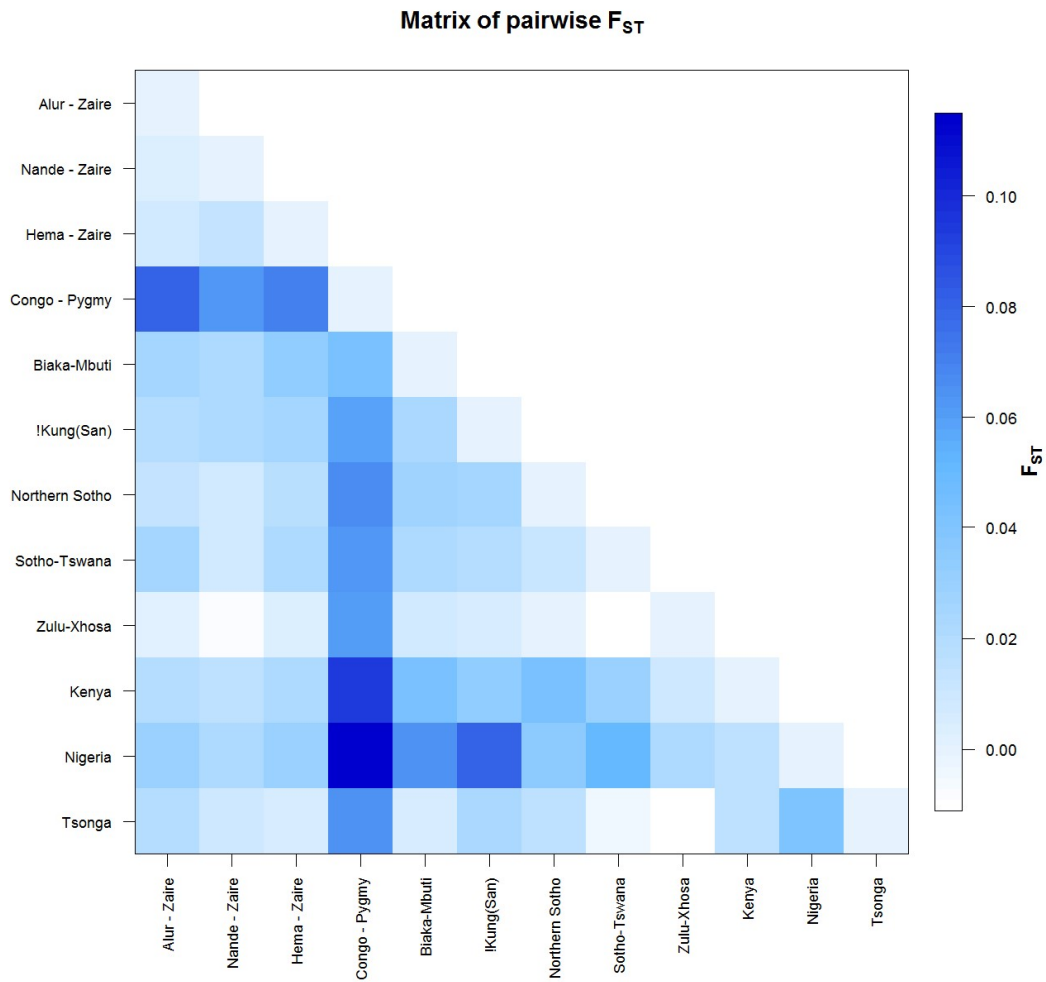


Figura I.10 Índice de diferenciação populacional (F_{ST}) entre os pares de populações subsaarianas.

Por outro lado, observou-se uma grande quantidade de subpopulações africanas que não apresentaram diferenciação populacional significativa. Essas populações, portanto, podem ser consideradas equivalentes com relação ao perfil genético. A tabela I.6 configura a lista de populações africanas que não apresentaram diferenciação populacional dentro do grupo africano, incluindo-se os valores relativos à significância do teste de hipótese (p-valor).

Tabela I. 6 Pares de populações africanas que não apresentaram diferenciação populacional significativa, com os respectivos valores que corroboram o teste de hipótese de não-diferenciação.

População 1	População 2	F _{ST}	p-Valor*
Alur	Nande	0,00270	0,19820 (0,0379)
Tsonga	Biaka	0,00534	0,18018 (0,0383)
	Hema	0,00606	0,05405 (0,0242)
	Sotho-Tswana	-0,00474	0,34234 (0,0327)
Northern Sotho	Alur	0,01344	0,08108 (0,0252)
	Hema	0,01793	0,07207 (0,0121)
	Nande	0,00731	0,07207 (0,0378)
	Sotho-Tswana	0,01182	0,08180 (0,0286)
Zulu-Xhosa	Alur	0,00164	0,34234 (0,0445)
	Biaka	0,00799	0,13514 (0,0244)
	Nande	-0,00838	0,93694 (0,0244)
	Northern Sotho	-0,00012	0,51351 (0,0651)
	San	0,00629	0,22523 (0,0389)
	Sotho-Tswana	-0,01117	0,90090 (0,0194)
	Kenya	0,00953	0,06306 (0,0237)
	Tsonga	-0,01025	0,65766 (0,0490)

*indica diferenciação populacional somente quando menor ou igual a 0,05.

Observando a tabela acima e a figura I.20 é possível verificar que existe uma subpopulação que pode ser considerada representativa das populações subsaarianas. Por não apresentar diferenciação populacional com oito das 13 populações subsaarianas e pequena diferenciação populacional com as outras populações (maior F_{ST}=0.06, tribos pigmeias do Congo), sugere-se que Zulu-Xhosa possa ser considerada uma referência genética das populações africanas subsaarianas.

Por outro lado, observa-se que as populações pigmeias do Congo (enquanto um único grupo populacional) apresentou diferença populacional com relação a absolutamente todas as outras populações africanas. Inclusive, o p-valor obtido para cada um dos testes de F_{ST} que envolviam a subpopulação pigmeus do Congo (Congo-Pygmy) foi igual a zero. Não curiosamente, o número de diferenças intrapopulacionais dessa população foi menor que o número de diferenças interpopulacionais envolvendo qualquer par de populações que contivesse o grupo de pigmeus do Congo (figura I.21).

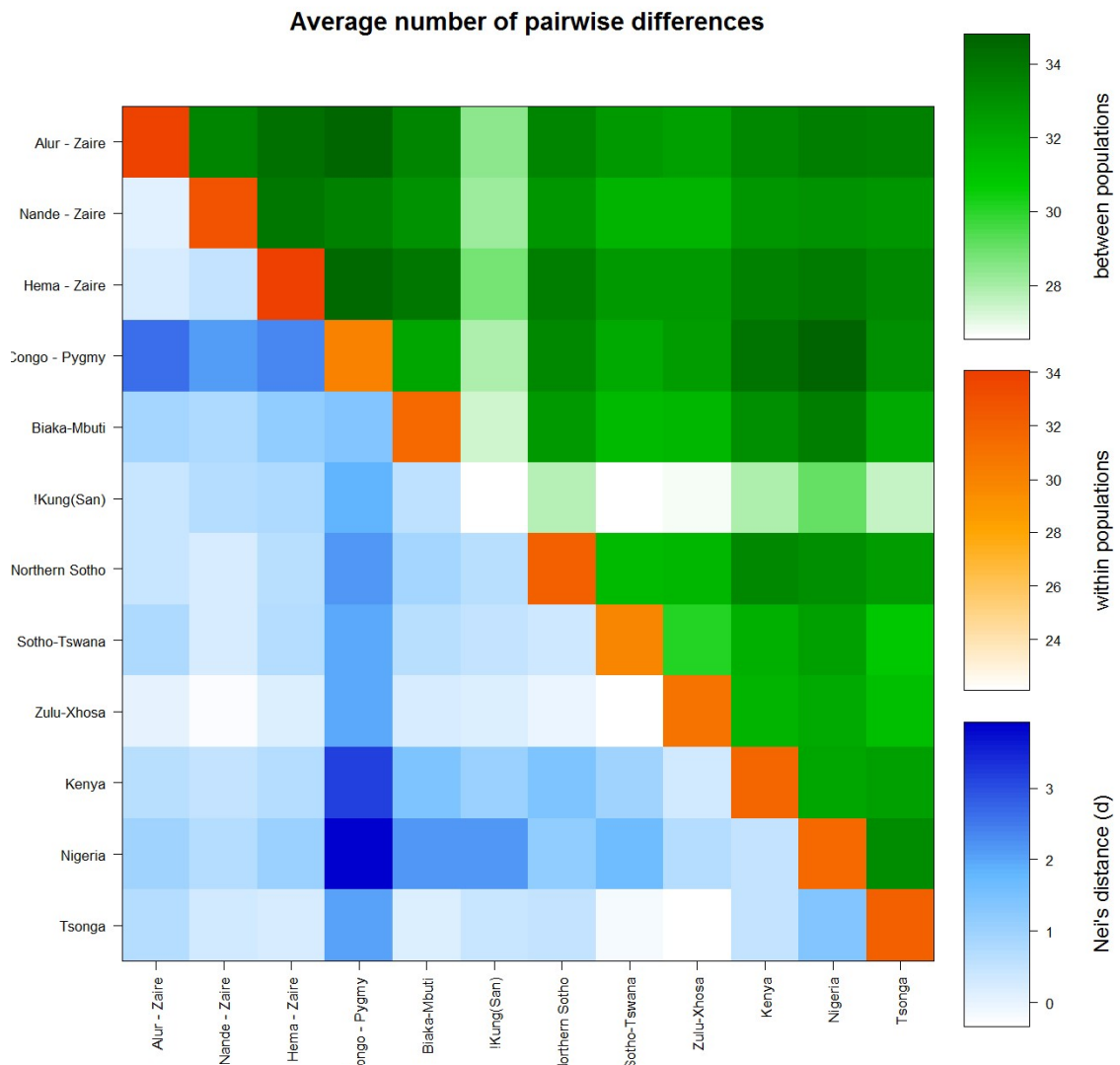


Figura I.11 Distância de Nei entre os pares de população (em azul, diagonal inferior), número médio de diferenças dentro de cada população (em laranja, diagonal) e o número médio de diferenças entre cada par de populações (em verde, diagonal superior).

Muito provavelmente, essas tribos pigmeias não se misturam e, muito menos, compartilham migrantes ou fluxo genético. Esse é o cenário mais congruente para explicar uma alta diferença intrapopulacional associada a uma baixa diferença interpopulacional, já que o comum no estudo da variação genética em populações humanas é encontrar mais diferenças dentro das populações do que entre as populações (Lewontin, 1972). Por outro lado, observa-se que os !Kung San apresentam baixas estimativas para os dois parâmetros (diferenças intrapopulacionais e diferenças interpopulacionais). Essa observação pode ser consequência de que os dados colhidos para a população !Kung San estão incompletos, sendo que apenas 24 dos 100 *Alus* foram genotipados para esses indivíduos.

Ao longo desse trabalho, as populações pigmeias foram agrupadas e denominadas populações pigmeias do Congo, como descrito em material e métodos. A análise de componentes principais para o grupo africano mostrou que as populações pigmeias do Congo se agrupam todas em um mesmo ponto do gráfico (Figura 1.22), o que corrobora a decisão de agrupamento dessas populações. Observando a análise em questão, observam-se de três a quatro grupos de populações, quando considerados o primeiro e o segundo componentes principais de variação genética dentro do grupo africano. O primeiro componente (que captura a maior variação) separa as populações pigmeias de todas as outras populações africanas, enquanto que o segundo componente separa os grupos Nigéria-Quênia, Sotho-Xhosa e todas as outras populações africanas (intermediárias).

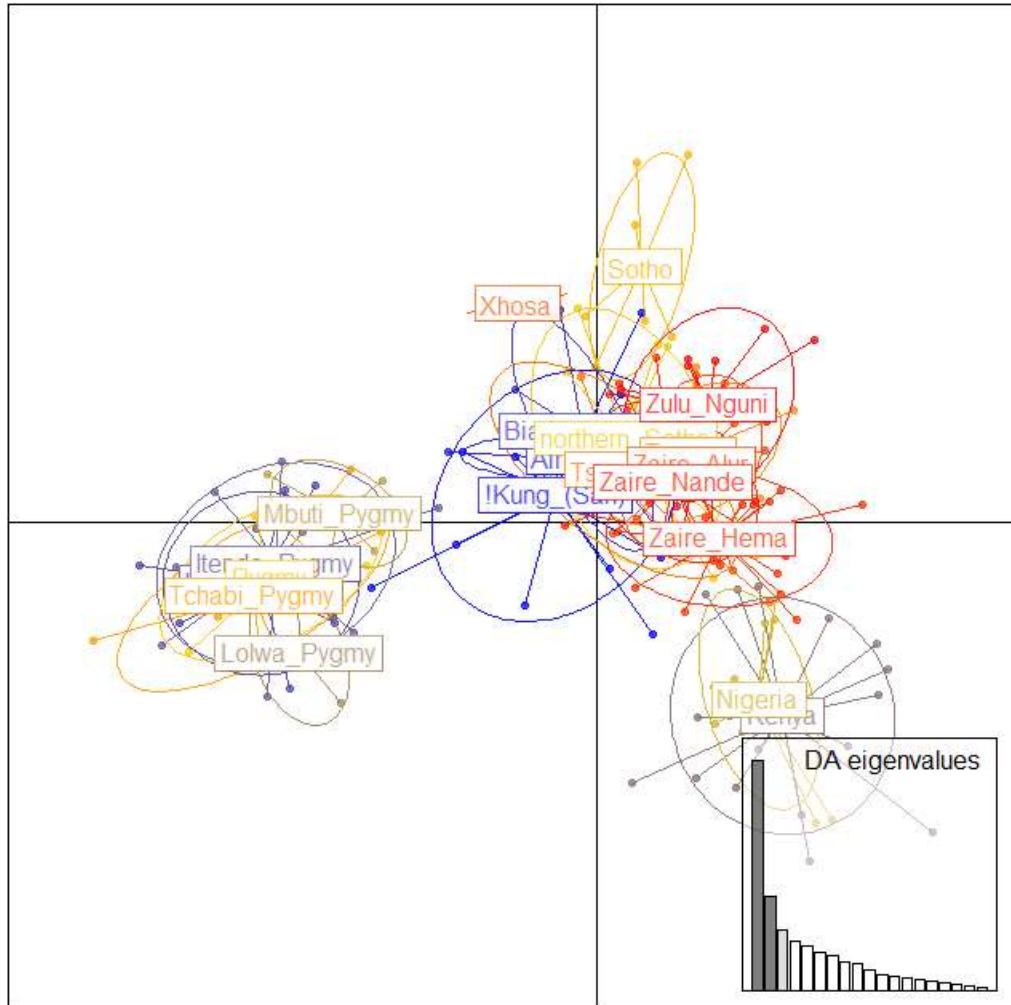


Figura I.12 Representação gráfica da análise de componentes principais da variação genética dentro do grupo africano. O eixo X representa o componente principal 1 (que captura a maior variação da amostra) e o eixo Y representa o componente principal 2 (que captura a segunda maior variação dentro da amostra).

1.3.4.2. Ásia

O grupo asiático é composto por oito populações, todas pertencentes à região sudeste da Ásia: Japão, Combódia, Vietnã, Sudeste da Ásia, Malásia, China, Taiwan e indivíduos sem origem definida. Com relação à diferenciação populacional dentro do grupo asiático, foi possível perceber que a distribuição de frequências alélicas neste grupo foi mais homogênea do que o que se esperaria (Figura I.23). A revisão feita por Xu (2012) evidenciou que mesmo a etnia Han (chineses) apresenta uma grande subestruturação populacional, estando dividida por grupos linguísticos (altaicos vs Sino-tibetanos/Tai-kadai). Essa subdivisão, no entanto, não é corroborada

pelas estimativas geradas através do estudo das 100 inserções *Alu*, detalhado a seguir.

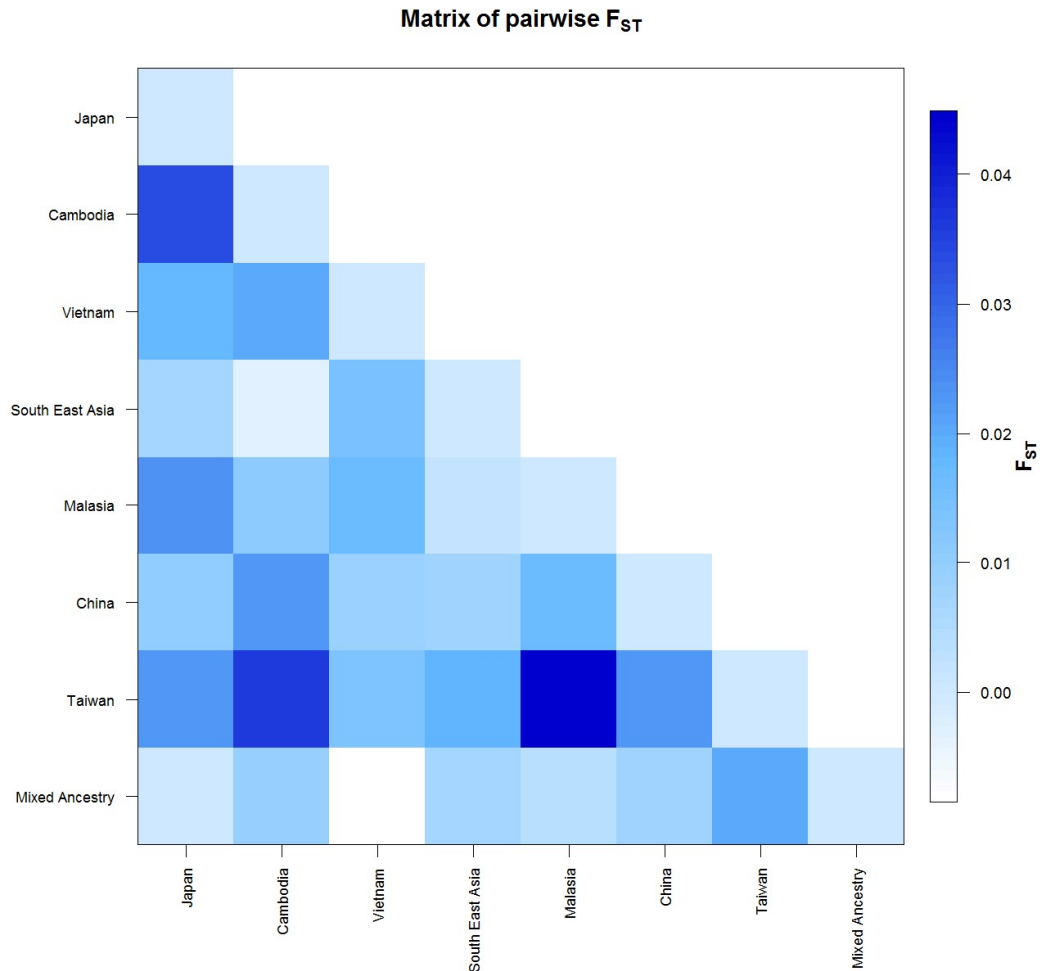


Figura I.13 Índice de diferenciação populacional baseado em F_{ST} entre pares de populações asiáticas, estimado através do painel de 100 inserções *Alu*.

No entanto, as estimativas evidenciadas pelo uso dos 100 marcadores do tipo inserções *Alu*, que são altamente conservadoras e com pequena taxa de mutação, mostraram que a maior diferenciação populacional encontrada dentro do grupo étnico asiático é pequena (0,045) e a maioria dos pares de populações sequer apresentaram diferenciação populacional significativa (tabela I.7).

Tabela I. 7 Pares de populações asiáticas que não apresentaram diferenciação populacional significativa

População 1	População 2	F _{ST}	p-valor
South East Asia	Japão	0,00731	0,17117 (0,0286)
	Cambódia	-0,00340	0,45946 (0,0550)
	Vietnã	0,01445	0,06306 (0,0237)
	Malásia	0,00188	0,39640 (0,0707)
	China	0,00745	0,18018 (0,0429)
	Ancestralidade mista	0,00721	0,18919 (0,0344)
Ancestralidade mista	Japão	0,00005	0,37838 (0,0354)
	Cambódia	0,00927	0,09910 (0,0252)
	Vietnã	-0,00852	0,76577 (0,0366)
	Malásia	0,00377	0,19820 (0,0227)
	China	0,00755	0,09910 (0,0316)
China	Vietnã	0,00885	0,10811 (0,0264)
	Malásia	0,01727	0,13514 (0,0279)
Malásia	Cambódia	0,01126	0,10811 (0,0264)
	Vietnã	0,01715	0,08108 (0,0316)

Pela tabela acima, pode-se observar que a única população que não aparece nem uma vez é a população de Taiwan. Essa população não somente se diferencia geneticamente de todas as populações do sudeste asiático, como apresenta os maiores F_{ST}s estimados para este grupo, considerando-se os pares de populações significativamente diferenciadas.

A análise de componentes principais revela uma estrutura populacional bastante interessante: seis grupos de populações são aparentes (figura I.24). O primeiro componente principal separa claramente dois grupos de populações: um grupo contendo amostras do Sudeste da Ásia, Taiwan e Cambódia; e outro grupo

contendo todas as outras populações. O segundo componente principal diferencia as populações sul-asiáticas, e separa o grupo misto em japoneses e outras populações.

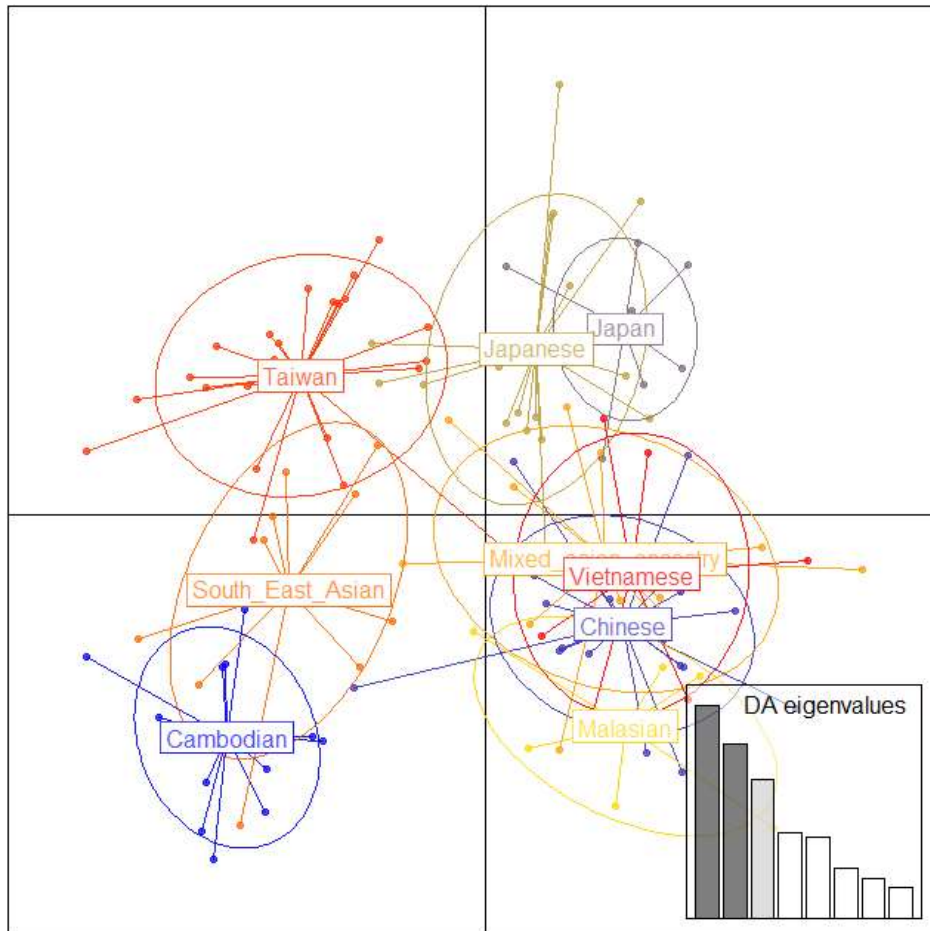


Figura I.24 Análise de componentes principais para o grupo de populações asiáticas. O eixo X contém a variação representada pelo primeiro componente principal, PC1, enquanto o eixo Y contém a variação representada pelo segundo componente principal, PC2

1.3.4.3. Europa

O grupo europeu é composto por sete populações, sendo uma delas localizada geograficamente na América do Norte (Utah, EUA): França, Ibéria, País Basco, Polônia (Europe), Finlândia, Utah e Rússia. A diferenciação populacional do grupo europeu mostrou que boa parte dos pares de populações apresentaram diferenças significativas com relação às 100 inserções *Alu* aqui estudadas (figura

I.25). Contudo, algumas observações podem ser feitas sobre a dinâmica de diferenciação populacional das amostras europeias aqui apresentadas.

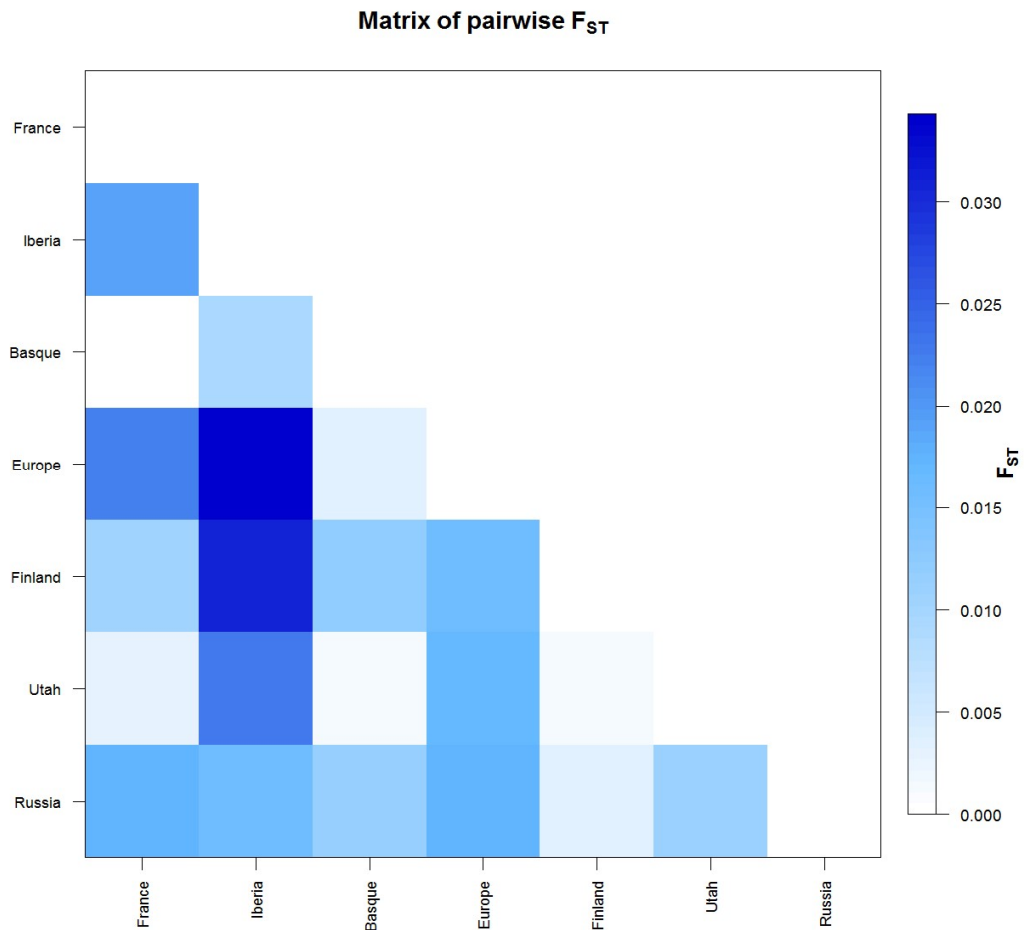


Figura I.14 Índice de diferenciação populacional entre as populações do grupo europeu, baseado em 100 inserções *Alu*.

Ainda que a diferenciação entre as populações europeias tenha sido, em sua maioria, significativa, pode-se notar que a maior diferenciação encontrada (entre Ibéria e Europa, que é composta principalmente de amostras polonesas) é apenas pequena (0,034). Além disso, vários outros pares de população não apresentaram diferenciação populacional significativa, como pode ser observado na tabela I.8.

Tabela I. 8 Índice de diferenciação populacional para os pares de populações europeias que não apresentaram diferenciação significativa.

População 1	População 2	F _{ST}	p-valor
País Basco	França	0,00006	0,39640 (0,0511)
	Ibéria	0,00911	0,13514 (0,0474)
	Europa	0,00366	0,40541 (0,0365)
	Utah	0,00150	0,27928 (0,0212)
Utah	França	0,00290	0,16216 (0,0402)
	Finlândia	0,00149	0,20721 (0,0333)
Rússia	Finlândia	0,00343	0,15315 (0,0360)

Estudos genéticos com marcadores de DNAm e genoma total mostraram que o País Basco é uma região semi-isolada que apresenta ancestralidade direta com antigos fazendeiros do Neolítico/Calcolítico (Behar et al., 2012; Günther et al., 2015). Esse semi-isolamento encontrado por marcadores mais dinâmicos diferencia os bascos do resto da Europa e corrobora a diferenciação linguística, já que se acredita que a língua basca é fruto das primeiras línguas dos fazendeiros do neolítico, e não apresenta compartilhamento léxico com as línguas atuais da Europa (Leonardi, 2005; Behar et al., 2012). No entanto, os dados encontrados para as 100 inserções *Alu* mostram um cenário completamente diferente: o País Basco apresenta grande similaridade genética com a maior parte das populações europeias aqui estudadas (4 de 6 populações).

Isso pode ser explicado pela baixa atividade de transposição de novas inserções e pelo fato das inserções *Alu* serem marcadores genéticos conservadores, especialmente importantes para detectar mistura genética antiga (Shedlock & Okada, 2000). As famílias de inserção *Alu* mais recentes no genoma humano (Yb) surgiram entre 1,11 e 4,81 milhões de anos atrás (Carter, et al., 2004), bem anteriores à expansão dos fazendeiros do neolítico, há 7 mil anos atrás (Behar et al., 2012). É possível que a perda de diversidade genética durante as migrações do humano moderno tenha sido mais contundente para a composição genética das populações

europeias do que as futuras subdivisões e isolamentos decorrentes da formação das fronteiras políticas e culturais, quando consideramos marcadores genéticos tão conservadores como as inserções *Alu*.

A análise de componente principal apresenta o cenário da composição genética europeia dividido em quatro grupos (Figura 1.26). O primeiro componente principal separa os países mais a oeste da Europa, cujas línguas são latinas (à exceção do País Basco), de países cujas línguas são urálicas (Finlândia e Rússia). No meio do caminho, o grupo de línguas do norte europeu e anglo-saxã. Um cenário similar foi encontrado por Nelis e colaboradores (2009) em um estudo com 270.000 SNPs autossômicos: nos vértices a Finlândia, as populações bálticas (junto com Polônia e Rússia), e Itália, e as populações da Europa Central e Oeste apresentando frequências intermediárias entre esses três vértices.

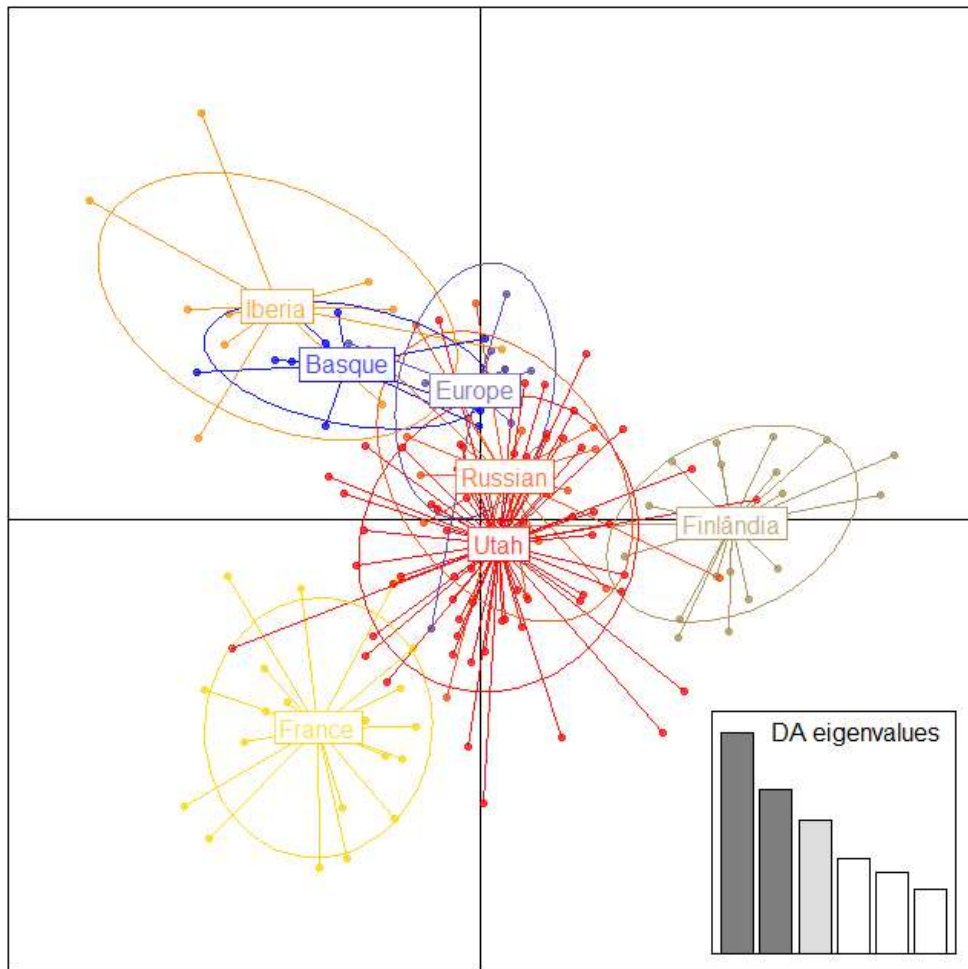


Figura I.15 Análise de componente principal para o grupo de populações europeias. O eixo X representa o componente principal 1 (captura a maior variação) e o eixo Y representa o componente principal 2 (captura a segunda maior variação).

1.3.4.4. Índia

Como dito anteriormente, o grupo indiano é o mais complexo de todos os grupos aqui considerados, e também o mais farto em número de populações (tabela I.2. Resultado de anos de miscigenação oriunda tanto de colonização quanto por interesses mercantis (não podemos esquecer que as grandes navegações tinham por principal objetivo “encontrar as Índias”), a formação genética do grupo indiano apresenta singularidades que são reflexo da alta miscigenação populacional. O primeiro ponto a se notar é que a maior parte da diferenciação entre os pares de populações é considerada moderada (de 0,03 a 0,06), refletindo subestruturação

populacional (figura 1.27), se comparado a outros grupos populacionais aqui estudados.

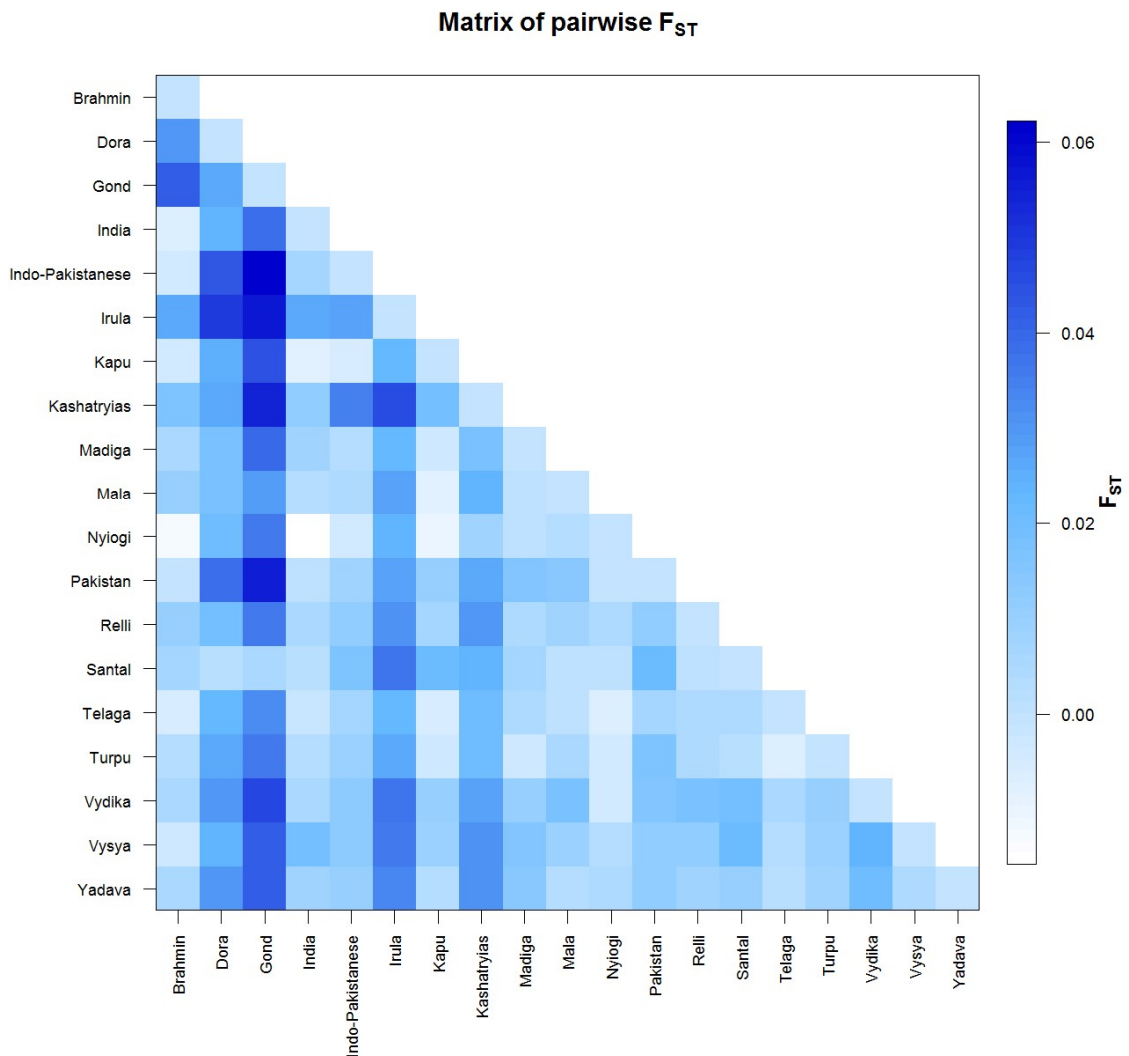


Figura I.16 Índice de diferenciação populacional para as subpopulações do grupo indiano, estimado a partir de 100 inserções *Alu*.

Outro ponto importante de se levantar é que a Índia é um país só. Ainda que seja um país gigante, o 7º maior do mundo com 3.287.263 km², (United Nations Statistics Division, 2014), seu tamanho, apenas, não é capaz de explicar a diferenciação populacional encontrada: pode-se observar que a diferenciação populacional dos indivíduos do Paquistão (outro país) com populações indianas é menor do que a diferenciação populacional entre algumas castas e tribos a

populações indianas. Por tanto, o sistema de organização social em castas é um importante fator para a formação desse cenário (Bamshad, et al., 2001). Contudo, a subestruturação populacional observada na Índia alcança índices menores do que os observados em algumas populações africanas subsaarianas (figura I.20). Isso mostra que, ainda que o sistema de castas seja um fator que cause estruturação populacional, o grupo indiano ainda é mais geneticamente coeso que o grupo africano.

Mesmo apresentando subestruturação populacional, pode-se observar que alguns pares de população não apresentam diferenciação populacional significativa (tabela I.9). Algumas populações apresentam um grande número de pares não significativos, como é o caso de Nyiogi (15 de 19), Índia (13 de 19), Brahmin (12 de 19), que não apresentam diferenciação populacional a mais de 50% das populações aqui estudadas. Além disso, cerca de metade dos possíveis pares de populações não apresentaram diferenciação populacional significativa, o que mostra que o grupo indiano, ainda que o mais complexo em termos de formação e organização social, apresenta grande coesão genética.

Tabela I. 9 Índice de diferenciação populacional para os pares de populações indianas que não apresentam diferenciação significativa. As populações 1 são apenas aquelas que não apresentaram diferenciação genética a mais de oito populações.

População 1	População 2	F _{ST}	p-valor
Nyiogi	Brahmin	-0,01321	0,99099 (0,0030)
	Índia	-0,01575	0,97297 (0,0184)
	Indo-Paquistaneses	-0,00471	0,68468 (0,0338)
	Kapu	-0,00969	0,79279 (0,0354)
	Kashatryias	0,00753	0,09910 (0,0286)
	Madiga	0,00057	0,36937 (0,0533)
	Mala	0,00342	0,13514 (0,0339)
	Paquistão	-0,00017	0,45045 (0,0359)
	Relli	0,00420	0,16216 (0,0264)
	Santal	0,00052	0,30631 (0,0338)
	Telaga	-0,00718	0,95495 (0,0203)

População 1	População 2	F _{ST}	p-valor
	Turpu	-0,00401	0,71171 (0,0395)
	Vydika	-0,00359	0,77477 (0,0279)
	Vysya	0,00329	0,19820 (0,0445)
	Yadava	0,00443	0,11712 (0,0273)
Índia	Brahmin	-0,00604	0,32432 (0,0624)
	Indo-Paquistaneses	0,00652	0,73874 (0,0446)
	Kapu	-0,00767	0,10811 (0,0353)
	Madiga	0,00761	0,10811 (0,0378)
	Mala	0,00350	0,40541 (0,0563)
	Paquistão	0,00121	0,45045 (0,0407)
	Relli	0,00565	0,24324 (0,0333)
	Santal	0,00181	0,24324 (0,0470)
	Telaga	-0,00160	0,69369 (0,0274)
	Turpu	0,0034	0,26126 (0,0252)
	Vydika	0,0052	0,32432 (0,036)
	Yadava	0,00773	0,09009 (0,0235)
Brahmin	Indo-Paquistaneses	-0,00397	0,75676 (0,0338)
	Kapu	-0,00472	0,71171 (0,0253)
	Madiga	0,00549	0,12613 (0,0201)
	Paquistão	-0,00083	0,60360 (0,0490)
	Santal	0,00638	0,10811 (0,0227)
	Telaga	-0,00488	0,90991 (0,0345)
	Turpu	0,00358	0,32432 (0,0433)
	Vydika	0,00594	0,11712 (0,0305)
	Vysya	-0,00347	0,67568 (0,0475)
	Yadava	0,00554	0,11712 (0,0194)
Indo-Paquistaneses	Kapu	-0,00544	0,71171 (0,0497)

População 1	População 2	F _{ST}	p-valor
	Madiga	0,00313	0,21622 (0,0388)
	Mala	0,00424	0,32432 (0,0492)
	Paquistão	0,00830	0,14414 (0,0309)
	Relli	0,01222	0,05405 (0,0242)
	Telaga	0,00715	0,15315 (0,0273)
	Turpu	0,00898	0,09910 (0,0252)
	Vydika	0,01299	0,06306 (0,0194)
	Vysya	0,01313	0,13514 (0,0203)
Kapu	Madiga	-0,00241	0,55856 (0,0327)
	Mala	-0,00813	0,89189 (0,0165)
	Relli	0,00725	0,14414 (0,0278)
	Telaga	-0,00596	0,77477 (0,0364)
	Turpu	-0,00293	0,62162 (0,0370)
	Vydika	0,01050	0,12613 (0,024)
	Vysya	0,00958	0,08108 (0,0091)
	Yadava	0,00356	0,19820 (0,0485)
Telaga	Madiga	0,00378	0,08108 (0,0316)
	Mala	0,00124	0,34234 (0,0504)
	Relli	0,00490	0,12613 (0,0242)
	Santal	0,00459	0,09009 (0,0303)
	Turpu	-0,00612	0,93694 (0,0244)
	Vydika	0,00533	0,06306 (0,0273)
	Vysya	0,00337	0,23423 (0,0309)
	Yadava	0,00249	0,11712 (0,0360)

A análise de discriminantes realizada mostrou como a diferenciação populacional e a variabilidade genética do grupo indiano se relacionam (figura I.28).

Analisando o primeiro e o segundo componentes principais de variabilidade genética, pode-se observar que são formados três grupos: um grupo contendo apenas a população Irula, um grupo contendo Gond, Dora e Santal; e um grupo contendo todas as outras 15 populações. Isso mostra que os primeiros dois componentes, que explicam a maior parte da variabilidade genética encontrada no grupo indiano, define os grupos tribais, mas coloca todas as outras populações em um único grupo. Essa dinâmica é característica de populações isoladas (as tribos indianas) inseridas em um contexto de populações altamente miscigenadas e homogêneas (todas as outras 15 populações). Santal, ocupa uma posição mais próxima dos grupos miscigenados, possivelmente devido à grande quantidade de genótipos faltantes para alguns indivíduos. Na estimativa de variabilidade, os genótipos faltantes tendem a ser estimados a partir dos mais frequentes para cada *locus*, e por isso Santal aparenta se agrupar com as populações mais numerosas.

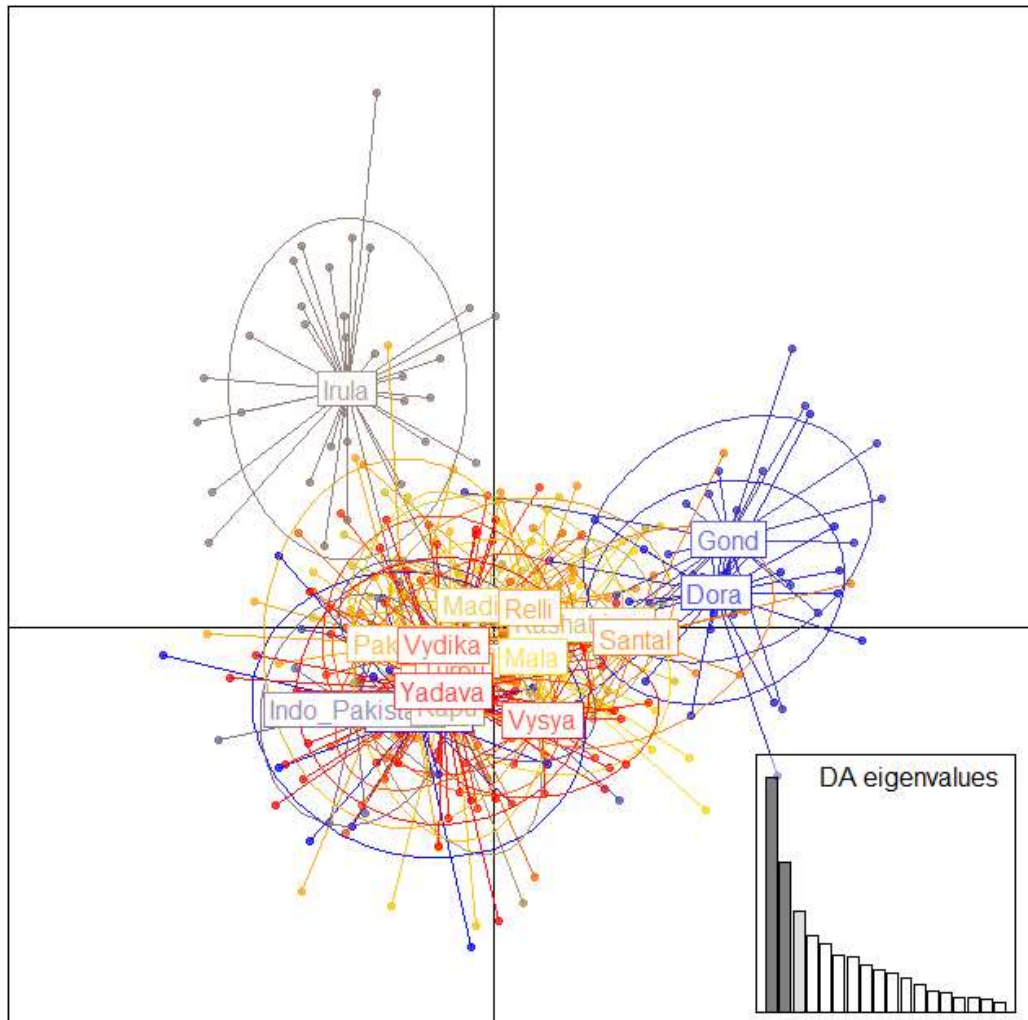


Figura I.17 Análise de discriminantes mostrando os primeiros dois componentes principais de variação dentro do grupo indiano. O eixo X representa o componente principal 1 (maior variação) e o eixo Y representa o componente principal 2 (segunda maior variação).

1.3.4.5. Oriente Médio

O grupo do Oriente Médio é composto por cinco populações: Jordânia, Bahrein, Emirados Árabes Unidos, Síria e Egito. A diferenciação populacional dentro do grupo do Oriente-Médio é a menor dos grupos aqui estudados (figura I.29). O par de populações com maior diferenciação populacional (Bahrein e Síria), apresentou apenas uma leve diferenciação populacional, ainda que significativa (0,02407) e a maioria dos pares testados não apresentaram diferenciação populacional significativa (tabela I.10). Isso mostra uma grande unidade genética das populações que compõem o grupo. São todos países predominantemente muçulmanos, de origem Árabe e que

ocupam uma região bastante específica e de passagem, estrategicamente localizado entre África, Europa e Ásia, o que pode contribuir para o fluxo gênico e a não diferenciação das populações.

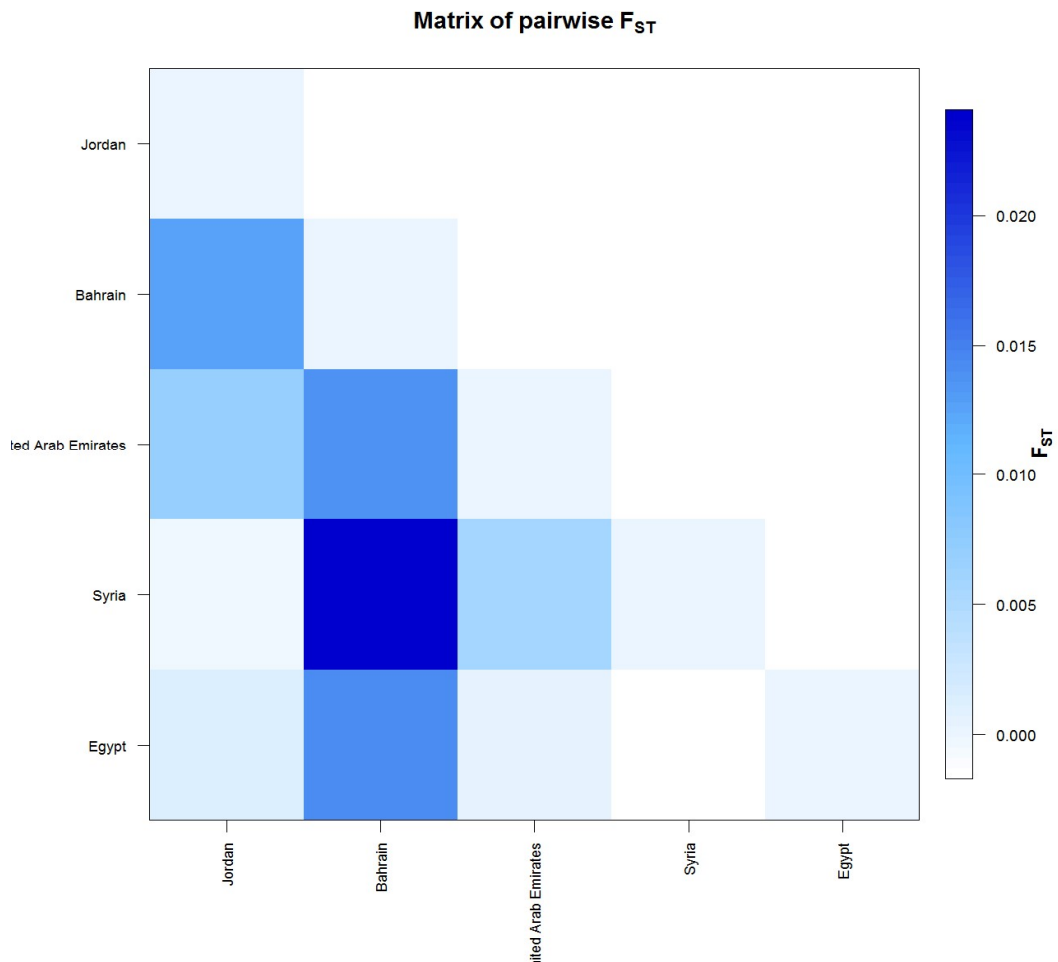


Figura I.18 Índice de diferenciação populacional para os pares de populações do grupo do Oriente-Médio, estimado por 100 inserções *Alu*.

Bahrain foi a única população que apresentou diferenciação populacional significativa a todas as outras populações da amostra, como pode ser observado na tabela I.10. Todos os outros pares testados (todos os pares de populações que não envolviam a população Bahrain) não apresentaram índice de diferenciação populacional significativo, indicando que todas as populações do grupo do Oriente Médio aqui estudadas são geneticamente iguais entre si, à exceção da população do Bahrain, cuja diferenciação populacional ao grupo é considerada leve.

Tabela I. 10 Índice de diferenciação populacional e significância do teste (p-Valor, desvio-padrão entre parênteses) para todos os pares de populações do grupo Oriente-Médio. Os valores que indicam diferenciação estão marcados com um asterisco (*).

População 1	População 2	FST	p-Valor
Bahrain	Jordânia	0,01273	0,02703 (0,0194)*
	Emirados Árabes Unidos	0,01382	0,01802 (0,0121)*
	Síria	0,02407	0,00000 (0,0000)*
	Egito	0,01426	0,00901 (0,0091)*
Jordânia	Emirados Árabes Unidos	0,00696	0,18919 (0,0316)
	Síria	-0,00019	0,58559 (0,0434)
	Egito	0,00127	0,42342 (0,0490)
Emirados Árabes Unidos	Síria	0,00561	0,30631 (0,0278)
	Egito	0,00068	0,53153 (0,0497)
Síria	Egito	-0,00172	0,63969 (0,0360)

A população de Bahrain é a mais diferente das populações do grupo do Oriente-Médio (figura I.29), inclusive registrando os maiores índices de diversidade gênica e heterozigose observada (tabela I.10). A análise de componente principal mostra dois grupos bem definidos: o primeiro grupo é composto apenas pela população do Bahrain, e o segundo grupo é composto por um aglomerado que se

diferencia em subgrupos a partir do segundo componente principal (figura I.30), que separa as populações por origem geográfica, distanciando-se da África.

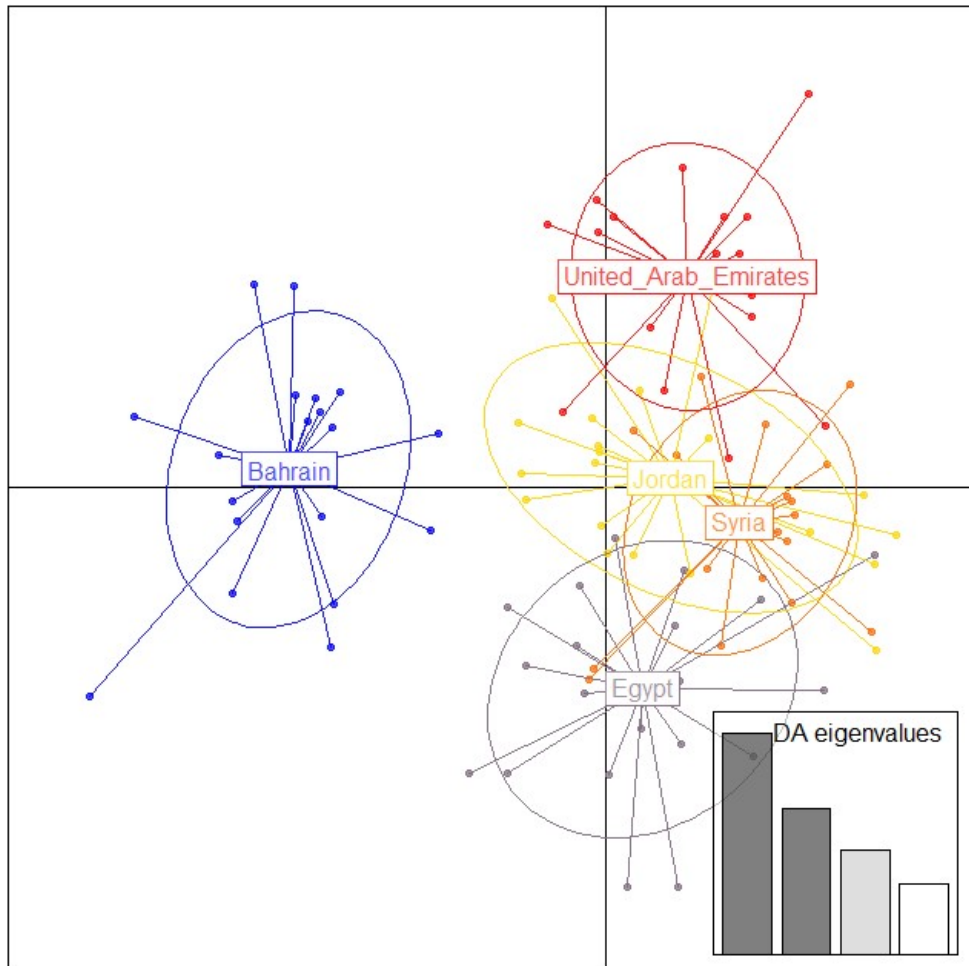


Figura I.19 Análise discriminante de componente principal para o grupo de populações do Oriente Médio, baseado em 100 inserções *Alu*. O eixo X representa a variação explicada pelo primeiro componente principal (maior variação) enquanto o eixo Y representa a variação explicada pelo segundo componente principal (segunda maior variação).

I.3.5. Teste de Neutralidade

A distribuição geográfica das amostras utilizadas neste trabalho permitiu delinear uma estruturação populacional baseada em continente/grupos étnicos *a priori*. Dessa forma, foi possível realizar testes de seleção natural para o conjunto total de marcadores utilizados. A partir do índice de diferenciação populacional estimado entre os grupos, foi possível verificar que oito marcadores *Alu* mostram valores de heterozigose sugestivos de estar sob seleção (figura I.31). Os marcadores alu35 e

alu36 apresentam alta heterozigose (entre 0,4 a 0,5), que pode ser indicativo de seleção estabilizadora, enquanto que o marcador alu21 apresenta heterozigose igual a 0, o que indica que esse locus deve sofrer seleção purificadora, uma vez que heterozigotos são inexistentes ou desfavorecidos. Os marcadores alu16, alu20, alu26, alu42 e alu50 apresentam baixa heterozigose, mas não tão extremo quanto no caso anterior e por isso podem estar sofrendo seleção balanceadora.

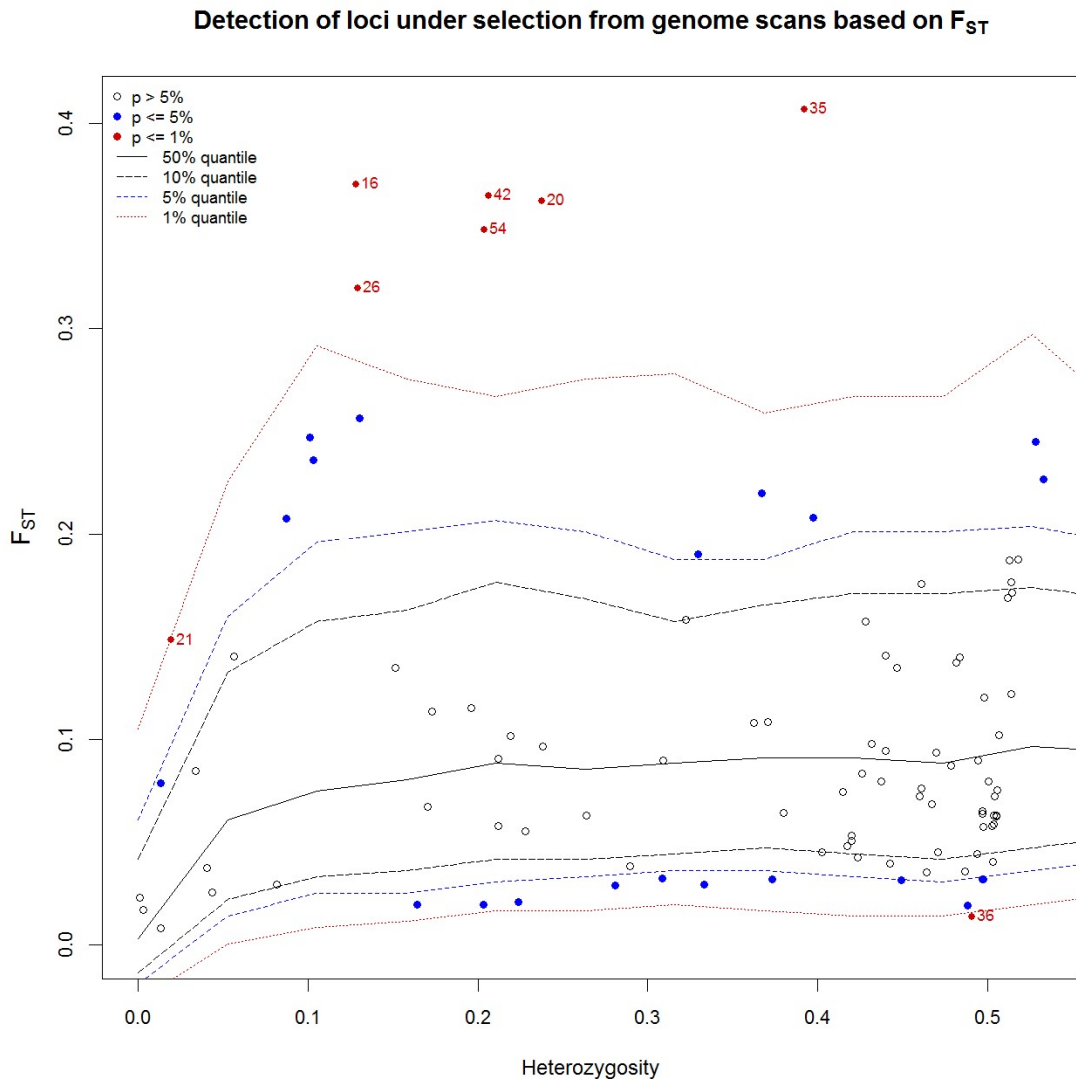


Figura 1.20 Inserções *Alu* que estão sob seleção para todos os grupos de populações analisadas. As linhas vermelhas indicam os limites de 1% de significância para a relação F_{ST} /heterozigose. Apenas marcadores que estejam sobre ou excedam os limites mínimo e máximo do intervalo são considerados sob seleção.

A inserção *Alu* de número 21 (Ya5NBC157) parece estar fixada em todas as populações não africanas, sugerindo um mecanismo de efeito fundador durante a

saída do humano moderno da África. Mesmo nas populações africanas, o estado ancestral, ausência de inserção, só é encontrado em cinco populações, todas localizadas no sul da África: Tsonga, Sotho-Tswana, Northern-Sotho, !Kung e Alur-Zaire. Já o estado ancestral da inserção Alu de número 26 (Ya5NBC212) é encontrado em populações da África Central e Sul como Zaire (Alur, Nande e Hema), todas as populações pigmeias, !Kung, Northern-Sotho, Sotho-Tswana e Zulu-Xhosa e em apenas duas populações não-africanas: Irula e Madiga, sugerindo gargalos de garrafa sucessivos durante a dispersão do homem moderno.

A inserção *Alu* de número 36 (Ya5NBC333), por sua vez, apresenta grande heterozigose e baixíssimo F_{ST} entre todas as populações aqui listadas, o que significa que as frequências para ausência e presença da inserção em diferentes locais do globo são essencialmente similares entre os grupos populacionais. Essa inserção está localizada numa grande região intergênica de 540Mb, entre os genes *DACT1* e *DAAM1*. Ambos os genes estão relacionados à gastrulação e formação do sistema nervoso em vertebrados (Sato, et al., 2006; Shi, et al., 2012) e, portanto, devem ser alvo de seleção. Sendo assim, o *Alu* Ya5NBC333 pode estar sofrendo um efeito “carona” em decorrência da sua proximidade aos genes *DACT1* e *DAAM1*.

Os comentários acima sobre duas das inserções *Alu* que apresentam sinais de seleção são pertinentes e suscitam boas hipóteses a serem testadas em trabalhos futuros como, por exemplo, se o mesmo resultado seria observado caso os genes *DACT1* e *DAAM1* fossem testados para seleção. Contudo, como apenas 8% dos marcadores apresentaram sinais de seleção, pode-se dizer que, essencialmente, o painel 100 inserções *Alu* é neutro.

1.4. Conclusões

1. Brasília e Kalunga não formam um grupo homogêneo de amostras e, por esse motivo, não configuram um grupo brasileiro ou sul-americano específico, fazendo com que essas populações pertençam aos conjuntos de amostras que

representem a maior contribuição genética para a composição dessas populações.

2. O desequilíbrio de ligação nas populações miscigenadas é maior e afeta mais *loci* do que nas populações não miscigenadas, sendo o desequilíbrio de ligação entre os marcadores do painel de 100 inserções *Alu* um indicativo de miscigenação da população.
3. À exceção do grupo populacional do Oriente-Médio, os grupos populacionais não podem ser considerados homogêneos e a maioria das populações que compõem cada um dos grupos não pode ser considerada referência genética para seu grupo de origem.
4. Determinadas populações de cada grupo, no entanto, podem ser consideradas representativas do seu grupo de origem por não apresentarem diferenciação populacional significativa: Zulu-Xhosa para o grupo africano; Sudeste Asiático para o grupo asiático; País Basco para o grupo europeu; Nyiogi para o grupo indiano; qualquer população, exceto Bahrein, para o grupo do Oriente-Médio.
5. Há uma grande similaridade genética entre a população brasileira e suas parentais africanas (Kalunga) e europeias (Brasília) para o painel de 100 inserções *Alu* autossômicos, corroborando estudos anteriores que utilizaram marcadores genéticos do tipo AIM, SNPs, do cromossomo Y e DNAm.
6. Há também grande similaridade genética entre a população de Brasília e o grupo de populações do Oriente-Médio, que, por sua vez, é geneticamente similar ao grupo de populações europeu, especialmente se consideradas as populações do centro e oeste europeus.
7. A análise de componentes principais agrupa Brasília a populações europeias e do Oriente-Médio e Kalunga às populações africanas, evidenciando que o painel de inserções *Alu* é capaz de agrupar as populações miscigenadas às suas populações doadoras, identificando as populações de maior contribuição genética.
8. As estimativas de composição genética feitas a partir do painel de 100 inserções *Alu* corroboram estimativas anteriores feitas com marcadores autossômicos.

9. Dentre os marcadores que constituem o painel de 100 inserções *Alu*, 8% mostraram sinais de estarem sob seleção nos diferentes grupos testados, sendo o painel essencialmente neutro.

Capítulo II – A variabilidade genética humana e sua influência na susceptibilidade à infecção severa pelo vírus da Influenza A

(Tradução do artigo: Role of the host genetic variability in the influenza A virus susceptibility, de Ana Carolina Arcanjo, Giovanni Mazzocco, Silviene Oliveira, Dariusz Plewczynski e Jan Radomski, publicado na Acta Biochimica Polonica em 2014; Anexo II.1)

II.1. Introdução

O vírus da gripe é uma das causas mais importantes de infecções que acometem o trato respiratório, com cerca de 3-5 milhões de infecções clínicas e 250.000-500.000 casos fatais a cada ano (Dawood, et al., 2012; Simonsen, et al., 2013; WHO, 2014). A resposta imune ao vírus da influenza tem sido alvo de pesquisa há mais de setenta anos (Andrewes, 1939) devido ao grande impacto que a doença tem no sistema de saúde, de uma forma global. Mas, ainda que alguns aspectos da resposta imunológica à infecção por influenza tenham sido esclarecidos nesses últimos 70 anos, ainda há muitas perguntas não-respondidas neste campo de pesquisa.

A influência da variabilidade genética do próprio vírus é, sem sombra de dúvidas, o tópico mais importante e que tem sido o maior alvo de pesquisas até hoje (Hatta, et al., 2001; Wagner, et al., 2002; Uiprasertkul, et al., 2005; Gambaryan, et al., 2006; Bateman, et al., 2008; Nicholls, et al., 2007; Das, et al., 2010; Jayaraman, et al., 2012; Sriwilaijaroen & Suzuki, 2012; Thakaramaran, et al., 2012; Guarnaccia, et al., 2013; Koel, et al., 2013; Sun, et al., 2013; Thakaramaran, et al., 2013; El Moussi, et al., 2014; Qi, et al., 2014; Su, et al., 2014). O alto grau de deriva antigênica observado no vírus da influenza é um dos principais motivos da recomendação para a constante atualização e vacinação contra a gripe sazonal. A variabilidade genômica do vírus da influenza tem sido largamente estudada ao longo dos anos, mas a variabilidade genômica do hospedeiro se mantém parcamente documentada, ainda que tenha um papel fundamental na resposta imune à infecção por influenza (Wijburg, et al., 1997; Matrosovich & Klenk, 2003; Schmitz, et al., 2005; Jayasekera, et al., 2007; Koyama, et al., 2007; Throsby, et al., 2008; Ichinohe, et al., 2009; Sabbah, et al., 2009; Kreijtz, et al., 2011; Zhou, et al., 2012; Henn, et al., 2013; Hertz, et al., 2013; Lin & Brass, 2013) e o desenvolvimento da infecção. O objetivo deste capítulo é, portanto, explorar a variabilidade genômica de genes do hospedeiro (o homem, no caso) que mediam a interação patógeno-hospedeiro. Para tal, desenvolveu-se uma abordagem genômica, focada em descrever a variabilidade genética no âmbito global das populações humanas existentes nos dados disponibilizados pelo projeto 1000 Genomes (Clarke,

et al., 2012; (The 1000 Genomes Project Consortium, 2015a), com o intuito de quantificar a variabilidade existente e o nível de conservação dos genes envolvidos nessa interação. Também tentou-se prever se as variantes encontradas nas grandes populações afetam a especificidade da ligação entre as proteínas virais hemaglutinina (HA) e neuraminidase (NA) e a respectiva resposta imune a essas proteínas.

Genes relacionados à entrada do vírus na célula do hospedeiro são, principalmente, responsáveis pela produção e montagem dos ácidos siálicos Neu5Ac- α 2,6-Gal β 1,4-GlcNAc (humano) e Neu5Ac- α 2,3-Gal β 1,4-GlcNAc (aviário) na membrana celular eucariótica. Tanto os receptores humanos quanto os aviários se ligam às proteínas NA e HA do vírus da influenza A durante o processo de infecção. A composição de ácidos siálicos na membrana celular do hospedeiro e as modificações nos sítios de ligação à NA e à HA influenciam o sucesso da infecção por Influenza A.

II.2. A ligação do vírus à célula

O vírus da influenza humano se liga à célula do hospedeiro por meio da interação com o ácido siálico Neu5Ac- α 2,6-Gal β 1,4-GlcNAc (Sia2,6, daqui para a frente), uma modificação de glicanos muito abundante em humanos, especialmente nas células epiteliais do trato respiratório superior (Skehel & Wiley, 2000). Além disso, o muco que protege as vias aéreas superiores também são igualmente ricos em Neu5Ac- α 2,3-Gal β 1,4-GlcNAc (Sia2,3, daqui para frente), o tipo de receptor reconhecido pelo vírus da gripe aviária (Cone, 2009; Cohen, et al., 2013). Os dois tipos de sialilglicanos estão presentes em diferentes proporções em diferentes regiões do trato respiratório humano, e são essenciais para o sucesso da infecção pelo vírus da Influenza A (Shynia, et al., 2006; Nicholls, et al., 2007).

Sialiltransferases são o grupo de enzimas responsáveis pela adição da porção Neu5Ac (Sia, daqui pra frente) aos glicanos (Cohen & Varki, 2010). Vinte genes que codificam sialiltransferases já foram descritos no genoma humano (Harduin-

Lepers, et al., 2005). Tanto sialiltransferases do tipo α 2,3 (ST3GAL1 a 6) quanto do tipo α 2,6 (ST6GAL1 e 2) são responsáveis pela sialilação dos glicanos que ficam ancorados à membrana plasmática e dos que são excretados para o muco. Alguns desses genes são expressos em apenas em tecidos específicos, e diferentes sialiltransferases adicionam o Sia a diferentes substratos. Por exemplo, a sialiltransferase codificada pelo gene ST6GAL1 adiciona a porção Sia na posição α 2,6 a glicanos do tipo II (Gal β 1,4GlcNAc) ligados pela porção amino (N-ligados), e, raras vezes, também a glicoproteínas ligadas pela porção ácido (O-ligados). Em contraste, ST6GAL2, que é expresso apenas no cérebro e em tecido fetal, adiciona o Sia na posição α 2,6 a oligossacarídeos, mas não a proteínas (Glaser, et al., 2007). Como nos interessa apenas sialiltransferases que sejam expressos no tecido epitelial respiratório (já que a forma de infecção do influenza em humanos se dá por essa via), selecionamos os genes ST6GAL1, ST3GAL1, ST3GAL2, ST3GAL4 e ST3GAL5 como relevantes para este trabalho (Kitagawa & Paulson, 1994; Harduin-Lepers, et al., 2001; Glaser, et al., 2007).

Para analisar os mecanismos e mutações que podem conferir imunidade celular completa ao vírus da influenza, Carette e colaboradores (2009) desenvolveram ensaios de mutagênese insercional em linhagens celulares humanas haplóides (exceto para o cromossomo 8). Eles mostraram que três inserções independentes no gene SLC35A2, e duas inserções independentes no gene CMAS, estão associadas com a imunidade celular completa ao vírus da Influenza. O gene CMAS codifica a enzima que liga o Sia a um resíduo de cistidina-monofosfato. As sialiltransferases não podem transferir o Sia para um aceptor a não ser que este esteja ligado a uma cistidina monofosfato, uma vez que Sia livres no citoplasma não podem ser transportados para a membrana do Complexo de Golgi, onde a transferência ocorre (Fleischer, 1998; Münster, et al., 1998). O gene SLC35A2 codifica um transportador de solutos da família 35, do tipo A2. Esse transportador é, justamente, responsável pelo transporte do CMP-Neu5Ac (Sia ligado à cistidina monofosfato) do citoplasma para o Complexo de Golgi.

Os receptores glicanos são extremamente importantes para sinalização e adesão celular, e, portanto, apresentam papel crucial para a manutenção do ciclo celular. Considerando o papel fundamental desses receptores para a adesão e

entrada do vírus na célula epitelial, decidiu-se investigar se a falta desses receptores ou defeitos na sua expressão estão associadas a patologias conhecidas. De fato, o mal funcionamento do gene SLC35A2 resulta em uma doença extremamente severa: defeito de glicosilação do tipo II_m (Ng, et al., 2013). Distúrbios causados pela expressão deficiente de SLC32A5 são muito severos e incluem atrasos de desenvolvimento, hipotonia, anomalias oculares, má formação cerebral associada a perfis anormais de transferrina, convulsões, funcionamento anormal do cérebro (hipsarritimia), deficiência nutricional, microcefalia, infecções recorrentes, dismorfismo, membros encurtados e defeitos de coagulação (Ng, et al., 2013). Todos os indivíduos estudados por Ng e colaboradores apresentaram proporções muito baixas de sialilglicanos. O profundo quadro fenotípico e curta expectativa de vida desse pacientes sugere que defeitos no SLC35A2 e/ou CMAS são incompatíveis com a vida. Portanto, conclui-se que o vírus da influenza se liga a um receptor essencial e obrigatoriamente presente nas células de indivíduos saudáveis.

Examinou-se a variabilidade dos genes responsáveis pela produção e montagem dos sialilglicanos, buscando evidenciar se são genes mais conservados, a partir da busca de variantes descritas no 1000 Genomes ENSEMBL Browser durante a fase piloto do projeto (The 1000 Genomes Project Consortium, 2010). O tamanho dos genes variou de 8,8Kb (SLC35A2) até 148Kb (ST6GAL1), tendo o último mostrado 0,5 variantes/Kb, enquanto o primeiro mostrou 7.2 variantes/Kb. A maior parte das variantes encontradas em todos os genes estudados foram mutações em regiões intrônicas dos genes (22 a 66,7% das variantes descritas), exceto para o gene ST6GAL1, para o qual 72% das variantes estavam localizadas na região 3'UTR (UnTranslated Region) com 9% de variância (tabela II.1).

Tabela II. 1 Lista de variantes para sete genes envolvidos na ligação do vírus da influenza à célula (dados retirados do banco de dados do 1000 Genomes Project).

Tipo de variante*	CMAS	SCL35A2	ST3GAL1	ST3GAL2	ST3GAL4	ST3GAL5	ST6GAL1	Total
Codante com frameshift	0	4	0	2	4	0	0	10
Codante não-sinônima	3	14	4	2	7	11	0	41
Sítio de splice	3	0	4	2	8	5	1	23
Codante sinônima	1	2	10	2	17	0	9	41
Intrônica	20	30	27	24	67	83	7	258
5'UTR	1	1	21	17	1	6	2	49
3'UTR	2	10	12	32	6	82	54	198
Upstream	1	0	0	1	3	5	0	10
Downstream	2	2	10	2	1	4	2	23
Total de variantes	33	63	88	84	114	196	75	653

*As variantes genéticas descritas na fase piloto do 1000 Genomes Project são: codante com frameshift (alteração do quadro de leitura); coding não-sinônima (alteração de um nucleotídeo na sequência codante que altera um aminoácido no peptídeo resultante; sítio de splice – 1 a 3 pb dentro de um íntron ou de um exon; codante sinônima (alteração na sequência nucleotídica que não altera o aminoácido no peptídeo); intrônica (mutação em um íntron); 5'UTR e 3'UTR mutação na região 5' ou 3' não traduzidas do gene; Upstream (mutação num intervalo de 5kb antes da região 5' de um transcrito; Downstream (mutação num intervalo de 5kb à frente da região 3' de um transcrito).

O gene SLC35A2 apresentou tanto mutações não-sinônimas quanto de mudança de quadro de leitura (frameshift mutations), que potencialmente poderiam alterar a estrutura e a função da proteína produzida. Interessante, porém, que este foi o único gene para o qual o mau funcionamento está associado a patologia severa e consequente morte precoce do indivíduo e que a ausência de uma cópia funcional leva a defeitos incompatíveis com a vida (Ng, et al., 2013). Não obstante, considerando a importância do transporte de CMP-Neu5Ac do citoplasma ao Complexo de Golgi, alguns estudos revelaram que o produto do gene SLC35A3 (que é um transportador de UDP-Gal/UDP-GlcNAc) pode auxiliar o transporte dessa molécula se SLC35A2 não está funcionando corretamente (Olczak, et al., 2013). Nenhuma das duas inserções no gene CMAS ou das três inserções no gene SLC35A2 descritas por Carette e colaboradores (2009) foram encontradas na busca por mutações conhecidas no banco de dados do 1000 Genomes Project.

Vários outros estudos relataram alterações na afinidade de ligação da molécula HA ao ácido siálico do hospedeiro que estão relacionadas a mudanças na estrutura da própria HA, levando a uma alteração conformacional no sítio de ligação da proteína (Ohuchi, et al., 1997; Matrosovich, et al., 2000; Hatta, et al., 2001;

Gambaryan, et al., 2006; Bateman, et al., 2008; Tria, et al., 2013). A ligação entre a HA do vírus da influenza A ao sialilglicano (tanto $\alpha 2,3$ quanto $\alpha 2,6$) é bastante fraca ($K_{diss} > 10^{-4}M$). Como resultado, o sucesso de entrada do vírus na célula hospedeira é facilitado quando várias HAs se ligam a vários sialilglicanos ao mesmo tempo, estabilizando a aderência do vírus à membrana da célula (Matrosovich & Klenk, 2003). Se qualquer uma das variantes encontradas nos genes de produção e montagem dos sialilglicanos afetar a disponibilidade ou a quantidade dessa molécula na superfície da célula hospedeira, afetará o sucesso da taxa de infecção do vírus da influenza. Nesse contexto, é sabido que a topologia e a densidade dos glicanos na membrana celular, bem como o tempo de incubação do vírus também estão envolvidos na ligação das proteínas aos glicanos (Lewallen, et al., 2009), e que as diferenças de especificidade de ligação dos vírus humano e aviário se apóiam não somente no tipo de ligação (se é $\alpha 2,3$ ou $\alpha 2,6$), mas também nas diferenças na estrutura dos glicanos que recebem a molécula de ácido siálico. Cepas humanas de influenza vírus se ligam a ácidos siálicos que apresentam uma forma de guarda-chuvas aberto, que normalmente é encontrado em ácidos siálicos ligados na posição $\alpha 2,6$ a longas cadeias de oligossacarídeos. Por outro lado, cepas aviárias do vírus influenza se ligam a ácidos siálicos ligados nas posições $\alpha 2,3$ ou $\alpha 2,6$ apenas se estiverem ligados a curtas cadeias de oligossacarídeos, que exibem uma forma de cone (Chandrasekaran, et al., 2008).

A ligação do vírus ao receptor e sua subsequente entrada na célula hospedeira (figura II.1) são duas etapas iniciais da interação vírus-hospedeiro, nas quais a variabilidade genética humana pode tanto predispor quanto proteger da infecção. A forma como o vírus acessa a célula e os numerosos mecanismos do hospedeiro que previnem a infecção viral também estão sujeitos às alterações genéticas. Um dos genes do hospedeiro que protege contra o vírus da influenza é o gene CD55, que codifica um fator de aceleração do decaimento que previne danos celulares causados por moléculas do complemento (Osuka, et al., 2007). O gene CD55 é expresso em todos os tipos de células, de duas formas: ancorado à membrana plasmática (mCD55), especialmente expresso nas células do sangue; e uma forma solúvel (sCD55) encontrada nos fluídos corporais (Medof, et al., 1987). Esse fator está associado à prevenção de lesões pulmonares, sendo amplamente expresso nesse

tipo de tecido (Osuka, et al., 2007). Um SNP na região promotora do gene CD55 (rs2564978) foi associado com uma forma severa da infecção por influenza A H1N1 pandêmica de 2009 em chineses e japoneses. O genótipo rs2564978T/T foi altamente associado à forma severa da infecção, enquanto que o genótipo rs2564978C/C foi associado a uma forma mais branda da infecção por influenza (Zhou, et al., 2012). Ao procurar as frequências desse SNP no banco de dados do 1000 Genomes Project, encontrou-se que o alelo rs2564978T é muito frequente nas populações chinesas (0,54 em chineses de Pequim e 0,63 em chineses Han do sudeste), enquanto que em outras populações mundiais as frequências desse alelo variaram de 0,017 (Africanos Yoribás de Ibadan) até 0,39 (Japoneses de Tóquio). Ainda que os africanos tenham apresentado maiores frequências do alelo protetor, a população africana foi uma das que mais sofreu com a epidemia de influenza A H1N1 em 2009, com uma taxa de mortalidade relativa de 7,8 mortes/100000 habitantes nos primeiros 12 meses de circulação do vírus (Dawood, et al., 2012). A segunda população mais afetada foram as populações do sudeste asiático, que apresentaram de 3,3 a 4,4 mortes/100000 habitantes (respectivamente, mortes por insuficiência respiratória e mortes por insuficiência respiratória + insuficiência cardíaca). Europeus e habitantes do Pacífico Oeste apresentaram as menores taxas de mortalidade associada à infecção por influenza vírus (1,8 e 1,7 mortes por 100000 habitantes), o que pode estar relacionado à alta frequência do alelo protetor C nessas populações: 0,7 a 0,9.

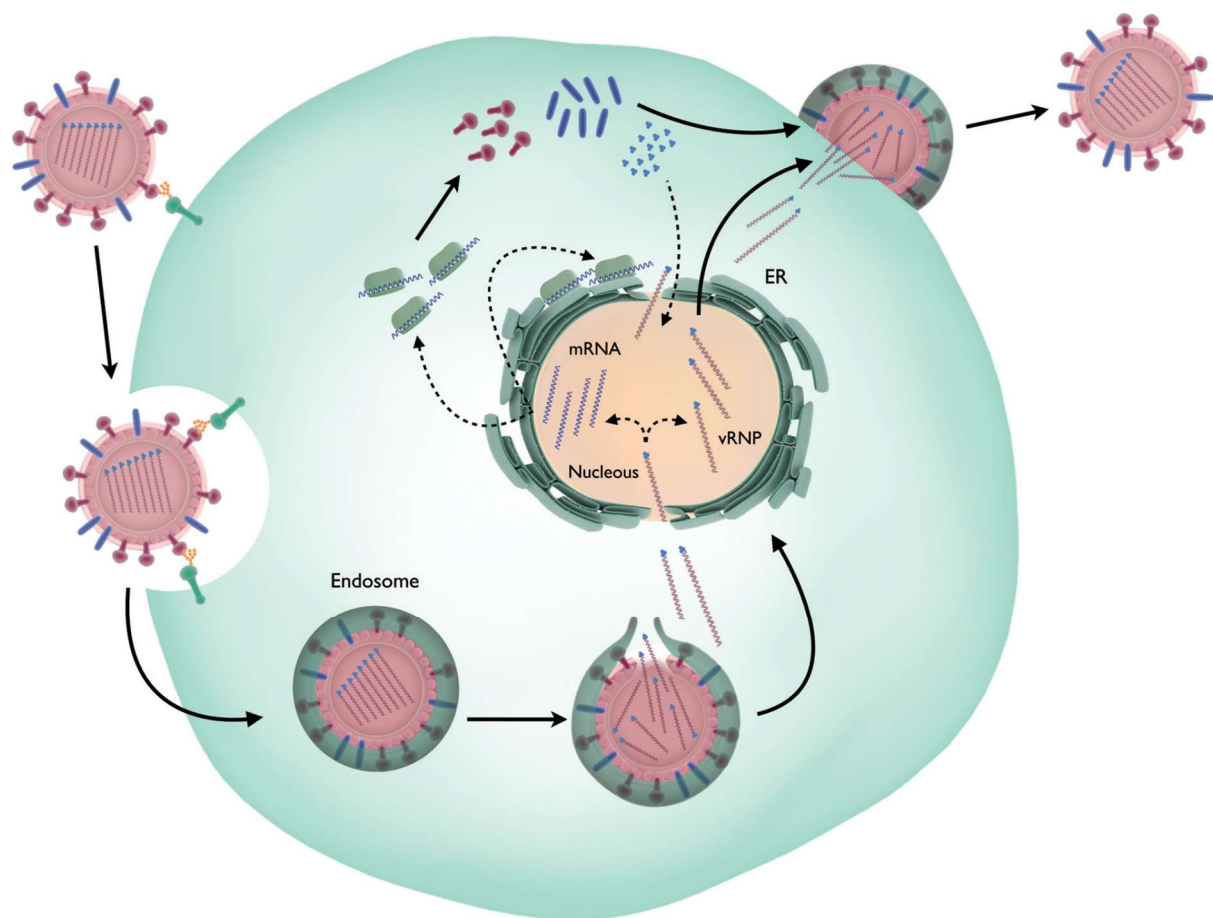


Figura II. 1 Ciclo de Vida do Vírus da influenza. Durante o primeiro passo da infecção pelo vírus da influenza, a glicoproteína viral HA se liga a receptores na superfície da célula hospedeira. O vírus entra na célula por meio de endocitose. O baixo pH dos endossomos causa uma alteração na conformação na HA, levando à fusão do vírus ao endossomo e a consequente liberação dos complexos virais de ribonucleoproteínas (vRNPs) no citoplasma. As vRNPs são transportadas para o núcleo, onde a transcrição do RNAm e a replicação/montagem de novos vRNPs ocorrem. A tradução do RNAm viral acontece tanto no citoplasma quanto no Retículo Endoplasmático (RE). Seguindo as modificações Pós-Traducionais (PTM) de determinadas proteínas virais, os componentes estruturais da partícula viral são transportados para a membrana plasmática junto com novos vRNPs, onde a montagem de novas partículas virais e o brotamento do vírus ocorrem.

II.3. As propriedades protetoras do muco contra o vírus da influenza A

O muco epitelial é uma barreira dinâmica e semipermeável que permite a troca de nutrientes, água, gases, partículas, hormônios e gametas, enquanto é impermeável à maior parte das bactérias e de muitos outros patógenos (Cone, 2009).

Ele está presente no trato respiratório, onde recebe o nome de *Airway Surface Liquid* (ASL, ou muco respiratório, em português) e constitui a primeira barreira inata a moléculas exógenas e patógenos (Lillehoj & Kim, 2002). O muco respiratório é composto por uma mistura estritamente regulada de água, sais e várias macromoléculas, tais como mucinas, proteoglicanos, lipídios e outras proteínas que conferem viscosidade e as funções de defesa ao muco. Quando qualquer um desses componentes está em falta, ou está presente em concentração subótima, existe uma grande probabilidade de surgimento de doenças, tais como asma, fibrose cística e outras síndromes respiratórias (Lillehoj & Kim, 2002). O muco respiratório é rico em α 2,3 e α 2,6-sialilglicanos, tanto na forma livre (secretada) quanto nas membranas das células epiteliais e outras células circulantes (Nicholls, et al., 2007).

Foi sugerido que o principal tipo de molécula presente no muco respiratório é o das mucinas, devido à grande concentração delas em sua forma secretada e ao grande número de genes que codificam tanto as formas ancoradas quanto secretadas de mucinas presentes no genoma humano. Já foram descritos treze genes diferentes que produzem tal molécula no nosso genoma, sendo que oito deles são expressos largamente no trato respiratório humano – cinco na forma secretada e três na forma ancorada (Lillehoj & Kim, 2002). Mucinas são glicoproteínas de grande peso molecular que contém um número grande e variável de repetições de aminoácidos em *tandem* (consecutivos) enriquecidos em resíduos de serina, treonina e prolina (Ser, Thr, Pro, respectivamente) (Voynow, et al., 1998; Ogasawara, et al., 2007). A presença desses resíduos repetitivos de aminoácidos é responsável pela alta glicosilação dessas moléculas, e subsequente heterogeneidade de tamanhos e carga eletrônica das mucinas (*dispersity*, sem tradução para o português). A glicosilação se dá entre os resíduos Ser/Thr da cadeia principal da proteína e do resíduo N-Acetilgalactosamina dos oligossacarídeos, característico de glicoproteínas O-ligadas (Lillehoj & Kim, 2002; Rose & Voynow, 2006). 95% de todas as mucinas secretadas são do tipo MUC5, que são altamente O-glicosiladas e servem de acceptor tanto para ácido siálico ligado na posição α 2,3 quanto α 2,6 (Lillehoj & Kim, 2002). O ácido siálico adicionado a essas mucinas conferem ao mucus suas características lubrificantes e de alta viscosidade (Nicholls, et al., 2012). Já MUC2 é uma forma secretada de mucinas que compreende cerca de 2,5% das mucinas produzidas no muco respiratório (Kirkham, et al., 2002;

Rose & Voynow, 2006), e que apresentam mais de 40% de oligossacarídeos sialilados (Karlsson, et al., 1997). Sua expressão é aumentada durante processos inflamatórios e de infecção, especialmente por meio de ativação por IL-4 e IL-9 (Lillehoj & Kim, 2002). O muco respiratório também é rico em outras proteínas de defesa, tais como inibidores de proteases, antioxidantes, proteases, proteínas antimicrobianas, IgA e citocinas. A alta concentração de IgA no muco é indicativa de início de infecção, sugerindo que anticorpos e imunoglobulinas são secretadas no muco para atuar nas primeiras fases da infecção (Kreijtz, et al., 2011).

Algumas proteases são capazes de ativar o vírus da influenza, possivelmente sendo responsáveis pela pneumopatogenicidade do vírus (Lillehoj & Kim, 2002). Mucinas que têm função antiprotease aparentemente têm a capacidade de melhorar o sistema imune do hospedeiro, melhorando o prognóstico da infecção e facilitando a eliminação do vírus (Nadziejko & Finkelstein, 1994). Nicholls e colaboradores (2007) hipotetizaram que a alta concentração de proteínas sialiladas serviriam como uma barreira física para o vírus da influenza, evitando que o vírus se ligasse aos ácidos siálicos ancorados às membranas das células epiteliais. A função fisiológica dos receptores de ácidos siálicos presentes nas células epiteliais é justamente promover a aderência do muco à superfície dessas, protegendo-as de desidratação, micróbios, e radicais livres produzidos ou por agentes exógenos ou pela ação de leucócitos (Ogasawara, et al., 2007). Entretanto, as sialoglicoproteínas são mediadoras da aderência celular e viscoelasticidade do muco, e servem como receptores para a ligação de elementos exógenos, incluindo a hemaglutinina do vírus da influenza.

A linha de defesa do muco respiratório compreende um gel viscoelástico de 50µm de espessura que imobilizam vírus e bactérias, que posteriormente são eliminados por meio dos batimentos ciliares executados pelas células epiteliais ciliadas do trato respiratório. A grande eficiência de defesa do muco provavelmente está relacionado à presença das proteínas sialiladas neste, já que mimetizam o receptor ao qual o vírus da influenza se liga, fazendo com que o vírus fique preso no muco e seja eliminado antes de chegar à célula epitelial (Duez, et al., 2008; Cohen, et al., 2013). Cohen e colaboradores (2013) descobriram que quanto maior a concentração de ácidos siálicos presentes no muco, menos células epiteliais era

infectadas pelo vírus da influenza em ensaios in vitro, especialmente se as culturas de célula eram tratadas com fosfato de oseltamivir, um inibidor de neuraminidase (NA). No entanto, na ausência do inibidor, as partículas virais conseguiam chegar até as células epiteliais, já que a neuraminidase é capaz de liberar o vírus das proteínas do muco quebrando a ligação da hemaglutinina ao ácido siálico. A taxa de eliminação dos vírions depende da interação entre o muco e os cílios das células epiteliais adjacentes. Ainda não é claro o que determina a taxa de eliminação do muco, que varia entre as porções proximal e distal do sistema respiratório. Certamente, a eficácia da atividade motora dos cílios é o primeiro determinante dessa eliminação, mas a quantidade e a viscoelasticidade das mucinas também são variáveis importantes (Knowles & Boucher, 2012).

Além do mais, proteínas com domínios ricos em cistidina (*cystidine-rich domain*, CRD) provavelmente interagem com as proteínas virais hemaglutinina e neuraminidase, já que essas duas proteínas contêm cadeias laterais de oligossacarídeos N-ligados. Yang e colaboradores (2011) mostraram que Galectina-1 pode se ligar ao vírus da influenza e inibir a infecção tanto in vitro quanto em ratos. Os autores observaram que a expressão do gene da galectina-1 é aumentada em animais durante o curso da infecção de acordo com a carga viral (quanto maior a carga viral, mais expresso é o gene). Além disso, ratos que foram infectados com uma cepa neurovirulenta de influenza A mostraram maior taxa de sobrevivência quando tratados intranasalmente com galectina-1, resultando em uma menor taxa apoptótica e inflamatória nos pulmões, corroborando que a galectina-1 é um eficiente inibidor da reprodução viral e reduz os sintomas da infecção.

II.4. Dentro da célula hospedeira: resposta imune inata à infecção pelo vírus da influenza A

O sistema imune inato é a primeira linha de defesa do organismo contra uma infecção viral (figura II.2). Durante esta fase, as células epiteliais do hospedeiro

conseguem reconhecer novas infecções por influenza a partir de receptores de reconhecimento de padrões (*pattern-recognition receptors*, PRRs). Os PRRs reconhecem padrões moleculares associados a patógenos (*pathogen-associated molecular patterns*, PAMPs) que são expressos durante a replicação do vírus, no interior da célula (Hale, et al., 2010). O principal PAMP do vírus da influenza é um RNA viral que contém um grupo trifosfato na região 5' (PPP-RNA), uma assinatura molecular que distingue RNA exógeno do RNA autóctone (Abbas, et al., 2013). Os receptores capazes de detectar o RNA viral são: receptores Toll-like (TLRs), receptores RIG-I-like (RLRs), receptores NOD-like (NLRs) (Pang & Iwasaki, 2011) e proteínas com repetições de tetratricopeptídeos induzidas por interferons (IFITs, Abbas, et al., 2013).

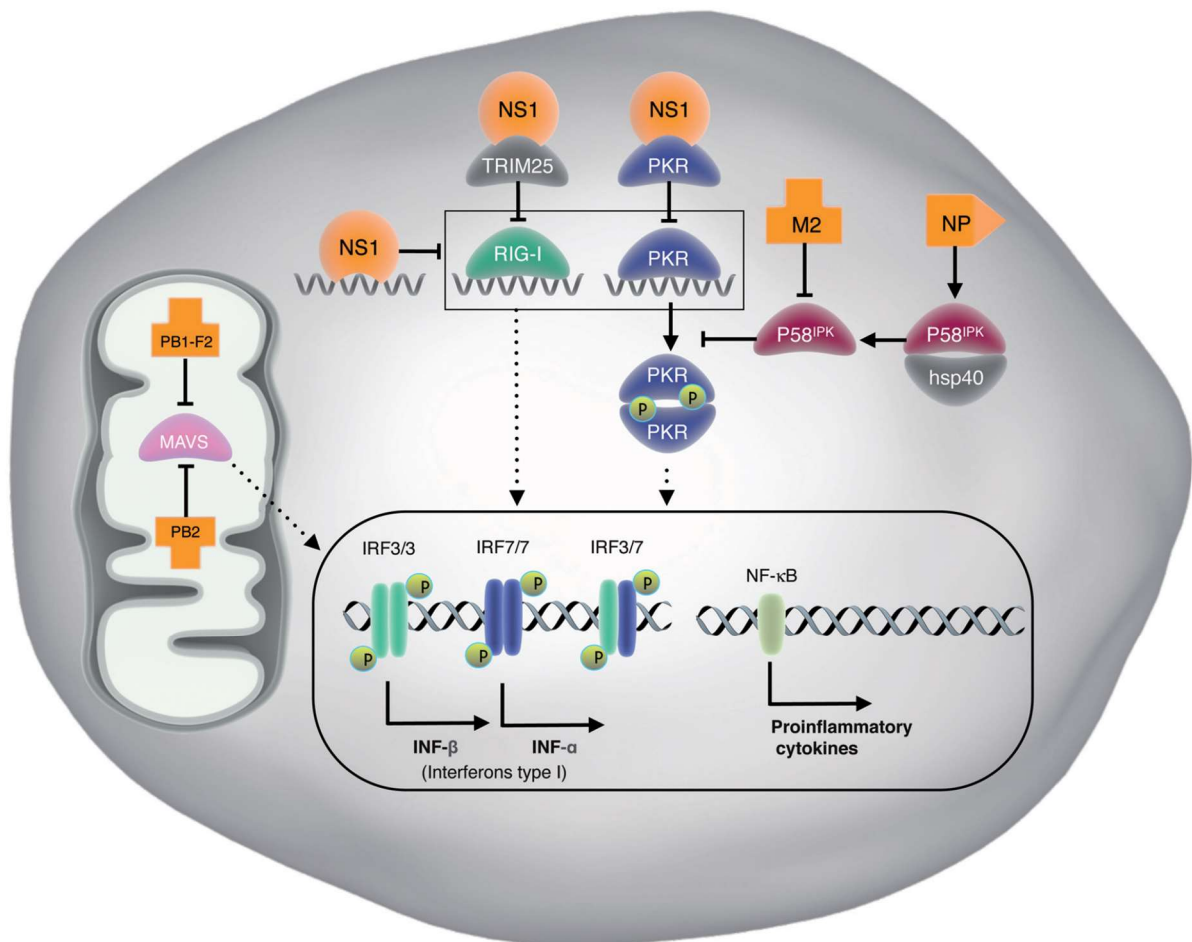


Figura II. 2 Escape da resposta adaptativa. Durante a infecção pelo vírus da influenza, o RNA viral é detectado e várias vias são ativadas. Isso resulta na produção de interferons tipo-I, citocinas pró-inflamatórias e indução de um estado antiviral da célula. As proteínas do vírus influenza interagem com essa cascata. NS1 se liga ao RNA viral para mascarar a detecção por PRRs. Ela bloqueia a ativação de RIG-I se ligando a TRIM25 e PKR, inibindo sua fosforilação, que ocorre quando PKR se liga ao RNA

viral. NP interage com o complexo P58IPK/hsp40, que então libera P58IPK e inibe a fosforilação de PKR. M2, por outro lado, restaura o complexo P58IPK/hsp40, inibindo a síntese proteica e induz a célula à apoptose, facilitando a liberação de partículas virais. PB1-F2 se liga à proteína sinalizadora antiviral mitocondrial (MAVS), inibindo essa proteína e a produção de interferon tipo-I. PB2 também se liga a MAVS e inibe sua função. Além disso, PB2 se liga ao estimulador de promotor interferon 1 (IPS-1), resultando na inativação do promotor que levaria à produção de interferon tipo-I. Setas pontilhadas representam vias e linhas sólidas e setas representam interações diretas.

A função desses receptores é iniciar a expressão de citocinas pró-inflamatórias e interferons do tipo I, IFN- α e IFN- β , responsáveis por inibir a síntese de proteínas na célula hospedeira, limitando a replicação viral (Heil, et al., 2004; Lund, et al., 2004). Além disso, esses receptores também estimulam células dendríticas a aumentar a apresentação de antígenos CD4+ e CD8+, consequentemente aumentando a resposta imune adaptativa (Kreijtz, et al., 2011). A resposta imune inata celular é feita principalmente pelas células dendríticas e pelos macrófagos. No entanto, os macrófagos alveolares exercem papéis aparentemente contraditórios após sua ativação nos pulmões. Por um lado, eles fagocitam células que estejam infectadas pelo vírus da influenza, limitando a dispersão dos vírus no organismo (Tumpey, et al., 2005; Kim, et al., 2008); por outro lado, eles produzem óxido nítrico sintase 2 (NOS2) e fator de necrose tumoral alfa (TNF α), contribuindo com a sintomatologia induzida pela infecção por influenza vírus (Jayasekera, et al., 2007; Lin, et al., 2008). Esses efeitos contraditórios claramente evidenciam o delicado equilíbrio que existe na regulação e atuação do sistema imune. Outra função dos macrófagos alveolares é o de regular a resposta imune, especialmente influenciando o desenvolvimento da imunidade mediada por células-T antígeno-específicas (Wijburg, et al., 1997). Cabe ressaltar que, embora macrófagos alveolares produzam apenas baixas quantidades de citocinas pró-inflamatórias (van Riel, et al., 2011), macrófagos derivados da corrente sanguínea produzem altas quantidades dessas citocinas quando infectados pelo vírus da influenza.

II.4.1. Receptores Toll-Like (TLRs)

Receptores Toll-like são uma família altamente conservada de glicoproteínas reconhecedoras de padrão que desempenham papel fundamental no reconhecimento de vírus pelo sistema imune inato (figura II.2). TLRs contém um domínio N-terminal ligante a PAMP, extracelular, rico em repetições de leucinas (*leucine-rich repeats*, LRRs), que está ligado a um domínio intracelular C-terminal com função sinalizadora, chamado de homologia Toll/IL-1R (TIR). O domínio TIR recebe esse nome pela sua similaridade ao receptor da interleucina 1 (IL-1R) que media cascatas sinalizadoras após sua ativação (Jensen & Thomsen, 2012). A ativação de TLR inicia uma cascata de sinalização que envolve proteínas adaptadoras, proteínas quinases e fatores de transcrição efetores, culminando na produção de interferons do tipo I (IFN- β). Esses processos induzem a célula hospedeira a entrar num estado antiviral, equipando-a com uma importante primeira linha de defesa contra esses patógenos. Outras citocinas pró-inflamatórias induzidas por receptores TLR, tais como o fator de necrose tumoral e a interleucina 6, também são importantes para o delicado equilíbrio entre a resposta imune eficiente e a imunopatologia.

Já foram descritos dez tipos diferentes de TLRs em humanos (TLR1-10). Dentre eles, os TLR3, 7, 8 e 10 são reconhecidamente capazes de reconhecer infecções por influenza vírus (Koyama, et al., 2007; Le Goffic, et al., 2007; Lee, et al., 2014). TLR7 e TLR8 são os receptores com maior similaridade de sequência, e são responsáveis pela detecção de oligonucleotídeos de RNA fita simples (ssRNA) que contenham sequências ricas em repetições de guanosina ou uridina. TLR7 é preferencialmente expresso em endossomos de células dendríticas plasmocitóides (pDCs) e células B, enquanto que TLR8 é preferencialmente expresso em células dendríticas mielóides e monócitos. Quando estimulados, TLR7 e TLR8 recrutam um adaptador que contém um motivo TIR chamado Gene 88 de Resposta a Diferenciação Mielóide Primária (*Myeloid Differentiation primary response gene 88*, MyD88) para que este interaja com o motivo TIR do receptor TLR, que fica voltado para o citoplasma (Medzhitov, et al., 1998). Essa cascata de sinalização leva à ativação dos complexos NF- κ B/AP-1 ou NF- κ B/PA-1/IRF7, que resultam na transcrição de citocinas pró-

inflamatórias e IFN- α , respectivamente. TLR3 fica localizado dentro do compartimento intracelular de macrófagos, linfócitos B e células dendríticas, e também na superfície da membrana de células NK (*Natural Killer*), epiteliais e fibroblastos. Esse receptor reconhece RNA viral de dupla fita (dsRNA) presente em alguns vírus, levando à expressão do complexo NF- κ B/AP-1/IRF3 e aumento da expressão de IFN- β (Jensen & Thomsen, 2012). Estudos recentes evidenciaram que há um aumento na expressão de TLR10 durante a infecção pelo vírus da influenza H5N1, especialmente em macrófagos. Os dados gerados em tal estudo sugerem que TLR10 é capaz de aumentar a expressão de IL-8 induzida por vRNP (Lee, et al., 2014).

II.4.2. Receptores RIG-I-like (RLRs)

Os receptores RIG-I-like são proteínas citosólicas que reconhecem RNA viral e são expressos na maioria dos tipos celulares humanos (figura II.2). Eles contêm uma estrutura central conservada, caracterizada por um domínio central de helicase DExD/H-box e um domínio C-terminal (CTD) ligante a ssRNA/dsRNA que, quando livre, funciona como um repressor de domínio (RD) (Gack, 2014). Três RLRs estão diretamente implicados no reconhecimento viral: o produto do gene I induzido por ácido retinóico (*Retinoic-acid Inducible Gene 1* – RIG-I – codificado pelo gene DDX58); o laboratório de genética e fisiologia 2 (*Laboratory of Genetics and Physiology 2* – LGP2 – codificado pelo gene DHX58); e o antígeno 5 associado à diferenciação de melanoma (*Melanoma-Differentiation-Associated antigen 5* – MDA5 – codificado pelo gene MDA5). RIG-I e MDA5 contêm dois domínios N-terminais de recrutamento e ativação de caspase (*Caspase Activation and Recruitment Domain*, CARD) que, quando da detecção do vírus, iniciam uma cascata de sinalização que culmina em expressão de genes de IFN-I. Em contraste, o LGP2 não tem o domínio CARD, mas exerce função regulatória da sinalização mediada por RLR, ainda que não se conheça exatamente o mecanismo (Kato, et al., 2006). RIG-I e LGP2 são importantes para o reconhecimento do vírus da influenza, mas MDA5 também é capaz de reconhecer RNA de outros patógenos também (Jensen & Thomsen, 2012). Estudos recentes mostraram que a atividade de RIG-I é estritamente regulada por modificações pós-

traducionais; em particular, a desfosforilação dos domínios CTD (Thr770, Ser854/855) e CARD (Ser8, Thr170) e subsequente poliubiquitinação das lisinas Lys63 e Lys172. Os CARDS ligados a ubiquitinas facilitam a oligomerização de RIG-I e sua ligação à MAVS, induzindo a sinalização antiviral (Gack, 2014).

II.4.3. Receptores NOD-like (NLRs)

Receptores que contém domínio de oligomerização de ligação a nucleotídeos (*Nucleotide-binding Oligomerization Domain*, NOD) são proteínas citosólicas envolvidas na regulação de respostas inflamatórias e apoptóticas, em particular durante respostas antivirais (figura II.2). Receptores NLR contém uma região C-terminal rica em repetições de leucina (*leucine-rich repeats*, LRRs), que é considerada a região capaz de detectar RNAs. Uma região central NACHT media a oligomerização e ativação do receptor, enquanto que a região N-terminal é responsável pela ligação à proteína efetora, normalmente um domínio CARD ou pirina (*pyrin domain*, PYD) que inicia a cascata de sinalização (Fritz, et al., 2006). Dos vários receptores NLRs atuantes no organismo humano, NLRC2 (ou NOD2) e NLRP3 estão diretamente envolvidos na detecção do RNA de vírus da influenza.

NLRC2, que contém dois domínios CARD na região N-terminal (Ogura, et al., 2001), recentemente foi implicado no reconhecimento de ssRNA pertencentes a RSV, vírus da parainfluenza e vírus da influenza A. Quando do reconhecimento do ssRNA viral, NLRC2 se associa à proteína MAVS (codificada pelo gene VISA) através de uma interação dependente de LRR e do domínio de ligação a nucleotídeo (*nucleotide-binding domain*, NBD). Essa interação inicia a cascata de sinalização dependente de MAVS (similar ao mecanismo utilizado pelos RLRs já descrito), culminando na expressão de interferons do tipo I e citocinas pró-inflamatórias. Diferente do mecanismo dos RLRs, contudo, a interação NLRC2-MAVS não envolve a ligação de domínios CARD-CARD (Sabbah, et al., 2009). NLRP3 contém um domínio pirina (PYD) que pode interagir com o domínio PYD da proteína ASC (*apoptosis-associated speck-like protein*), que também contém um domínio CARD. NLRP3 se oligomeriza e recruta ASC e uma procaspase-1 para formar um

inflamassomo. Esse inflamassomo ativa uma caspase-1, que subsequentemente media a conversão de pro-IL-1beta e pro-IL-18 em proteínas funcionais. Essa ativação também ocorre no caso de infecções por adenovírus (vírus de DNA), Sendai virus (ssRNA) e vírus da influenza A (Allen, et al., 2009; Ichinohe, et al., 2009).

II.4.4. Interferon-induced proteins with tetratricopeptide repeats (IFITs)

IFITs são proteínas efetoras do sistema imune inato que parecem conferir defesa contra vírus por meio da interrupção de interações proteína-proteína na maquinaria de iniciação da tradução do hospedeiro. Porém, IFITs são capazes de diretamente reconhecer o RNA viral que contenha um trifosfato na região 5' (5'-PPP-RNA). IFIT1, e IFIT5 em particular, são capazes de se ligar ativa e seletivamente a 5'-PPP-RNAs livres no citoplasma da célula (Pilchmair, et al., 2011).

II.4.5. Escape viral da imunidade inata

Dois processos evolutivos estão envolvidos na evolução viral e na evasão do sistema imune inato do hospedeiro: a deriva antigênica sofrida pelo vírus da influenza e a alta pressão seletiva exercida pela resposta imune humana (figura II.2). Esses mecanismos de escape envolvem interações entre as proteínas efetoras virais e alguns receptores humanos. A proteína viral NS1 é capaz de se ligar ao RNA do vírus, ocupando a região de interação do RNA com os receptores TLRs, evitando que a cascata de sinalização seja iniciada e que IFN-I seja produzido (García-Sastre, 2004; Guo, et al., 2007). Essas proteínas virais também são capazes de inibir a ligação entre RIG-I e TRIM25 (*tripartite-motif-containing protein 25*) e a ativação de uma proteína quinase detectora de ssRNA (PKR, codificada pelo gene EIF2AK2) (Tan &

Katze, 1998; García-Sastre, 2004; Gack, 2014). Somando a isso, a nucleoproteína e a proteína M2 de canal iônico do vírus da influenza podem se ligar e inibir a proteína P58^{IPK} e o complexo P58^{IPK}/hsp40, respectivamente. Esses dois últimos são efetores do hospedeiro que têm função crucial na regulação da atividade de PKR, e a inativação desses dois resulta em inibição da síntese proteica, indução de apoptose celular e liberação de partículas virais recém-montadas (Guan, et al., 2010).

A tabela II.2 mostra a variabilidade dos genes do hospedeiro que estão envolvidos nos mecanismos de escape viral. O gene EIF2AK2 é o mais variável dos quatro genes mais importantes para o mecanismo de detecção do vírus. Como era de se esperar, a maioria das variantes observadas estão localizadas nas regiões não-codificantes. Das variantes em região codificadora, a maior parte das mutações é do tipo mutação de nucleotídeo único (houve tanto mutações sinônimas quanto não-sinônimas).

Tabela II. 2 Genes do hospedeiro envolvidos nos mecanismos de escape viral. DNAJC3 (codifica o inibidor P58IPK); EIFAK2 (codifica PKR); TRIM25 (codifica a proteína que contém um motivo-tripartido); VISA (codifica a proteína MAVS); (dados retirados do banco de dados do 1000 Genomes Project).

Tipo de variante*	DNAJC3	EIF2AK2	TRIM25	VISA	Total
Codante com frameshift	0	0	0	0	0
Codante não-sinônima	2	15	4	16	37
Sítio de splice	0	8	0	4	12
Codante sinônima	5	16	3	6	30
Intrônica	29	177	9	49	264
5'UTR	2	24	1	0	27
3'UTR	3	67	24	15	109
Upstream	2	5	0	0	7
Downstream	3	3	0	0	6
Outras*	0	2	0	0	2
Total de variantes	46	317	41	90	494

*ver legenda da tabela II.1

II.4.6. Mutações relacionadas à infecção pelo vírus da influenza

Ainda que a variabilidade dos genes envolvidos na resposta imune inata possa ter um impacto importante no desenvolvimento da infecção, o entendimento de quanto essa variabilidade se correlaciona com a susceptibilidade a doenças infecciosas é muito limitado e, geralmente, não é suportado por abordagens experimentais.

Uma mutação com troca de sentido (não-sinônima, F303S) no gene TLR3 foi associada a uma condição neurológica severa e resultante da infecção por influenza (Esposito, et al., 2012), chamada encefalopatia associada a influenza (*influenza-associated encephalopathy*, IAE). O receptor TLR com a mutação F303S se mostrou menos eficiente na ativação do fator de transcrição NF- κ B, bem como em iniciar a cascata de sinalização através do receptor IFN-B. Nesse mesmo estudo, Esposito e colaboradores (2012) identificaram uma segunda mutação, um polimorfismo de nucleotídeo único rs5743313 (*Single Nucleotide Polymorphism*, SNP), ao estudar 51 crianças com diagnóstico confirmado de infecção por H1N1. O genótipo rs5743313C/T foi encontrado em 18/18 crianças que tiveram pneumonia aguda causada por influenza, mas significativamente menos frequente nas crianças que não tinham complicações pulmonares ($p < 0,0001$), corroborando a associação entre o receptor TLR e a patogenicidade da infecção por influenza. Ao buscar essas variantes no banco de dados do 1000 Genomes Browser, percebeu-se que nenhuma dos indivíduos sequenciados pelo projeto piloto carregavam a mutação F303S. Pela natureza deste estudo, inclusive, não foi possível acessar se os indivíduos do banco de dados que carregam o genótipo rs5743313C/T sofreram pneumonia aguda ou sequer foram, algum dia, infectados pelo vírus da influenza A. Por outro lado, foi possível verificar que as frequências dos alelos C e T são similares em todas as populações sequenciadas pelo 1000 Genomes Project, tendo o alelo C uma frequência média de 0,8, sugerindo que cerca de 32% dos indivíduos constantes no banco de dados possam ser portadores do genótipo rs5743313C/T (assumindo, claro, que todas as populações estejam em equilíbrio de Hardy-Weinberg para esse SNP). Mesmo assim, quatro das populações do 1000 Genomes se encontraram fora dessa

faixa de frequência para o alelo C: Britânicos (GBR, C=0,76 C/T=36%), Ibéricos (IBS, C=0,67 C/T=44%), Porto Riquenhos (PUR, C=0,71 C/T=41%), Toscanos (TSI, C=0,69 C/T=43%); o que indica que essas populações podem ser mais suscetíveis a manifestações severas da infecção por influenza, já que a possibilidade de genótipos heterozigotos aumenta quanto mais próximas forem as frequências de C e T. A população japonesa (JPT, C=0,978 C/T=4,3%) apresentou a maior frequência do alelo C, indicando que possivelmente apenas 4% da população deva ser suscetível a complicações pulmonares causadas pela Influenza.

Em um estudo caso-controle de associação com 91 pacientes com diagnóstico confirmado de pneumonia induzida por influenza, quatro outros SNPs foram identificados em genes que a priori não estariam diretamente envolvidos na infecção por influenza (Mänz, et al., 2013). O primeiro desses SNPs está localizado no gene da proteína interatuante na proteína de replicação A (*Replication Protein A Interacting protein*, RPAIN). RPAIN é uma proteína que facilita a localização da proteína RPA no núcleo, e esta tem papel fundamental na manutenção da homeostase e da integridade do DNA. O segundo SNP foi encontrado no gene do componente 1 do complemento, que codifica uma proteína que se liga a um subcomplemento (C1QBP), que tem por função inibir a ativação do sistema complemento, interrompendo a resposta imune. O terceiro SNP foi encontrado no gene que codifica um receptor de membrana que tem pouca afinidade a anticorpos IgG, chamado CD32 (ou FCGR2A). Esse receptor é encontrado na superfície de células fagocitadoras, como macrófagos e neutrófilos que estão envolvidos no processo de fagocitose e eliminação de complexos imunológicos. O último SNP está localizado numa região possivelmente intergênica do cromossomo 3. Uma vez que dois desses SNPs estão envolvidos ou na eliminação de complexos imunes (FCGR2A) ou na inativação do sistema complemento (C1QBP), é razoável hipotetizar que uma manifestação severa da infecção por influenza pode ser resultado de variações da resposta imune do hospedeiro.

Liu e colaboradores (2013) evidenciaram uma associação direta entre os polimorfismos rs17561 no gene IL1A e rs1143627 no gene IL1B e uma susceptibilidade aumentada à infecção por influenza A H1N1. As interleucinas IL1A e IL1B são muito importantes no combate a infecções, já que ambas induzem a expressão de uma

variedade de mediadores inflamatórios que são capazes de iniciar cascatas de respostas inflamatórias e a ativação de células T (Acosta-Rodriguez, et al., 2007). IL1B foi implicado na mediação de inflamação pulmonar aguda, que é uma das principais causas de morte associadas à infecção por influenza A; enquanto que a secreção de IL1A é regulada pelo inflamassomo NLRP3 (Schmitz, et al., 2005; Dawood, et al., 2012; Gross, et al., 2012). A variante rs17561T no gene IL1A está presente em baixas frequências no banco de sequências do 1000 Genomes Project. Essas frequências variam de 0,06 em Chineses Han a um máximo de 0,35 em Britânicos. As populações europeias mostraram as maiores frequências para essa variante, que está associada a um aumento na susceptibilidade à infecção pelo vírus da influenza A H1N1. Já o SNP rs1143627C no gene IL1B é muito frequente em todas as populações do 1000 Genomes Project, particularmente nas populações Mexicana, Porto Riquenha, Queniana e Nigeriana, variando de 0,52 em Mexicanos até 0,75 em Quenianos. Nas outras populações, as frequências desse alelo se mantêm relativamente altas, variando de 0,21 em Ibéricos até 0,48 em Americanos de ascendência Africana. Se, de fato, esse alelo confere um aumento na susceptibilidade à infecção pelo vírus da influenza, então essas altas frequências podem explicar o rápido desenvolvimento da gripe em uma pandemia e a inflamação respiratória aguda observada nas populações humanas com respeito à pandemia de H1N1/2009.

Os mecanismos de reconhecimento de RNA viral e de padrões específicos de sequências são muito importantes nas fases iniciais da infecção por influenza. Mutações nos genes que produzem as proteínas atuantes nesses dois processos podem interferir na eficiência desses reconhecimentos, levando a uma falha da resposta imune e um sucesso da infecção. Deste modo, é impressionante a baixíssima quantidade de mutações encontradas nesses genes, como pode ser observado na Tabela II.3. A maior parte da variação encontrada está localizada nas regiões intrônicas dos genes, que podem não influenciar a expressão ou a função desses genes. Isso não é verdade para os genes TLR, entretanto, já que encontramos uma grande quantidade de mutações sinônimas e não-sinônimas predominantemente em regiões codificadoras, que devem refletir um aumento na variedade de padrões reconhecidos por esses receptores, possivelmente como resultado a pressão seletiva causada pela própria variabilidade do vírus.

Tabela II. 3 Lista das variantes encontradas nos genes envolvidos na resposta imune inata à infecção pelo vírus da influenza (dados retirados do banco de dados do 1000 Genomes Project).

Tipo de variante	DDX58	DHX58	NLRP3	NOD2	IFIT1	IFIT5	TLR3	TLR7	TLR8	TLR10	Total
Codante com frameshift	3	1	0	2	1	0	1	1	0	2	11
Codante não-sinônima	30	8	42	28	5	5	10	11	10	34	183
Sítio de splice	3	3	3	0	0	1	1	0	2	0	13
Codante sinônima	24	1	54	13	6	2	10	6	33	22	171
Intrônica	124	22	234	16	0	2	10	1	0	12	421
5'UTR	12	2	40	3	1	0	4	0	1	8	71
3'UTR	43	0	55	10	1	26	1	13	27	12	188
Upstream	1	0	3	1	6	1	1	0	4	6	23
Downstream	5	0	5	0	3	1	1	0	2	8	25
Outras*	1	0	0	0	0	0	1	1	0	4	7
Total de variantes	246	37	436	73	23	38	40	33	79	108	1113

*Sítio de Splice Essencial (DDX58); ganho de parada (TLR3,TLR7,TLR10). Para descrição dos tipos de variante, ver tabela II.1.

II.5. Espalhando a infecção: a resposta imune adaptativa do hospedeiro ao vírus da influenza A

A fase de resposta imune adaptativa ao vírus da influenza é crucial caso o vírus seja capaz de infectar as células epiteliais do trato respiratório, e pode ser dividida em duas etapas: a resposta imune adaptativa humoral e a resposta imune adaptativa celular (figura II.3). A resposta imune humoral é caracterizada pela degradação de proteínas virais em peptídeos menores, que posteriormente serão apresentados na superfície de células apresentadoras de antígeno (*Antigen Presenting Cell*, APC) em complexos proteicos de proteínas da classe II do HLA (*Human Leukocyte Antigen*). O complexo HLA de classe II-epítipo que fica na superfície das APCs interagem com o receptor CD4+ das células T, que, por sua vez, ativam a expressão de anticorpos específicos contra as moléculas virais. Em contraste com a resposta humoral, a resposta celular resulta na ativação de células T CD8+ não

diferenciadas, diferenciando-as em linfócitos T citotóxicos (*Cytotoxic T Lymphocytes*, CTLs). Immunoproteassomas presentes no citoplasma das células hospedeiras degradam as proteínas virais, e os peptídeos resultantes dessa degradação são transportadas até o retículo endoplasmático via um transportador de antígenos em processamento (*Transporter of Antigen Processing*, TAP), onde são acoplados às moléculas de HLA classe I. O complexo HLA-I-epítipo é então transportado para o Complexo de Golgi e dali para a membrana celular, onde serão reconhecidos por CTLs antivirais (Guermonprez, et al., 2002). Através desse reconhecimento, as CTLs são capazes de matar as células do hospedeiro que estiverem infectadas, limitando o alastramento da infecção.

II.5.1. Resposta imune humoral ao vírus da influenza A H1N1

Pessoas idosas que foram expostas ao vírus da influenza A/H1N1 na década de 1950 apresentaram certa resistência ao vírus da influenza A/H1N1 durante a pandemia de 2009, uma vez que eles tinham desenvolvido anticorpos reativos da primeira exposição, na década de 50 (Yu, et al., 2008; Hancock, et al., 2009). Essa observação sugere que a resposta imune humoral gera uma proteção duradoura mediada por anticorpos contra cepas do vírus da influenza que sejam semelhantes à cepa da infecção original. Na fase humoral da resposta adaptativa, as células B são estimuladas a produzir anticorpos direcionados às proteínas virais HA, NA e M2 (Baumgarth, et al., 2000; Waffarn & Baumgarth, 2011). Anticorpos que são específicos contra a cabeça trimérica da HA se mostraram eficientes no bloqueio da endocitose mediada por receptores do vírus, justamente o mecanismo utilizado pelo vírus para se inserir na célula hospedeira (De Jong, et al., 2003). No entanto, a maioria dos anticorpos é produzido a partir da apresentação de um peptídeo retirado da proteína do vírus invasor, o que significa que vai ser eficiente contra essa cepa, mas possivelmente não a outras cepas do vírus da influenza (Yu, et al., 2008).

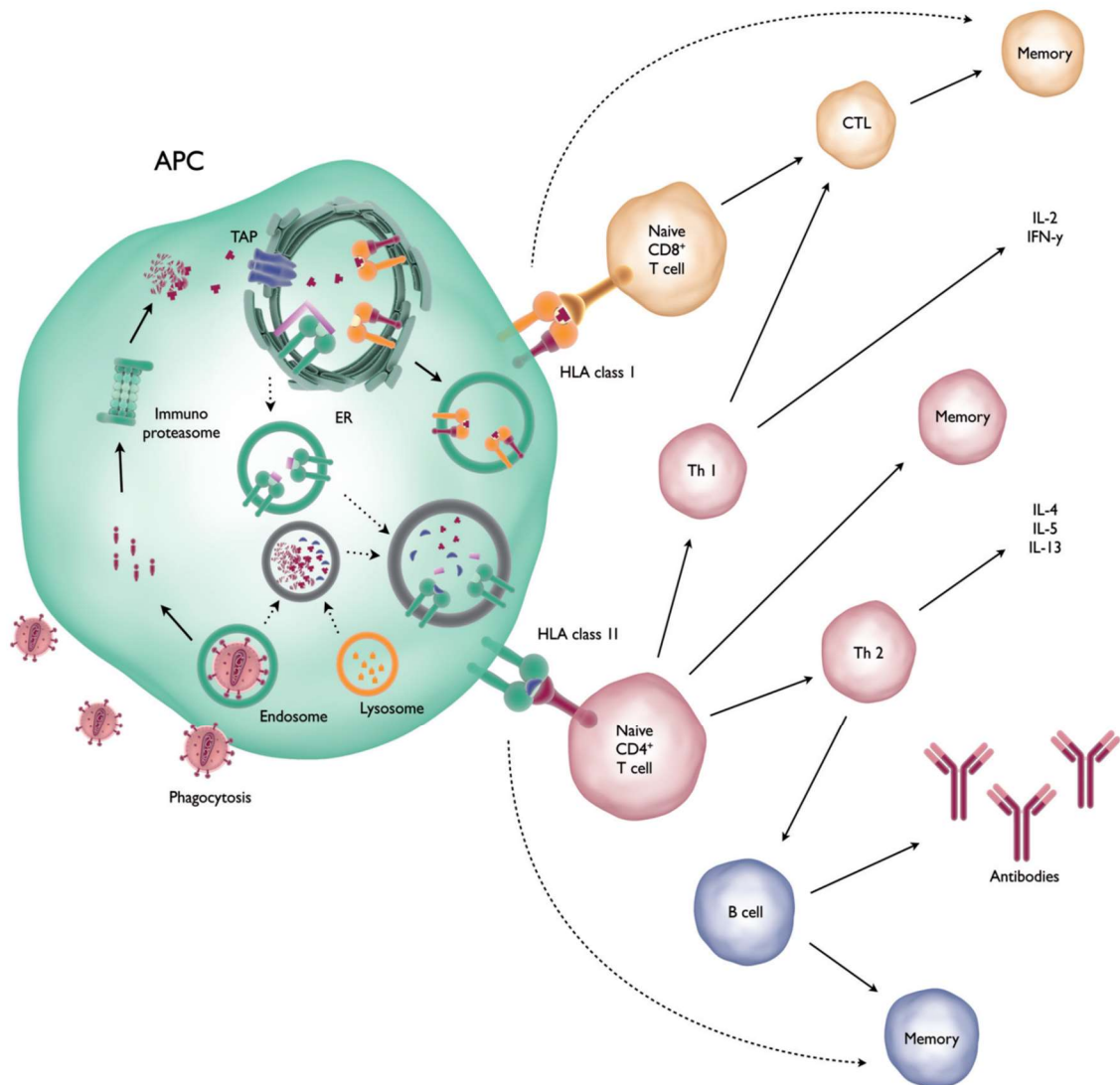


Figura II. 3 Vias de Apresentação antigênica por HLAs e ativação da resposta imune adaptativa.

A neuraminidase (NA) é uma hidrolase de glicosídeo viral que, ao catalizar a hidrólise do ácido siálico terminal tanto de novos vírions quanto da membrana da célula hospedeira, facilita a liberação e o espalhamento das partículas virais recém-formadas (von Itzstein, 2007). Anticorpos anti-NA inibem a atividade enzimática de NA, limitando o alcance da infecção e diminuindo a severidade e a duração da doença (Kilbourne, et al., 2004; Bosch, et al., 2010). Anticorpos específicos anti-NA também facilitam a eliminação do vírus, na medida em que facilitam a citotoxicidade mediada

por célula e depende de anticorpos (*Antibody-Dependent Cell-mediated Cytotoxicity*, ADCC) (van de Sandt, et al., 2012).

A proteína viral transmembrânica e tetramérica M2 funciona como um canal iônico e também é responsável por desenvolver o vírus quando este está dentro do endossomo (Schnell & Chou, 2008). O efeito protetor de anticorpos reativos anti-M2 foi primeiramente demonstrado em ratos após a transferência passiva de anticorpos monoclonais (Kim, et al., 2013), e é atribuído à ativação de ADCC (Mozdzanowska, et al., 1999; El Bakkouri, et al., 2011). O amplo espectro de ação dos anticorpos anti-M2 é provavelmente resultado da alta conservação da proteína M2 entre diferentes cepas de vírus da influenza (Tompkins, et al., 2007; Fiers, et al., 2009; Fu, et al., 2009; Schotsaert, et al., 2009; Wang, et al., 2009; Ebrahimi & Tebianian, 2011).

A resposta imune humoral também inicia a produção de anticorpos contra várias outras proteínas, inclusive contra a altamente conservada proteína viral nucleoproteína - NP (Carragher, et al., 2008). Ainda que esses anticorpos anti-NP não sejam capazes de neutralizar o vírus da influenza, foram encontradas evidências de que eles são capazes de estabelecer uma resposta imune protetora contra esses vírus (Carragher, et al., 2008; Lamere, et al., 2011). Esses mecanismos de proteção podem envolver tanto a ADCC de células infectadas quanto a opsonização da NP pelos anticorpos (Sambhara, et al., 2001; Bodewes, et al., 2010). Apesar das evidências, no entanto, a eficiência desse efeito protetor de anticorpos que não sejam anti-HA ou anti-NA ainda está sob debate.

As imunoglobulinas IgA, IgG e IgM são produzidos a partir de uma infecção primária pelo vírus da influenza, mas as respostas de IgM não são observadas após uma infecção secundária (van de Sandt, et al., 2012). Respostas séricas vírus-específicas de IgA parecem ser indicativas de infecção recente por influenza vírus (Voeten, et al., 1998; Koutsonanos, et al., 2011), enquanto respostas vírus-específicas de IgG se correlacionam com proteção duradoura (Koutsonanos, et al., 2011; Onodera, et al., 2012). Os IgM também podem inativar o vírus da influenza através da ativação do sistema complemento (Jayasekera, et al., 2007; Fernandez Gonzalez, et al., 2008). Além dos anticorpos séricos, a infecção pelo vírus da influenza também

induz a mucosa local do trato respiratório a produzir anticorpos IgA para proteger as células epiteliais da infecção (Onodera, et al., 2012).

II.5.2.1. Resposta imune celular ao vírus da influenza A H1N1 – Células T CD4+

A ativação das células T CD4+ se dá pelo reconhecimento de epítomos virais associados a moléculas de MHC-classe II e sua interação com outras moléculas co-estimulantes presentes na superfície das APCs. Células T CD4+ ingênuas (do inglês, *naïve*, ou não-especializadas) podem se diferenciar em células Th1 ou Th2 (do inglês, *helper*, ou auxiliares, figura II.3). As células Th1 são responsáveis, basicamente, pela regulação da atividade das CTLs (Zhu & Paul, 2010a), e pela indução de células T CD8+ de memória (Riberdy, et al., 2000; Belz, et al., 2002; Delyannis, et al., 2002). Células Th2 promovem a ativação das células B, resultando na produção de anticorpos específicos (Zhu & Paul, 2010b; Okoye & Wilson, 2011). Dois processos diferentes determinam a afinidade dos anticorpos anti-influenza produzidos, aumentando significativamente a eficiência desses anticorpos. São o chamados troca de classe antigênica e a hipermutação somática (Kamperschroer, et al., 2006). As células T CD4+ de memória que foram induzidas por uma primeira infecção pelo vírus da influenza contribuem para uma resposta mais rápida e eficiente no caso de uma segunda infecção (Strutt, et al., 2014), em especial as as células T CD4+ de memória residentes nos pulmões.

II.5.2.2. Resposta imune celular ao vírus da influenza A H1N1 – Células T CD8+

A ativação das células T CD8+ acontece a partir do reconhecimento de epítomos virais associados a moléculas do MHC de classe I presentes na superfície

das células APCs localizadas nos linfonodos. Esse processo leva à diferenciação das células T CD8⁺ em linfócitos T citotóxicos (do inglês *CytoToxic Lymphocytes*, CTLs), que vão migrar para o local da infecção para reconhecer e eliminar as células que estiverem infectadas, prevenindo que novos vírus sejam produzidos e liberados no meio extracelular (Nakanishi, et al., 2009). A interação do receptor de célula T (TCR) com o complexo HLA-classel-epítipo estimula as CTLs a liberarem perforinas e granzimas, induzindo a célula infectada à apoptose (Moffat, et al., 2009; Andrade, 2010). Também são produzidas citocinas pró-inflamatórias, tais como o TNF- α , que têm função de aumentar a função lítica e inibir a replicação do vírus dentro da célula (La Gruta, et al., 2004; La Gruta, et al., 2007; Metkar, et al., 2008; Andrade, 2010; van Domselaar & Bovenschen, 2011). Após a infecção, uma gama de produção de células T CD8⁺ de memória é produzida, constituindo a base para respostas imunes rápidas, fortes e eficientes caso uma segunda infecção ocorra (Zimmermann, et al., 1999; Chang, et al., 2007; Hikono, et al., 2007; DiSpirito & Shen, 2010; van Gisbergen, et al., 2011). As CTLs específicas para vírus são direcionadas principalmente a epítopos resultantes de proteínas altamente conservadas, tais como NP, M1 e PA. Isso significa que essas CTLs possivelmente serão eficientes contra cepas diferentes do vírus da influenza, efeito conhecido como reatividade cruzada (Assarsson, et al., 2008; Kreijtz, et al., 2008; Lee, et al., 2008). Estudos com ratos confirmaram essa hipótese, mostrando que células T CD8⁺ contribuem tanto para a imunidade homo- quanto heterosubtípica (Kreijtz, et al., 2007; Grebe, et al., 2008; Kreijtz, et al., 2009; Hillaire, et al., 2011a; Hillaire, et al., 2011b).

Contudo, as evidências que suportam a proteção imune ao vírus da influenza induzida por CTLs ainda são muito limitadas. A presença de células T de CD4⁺ e CD8⁺ de memória contra o vírus H1N1 da pandemia de 2009 também foi encontrada em indivíduos que, a princípio, não haviam contraído influenza (Sridhar, et al., 2012). Além disso, outras evidências circunstanciais sobre o papel protetor das células T CD8⁺ heterosubtípicas em infecções por influenza surgiram após estudos epidemiológicos: pessoas infectadas por influenza A H1N1 e mostraram sintomas antes da pandemia de 1957 foram parcialmente protegidos quando infectados pela cepa H2N2 pandêmica. Van de Sandt e colaboradores (2012) acabaram chegando a resultados semelhantes quando estudaram infecções isoladas de influenza H5N1.

II.5.3. A variabilidade genética na resposta imune adaptativa

A variabilidade genética nos *loci* HLA é reconhecidamente a maior em qualquer população humana estudada, comparativamente a qualquer outra região gênica (tabela II.4a,b). A hipervariabilidade dos loci HLA é, de fato, o mecanismo básico utilizado pelo hospedeiro para garantir o reconhecimento de um maior número de moléculas exógenas e potencialmente perigosas, iniciando a ativação da resposta imune contra essas moléculas (Tabela II.4a,b). O reconhecimento do complexo HLA-epítipo pelos receptores de célula T é crucial para a ativação da resposta imune adaptativa, e com esse grau de variabilidades dos genes HLA, é muito difícil associar mutações específicas a respostas diferentes à infecção pelo vírus da influenza. Normalmente, considera-se alelos inteiros ou haplótipos completos para tentar chegar a uma associação, já que esses blocos grandes são provavelmente mais propensos a influenciar diretamente os efeitos fenotípicos do que mutações isoladas. Mutações nessas regiões haplotípicas podem tanto atrapalhar a formação dos complexos HLA-epítopos (Price, et al., 2000; Voeten, et al., 2000; Berkhoff, et al., 2004; Rimmelzwaan, et al., 2004; Berkhoff, et al., 2007a), ou fazer com que os epítopos consigam escapar da detecção pelos TCRs, já que o epítipo deixa de ser complementar à nova especificidade do receptor de célula T (Price, et al., 2000; Boon, et al., 2002; Berkhoff, et al., 2007a; Berkhoff, et al., 2007b).

Tabela II. 4a Variabilidade genética dos três principais loci gênicos de genes HLA de classe I clássicos (dados retirados do bando de dados 1000 Genomes Project)

Tipo de variante	HLA-A	HLA-B	HLA-C	Total
Sítio de splice essencial	34	10	12	56
Ganho de parada	71	26	23	120
Codante com frameshift	577	127	94	798
Codante não-sinônima	1232	382	857	2471
Sítio de splice	206	28	48	282
Codante sinônima	486	148	339	973
Intrônica	1519	557	1230	3306
5'UTR	74	28	8	110
3'UTR	909	191	268	1368
Upstream	60	14	34	108
Downstream	70	60	59	189
Total de variantes	5238	1571	2972	9781

*para descrição das variantes, ver tabela II.1.

Tabela II. 4b Variabilidade genética das cadeias beta dos HLAs de classe II, nos três principais loci (retirado do banco de dados 1000 Genomes Project).

Tipo de variante	HLA-DPB1	HLA-DQB1	HLA-DQB2	HLA-DRB1	HLA-DRB3	HLA-DRB5	Total
Sítio de splice essencial	2	27	0	2	3	5	39
Ganho de parada	12	42	0	4	1	11	70
Codante com frameshift	0	150	0	35	3	24	212
Codante não-sinônima	176	1071	18	166	89	150	1670
Sítio de splice	10	138	4	9	2	2	165
Codante sinônima	69	491	33	90	38	62	783
Intrônica	192	3099	181	267	142	191	4072
5'UTR	3	60	1	16	13	4	97
3'UTR	117	633	48	113	86	130	1127
Upstream	7	100	3	20	15	11	156
Downstream	9	141	4	28	17	6	205
Total de variantes	597	5952	292	750	409	596	8501

*para descrição das variantes, ver tabela II.1.

Mutações no próprio epítipo viral mostraram ter um efeito direto na eficiência da resposta imune mediada por CTLs (van de Sandt, et al., 2012). Esse tipo de mutações foram observadas em vírus mutantes que infectam seus hospedeiros de uma forma crônica, como o HIV-1, capazes de escapar do sistema imune (Huet, et al., 1990; Cale, et al., 2011). Substituições dos epítipos reconhecidos pelas CTLs

também foram observadas durante a evolução do vírus da influenza periódico A/H3N2 (Voeten, et al., 2000; Rimmelzwaan, et al., 2004; Berkhoff, et al., 2007a). A substituição R384G, localizada no epítopo NP383-391 do HLA-B*2705-restrito (Voeten, et al., 2000) afeta consideravelmente a resposta específica de CTLs contra o vírus da influenza (Berkhoff, et al., 2007a), enquanto o epítopo NP418-426 no HLA-B*3501-restrito (Boon, et al., 2002; Berkhoff, et al., 2007b) apresenta sinais de deriva antigênica (Boon, et al., 2004), o que explicaria a reatividade cruzada de CTLs contra vírus contemporâneos que tenham evoluído de cepas históricas (Gras, et al., 2010). No entanto, cabe ressaltar que essas variações nos epítomos dos vírus pode ser considerada restrita à função. Por exemplo, o epítopo M1 58-66 é altamente conservado apesar de sua natureza imunodominantes, provavelmente porque mutações nessa região específica devem comprometer, de alguma forma, o *fitness* (p. ex. patogenicidade, capacidade de ligação aos receptores, replicabilidade, infectividade, e outros) do vírus (Berkhoff, et al., 2006).

Com relação aos genes do hospedeiro, é interessante notar que o gene mais variável dos não HLA estudados aqui (NLRP3 com 496 variantes no total) ainda é muito menos variável do que a média de variabilidade para os nove genes HLA, de acordo com os dados coletados no banco de sequências e variantes do 1000 Genomes Project. O gene mais conservado do sistema HLA, HLA-DQB2, tem 292 variantes descritas. O gene HLA-B, por outro lado, apresentou mais 1535 variantes (mas, ainda assim, menor do que o número de alelos funcionais descritos para esse gene, de acordo com o site do IPD-TGTM/HLA, que, em maio de 2014, registrava 3455 alelos que codificam 2577 proteínas descritas). Essa quantidade tão grande variabilidade reforça a importância da função desses genes na resposta imune, já que essas proteínas deverão ser capazes de reconhecer e se ligar a uma plethora de peptídeos possíveis (os epítomos) para que sejam apresentados aos linfócitos T citotóxicos e de memória. Quanto mais universal a molécula de HLA, mais eficiente será a apresentação dos peptídeos e a atividade da resposta imune adaptativa.

II.6. Considerações finais e conclusão

A motivação inicial para esta revisão foi o trabalho publicado por Horby e colaboradores (2012), onde os autores encerram o artigo com uma pergunta, a priori não respondida:

“Is susceptibility to severe influenza in humans heritable? – hard to answer, but not due to a lack of genotyping of analytic tools, nor because of insufficient evidence from severe influenza cases, but because of the absence of a coordinated effort to define and assemble cohorts of cases.” (Horby, et al., 2012)

Aqui apresentou-se uma visão geral atualizada da variabilidade dos genes do hospedeiro que possam, de alguma forma, estar envolvidos no processo de infecção do vírus da influenza, seja interagindo diretamente com o vírus ou na ativação de sistemas imunes. Além disso, discutiu-se como essa variabilidade pode influenciar a severidade da infecção, mostrando que mutações-chave podem ser a diferença entre a sobrevivência e a morte de indivíduos que recebam exatamente o mesmo tratamento. Toda interação entre o vírus e o hospedeiro se inicia a partir da entrada do vírus no organismo, e termina ou com a morte do indivíduo ou com a eliminação do vírus pelo sistema imune. A evolução da relação vírus hospedeiro é influenciada pelas variantes dos genes do hospedeiro, várias delas diretamente associadas a um bom ou mau prognóstico da infecção, aumentando a resistência e a resiliência do paciente contra subtipos específicos de epítomos virais. Essas variantes são polimórficas nas diferentes populações humanas, mesmo nos casos de infecções brandas e assintomáticas, e esse polimorfismo pode estar relacionado com as diferentes taxas de morbidade/mortalidade que resultam da infecção pelo vírus da influenza A. Considerando a revisão mostrada aqui e os dados coletados no 1000 Genomes Project, a resposta à pergunta feita por Horby e colaboradores (2012) é positiva, a susceptibilidade a infecção severa pelo vírus da influenza A é hereditária.

Por outro lado, isso não significa que todas as causas e variabilidade na sintomatologia e morbidade das gripes tenham sido descobertas ou possam ser atribuídas exclusivamente ao perfil genético do indivíduo, mas, claramente, há genes humanos que tem um impacto direto no curso da infecção. Em particular, quando se considerar a variabilidade genética associada a grupos étnicos, não se pode deixar de levar em conta os aspectos específicos de habitat e clima do local onde essas populações habitam. Em regiões temperadas existe um pico de infecções que acontece predominantemente nos meses de inverno e as epidemias acontecem com periodicidade, acompanhando as mudanças de estação. Experimentos com cepas de influenza adaptadas para ratos mostrou uma taxa de transmissão do vírus de 58,2% nos meses de inverno, contra 34,1% no verão (Schulman & Kilbourne, 1963). Já foi hipotetizado que essa variação periódica na infectividade do vírus da influenza esteja associada a a flutuações na eficiência do sistema imune do hospedeiro, como no caso das taxas de melatonina (Dowell, 2001) e vitamina D (Cannell, et al., 2006); nas mudanças de rotina do hospedeiro (frequência escolar, procura maior por ambientes fechados e com pouca circulação de ar, viagens aéreas, etc); e especialmente fatores ambientais, tais como a temperatura (Eccles, 2002), umidade relativa e a direção de ventos da estação na parte superior da atmosfera (Hammond, et al., 1989). Em regiões tropicais, as infecções por influenza se mantém mais ou menos constantes durante o ano inteiro, com um pequeno acréscimo de incidência nas estações chuvosas (Viboud et al., 2006, Shek & Lee, 2003, OMS). Florestas tropicais da África equatorial, a região amazônica do Brasil, Venezuela e Equador, são regiões potencialmente promissoras para esforços científicos no sentido de conectar possíveis fatores genéticos étnicos em comunidades isoladas e os padrões de prevalência e infectividade do vírus da influenza A. Na região do sudeste asiático, em particular em países que abrangem latitudes subtropicais (ao sul) até montanhosos (ao norte) passando por diversos tipos de clima, é um local de surgimento de novas cepas do vírus da influenza (Le, et al., 2013), e um campo de estudo filogeográfico dessas cepas (Wallace & Fitch, 2008).

Além disso, muitos outros fatores podem causar modificações no DNA e na forma como ele é expresso (alterações epigenéticas). Estudos suportam a teoria de programação epigenética pré-natal/infantil como origem para várias doenças do adulto

(Ahmed, 2010; Nicoletto & Rinaldi, 2011). Por exemplo, Yang e Huffman (2013) mostraram que indivíduos que sofreram algum tipo de privação nutricional durante a gestação tendem a ser adultos obesos. Pode ser pensado que se o padrão de metilação estabelecido durante os primeiros momentos de vida de um indivíduo pode afetar a forma como esse indivíduo responde à dieta enquanto adulto, esse mesmo padrão de metilação pode também afetar a forma como os genes relacionados ao sistema imune serão expressos durante processos infecciosos e inflamatórios (Wang, et al., 2013).

O trabalho de Guihot e colaboradores (2014) evidenciou um curioso caso que acabou trazendo mais perguntas do que respostas à discussão sobre a herdabilidade da susceptibilidade a formas severas da infecção pelo vírus da influenza. Os autores encontraram marcadores biológicos que podem prever se um indivíduo é mais ou menos propenso a sofrer consequências letais derivadas da infecção. Contudo, os mecanismos genéticos que influenciam esses marcadores biológicos ainda permanecem desconhecidos, já que os marcadores biológicos eram baseados na dinâmica de anticorpos presentes no pulmão: em casos letais, observaram um “sequestro” de anticorpos nos pulmões, enquanto que nos casos brandos e não-letais os anticorpos permanecem livres e circulantes no sangue. Worobey e colaboradores (2014) hipotetizaram que a alta mortalidade de adultos entre 20 e 40 anos durante a epidemia de gripe espanhola de 1918 deveria ser resultado de um encontro prévio desses indivíduos com um vírus supostamente heterosubtípico H3N8, que aparentemente circulou pelo globo entre 1889 e 1900. Sendo assim, outras faixas etárias estariam parcialmente protegidos da gripe espanhola por terem sido expostos a outros vírus da influenza N1 ou H1-relacionados, mas esses adultos que foram expostos durante a infância ao H3N8 não teriam desenvolvido memória imunitária contra o vírus H1N1 da gripe espanhola. Apesar dessa hipótese fazer bastante sentido, ela nos leva a outras perguntas imediatas: Que outros mecanismos imunológicos já descritos podem ser utilizados para explicar esse aumento discrepante na mortalidade de adultos relacionada à gripe espanhola de 1918? Para responder essa pergunta, talvez fosse necessário comparar os adultos que contraíram a gripe espanhola e que já tivessem sido expostos ao vírus H3N8 com crianças potencialmente desprovidas de imunidade (por não terem sido expostas) a todos os

tipos de HA e NA, para buscar cascatas de ativação imunológica que pudessem explicar essa inativação do sistema imune pelos tipos H3. Outra pergunta que pode ser levantada é: será que a infecção pelo subtipo H3N8 não causou alterações epigenéticas nos indivíduos expostos, que acabaram por dificultar a recuperação e a eliminação dos vírus H1N1 quando esses indivíduos foram infectados por esse subtipo? E será que a variabilidade genética dos próprios indivíduos também não pode ter afetado a forma como essa reprogramação epigenética ocorreu, mostrando perfis epigenéticos diferentes dependendo do indivíduo? Além dessa possibilidade epigenética, Worobey e colaboradores (2014) sugeriram um mecanismo parecido com o pecado original antigênico (do inglês, *Original Antigenic Sin* – OAS), inicialmente descrito por Francis (1960), que diz que os anticorpos de infecções subsequentes de influenza serão ditados pela primeira produção de anticorpos, a partir do encontro com o primeiro vírus da vida daquele indivíduo. Portanto, indivíduos que foram inicialmente expostos ao H3N8, terão seus anticorpos produzidos, majoritariamente e preferencialmente, a partir desse tipo de vírus, mostrando pouca ou nenhuma imunidade a vírus de cepas muito diferentes, como é o caso do H1N1.

Contudo, essa hipótese ainda carrega muitas dúvidas, principalmente sobre qual é o papel que o genoma do indivíduo, seu sistema imune e os sistemas moleculares interagem para criar essa memória preferencial, em detrimento de outras. É óbvio que muitos estudos nessa linha e que explorem a variabilidade genética humana no contexto de doenças infecciosas (em especial a influenza) sejam necessários para que as perguntas mencionadas acima sequer fiquem próximas de serem respondidas. Nesse sentido, esse capítulo teve o objetivo de começar a explorar esses ângulos, até então estudados de uma forma tão segmentada e compartimentalizada, para que a comunidade científica possa olhar as relações genéticas entre patógeno e hospedeiro de uma forma mais completa.

Considerações finais

O trabalho aqui apresentado comparou duas diferentes classes de marcadores genéticos quanto a localização genômica – regiões codificadoras e não codificadoras – buscando melhor compreender a distribuição da variabilidade genética humana. Classificou-se como uma situação genérica de saúde as regiões supostamente neutras e, as relacionadas à doença potencialmente sob seleção. Observou-se que 8% dos marcadores considerados neutros apresentam sinais de seleção, taxa essa baixa e que não os descaracteriza como um conjunto de marcadores neutros.

A variabilidade genética entre grupos de diferentes origens geográficas para marcadores neutros é maior do que a variabilidade genética para marcadores associados à infecção pelo vírus da influenza. Possivelmente, isso se dá pela importância da manutenção de sequências não deletérias ou que diminuem a susceptibilidade a infecções de uma forma global, uma vez que indivíduos que apresentem mutações associadas a prognósticos ruins da infecção têm menor chance de sobrevivência quando da manifestação do quadro clínico.

Sendo assim, observa-se uma maior variabilidade genética quando se considera marcadores neutros do que os em regiões codificadoras. Portanto, a maior variabilidade entre populações de origem geográficas distintas, aqui descrita, foi observada na saúde (painel de 100 inserções *Alu*) e a menor na doença (marcadores genéticos associados a bom ou mau prognóstico da infecção pelo vírus da influenza), resultado confirmatório da expectativa de seleção contra boa parte das variações em regiões codificadoras e da manutenção da variabilidade em regiões neutras do genoma.

Os marcadores genéticos neutros foram suficientes para a diferenciação de grupos de amostras quanto às ancestralidades geográficas. Exceção a isso foram as populações miscigenadas, isto é, as brasileiras usadas como modelo, que tenderam a se agrupar a populações que representassem o grupo com maior contribuição genética daquela população, sem que fosse possível traçar um perfil genético único para as duas populações brasileiras. Já a variabilidade genética entre

populações encontrada para os marcadores genéticos relacionados à susceptibilidade à infecção pelo vírus da influenza mostrou-se bastante homogênea entre os grupos populacionais. Quando capaz de diferenciar grupos, no entanto, estes eram também congruentes com a ancestralidade genética das amostras (africanos x não-africanos; asiáticos x não-asiáticos). Dessa forma, entender a ancestralidade genética de um grupo populacional facilitaria entender a susceptibilidade a doenças infecciosas, isto é, as duas classes de marcadores se complementam na busca de um melhor entendimento da transição entre a saúde e a doença.

Referências

- Abbas, Y. M. et al., 2013. Structural basis for viral 5'-PPP-RNA recognition by human IFIT proteins. *Nature*, Volume 494, pp. 60-64.
- Acosta-Rodriguez, E. V., Napolitani, G., Lanzavecchia, A. & Sallusto, F., 2007. Interleukins 1beta and 6 but not transforming growth factor-beta are essential for the differentiation of interleukin 17-producing human T helper cells. *Nat Immunol*, Volume 8, pp. 942-949.
- Ahmed, F., 2010. Epigenetics: Tales of adversity. *Nature*, Volume 468, p. S20.
- Allen, I. C. et al., 2009. The NLRP3 inflammasome mediates in vivo innate immunity to influenza A virus through recognition of viral RNA. *Immunity*, Volume 30, pp. 556-565.
- Andrade, F., 2010. Non-cytotoxic antiviral activities of granzymes in the context of the immune antiviral state. *Immunol Rev*, Volume 235, pp. 128-146.
- Andrewes, C., 1939. Immunity in influenza: the bearing of recent research work: (section of epidemiology and state medicine). *Proc R Soc Med*, pp. 145-152.
- Arcanjo, A. C., 2012. *Dinâmica de marcadores genéticos na região do Complexo de Histocompatibilidade Principal Humano em populações do Centro-Oeste do Brasil*. Dissertação de Mestrado ed. Brasília: Universidade de Brasília.
- Assarsson, E. et al., 2008. Immunomic analysis of the repertoire of T-cell specificities for influenza A virus in humans. *J Virol*, Volume 82, pp. 12241-12251.
- Astle, W. & Balding, D. J., 2009. Population structure and cryptic relatedness in genetic association studies. *Statistical Science*, pp. 451-471.
- Bamford, S. et al., 2004. The COSMIC(Catalogue of Somatic Mutations In Cancer) database and website. *Brittish Journal of Cancer*, pp. 355-358.
- Bamshad, M. J. et al., 2003. Human Population Genetic Structure and Inference of Group Membership. *Am. J. Hum. Genet*, Volume 72, pp. 578-589.
- Bamshad, M. et al., 2001. Genetic Evidence on the Origins of Indian Caste Populations. *Genome Research*, pp. 994-1004.
- Barbujani, G. & Colonna, V., 2010. Human genomic diversity: frequently asked questions. *Trends in Genetics*, Volume 26, pp. 285-295.
- Barcelos, R. S. S., 2006. *Contribuição genética de duas populações urbanas da região Centro-Oeste brasileira estimada por marcadores uniparentais*. Tese de Doutorado ed. Brasília: Universidade de Brasília.
- Bateman, A. C. et al., 2008. Amino acid 226 in the hemagglutinin of H4N6 influenza virus determines binding affinity for alpha2,6-linked sialic acid and infectivity levels in primary swine and human epithelial cells. *J Virol*, Volume 82, pp. 8204-8209.
- Batzer, M. A. & Deininger, P. L., 1991. A human-specific subfamily of Alu sequences. *Genomics*, pp. 481-487.

- Batzer, M. A. et al., 1994. African origin of human-specific Alu insertions. *PNAS*, pp. 12288-12292.
- Baumgarth, N. et al., 2000. B-1 and B-2 cell-derived immunoglobulin M antibodies are non redundant components of the protective response to influenza virus infection. *J Exp Med*, Volume 192, pp. 271-280.
- Behar, D M; Harmant, C; Manry, J; van Oven, M; Haak, W; Martinez-Cruz, B; Salaberria, J; Oyharçabal, B; Bauduer, F; Comas, D; Quintana-Murci, L; The Genographic Consortium, 2012. The Basque Paradigm: Genetic evidence of a maternal continuity in the Franco-Cantabrian region since Pre-Neolithic times. *The American Journal of Human Genetics*, pp. 486-493.
- Belz, G. T. et al., 2002. Compromised influenza virus-specific CD8(+)-T-cell memory in CD4(+)-T-cell-deficient mice. *J Virol*, Volume 76, pp. 12388-12393.
- Berkhoff, E. G. et al., 2004. A mutation in the HLA-B*2705-restricted NP383-391 epitope affects the human influenza A virus-specific cytotoxic T-lymphocyte response in vitro. *J Virol*, Volume 78, pp. 5216-5222.
- Berkhoff, E. G. et al., 2006. Fitness costs limit escape from cytotoxic T lymphocytes by influenza A viruses. *Vaccine*, Volume 24, pp. 6594-6596.
- Berkhoff, E. G. et al., 2007a. Assessment of the extent of variation in influenza A virus cytotoxic T-lymphocyte epitopes by using virus-specific CD8+ T-cell clones. *J Gen Virol*, Volume 88, pp. 53560-53566.
- Berkhoff, E. G. et al., 2007b. The loss of immunodominant epitopes affects interferon-gamma production and lytic activity of the human influenza virus-specific cytotoxic T lymphocyte response in vitro. *Clin Exp Immunol*, Volume 148, pp. 296-306.
- Bodewes, R., Osterhaus, A. D. & Rimmelzwaan, G. F., 2010. Targets for the induction of protective immunity against influenza A viruses. *Viruses*, Volume 2, pp. 166-188.
- Bonilla, C. et al., 2004. Admixture in the Hispanics of San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann Hum Genet*, pp. 139-153.
- Boon, A. C. et al., 2002. Sequence variation in a newly identified HLA-B35-restricted epitope in the influenza A virus nucleoprotein associated with escape from cytotoxic T lymphocytes. *J Virol*, Volume 76, pp. 2567-2572.
- Boon, A. C. et al., 2004. Recognition of homo- and heterosubtypic variants of influenza A viruses by human CD8+ T lymphocytes. *J Immunol*, Volume 172, pp. 2453-2460.
- Bosch, B. J. et al., 2010. Recombinant soluble, multimeric HA and NA exhibit distinctive types of protection against pandemic swine-origin 2009 A(H1N1) influenza virus infection in ferrets. *J Virol*, Volume 84, pp. 10366-10374.
- Budowle, B. et al., 1999. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans and Trinidadians. *J Forensic Sci*, pp. 1277-1286.
- Cale, E. M. et al., 2011. Mutations in a dominant Nef epitope of simian immunodeficiency virus diminish TCR-Epitope peptide affinity but not epitope peptide-MHC class I binding. *J Immunol*, Volume 187, pp. 3300-3313.

- Cannell, J. J. et al., 2006. Epidemic influenza and vitamin D. *Epidemiol Infect*, Volume 134, pp. 1129-1140.
- Carette, J. E. et al., 2009. Haploid genetic screens in human cells identify host factors used by pathogens. *Science*, Volume 326, pp. 1231-1235.
- Carragher, D. M. et al., 2008. A novel role for non-neutralizing antibodies against nucleoprotein in facilitating resistance to influenza virus. *J Immunol*, Volume 181, pp. 4168-4176.
- Carter, A. B. et al., 2004. Genome-wide analysis of the human Alu Yb-lineage. *Human Genomics*, pp. 1-13.
- Cerezo, M. et al., 2012. Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Research*, pp. 821-826.
- Chandrasekaran, A. et al., 2008. Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nature Biotechnology*, Volume 26, pp. 107-113.
- Chang, J. T. et al., 2007. Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science*, Volume 315, pp. 1687-1691.
- Clarke, L. et al., 2012. The 1000 Genomes Project: data management and community access. *Nature Methods*, Volume 9, pp. 1-4.
- Cohen, M. & Varki, A., 2010. The Sialome - Far more than the sum of its parts. *OMICS J Int Biol*, Volume 14, pp. 455-464.
- Cohen, M. et al., 2013. Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virology*, Volume 10, p. 321.
- Cone, R. A., 2009. Barrier properties of mucus. *Adv Drug Deliv Rev*, Volume 61, pp. 75-85.
- Consortium, T. 1. G. P., 2016. *1000 Genomes Project*. [Online] Available at: <http://www.1000genomes.org> [Acesso em 2014].
- Cordeaux, R., Hedges, D. J., Herke, S. W. & Batzer, M. A., 2006. Estimating the retrotransposition rate of human Alu elements. *Gene*, pp. 134-137.
- Das, S. R. et al., 2010. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. *PNAS Plus*, Volume early edition, pp. 1-6.
- Dausset, J. et al., 1990. Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative genetic mapping of the human genome. *Genomics*, pp. 575-577.
- Dawood, F. S. et al., 2012. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *Lancet Infect Dis*, Volume 12, pp. 687-695.
- Dawood, F. S. et al., 2012. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *Lancet Infect Dis*, Volume 12, pp. 687-695.

- De Jong, J. C. et al., 2003. Haemagglutination-inhibiting antibody to influenza virus. *Dev Biologicals*, Volume 115, pp. 63-73.
- Delyannis, G. et al., 2002. Induction of long-term memory CD8(+) T cells for recall of viral clearing responses against influenza virus. *J Virol*, Volume 76, pp. 4212-4221.
- Dillehay, T. D., 2003. Tracking the first Americans. *Nature*, pp. 23-24.
- DiSpirito, J. R. & Shen, H., 2010. Quick to remember, slow to forget: Rapid recall responses of memory CD8+ T cells. *Cell Res*, Volume 20, pp. 13-23.
- Dovichi, N. J., 1997. DNA sequencing by capillary electrophoresis. *Electrophoresis*, pp. 2392-2399.
- Dowell, S. F., 2001. Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerg Infect Dis*, Volume 7, pp. 369-374.
- Duez, J. M., Sixt, N. & Péchinot, A., 2008. Influenza virus infection: don't forget the role of the mucociliary system!. *J Antimicrobial Chemotherapy*, Volume 63, pp. 421-422.
- Ebrahimi, S. M. & Tebianian, M., 2011. Influenza A viruses: why focusing on M2e-based universal vaccines. *Virus Genes*, Volume 42, pp. 1-8.
- Eccles, R., 2002. An explanation for the seasonality of acute upper respiratory tract viral infections. *Acta Otolaryngol*, Volume 122, pp. 183-191.
- El Bakkouri, K. et al., 2011. Universal vaccine based on ectodomain of matrix protein 2 of influenza A: Fc receptors and alveolar macrophages mediate protection. *J Immunol*, Volume 186, pp. 1022-1031.
- El Moussi, A., Kacem, A. B. H. & Slim, A., 2014. Loss and gain of N-linked glycosylation sites in globular head and stem of HA found in A/H3N2 flu fatal and severe cases during 2013 Tunisia flu seasonal survey. *Virus Genes*, Volume 48, pp. 189-192.
- Espósito, S. et al., 2012. Toll-like receptor 3 gene polymorphisms and severity of pandemic A/H1N1/2009 influenza in otherwise healthy children. *Viral J*, Volume 9, p. 270.
- Excoffier, L. & Lischer, H. E. L., 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res*, Volume 10, pp. 564-567.
- Excoffier, L., Smouse, P. E. & Quattro, J. M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application from human mitochondrial DNA restriction data. *Genetics*, 131(2), pp. 479-491.
- Fernandez Gonzalez, S., Jayasekera, J. P. & Carroll, M. C., 2008. Complement and natural antibody are required in the long-term memory response to influenza virus. *Vaccine*, Volume 26, pp. 186-193.
- Fiers, W. et al., 2009. M2e-based universal influenza A vaccine. *Vaccine*, Volume 27, pp. 6280-6283.
- Firth, H. V. et al., 2009. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensemble Resources. *Am J Hum Genet*, pp. 524-533.
- Fleischer, B., 1998. Mechanism of glycosylation in the Golgi Apparatus. *The J of Histochem and Cytochem*, Volume 31, pp. 1033-1040.

- Francis, T., 1960. On the doctrine of original antigenic sin. *Proc Am Philos Soc*, Volume 104, pp. 572-578.
- Fritz, J. H., Ferrero, R. L., Philpott, D. J. & Girardin, S. E., 2006. Nod-like proteins in immunity, inflammation and disease. *Nature Immun*, Volume 7, pp. 1250-1257.
- Fu, T. M. et al., 2009. Characterizations of four monoclonal antibodies against M2 protein ectodomain of influenza A virus. *Virology*, Volume 385, pp. 218-226.
- Gack, M. U., 2014. Mechanisms of RIG-I-like receptor activation and manipulation by viral pathogens. *J Virol*, Volume 88, pp. 5213-5216.
- Gambaryan, A. et al., 2006. Evolution of the receptor binding phenotype of influenza A (H5) viruses. *Virology*, Volume 344, pp. 432-438.
- García-Sastre, A., 2004. Identification and characterization of viral antagonists of type I interferon in negative-strand RNA viruses. *Curr Top Microbiol Immunol*, Volume 283, pp. 249-280.
- Garrigan, D. et al., 2007. Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. *Genetics*, Volume 177, pp. 2195-2207.
- Gill, P., Jeffreys, A. J. & Werrett, D. J., 1985. Forensic application of DNA 'fingerprints'. *Nature*, pp. 577-579.
- Glaser, L., Conenello, G., Paulson, J. & Palese, P., 2007. Effective replication of human influenza viruses in mice lacking a major alpha2,6 sialyltransferase. *Virus Research*, Volume 126, pp. 9-18.
- Godinho, N., 2008. *O impacto das migrações na constituição genética de populações latino-americanas*. Brasília: Programa de Pós Graduação em Biologia Animal.
- Godinho, N. M. O. et al., 2009. Regional patterns of genetics admixture in South America. *Forensic Science International: Genetics Supplementary Series*, pp. 329-330.
- Gontijo, C. C. et al., 2014. Brazilian quilombos: a repository of Amerindian alleles. *Am J Hum Biol*, pp. 142-150.
- Gourraud, P. A. et al., 2006. The dbMHC microsatellite portal: a public resource for the storage and display of MHC microsatellite information. *Tissue Antigens*, pp. 395-401.
- Gras, S. et al., 2010. Cross-reactive CD8+ T-cell immunity between the pandemic H1N1-2009 and H1N1-1918 influenza A viruses. *PNAS USA*, Volume 107, pp. 12599-12604.
- Grebe, K. M., Yewdell, J. W. & Bennink, J. R., 2008. Heterosubtypic immunity to influenza A virus: Where do we stand?. *Microb Infect*, Volume 10, pp. 1024-1029.
- Gross, O. et al., 2012. Inflammasome activators induce interleukin-1alpha secretion via distinct pathways with differential requirement for the protease function of caspase-1. *Immunity*, Volume 36, pp. 388-400.
- Grotenhuis, A. J. et al., 2016. Independent replication of published germline polymorphisms associated with urinary bladder cancer prognosis and treatment response. *Bladder Cancer* 2, pp. 77-89.
- Guan, Y. et al., 2010. human TLRs 10 and 1 share common mechanisms of innate immune sensing but not signaling. *J Immunol*, Volume 184, pp. 5094-5103.

- Guarnaccia, T. et al., 2013. Antigenic drift of the pandemic 2009 A(H1N1) influenza virus in a ferret model. *PLOS Pathog*, Volume 9, p. e1003354.
- Guermonprez, P. et al., 2002. Antigen presentation and T cell stimulation by dendritic cells. *Annu Rev Immunol*, Volume 20, pp. 621-667.
- Guihot, A. et al., 2014. Low titers of serum antibodies inhibiting hemagglutination predict fatal fulminant influenza A (H1N1) 2009 infection. *American Journal of Respiratory and Critical Care Medicine*, 189(10), pp. 1240-1249.
- Günther, T. et al., 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *PNAS*, pp. 11917-11922.
- Guo, S. & Thomson, E., 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, Volume 48, pp. 361-372.
- Guo, Z. et al., 2007. NS1 protein of influenza A virus inhibits the function of intracytoplasmic pathogen sensor. *Am J Respir Cell Mol Biol*, Volume 23, pp. 263-269.
- Hale, B. G., Albrecht, R. A. & Garcia-Sastre, A., 2010. Innate immune evasion strategies of influenza viruses. *Future Microbiol*, Volume 5, pp. 23-41.
- Hammond, G. W., Raddatz, R. L. & Gelskey, D. E., 1989. Impact of atmospheric dispersion and transport of viral aerosols on the epidemiology of influenza. *Rev Infect Dis*, Volume 11, pp. 494-497.
- Hancock, K. et al., 2009. Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. *New Engl J Med*, Volume 361, pp. 1945-1952.
- Harduin-Lepers, A., Mollicone, R., Delannoy, P. & Oriol, R., 2005. The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach. *Glycobiology*, Volume 15, pp. 805-817.
- Harduin-Lepers, A. et al., 2001. The human sialyltransferase family. *Biochimie*, Volume 83, pp. 727-737.
- Hartl, D. L. & Clark, A. G., 2010. *Princípios de Genética de Populações*. 4ª ed. s.l.:Artmed.
- Hatta, M., Gao, P., Halfmann, P. & Hawaoka, Y., 2001. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science*, Volume 293, pp. 1840-1842.
- Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y., 2001. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science*, Volume 293, pp. 1840-1842.
- Heil, F. et al., 2004. Species-specific recognition of single-stranded RNA via Toll-like receptors 7 and 8. *Science*, Volume 303, pp. 1526-1529.
- Henn, A. D. et al., 2013. High-resolution temporal response patterns to influenza vaccine reveal a distinct human plasma cell gene signature. *Scientific Reports*, Volume 3, p. 2327.
- Hertz, T. et al., 2013. HLA targeting efficiency correlates with human T-cell response magnitude and with mortality from influenza A infection. *PNAS*, Volume 110, pp. 13492-13497.
- Hikono, H. et al., 2007. Activation phenotype, rather than central or effector-memory phenotype, predicts the recall efficacy of memory CD8+ T cells. *J Exp Med*, Volume 204, pp. 1625-1636.

- Hillaire, M. L., Osterhaus, A. D. & Rimmelzwaan, G. F., 2011b. Induction of virus-specific cytotoxic T lymphocytes as a basis for the development of broadly protective influenza vaccines. *J Biomed Biotechnol*, Volume 2011, p. 939860.
- Hillaire, M. L. et al., 2011a. Cross-protective immunity against influenza pH1N1 2009 viruses induced by seasonal influenza A (H3N2) virus is mediated by virus-specific T-cells. *J Gen Virol*, Volume 92, pp. 2339-2349.
- Horby, P., Nguyen, N. Y., Dunstan, S. J. & Baillie, J. K., 2012. The role of the host genetics in susceptibility to influenza: a systematic review. *PLOS ONE*, 7(3), p. e33180.
- Huber, W., Carey, V. J., Gentleman, R. & Morgan, M., 2015. Orchestrating high-throuput genomic analysis with Bioconductor. *Nature Methods*, Volume 12, p. 115.
- Huet, S. et al., 1990. Structural homologies between two HLA B27-restricted peptides suggest residues important for interaction with HLA B27. *Int Immunol*, Volume 2, pp. 311-316.
- Ichinohe, T. et al., 2009. Inflammasome recognition of influenza virus is essential for adaptive immune responses. *J Exp Med*, Volume 206, pp. 79-87.
- International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. *Nature*, pp. 931-945.
- Jagadeeswaran, P., Forget, B. G. & Weissman, S. M., 1981. Short interspersed repetitive DNA elements in Eucaryotes: transposable DNA elements generated by reverse transcription of RNA Pol III transcripts?. *Cell*, 26(2), pp. 141-142.
- Jakobsson, M. et al., 2008. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature*, pp. 998-1003.
- Jayaraman, A. et al., 2012. Glycosylation at Asn91 of H1N1 haemagglutinin affects binding to glycan receptors. *Biochem J*, Volume 444, pp. 429-435.
- Jayasekera, J. P., Moseman, E. A. & Carroll, M. C., 2007. Natural antibody and complement mediate neutralization of influenza virus in the absence of prior immunity. *J Virol*, Volume 81, pp. 3487-3494.
- Jensen, S. & Thomsen, A. R., 2012. Sensing of RNA viruses: a review of innate immune receptors involved in recognizing RNA virus invasion. *J Virol*, Volume 86, pp. 2900-2910.
- Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, Volume 24, pp. 1403-1405.
- Kamperschroer, C. et al., 2006. SAP is required for Th cell function and for immunity to influenza. *J Immunol*, Volume 177, pp. 5317-5327.
- Karlsson, N. G. et al., 1997. The glycosylation of rat intestinal Muc2 mucin varies between rat strains and the small and large intestine: a study of O-linked oligosaccharides by a mass spectrometry approach. *J Biol Chem*, Volume 272, pp. 27025-27034.
- Kashyap, V. K. et al., 2006. Genetic structure of Indian populations based on fifteen autosomal microsatellite loci. *BMC Genetics*.

- Kato, H. et al., 2006. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature*, Volume 44, pp. 101-105.
- Kilbourne, E. D. et al., 2004. Protection of mice with recombinant influenza virus neuraminidase. *J Infect Dis*, Volume 189, pp. 459-461.
- Kim, H. M. et al., 2008. Alveolar macrophages are indispensable for controlling Influenza viruses in lungs of pigs. *J Virol*, Volume 82, pp. 4265-4274.
- Kim, M. C. et al., 2013. Multiple heterologous M2 extracellular domains presented on virus-like particles confer broader and stronger M2 immunity than live influenza A virus infection. *Antiviral Res*, Volume 99, pp. 328-335.
- Kirkham, S. et al., 2002. Heterogeneity of airway mucus: variations in the amounts and glycoforms of the major oligomeric mucins MUC5AC and MUC5B. *Biochem J*, Volume 361, pp. 537-546.
- Kirk, H. & Freeland, J. R., 2011. Applications and implications of neutral versus non-neutral markers in molecular ecology. *In J Mol Sci*, pp. 3966-3988.
- Kirk, H. & Freeland, J. R., 2011. Applications and implications of neutral vs. non-neutral markers in molecular ecology. *Int J Mol Sci*, pp. 3966-3988.
- Kitagawa, H. & Paulson, J. C., 1994. Differential expression of five sialyltransferase genes in human tissues. *The J of Biol Chem*, Volume 269, pp. 17872-17878.
- Kitts, A., Church, D., Hefferon, T. & Phan, L., 2014. *The NCBI Handbook*, Bethesda (MD): National Center for Biotechnology Information.
- Knowles, M. R. & Boucher, R. C., 2012. Mucus clearance as a primary innate defense mechanism for mammalian airways. *J Clin Invest*, Volume 109, pp. 571-577.
- Koel, B. F. et al., 2013. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science*, Volume 342, pp. 976-979.
- Koutsonanos, D. G. et al., 2011. Serological memory and long-term protection to novel H1N1 influenza virus after skin vaccination. *J Infect Dis*, Volume 204, pp. 582-591.
- Koyama, S. et al., 2007. Differential role of TLR and RLR-signaling in the immune responses to influenza A virus infection and vaccination. *J Immunol*, Volume 179, pp. 4711-4720.
- Kreijtz, J. H. C. M. et al., 2007. Primary influenza A virus infection induces cross-protective immunity against a lethal infection with a heterosubtypic virus strain in mice. *Vaccine*, Volume 25, pp. 612-620.
- Kreijtz, J. H. C. M. et al., 2009. Infection of mice with a human influenza A/H3N2 virus induces protective immunity against lethal infection with influenza A/H5N1 virus. *Vaccine*, Volume 27, pp. 4983-4989.
- Kreijtz, J. H. C. M. et al., 2008. Cross-recognition of avian H5N1 influenza virus by human cytotoxic T-lymphocyte populations directed to human influenza A virus. *J Virol*, Volume 82, pp. 5161-5166.
- Kreijtz, J. H. C. M., Fouchier, R. A. M. & Rimmelzwaan, G. F., 2011. Immune responses to influenza virus infection. *Virus Res*, Volume 162, pp. 19-30.
- Kriegs, J. O. et al., 2007. Evolutionary history of the 7SL RNA-derived SINEs in Suprimate. *TRENDS in Genetics*, pp. 158-161.

- La Gruta, N. L., Kedzierska, K., Stambas, J. & Doherty, P. C., 2007. A question of self-preservation: Immunopathology in influenza virus infection. *Immunol Cell Biol*, Volume 85, pp. 85-92.
- La Gruta, N. L., Turner, S. J. & Doherty, P. C., 2004. Hierarchies in cytokine expression profiles for acute and resolving influenza virus-specific CD8+ T cell responses: Correlation of cytokine profile and TCR avidity. *J Immunol*, Volume 172, pp. 5553-5560.
- Lamere, M. W. et al., 2011. Regulation of antinucleoprotein IgG by systemic vaccination and its effects on influenza virus clearance. *J Virol*, Volume 85, pp. 5027-5035.
- Landrum, M. J. et al., 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acid Res*, pp. D980-985.
- Le Goffic, R. et al., 2007. Cutting edge: influenza A virus activates TLR3-dependent inflammatory and RIG-I-dependent antiviral responses in human lung epithelial cells. *J Immunol*, Volume 178, pp. 3368-3372.
- Lee, L. Y. et al., 2008. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. *J Clin Invest*, Volume 118, pp. 3478-3490.
- Lee, S. M. Y. et al., 2014. Toll like receptor 10 is involved in induction of innate immune responses to influenza virus infection. *PNAS*, Volume 111, pp. 3793-3798.
- Le, M. Q. et al., 2013. Migration and persistence of human influenza A viruses, Vietnam, 2001-2008. *Emerg Infect Dis*, Volume 19, pp. 1756-1765.
- Leonardi, V., 2005. Sírios, árabes e mouros em nossa herança ibérica. Em: *Os navegantes e o sonho: Presença do Oriente na História do Brasil*. s.l.:Paralelo 15, p. 172.
- Lewallen, D. M., Siler, D. & Iyer, S. S., 2009. Factors affecting protein-glycan specificity: Effect of spacers and incubation time. *Chem Bio Chem*, Volume 10, pp. 1486-1489.
- Lewontin, R. C., 1972. An apportionment of human diversity. Em: T. Dobzhansky, ed. *Evolutionary Biology*. s.l.:s.n., pp. 381-398.
- Librado, P. & Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, pp. 1451-1452.
- Ligges, U. & Mächler, M., 2003. Scatterplot3d - an R package for visualizing multivariate data. *Journal of Statistical Software*, 8(11), pp. 1-20.
- Li, J. Z. et al., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, pp. 1100-1104.
- Lillehoj, E. P. & Kim, K. C., 2002. Airway mucus: its components and function. *Arch Pharma Res*, Volume 25, pp. 770-780.
- Lin, K. L. et al., 2008. CCR2+ monocyte-derived dendritic cells and exudate macrophages produce influenza-induced pulmonary immune pathology and mortality. *J Immunol*, Volume 180, pp. 2562-2572.
- Lin, T. Y. & Brass, A. L., 2013. Host genetic determinants of influenza pathogenicity. *Current Opinion in Virology*, Volume 3, pp. 531-536.

- Liu, H., Prugnolle, F., Manica, A. & Balloux, F., 2006. A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet*, pp. 230-237.
- Liu, Y. et al., 2013. Genetic variants in IL1A and IL1B contribute to the susceptibility to 2009 pandemic H1N1 influenza A virus. *BMC Immunology*, Volume 14, p. 37.
- Lund, E. et al., 2004. Nuclear export of microRNA precursors. *Science*, Volume 303, pp. 95-98.
- Mailman, M. D. et al., 2007. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*, pp. 1181-1185.
- Mänz, B. et al., 2013. Pandemic influenza A viruses escape from restriction by human MxA through adaptive mutations in the nucleoprotein. *PLOS Pathogens*, Volume 9, p. e1003279.
- Markert, C. L., 1968. The molecular bases for isozymes. *Annals New York Academy of Sciences*, pp. 1-38.
- Matrosovich, M. & Klenk, H. D., 2003. Natural and synthetic sialic acid-containing inhibitors of influenza virus receptor binding. *Rev Med Virol*, Volume 13, pp. 85-97.
- Matrosovich, M. et al., 2000. Early alterations of the receptor-binding properties of H1, H2 and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J Virol*, Volume 74, pp. 8502-8512.
- Medof, M. E. et al., 1987. Identification of the complement decay-accelerating factor (DAF) on the epithelium and glandular cells and in body fluids. *J Exp Med*, Volume 165, pp. 848-864.
- Medzhitov, R. et al., 1998. MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. *Mol Cell*, Volume 2, pp. 253-258.
- Metkar, S. S. et al., 2008. Human and mouse granzyme A induce proinflammatory cytokine response. *Immunity*, Volume 29, pp. 720-733.
- Ministério da Cultura, 2004. *portaria 6/2004, artigo 2º*. [Online] Available at: <http://www.cpis.org.br/htm/leis/page.aspx?LeiID=141> [Acesso em 11 07 2016].
- Moffat, J. M. et al., 2009. Granzyme A expression reveals distinct cytolytic CTL subsets following influenza A virus infection. *Eur J Immunol*, Volume 39, pp. 1203-1210.
- Montinaro, F. et al., 2015. Unravelling the hidden ancestry of American admixed populations. *Nature Communications*, p. 6596.
- Mozdzanowska, K., Maiese, K., Furchner, M. & Gerhard, W., 1999. Treatment of influenza virus-infected SCID mice with nonneutralizing antibodies specific for the transmembrane proteins matrix 2 and neuraminidase reduces the pulmonary virus titer but fails to clear the infection. *Virology*, Volume 254, pp. 138-146.
- Münster, A. K. et al., 1998. Mammalian cystidine 5'-monophosphate N-acetylneuraminic acid synthetase: A nuclear protein with evolutionarily conserved structural motifs. *PNAS USA*, Volume 95, pp. 9140-9145.
- Nadziejko, C. & Finkelstein, I., 1994. Inhibition of neutrophil elastase by mucus glycoprotein. *Am J Respir Cell Mol Biol*, Volume 11, pp. 103-107.

- Nakanishi, Y., Lu, B., Gerard, C. & Iwasaki, A., 2009. CD8(+) T lymphocyte mobilization to virus-infected tissue requires CD4(+) T-cell help. *Nature*, Volume 462, pp. 510-513.
- Nakatsuji, T., Burnley, M. S. & Huisman, T. H., 1985. Fetal hemoglobin variants identified in adults through restriction endonuclease gene mapping methodology. *Blood*, pp. 803-807.
- Neel, J. V., 1973. "Private" genetic variants and the frequency of mutation among South American Indians. *PNAS*, pp. 3311-3315.
- Nei, M. & Li, W. H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS*, 76(10), pp. 5269-5273.
- Nelis, M. et al., 2009. Genetic structure of Europeans: A view from the North-East. *PLoS One*.
- Ng, B. G. et al., 2013. Mosaicism of the UDP-Galactose transporter SLC35A2 causes a congenital disorder of glycosylation. *The Amer J of Hum Genet*, Volume 92, pp. 632-636.
- Nicholls, J. M. et al., 2007. Sialic acid receptor detection in the human respiratory tract: evidence for widespread distribution of potential binding sites for human and avian influenza viruses. *Respiratory Research*, Volume 8, p. 73.
- Nicholls, J. M., Lai, J. & Garcia, J. M., 2012. Investigating the interaction between influenza A and sialic acid: making and breaking the link. Em: M. von Itzstein, ed. *Influenza Virus Sialidase - A drug discovery target*. s.l.:s.n.
- Nicoletto, S. F. & Rinaldi, A., 2011. In the womb's shadow. *EMBO Reports*, Volume 12, pp. 30-34.
- Ogasawara, Y. et al., 2007. Sialic acid is an essential moiety of mucin as a hydroxyl radical scavenger. *FEBS Letters*, Volume 581, pp. 2473-2477.
- Ogura, Y. et al., 2001. Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF- κ B. *J Biol Chem*, Volume 276, pp. 4812-4818.
- Ohuchi, M., Ohuchi, R., Feldmann, A. & Kelnk, H. D., 1997. Regulation of receptor binding affinity of influenza virus hemagglutinin by its carbohydrate moiety. *J Virol*, Volume 71, pp. 8377-8384.
- Okoye, I. S. & Wilson, M. S., 2011. CD4+ T helper 2 cells - Microbial triggers, differentiation requirements and effector functions. *Immunology*, Volume 134, pp. 368-377.
- Olczak, M. et al., 2013. UDP-Gal/UDP-GlcNAc chimeric transporter complements mutation defect in mammalian cells deficient in UDP-Gal transporter. *Biochem and Biophys Res Commun*, Volume 434, pp. 473-478.
- Oliveira, S. F. et al., 2004. Uni- and bi-parental analyses of the genetic contribution in an afro-descendent community in central Brazil. Em: J. E. Egocheaga Rodríguez, M. Egocheaga Rodríguez & J. Egocheaga Rodríguez, eds. *Biología de poblaciones humanas: diversidad, tiempo, espacio*. s.l.:Sociedad Española de Antropología Biológica.
- Onodera, T. et al., 2012. Memory B cells in the lung participate in protective humoral immune responses to pulmonary influenza virus infection. *PNAS USA*, Volume 109, pp. 2485-2490.
- Osuka, F. et al., 2007. Molecular cloning and characterization of novel splicing variants of human decay-accelerating factor. *Genomics*, Volume 88, pp. 316-322.

- Pang, I. K. & Iwasaki, A., 2011. Control of antiviral immunity by pattern recognition and the microbiome. *Immun Rev*, Volume 245, pp. 209-226.
- Paradis, E., Claude, J. & Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, Volume 20, pp. 289-290.
- Parra, E. J. et al., 1998. Estimating African American admixture proportions by use of population-specific alleles. *Am. J. Hum. Genet.*, pp. 1839-1851.
- Parra, F. C. et al., 2003. Color and genomic ancestry in Brazilians. *PNAS*, pp. 177-182.
- Patin, E. et al., 2009. Inferring the demographic history of African farmers and Pygmy hunter-gatherers using multilocus resequencing data set. *PLoS Genetics*, p. e1000448.
- Patnaik, S. K., Helmberg, W. & Blumenfeld, O. O., 2012. BGMUT: NCBI dbRBC database of allelic variations of genes encoding antigens of blood group systems. *Nucleic Acid Res*, pp. D1023-1029.
- Pedrosa, M. A. F., 2006. *Composição genética de quatro populações remanescentes de quilombos do Brasil com base em microssatélites e marcadores de ancestralidade*. Tese de Mestrado ed. Brasília: Universidade de Brasília.
- Pedrosa, M. A. F., 2006. *Composição genética de quatro populações remanescentes de quilombos do Brasil com base em microssatélites e marcadores de ancestralidade*. Tese de Mestrado ed. Brasília: Universidade de Brasília.
- Pfaff, C. L. et al., 2001. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am J Hum Genet*, pp. 198-207.
- Piel, F. B. et al., 2010. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nature Communications*, pp. 1-7.
- Pilchmair, A. et al., 2011. IFIT1 is an antiviral protein that recognizes 5'-triphosphate RNA. *Nature Immunology*, Volume 12, pp. 624-630.
- Porrás-Hurtado, L. et al., 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*, pp. 1-13.
- Price, G. E. et al., 2000. Viral escape by selection of cytotoxic T cell-resistant variants in influenza-A virus pneumonia. *J Exp Med*, Volume 191, pp. 1853-1867.
- Pritchard, J. K., Stephens, M. & Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, Volume 155, pp. 945-959.
- Pritchard, J. K., Wen, X. & Falush, D., 2010. *Documentation for structure software: Version 2.3*. s.l.:s.n.
- Qi, Y. et al., 2014. A novel pyrosequencing assay for the detection of neuraminidase inhibitor resistance-conferring mutations among clinical isolates of avian H7N9 influenza virus. *Virus Research*, Volume 179, pp. 119-124.
- Quintana-Murci, L. et al., 2007. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu speaking farmers. *PNAS*, pp. 1596-1601.
- R Core Team, 2015. *R: a language environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

- Raghavan, M. et al., 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, pp. 87-91.
- Ramachandran, S. et al., 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *PNAS*, pp. 15942-15947.
- Ray, D. A., Walker, J. A. & Batzer, M. A., 2007. Mobile element-based forensic genomics. *Mutation Research*, Volume 616, pp. 24-33.
- Ray, D. A. et al., 2005. Inference of human geographic origins using Alu insertion polymorphisms. *Forensic Science International*, Volume 153, pp. 117-124.
- Reich, D. et al., 2012. Reconstructing Native American population history. *Nature*, pp. 370-375.
- Reynolds, J., Weir, B. S. & Cockerham, C. C., 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, 105(3), pp. 767-779.
- Ribeiro, G. G. B. L., 2009. *Análise genética de marcadores do tipo STR e indel em cromossomos sexuais humanos em populações remanescentes de quilombo*. Tese de Doutorado ed. Brasília: Universidade de Brasília.
- Riberdy, J. M., Christensen, J. P., Branum, K. & Doherty, P. C., 2000. Diminished primary and secondary influenza virus-specific CD8(+) T-cell responses in CD4-depleted Ig(-/-) mice. *J Virol*, Volume 74, pp. 9862-9765.
- Rimmelzwaan, G. F. et al., 2004. Sequence variation in the influenza A virus nucleoprotein associated with escape from cytotoxic T lymphocytes. *Virus Res*, Volume 103, pp. 97-100.
- Rose, M. C. & Voynow, J. A., 2006. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiol Rev*, Volume 86, pp. 245-278.
- Sabbah, A. et al., 2009. Activation of immune antiviral responses by Nod2. *Nature Immunol*, Volume 10, pp. 1073-1080.
- Saitou, N., 1995. A genetic affinity analysis of human populations. *Hum Evol*, pp. 17-33.
- Saitou, N. & Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, Jul, 4(4), pp. 406-425.
- Salem, A. H., Ray, D. A. & Batzer, M. A., 2005. Identity by descent and DNA sequence variation of human SINE and LINE elements. *Cytogenet Genome Res*, pp. 63-72.
- Salzano, F. M. & Callegari-Jacques, S. M., 1988. South American Indians: a case study in evolution. Em: *Research Monographs on Human Population Biology*. Oxford: Clarendon Press.
- Sambhara, S. et al., 2001. Heterosubtypic immunity against human influenza viruses, including recently emerged avian H5 and H9 viruses, induced by FLU-ISCOM vaccine in mice requires both cytotoxic T-lymphocyte and macrophage function. *Cell Immunol*, Volume 211, pp. 143-153.
- Sanger, F., Nicklen, S. & Coulson, R., 1977. DNA sequencing with chain-terminating inhibitors. *PNAS*, pp. 5463-5467.
- Sato, A. et al., 2006. Profilin is an effector for Daam1 in non-canonical Wnt signaling and is required for vertebrate gastrulation. *Development*, pp. 4219-4231.

- Schmitz, N., Kurrer, M., Bachmann, M. F. & Kopf, M., 2005. Interleukin-1 is responsible for acute lung immunopathology but increases survival of respiratory Influenza virus infection. *J Virol*, Volume 79, pp. 6441-6448.
- Schnell, J. R. & Chou, J. J., 2008. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature*, Volume 45, pp. 591-595.
- Schotsaert, M., de Fillette, M., Fiers, W. & Saelens, X., 2009. Universal M2 ectodomain-based influenza A vaccines: Preclinical and clinical developments. *Expert Rev Vaccine*, Volume 8, pp. 499-508.
- Schoumans, J. et al., 2004. The performance of CGH array for the detection of cryptic constitutional chromosome imbalances. *J Med Genet*, pp. 198-202.
- Schulman, J. L. & Kilbourne, E. D., 1963. Transmission of influenza virus infection in mice. II. Some factors affecting the incidence of transmitted infection. *J Exp Med*, Volume 125, pp. 479-488.
- Shedlock, A. M. & Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. *Bioessays*, pp. 148-160.
- Shedlock, A. M. & Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. *BioEssays*, pp. 148-160.
- Sherry, S. T. et al., 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, pp. 308-311.
- Shi, Y. et al., 2012. Identification of novel rare mutations of DACT1 in human neural tube defects. *Hum Mutat*, pp. 1450-1455.
- Shriver, M. D. et al., 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. *Am. J. Hum. Genet.*, pp. 957-964.
- Shynia, K. et al., 2006. Avian flu: Influenza virus receptors in the human airway. *Nature*, Volume 440, pp. 435-436.
- Simonsen, L. et al., 2013. Global mortality estimates for the 2009 influenza pandemic from the GlAMOR Project: A modelling study. *PLOS Medicine*, Volume 11, p. e1001558.
- Skehel, J. & Wiley, D. C., 2000. Receptor binding and membrane fusion in virus entry: the influenza Hemagglutinin. *Annu Rev Biochem*, Volume 69, pp. 531-569.
- Slatkin, M., 1991. Inbreeding coefficients and coalescence times. *Genet Res*, Volume 2, pp. 167-175.
- Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139(1), pp. 457-462.
- Slatkin, M. & Excoffier, L., 1996. Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity*, Volume 76, pp. 377-383.
- Sridhar, S. et al., 2012. Predominance of heterosubtypic IFN-gamma-only-secreting effector memory T cells in a pandemic H1N1 naïve adults. *Eur J Immunol*, Volume 2012.
- Sriwilaijaroen, N. & Suzuki, Y., 2012. Molecular basis of the structure and function of the H1 hemagglutinin of influenza virus. *Proc Jpn Acad Ser B*, Volume 88, pp. 226-249.

Stoneking, M. et al., 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Research*, pp. 1061-1071.

Strutt, T. M. et al., 2014. Memory CD4+ T cells induce innate immune responses independently of pathogen. *Nat Med*, Volume 16, pp. 558-564.

Su, C. T. T., Schönbach, C. & Kwoh, C. K., 2014. Molecular docking analysis of 2009-H1N1 and 2004-H5N1 influenza virus HLA-B*4405-restricted HA epitope candidates: implications for TCR cross-recognition and vaccine development. *BMC Bioinformatics*, Volume 14 (suppl 2), p. S21.

Sun, X. et al., 2013. N-linked glycosylation of the hemagglutinin protein influences virulence and antigenicity of the 1918 pandemic and seasonal H1N1 influenza A viruses. *J Virol*, Volume 87, pp. 8756-8766.

Tajima, F., 1993. Measurement of DNA polymorphism. In: N. Takahata & A. Clark, eds. *Mechanisms of Molecular Evolution. Introduction to Molecular Paleopopulation Biology*. Tokyo; Sunderland, MA: Japan Scientific Societies Press. Sinauer Associates, Inc., pp. 37-59.

Tan, S. L. & Katze, M. C., 1998. Using genetic means to dissect homologous and heterologous protein-protein interactions of PKR, the Interferon-induced protein kinase. *Methods*, Volume 15, pp. 207-223.

Thakaramaran, K. et al., 2012. Structural determinants for naturally evolving H5N1 hemagglutinin to switch its receptor specificity. *Cell*, Volume 153, pp. 1475-1485.

Thakaramaran, K. et al., 2013. Glycan receptor binding of the influenza A virus H7N9 hemagglutinin. *Cell*, Volume 153, pp. 1486-1493.

The 1000 Genomes Project Consortium, 2010. A map of human genome variation from population-scale sequencing. *Nature*, Volume 467, pp. 1061-1073.

The 1000 Genomes Project Consortium, 2015a. A global reference for human genetic variation. *Nature*, pp. 68-74.

The 1000 Genomes Project Consortium, 2015b. An integrated map of structural variation in 2504 human genomes. *Nature*, pp. 75-81.

The International HapMap Consortium, 2003. The International HapMap Consortium. *Nature*, pp. 789-796.

The International HapMap Consortium, 2005. A haplotype map of the human genome. *Nature*, pp. 1299-1320.

The International HapMap Consortium, 2005. A haplotype map of the human genome. *Nature*, pp. 1299-1320.

Throsby, M. et al., 2008. Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. *PLoS ONE*, 3(12), p. e3492.

Tishkoff, S. A. & Verrelli, B. C., 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Ann Rev Genomics Hum Genet*, pp. 293-340.

Toledo, R., 2011. *Região 3'UTR do gene HLA-G em populações humanas do Centro-Oeste*. Brasília: Dissertação de Mestrado.

- Tompkins, S. M. et al., 2007. Matrix 2 protein vaccination and protection against influenza viruses, including subtype H5N1. *Emerg Infect Dis*, Volume 13, pp. 426-435.
- Tria, F., Pompei, S. & Loreto, V., 2013. Dynamically correlated mutations drive human Influenza A evolution. *Nature Scient Rep*, Volume 3, p. 2705.
- Tumpey, T. M. et al., 2005. Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science*, Volume 310, pp. 77-80.
- Uiprasertkul, M. et al., 2005. Influenza A H5N1 replication sites in humans. *Emerg Infect Dis*, Volume 11, pp. 1036-1041.
- Ullu, E. & Tschudi, C., 1984. Alu sequences are processed 7SL RNA genes. *Nature*, pp. 171-172.
- United Nations Statistics Division, 2014. *Demographic Yearbook*, s.l.: s.n.
- van de Sandt, C. E., Kreijtz, J. H. C. M. & Rimmelzwaan, G. F., 2012. Evasion of influenza A viruses from innate and adaptive immune responses. *Viruses*, Volume 4, pp. 1438-1476.
- van Domselaar, R. & Bovenschen, N., 2011. Cell death-independent functions of granzymes: Hit viruses where it hurts. *Rev Med Virol*, Volume 21, pp. 301-314.
- van Gisbergen, K. P. et al., 2011. The costimulatory molecule CD27 maintains clonally diverse CD8(+) T cell responses of low antigen affinity to protect against viral variants. *Immunity*, Volume 35, pp. 97-108.
- van Riel, D. et al., 2011. Highly pathogenic avian influenza virus H5N1 infects alveolar macrophages without virus production or excessive TNF-alpha induction. *PLoS Pathog*, Volume 7:(6), p. e1002099.
- Vansina, J. & Obenga, T., 2011. *Coleção História Geral da África UNESCO*. 2 ed. s.l.:Cortez & UNESCO.
- Verdu, P. et al., 2009. Origins and genetic diversity of Pygmy hunter-gatherers from Western Central Africa. *Current Biology*, pp. 312-318.
- Voeten, J. T. et al., 2000. Antigenic drift in the influenza A virus (H3N2) nucleoprotein and escape from recognition by cytotoxic T lymphocytes. *J Virol*, Volume 74, pp. 6800-6807.
- Voeten, J. T. et al., 1998. Use of recombinant nucleoproteins in enzyme-linked immunosorbent assays for detection of virus-specific immunoglobulin A (IgA) and IgG in antibodies in influenza A- or B-infected patients. *J Clin Microbiol*, Volume 36, pp. 3527-3531.
- von Itzstein, M., 2007. The war against influenza: discovery and development of sialidase inhibitors. *Nat Rev Drug Disc*, Volume 6, pp. 967-974.
- Voynow, J. A., Selby, D. M. & Rose, M. C., 1998. Mucin gene expression (MUC1, MUC2 and MUC5/5AC) in nasal epithelial cells of cystic fibrosis, allergic rhinitis and normal individuals. *Lung*, Volume 176, pp. 345-354.
- Waffarn, E. E. & Baumgarth, N., 2011. Protective B cell responses to flu - No fluke!. *J Immunol*, Volume 186, pp. 3823-3829.
- Wagner, R., Matrosovich, M. & Klenk, H. D., 2002. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev Med Virol*, Volume 12, pp. 159-166.

Wallace, R. G. & Fitch, W. M., 2008. Influenza A H5N1 immigration is filtered out at some international borders. *PLoS ONE*, Volume 3, p. e1697.

Wang, G. et al., 2013. Epigenetics and early life origins of chronic noncommunicable diseases. *Journal of Adolescent Health*, 52(2 (suppl 2)), pp. S14-S21.

Wang, J., 2006. Informativeness of genetic marker for pairwise relationship and relatedness inference. *Theoretical Population Biology*, pp. 300-321.

Wang, J. et al., 2006. dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum Mut*, pp. 323-329.

Wang, S. et al., 2007. Genetic variation and population structure in Native Americans. *PLoS Genetics*, p. e185.

Wang, Y. et al., 2009. Monoclonal antibody recognizing SLLTEVET epitope of M2 protein potentially inhibited replication of influenza A viruses in MDCK cells. *Biochem Biophys Res Comm*, Volume 385, pp. 118-122.

Wang, Z. et al., 2010. Genetic admixture and population substructure in Guanacaste Costa Rica. *PLoS ONE*, p. e13336.

Watkins, W. S. et al., 2001. Patterns of Ancestral Human Diversity: An Analysis of Alu-Insertion and Restriction-Site Polymorphisms. *American Journal of Human Genetics*, Volume 68, pp. 738-752.

Watkins, W. S. et al., 2003. Genetic Variation Among World Populations: Inferences from 100 Alu Insertion Polymorphisms. *Genome Res.*, pp. 1-12.

Weir, B. S. & Cockerham, C. C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, pp. 1358-1370.

WHO, s.d. *World Health Organization*. [Online] Available at: <http://www.who.int/en/> [Acesso em 10 01 2016].

Wijburg, O. L. C., Heemskerk, M. H. M., Boog, C. J. P. & van Rooijen, N., 1997. Role of spleen macrophages in innate and acquired immune responses against mouse hepatitis virus strain A59. *Immunology*, Volume 92, pp. 252-258.

Witherspoon, D. J. et al., 2006. Human population genetic structure and diversity inferred from polymorphic L1 (LINE-1) and Alu insertions. *Hum Hered*, pp. 30-46.

Worobey, M., Guan-Zhu, H. & Rambaut, A., 2014. Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *PNAS*, Volume 111, pp. 8107-8112.

Xu, S., 2012. Human population admixture in Asia. *Genomics and Informatics*, pp. 133-144.

Yang, M. L. et al., 2011. Galectin-1 binds to influenza virus and ameliorates influenza virus pathogenesis. *J Virol*, Volume 85, pp. 10010-10020.

Yang, Z. & Huffman, F. L., 2013. Nutrition in pregnancy and early childhood and associations with obesity in developing countries. *Maternal and Child Nutrition*, 9(suppl 1), pp. 105-119.

Yu, X. et al., 2008. Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature*, Volume 455, pp. 532-536.

Zhou, J. et al., 2012. A functional variation in CD55 gene increases the disease severity of pandemic A(H1N1)2009 influenza infection. *J Infect Dis*, Volume 206, pp. 495-503.

Zhu, J. & Paul, W. E., 2010a. Heterogeneity and plasticity of T helper cells. *Cell Res*, Volume 20, pp. 4-12.

Zhu, J. & Paul, W. E., 2010b. Peripheral CD4+ T-cell differentiation regulated by networks of cytokines and transcription factors. *Immunol Rev*, Volume 238, pp. 247-262.

Zimmermann, C., Prevost-Blondel, A., Blaser, C. & Pircher, H., 1999. Kinetics of the response of naïve and memory CD8 T cells to antigen: similarities and differences. *Eur J Immunol*, Volume 29, pp. 284-290.



Universidade de Brasília
Faculdade de Ciências da Saúde
Comitê de Ética em Pesquisa – CEP/FS

PROCESSO DE ANÁLISE DE PROJETO DE PESQUISA

Registro do Projeto no CEP: 151/07

Título do Projeto: ANCESTRABILIDADE BIOLÓGICA DE POPULAÇÕES REMANESCENTES DE QUILOMBOLAS BASEADA EM MARCADORES MOLECULARES LOCALIZADOS EM CROMOSSOMOS SEXUAIS E AUTOSSÔMICOS.

Pesquisador Responsável: Silviene Fabiana de Oliveira

Data de Entrada: 30/01/2007

Com base nas Resoluções 196/96, do CNS/MS, que regulamenta a ética da pesquisa em seres humanos, o Comitê de Ética em Pesquisa com Seres Humanos da Faculdade de Ciências da Saúde da Universidade de Brasília, após análise dos aspectos éticos e do contexto técnico-científico, resolveu **APROVAR** o projeto 151/2007 com o título: "Ancestralidade Biológica de Populações Remanescentes de Quilombolas Baseada em Marcadores Moleculares Localizados em Cromossomos Sexuais e Autossômicos", analisado na 11ª Reunião Ordinária, realizada no dia 11 de Dezembro de 2007.

A pesquisadora responsável fica, desde já, notificada da obrigatoriedade da apresentação de um relatório semestral e relatório final sucinto e objetivo sobre o desenvolvimento do Projeto, no prazo de 1 (um) ano a contar da presente data (item VII.13 da Resolução 196/96).

Brasília, 12 de Dezembro de 2008.

Prof. Volnei Garrafa
Coordenador do CEP-FS/UnB



Universidade de Brasília
Faculdade de Ciências da Saúde
Comitê de Ética em Pesquisa – CEP/FS

PROCESSO DE ANÁLISE DE PROJETO DE PESQUISA

Registro do Projeto: 028/2001

Título do Projeto: "Brasília – Reflexo da Constituição Genética Brasileira",

Data da Entrada: 21/05/2001

Responsável Principal: Silviene Fábiana de Oliveira

Com base nas Resoluções 196/96, do CNS/MS, que regulamenta a ética da pesquisa em seres humanos, o Comitê de Ética em Pesquisa da Faculdade de Ciências da Saúde da Universidade de Brasília, após análise dos aspectos éticos e do contexto técnico-científico e julgando satisfatórias as justificativas e pendências apresentadas a(o) relator(s), resolveu **APROVAR** o projeto 028/01 com o título: "Brasília – Reflexo da Constituição Genética Brasileira", já analisado em sua 39ª Reunião, realizada no dia 08 de junho de 2001, retomando a 41ª Reunião, realizada no dia 14 de agosto de 2001.

Como o projeto está enquadrado na área temática especial – a de pesquisa genética, deverá ser encaminhado à CONEP, aguardando a apreciação final, para início da execução

Brasília, 14 de agosto de 2001

Prof. Volnei Garrafa
Coordenador do CEP-FSA/UnB

Campus Universitário Darcy Ribeiro
Faculdade de Ciências da Saúde
Cep: 70 910-900

Tel: 307-32292113

23/03/05



MINISTÉRIO DA SAÚDE
Conselho Nacional de Saúde
Comissão Nacional de Ética em Pesquisa - CONEP

PARECER Nº 1259/2001

Registro CONEP = 2932 (Este nº deverá ser citado nas correspondências referentes a este projeto)

Protocolo CEP = 028/2001

Processo nº 25000.102003/2001-05

Projeto de Pesquisa: *Brasília: um reflexo de constituição genética brasileira?*

Pesquisador Responsável: Dra. Silviene Fabiana de Oliveira

Instituição: Departamento de Genética e Morfologia / Universidade de Brasília / UNB / DF

Área Temática Especial: Genética Humana

Ao se proceder à análise do protocolo em questão, cabem as seguintes considerações:

a) as informações enviadas atendem aos aspectos fundamentais da Resolução CNS 196/96, sobre Diretrizes e Normas Regulamentadoras de Pesquisas Envolvendo Seres Humanos;

b) o projeto foi aprovado pelo Comitê de Ética em Pesquisa - CEP da instituição supracitada.

Diante do exposto, a Comissão Nacional de Ética em Pesquisa - CONEP, de acordo com as atribuições definidas na Res. CNS 196/96, manifesta - se pela aprovação do projeto de pesquisa proposto.

Situação: Projeto aprovado.

Brasília, 19 de outubro de 2001


WILLIAM SAAD HOSSNE
Coordenador da CONEP-MS

Anexo I.2 Protocolo para Amplificação de Genoma Total com o kit GenomiPhi®

1. Aliquotar uma pequena quantidade da amostra a ser amplificada para obter alíquotas com concentração 10ng/μl;
2. Degelar os reagentes do kit GenomiPhi em gelo (reagentes estocados no freezer a -80°C);
3. Preparar 2 tubos de PCR de 0,2μl por amostra;
4. Adicionar 9μl do tampão de amostra (tubo de tampa verde) a cada tubo de PCR;
5. Adicionar 1μl da alíquota de DNA feita anteriormente ao seu respectivo tubo;
6. Aquecer os tubos de DNA + tampão a 95°C no termociclador (BioRad®) por exatamente 3 minutos e resfriar imediatamente a 4°C no gelo;
7. Preparar o máster mix para a quantidade de tubos:
 - a. 9μl de tampão de reação (tubo de tampa azul) por tubo;
 - b. 1μl de enzima (tubo de tampa amarela) por tubo.
8. Adicionar 10μl desse máster mix imediatamente a cada um dos tubos com DNA;
9. Colocar no termociclador seguindo o seguinte programa:
 - a. 30°C por 1h30 min;
 - b. Desativação da enzima a 65°C por 10 minutos;
 - c. Resfriamento das amostras a 4°C.

Anexo I.3 Purificação dos genomas amplificados.

1. Adicionar 20 μ l (mesmo volume) de água destilada e deionizada estéril a cada tubo de genoma amplificado, centrifugar se necessário;
2. Combinar os conteúdos dos tubos de mesma amostra (80 μ l total) em um tubo de 1,5ml de tampa de rosca;
3. Adicionar 8 μ l (1/10 do volume) de tampão acetato de sódio/EDTA (1,5M de acetato de sódio; pH > 8,0; 250mM EDTA) temperatura ambiente a cada tubo;
4. Adicionar 400 μ l de Etanol a 100% e misturar por inversão para precipitação do DNA;
5. Centrifugar a 12.000 RPM por 15 minutos;
6. Cuidadosamente descartar o sobrenadante e lavar o precipitado com etanol 70%;
7. Centrifugar novamente a 12.000 RPM por 3 minutos;
8. Descartar o sobrenadante, remover o excesso de álcool com ponteira de pipeta e deixar secar aberto para remover todo o etanol visível;
9. Suspender o DNA amplificado em 80 μ l de tampão Tris-EDTA (TE)
10. Incubar a 60°C por uma hora, ou a temperatura ambiente overnight.

Anexo I.4 Tabela contendo os nomes das inserções *Alu*, o cromossomo em que se encontra, os iniciadores 5' (forward) e 3' (reverse), os tamanhos dos fragmentos amplificados pela PCR (presença de inserção e ausência de inserção, respectivamente), a temperatura de anelamento utilizada no programa da MyCycler™ e a concentração de cloreto de magnésio (MgCl₂) exigida para cada inserção *Alu*.

# <i>Alu</i>	Nome	# Cr.	Iniciadores (todos na orientação 5'-3')		Tamanho (pb)	T°	[] mM MgCl ₂
			Forward	Reverse			
1	ACE	17	CTGGAGACCACTCCCATCCTTTCT	GATGTGGCCATCACATTCGTCAGAT	480-191	58	1,5
2	APO	11	AAGTGCTGTAGGCCATTTAGATTAG	AGTCTTCGATGACAGCGTATACAGA	409-96	55	1,5
3	B65	11	ATATCCTAAAAGGGACACCA	AAAATTTATGGCATGCGTAT	420-81	60	3,0
4	COL3A1	2	ACCTGCAGCACCAGGAGGTCCTGGAGGGC C	GAGTCCTTTAGAAGGATATGCTCTG	1200-900	60	1,5
5	HS2,43	1	ACTCCCCACCAGGTAATGGT	AGGGCCTTCATCCAGTTTGT	482-184	62	1,5
6	HS4,32	12	GTTTATTGGGCTAACCTGGG	TGACCAGCTAACTTCTACTTTAACC	601-289	60	1,5
7	HS4,65	9	TGAAGCCAATGGAAAGAGAG	ACAGGAGCATCTAAACCTTGG	650-329	60	1,5
8	HS4,75	3	CAGCATTACATACAATAGTTAGGAGC	GTGATATTTGTCTTTCTGTACCTGG	520-194	60	1,5
9	PV92	16	AACTGGGAAAATTTGAAGAAAGT	TGAGTTCTCAACTCCTGTGTGTTAG	437-122	55	1,5
10	Sb22777/Sb19.12	19	TTAACATCCCTGCAACCCATC	GATTATAGTCACCCTGTTGTGC	761-448	65	1,5
11	Sb23467/Sb19.3	19	TCTAGGCCCAGATTTATGGTAACTG	AAGCACAATTGGTTATTTTCTGAC	456-135	60	1,5
12	TPA25	8	GTAAGAGTTCCGTAACAGGACAGCT	CCCCACCCTAGGAGAACTTCTCTTT	424-113	60	1,5
13	Ya5NBC102	17	TCCATTTCTCTAGACCTGCTG	CCCATAACAGGTCTTCATATTTCC	483-194	55	1,5

# <i>Alu</i>	Nome	# Cr.	Iniciadores (todos na orientação 5'-3')		Tamanho (pb)	T°	[] mM MgCl ₂
			Forward	Reverse			
14	Ya5NBC120	17	GGACCACATGACTGAGTGTAAGT	GAGGTGGCCTCTTAACCATAATTC	518-199	55	1,5
15	Ya5NBC123	5	ATCAAGTTGACACTCAGTATTCACCAC	CTAGTCTGCAGAACTGTGAGAAATGTA	490-186	55	1,5
16	Ya5NBC132	6	CTCGTGATTCACAGAAGTGTGTAAG	CGGGGTTTCATCCTTAATACATACAT	459-139	60	1,5
17	Ya5NBC135	11	ATTAAGCTCATGGTAACCAGCAC	GACTCTCCTCTCTGGATTAGAAACAG	436-117	60	1,5
18	Ya5NBC147	12	TAGCTGGGGGAGGTAGATAATAAAC	AAATATCACCTTATCAGTGGGACCT	493-155	60	1,5
19	Ya5NBC148	20	ACAAGATGACAGATGTAAACCCAAC	AAGGTGTTGTCAGACTAATCTATCG	505-193	60	1,5
20	Ya5NBC150	19	AAATGGAGACACAGAGGTGTAAAGA	CCCAAACCTGCATATTTAAAGGGTAG	491-169	60	1,5
21	Ya5NBC157	17	CATACGTAAATCACTCGGTACTIONCA	TCAGAAAAGTATACAGGTGATGTGC	516-207	60	1,5
22	Ya5NBC159	1	GAGGGTCTTTCTAGGTTTTGTTT	ATGTCAGGTGTCCCTTATGGAGTAT	488-166	60	1,5
23	Ya5NBC171	6	TCTAGAATTACAAGTGCAAGCCATC	CTTCTCATCCCTGCTAACATAACAT	451-130	55	1,5
24	Ya5NBC182	7	GAAGGACTATGTAGTTGCAGAAGC	AACCCAGTGGAAACAGAAGATG	563-287	60	1,5
25	Ya5NBC208	14	AATACCTTGTACATCTTCACCCCTA	TCTCTCTGCTGCACAGTTTGTT	441-115	50	1,5
26	Ya5NBC212	7	CATTTGGCGCAAGTGGTATT	ATCCCAAAGAAACCCACGA	502-190	60	1,5
27	Ya5NBC216	7	GATGTGACCCTGGCTTGTAAG	CAGAGTCCCTGTGCAAAATG	456-141	55	1,5
28	Ya5NBC221	22	CAGTTTTCCATATACATGTGGGTTC	TAGTGTTAAGAGGCCCATTTTCTAC	640-313	60	1,5
29	Ya5NBC237	1	CCCATGGAGGGTCTTTCTA	CTGGAAACCATCCTTCACAGT	410-88	60	1,5

# <i>Alu</i>	Nome	# Cr.	Iniciadores (todos na orientação 5'-3')		Tamanho (pb)	T°	[] mM MgCl ₂
			Forward	Reverse			
30	Ya5NBC239	9	CAGCTGAGAAGTGTACAAATAGAA	ATCAATGACTGACTTGTGCTGAGT	531-198	55	1,5
31	Ya5NBC241	15	GGTTCCAATAGAGAGCAACAGAA	ACCTTAAGCTTTCCCCCAGA	392-66	55	1,5
32	Ya5NBC242	7	AACAAAATTCCCTTTCCTCCA	GGCAATCTGACCTTGGGTAA	503-192	55	1,5
33	Ya5NBC27	5	CTGAATACAGGTATCACTGAACAGAAC	ACAGTGTAAGTCTAACCTACCAGAGGAT	591-265	55	1,5
34	Ya5NBC311	5	TCTTGGAAGGAGATGTGAA	AATCACATCCGAGGGTGTCT	584-279	60	1,5
35	Ya5NBC327	6	AGGCAGGTTCAATGTTCAA	TTGTCTTATTGTGCTGGCTAGA	668-339	60	1,5
36	Ya5NBC333	14	GGCATGCTATCATTCCCAA	CCAAACTTCTGTTTGAGAGAATACG	588-281	60	1,5
37	Ya5NBC335	20	TGGGTAAGGAAATATGTGCTGCTTTAG	TTCACAGCATTAGAGAGAGTTGATG	361-128	60	1,5
38	Ya5NBC345	3	GCCATGAGAGTGGTCAGCAT	AGTCTCCACCATCTCTGCTGT	391-94	60	1,5
39	Ya5NBC347	1	CATGCCCATGCTTTACGTT	TGGGGTAGATGGACTCATCC	465-140	60	1,5
40	Ya5NBC351	21	TTCCTCCCCTTTTCCTGTT	TGTCAGTATGTAAACCCATGCT	437-123	55	1,5
41	Ya5NBC354	7	GTAGCTTGGCCTGTGCTCTT	CCTCTGGGCTGAGAACTCTT	466-148	65	1,5
42	Ya5NBC45	20	TAGGGTAAGGAAATATGTGCTGCTTTAG	GTCTCTGAACGACTATGTGAGCAG	411-98	60	1,5
43	Ya5NBC51	3	ATATTCCAGAAGTTTCTTACATCTAGTGC	AAAGCTTTAAGTCTCCACCATCTCT	437-140	60	1,5
44	Ya5NBC54	6	GTTTATGTCAGTAGGAGTTTTCTCGTGTAG	TCATTGTATCATCTGCTGTACCTGT	433-130	60	1,5
45	Ya5NBC61	9	TGAAATAATCCAGTTGGGGAAG	GTATATCTCTACCGAGACTCAGTTTTTAGC	493-180	55	1,5

# <i>Alu</i>	Nome	# Cr.	Iniciadores (todos na orientação 5'-3')		Tamanho (pb)	T°	[] mM MgCl ₂
			Forward	Reverse			
46	Ya5NBC96	7	TAGATGAGATAGAGCCATCAAACACTC	GTACCCTGTGAGAAAATATTAGGAGCTATG	509-169	60	4,0
47	Yb8NBC106	21	TCACAGCACAATTCACAACACTG	CTGGGTTGCATTTTCATGGTA	558-233	60	1,5
48	Yb8NBC120	22	CAGTGGATCTCCATTTTACCTCTC	GGAAAGGTTTCAGGAAGAAAGTG	466-109	60	1,5
49	Yb8NBC125	22	AGCCAGAAACCCTGAACAAG	AAAGGCCCCAGAAGTATACCA	415-97	60	1,5
50	Yb8NBC13	16	TCTGGGTTTCTCTGGTGGAC	CTGGCAAATGCTACCCAAGT	510-168	60	1,5
51	Yb8NBC146	16	CTCTTCTCTCCAGGAAACGTC	GGAGCTCTGCCTTACACTCAA	887-592	60	1,5
52	Yb8NBC148	14	CCAGGCCTCCATCTTTGATA	TCACTTTTGGGCATGTCAAG	537-218	60	1,5
53	Yb8NBC157	10	TATGGTTCTCAGCCATCACG	ATTCTTCCCCAAAGGGAGTC	712-423	60	3,0
54	Yb8NBC181	3	CATGTACCTTAGAATTCACACTCTCA	CCCCAAAGTTTATAGTCTGTTGTCT	487-151	55	1,5
55	Yb8NBC192	7	CTGCTCTACCCTAGGCTCTTCTATC	GCTCCTCTGCTTTTATGTGTTCTAC	423-132	55	1,5
56	Yb8NBC201	7	GGAGAAAATGTAAGGTTTCTAGCAC	ACCAATGCAACTATCTACACTGACA	476-145	60	1,5
57	Yb8NBC207	7	GTAATATGAGGTGATGGGGGTTACT	GGTGAAAGAAGAACCCCTAAGTTAT	474-138	60	1,5
58	Yb8NBC227	17	AAGAAAAGGGAAGCCTGGAG	CAGTCATCACCAGCCATGAG	881-546	60	3,0
59	Yb8NBC237	7	GCCAAAATCAACTGCCAAAC	TGCTGAGGATAGAGCTATAGCAGA	491-164	60	1,5
60	Yb8NBC243	16	GAACCCCATCCATTCTCTTACA	GTGGCAAATATTGGCGACT	508-156	60	1,5
61	Yb8NBC405	3	GCCCATCCCCTATTATAGCC	ACCAAACCCCATGACACTA	463-155	60	1,5

# <i>Alu</i>	Nome	# Cr.	Iniciadores (todos na orientação 5'-3')		Tamanho (pb)	T°	[] mM MgCl ₂
			Forward	Reverse			
62	Yb8NBC412	7	CAAAGATGGTTGTTGAGGTTGA	CCCAGCAACTTCCCCTTAAT	500-185	60	1,5
63	Yb8NBC419	5	CATCTCCTGGCAACTGAG	ACAAAGCAAGGGTATTTACAGC	572-241	60	3,0
64	Yb8NBC420	2	AAATGCCCAAGTTTCATTGC	AACTGCCACAGCGATTCTTT	430-103	60	1,5
65	Yb8NBC435	20	TGAATGATTGGGACTGGGTA	TGGCTGGATGAACTTTCACA	444-113	55	1,5
66	Yb8NBC437	13	GGCGGTGATGGTAAAACAAC	CTTCCCCAAGGAGCCTTTTA	498-183	60	1,5
67	Yb8NBC441	6	CTCCTGGCATGTCTTCAGGT	TCTCAGCCTAGACCAATACCAA	488-160	55	1,5
68	Yb8NBC450	13	TGAAATCTATCTCGTAGGAAGGCTA	CCGCTGGTTACCAAAGATT	500-186	60	1,5
69	Yb8NBC461	1	CCAAAGTCATTCTTCATTCTGC	GACACCCGAAAAGACTAAAGACA	529-205	55	1,5
70	Yb8NBC463	13	GCCAGTGCTTGGGTTTTAGA	CTGGCAATGAATTTCCCTTT	543-220	55	1,5
71	Yb8NBC466	14	TTGAGGCACTAGACTTACAGAATTG	CAGGAGCTGCTTTCACCTCT	427-92	60	1,5
72	Yb8NBC479	20	CATCCTGTTTCAACATCAGCA	GTTCCCAGCAGGAATCTGAG	398-182	60	1,5
73	Yb8NBC480	13	CCTCTCTCACAAACAGTGCAG	TCGCAAGACACAGGCTATCA	457-114	60	1,5
74	Yb8NBC485	1	TGTTCTTGCCAGAAAGTTTGC	CCAATCCAGGACTCGACATT	433-95	55	1,5
75	Yb8NBC49	6	GCAGTGGATTGGTTTTTCTG	GCTGAAAGAGGCATTGAAATC	542-205	60	1,5
76	Yb8NBC5	7	AAGGTCTAAGCGCAGTGGAA	TGTATGCAGGTTGCTTGCTC	503-167	60	1,5
77	Yb8NBC505	5	TGAGCCTATGACTGAGCATGA	GGGGCTCTCATCAGCATTTA	764-427	60	1,5

# <i>Alu</i>	Nome	# Cr.	Iniciadores (todos na orientação 5'-3')		Tamanho (pb)	T°	[] mM MgCl ₂
			Forward	Reverse			
78	Yb8NBC516	20	GGGCTCAGGGATACTATGCTC	GCCTAGGCCTACCACTCAGA	445-124	60	1,5
79	Yb8NBC547	5	GCCCATGCTCAGTCTAAACC	GATTGGAGCCCTTGTCTACG	413-238	60	1,5
80	Yb8NBC568	6	AAACCCAACAAATGTGCTTC	GGCAACCTACACAAAGCATGT	501-213	60	1,5
81	Yb8NBC576	6	GGGAECTAAGTGGGCAAA	GTCATGTACTAAGGTATGCAAAAG	475-149	55	1,5
82	Yb8NBC585	9	CATTGGGTTAACATTCGCTCT	CACGTGTGCAGCAATGTATG	585-260	60	1,5
83	Yb8NBC589	6	AGTCTTAATGGGCGCTGAGA	AGTGCCTCACCCAGTAGCAC	472-154	60	1,5
84	Yb8NBC596	1	TCCAGGGCCAAGTAGTGAAT	CTGCCCAAATGCTTACT	543-203	55	1,5
85	Yb8NBC597	5	TGAGGTGTTGCAGACGATGT	CGCATGCTTTAGAGAATACCC	429-108	60	1,5
86	Yb8NBC598	19	TGGGTCCTATCATCCCCTATC	CCAGAAGGCATCTCATGGTT	691-420	60	1,5
87	Yb8NBC605	1	GCCTCTAGGTGGAGCCCTTT	GCCCCATTTATGGCTGTTTA	723-392	60	1,5
88	Yb8NBC622	11	TCAAACTTGCGGATTTTCC	TGCTGAGCTATACTGGTGCAA	451-357	60	1,5
89	Yb8NBC636	1	CCTCTGGCAAGCTGCTTAAT	TCACAGCTAGAGGAGACATGAAA	498-169	60	1,5
90	Yb8NBC65	6	ATCTCATCTCCCTGCCTCTG	GGGAGGTCTGGAGATCTGTG	517-186	60	1,5
91	Yb8NBC77	12	CGGAATGTTCTGAGGATCAAA	GGAAGCTCTGCACAACCTCCTA	547-218	55	1,5
92	Yb8NBC80	18	ATTTACAGTGCCCTGTCCT	TCCAGGCAGATGAATTGACA	456-123	60	3,0
93	Yb8NBC93	20	AAGTGAGTCCCAGGGCCTTCT	CACACAGGCACTTGTGGT	518-189	60	1,5

# <i>Alu</i>	Nome	# Cr.	Iniciadores (todos na orientação 5'-3')		Tamanho (pb)	T°	[] mM MgCl ₂
			Forward	Reverse			
94	Yb9NBC10	4	GTTTTCTGGTGTGCCCTAAATA	TTTACCTAACTCACAAGACCCAAAG	524-197	60	4,0
95	Yb9NBC50	11	GTTCCACAAGTACAGGAGAAAATGT	GAAGCTCTTTAGGAAACCAAATCTC	460-138	55	1,5
96	Yc1NBC2	7	TCTCTCATGAACATAGATACAAA	CGTGCATTCTTGAGATAAAT	443-115	60	3,0
97	Yc1NBC35	16	CCCATTCTCCATGCCGTGAT	TGCAAGGCATTGGGGATACA	481-162	60	1,5
98	Yc1NBC53	7	AAAGCTATCAACCATGCCAACA	GAAAATGCTATTTTGGGGAATG	505-186	58	1,5
99	Yc1NBC63	6	GGTACTCAGTAACACATCAAGA	AAGCTGGGTGGTGGGTTTAC	502-181	60	1,5
100	Yc1RG68	7	ATGGTGTCCACAAGAACTGAG	GGAAGGCTCCATTATAGGTCTTG	480-166	60	1,5

Role of the host genetic variability in the influenza A virus susceptibility

Ana Carolina Arcanjo^{1*}, Giovanni Mazzocco^{2,3*}, Silviene Fabiana de Oliveira^{1,4}, Dariusz Plewczynski^{2,4,5,6} and Jan P. Radomski^{2,7-11}¹Laboratório de Genética, Departamento de Genética e Morfologia, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF, Brazil; ²Interdisciplinary Center for Mathematical and Computational Modeling, Warsaw University, Warsaw, Poland; ³Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland; ⁴The Jackson Laboratory for Genomic Medicine, c/o University of Connecticut Health Center, Farmington, CT USA; ⁵Yale University, New Haven, CT, USA; ⁶Center for Innovative Research, Medical University of Białystok, Poland; ⁷Institute of Biotechnology and Antibiotics, Warsaw, Poland

The aftermath of influenza infection is determined by a complex set of host-pathogen interactions, where genomic variability on both viral and host sides influences the final outcome. Although there exists large body of literature describing influenza virus variability, only a very small fraction covers the issue of host variance. The goal of this review is to explore the variability of host genes responsible for host-pathogen interactions, paying particular attention to genes responsible for the presence of sialylated glycans in the host endothelial membrane, mucus, genes used by viral immune escape mechanisms, and genes particularly expressed after vaccination, since they are more likely to have a direct influence on the infection outcome.

Key words: Influenza A virus, host-pathogen interactions

Received: 21 May, 2014; revised: 13 August, 2014; accepted: 28 August, 2014; available on-line: 04 September, 2014

INTRODUCTION

Influenza virus is one of the most important cause of infections of the respiratory tract, with 3–5 million clinical infections and 250 000–500 000 fatal cases annually (Datwood *et al.*, 2012; Simonsen *et al.*, 2013; WHO, 2014). Immunity to influenza virus infection has been a research topic for more than 70 years (Andrewes, 1939) due to the relevant impact the illness has on global health care. Although some aspects of the immunological response to influenza are well understood, there are still many open research questions in this field.

The influence of viral genetic variability on infection is undoubtedly the most important and heavily investigated topic (Hatta *et al.*, 2001; Wagner *et al.*, 2002; Uiprasertkul *et al.*, 2005; Gambaryan *et al.*, 2006; Bateman *et al.*, 2008; Nicholls *et al.*, 2008; Das *et al.*, 2010; Jayaraman *et al.*, 2012; Sriwilajaroen & Suzuki, 2012; Thakaramaran *et al.*, 2012; Guarnaccia *et al.*, 2013; Koel *et al.*, 2013; Sun *et al.*, 2013; Thakaramaran *et al.*, 2013; El Moussi *et al.*, 2014; Qi *et al.*, 2014; Su *et al.*, 2014). The high antigenic drift rate observed in the influenza virus is one of the primary reasons why constantly updated seasonal influenza vaccination is recommended. Genomic variability of the influenza virus has been largely investigated over the years, yet the variability of the host remains a sparsely documented topic, despite its crucial impact on the immune response (Wijburg *et al.*, 1997; Matrosovich & Klenk, 2003;

Schmitz *et al.*, 2005; Jayasekera *et al.*, 2007; Koyama *et al.*, 2007; Throsby *et al.*, 2008; Ichinohe *et al.*, 2009; Sabbah *et al.*, 2009; Kreijtz *et al.*, 2011; Zhou *et al.*, 2012; Henn *et al.*, 2013; Hertz *et al.*, 2013; Lin & Brass, 2013) and the course of infection. The goal of this review is to explore genomic variability in the host genes that mediate host-pathogen interactions. Implemented genomic approach focuses on the major gene variants found in the 1000 Genomes Project data (the 1000 Genomes Project Consortium, 2010; Clarke *et al.*, 2012), and is aimed at understanding how much variability is present across different human populations, and at identifying conservation level of such genes. We attempted to predict whether the genetic variants found in large human populations affect the specificity of virus binding and subsequent effectiveness of the immune response. Genes related to virus entry into the host cell are responsible for the production and assembly of Neu5Ac- α 2,6-Gal β 1,4-GlcNAc (human), and the Neu5Ac- α 2,3-Gal β 1,4-GlcNAc (avian) receptors in the host membrane. Both the human and avian influenza A virus receptor proteins, hemagglutinin (HA) and neuraminidase (NA), interact with the above glycans at the beginning of infection. Sialic acid composition and

e-mail: jan@icm.edu.pl

*Both authors contributed equally

Abbreviations: ADCC, antibody-dependent cell-mediated cytotoxicity; AM, alveolar macrophage; APC, antigen-presenting cell; ASC, apoptotic-associated speck-like protein; ASL, airway surface liquid; C1QBp, component 1 binding protein; CARD, caspase activation and recruitment domain; CD32, cell surface receptor; CD55, gene, decay-accelerating factor 55; CMAS, gene, cystidine monophosphate N-acetylneuraminic acid synthase; CMP, cystidine monophosphate; CRD, carbohydrate recognizing domain; CTD, C-terminal domain; CTL, cytotoxic T lymphocyte; DC, dendritic cell; DDX58, gene, encodes RIG-I; DHX58, gene, encodes LGP2; DNAJC3, gene, encodes P58^{IPK}; EIF2AK2, gene, encoded PKR; ER, endoplasmic reticulum; HLA, human leukocyte antigen; human version of the major histocompatibility complex (MHC); IAV, influenza A virus; IFTI, interferon-induced proteins with tetratricopeptide; IFN, Interferon; IgA, immunoglobulin A; IgG, immunoglobulin G; IgM, immunoglobulin M; LGP2, laboratory of genetics and physiology 2 protein; LRR, leucine-rich-repeat; MAVS, mitochondrial antiviral-signaling protein; MUC, gene, mucin; NBD, nucleotide-binding domain; Neu5Ac, sialic acid (also known as neuraminic acid or N-acetylneuraminic acid); NK, natural killer cell; NLR, NOD-like receptor; NOS2, nitric oxide synthase 2; NP, influenza virus nucleoprotein; PAMP, pathogen-associated molecular pattern; PKR, ssRNA-binding protein kinase; PRR, pattern recognition receptor; PYD, pyrin domain; RIG-I, retinoic acid-inducible gene I product; RLR, RIG-I-like receptor; RPAIN, replication protein A interacting protein; SLC35A2, gene, solute carrier family 35 type A2; ST3GAL, gene, sialyltransferase that binds sialic acid to a 2,3 positioned galactose; ST6GAL, gene, sialyltransferase that binds sialic acid to a 2,6 positioned galactose; TAP, transporter of antigen processing; TCR, T-cell receptor; TLR, toll-like receptor; TNF, tumor necrosis factor; TRIM25, tripartite-motif-containing protein 25; VISA, gene, encodes MAVS.

Apêndice I.1 Amostras selecionadas para amplificação de genoma total devido à pequena quantidade de DNA disponível para caracterização dos 100 *Alus*.

Brasília		Kalunga	
Amostra	Quantidade inicial de DNA (µg)	Amostra	Quantidade inicial de DNA (µg)
BSB002	0,63	4A003	0,2835
BSB004	0,5745	4A004	0,162
BSB005	0,5625	4A008	0,3945
BSB006	0,513	4A009	0,2655
BSB007	1,6425	4A015	0,036
BSB010	0,489	4A017	0,011
BSB012	0,387	4A019	0,033
BSB013	0,672	4A023	0,525
BSB015	0,699	4A024	0,273
BSB016	0,0555	4A025	0,228
BSB018	0,363	4A029	0,0465
BSB022	0,5085	4A035	0,2535
BSB024	1,0005	4A038	0,021
BSB027	4,3695	4A040	0,9285
BSB035	0,423	4A044	1,2135
BSB040	0,213	4A055	0,2625
BSB043	0,351	4A058	0,009
BSB044	0,378	4A061	0,5055
BSB046	0,285	4A062	0,111
BSB047	0,2805	4A065	2,975
BSB052	0,8055	4A076	0,036
BSB054	0,3	4A080	0,1215
BSB062	0,387	4A081	0,012

Brasília		Kalunga	
Amostra	Quantidade inicial de DNA (µg)	Amostra	Quantidade inicial de DNA (µg)
BSB067	0,3945	4A082	0,6255
BSB070	0,318	4A086	0,7215
BSB071	0,81	4A087	0,207
BSB077	0,4875	4A088	0,2085
BSB081	0,228	4A089	10,4
BSB085	0,576	4A090	0,6405
BSB089	0,3465	4A092	0,2775
BSB092	4,08	4A096	0,2145
BSB093	0,877	4A093	0,3405
BSB096	0,585	4A098	0,699
BSB098	0,264	4B106	0,636
BSB101	0,444	4B108	0,0555
BSB102	0,288	4B109	0,1905
BSB105	0,108	4B110	1,5885
BSB108	0,2535	4B114	0,102
BSB120	0,247	4B115	0,24
BSB121	0,885	4B116	0,1785
BSB124	0,147	4B119	0,312
BSB142	0,327	4B122	0,306
BSB143	0,36	4B123	0,126
BSB160	1,149	4B129	2,15
BSB161	1,3725	4B130	0,0135
BSB165	0,831	4B132	0,336
BSB166	0,642	4B134	0,2385
BSB172	0,3495	4B135	0,4305
BSB173	0,357	4B136	0,318

Brasília		Kalunga	
Amostra	Quantidade inicial de DNA (µg)	Amostra	Quantidade inicial de DNA (µg)
BSB174	1,05	4B137	0,1125
BSB177	0,462	4B139	0,339
		4B140	0,0915
		4B141	0,606
		4B142	0,243
		4B143	0,525
		4B144	0,249
		4B158	1,525
		4B172	1,35
		4B180	1,55
		4B181	1,475
		4B189	2,1
		4B194	3,95
		4C203	5,475
		4C204	3,35

Apêndice I.2. P-valor para diferenciação populacional entre as populações ameríndias e asiáticas para cada inserção *Alu* cujos dados estavam disponíveis. As células em cinza mostram diferenciação populacional significativa, o desvio-padrão da estimativa está entre parênteses.

ACE												
	Ache	Cinta Larga	Gavião	Guarani	Caingangue	Quéchua A	Quéchua QT	Surui	Waiwai	Xavante	Ianomâmi	Zoro
Vietnã	0,00000 (0,0000)	0,09009 (0,0192)	0,00901 (0,0091)	0,00901 (0,0091)	0,18919 (0,0370)	0,09910 (0,0286)	0,99099 (0,0030)	0,05405 (0,0201)	0,00901 (0,0091)	0,42342 (0,0636)	0,40541 (0,0204)	0,00000 (0,0000)
Ásia Misto	0,00000 (0,0000)	0,33333 (0,0305)	0,00901 (0,0091)	0,07207 (0,0264)	0,54955 (0,0670)	0,23423 (0,0364)	0,70270 (0,0504)	0,09910 (0,0316)	0,00000 (0,0000)	0,72072 (0,0304)	0,55856 (0,0539)	0,00000 (0,0000)
Camboja	0,15315 (0,0385)	0,27928 (0,0370)	0,99099 (0,0030)	0,71171 (0,0533)	0,04505 (0,0152)	0,40541 (0,0434)	0,00901 (0,0091)	0,60360 (0,0592)	0,99099 (0,0030)	0,06306 (0,0194)	0,08108 (0,0316)	0,99099 (0,0030)
Malásia	0,99099 (0,0030)	0,51351 (0,0651)	0,99099 (0,0030)	0,54955 (0,0345)	0,34234 (0,0402)	0,57658 (0,0474)	0,18018 (0,0332)	0,61261 (0,0354)	0,99099 (0,0030)	0,28829 (0,0297)	0,31532 (0,0546)	0,99099 (0,0030)
China	0,00000 (0,0000)	0,01802 (0,0121)	0,00000 (0,0000)	0,00000 (0,0000)	0,03604 (0,0148)	0,00901 (0,0091)	0,39640 (0,0454)	0,01802 (0,0121)	0,00000 (0,0000)	0,08108 (0,0316)	0,05405 (0,0201)	0,00000 (0,0000)
Japão	0,00000 (0,0000)	0,01802 (0,0121)	0,00000 (0,0000)	0,00901 (0,0091)	0,01802 (0,0121)	0,00901 (0,0091)	0,47748 (0,0592)	0,00000 (0,0000)	0,00000 (0,0000)	0,09910 (0,0316)	0,13514 (0,0279)	0,00000 (0,0000)
SEA	0,00000 (0,0000)	0,04505 (0,0203)	0,00000 (0,0000)	0,00901 (0,0091)	0,08108 (0,0286)	0,02703 (0,0194)	0,80180 (0,0395)	0,02703 (0,0139)	0,00000 (0,0000)	0,16216 (0,0297)	0,17117 (0,0316)	0,00000 (0,0000)
Taiwan	0,00000 (0,0000)	0,16216 (0,0326)	0,01802 (0,0121)	0,03604 (0,0201)	0,17117 (0,0316)	0,11712 (0,0194)	0,99099 (0,0030)	0,04505 (0,0244)	0,00901 (0,0091)	0,35135 (0,0286)	0,37838 (0,0402)	0,00000 (0,0000)
Nigéria	0,00000 (0,0000)	0,41441 (0,0454)	0,03604 (0,0201)	0,14414 (0,0364)	0,48649 (0,0309)	0,26126 (0,0533)	0,99099 (0,0030)	0,12613 (0,0149)	0,00901 (0,0091)	0,74775 (0,0451)	0,40541 (0,0455)	0,00901 (0,0091)
Pigmeus	0,03604 (0,0148)	0,70270 (0,0504)	0,06306 (0,0237)	0,35135 (0,0478)	0,71171 (0,0345)	0,39640 (0,0493)	0,99099 (0,0030)	0,33333 (0,0408)	0,04505 (0,0152)	0,99099 (0,0030)	0,99099 (0,0030)	0,05405 (0,0148)
APO												
	Ache	Cinta Larga	Gavião	Guarani	Caingangue	Quéchua A	Quéchua QT	Surui	Waiwai	Xavante	Ianomâmi	Zoro
Vietnã	0.21622	0.99099	0.50450	0.99099	0.62162	0.47748	0.99099	0.57658	0.53153	0.22523	0.47748	0.99099

	(0.0454)	(0.0030)	(0.0388)	(0.0030)	(0.0479)	(0.0651)	(0.0030)	(0.0364)	(0.0286)	(0.0279)	(0.0651)	(0.0030)
Ásia Misto	0.00000 (0.0000)	0.33333 (0.0385)	0.13514 (0.0311)	0.43243 (0.0265)	0.18018 (0.0235)	0.17117 (0.0478)	0.39640 (0.0309)	0.09910 (0.0316)	0.09910 (0.0286)	0.05405 (0.0242)	0.14414 (0.0492)	0.36036 (0.0450)
Camboja	0.00000 (0.0000)	0.10811 (0.0297)	0.01802 (0.0182)	0.14414 (0.0309)	0.06306 (0.0305)	0.01802 (0.0182)	0.10811 (0.0297)	0.02703 (0.0139)	0.00901 (0.0091)	0.00901 (0.0091)	0.01802 (0.0121)	0.09009 (0.0235)
Malásia	0.08108 (0.0286)	0.29730 (0.0508)	0.14414 (0.0278)	0.44144 (0.0344)	0.30631 (0.0278)	0.19820 (0.0402)	0.37838 (0.0504)	0.17117 (0.0286)	0.19820 (0.0402)	0.09910 (0.0212)	0.12613 (0.0388)	0.46847 (0.0182)
China	0.07207 (0.0227)	0.61261 (0.0485)	0.21622 (0.0388)	0.63063 (0.0359)	0.28829 (0.0297)	0.19820 (0.0297)	0.52252 (0.0402)	0.20721 (0.0451)	0.17117 (0.0344)	0.09009 (0.0271)	0.21622 (0.0278)	0.63964 (0.0459)
Japão	0.00901 (0.0091)	0.11712 (0.0237)	0.03604 (0.0148)	0.11712 (0.0385)	0.05405 (0.0242)	0.02703 (0.0139)	0.12613 (0.0337)	0.01802 (0.0121)	0.03604 (0.0148)	0.02703 (0.0139)	0.04505 (0.0203)	0.16216 (0.0264)
SEA	0.00000 (0.0000)	0.12613 (0.0201)	0.01802 (0.0121)	0.09009 (0.0332)	0.04505 (0.0203)	0.02703 (0.0194)	0.13514 (0.0412)	0.00901 (0.0091)	0.02703 (0.0139)	0.02703 (0.0139)	0.02703 (0.0139)	0.09910 (0.0370)
Taiwan	0.11712 (0.0237)	0.99099 (0.0030)	0.52252 (0.0522)	0.99099 (0.0030)	0.52252 (0.0327)	0.53153 (0.0317)	0.99099 (0.0030)	0.46847 (0.0354)	0.54955 (0.0550)	0.27928 (0.0478)	0.45946 (0.0438)	0.99099 (0.0030)
Nigéria	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.00901 (0.0091)	0.02703 (0.0139)	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.00901 (0.0091)
Pigmeus	0.29730 (0.0273)	0.99099 (0.0030)	0.27027 (0.0332)	0.99099 (0.0030)	0.44144 (0.0417)	0.35135 (0.0417)	0.99099 (0.0030)	0.30631 (0.0364)	0.25225 (0.0297)	0.20721 (0.0430)	0.21622 (0.0309)	0.51351 (0.0273)

HS4.32

	Ache	Cinta Larga	Gavião	Guarani	Caingangue	Quéchua A	Quéchua QT	Surui	Waiwai	Xavante	Ianomâmi	Zoro
Vietnã	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.00901 (0.0091)	0.16216 (0.0353)	0.06306 (0.0194)	0.00901 (0.0091)	0.00000 (0.0000)	0.00000 (0.0000)	0.01802 (0.0121)	0.00901 (0.0091)
Ásia Misto	0.12613 (0.0364)	0.02703 (0.0194)	0.03604 (0.0201)	0.03604 (0.0201)	0.15315 (0.0333)	0.73874 (0.0354)	0.67568 (0.0311)	0.11712 (0.0360)	0.00000 (0.0000)	0.31532 (0.0455)	0.36937 (0.0317)	0.23423 (0.0388)
Camboja	0.00000 (0.0000)	0.00901 (0.0091)	0.00000 (0.0000)	0.00000 (0.0000)	0.01802 (0.0182)	0.09910 (0.0417)	0.09910 (0.0316)	0.01802 (0.0121)	0.00000 (0.0000)	0.01802 (0.0121)	0.04505 (0.0203)	0.01802 (0.0121)
Malásia	0.14414 (0.0242)	0.02703 (0.0194)	0.10811 (0.0353)	0.09009 (0.0235)	0.28829 (0.0264)	0.66667 (0.0338)	0.64865 (0.0379)	0.31532 (0.0493)	0.03604 (0.0201)	0.23423 (0.0411)	0.35135 (0.0394)	0.36937 (0.0566)
China	0.06306 (0.0273)	0.01802 (0.0121)	0.00901 (0.0091)	0.00901 (0.0091)	0.05405 (0.0148)	0.45946 (0.0515)	0.27027 (0.0429)	0.06306 (0.0237)	0.00000 (0.0000)	0.10811 (0.0326)	0.19820 (0.0379)	0.11712 (0.0305)

Japão	0.04505 (0.0203)	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.10811 (0.0264)	0.39640 (0.0562)	0.23423 (0.0278)	0.10811 (0.0121)	0.00901 (0.0091)	0.17117 (0.0286)	0.18919 (0.0438)	0.06306 (0.0237)
SEA	0.34234 (0.0424)	0.12613 (0.0309)	0.23423 (0.0243)	0.21622 (0.0201)	0.80180 (0.0395)	0.67568 (0.0563)	0.99099 (0.0030)	0.54054 (0.0272)	0.08108 (0.0316)	0.77477 (0.0389)	0.76577 (0.0311)	0.72973 (0.0287)
Taiwan	0.04505 (0.0152)	0.00000 (0.0000)	0.00901 (0.0091)	0.00901 (0.0091)	0.09009 (0.0271)	0.40541 (0.0546)	0.24324 (0.0305)	0.05405 (0.0201)	0.00000 (0.0000)	0.12613 (0.0388)	0.13514 (0.0412)	0.12613 (0.0309)
Nigéria	0.03604 (0.0148)	0.00901 (0.0091)	0.00901 (0.0091)	0.01802 (0.0121)	0.09009 (0.0303)	0.31532 (0.0389)	0.18018 (0.0407)	0.06306 (0.0194)	0.00000 (0.0000)	0.04505 (0.0152)	0.08108 (0.0163)	0.04505 (0.0203)
Pigmeus	0.64865 (0.0327)	0.19820 (0.0402)	0.32432 (0.0411)	0.39640 (0.0562)	0.57658 (0.0594)	0.99099 (0.0030)	0.99099 (0.0030)	0.68468 (0.0474)	0.08108 (0.0370)	0.70270 (0.0327)	0.99099 (0.0030)	0.63063 (0.0489)

HS4.65

	Ache	Cinta Larga	Gavião	Guarani	Caingangue	Quéchua A	Quéchua QT	Surui	Waiwai	Xavante	Ianomâmi	Zoro
Vietnã	0.99099 (0.0030)	0.52252 (0.0297)	0.02703 (0.0139)	0.99099 (0.0030)	0.04505 (0.0203)	0.99099 (0.0030)	0.18919 (0.0370)	0.12613 (0.0309)	0.45946 (0.0515)	0.02703 (0.0139)	0.99099 (0.0030)	0.99099 (0.0030)
Ásia Misto	0.08108 (0.0212)	0.54054 (0.0450)	0.33333 (0.0451)	0.35135 (0.0317)	0.72072 (0.0407)	0.19820 (0.0379)	0.99099 (0.0030)	0.99099 (0.0030)	0.69369 (0.0526)	0.55856 (0.0522)	0.33333 (0.0592)	0.58559 (0.0434)
Camboja	0.37838 (0.0402)	0.99099 (0.0030)	0.09009 (0.0303)	0.99099 (0.0030)	0.18018 (0.0429)	0.99099 (0.0030)	0.50450 (0.0243)	0.58559 (0.0434)	0.99099 (0.0030)	0.17117 (0.0316)	0.99099 (0.0030)	0.99099 (0.0030)
Malásia	0.99099 (0.0030)	0.99099 (0.0030)	0.36036 (0.0489)	0.99099 (0.0030)	0.54054 (0.0542)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.57658 (0.0310)	0.99099 (0.0030)	0.99099 (0.0030)
China	0.00000 (0.0000)	0.02703 (0.0194)	0.50450 (0.0243)	0.00000 (0.0000)	0.11712 (0.0273)	0.00901 (0.0091)	0.08108 (0.0212)	0.08108 (0.0286)	0.03604 (0.0278)	0.29730 (0.0451)	0.00000 (0.0000)	0.00000 (0.0000)
Japão	0.01802 (0.0121)	0.32432 (0.0364)	0.52252 (0.0618)	0.16216 (0.0353)	0.99099 (0.0030)	0.09009 (0.0235)	0.63964 (0.0598)	0.54054 (0.0606)	0.39640 (0.0278)	0.71171 (0.0417)	0.07207 (0.0297)	0.34234 (0.0379)
SEA	0.01802 (0.0121)	0.17117 (0.0286)	0.67568 (0.0455)	0.18018 (0.0407)	0.99099 (0.0030)	0.03604 (0.0148)	0.72973 (0.0253)	0.45045 (0.0678)	0.22523 (0.0311)	0.99099 (0.0030)	0.04505 (0.0152)	0.14414 (0.0309)
Taiwan	0.43243 (0.0633)	0.99099 (0.0030)	0.05405 (0.0201)	0.99099 (0.0030)	0.1621 (0.0297)	0.99099 (0.0030)	0.59459 (0.0411)	0.45946 (0.0550)	0.99099 (0.0030)	0.09910 (0.0316)	0.99099 (0.0030)	0.99099 (0.0030)
Nigéria	0.03604 (0.0201)	0.22523 (0.0311)	0.99099 (0.0030)	0.21622 (0.0338)	0.99099 (0.0030)	0.09009 (0.0303)	0.55856 (0.0675)	0.60360 (0.0385)	0.18919 (0.0370)	0.99099 (0.0030)	0.03604 (0.0201)	0.13514 (0.0244)
Pigmeus	0.17117	0.99099	0.69369	0.46847	0.99099	0.24324	0.99099	0.99099	0.99099	0.66667	0.31532	0.34234

(0.0370) (0.0030) (0.0274) (0.0539) (0.0030) (0.0237) (0.0030) (0.0030) (0.0030) (0.0578) (0.0339) (0.0485)

HS4.75

	Ache	Cinta Larga	Gavião	Guarani	Caingangue	Quéchua A	Quéchua QT	Surui	Waiwai	Xavante	Ianomâmi	Zoro
Vietnã	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)
Ásia Misto	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)
Camboja	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)
Malásia	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)
China	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)
Japão	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)
SEA	0.15315 (0.0430)	0.50450 (0.0364)	0.40541 (0.0595)	0.52252 (0.0618)	0.18018 (0.0429)	0.48649 (0.0388)	0.45045 (0.0303)	0.43243 (0.0227)	0.53153 (0.0370)	0.20721 (0.0430)	0.40541 (0.0579)	0.46847 (0.0466)
Taiwan	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)	0.99099 (0.0030)
Nigéria	0.00000 (0.0000)	0.00901 (0.0091)	0.00901 (0.0091)	0.00901 (0.0091)	0.00000 (0.0000)	0.05405 (0.0201)	0.01802 (0.0121)	0.01802 (0.0121)	0.01802 (0.0121)	0.01802 (0.0121)	0.00901 (0.0091)	0.01802 (0.0121)
Pigmeus	0.16216 (0.0353)	0.30631 (0.0364)	0.28829 (0.0264)	0.32432 (0.0411)	0.29730 (0.0360)	0.26126 (0.0478)	0.33333 (0.0526)	0.34234 (0.0424)	0.42342 (0.0490)	0.19820 (0.0297)	0.25225 (0.0485)	0.26126 (0.0344)

PV92

	Ache	Cinta Larga	Gavião	Guarani	Caingangue	Quéchua A	Quéchua QT	Surui	Waiwai	Xavante	Ianomâmi	Zoro
Vietnã	0.53153 (0.0317)	0.01802 (0.0121)	0.99099 (0.0030)	0.33333 (0.0526)	0.40541 (0.0474)	0.09910 (0.0252)	0.19820 (0.0503)	0.61261 (0.0504)	0.99099 (0.0030)	0.39640 (0.0546)	0.58559 (0.0494)	0.71171 (0.0287)
Ásia Misto	0.50450 (0.0433)	0.00000 (0.0000)	0.99099 (0.0030)	0.45045 (0.0606)	0.25225 (0.0445)	0.06306 (0.0194)	0.10811 (0.0402)	0.59459 (0.0433)	0.99099 (0.0030)	0.40541 (0.0434)	0.58559 (0.0475)	0.67568 (0.0530)
Camboja	0.00901	0.00000	0.09009	0.03604	0.02703	0.00000	0.00000	0.42342	0.05405	0.00000	0.52252	0.04505

	(0.0091)	(0.0000)	(0.0271)	(0.0201)	(0.0139)	(0.0000)	(0.0000)	(0.0560)	(0.0242)	(0.0000)	(0.0265)	(0.0152)
Malásia	0.09009 (0.0271)	0.99099 (0.0030)	0.03604 (0.0148)	0.39640 (0.0609)	0.15315 (0.0385)	0.66667 (0.0389)	0.33333 (0.0490)	0.00901 (0.0091)	0.09009 (0.0192)	0.05405 (0.0242)	0.00901 (0.0091)	0.19820 (0.0445)
China	0.52252 (0.0425)	0.00000 (0.0000)	0.99099 (0.0030)	0.23423 (0.0338)	0.35135 (0.0459)	0.06306 (0.0194)	0.11712 (0.0273)	0.64865 (0.0425)	0.99099 (0.00300)	0.44144 (0.0566)	0.57658 (0.0511)	0.69369 (0.0334)
Japão	0.99099 (0.0030)	0.02703 (0.0139)	0.54054 (0.0407)	0.99099 (0.0030)	0.99099 (0.0030)	0.54955 (0.0515)	0.77477 (0.0364)	0.09910 (0.0252)	0.69369 (0.0238)	0.99099 (0.0030)	0.09009 (0.0303)	0.99099 (0.0030)
SEA	0.64865 (0.0379)	0.09910 (0.0286)	0.09910 (0.0370)	0.99099 (0.0030)	0.59459 (0.0388)	0.66667 (0.0364)	0.99099 (0.0030)	0.08108 (0.0286)	0.32432 (0.0364)	0.63063 (0.0384)	0.02703 (0.0194)	0.52252 (0.0327)
Taiwan	0.43243 (0.0522)	0.00000 (0.0000)	0.99099 (0.0030)	0.29730 (0.0333)	0.37838 (0.0466)	0.07207 (0.0378)	0.18919 (0.0370)	0.66667 (0.0412)	0.99099 (0.0030)	0.46847 (0.0446)	0.60360 (0.0360)	0.70270 (0.0539)
Nigéria	0.00901 (0.0091)	0.53153 (0.0533)	0.00901 (0.0091)	0.02703 (0.0139)	0.00901 (0.0091)	0.09910 (0.0212)	0.03604 (0.0201)	0.00000 (0.0000)	0.00901 (0.0091)	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)
Pigmeus	0.03604 (0.0201)	0.99099 (0.0030)	0.00901 (0.009)	0.10811 (0.0227)	0.09009 (0.0303)	0.42342 (0.0273)	0.26126 (0.0417)	0.00000 (0.0000)	0.05405 (0.0201)	0.05405 (0.0148)	0.00000 (0.0000)	0.12613 (0.0309)

TPA25

	Ache	Cinta Larga	Gavião	Guarani	Caingangue	Quéchua A	Quéchua QT	Surui	Waiwai	Xavante	Ianomâmi	Zoro
Vietnã	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.00901 (0.0091)	0.00000 (0.0000)	0.01802 (0.0121)	0.00901 (0.0091)	0.36036 (0.0429)	0.00000 (0.0000)	0.16216 (0.0326)	0.00000 (0.0000)	0.00000 (0.0000)
Ásia Misto	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.00901 (0.0091)	0.03604 (0.0201)	0.00901 (0.0091)	0.44144 (0.0478)	0.00000 (0.0000)	0.32432 (0.0562)	0.01802 (0.0121)	0.00000 (0.0000)
Camboja	0.00000 (0.0000)	0.04505 (0.0203)	0.00901 (0.0091)	0.16216 (0.0326)	0.27027 (0.0383)	0.44144 (0.0394)	0.23423 (0.0454)	0.48649 (0.0338)	0.05405 (0.0201)	0.59459 (0.0364)	0.36937 (0.0394)	0.27027 (0.0429)
Malásia	0.00000 (0.0000)	0.00901 (0.0091)	0.00000 (0.0000)	0.07207 (0.0182)	0.07207 (0.0227)	0.21622 (0.0309)	0.09009 (0.0303)	0.67568 (0.0280)	0.02703 (0.0194)	0.70270 (0.0446)	0.09910 (0.0344)	0.05405 (0.0278)
China	0.00000 (0.0000)	0.00000 (0.0000)	0.00000 (0.0000)	0.09910 (0.0252)	0.12613 (0.0388)	0.28829 (0.0521)	0.13514 (0.0244)	0.78378 (0.0408)	0.00901 (0.0091)	0.99099 (0.0030)	0.12613 (0.0309)	0.15315 (0.0237)
Japão	0.00000 (0.0000)	0.04505 (0.0152)	0.00000 (0.0000)	0.18919 (0.0550)	0.30631 (0.0454)	0.33333 (0.0333)	0.26126 (0.0459)	0.40541 (0.0434)	0.03604 (0.0148)	0.60360 (0.0273)	0.34234 (0.0424)	0.31532 (0.0412)
SEA	0.00000 (0.0000)	0.15315 (0.0139)	0.00901 (0.0091)	0.39640 (0.0562)	0.61261 (0.0425)	0.81081 (0.0489)	0.68468 (0.0529)	0.30631 (0.0309)	0.27928 (0.0459)	0.17117 (0.0459)	0.76577 (0.0494)	0.61261 (0.0618)

Taiwan	0.00000	0.00000	0.00000	0.00000	0.00901	0.03604	0.03604	0.76577	0.01802	0.68468	0.01802	0.03604
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0091)	(0.0201)	(0.0148)	(0.0366)	(0.0182)	(0.0389)	(0.0121)	(0.0201)
Nigéria	0.00000	0.05405	0.00000	0.15315	0.29730	0.40541	0.26126	0.68468	0.05405	0.80180	0.37838	0.21622
	(0.0000)	(0.0242)	(0.0000)	(0.0194)	(0.0360)	(0.0339)	(0.0212)	(0.0579)	(0.0278)	(0.0345)	(0.0504)	(0.0243)
Pigmeus	0.00000	0.00000	0.00000	0.00000	0.00000	0.02703	0.01802	0.37838	0.00901	0.30631	0.04505	0.04505
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0139)	(0.0121)	(0.0504)	(0.0091)	(0.0309)	(0.0203)	(0.0152)