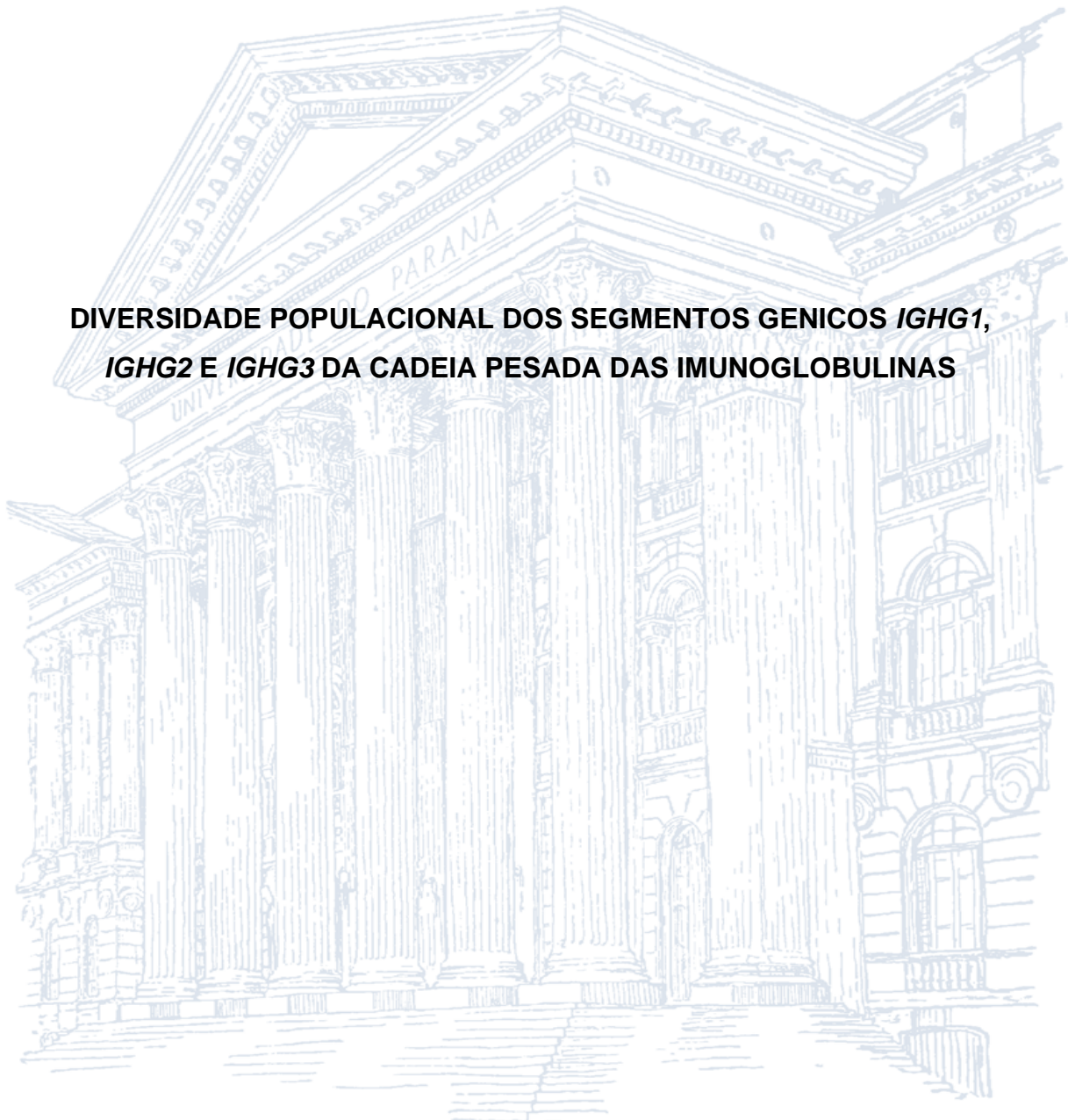


UNIVERSIDADE FEDERAL DO PARANÁ

VERÓNICA CALONGA SOLÍS

**DIVERSIDADE POPULACIONAL DOS SEGMENTOS GENICOS *IGHG1*,
IGHG2 E *IGHG3* DA CADEIA PESADA DAS IMUNOGLOBULINAS**



CURITIBA

2016

UNIVERSIDADE FEDERAL DO PARANÁ

VERÓNICA CALONGA SOLÍS

**DIVERSIDADE POPULACIONAL DOS SEGMENTOS GENICOS *IGHG1*,
IGHG2 E *IGHG3* DA CADEIA PESADA DAS IMUNOGLOBULINAS**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Genética ao Programa de Pós-Graduação em Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná.

Orientador: Dr. Danilo Gardenal Augusto

Coorientadora: Prof^a Dra. Danielle Malheiros Ferreira

CURITIBA

2016

Universidade Federal do Paraná
Sistema de Bibliotecas

Calonga Solís, Verónica

Diversidade populacional dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* da cadeia pesada das imunoglobulinas. / Verónica Calonga Solís.
– Curitiba, 2016.

118f. : il. ; 30cm.

Orientador: Danillo Gardenal Augusto

Coorientadora: Danielle Malheiros Ferreira

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Imunoglobulinas. 2. Genética de populações. I. Título. II. Augusto, Danillo Gardenal. III. Ferreira, Danielle Malheiros. IV. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

CDD (20. ed.) 575.1



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA
CENTRO POLITÉCNICO - JARDIM DAS AMÉRICAS - CAIXA
POSTAL 19071 - CEP 81531-990 - CTBA, BRASIL 41 3361 1587
PPG-GEN@UFPR.BR



PARECER

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Genética da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de VERÓNICA CALONGA SOLÍS, intitulada: "Diversidade populacional dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* da cadeia pesada das imunoglobulinas", após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua aprovação.

Curitiba, 30 de agosto de 2016

Dr. Danilo Gardenal Augusto
Dep. Genética/UFPR – presidente

Dr.ª Marcia Regina Pincerati
Núcleo de Ciências Biológicas e da Saúde/Universidade Positivo

Dr.ª Marcia Holsbach Beltrame
Dep. Genética/Universidade da Pensilvânia/EUA

Dr. Ricardo Lehtonen Rodrigues de Souza
Dep. Genética/UFPR

Dr.ª Danielle Malheiros Ferreira
Dep. Genética/UFPR

Aos meus pais
Magda e Ignacio

AGRADECIMENTOS

Agradeço a professora Luiza Petzl-Erler, minha primeira orientadora, por abrir as portas do laboratório, pelo acompanhamento, apoio e carinho.

Agradeço imensamente aos meus orientadores Danillo Augusto e Danielle Malheiros, que com infinita paciência e amor me guiaram desde o começo deste projeto. Obrigada pela sua amizade, por todas as horas que investiram no meu aprendizado e desculpas pelos apertos com o tempo.

Agradeço as Professoras Angelica Boldt e Karin Braun pela amizade e predisposição em ensinar e ajudar sempre que precisei.

Agradeço aos meus pais e as minhas irmãs, Valeria e Victoria, que estão sempre ao meu lado, me apoiando, me dando força, amor e luz para continuar. E agradeço a minha grande família pelo amor e carinho.

Agradeço a Leila Dominguez e a Fátima Vieira, que me adotaram desde o momento que cheguei a Curitiba.

Agradeço a Renata Dourado e Luciana Vargas, pela grande amizade e pela participação muito valiosa neste projeto; e a Ticiane Farias, pela amizade, carinho e apoio de sempre. Agradeço a equipe do LGMH, e as amigas que nasceram no laboratório: Amanda Salviano, Andressa Matoso, Andressa Valengo, Carolina D'Errico, Eleonora Paulini, Karoliny Cavalheiro, Debora Roepke, Gabriel Cipolla, Gabriela Canalli, Leonardo Amorin, Liana Oliveira, Luana Oliveira, Mariana Castro, Rodrigo Almeida, Sara Alves, Simone Fachin. Obrigada pelas risadas, pelos doces e por tudo que aprendi e aprendemos juntos.

Agradeço aos amigos daqui, que são um grande apoio: Iriel Joerin, Hugo Rojas, Max Maronna, Juliana Jordão, Andrei Strickler. Agradeço as meninas do pensionato pelo seu companheirismo muito valioso e aos meus amigos de outros lugares pelo apoio apesar da distância.

Agradeço ao Valter Baura pela ajuda e assessoramento com o sequenciamento neste projeto. Agradeço as instituições de fomento que fazem possível o desenvolvimento dos nossos projetos de pesquisa. E agradeço ao Brasil, por me receber e fazer sentir confortável, mesmo estando longe de casa.

“Live as if you were to die tomorrow.
Learn as if you were to live forever.”

Mahatma Gandhi

RESUMO

As imunoglobulinas (Ig) são moléculas produzidas a partir de recombinação somática de segmentos gênicos *IGHG* na linhagem germinativa de linfócitos B. São bem conhecidos os mecanismos que geram variabilidade nas cadeias leve e pesadas que formam os domínios de reconhecimento de antígenos. Entretanto, os domínios constantes das cadeias pesadas também exibem grande diversidade, sendo esta conhecida principalmente por técnicas sorológicas que definiram os alótipos de Ig. Já a variabilidade genética dos segmentos gênicos que codificam a cadeia pesada das IgG, os *IGHG*, é pouco conhecida, devido a limitações técnicas produzidas pela elevada similaridade entre os segmentos gênicos que os codificam e, também ao fato que os bancos de dados genômicos foram produzidos principalmente partir de DNA proveniente de linhagens de linfócitos B, que são as células nas quais o rearranjo somático ocorre e, portanto, inadequadas para análise desses segmentos gênicos. O estudo da variabilidade alélica de *IGHG* nas populações humanas pode contribuir para compreender o papel funcional dos polimorfismos, implicação na susceptibilidade a doenças e sua evolução. O objetivo deste trabalho é caracterizar a diversidade genética dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* da cadeia pesada das imunoglobulinas em uma população de descendentes de Japoneses e em cinco populações ameríndias brasileiras: Guarani Nandeva, Guarani Kaiowa, Guarani Mbya, Kaingang de Ivai e Kaingang de Rio das Cobras. Foram encontrados 25 alelos novos nas populações estudadas, quatro deles foram confirmados. Os alelos mais comuns nas populações foram *IGHG1*02*, *IGHG2*03* e *IGHG3*14*. Os alótipos mais comuns em Ameríndios foram: Gm21,26,27,28;17,1;(..) e Gm21,26,27,28;17,1,2;(..), e os mais comuns nos descendentes de Japoneses foram: Gm21;17,1;(..), Gm11,13,15,16;17,1;(..), Gm21;17,1,2;(..) e Gm5,11,13;3,1;23. A partir dos testes de neutralidade, sugere-se que há evidência de que a diversidade observada em *IGHG* seja produto do efeito carona por seleção natural positiva nas regiões flanqueadoras do segmento gênico *IGHG*.

Palavras chave: Segmento gênico de imunoglobulinas, *IGHG*, alelos novos, Guarani, Kaingang, descendentes de Japoneses, genética de populações

ABSTRACT

Immunoglobulins are molecules produced as a result of somatic recombination of the *IGHG* gene segments in the germ line of B cells. The mechanisms that generate variability at the antigen recognition site in the heavy and light chain are well known. Despite the fact that the constant domains also exhibit a high diversity, this variation was mainly characterized by serological methods which defined the Ig allotypes. The genetic variability of the gene segments that encode IgG heavy chains, *IGHG*, is poorly known, due to the technical constraints caused by the high similarity between gene segments, and because this genetic region is not well covered in genomic databases since DNA samples are often extracted from B cell lines, which have undergone rearrangement within the *locus*; therefore, they are not suitable for the study of these gene segments. To study the diversity study of *IGHG* gene segments can contribute to a better understanding of Ig polymorphism and its functional role as well as how they contribute in disease susceptibility. The aim of this study was to characterize the genetic diversity of the *IGHG1*, *IGHG2* and *IGHG3* gene segments in one population of Japanese-descendant and five Brazilian Amerindian populations: Guarani Nandeva, Guarani Kaiowa, Guarani Mbya, Kaingang from Rio das Cobras and Kaingang from Ivai. Twenty five new alleles were found for the *IGHG* gene segments, and four of them have been confirmed. The most frequent alleles are *IGHG1*02*, *IGHG2*03* and *IGHG3*14*. The most common allotypes in Amerindians are: Gm21,26,27,28;17,1;(..) and Gm21,26,27,28;17,1,2;(.), and in Japanese-descendant: Gm21;17,1;(.), Gm11,13,15,16;17,1;(.), Gm21;17,1,2;(.) and Gm5,11,13;3,1;23. Neutrality tests suggest that *IGHG* diversity is a product of genetic hitchhiking due to positive selection in flanking regions of these gene segments.

Key words: Immunoglobulin gene segments, *IGHG*, new alleles, Guarani, Kaingang, Japanese-descendant, population genetics

LISTA DE FIGURAS

Figura 1 Constituição básica de uma molécula de Imunoglobulina G	18
Figura 2 Representação esquemática de Ig.....	19
Figura 3 Localização dos genes <i>IGH</i> no cromossomo 14	20
Figura 4 Disposição dos segmentos gênicos de <i>IGH</i>	21
Figura 5 Diferenciação das células B a partir das células-tronco.....	22
Figura 6 Rearranjo dos genes V(D)J.....	23
Figura 7 Formação dos nucleotídeos P e N.....	24
Figura 8 Mudança de classe por recombinação.....	26
Figura 9 Comparação entre alelos de <i>IGHG3</i>	27
Figura 10 Relação entre os genes de <i>IGHG</i> e região de dobradiça.....	28
Figura 11 Ativação da via clássica do complemento.....	29
Figura 12 Troncos e famílias linguísticas do Brasil	42
Figura 13 Mapa de populações indígenas analisadas	48
Figura 14 Segmentos gênicos sequenciados.....	49
Figura 15 Arvore filogenética do segmento gênico <i>IGHG1</i>	66
Figura 16 Arvore filogenética do segmento gênico <i>IGHG2</i>	67
Figura 17 Arvore filogenética do segmento gênico <i>IGHG3</i>	68
Figura 18 Rede Median-Joining dos alelos do segmento gênico <i>IGHG1</i>	70
Figura 19 Rede Median-Joining dos alelos do segmento gênico <i>IGHG2</i>	70
Figura 20 Rede Median-Joining dos alelos do segmento gênico <i>IGHG3</i>	71
Figura 21 Desequilíbrio de ligação entre SNPs em GW.	72
Figura 22 Desequilíbrio de ligação entre SNPs em GND.....	73
Figura 23 Desequilíbrio de ligação entre SNPs em GRC.....	74
Figura 24 Desequilíbrio de ligação entre SNPs em KIV	75
Figura 25 Desequilíbrio de ligação entre SNPs em Kaingang KRC	76
Figura 26 Desequilíbrio de ligação entre SNPs em JAP.	77
Figura 27 Distância genética entre as populações.....	82
Figura 28 Análise de Componente Principal das populações	82
Figura 29 Distância genética entre todas as populações	83
Figura 30 Análise de componente principal entre todas as populações	83
Figura 31 Análise de componente principal entre indivíduos das populações .	84
Figura A1 Frequências nucleotídica dos SNPs de <i>IGHG1</i>	110

Figura A2	Frequências nucleotídica dos SNPs de <i>IGHG2</i>	111
Figura A3	Frequências nucleotídica dos SNPs de <i>IGHG3</i>	112
Figura A4	Desequilíbrio de ligação na população GWK.....	113
Figura A5	Desequilíbrio de ligação na população GND.	114
Figura A6	Desequilíbrio de ligação na população GRC	115
Figura A7	Desequilíbrio de ligação na população KIV	116
Figura A8	Desequilíbrio de ligação na população KRC.....	117
Figura A9	Desequilíbrio de ligação, na população JAP.....	118

LISTA DE TABELAS

Tabela 1 Nomenclatura dos alótipos de Ig.	32
Tabela 2 Associação entre alótipos de imunoglobulina e diversos fenótipos ...	37
Tabela 3 Correspondência entre alótipos e alelos de Ig.	38
Tabela 4 Dados demográficos de Indigenas no Brasil	41
Tabela 5 Descrição das populações analisadas.	46
Tabela 6 Descrição dos primers da amplificação.	48
Tabela 7: Descrição primers de sequenciamento	51
Tabela 8: Sítios variáveis encontrados.....	57
Tabela 9: Alelos novos identificados em <i>IGHG1</i>	59
Tabela 10 Alelos novos identificados em <i>IGHG2</i>	60
Tabela 11 Alelos novos identificados em <i>IGHG3</i>	61
Tabela 12 Frequência alélica nas populações	63
Tabela 13 Frequência alotípica nas populações	64
Tabela 14 Frequência genotípica nas populações	65
Tabela 15: Frequência haplotípicas nas populações	69
Tabela 16: Diferenciação genética entre as populações em <i>IGHG1</i>	78
Tabela 17 Diferenciação genética entre as populações em <i>IGHG2</i>	79
Tabela 18 Diferenciação genética entre as populações em <i>IGHG3</i>	79
Tabela 19 Diferenciação genética entre as populações.....	79
Tabela 20: Diferenciação genética entre as populações para o G1m.....	80
Tabela 21: Diferenciação genética entre as populações para o G2m.....	80
Tabela 22: Diferenciação genética entre as populações para o G3m.....	81
Tabela 23 Parâmetros de diversidade.....	85
Tabela 24 Teste de neutralidade seletiva.....	87
Tabela 25 Teste de neutralidade seletiva com grupo externo.....	88
Tabela 26 Teste de seleção dN/dS dentro das populações	88
Tabela A1 SNPs com dados extraídos do projeto <i>1000 genomes</i>	109

LISTA DE ABREVIATURAS

Am	Marcador alfa
BCR	Receptores de células B
BEB	Bengalis em Bangladesh
CCDA	Citotoxicidade celular dependente de anticorpo
CDC	Citotoxicidade dependente de complemento
CDR	Regiões determinantes de complementaridade
C_H	Domínios ou segmento constante da cadeia pesada das imunoglobulinas
CSR	Mudança de classe por recombinação
D_H	Segmento de diversidade da cadeia pesada das imunoglobulinas
ELB	Método Excoffier–Laval–Balding
Fab	Fragmentos de união ao antígeno
Fc	Fragmento cristalizável
FcγR	Receptores do fragmento Fc das imunoglobulinas
FIN	Finlandeses em Finlândia
GKW	Guarani Kaiowa
Gm	Marcador gama
GND	Guarani Nandeva
GRC	Guarani Mbya
GWAS	Estudos de associação de genoma total
H	Região de dobradiça
HCMV	Citomegalovírus humano
HLA	Antígeno leucocitário humano
HSV	Herpesvírus humano
Ig	Imunoglobulinas
IGH	Gene da cadeia pesada das imunoglobulinas
IMGT	International ImMunoGeneTics Information System
JAP	Descendentes de Japoneses
J_H	Segmento de junção da cadeia pesada das imunoglobulinas
KIV	Kaingang de Ivai
Km	Marcador Kappa
KRC	Kaingang de Rio das Cobras

MAC	complexo de ataque a membrana
MJ	Median-Joining
NT	Nucleotídeo
PCR	Reação em cadeia da polimerase
PEL	Peruanos em Peru
RAG	Genes ativadores de recombinação
RNA_m	RNA mensageiros
RSS	Sequência sinal de recombinação
SNP	Polimorfismos de nucleotídeo único
SSP	Oligonucleotídeo iniciador de sequência específica
TI	Terras Indígenas
V_H	Domínios ou segmento variável da cadeia pesada das imunoglobulinas
YRI	Iorubás em Nigéria

SUMÁRIO

1 INTRODUÇÃO	16
2 REVISÃO BIBLIOGRÁFICA.....	17
2.1 IMUNOGLOBULINAS.....	17
2.2 ESTRUTURA E REARRANJO DOS GENES DA CADEIA PESADA DE IMUNOGLOBULINAS	20
2.3 MUDANÇA DE CLASSE POR RECOMBINAÇÃO E HIPERMUTAÇÃO SOMÁTICA.....	25
2.4 VARIAÇÃO ESTRUTURAL NA REGIÃO DE DOBRADIÇA DAS IgG3.....	27
2.5 FUNÇÃO DAS IMUNOGLOBULINAS	28
2.6 ALÓTIPOS DE IMUNOGLOBULINAS.....	31
2.7 IMPLICAÇÕES FUNCIONAIS DA VARIABILIDADE GENÉTICA DE GM..	33
2.8 POLIMORFISMOS EM IG E SUSCEPTIBILIDADE A DOENÇAS	34
2.9 CARACTERIZAÇÃO DA VARIABILIDADE DAS IG	36
2.10 POPULAÇÕES AMERÍNDIAS NO BRASIL	40
2.10.1 Kaingang	42
2.10.2 Guarani.....	43
3 JUSTIFICATIVA E HIPÓTESES.....	44
4 OBJETIVOS.....	45
4.1 OBJETIVO GERAL	45
4.2 OBJETIVOS ESPECÍFICOS	45
5 METODOLOGIA	46
5.1 AMOSTRAS POPULACIONAIS	46
5.2 AMPLIFICAÇÃO DOS FRAGMENTOS PELA REAÇÃO EM CADEIA DA POLIMERASE	48
5.3 REAÇÃO DE SEQUENCIAMENTO PELO MÉTODO DE SANGER	50

5.4 LEITURA DAS SEQUÊNCIAS	52
5.5 ANÁLISES ESTATÍSTICAS	52
6 RESULTADOS.....	56
7 DISCUSSÃO	89
8 CONCLUSÕES	99
BIBLIOGRAFIA	100
ANEXO 1.....	109

1 INTRODUÇÃO

As imunoglobulinas (Ig) são moléculas encarregadas de reconhecer antígenos. São produzidas pelas células B e participam da resposta imune humoral. São compostas de duas cadeias pesadas idênticas, e duas cadeias leves idênticas. É ainda composta por domínios variáveis (V), que reconhecem e se ligam aos antígenos, e domínios constantes (C), que especificam funções efetoras.

A cadeia pesada das Ig são codificadas por segmentos gênicos variáveis e constantes que sofrem recombinação somática no gene da cadeia pesada das imunoglobulinas (*IGH*). A diversidade dos domínios constantes foi caracterizada principalmente por métodos sorológicos, enquanto que sua diversidade genética é pouco conhecida. A elevada similaridade entre os segmentos gênicos que os codificam dificulta a elaboração de estratégias de genotipagem. Adicionalmente, os bancos de dados genômicos foram majoritariamente produzidos partir de DNA proveniente de linhagens de linfócitos B, nas quais ocorre recombinação somática nos genes *IGHG*, não sendo adequado o uso do DNA dessas células para o estudo dos genes das imunoglobulinas.

Por isto, existe pouca informação sobre a diversidade genética das IgG nas populações. O objetivo deste trabalho foi caracterizar a diversidade genética dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* da cadeia pesada das imunoglobulinas em indivíduos de cinco populações ameríndias brasileiras e uma população de descendentes de japoneses e verificar se há evidências de fatores evolutivos que estejam atuando sobre eles.

Neste trabalho foram sequenciados os segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* da cadeia pesada das imunoglobulinas nas populações Guarani Kaiowa, Guarani Mbyá, Guarani Nandeva, Kaingang de Rio das Cobras, Kaingang de Ivai, e uma população de descendentes de Japoneses. Posteriormente foram identificadas as variantes nucleotídicas, alelos e alótipos de cada indivíduo, e a partir das suas frequências foram realizadas comparações entre as populações estudadas e trabalhos descritos na literatura. Finalmente, foram realizadas análises estatísticas para verificar a ação de fatores evolutivos moldando a diversidade destes segmentos gênicos.

2 REVISÃO BIBLIOGRÁFICA

2.1 IMUNOGLOBULINAS

O sistema imune conta com um conjunto de moléculas encarregadas de reconhecer antígenos, como patógenos e estruturas não próprias. Dentre estas, incluem-se as moléculas do complexo principal de histocompatibilidade, receptores de células T e as imunoglobulinas. Estas últimas serão o foco deste trabalho.

Chama-se de antígeno uma molécula capaz de unir-se com especificidade a uma molécula de anticorpo ou um receptor de células T. Os anticorpos podem reconhecer como antígeno quase todos os tipos de moléculas biológicas, carboidratos, proteínas, lipídios e ácidos nucleicos, mas nem todos os antígenos podem induzir uma resposta imunológica. Os que possuem esta capacidade são chamados de imunógenos. Somente as macromoléculas são capazes de estimular um linfócito B para que iniciem uma resposta imune humoral. Dessa forma, as imunoglobulinas reconhecem somente a uma porção desta macromolécula, denominada epítipo ou determinante antigênico (MURPHY, 2011; ABBAS et al., 2012).

As imunoglobulinas (Ig) são produzidas pelas células B (ou linfócitos B), que são células importantes do sistema imune adaptativo. As células B originam-se e amadurecem na medula óssea. Posteriormente, circulam pelos sistemas sanguíneo e linfático ou são mantidas nos órgãos linfoides periféricos, onde suas respostas imunes são iniciadas. Quando as Ig se encontram ligadas à membrana plasmática das células B, são chamadas de receptores de células B (BCR, do inglês *B-cell receptor*). Quando as Ig são secretadas aos espaços extracelulares são chamadas de anticorpos, e produzem um tipo de resposta denominada resposta imune humoral (MURPHY, 2011; ABBAS et al., 2012).

As Ig são proteínas compostas de duas cadeias pesadas idênticas, e duas cadeias leves idênticas. A molécula é composta por domínios variáveis (V), que reconhecem e se ligam aos antígenos, e domínios constantes (C) (Figura 1), que especificam funções efetoras, como a ativação do complemento

ou a união a receptores da porção constante das Ig das células efetoras do sistema imune (MURPHY, 2011; ABBAS et al., 2012).

As Ig são classificadas de acordo com sua região constante, em diferentes isotipos ou classes: IgM, IgD, IgG, IgA e IgE. As cadeias pesadas das imunoglobulinas IgD, IgG e IgA possuem um domínio V, uma região C constituída por três domínios CH (*constant heavy*), CH1, CH2 e CH3, e uma região de dobradiça (H, *hinge*) entre CH1 e CH2. A região H é uma região rica em prolina, e contém as pontes dissulfeto que unem as duas cadeias pesadas das moléculas. Já as imunoglobulinas IgM e IgE, além do domínio variável, possuem quatro domínios CH - CH1, CH2, CH3 e CH4 - mas não apresentam nenhuma sequência análoga à região de dobradiça das outras subclasses. Entretanto, as cadeias pesadas se encontram unidas por pontes dissulfeto presente no domínio CH2 (Figura 2) (SCHROEDER; CAVACINI, 2010; MURPHY, 2011; ABBAS et al., 2012).

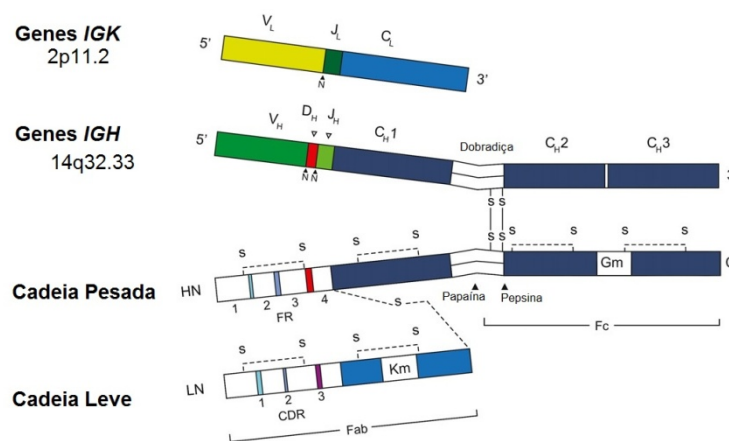


Figura 1. Constituição básica de uma molécula de Imunoglobulina G e sua correspondência com os genes que as codificam. Note-se que as cadeias pesadas estão unidas entre si e às cadeias leves por pontes dissulfeto. A região de dobradiça localiza-se entre CH1 e CH2, e é o sítio de clivagem pelas enzimas papaína e pepsina. Na extremidade N encontra-se o domínio variável, e as regiões determinantes de complementaridade (CDR1, CDR2 e CDR3). Os marcadores Gm encontram-se na cadeia pesada, e os Km na cadeia leve κ . Fonte: SCHROEDER; CAVACINI, 2010

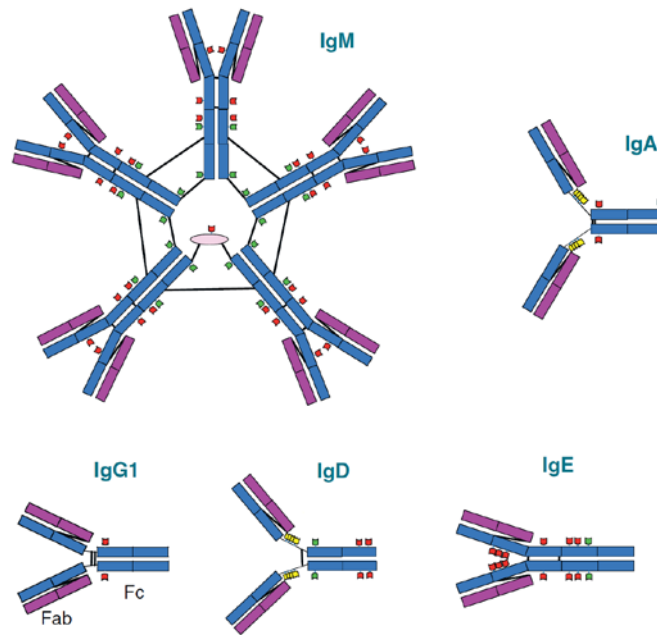


Figura 2: Representação esquemática de Ig mostrando os domínios das diferentes subclasses. IgD, IgG e IgA: domínios V, CH1, H, CH2 e CH3. IgM e IgE: V, CH1, CH2, CH3 e CH4. Na superfície das cadeias pesadas e leves, tanto em Fc como em Fab, encontram-se oligossacarídeos ligados covalentemente. Fonte: Modificado de ARNOLD et al. (2007)

A partir desta unidade básica, algumas Ig podem formar polímeros, assim as IgA secretadas são encontradas formando dímeros e as IgM formando pentâmeros. As cadeias leves podem ser de dois tipos, kappa e lambda (κ e λ) e possuem um domínio V e um C (BEALE; FEINSTEIN, 1976; MURPHY, 2011; ABBAS et al., 2012).

Os primeiros estudos para compreender a estrutura das imunoglobulinas utilizaram enzimas que fragmentam as moléculas de IgG. Por exemplo, a papaína cliva a molécula em dois fragmentos Fab (fragmentos de união ao antígeno, do inglês *fragment antigen binding*), a qual tinha capacidade de ligação a determinantes antigênicos, e em um fragmento Fc (fragmento cristalizável), o qual mostrou pouca afinidade com o antígeno (PORTER, 1959; SCHROEDER; CAVACINI, 2010). Posteriormente, foi descoberto que a porção Fab contém as cadeias leves e as porções V e CH1 das cadeias pesadas, e que Fc contém os fragmentos CH2 e CH3 (Figura 1).

As Ig também são chamadas de glicoproteínas por apresentarem em sua superfície oligossacarídeos ligados covalentemente tanto em Fc como em

Fab, os quais tem um papel crucial na função efetoras destas moléculas (ARNOLD et al., 2007).

2.2 ESTRUTURA E REARRANJO DOS GENES DA CADEIA PESADA DE IMUNOGLOBULINAS

Em humanos, os genes da cadeia pesada das imunoglobulinas (*IGH*) foram mapeados no cromossomo 14q32 (Figura 3). Os genes *IGH* consistem de quatro segmentos: variável (V_H , *variable*), diversidade (D_H , *diversity*), junção (J_H , *joining*) e constante (C_H , *constant*) (CROCE et al., 1979; MCBRIDE et al., 1982).

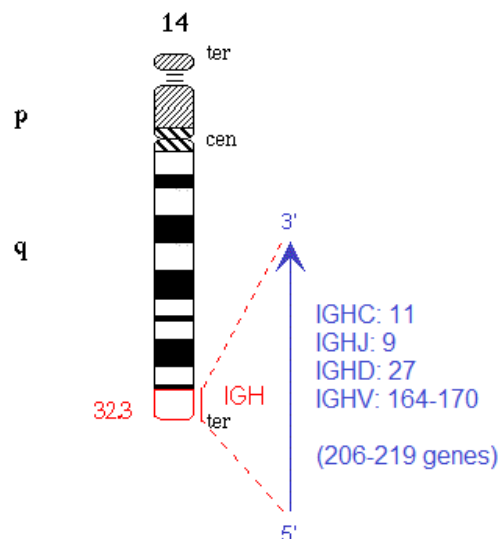


Figura 3: Localização dos genes *IGH* no cromossomo 14q32.3, e a quantidade de genes para os segmentos V, D, J e C. Fonte: Modificado de LEFRANC et al. (2015)

Até o momento foram descritos entre 206 e 219 segmentos gênicos dentro do *locus IGH*, sendo 164-170 segmentos gênicos *IGHV*, 27 *IGHD*, nove *IGHJ* e 11 *IGHC*. Os segmentos gênicos *IGHC* que codificam a região constante das distintas classes de Ig, por sua vez, compreendem nove segmentos gênicos funcionais e dois pseudossegmentos, na seguinte ordem: *IGHM*, *IGHD*, *IGHG3*, *IGHG1*, *IGHEP1*, *IGHA1*, *IGHGP*, *IGHG2*, *IGHG4*, *IGHGE* e *IGHA2*, também referidos em alfabeto grego: mi ($C\mu$), delta ($C\delta$), gama3 ($C\gamma3$), gama1 ($C\gamma1$), pseudo épsilon ($C\psi\epsilon1$), alfa 1 ($C\alpha1$), pseudo gama ($C\psi\gamma$), gama2 ($C\gamma2$), gama4 ($C\gamma4$), épsilon ($C\epsilon$) e alfa2 ($C\alpha2$), respectivamente

(Figura 4) (LEFRANC et al., 2015). Note que o termo pseudossegmento é utilizado neste trabalho para descrever segmentos gênicos que perderam sua função, de maneira análoga ao conceito de pseudogene.

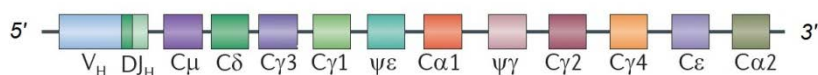


Figura 4: Disposição dos segmentos gênicos de *IGH*. O éxon V(D)J rearranjado encontra-se na extremidade 5', seguido dos segmentos gênicos C. Fonte: Modificado de XU et al. (2012)

A disposição dos segmentos gênicos *IGH* encontrados na linhagem germinativa não permite que estes possam ser transcritos a RNA mensageiros (RNAm) e, portanto não podem codificar imunoglobulinas funcionais. Para tal, o DNA precisa passar por um processo de rearranjo que acontece durante o desenvolvimento das células B. Este processo é chamado de recombinação V(D)J, e acontece tanto nos genes de imunoglobulinas quanto nos genes do receptor de antígenos dos linfócitos T (Revisado em DUDLEY et al., 2005). Um gene de imunoglobulina, como unidade funcional, existe apenas nas linhagens de células B e somente após o processo de recombinação somática. Em outras palavras, não existem genes de imunoglobulinas nas linhagens germinativas. Existem apenas segmentos gênicos que serão rearranjados para originar genes nas linhagens de células B. Por esse motivo, as regiões analisadas neste estudo, *IGHG1*, *IGHG2* e *IGHG3*, muitas vezes chamadas de genes na literatura, são aqui cuidadosamente chamadas de segmentos gênicos. A seguir, o mecanismo de recombinação será explicado em maiores detalhes.

As células pró-B, as primeiras da medula óssea comprometidas com as linhagens das células B, não produzem Ig, e no seu processo de maturação inicia-se a recombinação dentro do *locus* da cadeia pesada das imunoglobulinas (Figura 5). A recombinação requer a seleção aleatória de um dos segmentos gênicos V_H, um dos segmentos D_H, e um J_H em cada linfócito, que logo formarão o éxon V(D)J, codificador da região variável da proteína. Neste éxon V(D)J, o segmento V contém duas regiões determinantes de complementaridade com o antígeno (CDR1 e CDR2). Estas sequências de CDRs são diferentes entre os segmentos gênicos V_H germinais, o que

proporciona a primeira fonte de diversidade entre as moléculas: a diversidade provinda da linhagem germinativa. Adicionalmente, a união do segmento V_H , D_H e J_H no éxon V(D)J rearranjado formam a terceira região determinante de complementaridade (CDR3), como observado na Figura 1 (Revisado em JUNG et al., 2006). Assim, a junção aleatória de diferentes segmentos da linhagem germinativa V_H , D_H e J_H é uma segunda fonte de diversidade.

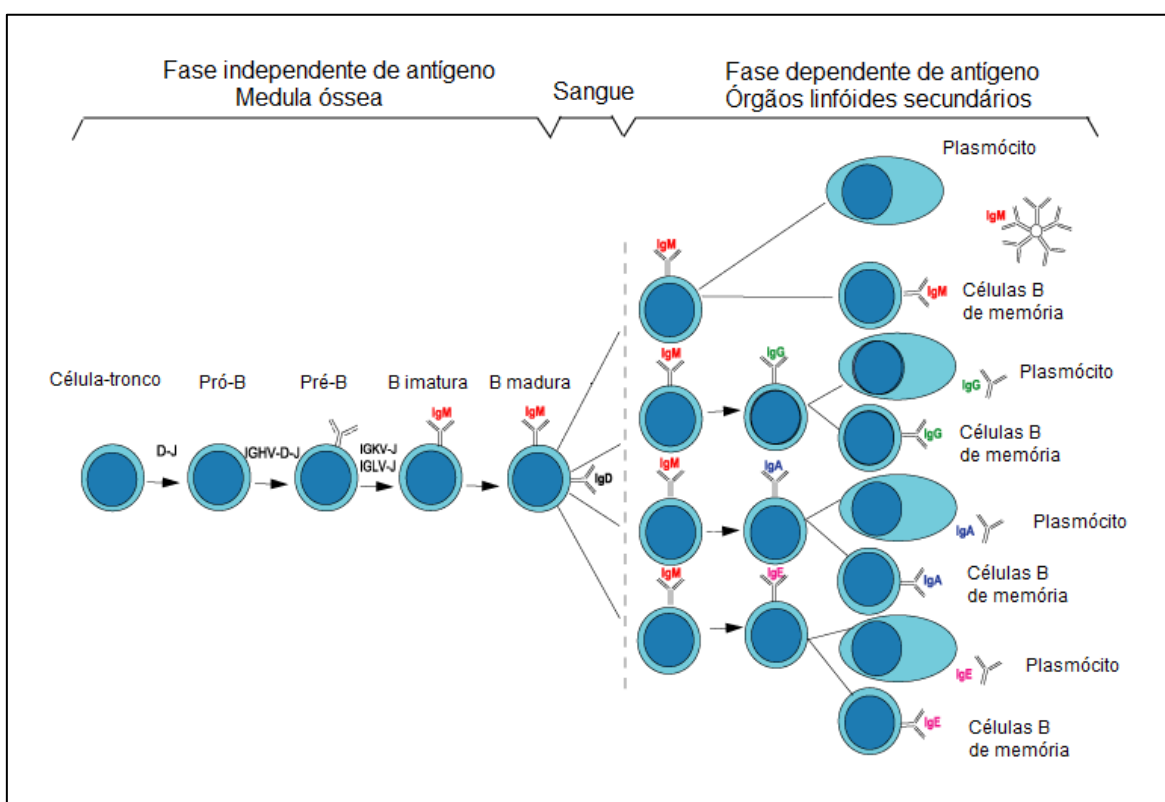


Figura 5: A diferenciação das células B a partir das células-tronco até as células B maduras na medula óssea é independente de antígeno. As etapas finais da diferenciação até as células de memória e plasmócitos acontece nos órgãos secundários e são dependentes de antígeno. Fonte: Modificado de LEFRANC et al. (2015)

A recombinação dos segmentos é produzida pelas endonucleases RAG1 e RAG2, também conhecidas como recombinases V(D)J, codificados pelos genes ativadores de recombinação – RAG1 e 2. As RAG1/2 clivam a fita dupla de DNA em sequências específicas, chamadas de sinais de recombinação (RSS, do inglês *recombination signal sequences*), as quais flanqueiam os segmentos codificadores (V_H , D_H ou J_H). RSS são compostas por

um heptâmero, com uma sequência altamente conservada de sete nucleotídeos (nt), localizado adjacente à sequência codificadora, seguido de um espaçador de 12 ou 23 nt, e seguidos por um nonâmero, segmento altamente conservado de nove nucleotídeos (Figura 6a). Os segmentos V_H e J_H são flanqueados por um RSS de 23 nt, e os segmentos D_H por RSS de 12 nt em ambos os lados. Para uma quebra efetiva da fita dupla, as proteínas RAG requerem que no momento da junção entre dois segmentos codificadores, um RSS tenha um espaçador de 12 nt e o outro de 23 nt, isto é chamado de “regra de 12/23”. Isto garante que a junção seja feita na ordem V_H , D_H , J_H . Desta forma são produzidas quebras de fita dupla de DNA entre um heptâmero de D e um heptâmero de J (recombinação DJ), seguida de uma quebra na fita de DNA entre V e DJ (Figura 6b) (JUNG et al., 2006)

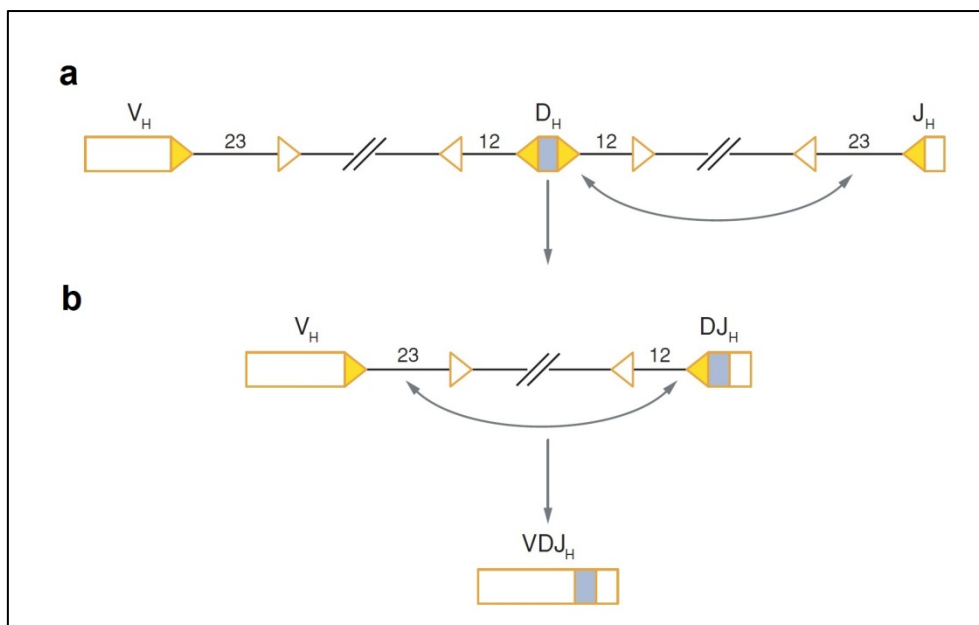


Figura 6: Rearranjo dos genes V(D)J na porção variável de *IGH*. Na primeira etapa é produzida uma quebra e união na fita dupla entre D e J, e posteriormente uma quebra e união entre V e DJ. Fonte: Modificado de JUNG et al. (2006)

Quando o DNA é clivado pelas RAG, a fita dupla forma uma alça em forma de grampo, as quais por sua vez são clivadas pelas enzimas Artemis e DNA-PKcs (do inglês, *DNA-dependent protein kinase, catalytic subunit*). A clivagem acontece de forma assimétrica, de modo que um filamento de DNA é

mais longo do que o outro. O filamento mais curto é alongado com nucleotídeos complementares, gerando sequências palindrômicas curtas (nucleotídeos P), e posteriormente a enzima desoxinucleotidil transferase terminal (TdT) adiciona aleatoriamente nucleotídeos sem uso de molde à extremidade 3' e introduz os chamados nucleotídeos N (Figura 7). Devido a este mecanismo de geração de diversidade, chamada de diversidade juncional, as CDR3 exibem a maior variabilidade, e são chamadas de região hipervariável. Estas constituem as porções mais importantes das imunoglobulinas para especificidade da ligação ao antígeno (DUDLEY et al., 2005)

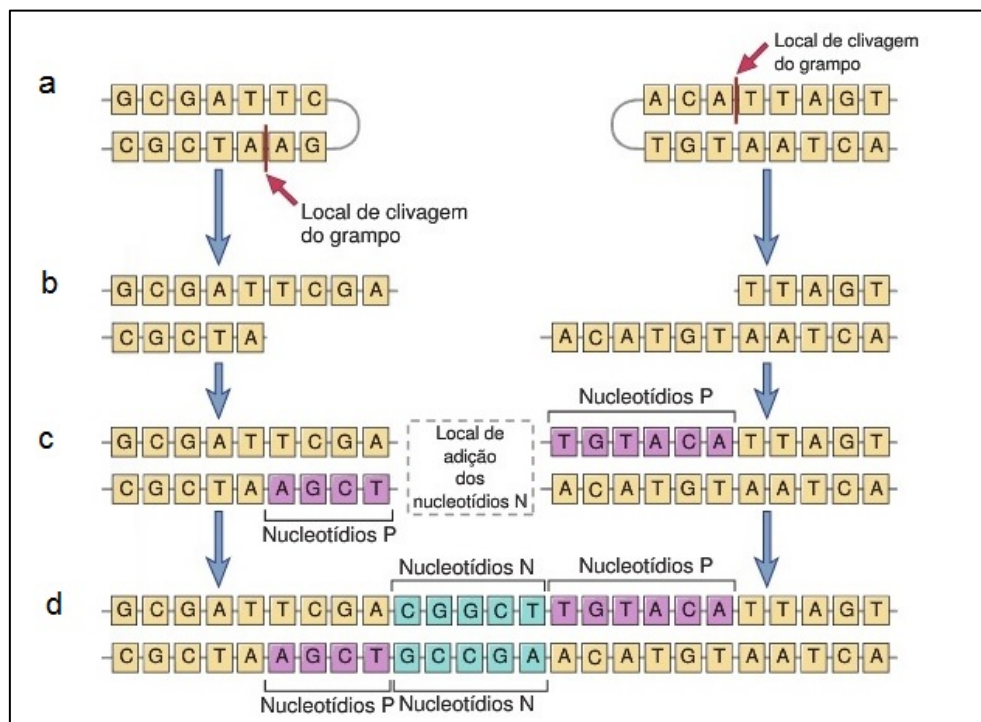


Figura 7: a) Após a quebra da fita dupla pela RAG e formação de um grampo no DNA, a enzima Artemisa/DNA-PKcs cliva a fita de forma assimétrica. b) Formação de um filamento mais curto do que o outro. c) Alongamento da fita curta (Nucleotídeos P). d) A enzima TdT adiciona nucleotídeos de forma aleatória (Nucleotídeos N). Fonte: ABBAS et al., 2012

Ao término do rearranjo do gene *IGH*, as células entram no estágio pré-B, e expressam em sua superfície a cadeia pesada rearranjada de imunoglobulina M, codificado pelo segmento gênico *IGHM* (Figura 5),

associadas a proteínas chamadas cadeias leves substitutas, estruturalmente homologas às cadeias leves κ e λ .

Neste estágio tem início o rearranjo do gene da cadeia leve, de forma similar ao rearranjo na cadeia pesada. Se o rearranjo da cadeia leve for bem-sucedido, seu produto gênico se unirá ao da cadeia pesada, para formar a proteína IgM. Neste momento, as células são denominadas células B imaturas (ABBAS et al., 2012).

Como resultado da associação das duas cadeias rearranjadas (pesadas e leves), obtêm-se a diversidade combinatória, onde há justaposição das regiões V_H e V_L , ambas geradas de forma independente e aleatória, para formar o sítio de união ao antígeno (Revisado em LEFRANC; LEFRANC, 2001).

2.3 MUDANÇA DE CLASSE POR RECOMBINAÇÃO E HIPERMUTAÇÃO SOMÁTICA

Como consequência de uma recombinação V(D)J bem sucedida nos genes de *IGH* e *IGL*, as células B expressam IgM e/ou IgD, e são denominadas células B maduras. Após a estimulação pelo imunógeno, auxílio de células T e sinais coestimulatórios, no transcorrer da resposta imune, as células B podem proliferar e diferenciar-se em células que expressam Igs de outros isotipos (IgA, IgG e IgE), por um processo chamado mudança de classe por recombinação (CSR, do inglês *Class switch recombination*) (DUDLEY et al., 2005).

Isto acontece por um evento de recombinação-deleção intracromossômica, que envolve a substituição dos segmentos gênicos *IGHM* e *IGHD*, por outros segmentos gênicos constantes da cadeia pesada, sem alteração da porção variável, nem do gene da cadeia leve daquela célula. Esta recombinação acontece entre sequências repetitivas chamadas de regiões S (*switch*), que se encontram a montante dos segmentos gênicos C_H (Revisado em SELSING, 2006; STAVNEZER, 2008; XU et al., 2012). CSR também requer uma citidina desaminase induzida por ativação (AID, do inglês *activation-induced cytidine deaminase*) a qual promove quebra na fita dupla de DNA nas regiões S_{μ} a jusante de V(D)J e a montante do segmento gênico C_H que

codifica a porção constante que será expressa pela célula, posteriormente o DNA intermediário, que forma um segmento circular extracromossômico, é descartado. Forma-se assim uma junção de regiões S-S que aproximam V(D)J ao C_H que formara o novo isotipo da Ig (Figura 8) (XU et al., 2012). Desta forma, os segmentos gênicos *IGHM* e *IGHD* são retirados. Por exemplo, se o éxon V(D)J for aproximado ao segmento gênico *IGHA*, a célula começará a produzir IgA, e não mais IgM.

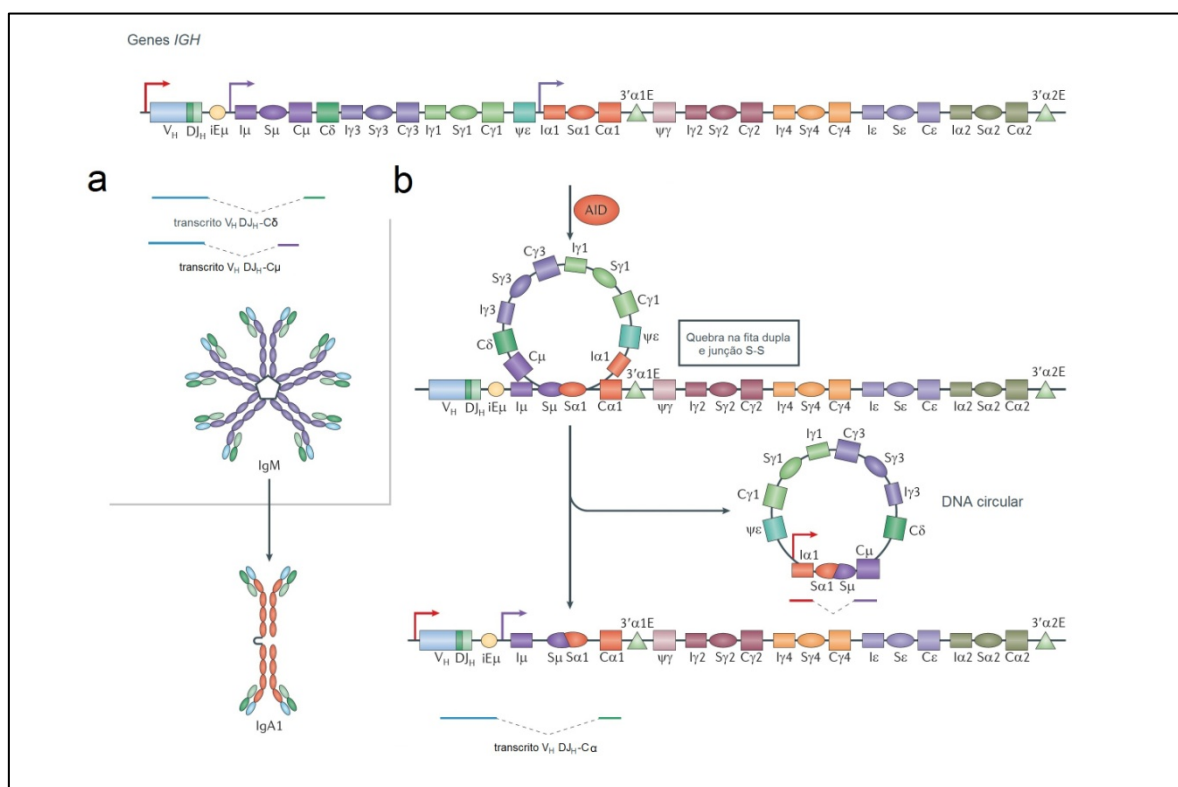


Figura 8: Mudança de classe por recombinação (CSR). a) Produção de IgM antes da mudança de classe. b) Substituição do gene *IGHM* pelo gene *IGHA*, pela enzima AID, resultando na produção de IgA1. Fonte: Modificado de XU et al. (2012)

Além da mudança de classe, as células B também passam por uma maturação na afinidade pelo antígeno. Isto acontece graças à hipermutação somática, onde acontecem mutações no DNA com uma alta taxa (10^{-3} a 10^{-4} por nt) no éxon V(D)J (segmentos variáveis) das cadeias pesadas e leves. O que permite uma posterior seleção positiva das células B que produzem as

imunoglobulinas mutadas com uma maior afinidade por um determinado antígeno, e assim uma resposta imune mais efetiva (DUDLEY et al., 2005).

Uma parte das células B que sofreram mudanças em resposta ao antígeno se diferenciam em plasmócitos e secretam anticorpos. Cada célula B secreta apenas um isotipo de Ig com alta afinidade por um antígeno (Figura 5) (LEFRANC; LEFRANC, 2001)

2.4 VARIAÇÃO ESTRUTURAL NA REGIÃO DE DOBRADIÇA DAS IgG3

Na estrutura proteica das IgG, a região de dobradiça, codificada pelo éxon *H* (*hinge*), forma uma união flexível entre o braço Fab e o Fc. O comprimento e flexibilidade da dobradiça varia entre as subclasses de IgG. Isto possivelmente afeta a localização do braço Fab em relação ao domínio Fc. O segmento gênico *IGHG3* possui de dois a quatro éxons codificadores da dobradiça (*H1*, *H2*, *H3* e *H4*) (Figura 9), enquanto todos os demais segmentos gênicos *IGHG* possuem apenas um éxon H (DARD et al., 1997).

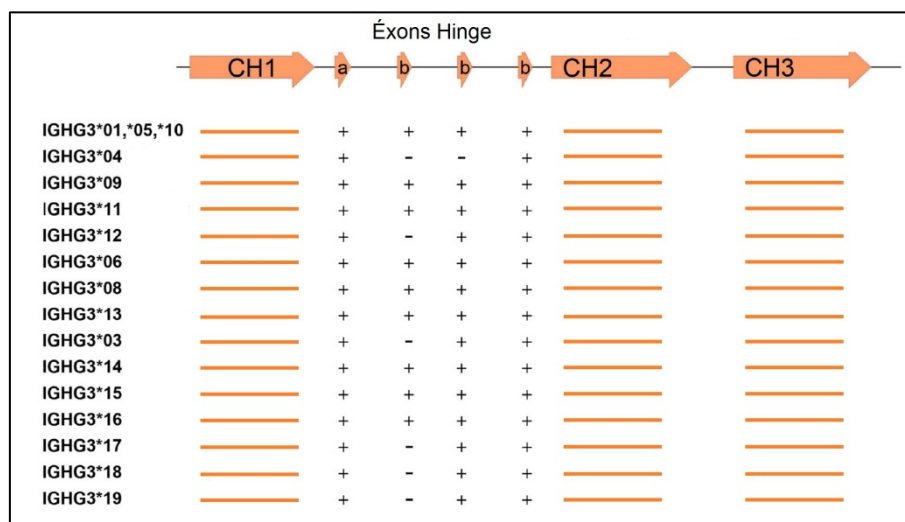


Figura 9: Comparação entre alelos de *IGHG3* em relação ao número de cópias dos éxons *H* (*hinge*). O alelo *IGHG3*04* possui dois éxons *H*, e os alelos *IGHG3*12*, **03*, **17*, **18* e **19* possuem três alelos *H*. Os demais possuem quatro éxons *H*. Fonte: Modificado de VIDARSSON et al. (2014)

A dobradiça de IgG1 tem 15 aminoácidos e é muito flexível. A de IgG2 tem 12 aminoácidos, a menor de todas as subclasses, e pontes dissulfeto

adicionais, o que a faz a menos flexível dentre as IgG. A dobradiça de IgG4 também contém 12 aminoácidos e possui uma flexibilidade intermediária entre IgG1 e IgG2. IgG3 tem uma região *H* contendo mais de 62 aminoácidos (Figura 10). Assim, a porção Fab se encontra longe de Fc, conferindo maior flexibilidade (Revisado em VIDARSSON et al., 2014). O comprimento exato da dobradiça depende do alelo de IgG3.

A comparação dos éxons *H* de *IGHG3* com os de outros segmentos gênicos *IGHG* mostra que seu éxon *H1* é homólogo ao do pseudogene *IGHPG*, e que seus outros éxons (*H2*, *H3* e *H4*) são homólogos ao éxon *H* de *IGHG1*, sugerindo que o ancestral do segmento gênico *IGHG3* é um híbrido criado por permuta desigual entre o ancestral *IGHG1* e o pseudogene *IGHPG* (TAKAHASHI et al., 1982).

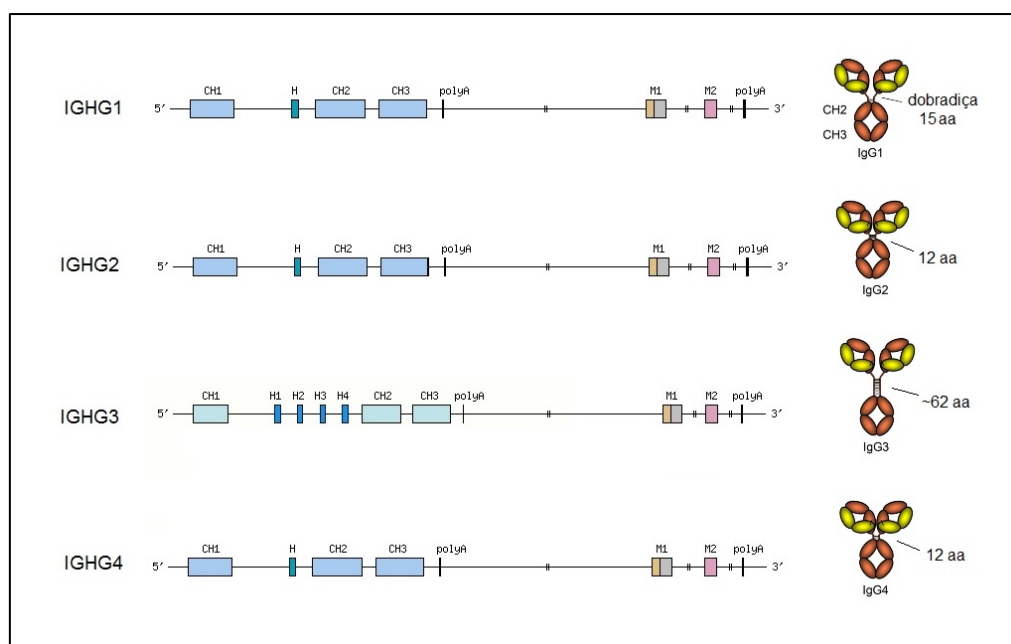


Figura 10: Relação entre os genes de *IGHG* e o comprimento da região de dobradiça das moléculas. Note-se que o gene *IGHG3* possui quatro éxons *H*, enquanto que os demais genes possuem um éxon *H*. Fonte: Modificado de LEFRANC et al. (2015) e VIDARSSON et al. (2014)

2.5 FUNÇÃO DAS IMUNOGLOBULINAS

As IgG são o isotipo predominante no soro e podem neutralizar toxinas e vírus, opsonizar patógenos, induzir a citotoxicidade celular dependente de

anticorpo (CCDA) e a citotoxicidade dependente de complemento (MURPHY, 2011; ABBAS et al., 2012).

O complexo imune antígeno:anticorpo pode ativar a via clássica do sistema complemento quando a molécula C1q, a primeira proteína da cascata do sistema complemento, reconhece a porção Fc das IgG ou IgM (Figura 11). Isto leva à deposição de C3b, opsonização das células alvo, e também à formação do complexo de ataque a membrana (MAC). As diferentes subclasses de IgG diferem em sua potencia para ativar o sistema complemento, na seguinte ordem: $IgG3 > IgG1 > IgG2$, entanto que IgG4 apresenta baixa ou nenhuma afinidade com C1q. O sítio de ligação para C1q é o domínio CH2 das IgG (Revisado em KARSTEN; KÖHL, 2012; VIDARSSON et al., 2014).

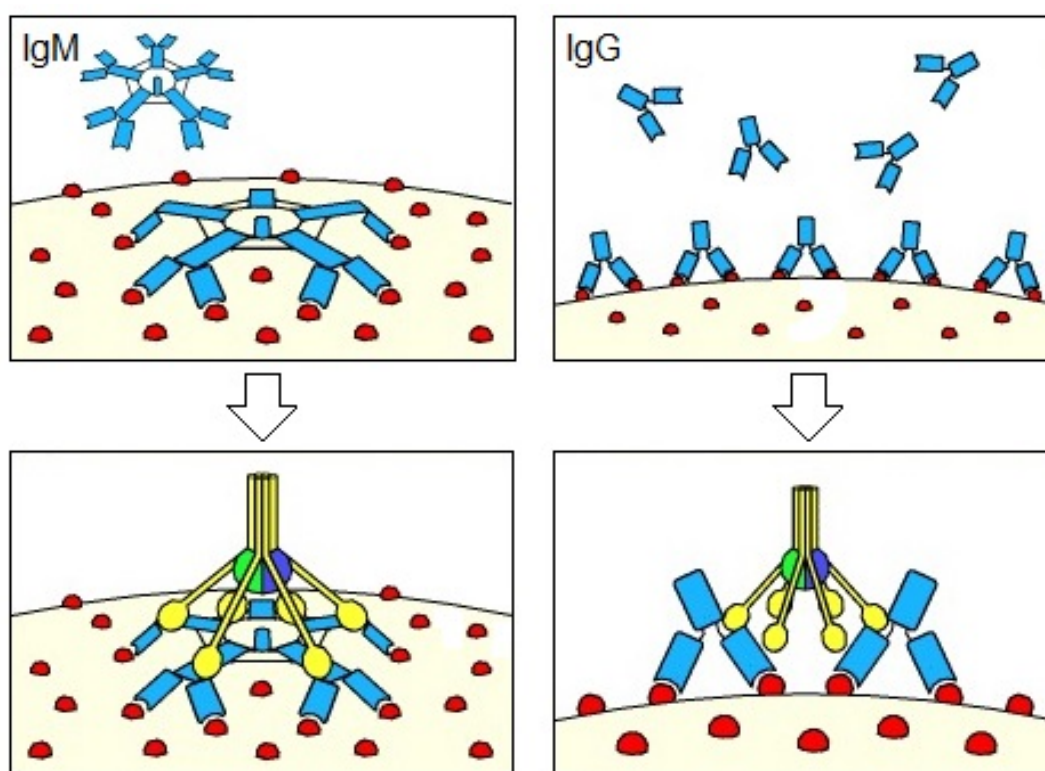


Figura 11: A ativação da via clássica do complemento é iniciada pela interação entre C1q com os anticorpos numa superfície celular, como por exemplo uma célula bacteriana. A esquerda, uma molécula de IgM na conformação em “grampo” unida a vários epítomos idênticos na superfície une-se a molécula C1q. A direita, duas ou mais moléculas de IgG devem estar unidas aos epítomos para permitir a ligação com C1q. Fonte: Modificado de MURPHY, (2011).

Um dos possíveis motivos para a diferença na força de ativação do sistema complemento pelas diferentes subclasses de IgG é a diversidade que elas apresentam no comprimento da região de dobradiça da molécula, a qual separa a região de união ao antígeno Fab do fragmento Fc, que se une ao C1q, conferindo diferença na flexibilidade dos braços dos anticorpos. Como já mencionado, IgG3 é a subclasse que apresenta a dobradiça mais comprida dentre as quatro subclasses e é também a que ativa o complemento com mais efetividade, em contraste a IgG4 apresenta uma menor dobradiça e menor flexibilidade. Isso confere uma menor acessibilidade para C1q, e menor efetividade para ativar o complemento (Revisado em KARSTEN; KÖHL, 2012). Além disso, outras características também foram sugeridas como fatores que interferem na efetividade da ativação do complemento, como a quantidade de pontes dissulfeto entre as cadeias. A menor quantidade de pontes dissulfeto diminuiria a eficiência de fixação do complemento (SCHUR; CHRISTIAN, 1964).

Do mesmo modo, a diferença na sequência de aminoácidos entre as subclasses no domínio CH2 também seria responsável pela diferença na ativação do sistema complemento, por exemplo: Glu318-Lys320-Lys322 são os resíduos em CH2 em que o C1q se une aos anticorpos, como foi sugerido por TAO et al. (1991, 1993). Portanto, variações nessas posições podem interferir na capacidade de um anticorpo de ativar o complemento. Assim também, foi encontrado que os oligossacarídeos acoplados aos anticorpos podem exercer um papel importante na ativação do complemento, assim como na formação de rosetas e na CCDA (KOIDE et al., 1977).

Por outro lado, as imunoglobulinas IgG podem opsonizar patógenos, estimulando assim sua fagocitose através da sua união aos receptores de Fc (FcγR) das células fagocitárias (neutrófilos, monócitos, macrófagos, células dendríticas). Os receptores FcγRs servem como ponte entre o sistema imune adaptativo com o inato, já que se unem às Ig e ativam várias células do sistema imune inato. Para isto deve existir uma união específica das Ig com o antígeno alvo, que resulta na formação dos complexos imunes (ICs) e na adição dos FcγRs na superfície celular. Posteriormente, inicia-se a fagocitose, CCDA, e liberação de citocinas e mediadores inflamatórios. Foi encontrado que mudanças na configuração dos anticorpos devido a ligação de sacarídeos no

domínio CH2 de IgG2 pode afetar sua afinidade aos FcγRs (Revisado em SCHROEDER; CAVACINI, 2010; KARSTEN; KÖHL, 2012).

As IgM são os receptores antigênicos das células B virgens, estão associadas com uma resposta imune primária e a ativação da via clássica do complemento. As IgA estão encarregadas pela proteção nas mucosas contra toxinas, vírus, e bactérias, onde atuam neutralizando os patógenos ou impedindo sua união à superfície das mucosas. As IgE, apesar de encontrar-se em baixa concentração, possuem respostas muito potentes especialmente contra parasitas e vermes. E são associadas a reações alérgicas ou de hipersensibilidade (Revisado em SCHROEDER; CAVACINI, 2010).

2.6 ALÓTIPOS DE IMUNOGLOBULINAS

Os alótipos de Ig são variações na região constante das cadeias polipeptídicas de uma determinada subclasse de imunoglobulina, resultantes da diversidade alélica. São denominados como Gm (marcador gama), Am (marcador alfa) e Km (marcador Kappa), já que encontram-se nas cadeias pesadas das Ig - gama1, gama2, gama3, alfa2 (alótipos G1m, G2m, G3m e A2m, respectivamente) e na cadeia leve kappa (alótipos Km). A localização dos marcadores antigênicos de cada alótipo nas Ig e a nomenclatura alotípica atual e antiga, podem ser encontradas na Tabela 1. Os alótipos apresentam determinantes antigênicos únicos reconhecidos por anticorpos específicos, portanto são sequências de aminoácidos nas moléculas de Ig, cuja identificação foi inicialmente feita por métodos sorológicos. E em termos de imunogenicidade, representam os epítomos das células B e podem ser detectadas como imunógenos em indivíduos que possuam Ig com alótipos diferentes (JEFFERIS; LEFRANC, 2009; LEFRANC; LEFRANC, 2012; PANDEY; LI, 2013).

A nomenclatura dos haplótipos Gm leva em consideração a ordem dos segmentos gênicos no *locus IGHG*: *IGHG3*, *IGHG1* e *IGHG2*, com ponto e vírgula separando os alótipos dentro das subclasses e vírgulas separando os alótipos. O G2m23 é o único alótipo encontrado em gama2, e são utilizados dois pontos (..) para indicar que um indivíduo foi testado e apresentou resultado sorológico negativo (LEFRANC et al., 2015).

Tabela 1: Nomenclatura dos alótipos atual (WHO) e antiga dos marcadores Gm, A2m e Km e localização dos marcadores antigênicos nas moléculas de Ig.

Domínio na Cadeia Pesada de IG		Nomenclatura WHO	Nomenclatura antiga
H-GAMA1	CH3	G1m1	G1m(a)
	CH3	G1m2	(x)
	CH1	G1m3	(f)
	CH1	G1m17	(z)
	CH3	G1m27	
	CH3	G1m28	
H-GAMA2	CH2	G2m23	G2m(n)
H-GAMA3	CH3	G3m5	G3m(b1)
	CH3	G3m6	(c3)
	CH3	G3m10	(b5)
	CH3	G3m11	(b0)
	CH3	G3m13	(b3)
	CH3	G3m14	(b4)
	CH3	G3m15	(s)
	CH2	G3m16	(t)
	CH2	G3m21	(g1)
	CH3	G3m24	(c5)
	CH3	G3m26	(u)
	CH3	G3m27	(v)
	CH3	G3m28	(g5)
H-ALFA2	CH1	A2m1	A2m1
	CH1	A2m2	A2m2
	CH3	A2m3	
L-KAPPA	C-KAPPA	Km1	Km1
	C-KAPPA	Km2	Km2
	C-KAPPA	Km3	Km3

H=cadeia pesada; L=cadeia leve das imunoglobulinas. Fonte: LEFRANC e LEFRANC (2012).
WHO: organização mundial da saúde (world health organization)

Atualmente são conhecidos 26 alótipos de Ig, 20 de Gm (18 “clássicos” e dois “supranumerários”), três alótipos A2m e três Km. Os 18 alótipos clássicos integram quatro G1m (1, 2, 3, 17); um G2m (23); e 13 G3m (5, 6, 10, 11, 13, 14, 15, 16, 21, 24, 26, 27, 28). Os alótipos supernumerários G1m27 e G1m28 foram descritos em populações de origem predominantemente africana; são alótipos de gama1 invés de estar em gama3 como seria esperado. A2m possui três alótipos (1, 2, 3) e Km três alótipos (1, 2, 3) (LEFRANC; LEFRANC, 2012).

Os alótipos Gm são herdados em blocos chamados de “haplótipos Gm” ou “haplótipos Gm-Am”, quando os alótipos Am também são testados. Isso ocorre devido ao elevado desequilíbrio de ligação entre os segmentos gênicos no *locus IGH*. Apesar de ocorrer em baixa frequência, eventos de

recombinação e conversão gênica aconteceram durante a evolução, resultando no aumento da diversidade haplotípica em populações. Assim, este sistema é bastante apropriado para caracterizar as populações humanas por seus conjuntos específicos de haplótipos. Por exemplo, o haplótipo Gm 5,10,11,13,14,26;3;23 é comum em europeus e eurodescendentes, enquanto Gm 5,10,11,13,14,26,27;17,1 em africanos (e afrodescendentes) (JOHNSON et al., 1977; LEFRANC; LEFRANC, 2012; PANDEY; LI, 2013).

2.7 IMPLICAÇÕES FUNCIONAIS DA VARIABILIDADE GENÉTICA DE GM

PANDEY e colaboradores (1979) observaram que crianças eurodescendentes negativas para o alótipo Km1 (Km1-/1-) (n=12) desenvolveram imunidade após imunização com uma dose da vacina contra *Haemophilus influenzae* tipo b. Em contraste, crianças Km1 positivas (n=8) não desenvolveram imunidade ($p=0,008$). O mesmo não foi verificado em crianças afrodescendentes (n=30). Ainda no mesmo estudo foi demonstrado que crianças eurodescendentes que apresentam o alótipo Km1 (Km1+/1+) (n=8) desenvolveram imunidade após aplicação da vacina contra Meningococo C, por outro lado, crianças Km1 negativas (n=6) não desenvolveram imunidade ($p=0,005$).

Associações entre alótipos de Gm e a concentração de anticorpos no soro já foram verificadas. A presença do alótipo Gm23 foi associado a uma maior concentração média de IgG2, e a presença do haplótipo Gm5,10,11,13,14;3 a uma maior concentração média de IgG3 (SARVAS et al., 1991). Também foi visto que indivíduos homozigotos para Gm23 produzem quatro vezes mais anticorpos IgG2 em resposta à vacinas contra *Haemophilus influenzae* tipo b do que indivíduos homozigotos para Gm(..) (Gm23 negativo) (GRANOFF; HOLMES, 1992).

Ainda não está completamente claro como os alótipos de Ig influenciam a resposta imunológica e o desenvolvimento de imunidade. Contudo, PANDEY (2000) sugere os seguintes mecanismos: 1) os polimorfismos que definem os alótipos podem estar em desequilíbrio de ligação com variações da região variável associados à resposta imunológica; 2) a variação estrutural atribuída

aos alótipos pode modular a cinética dos sítios de união ao antígeno; 3) os polimorfismos na cadeia constante podem contribuir diretamente para a formação dos sítios de união ao antígeno, como foi relatado em camundongos (MORAHAN et al., 1983).

2.8 POLIMORFISMOS EM IG E SUSCEPTIBILIDADE A DOENÇAS

Em um estudo de associação caso-controle em indivíduos Brasileiros de diversas etnias e indivíduos Japoneses, foram genotipadas as variantes nucleotídicas que definem os alótipos Gm 3, 17, 23, 5 e 21. Foi sugerido que indivíduos eurobrasileiros que possuem a variante alélica que define o alótipo Gm3 em IgG1 tem duas vezes mais probabilidade de desenvolver câncer de mama, em comparação àqueles indivíduos homocigotos para Gm17. Porém, essa associação não foi corroborada em amostras populacionais de outros grupos étnicos analisadas no mesmo estudo. Os autores sugerem que o alótipo Gm3 poderia mediar diretamente o risco de desenvolver a doença. Alternativamente, poderia existir outro *locus* de susceptibilidade ao câncer no cromossomo 14 que estaria em desequilíbrio de ligação com a variante que codifica os alótipos de Ig. O desequilíbrio de ligação entre alótipos Gm em Japoneses é diferente daquele em eurodescendentes ou afrodescendentes, o que também sugere que o desequilíbrio com outros *loci* que possam conferir maior risco ao câncer de mama também seja diferente entre os grupos, contribuindo para as diferenças nas associações genéticas entre indivíduos pertencentes a diferentes grupos populacionais (PANDEY et al., 2012).

Diversos vírus, incluído o citomegalovírus humano (HCMV), desenvolveram estratégias evolutivas para evadir às respostas imunológicas dos seus hospedeiros. Uma dessas estratégias é a produção de proteínas que possuem características similares aos receptores FcγR do hospedeiro, que se une a região Fc das IgG. Desse modo, evitam que os anticorpos se unam aos FcγR do hospedeiro e que desencadeie a função efetora destes receptores, necessária para a eliminação do patógeno (Revisado em MILLER-KITTRELL; SPARER, 2009).

O gene *TRL11* de HCMV, codifica uma proteína viral, semelhante a FcγR, que se une as Fc das Ig do hospedeiro. Em um estudo para determinar se a proteína codificada por *TRL11* se une diferencialmente a anticorpos IgG1 dependendo do seu alótipo, foi encontrado que esta proteína possui maior afinidade com anticorpos IgG1 com alótipo Gm3 do que com Gm1,2,17. Como resultado, indivíduos com alótipo Gm3 seriam mais propensos a ter os domínios Fc de seus anticorpos inativados por esta proteína viral, reduzindo sua competência para neutralizar o vírus. Entretanto, IgG1 com Gm1,2,17 tem menor afinidade com a proteína codificada por *TRL11*, e menor risco de desenvolver doenças mediadas pelo vírus HCMV (NAMBOODIRI; PANDEY, 2011). Um resultado similar foi encontrado para uma proteína FcγR codificada pelo herpesvírus HSV1, em que foi sugerido que as moléculas de IgG com alótipo Gm1,17 ligam-se com mais força ao receptor em comparação aos IgG com alótipo Gm3 (ATHERTON et al., 2000). A partir disto, NAMBOODIRI e PANDEY (2011) sugerem que os indivíduos heterozigotos em IgG1 (Gm3 / Gm1,2,17) teriam uma vantagem evolutiva em relação a aqueles homozigotos para algum desses alótipos, já que teriam proteção contra diversos vírus, e que a seleção balanceadora, por vantagem do heterozigoto, seria o mecanismo pelo qual os dois alelos persistem dentro das populações.

Em um estudo de associação caso-controle foi analisada a associação entre interações genéticas dos alótipos de Ig e *loci* localizados no MHC (complexo principal de histocompatibilidade) com suscetibilidade a meningite por *Haemophilus* e epiglote, em uma população de eurodescendentes nos Estados Unidos. Nenhum alótipo de Ig foi associado com um aumento ou diminuição a suscetibilidade a estas doenças, similarmente, nenhum alelo de *HLA* foi associado a susceptibilidade no mesmo grupo de pacientes e controles. Porém, foram encontradas frequências significativamente diferentes de combinações de grupos alélicos de *HLA* e alótipos de Ig entre pacientes e controles. Assim, em pacientes com o haplótipo Gm1,3,17; 5,13,21;23 o risco de desenvolver meningite ou epiglote foi menor naqueles com *HLA-B*51* ou *HLA-B*52* (grupo sorológico HLA-B5) do que em pacientes sem estes alelos (OR < 0,1; $p < 0,004$). Por outro lado, em indivíduos com o haplótipo Gm1,3,17;5,13,21 (sem Gm23), o risco de desenvolver estas doenças foi maior naqueles que apresentavam *HLA-DR3* (OR = 11; $p = 0,02$) (GRANOFF et al.,

1984). Resultados semelhantes foram encontrados em um estudo sobre hepatite crônica autoimune onde encontraram um risco aumentado de desenvolver a doença em indivíduos com HLA-B8 e Gm1,2 (WHITTINGHAM et al., 1981).

A Tabela 2 resume outros trabalhos nos quais os alótipos de Ig foram associados à susceptibilidade a doenças. Na maioria dos casos a etiologia destes fenômenos ainda não é conhecida. Além da associação direta de algum mecanismo envolvendo as Ig com o desenvolvimento das doenças, uma das hipóteses proposta pelos autores e que estes polimorfismos estão ligados a outros genes do cromossomo 14 que realmente estariam envolvidos na doença. Portanto o conhecimento da variabilidade e do desequilíbrio de ligação no *locus IGH* e genes adjacentes nas populações, permitirá desenvolver estratégias para a realização de imputações entre as variantes, e ajudará a compreender a relação com doenças complexas.

2.9 CARACTERIZAÇÃO DA VARIABILIDADE DAS IG

Ao longo dos anos, os alótipos têm sido caracterizados principalmente por métodos sorológicos, enquanto que técnicas de genética molecular foram de uso bastante limitado. Dessa forma, a diversidade alélica destes genes está subestimada. Os alelos de um segmento gênico de Ig correspondem a variantes polimórficas na sua sequência nucleotídica, que difere por pelo menos um nucleotídeo na sequência da região codificante. A nomenclatura é feita com o símbolo do IMGT (International ImMunoGeneTics Information System) do segmento gênico, seguido de um asterisco e de um número começando por *01, por exemplo, IGHG1*01. Os polimorfismos de nucleotídeo único (SNP) responsáveis pelos alótipos clássicos ainda não foram totalmente identificados (JEFFERIS; LEFRANC, 2009; PANDEY; LI, 2013; LEFRANC et al., 2015). Entretanto, é possível relacionar os alelos descritos até o momento com os alótipos de Ig conhecidos (Tabela 3).

Tabela 2: Associação entre alótipos de imunoglobulina e diversos fenótipos

Fenótipo	Alótipos associados	Associação	Técnica utilizada	População	Referencia
Susceptibilidade a Esquizofrenia	Gm3;(..)/3;(..)	OR=3,4 $p = 0,0002$	Genotipagem de SNPs	Eurodescendentes Pais:Estados Unidos	(PANDEY et al., 2016)
Alta concentração de IgG no líquido cefalorraquidiano em pacientes com MS	GM21;1,17;(..) e GM21;1,2,17;(..)	$p < 1 \times 10^{-5}$	GWAS e sequenciamento	Europeus Pais: Alemanha e Bélgica	(BUCK et al., 2013)
Susceptibilidade a esclerose múltipla (MS)	Gm1,17;21/ 1,17;21	RR = 3,6 $p = 0,02$	Inibição da hemaglutinação	Eurodescendentes Pais:Estados Unidos	(PANDEY et al., 1981)
Polipose adenomatosa familiar	Gm3;5,13	RR = 2,34 $p = 0,03$	Inibição da hemaglutinação	Europeus Pais: Dinamarca	(PANDEY et al., 1986)

Abreviações: OR: Odds Ratio. RR: Risco relativo. p: significância estatística. GWAS: Estudo de associação de genoma total. SNP: Polimorfismo de nucleotídeo único
Fonte: a autora

Tabela 3: Correspondência entre os alótipos G1m, G2m e G3m e os alelos de *IGHG1*, *IGHG2* e *IGHG3*.

Alótipo G1m	Alelos <i>IGHG1</i>	Exon/Domínio IMGT Posição								
		CH1 103	120	CH3 12	14	110				
G1m3	<i>IGHG1*03</i>	Ile atc	Arg aga	Glu gag	Met atg	Ala gct				
G1m3,1	<i>IGHG1*08</i>	Ile atc	Arg aga	Asp gat	Leu ctg	Ala gct				
G1m17,1	<i>IGHG1*01, IGHG1*02</i>	Ile atc	Lys aaa	Asp gat	Leu ctg	Ala gct				
G1m17,1,27	<i>IGHG1*04</i>									
G1m17,1,28	<i>IGHG1*05</i>									
G1m17,1,27,28	<i>IGHG1*06</i>									
G1m17,1,2	<i>IGHG1*07</i>	Ile atc	Lys aaa	Asp gat	Leu ctg	Gly ggt				
Alótipo G2m	Alelos <i>IGHG2</i>	CH2 45.1								
G2m23	<i>IGHG2*02</i>	Met atg								
G2m..	<i>IGHG2*01, IGHG2*03, IGHG2*04, IGHG2*05, IGHG2*06</i>	Val gtg								
Alótipo G3m	Alelos <i>IGHG3</i>	CH2 82	83	CH3 39	44	84	98	101	115	116
G3m5*(5,10,11,13,1,26,27)	<i>IGHG3*01, IGHG3*05, IGHG3*06, IGHG3*07, IGHG3*09, IGHG3*10, IGHG3*11, IGHG3*12</i>	Pro cgg	Arg cgg	Val gtg	Ser agc	Met atg	Gln cag	Ileu atc	Arg cgc	Phe ttc
G3m6*(5,6,10,11,14,26,27)	<i>IGHG3*13</i>	Pro cgg	Arg cgg	Val gtg	Ser agc	Met atg	Glu gag	Ileu atc	Arg cgc	Phe ttc
G3m24*(5,6,11,24,26)	<i>IGHG3*03</i>	Pro cgg	Arg cgg	Val gtg	Ser agc	Val gtg	Glu gag	Val gtg	Arg cgc	Phe ttc
G3m15*(10,11,13,15,27)	<i>IGHG3*17</i>	Pro cgg	Arg cgg	Met atg	Ser agc	Val gtg	Gln cag	Ileu atc	His cac	Tyr tac
G3m16*(10,11,13,15,16,27)	<i>IGHG3*18, IGHG3*19</i>	Pro cgg	Trp tgg	Met atg	Ser agc	Val gtg	Gln cag	Ileu atc	His cac	Tyr tac
G3m21*(21,26,27,28)	<i>IGHG3*14, IGHG3*15, IGHG3*16</i>	Leu ctg	Arg cgg	Val gtg	Asn aat	Met atg	Gln cag	Ileu atc	Arg cgc	Tyr tac

Fonte: LEFRANC e LEFRANC (2012)

Os estudos que utilizam marcadores Gm como candidatos a fatores de risco para doenças infecciosas, autoimunes ou cânceres, foram principalmente executados com métodos sorológicos, ou por genotipagem de SNPs específicos. Entretanto, a maioria deles não pode ser confirmada, ou refutada, por metodologias mais modernas, como são os estudos de associação de genoma total (GWAS, *Genome-Wide Association Study*). Um dos principais motivos é a ausência de marcadores no *locus IGH* na maioria das plataformas de genotipagem (PANDEY; LI, 2013).

As plataformas de GWAS são desenhadas para detectar SNPs cujas frequências são de pelo menos 5%, os quais também são selecionados por ter a capacidade de identificação de variantes herdadas em blocos, conhecidos com o nome de *tag SNPs* (FEERO et al., 2010). Apesar de que alguns SNPs encontrados nos alelos de *IGHG* possuem uma frequência maior de 5% em diferentes populações, eles não se encontram nas plataformas de genotipagem por microarranjo devido a alta similaridade entre os segmentos gênicos adjacentes a eles, o que dificulta o desenho de sondas de DNA. E finalmente, como estes genes não foram completamente estudados no projeto *HapMap* (Mapa de haplótipos do genoma humano) (INTERNATIONAL HAPMAP CONSORTIUM, 2005), eles não podem ser imputados por desequilíbrio de ligação com outros SNPs presentes nas plataformas de GWAS disponíveis (PANDEY; LI, 2013). Por este motivo, estas plataformas não detectam alelos comuns do *locus IGHG* nem *tag SNPs* em desequilíbrio de ligação com eles, o que não permite confirmar nem refutar resultados obtidos utilizando alótipos Gm como candidatos (PANDEY, 2010). Ainda, esta região do genoma não está bem coberta em estudos de genoma total e nas bases de dados genômicos porque as amostras de DNA usadas para a construção destes bancos geralmente é proveniente de linhagens de células B, as quais sofreram recombinação somática e, portanto, não são adequadas para esse tipo de estudo.

Assim, a fim de conhecer a base genética do polimorfismo existentes nas moléculas de Ig é necessário o desenho de oligonucleotídeos iniciadores (*primers*) específicos para cada um dos segmentos gênicos, fazer uso de técnicas que permitam conhecer a sua sequência nucleotídica, e utilizar DNA extraído de tecidos nos quais estes genes não sofreram rearranjo.

2.10 POPULAÇÕES AMERÍNDIAS NO BRASIL

A maioria das populações nativas americanas, desde o Canadá até o sul do Chile, provém de uma população ancestral homogênea de “primeiros ancestrais”, que atravessou o estreito de Bering provavelmente há mais de 15.000 anos. Também foi sugerido que pelos menos duas ondas de migração vindas da Ásia tenham acontecido posteriormente, estabelecendo três linhagens asiáticas diferentes: os Paleo-Ameríndios ou primeiros americanos, os Esquimó-Aleútes e os Na-dene (WILLIAMS et al., 1985; SCHANFIELD, 1992; REICH et al., 2012).

SKOGLUND e colaboradores (2015) analisando dados de genoma total encontraram evidências genéticas de que populações nativas americanas da Amazônia (Suruí e Karitiana, da família linguística Tupi) e do Centro do Brasil (Xavante, da família linguística Ge), possuem sinais genéticos compartilhado com populações indígenas da Australásia (Austrália, Nova Guiné e Ilhas Andaman). Estes autores propõem que uma população ancestral, chamada “População Y” (*Ypykuéra*), que deram origem aos atuais nativos da Australásia, também contribuiu para o atual *pool* gênico dos Ameríndios, junto com as populações vindas da Sibéria já nas primeiras ondas migratórias que colonizaram a Amazônia.

Em análises realizadas com alótipos de Ig e seu polimorfismo dentro de populações, foi observado que este sistema não é capaz diferenciar as populações de nativos da América do Sul das populações de aborígenes da Austrália (SCHANFIELD, 1980). Em ambos grupos os haplótipo Gm1,17,21 e Gm1,2,17,21 se encontram com frequências altas e semelhantes, de 0,80 e 0,20, respectivamente (JOHNSON et al., 1977; SCHANFIELD, 1977, 1992).

Em relação aos nativos americanos que habitavam o Brasil antes da colonização europeia, estima-se que existiam mais de 1.000 povos indígenas, somando entre 2 e 4 milhões de indivíduos. Hoje restam somente 246 desses povos, somando 896.917 indivíduos, o que representa aproximadamente 0,47% da população total do país (Censo Demográfico IBGE 2010). A maior parte dessa população distribui-se por centenas de aldeias, situadas no interior de 699 Terras Indígenas (TI), de norte a sul do território nacional, o que

representa somente 13,6% das terras do país reservadas aos povos indígenas (INSTITUTO-SOCIOAMBIENTAL, 2016).

Dentre essas populações indígenas, as mais numerosas são Guarani, com cerca de 58 mil indivíduos, Ticuna com aproximadamente 36 mil e Kaingang com aproximadamente 33 mil representantes (Tabela 4) (INSTITUTO-SOCIOAMBIENTAL, 2016).

Tabela 4: Distribuição geográfica e tamanho populacional de Kaingang, Guarani e Tikuna no Brasil

População	Nº de indivíduos no Brasil	Família linguística	Estados de residência
Kaingang	33.064*	Jê	PR, RS, SC, SP
Guarani Nandeva	13.000*	Tupi-Guarani	MS, PR, RS, SC, SP
Guarani Kaiowa	31.000*	Tupi-Guarani	MS
Guarani Mbya	7.000*	Tupi-Guarani	ES, PA, PR, RJ, RS, SC, SP, TO
Tikuna	36.377**	Tikuna	AM

Fonte: INSTITUTO-SOCIOAMBIENTAL, 2016

*(Funasa, Funai, 2008)

** (Funasa, Funai, 2009)

A origem e processos de diversificação ocorridas ao longo do tempo nas populações indígenas podem ser estudados a partir das suas línguas. Antes da chegada dos colonizadores portugueses, havia cerca de mil línguas diferentes no então território brasileiro. Hoje, restaram cerca de 150 línguas indígenas, algumas mais semelhantes entre si do que outras. Essas relações podem ser representadas através de troncos e famílias linguísticas (Figura 12). Entre línguas de uma mesma família, as semelhanças são maiores, resultado de separações ocorridas há menos tempo. As línguas indígenas no Brasil possuem dois grandes troncos, Tupi e Macro-Jê, além de outras 19 famílias linguísticas que não são agrupadas em troncos pela baixa semelhança entre elas (INSTITUTO-SOCIOAMBIENTAL, 2016).

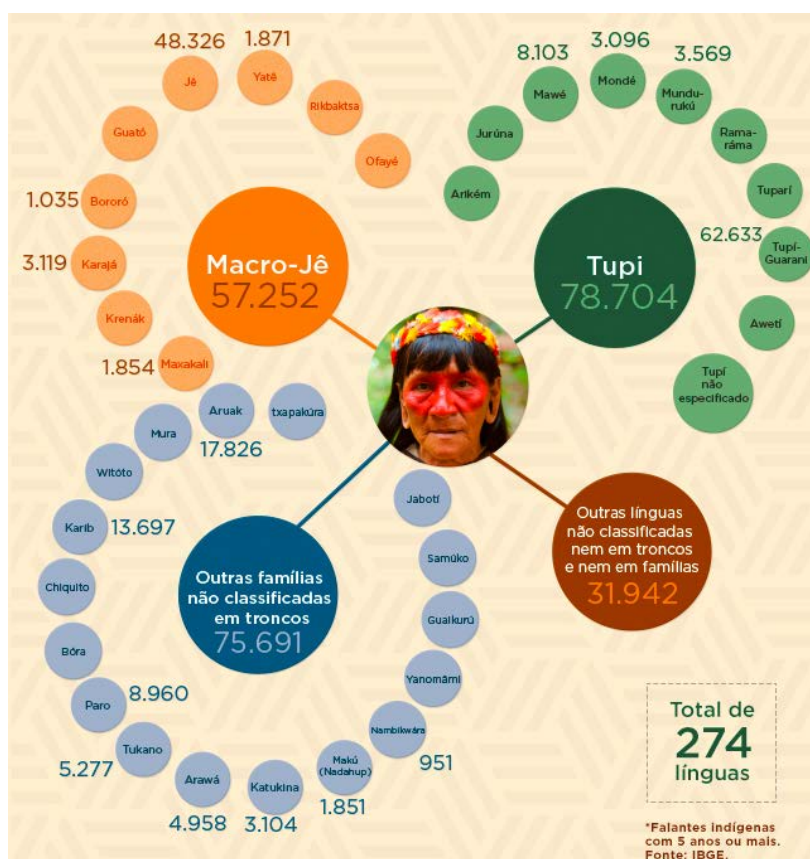


Figura 12 Troncos e famílias linguísticas do Brasil. Dentro da família Jê inclui-se o Kaingang, e dentro de Tupi Guaraní: Guaraní Mbya, Kaiowa e Nãndeva. *274 línguas identificadas em todo o Brasil (PORTAL-AMAZÔNIA, 2016)

No Brasil, os indígenas habitam as Terras Indígenas (TIs), as quais, segundo a Constituição Federal, são aquelas "por eles habitadas em caráter permanente, as utilizadas para suas atividades produtivas, as imprescindíveis à preservação dos recursos ambientais necessários a seu bem-estar e as necessárias a sua reprodução física e cultural, segundo seu usos, costumes e tradições". No estado do Paraná existem 26 TIs, onde habitam as etnias Kaingang, Guaraní (Nãndeva e Mbyá) e Xetá. No Mato Grosso do Sul existem 53 TIs, onde residem Guaraní Kaiowá e Nãndeva (INSTITUTO-SOCIOAMBIENTAL, 2016).

2.10.1 Kaingang

Antes da colonização europeia, os Kaingang ocupavam grandes áreas no sul do Brasil e parte de Misiones, na Argentina. Viviam principalmente da agricultura, além da caça, pesca e coleta. Hoje vivem em mais de 30 TIs, o que

representa uma pequena porcentagem do seu território original. São a terceira maior etnia indígena no Brasil. Sua língua é o Kaingang, classificado na família Jê, tronco lingüístico Macro-Jê (INSTITUTO-SOCIOAMBIENTAL, 2016; PORTAL KAINGANG, 2016).

2.10.2 Guarani

Antes da chegada dos colonizadores, os Guaranis viviam perto de grandes rios, praticavam a agricultura e eram nômades. Habitavam o Rio Grande do Sul, Santa Catarina, Paraná, São Paulo e Mato Grosso do Sul, assim como Paraguai, Argentina e Bolívia. Hoje continuam vivendo no mesmo território, e são a maior etnia indígena brasileira. Sua língua é o Guarani, da família Tupi-Guarani, tronco linguístico Tupi, e estão divididos em três grupos: Guarani Mbyá; Guarani Kaiowá, e Guarani Ñandeva segundo o dialeto que falam (AZEVEDO et al., 2008; INSTITUTO-SOCIOAMBIENTAL, 2016).

As populações de Guarani e Kaingang têm sido estudadas em vários trabalhos de genética de populações. Em um estudo com base em polimorfismos de *HLA-A*, *-B*, *-C*, *-DR* e *-DQ* foi estimado que 1,4% do *pool* gênico presente em GRC (Guarani Mbya) provêm da população Kaingang, e que 0,5% da variabilidade de Kaingang é procedente de populações de GRC. No mesmo estudo foi sugerido que 5,3% da variação presente em Kaingang e 2,5% da presente em Guarani pertencem a populações não-Ameríndias (PETZL-ERLER et al., 1993).

Em um trabalho posterior em que se analisaram genes *HLA* de classe II em Guarani Kaiowa e Guarani Ñandeva, TSUNETO e colaboradores (2003) encontraram que, em relação à populações exogâmicas, elas apresentam uma diversidade alélica baixa. Também foi estimado que a porcentagem de miscigenação das populações nativas com não-Ameríndios foi de 14,3% em Guarani Ñandeva, 7,2% em Kaingang e de 3,7% em Guarani Mbya, sendo que não foi encontrada evidência de fluxo gênico com não-Ameríndios em Guarani Kaiowa.

3 JUSTIFICATIVA E HIPÓTESES

A diversidade presente nas imunoglobulinas IgG é conhecida principalmente por técnicas sorológicas, enquanto que sua variabilidade genética é pouco conhecida. Isto é devido à limitação técnica produzida pela elevada similaridade entre os segmentos gênicos que os codificam, o que dificulta a elaboração de estratégias de genotipagem. Adicionalmente, os bancos de dados genômicos foram majoritariamente produzidos partir de DNA proveniente de linhagens de linfócitos B, nas quais ocorre recombinação somática dos genes *IGHG*, não sendo adequado o uso do DNA dessas células para o estudo dos genes das imunoglobulinas.

Também por este motivo, existe pouca informação sobre a diversidade genética das IgG nas populações. Assim, a hipótese deste trabalho é que existem alelos *IGHG* ainda não conhecidos nas populações de Ameríndios e de descendentes de Japoneses, e que o conhecimento da variabilidade nessas populações permitira compreender quais fatores evolutivos estão moldando a sua diversidade genética.

Devido à sua importância biológica, envolvimento na resposta imune, tanto adaptativa quanto inata, o estudo da variabilidade nucleotídica dos segmentos gênicos *IGHG* pode contribuir para compreender o papel funcional dos seus polimorfismos, implicação na susceptibilidade a doenças e sua evolução. Também pode contribuir para o desenvolvimento e humanização de anticorpos monoclonais não imunogênicos, utilizados em imunoterapias quando administrados a indivíduos de diversas populações humanas.

4 OBJETIVOS

4.1 OBJETIVO GERAL

Caracterizar a diversidade genética dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* da cadeia pesada das imunoglobulinas em indivíduos de cinco populações ameríndias brasileiras (Guarani e Kaingang) e uma população de descendentes de Japoneses.

4.2 OBJETIVOS ESPECÍFICOS

- Caracterizar a diversidade dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* em uma população de descendentes de Japoneses e cinco populações de ameríndios brasileiros: Guarani Kaiowa, Guarani Mbyá, Guarani Nandeva, Kaingang de Rio das Cobras e Kaingang de Ivai.
- Identificar os alelos e alótipos de cada segmento gênico a partir da sequência nucleotídica de cada indivíduo, e descrever os alelos não descritos anteriormente.
- Descrever a distribuição alélica, genotípica e alotípica de cada segmento gênico dentro das populações
- Descrever o desequilíbrio de ligação entre as variantes nucleotídicas e entre alelos dos segmentos gênicos.
- Comparar a diversidade dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* entre as populações estudadas.
- Comparar as frequências alotípicas de cada população com as frequências descritas em populações previamente estudadas.
- Verificar evidências de fatores evolutivos atuando sobre os segmentos gênicos.

5 METODOLOGIA

5.1 AMOSTRAS POPULACIONAIS

Foram analisados 299 indivíduos, dentre eles, 245 indivíduos são Ameríndios, das tribos Kaingang e Guarani (Tabela 5). Os Kaingang pertencem a duas terras indígenas do Paraná: 1) Rio das Cobras (25°18'S, 52°32'O; 18.682 ha), a maior do estado, localizada no município de Nova Laranjeiras, com 2.917 habitantes, sendo cerca de um terço Kaingang e os restantes Guarani e Xetá; e 2) Ivaí (24°30'S, 51°40'O; 7.306 ha), no município de Manoel Ribas, com 1.552 pessoas. Os Guarani analisados neste estudo pertencem aos grupos: 1) Guarani Mbyá (de Rio das Cobras); 2) Guarani Kaiowá, das aldeias Amambai (23°06'S; 55°12'O; 2.429 ha) no município de Amambaí-MS, e Limão Verde (23°12'S; 55°06'O; 5.377 ha) pertencente ao município de Aquidauana-MS, que abrigam cerca de 7.126 e 1.335 indivíduos, respectivamente e 3) Guarani Ñandeva, das áreas indígenas Amambai e Porto Lindo-MS (23°48'S, 54°30'O) (Figura 13). Os dados demográficos foram obtidos a partir de IBGE-Censo Demográfico 2010.

Também formaram parte da amostra deste estudo 54 indivíduos descendentes de Japoneses, residentes em Curitiba (PR). Estes indivíduos são brasileiros que possuem os quatro avós ou ambos os pais nascidos no Japão. Por tanto, é pouco provável que tenham miscigenação de outras etnias.

Tabela 5: Nome, código e tamanho amostral de cada população analisada.

População	Abreviação	n
Guarani Kaiowá	GKW	51
Guarani Ñandeva	GND	51
Guarani Mbyá	GRC	48
Kaingang de Ivaí	KIV	49
Kaingang de Rio das Cobras	KRC	46
Descendentes de Japoneses	JAP	54

As coletas do material biológico destes indivíduos foram realizadas previamente por integrantes do Laboratório de Genética Molecular Humana da Universidade Federal do Paraná. O DNA das amostras foi extraído a partir do sangue total, através dos métodos fenol-clorofórmio-álcool isoamílico (SAMBROOK et al., 1989) ou *salting-out* (LAHIRI; NURNBERGER, 1991) e a concentração de suas soluções uso foram ajustadas a 20 ng/μl.

Este trabalho foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos do Setor de Ciências Biológicas. Os indivíduos descendentes de Japoneses participaram de forma voluntária e assinaram um termo de consentimento para participação em pesquisa e um questionário para obtenção de dados pessoais. Para a utilização do material biológico de indivíduos indígenas, o projeto também foi aprovado pela Fundação Nacional do Índio (FUNAI) e pelos líderes das tribos indígenas estudadas. As amostras das populações indígenas foram colhidas anteriormente à resolução CNS 196/96 do Ministério da Saúde, quando ainda não se obtinha o consentimento informado por escrito, e seguiram o código de conduta de informação prévia dos participantes e de seus líderes comunitários, participação voluntária e garantia de anonimato.

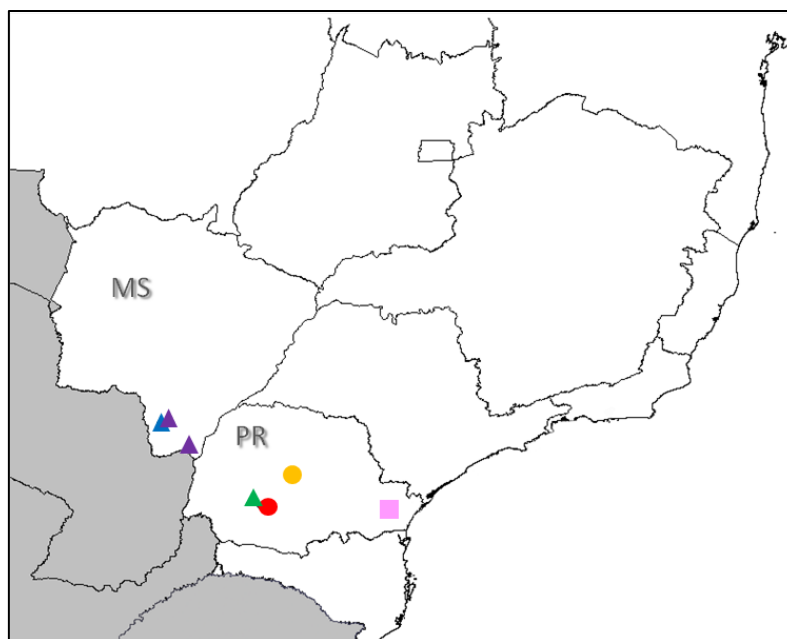


Figura 13: Mapa físico do Sul do Brasil. Em destaque estão as populações indígenas analisadas e sua localização nas terras indígenas onde habitam. Guarani Nandeva (▲), Guarani Kaiowa (▲), Guarani Mbya (▲), Kaingang de Rio Das Cobras (●), Kaingang de Ivaí (●), Descendentes de Japoneses de Curitiba (■).

5.2 AMPLIFICAÇÃO DOS FRAGMENTOS PELA REAÇÃO EM CADEIA DA POLIMERASE

Para a amplificação dos éxons *CH1*, *H*, *CH2* e *CH3* dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* foi realizada a técnica de reação em cadeia da polimerase (PCR, do inglês *Polymerase Chain Reaction*). Os oligonucleotídeos iniciadores (*primers*) específicos para cada um dos éxons se encontram na Tabela 6. A localização dos *primers* no *locus IGH* se encontra na Figura 14.

Tabela 6: Descrição dos *primers* da amplificação de cada fragmento e suas respectivas temperaturas de fusão (T_m).

Segmento Gênico	Éxon	Primer	Sequência 5'-3'	T_{mA} (°C)	T_{mB} (°C)	Tamanho do amplicom em pb
<i>IGHG1</i>	CH1+H+	IGHG1_F	ctgaacctcgcggacagtt	64	60	1846
	CH2+CH3	IGHG1_R	gaaagaaccatcacagtctcg			
<i>IGHG2</i>	CH1+H	IGHG2_F	gaccctgcctggaccc	64	60	834

		IGHG2_H_R2	gcactcgacacaacattgc			
	CH2+CH3	IGHG2_Ex1_F5	cccaaaggccaaactg	64	60	1170
		IGHG_R1	cactcaggcctcagactcg			
IGHG3	CH1	IGHG3_F1	gctctgtcccacaccgca	67	65	633
		IGHG3_Ex1_R1	aggggcagggtcctcct			
	H	IGHG3_Ex1_F2	ccaaaggccaaactcttactca	65	63	893
		IGHG3_CH2_R1	ggtccgggaaatcataagggta			
	CH2+CH3	IGHG3_Ex5_F2	aggacaggtgcctagagtg	64	60	895
		IGHG3_R1	cgtgccaagcatcctcg			

T_m A indica a temperatura de hibridação do *primer* nos primeiros 10 ciclos da reação de PCR e T_m B a temperatura nos 25 ciclos subsequentes da reação de PCR. pb: pares de bases

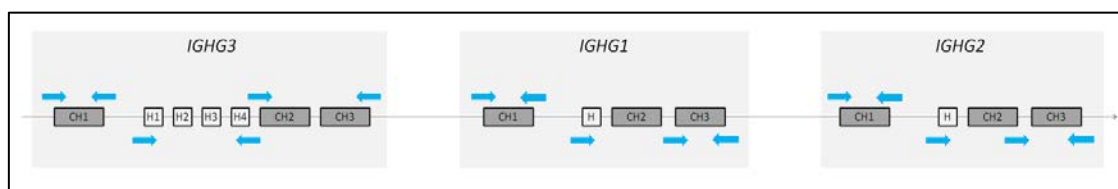


Figura 14: Representação dos segmentos gênicos sequenciados e a localização dos *primers* utilizados para a amplificação e sequenciamento.

As amplificações foram realizadas em um volume final de 10 μ l, contendo: tampão 1X (Invitrogen™); 0,2 mM de cada dNTP (Life Technologies); 1,5 mM de $MgCl_2$ (Invitrogen™); 0,3 μ M de cada *primer*, 0,5 U de Taq polimerase Platinum® (Invitrogen™) e 20 ng de DNA genômico.

As reações foram executadas no termociclador Mastecycler ep Gradient S (Eppendorf), utilizando a seguinte condição de ciclagem: um passo de desnaturação inicial a 94 °C por 2 minutos, 10 ciclos de 94 °C por 15 segundos, T_m A por 15 segundos (Tabela 6), 72 °C por 60 segundos, seguidos de 25 ciclos de 94 °C por 15 segundos T_m B por 15 segundos (Tabela 6), 72 °C por 60 segundos. E finalmente, um passo de extensão final a 72 °C por 1 minuto.

Para confirmação da amplificação do fragmento foram realizadas eletroforeses a 100 V por 40 minutos, aplicando 2,0 μ l do produto de PCR misturado a 1,0 μ l de tampão de corrida em gel de agarose 1,5% com o corante de ácidos nucleicos Diamond™ (Promega) a 1%. Juntamente às amostras foram aplicados 2 μ l do marcador Gene Ruler 1Kb (Fermentas) a fim de semi-

quantificar a concentração dos produtos de PCR de cada amostra. Posteriormente os géis de agarose foram visualizados e fotografados através de um sistema de fotodocumentação computadorizado. Com base na semi-quantificação eram utilizados volume de produto de PCR que continham aproximadamente 100 ng de DNA.

A fim de evitar que reagentes utilizados na PCR interfiram na reação de sequenciamento os produtos de amplificação foram purificados por método enzimático. Assim foram utilizadas as enzimas exonuclease I que degrada DNA de fita simples (*primers*) e a fosfatase alcalina que elimina grupos fosfato e inativa os nucleotídeos da reação anterior. A cada 5 µl de produto de PCR foram adicionados 2 µl de solução de purificação contendo 4 U da enzima exonuclease I (Fermentas) e 0,7 U da enzima fosfatase alcalina (Thermo Scientific). Esta solução foi levada ao termociclador com a seguinte condição de ciclagem: ativação enzimática a 37 °C por 60 minutos e inativação enzimática a 80 °C por 20 minutos para degradação das enzimas de purificação.

5.3 REAÇÃO DE SEQUENCIAMENTO PELO MÉTODO DE SANGER

As reações de sequenciamento dos segmentos gênicos foram realizadas pelo método de terminação da cadeia de Sanger. Para um volume final de 10 µl utilizou-se: 0,3 µl de Big Dye Terminator Cycle Sequencing Standart v3.1 (Life Technologies), 0,2 µl de oligonucleotídeo a uma concentração inicial de 10pmol/µl, 1,8 µl de Big Dye Terminator v.3.1 5X Sequencing Buffer e 1 µl de produto de PCR purificado. As reações foram realizadas em placas de reação de 96 poços Microamp Optical (Life Technologies). As condições de ciclagem consistiram em uma etapa de desnaturação inicial a 95 °C por 1 minuto; seguida por 25 ciclos de desnaturação a 95 °C por 10 segundos, hibridação a 50 °C por 5 segundos e extensão a 60 °C por 4 minutos.

A maioria dos oligonucleotídeos utilizados nas reações de sequenciamento foi diferente daqueles utilizados na amplificação, para aumentar a especificidade e confiabilidade das sequências obtidas (Tabela 7). Após a reação de sequenciamento, foi realizada uma purificação do DNA antes

da injeção das amostras no sequenciador a fim de que os reagentes utilizados não interfiram na eletroforese capilar e leitura das bases. Em cada amostra foram adicionados 2,5 µl de EDTA 125 mM (Merck) e 30 µl de etanol absoluto (Merck). Após a incubação em temperatura ambiente por 15 minutos, seguiu-se uma centrifugação a 3.220 xg por 45 minutos a 18 °C. O sobrenadante foi descartado por inversão e a placa com as amostras foi centrifugada de maneira invertida rapidamente até 300 xg. Foram adicionados 30 µl de etanol 70% à temperatura ambiente e feita uma centrifugação a 3.220 xg por 5 minutos a 25 °C. O sobrenadante foi novamente descartado por inversão, e a placa centrifugada invertida até 300 xg. A placa foi coberta apenas com papel alumínio, a temperatura ambiente por 20 minutos, para que todo o álcool evaporasse.

Em seguida a placa foi selada e armazenada a -20 °C. Antes da injeção das amostras no sequenciador, estas foram ressuspensas em 10 µl de Hi-Di Formamida (Life Technologies). A eletroforese capilar para leitura das seqüências nucleotídicas foram realizadas no sequenciador 3500xl Genetic Analyzer (Life Technologies) com os seguintes reagentes: POP-7 Polymer for 3500/3500xl Genetic Analyzers, Anode Buffer Container (ABC) 3500 Series, Cathode Buffer Container (CBC) 3500 Series e Conditioning Reagent 3500 Series (Life Technologies).

Tabela 7: *Primers* utilizados no sequenciamento de cada éxon

Segmento Gênico	Éxon	Primer	Seqüência 5-3'	Tamanho do éxon em pb
<i>IGHG1</i>	CH1	IGHG1_CH1_F1	ctggcaccctcctccaa	294
		IGHG1_Ex1_R3	cctgggttagggcacct	
		IGHG1_CH1_R1	gcttgattcacgtgcaga	
	CH2	IGHG1_Ex1_F4	acacctctcctcccagatt	330
		IGHG1_CH2_R1	tgactgtactccttgcca	
	CH3	IGHG1_CH2_F1	cggaccctgaggtcaca	323
IGHG1_Ex3_F1		ccaaaggtgggaccctg		
<i>IGHG2</i>	CH1	IGHG2_Ex1_R3	gacaccctccctccctga	294
		IGHG2_H_R2	gcactcgacacaacatttcg	
	CH2	IGHG2_Ex1_F5	cccaaaggccaaactg	327
		IGHG2_H_F1	agcgcaaattgtgtctg	
	CH3	IGHG2_CH2_F2	ataatgccaagacaaagcca	323

		IGHG2_CH2_F3	cagcgtcctcaccgtyg	
		IGHG_R1	cactcaggcctcagactcg	
IGHG3	CH1	IGHG3_F1	gctctgtcccacaccgca	294
		IGHG3_Ex1_R2	ccgggtgaggagtcagac	
	CH2	IGHG3_Ex6_R1	cagggcagagggtgggt	330
		IGHG3_CH3_R1	gtggttctcggggctgt	
	CH3	IGHG3_R2	atcctcgcgcgaccc	320
		IGHG3_Ex6_F1	ggacagaggccagcttga	

pb: pares de bases

5.4 LEITURA DAS SEQUÊNCIAS

A visualização dos cromatogramas e verificação da qualidade das sequências foram realizadas através dos programas Chromas Lite versão 2.1 (Technelysium). O alinhamento das sequências e identificação dos sítios com variações foi feito com Mutation Surveyor® DNA Variant Analysis Software v5.0.1 (Softgenetics). Com base nos sítios com variações os alelos de cada indivíduo foram definidos utilizando a base de dados do IMGT (LEFRANC et al., 2015), e os alótipos de cada gene foram definidos com a utilização dos dados descritos por LEFRANC e LEFRANC, 2012 (Tabela 3).

5.5 ANÁLISES ESTATÍSTICAS

Neste trabalho foram estudados diversos níveis de diversidade genética: variantes nucleotídicas, alelos, alótipos e haplótipos de alelos e alótipos. A partir do sequenciamento dos éxons dos segmentos gênicos foram encontradas variantes nucleotídicas, e elas foram associadas, quando possível a seus identificadores *rs*. Por sua vez, os alelos de um segmento gênico correspondem a variantes polimórficas na sua sequência nucleotídica e foram descritos na base de dados do IMGT. A partir da sequência nucleotídica é possível relacionar os alelos com os alótipos de Ig conhecidos. E finalmente os haplótipos de alelos (ou alótipos) são combinações de alelos (ou alótipos) de segmentos gênicos em um cromossomo.

As frequências alélicas e alótípicas foram obtidas por contagem direta e confirmadas utilizando o programa Genalex v6.502 (PEAKALL; SMOUSE, 2006), o qual também foi utilizado para gerar os arquivos de entrada para o programa Arlequin. O teste de equilíbrio de Hardy-Weinberg em relação as

frequências genotípicas e haplotípicas foi realizado para descartar erros na genotipagem das amostras (XU et al., 2002), de acordo com o método de GUO; THOMPSON (1992) utilizando o programa Arlequin v3.5.2 (EXCOFFIER; LISCHER, 2010).

Através do algoritmo ELB (Excoffier–Laval–Balding), um método pseudo-bayesiano para a inferência de haplótipos, foi reconstruída a fase gamética dos alelos de cada indivíduo (EXCOFFIER et al., 2005), sendo considerado um haplótipo como a combinação de alelos em cada cromossomo.

Foram realizadas análises filogenéticas com as sequências nucleotídicas dos alelos dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* com o método Neighbor-Joining (modelo Kimura 2-parâmetros), utilizando o programa MEGA v7 (KUMAR et al., 2016).

Para representar de forma gráfica a relação evolutiva entre os alelos de cada segmento gênico, foram construídas redes mediante o algoritmo Median-Joining (MJ). Estas redes conectam os alelos conforme suas semelhanças, representando em cada vetor a quantidade de variações, ou passos mutacionais, que explicam as transições de um alelo a outro (BANDELT et al., 1999). As redes foram realizadas utilizando o programa Network v5.0.

O desequilíbrio de ligação entre as variantes encontradas dentro das populações foi estimado com o software Haploview (BARRETT et al., 2005). Dessa maneira, são representados de forma gráfica os parâmetros D' /LOD e r^2 . O D' pode ser definido como um parâmetro do desequilíbrio de ligação influenciado principalmente pela quantidade de recombinação e LOD (*log of the likelihood odds ratio*) é uma medida da confiança no valor de D' . Por outro lado r^2 é o coeficiente de correlação entre dois *loci*, e pode assumir qualquer valor entre zero e D' .

A fim de estimar a diferenciação populacional devida a estruturação genética entre as populações foi utilizada a estatística F (F_{ST}) (REYNOLDS et al., 1983; SLATKIN, 1995) e realizado o teste exato de diferenciação populacional (RAYMOND; ROUSSET, 1995), com o programa Arlequin.

A partir das frequências alélicas dos segmentos gênicos de cada população, foram estimadas as distâncias genéticas entre as mesmas pelo método de Nei (NEI, 1972), utilizando 100.000 reamostragens pela técnica de *bootstraps*. Posteriormente, as distâncias foram representadas em um

dendrograma desenhado pelo método Neighbor-Joining (SAITOU; NEI, 1987) e visualizadas com o software TreeView version 3.5c (PAGE, 1996). Estas análises foram realizadas através do pacote de programas PHYLIP v3.695 (FELSENSTEIN, 2004).

A análise de componente principal a nível populacional foi realizada a partir das frequências alélicas dos segmentos gênicos de cada população com o programa estatístico Minitab 17 *Statistical Software* (2010). Também foi realizada uma análise de componente principal em nível de indivíduos, utilizando o genótipo de SNPs da Tabela A1, com o pacote PLINK v1.9 (PURCELL et al., 2007), e visualizada com o programa R (R CORE TEAM, 2008).

A estimativa da diversidade nucleotídica dentro das populações foi calculada através dos seguintes parâmetros: S , número de sítios segregantes, ou com variação, em um grupo de sequências alinhadas; Π (π), diferença média de nucleotídeos entre pares de sequências; e θ , diversidade genética esperada, calculada como $\theta = 4N_e\mu$, onde N_e é o tamanho efetivo da população e μ é a taxa de mutação por sítio por geração.

A hipótese de neutralidade seletiva dos alelos dos segmentos gênicos dentro das populações foi avaliada com os testes estatísticos D de Tajima (TAJIMA, 1989), D^* , F^* , D e F de Fu e Li (FU; LI, 1993), e H de Fay e Wu (FAY; WU, 2000). Estes testes foram realizados com o programa DnaSP (LIBRADO; ROZAS, 2009). O teste D de Tajima compara o número de sítios segregantes e a média de diferenças nucleotídicas entre as sequências (TAJIMA, 1989). O teste D^* de Fu e Li se baseia na diferença entre o número de mutações que aparecem somente uma vez (*singletons*) entre as sequências, e o número total de variantes nucleotídicas derivadas (* indica que não foi utilizada uma espécie como grupo externo). Espera-se que estes parâmetros tomem valores próximos a zero quando os alelos dentro de uma população são seletivamente neutros, ou não sofreram variação devido a fatores demográficos. Os estatísticos de Fu e Li se baseiam na genealogia dos genes, e comparam o número de mutações nos ramos internos (primitivos) e externos (derivados) com o esperado sob neutralidade seletiva. Por tanto, a presença de um grupo externo nas análises proporciona uma melhor definição na genealogia (FU; LI, 1993).

A significância destes testes é dada pela quantidade de vezes que esses parâmetros adotam valores iguais, ou mais extremos que o calculado, em simulações de coalescência, sob a hipótese de neutralidade seletiva e equilíbrio populacional. Todas as simulações de coalescência foram realizadas no programa DnaSP, considerando tanto theta ($\theta = 4N_e\mu$) como sítios segregantes, e recombinação elevada e ausência de recombinação.

O estatístico H de Fay e Wu compara mutações com frequências intermediárias e altas. Este teste também utiliza um grupo externo para polarizar as genealogias. Quando há neutralidade seletiva, espera-se que existam mutações de baixa frequência e não mutações de alta frequência.

Outra abordagem para testar o efeito da seleção natural em um gene é comparar a taxa substituições sinônimas (dS) e não sinônimas (dN) entre as sequências de seus alelos. Foi avaliada a hipótese nula de ausência de diferença entre essas taxas ($dN = dS$); e as hipóteses alternativas de seleção positiva ($dN > dS$) e seleção purificadora ($dN < dS$), utilizando o programa MEGA v7.

Foram realizados os testes F_s de Fu e R_2 de Ramos-Onsins e Rozas para detectar expansão populacional (RAMOS-ONSINS; ROZAS, 2002) utilizando o programa DnaSP (LIBRADO; ROZAS, 2009). O estatístico F_s de Fu estima, a partir do nível de diversidade observado, a probabilidade de obter uma amostra aleatória com um número de alelos igual ou menor que o observado, assumindo que todos os alelos são seletivamente neutros. R_2 compara a diferença entre o número de *singletons* e o número médio de diferenças nucleotídicas. A significância estatística destes testes também é obtida por simulações de coalescência.

6 RESULTADOS

Foram sequenciados os éxons *CH1*, *CH2* e *CH3* dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* em indivíduos pertencentes a cinco populações de ameríndios (GKW, GND, GRC, KIV e KRC) e em descendentes de Japoneses (JAP).

As sequências obtidas foram comparadas com as sequências de referência do IMGT e foram registrados os sítios com variação em cada indivíduo. Os sítios variáveis encontrados nos três segmentos gênicos nas populações deste estudo se encontram na Tabela 8. Foram encontradas duas variantes nucleotídicas não descritas previamente no éxon *CH1* do segmento gênico *IGHG3*, e que por tanto, não possuem identificador *rs*. As frequências dos SNPs que resultaram em substituições não sinônimas encontrados neste trabalho, foram comparados com as disponíveis no banco de dados do projeto *1000 genomes*. Estas populações foram escolhidas para representar europeus (FIN, Finlandeses em Finlândia), africanos (YRI, Iorubás em Nigéria), indosarianos (BEB, Bengalis em Bangladesh) e hispânicos (PEL, Peruanos em Peru) (AUTON et al., 2015) (Tabela A1 e Figura A1 a A3).

A partir dos sítios variáveis de cada indivíduo, foi realizada a inferência dos seus alelos, com base no banco de dados do IMGT (Tabela 3). Quando não foi possível atribuir dois alelos já descritos a um indivíduo, foi inferido um alelo novo (ainda não descrito na base de dado do IMGT), que pudesse se ajustar ao genótipo do indivíduo junto com o alelo mais comum da sua população, e foram nomeados com a letra “n”, seguida de um número consecutivo ao último alelo descrito no IMGT. A descrição dos sítios variantes nos alelos novos, os alótipos que codificam, e o número de indivíduos que apresentaram esse alelo, se encontram nas Tabelas 9 a 11. Os alótipos codificados por estes alelos foram inferidos com base na literatura (Tabela 3). Alguns alelos codificam combinações de alótipos não relatadas em outros estudos, como o alótipo Gm5,10,11,13,14,21,26,27 codificado pelo alelo *IGHG3*n24*, onde os alótipos Gm5 e Gm21 se encontram juntos.

Tabela 8: Sítios variáveis encontrados nos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* (continua)

Segmento	Éxon	rs	Localização ¹	Posição no éxon ²	Número IMGT ³	Nucleotídeo substituído	Aminoácido substituído	Alotipo ⁴	D de Grantham ⁵	Frequência ⁶
IGHG1	CH1	rs11552998	106209340	68	19	G>A				a
		rs1071803	106209119	289	120	A>G	K>R	Gm17>Gm3	26	c
	CH2	rs587690960	106208471	89	22	A>G				a
		rs377538050	106208364	196	84.3	A>T	Y>F		22	a
		rs193160354	106208327	233	91	C>T				b
		rs1043109	106208326	234	92	C>G	L>V		32	b
		rs1043249	106208306	254	98	T>C				b
	CH3	rs11557940	106208107	26	6	C>T				a
		rs1045853	106208086	47	12	T>G	D>E	Gm1>nGm1	45	b
		rs11621259	106208082	51	14	C>A	L>M	Gm1>nGm1	15	b
		rs17841087	106207933	200	86	C>T				b
		rs113804727	106207862	271	110	C>G	A>G	>Gm2	60	c
		rs12879979	106207822	311	124	T>C				b
IGHG2	CH1	rs771078711	106111097	30	7	C>G	L>V		32	a
		rs189328740	106111071	56	15	C>T				b
		rs587648672	106111069	58	16	A>G	E>G		98	b
		rs773818177	106111067	60	17	A>G	S>G		56	b
		rs11557955	106110966	161	82	A>G				c
		rs11627594	106110914	213	92	C>A	P>T		38	c
	CH2	rs8009156	106110137	150	45,1	G>A	V>M	Gm(..)>23	21	c
		rs11160859	106110057	230	91	T>C				c
		rs113678609	106110056	231	92	G>C	V>L		32	a
	CH3	rs587682450	106109825	38	9	A>C				b
		rs368359789	106109708	155	79	G>C	K>N		94	a
		rs1049810	106109702	161	81	A>G				c
		rs28371022	106109573	290	117	G>A				c

1 Localização no cromossomo 14 descrito no *GRCh37.p13 primary assembly*.

2 Posição a partir do primeiro nucleotídeo de cada éxon

3 Posição do aminoácido de acordo com o IMGT (International ImMunoGeneTics Information System). (LEFRANC et al., 2015).

4 A substituição de aminoácidos resulta no determinante antigênico do alotipo mencionado quando encontrar-se após >, e na perda do alotipo quando encontrar-se antes de >. Quando não indicada a substituição de aminoácidos a substituição nucleotídica gera uma mudança sinônima.

5 Distância evolutiva entre os aminoácidos envolvidos na substituição de acordo com (GRANTHAM, 1974). Esta escala pode ter valores desde 5 até 215 baseadas nas diferenças químico-físicas entre aminoácidos. Quanto a maior valor, maior diferença entre aminoácidos.

6 Frequência do menor alelo (MAF) na amostra total (Ameríndios+Japoneses) a, MAF < 1%; b, 1 > MAF > 5%; c, 5 > MAF > 30%; d, MAF >30%.

Tabela 8: Sítios variáveis encontrados nos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* (continuação)

Segmento	Éxon	rs	Localização ¹	Posição no éxon ²	Número IMGT ³	Nucleotídeo substituído	Aminoácido substituído	Alótipo ⁴	D de Grantham ⁵	Frequência ⁶	
IGHG3	CH1	rs373549707	106237687	56	15	T>C				a	
		rs756482499	106237685	58	16	G>A	G>E		98	a	
		rs750821345	106237683	60	17	G>A	G>S		56	a	
		rs369012923	106237675	68	19	G>C				a	
		rs12050095	106237624	119	40	G>A				b	
		novo*	106237514	129	44	G>T	G>Y		147	a	
	novo*	106237513	130	44	G>A	G>Y		147	a		
	CH2	rs138869693	106236202	122	35	C>T					a
		rs145035200	106236195	129	38	C>A	Q>K		53	a	
		rs74093865	106236143	181	82	C>T	P>L	>Gm21	98	d	
		rs60746425	106236141	183	83	C>T	R>W	>Gm16	101	b	
		rs12890621	106236128	196	84,3	A>T	Y>F		22	b	
		rs141959627	106236000	324	124	A>G	T>A		58	a	
	CH3	rs113169458	106235783	114	39	G>A	V>M	>Gm15	21	b	
		rs77307099	106235767	130	44	G>A	S>N	Gm11>nGm11	46	d	
		rs78376194	106235766	131	44	C>T	S>N	Gm11>nGm11	46	d	
		rs587739524	106235758	139	45,2	C>G	P>R		103	a	
		rs149653267	106235742	155	79	C>G	N>K		94	b	
		rs139413052	106235729	168	84	A>G	M>V	Gm14>	21	b	
		rs4042056	106235614	283	115	G>A	R>H	>Gm15	29	c	
rs1051112		106235611	286	116	T>A	F>Y	>Gm15	22	c		

* Identificador rs não descrito para o SNP;

1 Localização no cromossomo 14 descrito no *GRCh37.p13 primary assembly*.

2 Posição a partir do primeiro nucleotídeo de cada éxon

3 Posição do aminoácido de acordo com o IMGT (International ImMunoGeneTics Information System). (LEFRANC et al., 2015).

4 A substituição de aminoácidos resulta no determinante antigênico do alótipo mencionado quando encontrar-se após >, e na perda do alótipo quando encontrar-se antes de > Para a formação do alótipo Gm15 e necessária a substituição simultânea nos sítios CH3 39, 115 e 116.

Quando não indicada a substituição de aminoácidos a substituição nucleotídica gera uma mudança sinônima.

5 Distância evolutiva entre os aminoácidos envolvidos na substituição de acordo com (GRANTHAM, 1974). Esta escala pode ter valores desde 5 até 215 baseadas nas diferenças químico-físicas entre aminoácidos. Quanto a maior valor, maior diferença entres aminoácidos.

6 Frequência do menor alelo (MAF) na amostra total (Ameríndios+Japoneses) a, MAF < 1%; b, 1 > MAF > 5%; c, 5 > MAF > 30%; d, MAF >30%.

Tabela 9: Alelos novos identificados em *IGHG1*

	Éxon	CH1							CH2					CH3					#
	N° IMGT*	19	22	84.3	85,1	91	92	98	6	12	14	86	124						
Referencia	G	A	A	C	C	C	T	C	T	C	C	T							
Alelo/Alótipo																			
<i>IGHG1*n9</i> / <u>G1m17,1</u>	A	.	.	T	1					
<i>IGHG1*n10</i> / G1m17,1	.	G	.	T	2					
<i>IGHG1*n11</i> / G1m17,1	.	.	.	T	T	G	C	7					
<i>IGHG1*n12</i> / G1m17,1	.	.	.	T	.	.	.	T	1					
<i>IGHG1*n13</i> / <u>G1m17</u>	.	.	.	T	G	A	T	C	.	2					
<i>IGHG1*n14</i> / G1m17,1	.	.	T	T	1					

Alelos com fase inferida por desequilíbrio de ligação e frequência de alelos comuns. Esses alelos não foram confirmados por sequenciamento de uma única fita.

* Posição do nucleotídeo de acordo com o IMGT (International ImMunoGeneTics Information System). (LEFRANC et al., 2015).

Número de indivíduos que apresentaram este alelo

Alótipos sublinhados são combinações de alótipos novas.

Apenas os nucleotídeos diferentes da sequência consenso são mostrados.

Tabela 10 Alelos novos identificados em *IGHG2*

Alelo/Alótipo	Éxon CH1					CH2					CH3					#
	N° IMGT*	7	15	16	17	19	82	92	45,1	91	92	9	79	81	117	
	Referencia	C	C	A	A	C	A	C	G	T	G	A	G	A	G	
<i>IGHG2*n7/G2m..</i>	G	.	.	.	C	.	.	.	G	.	6	
<i>IGHG2*n8/G2m..</i>	G	G	.	.	C	23	
<i>IGHG2*n9/G2m..</i>	G	G	.	.	C	.	.	C	.	.	1	
<i>IGHG2*n10/G2m..</i>	.	T	G	G	G	14	
<i>IGHG2*n11/G2m..</i>	G	C	.	.	.	12	
<i>IGHG2*n12/G2m..</i>	G	G	.	.	C	C	.	.	G	.	2	
<i>IGHG2*n13/G2m23</i>	G	.	.	.	G	G	A	A	C	A	1	
<i>IGHG2*n14/G2m..</i>	G	G	3	
<i>IGHG2*n15/G2m..</i>	G	G	A	2	

Alelos com fase inferida por desequilíbrio de ligação e frequência de alelos comuns. Esses alelos não foram confirmados por sequenciamento de uma única fita.

Os alelos em negrito foram confirmados por sequenciamento em indivíduos homocigotos (ausência de sítios heterocigotos).

* Posição do nucleotídeo de acordo com o IMGT (International ImMunoGeneTics Information System). (LEFRANC et al., 2015).

Número de indivíduos que apresentaram este alelo

Apenas os nucleotídeos diferentes da sequência consenso são mostrados.

Tabela 11 Alelos novos identificados em *IGHG3*

Alelo/Alótipo	Éxon CH1				CH2				CH3								#				
	N°	15	16	17	19	40	44	44	35	38	82	83	39	44	44	45,2		79	84	115	116
	Referencia	T	G	G	G	G	G	G	C	C	C	C	G	G	C	C		C	A	G	T
<i>IGHG3</i> *n20/ <u>Gm28*</u> (10,11,13,16,26,27,28)	T	A	1
<i>IGHG3</i>*n21/<u>Gm27*</u> (21,27)	T	.	.	A	T	.	.	.	A	A	21
<i>IGHG3</i>*n22/ G3m5* (Gm5,10,11,13,14,26,27)	A	14
<i>IGHG3</i> *n23/ G3m21* (Gm21,26,27,28)	A	T	.	.	A	T	A.	1
<i>IGHG3</i> *n24/ <u>G3m5,21*</u> (5,10,11,13,14,21,26,27)	T	1
<i>IGHG3</i> *n25/ G3m21* Gm21,26,27,28	T	.	.	A	T	G	.	.	.	A	2
<i>IGHG3</i> *n26/ <u>G3m13*</u> (10,11,13,16,27)	T	G	G	A	A	1
<i>IGHG3</i> *n27/ G3m5* (5,10,11,13,14,26,27)	A	T	A	1
<i>IGHG3</i> *n28/ G3m21* (21,26,27,28)	.	C	A	A	C	T	.	A	T	G	1
<i>IGHG3</i> *n29/ G3m21* (21,26,27,28)	T	A	T	.	.	A	T	G	3

Alelos com fase inferida por desequilíbrio de ligação e frequência de alelos comuns. Esses alelos não foram confirmados por sequenciamento de uma única fita.

Os alelos em negrito foram confirmados por sequenciamento em indivíduos homocigotos (ausência de sítios heterocigotos).

Alótipos sublinhados são combinações de alótipos novas.

* Posição do nucleotídeo de acordo com o IMGT (International ImMunoGeneTics Information System). (LEFRANC et al., 2015).

Número de indivíduos que apresentaram este alelo

Apenas os nucleotídeos diferentes da sequência consenso são mostrados.

Os alelos novos *IGHG2**n8, *n10 e *IGHG3**n21, *n22 foram encontrados em homozigose (Tabela 14). Os alelos *IGHG1**n11 e *IGHG2**n11 não foram encontrados em homozigose, mas se encontram em um número elevado de indivíduos e em diferentes combinações genóticas (Tabela 10 a 11).

A partir dos sítios variáveis de cada indivíduo, foi realizada a inferência de fases dos nucleotídeos, utilizando o algoritmo ELB (Excoffier–Laval–Balding) (EXCOFFIER et al., 2003). A inferência de todos os alelos feita com o programa correspondeu com aquela feita manualmente, com exceção do alelo *IGHG2**05, encontrado na população de descendentes de Japoneses. De acordo com a inferência do algoritmo, a combinação de SNPs sugerida criaria novos alelos. Optamos por manter alelo *IGHG2**05, já que ele foi previamente confirmado e se encontra na base de dados do IMGT (Tabela 14).

As frequências alélicas, genóticas e alotípicas de cada segmento gênico se encontram nas Tabelas 12 a 14. As populações se encontram em equilíbrio de Hardy-Weinberg (HW) para os três segmentos gênicos estudados, com exceção da população de descendentes de Japoneses para *IGHG2* ($p=0,00418$). Porém, quando analisamos o equilíbrio de Hardy-Weinberg para os haplótipo formados pelos três segmentos, todas as populações se encontram em equilíbrio de HW (Tabela 15). Todos os alelos novos de *IGHG1* e *IGHG3* são exclusivos ou da população de ameríndios ou dos descendentes de Japoneses, já para *IGHG2*, as populações de ameríndios e descendentes de Japoneses compartilham quatro dos nove alelos novos, *IGHG2**n07, *n08 e *n15 (Tabela 12).

Tabela 12 Frequência alélica dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* nas populações

Alelo	GKW	GND	GRC	KIV	KRC	JAP
n	51	51	48	49	46	54
<i>IGHG1*02</i>	0,471	0,451	0,740	0,765	0,620	0,611
<i>IGHG1*03</i>	0,010	0,078	0,021	0,020	0,011	
<i>IGHG1*07</i>	0,510	0,461	0,229	0,204	0,359	0,120
<i>IGHG1*08</i>	0,010			0,010		0,167
<i>IGHG1*n09</i>						0,009
<i>IGHG1*n10</i>						0,019
<i>IGHG1*n11</i>						0,065
<i>IGHG1*n12</i>			0,010			
<i>IGHG1*n13</i>		0,010			0,011	
<i>IGHG1*n14</i>						0,009
<i>IGHG2*02</i>		0,078	0,021	0,031	0,022	0,139
<i>IGHG2*03</i>	0,961	0,882	0,740	0,949	0,946	0,528
<i>IGHG2*05</i>						0,056
<i>IGHG2*n07</i>	0,029				0,011	0,019
<i>IGHG2*n08</i>				0,020	0,011	0,194
<i>IGHG2*n09</i>						0,009
<i>IGHG2*n10</i>	0,010	0,010	0,135			
<i>IGHG2*n11</i>		0,029	0,094			
<i>IGHG2*n12</i>						0,019
<i>IGHG2*n13</i>						0,009
<i>IGHG2*n14</i>					0,011	0,019
<i>IGHG2*n15</i>			0,010			0,009
<i>IGHG3*11</i>	0,010	0,078	0,010	0,031	0,022	
<i>IGHG3*14</i>	0,990	0,843	0,844	0,959	0,967	0,491
<i>IGHG3*15</i>						0,009
<i>IGHG3*16</i>				0,010		
<i>IGHG3*19</i>						0,250
<i>IGHG3*n20</i>						0,009
<i>IGHG3*n21</i>		0,069	0,146		0,011	
<i>IGHG3*n22</i>						0,157
<i>IGHG3*n23</i>						0,009
<i>IGHG3*n24</i>		0,010				
<i>IGHG3*n25</i>						0,019
<i>IGHG3*n26</i>						0,009
<i>IGHG3*n27</i>						0,009
<i>IGHG3*n28</i>						0,009
<i>IGHG3*n29</i>						0,028

n: Número de indivíduos analisados em cada população

Os alelos de com a letra "n" antes do número representam alelos novos

Os alelos em negrito foram confirmados por sequenciamento em indivíduos homocigotos (ausência de sítios heterocigotos).

Tabela 13 Frequência alotípica G1m, G2m e G3m nas populações

Alótipo Gm	GKW	GND	GRC	KIV	KRC	JAP
n	51	51	48	49	46	54
G1m 17,1,2	0,510	0,461	0,229	0,204	0,359	0,120
G1m 17,1	0,471	0,451	0,750	0,765	0,620	0,713
G1m 17		0,010			0,011	
G1m 3	0,010	0,078	0,021	0,020	0,011	
G1m 3,1	0,010			0,010	0,000	0,167
G2m (..)	1,000	0,922	0,979	0,969	0,978	0,796
G2m 23		0,078	0,021	0,031	0,022	0,204
<u>G3m21*</u> (21,26,27,28)	0,990	0,843	0,844	0,969	0,967	0,565
<u>G3m5*</u> (5,10,11,13,14,26,27)	0,010	0,078	0,010	0,031	0,022	0,167
<u>G3m16*</u> (10,11,13,15,16,27)						0,250
<u>G3m27*</u> (21,27)		0,069	0,146	0,000	0,011	
<u>G3m13*</u> (10,11,13,16,27)						0,009
<u>Gm28*</u> (10,11,13,16,26,27,28)						0,009
<u>G3m5,21*</u> (5,10,11,13,14,21,26,27)		0,010				

Alótipos sublinhados são combinações de alótipos novas

Tabela 14 Frequência genotípica dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3*

Segmento gênico	Genotipos	GKW	GND	GRC	KIV	KRC	JAP
	<i>n</i>	51	51	48	49	46	54
<i>IGHG1</i>	*02 / *02	0,2	0,29	0,52	0,55	0,39	0,41
	*02 / *03		0,04	0,04	0,04		
	*02 / *07	0,55	0,27	0,38	0,37	0,43	0,19
	*02 / *08				0,02		0,11
	*03*07	0,02	0,12			0,02	
	*07*07	0,22	0,25	0,04	0,02	0,13	
	*07 / *08	0,02					0,06
	*08 / *08						0,04
	*02 / *n9						0,02
	*02 / *n10						0,04
	*02 / *n11						0,04
	*08 / *n11						0,09
	*02 / *n12				0,02		
	*02 / *n13						0,02
*07 / *n13			0,02				
*02 / *n14						0,02	
	HW (<i>p</i>)	0,43712	0,06915	1,000	0,84218	0,74120	0,14596
<i>IGHG2</i>	*02 / *02						0,04
	*02 / *03		0,16	0,04	0,06	0,04	0,15
	*02 / *05						0,04
	*03 / *03	0,92	0,76	0,52	0,9	0,89	0,24
	*03 / *n7	0,06				0,02	
	*02 / *n8						0,02
	*03 / *n8				0,04	0,02	0,33
	*03 / *n10	0,02	0,02	0,21			
	*03 / *n11		0,06	0,17			
	*03 / *n12						0,04
	*03 / *n13						0,02
	*03 / *n14					0,02	0,02
	*03 / *n15			0,02			0,02
	*05 / *n7						0,04
	*05 / *n9						0,02
	*05 / *n14						0,02
*n8 / *n8						0,02	
*n10 / *n10				0,02			
*n10 / *n11				0,02			
	HW (<i>p</i>)	1,000	1,000	1,000	1,000	1,000	0,00418
<i>IGHG3</i>	*11 / *14	0,02	0,16	0,02	0,06	0,04	
	*14 / *14	0,98	0,69	0,71	0,92	0,93	0,28
	*14 / *16				0,02		
	*14 / *19						0,2
	*15 / *19						0,02
	*19 / *19						0,06
	*14 / *n20						0,02
	*14 / n21		0,14	0,25		0,02	
	*14 / n24		0,02				
	*n21 / *n21			0,02			
	*14 / *n22						0,13
	*19 / *n22						0,11
	*n22 / *n22						0,02
	*19 / *n23						0,02
	*14 / *n25						0,04
	*14 / *n26						0,02
	*n22 / *n27						0,02
	*n22 / *n28						0,02
	*14 / *n29						0,02
*19 / *n29						0,04	
	HW (<i>p</i>)	1,000	1,000	1,000	1,000	1,000	0,59500

Os alelos com a letra "n" antes do número representam alelos novos

n: Número de indivíduos analisados em cada população

HW (*p*): significância estatística do teste de equilíbrio de Hardy-Weinberg

Os alelos mais comuns foram *IGHG1*02* ($fr = \sim 0,7$ em GRC, KIV e KRC e $fr = 0,6$ em JAP), *IGHG1*07* ($fr = \sim 0,5$ em GWK e GND), *IGHG2*03* ($fr = \sim 0,9$ em ameríndios e $fr = 0,52$ em JAP) e *IGHG3*014* ($fr = \sim 0,9$ em ameríndios e $fr = 0,49$ em Japoneses). E os alótipos mais comuns foram G1m17,1 (~71% em GRC, KIV e KRC e 73% em JAP), G1m17,1,2 (~48% em GWK e GND), G2m(..) (~97% em ameríndios e 79,6% em JAP) e G3m21* (~92% em ameríndios e 56,5% em JAP).

Foram realizadas análises filogenéticas entre os alelos dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* com o método Neighbor-Joining (modelo Kimura 2-parâmetros) (Figura 15 a 17). Para a realização das árvores foram incluídos todos os alelos de cada segmento gênico já descrito no IMGT, e os alelos novos com frequência maior a 0,05 ou aqueles encontrados em homozigose. Cada alelo foi relacionado com o alótipo que codifica.

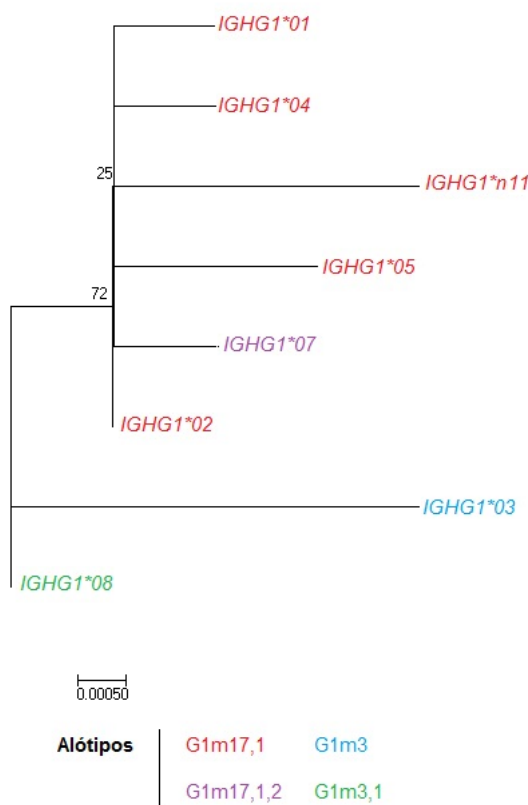


Figura 15: Árvore Neighbor-Joining (modelo Kimura 2-parâmetros) gerada com as sequências alélicas do segmento gênico *IGHG1*. A porcentagem de árvores que agruparam os alelos do mesmo modo usando *bootstrap* de 1000 é mostrada próximo aos ramos. A árvore é escalonada, de modo que o comprimento dos ramos representa as distâncias evolutivas usadas na inferência da árvore. As cores representam o alótipo que cada alelo codifica. Os alelos com "n" no nome são alelos novos.

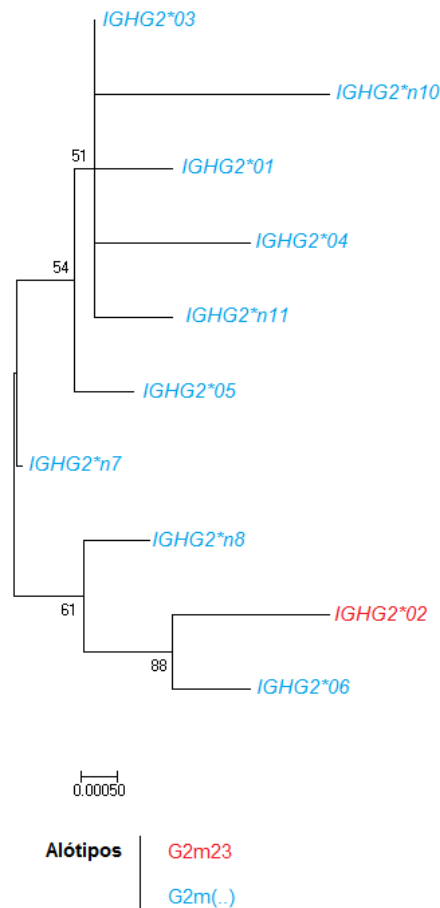


Figura 16: Árvore Neighbor-Joining (modelo Kimura 2-parâmetros) gerada com as sequências alélicas do segmento gênico *IGHG2*. A porcentagem de árvores que agruparam os alelos do mesmo modo usando *bootstrap* de 1000 é mostrada próximo aos ramos. A árvore é escalonada, de modo que o comprimento dos ramos representa as distâncias evolutivas usadas na inferência da árvore. As cores representam o alótipo que cada alelo codifica. Os alelos com “n” no nome são alelos novos.

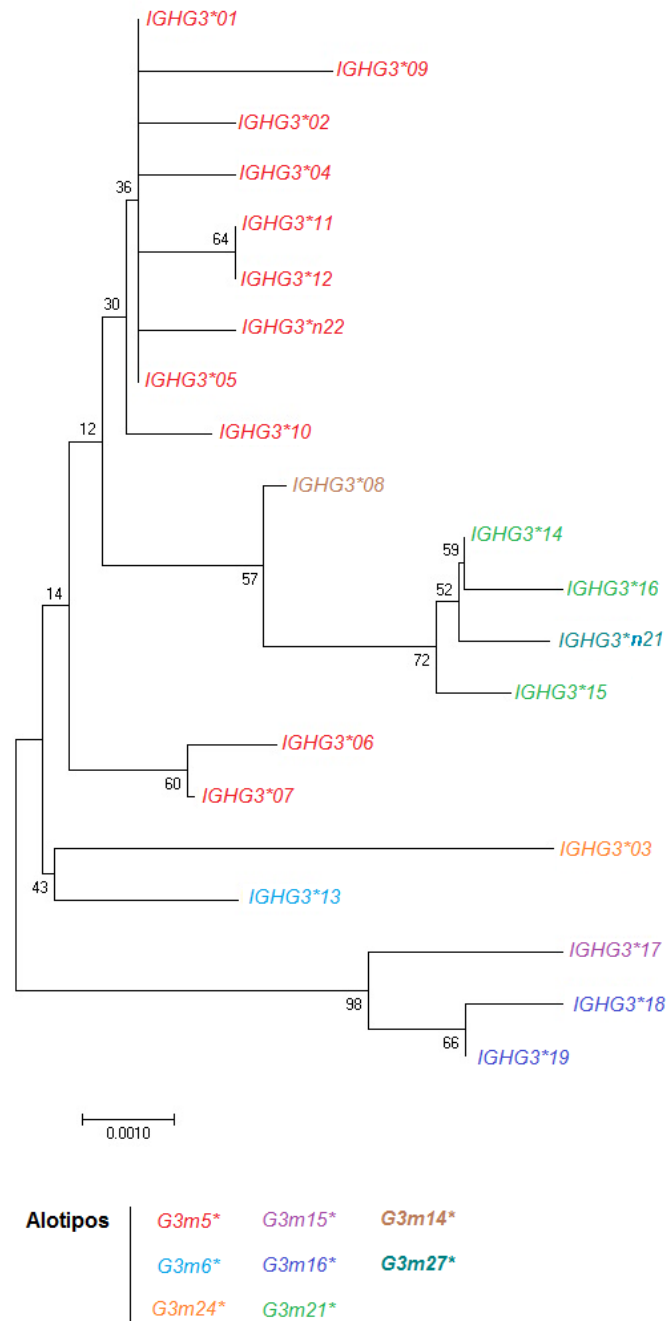


Figura 17 Árvore Neighbor-Joining (modelo Kimura 2-parâmetros) gerada com as sequências alélicas do segmento gênico *IGHG3*. A porcentagem de árvores que agruparam os alelos do mesmo modo usando *bootstrap* de 1000 é mostrada próximo aos ramos. A árvore é escalonada, de modo que o comprimento dos ramos representa as distâncias evolutivas usadas na inferência da árvore. As cores representam o alótipo que cada alelo codifica. Os alelos com “n” no nome são alelos novos. Os alótipos em negrito (G3m14 e G3m27) representam combinações alotípicas novas (Tabela 13)

Foi realizada a inferência de haplótipos de alelos e alótipos de todos os indivíduos em relação a cada um dos três segmentos gênicos. A frequência de haplótipos de alótipos dentro de cada população se encontra na Tabela 15. A vírgula separa diferentes alótipos do mesmo segmento gênico; o ponto e vírgula separa alótipos de segmentos gênicos diferentes. Estes se encontram na mesma ordem que no cromossomo, assim primeiro se descrevem alótipos de G3m, logo de G1m e finalmente de G2m. Os dois pontos (..) indicam que o alótipo G2m23, único de G2m, está ausente naquele haplótipo. Neste estudo foram encontrados haplótipos que não foram relatados em outras publicações.

Tabela 15: Frequência haplotípicas dos alótipos de Ig dentro de cada população

Haplótipos	GKW	GND	GRC	KIV	KRC	JAP
<i>n</i>	51	51	48	49	46	54
21,26,27,28;17,1;(..)	0,4704	0,3824	0,6042	0,7656	0,6088	0,43512
21,26,27,28;17,1,2;(..)	0,5098	0,4608	0,2287	0,204	0,3589	0,12038
10,11,13,15,16,27;17,1;(..)						0,26858
5,10,11,13,14,26,27;3,1;23						0,12952
5,10,11,13,14,26,27;3;(..)	0,0098					
21,26,27,28;3,1;(..)	0,0098					
<u>21,27;17,1;(..)</u>		0,0686	0,146		0,0109	
5,10,11,13,14,26,27;3;23		0,0686	0,0104	0,0204	0,0109	
<u>5,10,11,13,14,26,27;17;23</u>		0,0098			0,0109	
<u>5,10,11,13,14,21,26,27;3;(..)</u>		0,0098				
21,26,27,28;3;23			0,0104			
5,10,11,13,14,26,27;3,1;(..)						0,00926
5,10,11,13,14,26,27;17,1;23						0,00926
<u>10,11,13,16,26,27,28;17,1;(..)</u>						0,00926
<u>10,11,13,16,27;17,1;(..)</u>						0,00926
21,26,27,28;3,1;23						0,00926
HW (ρ):	0.68288	0.57431	0.87647	0.88990	0.96539	0.11245

n: Número de indivíduos analisados em cada população

Alótipos sublinhados são combinações de alótipos novas

HW (ρ): significância estatística do teste de equilíbrio de Hardy-Weinberg

A partir das sequências dos alelos, e considerando cada alelo como um haplótipo de SNPs, foram geradas redes de haplótipos (ou de alelos) de cada segmento gênico (Figuras 18 a 20). Estas redes mostram cada alelo como um nó e o tamanho do nó indica o número de vezes que ele está presente na

amostra total, sendo diferenciada a frequência dentro de cada população por uma cor diferente. Os ramos da rede representam os passos mutacionais que diferenciam um alelo do outro.

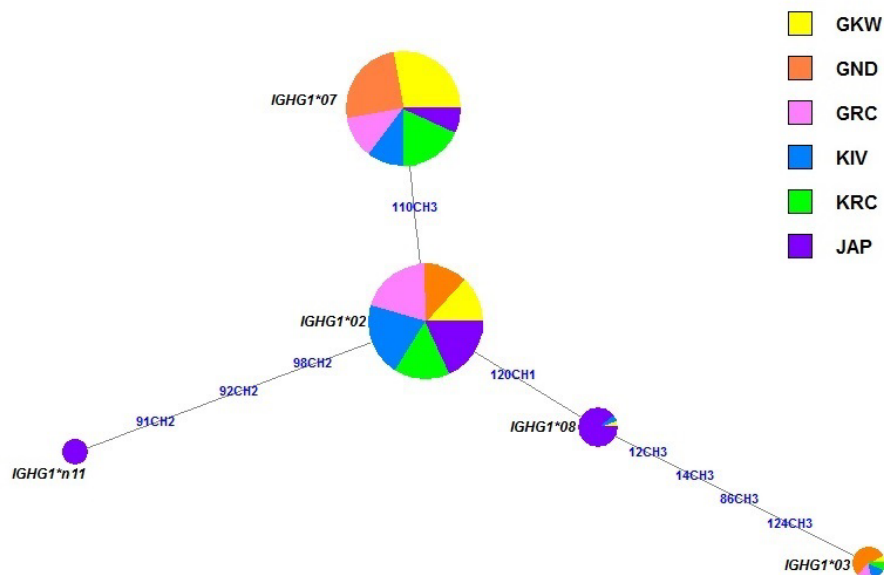


Figura 18: Rede Median-Joining dos alelos do segmento gênico *IGHG1*. Cada nó representa um alelo e o tamanho de nó é proporcional ao número de vezes que o alelo aparece na população agrupada. O comprimento dos ramos é proporcional ao número de mudanças de nucleotídeos entre alelos, indicados pela posição e pelo éxon onde acontecem.

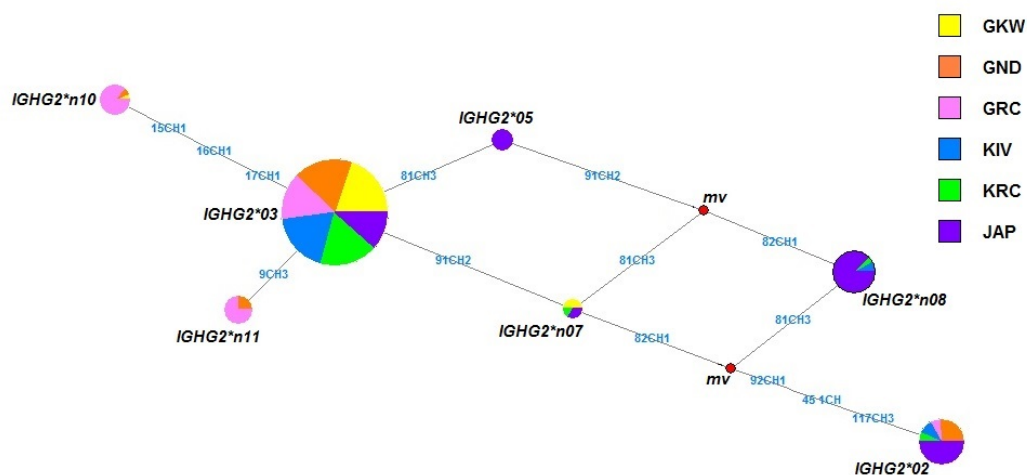


Figura 19 Rede Median-Joining dos alelos do segmento gênico *IGHG2*. Cada nó representa um alelo e o tamanho de nó é proporcional ao número de vezes que o alelo aparece na população agrupada. O comprimento dos ramos é proporcional ao número de mudanças de nucleotídeos entre alelos, indicados pela posição e pelo éxon onde acontecem. Os nós "mv" (do inglês *median vectors*) são alelos gerados pelo algoritmo MJ para conectar os alelos das populações.

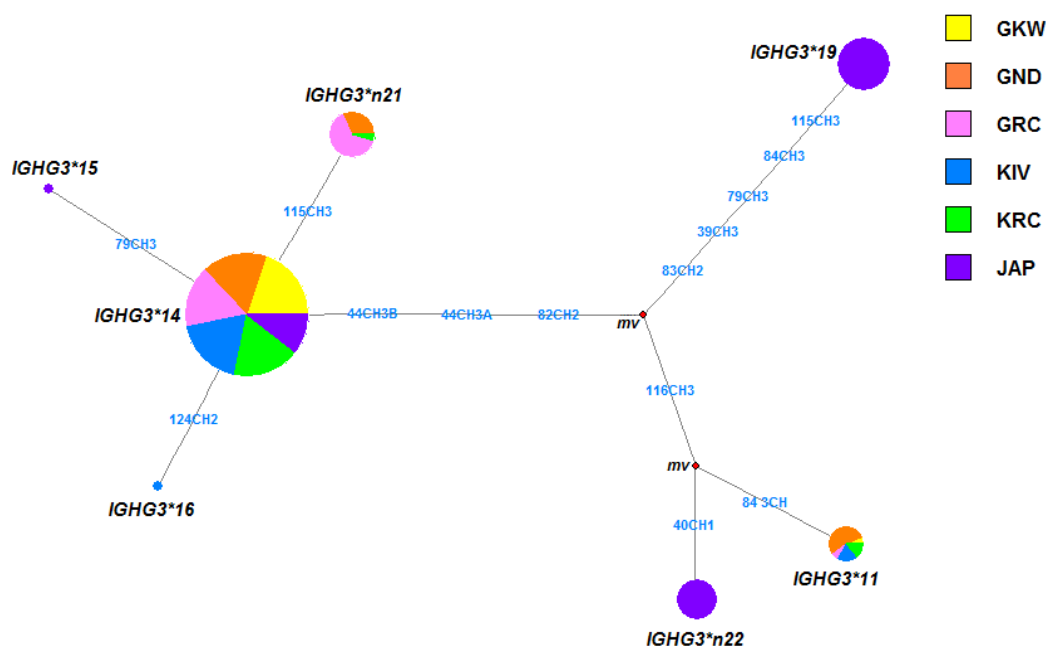


Figura 20: Rede Median-Joining dos alelos do segmento gênico *IGHG3*. Cada nó representa um alelo e o tamanho de nó é proporcional ao número de vezes que o alelo aparece na população agrupada. O comprimento dos ramos é proporcional ao número de mudanças de nucleotídeos entre alelos, indicados pela posição e pelo éxon onde acontecem.

O desequilíbrio de ligação entre os SNPs dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* dentro das populações foi estimado utilizando o programa Haploview (Figuras 21 a 26). Nele é possível observar que os valores máximos de D' , entre dois SNPs, estão representados pela cor vermelha, quando o LOD (*likelihood of D'*) é alto, e os valores mínimos de D' pela cor cinza, quando o LOD é alto, e branca quando o LOD é baixo. E dentro de cada quadrado estão representados os valores de r^2 (de 0 a 1) entre dois SNPs. Os blocos representados com um triângulo dividem cada segmento gênico.

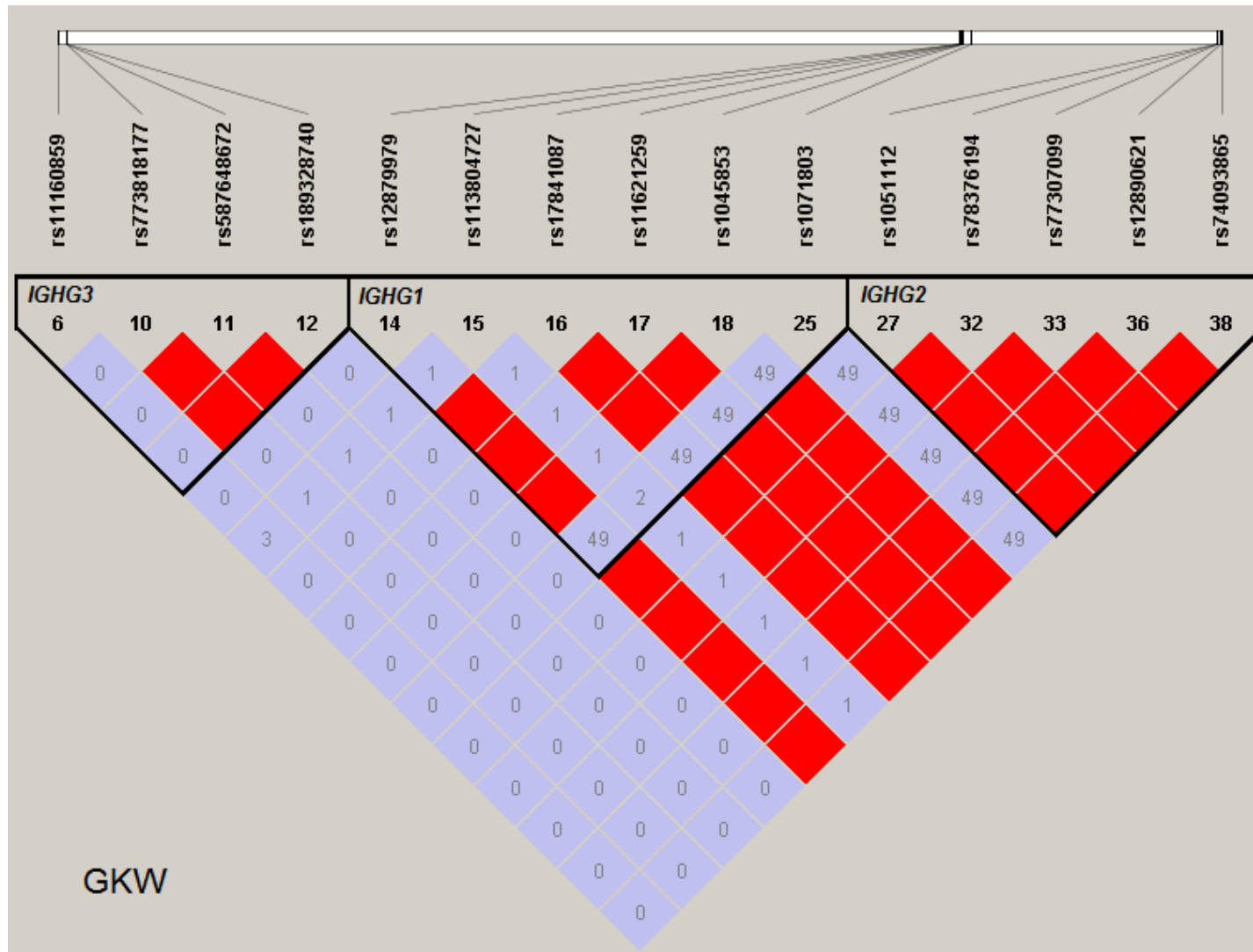


Figura 21: Desequilíbrio de ligação entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* na população Guarani Kaiowa (GKW). Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza. Dentro de cada losango estão representados os valores de $r^2 \times 100$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).

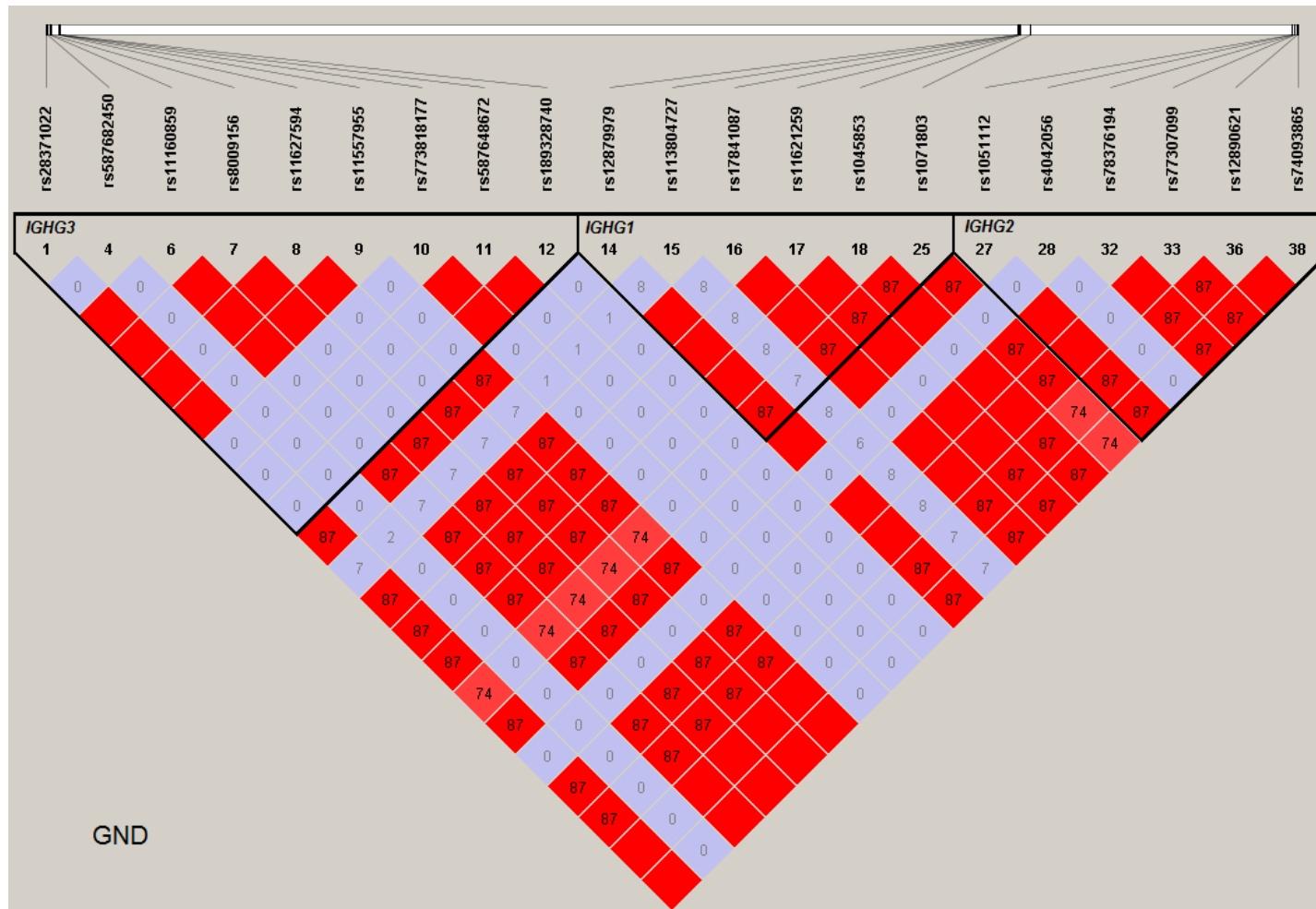


Figura 22 Desequilíbrio de ligação entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* na população Guarani Nandeva (GND). Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza. Dentro de cada losango estão representados os valores de $r^2 \times 100$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).

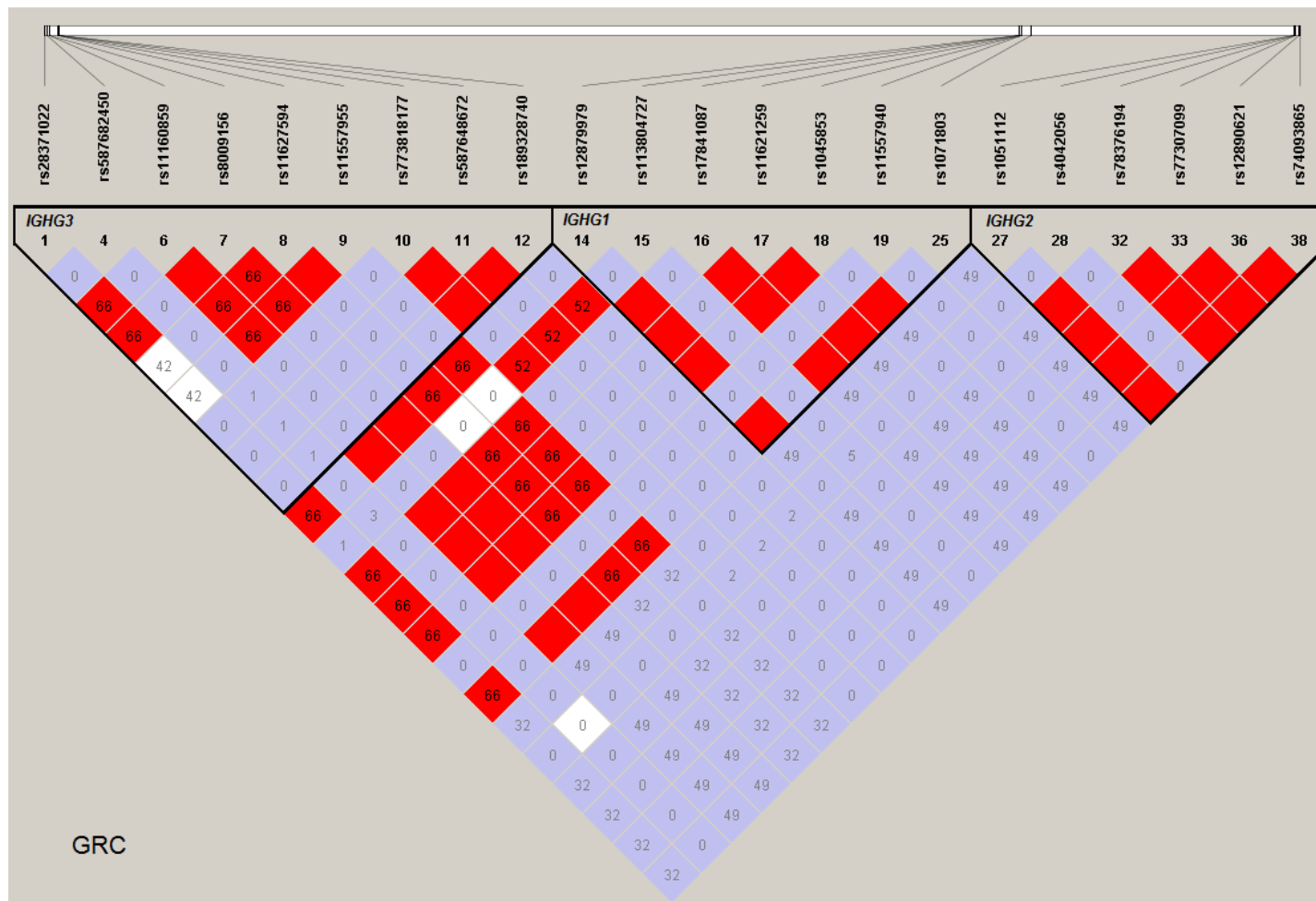


Figura 23: Desequilíbrio de ligação entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* na população Guarani Mbya (GRC). Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza ou branco. Dentro de cada losango estão representados os valores de $r^2 \times 100$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).



Figura 24: Desequilíbrio de ligação entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* na população Kaingang de Ivai (KIV). Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza. Dentro de cada losango estão representados os valores de $r^2 \times 100$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).



Figura 25: Desequilíbrio de ligação entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* na população Kaingang de Rio das Cobras (KRC). Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza. Dentro de cada losango estão representados os valores de $r^2 \times 100$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).

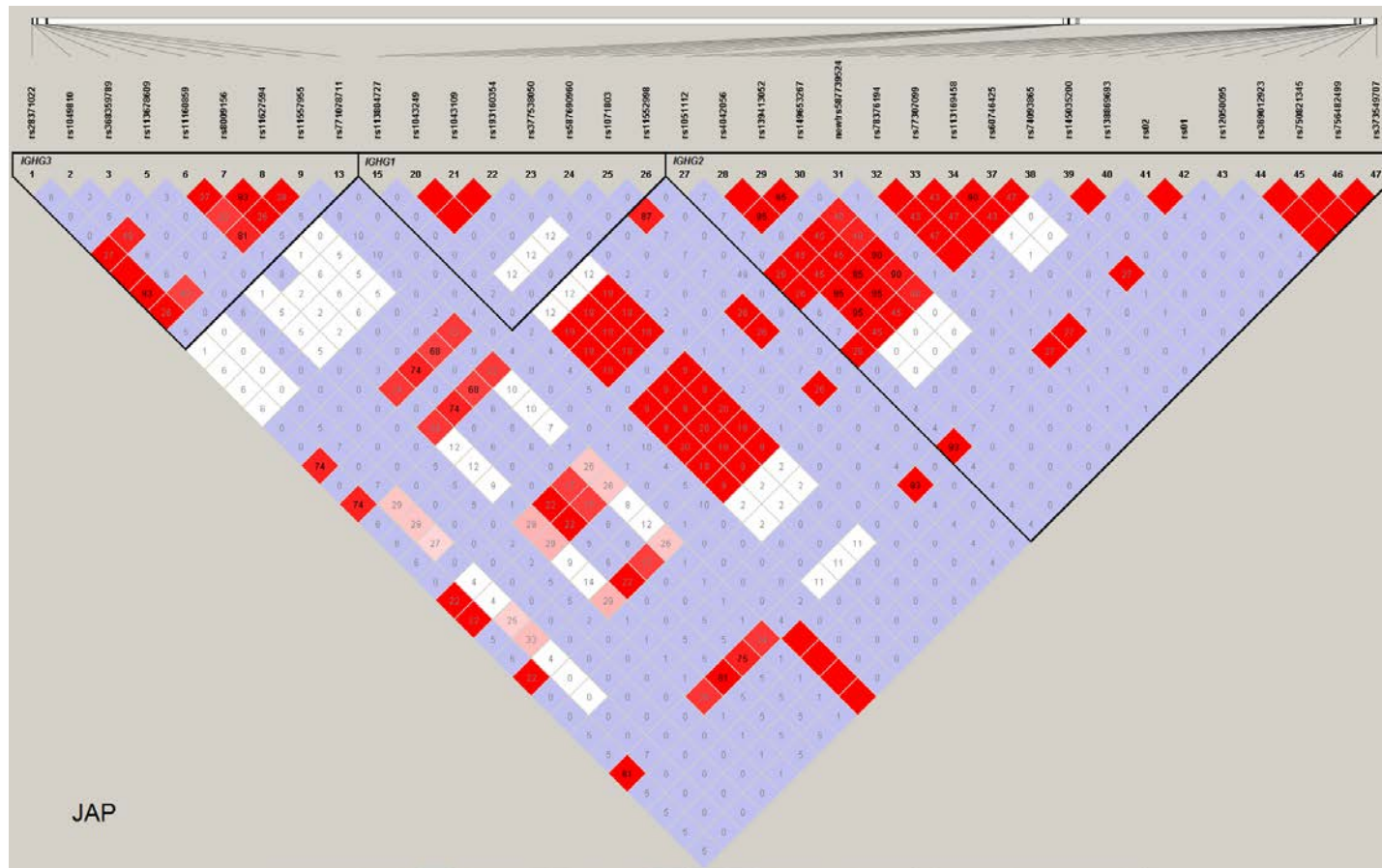


Figura 26: Desequilíbrio de ligação entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* na população de descendentes de Japoneses (JAP). Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza ou branco. Dentro de cada losango estão representados os valores de $r^2 \times 100$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).

Para avaliar a diferenciação genética entre as populações foi estimado o valor de F_{ST} entre pares de populações utilizando as frequências alélicas de cada segmento gênico e, também, combinando as frequências dos três segmentos gênicos. A estatística F_{ST} compara a variância da frequência alélica dentro das populações com a variância entre as populações, e pode tomar valores de 0 a 1. WRIGHT (1978) sugeriu que valores de F_{ST} entre 0 e 0,05, entre 0,05 e 0,15 e superiores a 0,15 indicam baixa, média e alta diferenciação genética, respectivamente.

Os valores de F_{ST} entre as populações para *IGHG1*, *IGHG2* e *IGHG3*, e para os três segmentos gênicos analisados em conjunto, se encontram na parte inferior das Tabela 16 a 19. Assim, as populações mostraram valores que variam entre pequena diferenciação a uma diferenciação elevada entre elas (de 0,015 a 0,316).

Também foi realizado o teste exato de diferenciação populacional que realiza um teste de contingência baseado nas frequências genotípicas das populações. O valor da significância estatística deste teste se encontra na parte superior das Tabela 16 a 19. É possível notar que todas as comparações realizadas entre descendentes de Japoneses e ameríndios resultaram em diferenças significativas, já nos ameríndios algumas comparações resultaram não significativas, especialmente ao comparar KRC com outras populações.

Tabela 16: Diferenciação genética entre as populações para o segmento gênico *IGHG1*

IGHG1	GKW	GND	GRC	KIV	KRC	JAP
GKW		*	***	***	ns	***
GND	-0,0029 ^{ns}		***	***	*	***
GRC	0,1323***	0,1160***		ns	ns	***
KIV	0,1596***	0,1413***	-0,0084 ^{ns}		ns	**
KRC	0,0330*	0,0290 ^{ns}	0,0240*	0,0403*		***
JAP	0,1463***	0,1264***	0,0490***	0,0498***	0,0676***	

Na parte superior da tabela: significância estatística do teste exato de diferenciação populacional. Na parte de inferior da tabela: Valores de F_{ST} entre populações. Abreviações: KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva; JAP: descendentes de Japoneses. * $p < 0,01-0,05$; ** $p < 0,001-0,01$; *** $p < 0,001$; ns $p > 0,05$. Escala de F_{ST} de Wright de diferenciação: entre 0 e 0,05, entre 0,05 e 0,15 e superiores a 0,15 indicam baixa, média e alta diferenciação genética, respectivamente.

Tabela 17 Diferenciação genética entre as populações para o segmento gênico IGHG2

	GKW	GND	GRC	KIV	KRC	JAP
GKW		***	***	*	ns	***
GND	0,0367***		***	*	*	***
GRC	0,1236***	0,0550***		***	***	***
KIV	0,0040 ^{ns}	0,0152*	0,1118***		ns	***
KRC	-0,0028 ^{ns}	0,01543 ^{ns}	0,1058***	-0,00857 ^{ns}		***
JAP	0,2413***	0,1547***	0,0954***	0,21413***	0,21024***	

Na parte superior da tabela: significância estatística do teste exato de diferenciação populacional. Na parte de inferior da tabela: Valores de F_{ST} entre populações. Abreviações: KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva. JAP: descendentes de Japoneses. * $p < 0,01-0,05$; ** $p < 0,001-0,01$; *** $p < 0,001$; ns $p > 0,05$. Escala de F_{ST} de Wright de diferenciação: entre 0 e 0,05, entre 0,05 e 0,15 e superiores a 0,15 indicam baixa, média e alta diferenciação genética, respectivamente.

Tabela 18 Diferenciação genética entre as populações para o segmento gênico IGHG3

	GKW	GND	GRC	KIV	KRC	JAP
GKW		***	***	ns	ns	***
GND	0,0858***		*	**	*	***
GRC	0,1241***	0,0092 ^{ns}		***	**	***
KIV	0,0052 ^{ns}	0,0446***	0,0832***		ns	***
KRC	-0,0008 ^{ns}	0,0491***	0,0816**	-0,0081 ^{ns}		***
JAP	0,3168***	0,1815***	0,1885***	0,2775***	0,2809***	

Na parte superior da tabela: significância estatística do teste exato de diferenciação populacional. Na parte de inferior da tabela: Valores de F_{ST} entre populações. Abreviações: KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva. JAP: descendentes de Japoneses. * $p < 0,01-0,05$; ** $p < 0,001-0,01$; *** $p < 0,001$; ns $p > 0,05$. Escala de F_{ST} de Wright de diferenciação: entre 0 e 0,05, entre 0,05 e 0,15 e superiores a 0,15 indicam baixa, média e alta diferenciação genética, respectivamente.

Tabela 19 Diferenciação genética entre as populações comparando os segmentos gênicos IGHG3, IGHG1 e IGHG2

	GKW	GND	GRC	KIV	KRC	JAP
GKW		**	***	***	ns	***
GND	0,02083*		***	***	**	***
GRC	0,12844***	0,07332***		***	***	***
KIV	0,12800***	0,09882***	0,05228***		*	***
KRC	0,02599**	0,03062**	0,06146***	0,02681*		***
JAP	0,22799***	0,15300***	0,11341***	0,18238***	0,18347***	

Na parte superior da tabela: significância estatística do teste exato de diferenciação populacional. Na parte de inferior da tabela: Valores de F_{ST} entre populações. Abreviações: KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva. JAP: descendentes de Japoneses. * $p < 0,01-0,05$; ** $p < 0,001-0,01$; *** $p < 0,001$; ns $p > 0,05$. Escala de F_{ST} de Wright de diferenciação: entre 0 e 0,05, entre 0,05 e 0,15 e superiores a 0,15 indicam baixa, média e alta diferenciação genética, respectivamente.

As frequências alóticas entre as populações também foram comparadas com o estatístico F_{ST} e teste exato de diferenciação (Tabela 20 a 22). Neste caso foi observado que as populações também diferem em nível de alótipos, com exceção dos alótipos de G2m, onde há diferenças significativas entre ameríndios e descendentes de Japoneses, mas não entre as populações de ameríndios.

Tabela 20: Diferenciação genética entre as populações para alótipos de G1m

	GKW	GND	GRC	KIV	KRC	JAP
GKW		ns	***	***	ns	***
GND	-0.00287ns		***	***	*	***
GRC	0.13888***	0.12250***		ns	ns	***
KIV	0.15961***	0.14130***	-0.00913ns		*	***
KRC	0.03299*	0.02895**	0.02728*	0.04033*		***
JAP	0.18802***	0.16798***	0.03765***	0.03086*	0.08183***	

Na parte de cima da tabela: significância estatística do teste exato de diferenciação populacional. Na parte de baixo da tabela: Valores de F_{ST} entre populações. Abreviações: KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva. JAP: descendentes de Japoneses. * $p < 0,01-0,05$; ** $p < 0,001-0,01$; *** $p < 0,001$; ns $p > 0,05$. Escala de F_{ST} de Wright de diferenciação: entre 0 e 0,05, entre 0,05 e 0,15 e superiores a 0,15 indicam baixa, média e alta diferenciação genética, respectivamente.

Tabela 21: Diferenciação genética entre as populações para alótipos de G2m

	GKW	GND	GRC	KIV	KRC	JAP
GKW		**	ns	ns	ns	***
GND	0.06931*		ns	ns	ns	***
GRC	0.01148ns	0.02414ns		ns	ns	***
KIV	0.02143ns	0.01187ns	-0.00850ns		ns	***
KRC	0.01269ns	0.02223ns	-0.01074ns	-0.00910ns		***
JAP	0.19147***	0.05306**	0.14058***	0.12271***	0.13663***	

Na parte de cima da tabela: significância estatística do teste exato de diferenciação populacional. Na parte de baixo da tabela: Valores de F_{ST} entre populações. Abreviações: KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva. JAP: descendentes de Japoneses. * $p < 0,01-0,05$; ** $p < 0,001-0,01$; *** $p < 0,001$; ns $p > 0,05$. Escala de F_{ST} de Wright de diferenciação: entre 0 e 0,05, entre 0,05 e 0,15 e superiores a 0,15 indicam baixa, média e alta diferenciação genética, respectivamente.

Tabela 22: Diferenciação genética entre as populações para alótipos de G3m

G3m	GKW	GND	GRC	KIV	KRC	JAP
GKW		***	***	ns	ns	***
GND	0.08577***		*	**	*	***
GRC	0.12407***	0.00920ns		***	**	***
KIV	0.00099ns	0.05365**	0.09432***		ns	***
KRC	0.00076ns	0.04909**	0.08156***	-0.00900ns		***
JAP	0.29313***	0.14003***	0.16703***	0.25879***	0.25347***	

Na parte de cima da tabela: significância estatística do teste exato de diferenciação populacional. Na parte de baixo da tabela: Valores de F_{st} entre populações. Abreviações: KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãdeva. JAP: descendentes de Japoneses. * $p < 0,01-0,05$; ** $p < 0,001-0,01$; *** $p < 0,001$; ns $p > 0,05$. Escala de F_{ST} de Wright de diferenciação: entre 0 e 0,05, entre 0,05 e 0,15 e superiores a 0,15 indicam baixa, média e alta diferenciação genética, respectivamente.

A distância genética de Nei entre as populações, considerando as frequências alélicas dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3*, está representada na forma de um dendrograma (Figura 27). A análise de componente principal entre estas populações se encontra na Figura 28. Também foi realizado um dendrograma baseado nas distâncias genéticas de Nei entre as populações estudadas neste trabalho e populações disponíveis no projeto *1000 genomes*. Para essa análise, foram utilizados os dados de SNPs genotipados nesse trabalho e no projeto *1000 genomes* (Figura 29). O resultado foi consistente com a análise de componentes principais em nível populacional e individual realizado com as mesmas populações. A análise de componente principal em nível individual mostra que indivíduos da mesma população se encontram agrupadas formando conjuntos (Figura 30 a 31).

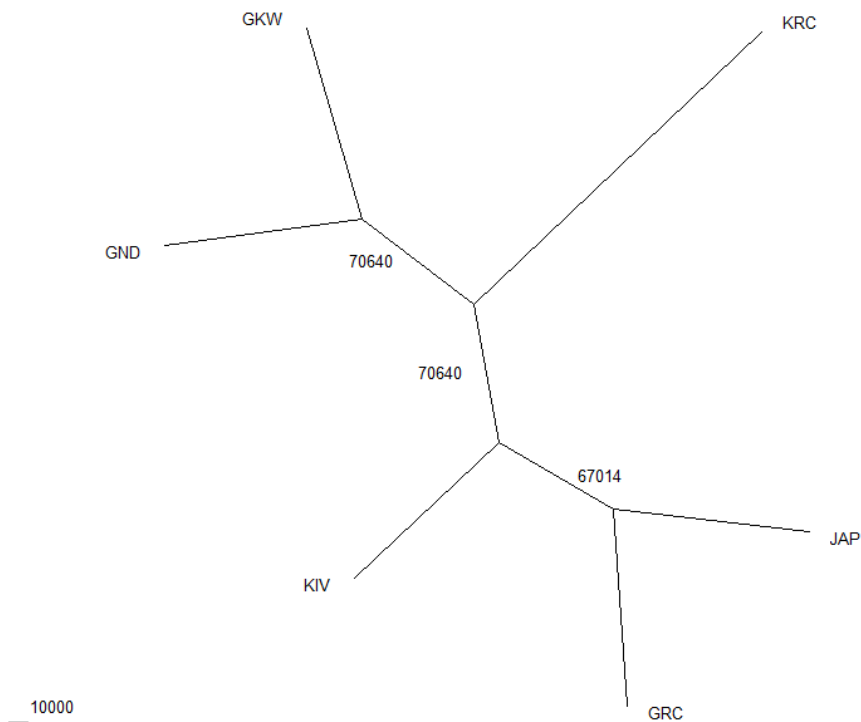


Figura 27: Dendrograma representando a distância genética de Nei para *IGHG1*, *IGHG2* e *IGHG3* entre as populações, realizadas com *bootstrap* de 100.000 reamostragens. KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva. JAP: Descendentes de Japoneses.

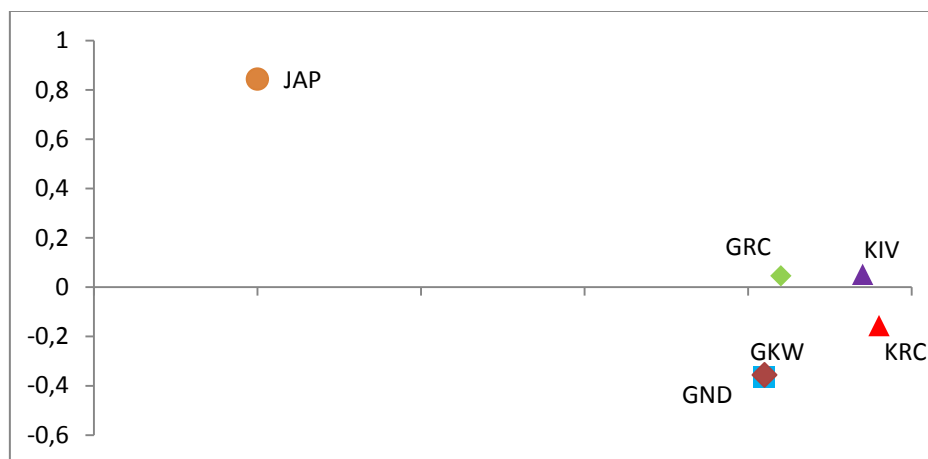


Figura 28: Análise de Componente Principal utilizando frequências alélicas dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* das populações KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva. JAP: Descendentes de Japoneses.

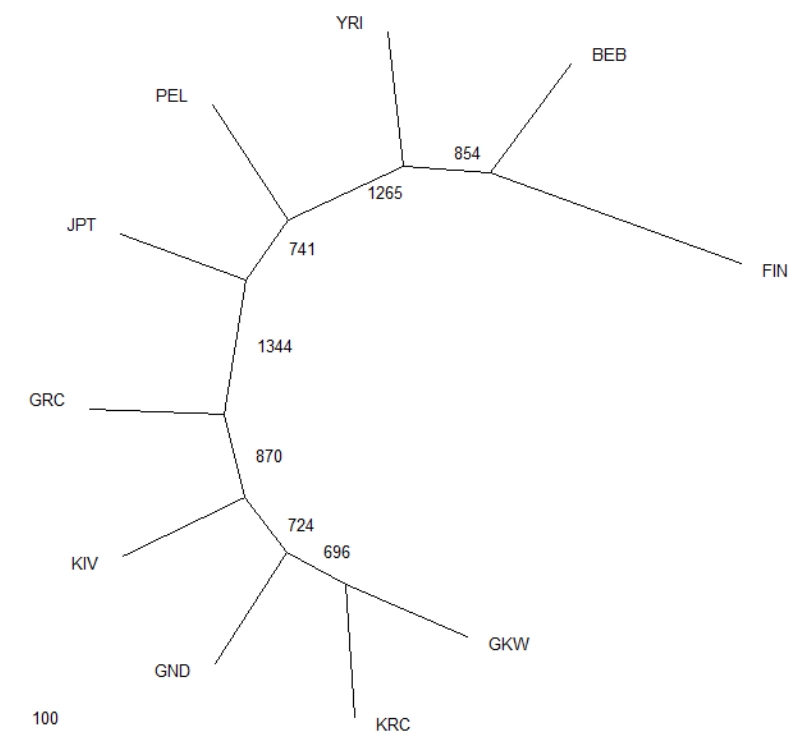


Figura 29: Dendrograma representando a distância genética de Nei realizadas com *bootstrap* de 10.000 reamostragens para *SNPs* genotipados neste estudo em ameríndios e descendentes de Japoneses, e populações disponíveis no banco de dados do projeto *1000 genomes*. KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GWK: Guarani Kaiowá; GND: Guarani Nãndeva; JAP: Descendentes de Japoneses; FIN: Finlandeses em Finlândia; YRI: Iorubás em Nigéria; BEB: Bengalis em Bangladesh; PEL: Peruanos em Peru. Os *SNPs* analisados se encontram na Tabela A1.

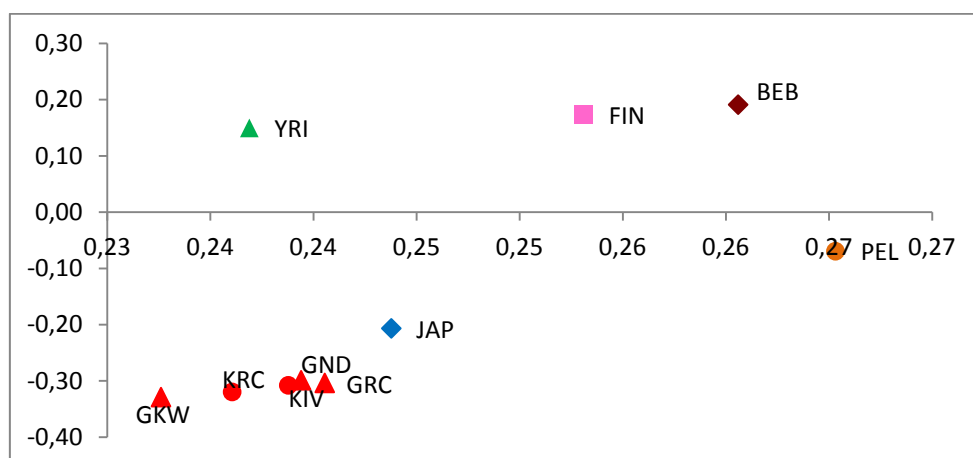


Figura 30: Análise de componente principal das populações utilizando frequências nucleotídicas de *SNPs* nas populações KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GWK: Guarani Kaiowá; GND: Guarani Nãndeva. JAP: Descendentes de Japoneses, e nas populações FIN: Finlandeses em Finlândia; YRI: Iorubás em Nigéria; BEB: Bengalis em Bangladesh; PEL: Peruanos em Peru do projeto *1000 genomes*. Os *SNPs* analisados se encontram na Tabela A1

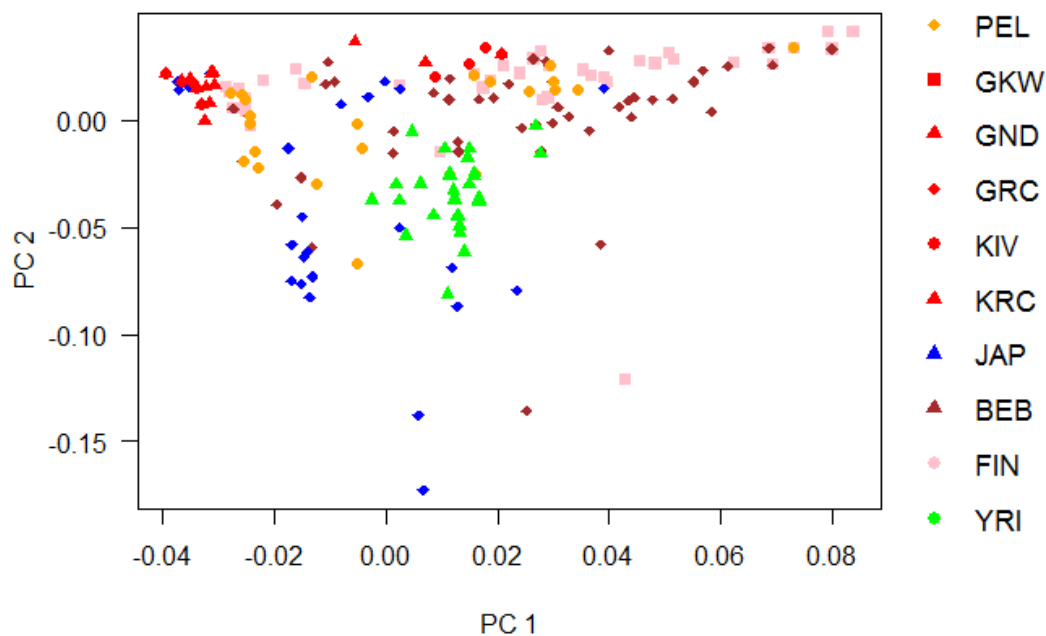


Figura 31: Análise de componente principal gerado a partir do genótipo de SNPs dos indivíduos. Cada ponto representa um indivíduo das populações KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva; JAP: Descendentes de Japoneses; e das populações FIN: Finlandeses em Finlândia; YRI: Iorubás em Nigéria; BEB: Bengalis em Bangladesh; PEL: Peruanos em Peru do projeto *1000 genomes*. Os SNPs analisados se encontram na Tabela A1

A partir do sequenciamento dos três segmentos gênicos, e da inferência de haplótipos realizada anteriormente foi construída uma sequência nucleotídica contendo os três segmentos gênicos para cada cromossomo em todos os indivíduos. A partir disto foram estimados parâmetros de diversidade molecular das populações (Tabela 23). A população de descendentes de Japoneses apresentou o maior número de sítios segregantes ($S = 36$), maior número de haplótipos ($h = 33$), uma maior divergência nucleotídica ($k = 7,411$) e maior diversidade nucleotídica por sítio ($\pi = 0,00261$).

Tabela 23 Parâmetros de diversidade

Parâmetros	GKW	GND	GRC	KIV	KRC	JAP
N de cromossomos	102	102	96	98	92	108
Número de sítios analisados	2838	2838	2838	2838	2838	2838
Sítios segregantes	15	21	22	18	18	36
Número de haplótipos (h)	6	8	9	6	8	33
Diversidade haplotípica (Hd)	0,560	0,668	0,702	0,422	0,422	0,886
Media de divergencia nucleotídica (k)	0,837	3,053	2,068	1,286	1,215	7,411
Diversidade nucleotídica por sítio (π)	0,00029	0,00108	0,00073	0,00045	0,00043	0,00261
Theta-W (por sequência)	2,886	4,041	4,283	3,491	3,534	6,851
Theta-W (por sítio)	0,00102	0,00142	0,00151	0,00123	0,00125	0,00241

Abreviatura das populações KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãdeva; JAP: Descendentes de Japoneses.

Adicionalmente foram realizados testes para avaliar a hipótese de evolução neutra dos segmentos gênicos nas populações. Estes testes foram realizados considerando as sequências dos três segmentos gênicos como um haplótipo, resultando em 2.838 pb analisados; considerando cada segmento gênico separadamente (media de 945 pb); e os éxons *CH1* de *IGHG1*, *IGHG2* e *IGHG3* (882 pb), os éxons *CH2* de *IGHG1*, *IGHG2* e *IGHG3* (984 pb), e os éxons *CH3* de *IGHG1*, *IGHG2* e *IGHG3* (966 pb), para investigar se todos os segmentos gênicos e todos os domínios mostram os mesmos sinais evolutivos.

Na Tabela 24 encontram-se os valores calculados dos parâmetros dentro das populações. A população GKW mostrou valores negativos significativos considerado tanto o haplótipo inteiro, como os segmentos gênicos e os éxons, com exceção de *CH2*. O éxon *CH2* exibiu valores não significativos em todas as demais populações.

Foram realizados os testes *D* e *F* de Fu e Li, utilizando um grupo externo, a espécie *Macaca mulatta* (macaco Rhesus) com números de acesso (NW_001121238, AY292519, AY292512). Este teste apresentou valores negativos e significativos somente em *IGHG2* na população Guarani Kaiowa (GKW) (Tabela 25).

Para todos os segmentos gênicos analisados, o estatístico *H* de Fay e Wu resultou em valores negativos e significativos em todas as populações, somente com exceção de *IGHG2* na população de descende de Japoneses.

O teste dN/dS, que compara a taxa de substituição nucleotídica não sinônima com a sinônima, foi realizado utilizando a sequência dos alelos de cada segmento gênico dentro das populações. Com exceção do *IGHG2* em populações de Guarani e Kaingang, nos demais casos, o teste apontou valores globais não significativos (Tabela 26).

Foram realizados os testes F_s de Fu e R_2 de Ramos-Onsins e Rozas que estimam mudanças no tamanho populacional. Todas as populações mostraram valores não significativos do parâmetro F_s ; e os valores significativos de R_2 , observado nas populações de Kaingang, resultaram em valores próximos a zero (Tabela 24).

Tabela 24 Teste de neutralidade seletiva para os segmentos gênicos

Segmento	Teste	GKW	GND	GRC	KIV	KRC	JAP
	n	51	51	48	49	46	54
IGHG1	Tajima's D	-	-	-	-	-	-
	Fu and Li's D*	-2,7187 $p < 0.05$	-	-	-	-	-
	Fu and Li's F*	-2,5481 $p < 0.05$	-	-	-	-	-
IGHG2	Tajima's D	-	-	-	-	-	-
	Fu and Li's D*	-2,6853 $p = 0,005$	-	-	-	-	-
	Fu and Li's F*	-2,7767 $p = 0,005$	-	-	-	-	-
IGHG3	Tajima's D	-1,9060 $p < 0,001$	-	-	-	-1,8200 $p < 0,001$	-
	Fu and Li's D*	-4,2919 $p < 0,001$	-	-3,6345 $p = 0,0001$	-	-	-
	Fu and Li's F*	-4,1376 $p < 0,001$	-	-3,4825 $p < 0,001$	-	-	-
CH1	Tajima's D	-1,7254 $p = 0,004$	-	-	-	-1,3616 $p = 0,0082$	-
	Fu and Li's D*	-2,6853 $p = 0,02$	-	-	-	-	-3,6191 $p = 0,002$
	Fu and Li's F*	-2,7951 $p = 0,002$	-	-	-	-	-3,1639 $p = 0,001$
	Fs de Fu	-	-	-	-	-3,3505 $p = 0,0173$	-
CH2	Tajima's D	-	-	-	-	-	-
	Fu and Li's D*	-2,0060 $p = 0,001$	-	-	-	-	-
	Fu and Li's F*	-	-	-	-	-	-
CH3	Tajima's D	-	-	-	-	-	-
	Fu and Li's D*	-4,4060 $p < 0,001$	-	-	-	-	-
	Fu and Li's F*	-4,0056 $p < 0,001$	-	-	-	-	-
IGHG3; IGHG1; IGHG2	Tajima's D	-1,96 $p < 0,001$	-	-1,525 $p = 0,008$	-1,805 $p = 0,005$	-1,891 $p = 0,003$	-
	Fu and Li's D*	-4,942 $p < 0,001$	-	-	-	-	-
	Fu and Li's F*	-4,595 $p < 0,001$	-	-	-	-	-
	Fs de Fu	-	-	-	-	-	-
	R2 de Ramos- Onsins e Rozas	-	-	0,036 $p = 0,0246$	0,033 $p = 0,0159$	-	-

Abreviaturas das populações KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva; JAP: Descendentes de Japoneses.

- : resultados não significativos

Tabela 25: Teste de neutralidade seletiva para os segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* utilizando como grupo externo a espécie *Macaca mulatta*

Segmentos gênicos	Teste	GKW	GND	GRC	KIV	KRC	JAP
	n	51	51	48	49	46	54
<i>IGHG1</i>	H de Fay e Wu	-3,8253 <i>p</i> =0,0040	-2,5727 <i>p</i> =0,017	-3,4004 <i>p</i> =0,001	-3,3747 <i>p</i> =0,0011	-3,5174 <i>p</i> =0,001	2,3936 <i>p</i> =0,011
	Fu and Li's D	-	-	-	-	-	-
	Fu and Li's F	-	-	-	-	-	-
<i>IGHG2</i>	H de Fay e Wu	-1,7868 <i>p</i> =0,009	-2,8134 <i>p</i> =0,020	-3,0338 <i>p</i> =0,016	-5,2262 <i>p</i> <0,001	-5,40420 <i>p</i> <0,001	-
	Fu and Li's D	-2,733 <i>p</i> =0,007	-	-	-	-	-
	Fu and Li's F	-2,821 <i>p</i> <0,001	-	-	-	-	-
<i>IGHG3</i>	H de Fay e Wu	-5,8435 <i>p</i> =0,029	-5,9453 <i>p</i> <0,001	-7,0566 <i>p</i> <0,001	-7,3596 <i>p</i> <0,001	-7,5485 <i>p</i> <0,001	-9,2077 <i>p</i> =0,0001
	Fu and Li's D	-	-	-	-	-	-
	Fu and Li's F	-	-	-	-	-	-

Abreviaturas das populações KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva; JAP: Descendentes de Japoneses.
- : resultados não significativos

Tabela 26: Teste de seleção dN/dS dentro das populações

Segmento gênico	Hipótese testada	GKW	GND	GRC	KIV	KRC	JAP
	n	51	51	48	49	46	54
<i>IGHG1</i>	dN=dS	0,705	0,547	0,339	0,705	0,547	0,180
	dN>dS	1,000	1,000	1,000	1,000	1,000	1,000
	dN<dS	0,350	0,283	0,169	0,350	0,283	0,095
<i>IGHG2</i>	dN=dS	0,341	0,090	0,090	0,080	0,084	0,130
	dN>dS	1,000	1,000	1,000	1,000	1,000	1,000
	dN<dS	0,166	0,05	0,050	0,044	0,041	0,068
<i>IGHG3</i>	dN=dS	0,730	0,646	0,596	0,608	0,596	0,820
	dN>dS	0,360	0,310	0,299	0,302	0,299	1,000
	dN<dS	1,000	1,000	1,000	1,000	1,000	0,410

São mostrados os valores da significância estatística global (*p*) dos testes dentro das populações. As hipóteses testadas foram: H. nula (dN=dS: Neutralidade seletiva); H. alternativa (dN>dS: Seleção Positiva); H. alternativa (dN<dS: Seleção purificadora). Em negrito de *p* significativos.

Abreviaturas das populações KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; GRC: Guarani Mbyá; GKW: Guarani Kaiowá; GND: Guarani Nãndeva; JAP: Descendentes de Japoneses.

Em negrito os valores em negrito indicam que o teste resultou estatisticamente significativo.

7 DISCUSSÃO

Este é o primeiro trabalho de caracterização da diversidade genética dos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* da cadeia pesada das imunoglobulinas em populações ameríndias e um dos poucos a estudar populações através de sequenciamento de DNA. Estudos anteriores analisaram diversidade nestas populações indígenas através de métodos sorológicos de genotipagem dos alótipos de Ig (SALZANO; STEINBERG, 1965; LÚCIA et al., 1984; WILLIAMS et al., 1985). O sequenciamento da região exônica destes segmentos gênicos permitiu observar que muitos dos sítios variáveis encontrados neste trabalho não estavam presentes nos alelos descritos no IMGT (International ImMunoGeneTics Information System). Foram encontrados dois sítios variáveis no éxon *CH1* de *IGHG3*, que não possuem identificador *rs*. Apesar da maioria dos SNPs já terem sido descritos anteriormente na base de dados do projeto *1000 genomes*, não se sabe a fase gamética em que eles se encontram, o que caracteriza os alelos dos segmentos *IGHG*. Ainda, pouco se sabe sobre o impacto fenotípico desses polimorfismos. Encontramos 30 SNPs de variação não-sinônima, dos quais 23 foram encontrados no banco de dados do projeto *1000 genomes*. Desses 23, sete foram monomórficos nas populações disponíveis no banco de dados. Os dados do projeto *1000 genomes* foram utilizados em diversas análises deste estudo para comparação da diversidade nucleotídica entre as populações.

Considerando somente as populações de ameríndios, foram encontrados dez alelos novos, sendo dois em *IGHG1*, seis em *IGHG2* e dois em *IGHG3*. Alguns desses novos alelos foram vistos em alta frequência ($fr > 0,03$) e compartilhadas por duas ou mais populações de ameríndios. Como exemplos, os alelos *IGHG2*n07*, **n08*, **n10*, **n11* e *IGHG3*n21*. Na população de descendentes de Japoneses foram encontrados 21 alelos novos, sendo os de maior frequência *IGHG2*n8* ($fr = 0,194$), *IGHG3*n22* ($fr = 0,157$) e *IGHG1*n11* ($fr = 0,065$). Estes resultados podem refletir que a diversidade alélica destes segmentos gênicos é pouco conhecida, e que serão encontrados muitos outros ao se estudar outras populações.

É interessante comparar as frequências alélicas entre Ameríndios e Japoneses, uma população do leste Asiático. O fato de os ancestrais dos Ameríndios serem do continente asiático pode explicar o compartilhamento de alelos. Alguns alelos de alta frequência em descendentes de Japoneses (*IGHG2*n08*, $fr = 0,194$), se encontram em baixa frequência ou ausentes nas populações de Ameríndios ($0 < fr < 0,02$), podendo ter se perdido por efeito fundador e gargalo de garrafa nos processos de colonização das Américas. Por outro lado, alguns alelos de frequência alta em Ameríndios (*IGHG3*14*, $fr > 0,84$), se encontram em menor frequência em descendentes de Japoneses ($fr = 0,49$). A deriva genética, portanto, provavelmente teve efeito evolutivo intenso na história evolutiva dessas populações, impactando nas frequências alélicas que hoje são observadas nessas populações.

Os alelos *IGHG2*n8*, **n10*, *IGHG3*n21* e **n22* foram confirmados por sequenciamento de indivíduos homocigotos, portanto não restam dúvidas em relação à fase das variantes de nucleotídeo único. Os demais alelos de imunoglobulinas foram encontrados em heterocigose, e sua sequência foi inicialmente inferida com base na combinação mais provável levando em conta os alelos mais frequentes dentro de cada população. Posteriormente foram inferidos com a utilização do algoritmo ELB (Excoffier–Laval–Balding) (EXCOFFIER et al., 2005). Os dois métodos apresentaram resultados muito similares, o que garante uma robustez à inferência destes alelos, principalmente para aqueles encontrados em múltiplos indivíduos.

Alguns alelos foram encontrados somente em um indivíduo, e de acordo com a sua sequência nucleotídica, codificariam alótipos que não foram previamente descritos. Este é o caso do alelo *IGHG3*n24*, que codificaria o alótipo G3m5,10,11,13,14,21,26,27, encontrado em um indivíduo na população Guarani Ñandeva. Mesmo tendo determinantes antigênicos afastados um do outro, a combinação dos alótipos G3m5 (posição 115/116 do éxon CH3) e G3m21 (posição 82 do éxon CH2) não é usual. Essa combinação não foi encontrada em trabalhos de caracterização de alótipos de Ig em populações de Ameríndios (SALZANO; STEINBERG, 1965; LÚCIA et al., 1984; WILLIAMS et al., 1985). Esse alelo poderia ser consequência de um erro na genotipagem ou erro na inferência de fase. Por outro lado, a maioria dos estudos foi feita por

técnicas sorológicas, nas quais nunca foi possível distinguir indivíduos homocigotos dos heterocigotos, sendo feitas inferências dos haplótipos que corresponderiam a cada cromossomo. Assim, se algum indivíduo apresentava os alótipos 5 e 21, eram considerados como pertencentes a dois cromossomos diferentes porque se assumia que eles nunca estavam juntos. Portanto, não se pode descartar a possibilidade do alelo *IGHG3*n24* ser verdadeiro. WILLIAMS e colaboradores (1985) relatam que a combinação de alótipos Gm(1,3,17;5,13,14,21) foi encontrada nas população de nativos americanos Pima (1,9%) e Papago (1,4%), e que seriam produto da combinação dos haplótipos Gm1,17;21 / Gm3;5,13,14. Para resolver esse tipo de dúvidas desenhamos oligonucleotídeos iniciadores específicos para sequenciamento de uma fita além de oligonucleotídeos iniciadores para PCR-SSP (reação da cadeia da polimerase com oligonucleotídeos de sequência específica) para confirmação de fase todos os alelos novos. Essa etapa ocorrerá subsequentemente.

Em populações de Kaingang e Guarani de Duque de Caxias e Xapecó (Santa Catarina, Brasil), na qual foram testados os alótipos Gm 1, 2, 5, 6 e 24, foi relatado que o alótipo Gm1 se encontra em 69% da população de Kaingang e em 72% dos Guarani. Gm1,2 em 28% dos Kaingang e em 22% dos Guarani (SALZANO; STEINBERG, 1965). Mesmo que nesse trabalho tenham sido testados poucos alótipos, Gm1 e Gm1,2 podem ser relacionados com os haplótipos completos Gm21,26,27,28;17,1;(..) e Gm21,26,27,28;17,1,2;(..), respectivamente. No presente estudo Gm21,26,27,28;17,1;(..) encontrou-se em ~68% dos Kaingang e em ~48% dos Guarani, e Gm21,26,27,28;17,1,2;(..) em ~27% dos Kaingang e em ~40% dos Guarani (Tabela 15). As frequências alotípicas foi semelhante entre as populações de Kaingang de ambos estudos, mas não as frequências das populações de Guarani deste trabalho comparadas com às relatadas na literatura. Porém, elas mostram que esses alótipos são os mais comuns dentro das populações Guarani e Kaingang.

As frequências haplotípicas encontradas na população de descendentes de Japoneses são muito similares às encontradas em um estudo anterior envolvendo famílias japonesas (LOGHEM et al., 1970). Naquele trabalho, o haplótipo Gm21;17,1;(..) foi encontrado com frequência de 40,7%, Gm11,13,15,16;17,1;(..) de 27,7%, Gm21;17,1,2;(..) de 16,4%, e

Gm5,11,13;3,1;23 de 15,2%. Na população de descendente de Japoneses deste estudo, esses haplótipos foram encontrados em frequências similares, de 43,5%, 26,8%, 12% e 12,9%, respectivamente.

A partir desta comparação, é possível observar que há semelhança entre as frequências dos haplótipos relatada na literatura e os haplótipos inferidos neste trabalho pelo método ELB. Essa comparação mostra que a inferência da fase dos haplótipos realizada nesse trabalho é confiável. Mais ainda, é uma maneira indireta de validar a genotipagem. Isto é importante porque a partir desta inferência de fase foram feitos os testes de neutralidade que utilizaram os haplótipos *IGHG3;IGHG1;IGHG2*.

Os haplótipos mais comuns em ameríndios foram Gm21,26,27,28;17,1;(..) (38-76%) e Gm21,26,27,28;17,1,2;(..) (20%-51%) (Tabela 15), coincidentes com as frequências relatadas para outras populações de nativos da América do Sul de 55% a 89% e 11% a 34%, respectivamente, (Revisado em JOHNSON et al., 1977; SCHANFIELD, 1992).

Na população GKW foi encontrado um indivíduo com o haplótipo Gm5,10,11,13,14,26,27;3;(..) (1%) e em indivíduos de GND, GRC, KIV e KRC foi encontrado o haplótipo Gm5,10,11,13,14,26,27;3;23 (1-6%) (Tabela 15). Estes são haplótipos encontrados em alta frequência em indivíduos de origem europeia (~60%), mas com frequência inferior a 1% em populações de nativos americanos (Revisado em JOHNSON et al., 1977; LÚCIA et al., 1984; SCHANFIELD, 1992; LEFRANC; LEFRANC, 2012). Portanto, esses haplótipos podem ter sido introduzidos nas populações de Ameríndios por miscigenação com não nativos após a colonização europeia. Os genes *HLA-A*, *-B*, *-C*, *DR* e *-DQ* foram previamente caracterizados em indivíduos destas populações (PETZL-ERLER et al., 1993). O estudo mostra que os indivíduos que apresentaram os alótipos de Ig europeus, também possuem alelos de *HLA* comuns em eurodescendentes e incomuns em ameríndios: *HLA-B*14*, *-B*18*, *-B*45*, *-DQB*05*, *-DQA*01*. Isto ratifica a provável origem europeia desses haplótipos de Ig mencionados anteriormente.

As análises filogenéticas foram realizadas apenas com alelos já descritos e alelos novos com alta frequência. As árvores mostram vários ramos polifiléticos, com valores de *bootstrap* baixos, nos quais não é possível definir a distância e relação entre alelos. Isto pode acontecer devido aos poucos sítios

variáveis entre sequências nucleotídicas nesses alelos, e portanto, pouca informação para a realização dos cálculos de distância genética eles. Deve destacar-se que, na maioria dos casos, alelos que se encontram agrupados codificam para os mesmos alótipos, como é o caso de alelos de *IGHG3* que codificam o alótipo G3m5* (vermelho). Inclusive um alelo novo *IGHG3*n22*, que também codifica este alótipo, localizou-se no mesmo agrupamento (Figura 17). Os alelos *IGHG3*14*, *15 e *16 que codificam o alótipo G3m21*(21, 26,27,28) (verde), localizaram-se no mesmo ramo que o alelo novo *IGHG*n21*, que possui o alótipo G3m21,27 (verde escuro) e que, portanto, perdeu os determinantes antigênicos dos alótipos 26 e 28 (Figura 17). Também é possível observar a proximidade entre o alelo *IGHG3*n21* com *14, *15 e *16 na rede de haplótipo realizada pelo método Median-Joining (MJ). Ela mostra que a partir do alelo *14, é necessária apenas uma mudança nucleotídica para a formação dos demais alelos.

As redes de haplótipo de SNPs mostram como alguns alelos se encontram em alta frequência em todos os grupos (ex: *IGHG1*02* e *07), sendo que outros são exclusivos de uma população (ex: *IGHG1*n11* exclusivo de descendentes de Japoneses ou *IGHG1*03* que somente está presente em populações de Ameríndios). As redes de haplótipos de SNPs de MJ não possuem raiz. Considerando somente as populações analisadas neste trabalho, e a partir da topologia das redes, poderíamos sugerir que os alelos ancestrais seriam *IGHG1*02*, *IGHG2*03* e *IGHG3*14*, já que eles encontram-se em alta frequência em todas as populações e possuem o maior número de conexões com os demais alelos. Portanto, a partir dele teriam se originado os demais alelos, seguindo os passos mutacionais indicados pelos ramos das redes.

No caso do segmento gênico *IGHG2*, nem sempre é possível resolver qual é o caminho mais provável que levou a geração dos alelos. Podemos sugerir que a configuração destes ramos se deve a eventos de recombinação entre os alelos envolvidos no ciclo. Outra alternativa seria que a mesma mudança nucleotídica tenha acontecido duas vezes na evolução destes alelos, por exemplo a mudança na posição 91 do éxon CH1 (91CH1), indicada entre os alelos *IGHG2*02* e *n07, e entre *05 e *n08. Porém, é menos provável que mutações pontuais aconteçam no mesmo sítio em eventos independentes. Os

nós indicados com “mv” (*mediam vectors*) foram gerados pelo algoritmo MJ para representar alelos que não fazem parte da amostra, mas que são necessários para conectar os alelos das populações. Ele representa alelos ancestrais extintos, ou alelos em baixa frequência (não identificados) na população em estudo.

O padrão geral de desequilíbrio de ligação entre SNPs se mostrou diferente entre as populações. Porém, entre KRC e KIV o desequilíbrio de ligação entre SNPs é similar, o que é esperado pela proximidade genética entre as populações. Na população GKW existe um forte desequilíbrio entre os SNPs de *IGHG1* e *IGHG2*, já em GND o desequilíbrio de ligação é forte entre os três segmentos gênicos estudados, e em GRC, entre *IGHG3* e *IGHG1*. As diferenças encontradas entre as populações de Guarani podem ser devidas a fatores evolutivos como deriva genética, seleção natural e também miscigenação com outras populações.

O gráfico de desequilíbrio de ligação realizado apenas com os SNPs que são polimórficos em todas as populações de Ameríndios (Figura A4 a A9), mostra forte desequilíbrio de ligação entre eles. Porém, o da população GRC apresentou-se diferente às demais populações de Ameríndios, e é possível observar que nesta população não há desequilíbrio de ligação entre SNPs de *IGHG2* com os demais segmentos gênicos.

Observamos forte desequilíbrio de ligação entre alelos de SNPs de diferentes segmentos gênicos, apesar da sua distância no genoma. Os segmentos gênicos *IGHG3* e *IGHG1* se encontram a 96 kb, *IGHG1* e *IGHG2* a 26 kb, e *IGHG3* e *IGHG2* a 124 kb de distância. Isto está de acordo com relatos da literatura que indicam baixa recombinação entre estes segmentos gênicos (Revisado em ZELASCHI et al., 1983).

De forma geral, a estruturação populacional estimada pelo F_{ST} mostra níveis altos de diferenciação, especialmente entre as populações de ameríndios e a de descendentes de Japoneses, $F_{ST} > 15$. Em Ameríndios os níveis de estruturação entre populações são de baixo a moderado (de 1,5% a 15%, ou não significativas), tanto nas comparações de frequências alélicas quando alotípicas.

Entre KRC e KIV, o $F_{ST} < 4\%$ mostra níveis baixos ou não significativos de diferenciação, o que é esperado pela proximidade genética conhecida entre

elas. Este também é o caso de GKW e GND, entretanto em alguns casos elas mostram F_{ST} de 8,5% indicando moderada diferenciação. Estas diferenças genéticas moderadas (>5%) entre populações que tiveram origem genética comum podem ser produto de eventos estocásticos, como a deriva genética, assim como seleção natural, ou ainda fluxo gênico com outras populações, uma vez que fluxo gênico entre as populações do nosso estudo diminuiriam os valores do estatístico F_{ST} .

Os dendrogramas construídos a partir das frequências alélicas dos três segmentos gênicos mostra a distância genética entre as populações, a qual apresenta uma topologia semelhante à distribuição das populações no plano cartesiano dos componentes principais da variabilidade genética das populações. É possível verificar que as populações de Ameríndios são próximas entre si, e a de descendentes de Japoneses se encontra mais distante delas, principalmente na análise de componente principal. No dendrograma, a população KIV mostra a mesma distância com GRC e com JAP, porém, a população de Japoneses se encontra afastada de todos os demais Ameríndios. O dendrograma baseado em distâncias genéticas utilizando frequências de SNPs e a análise de componente principal utilizando frequências de SNPs apresentaram resultados semelhantes. Nessas análises, além das populações de nosso estudo, incluímos populações do banco de dados do projeto *1000 genomes*. Os nativos americanos se encontram agrupados e próximos às populações de descendentes de Japoneses (JAP) e peruanos (PEL) e mais distantes das populações de europeus (FIN), africanos (YRI) e indo-arianos (BEB).

Tanto a análise de componente principal, como os dendrogramas de distância de Nei não indicam origem comum. Cada clado não indica ancestral comum e não se trata de uma árvore filogenética. O resultado dessas análises é apenas um retrato das diferenças das frequências alélicas desses segmentos gênicos nessas populações. Os agrupamentos e clados formados podem ser explicados não apenas por origem comum, mas também por efeitos demográficos e fatores evolutivos. Para estudar as relações genéticas entre populações é necessário utilizar informações de diversidade genética de um maior número de genes e realizar análises mais robustas, o que não é objetivo do presente estudo.

Foram realizados testes de neutralidade seletiva dentro das populações de estudo em diferentes abordagens: utilizando os segmentos gênicos inteiros (sem os íntrons); utilizando cada éxon separadamente (*CH1*, *CH2* e *CH3* de cada segmento gênico); e também utilizando haplótipos inferidos de *IGHG3;IGHG1;IGHG2*. Quando foi analisado o haplótipo *IGHG3;IGHG1;IGHG2* dentro de cada população, o *D* de Tajima mostrou valores negativos e significativos tanto em Kaingang, como em Guarani ($D = -1,795$; $p < 0,01$). Este resultado sugere um excesso de alelos de frequência baixa como consequência de seleção purificadora, no *locus* em estudo ou de uma região próxima no genoma. Alternativamente, pode indicar expansão populacional recente (TAJIMA, 1989). Ao realizar este teste utilizando a sequência nucleotídica dos éxons e de cada segmento gênico, o parâmetro *D* resultou não significativo em todas as populações, exceto para *IGHG3* em GKW ($-1,906$; $p < 0,001$) e KRC ($-1,820$; $p < 0,001$), e para *CH1* em GKW ($-1,72543$; $p=0,004$) e KRC ($-1,3616$; $p=0,0082$). O que pode ser consequência da diminuição de sítios disponíveis para análise, com perda do poder estatístico do *D* de Tajima.

A espécie *Macaca mulatta* foi escolhida como grupo externo para os testes *D* e *F* de Fu e Li, por ser o único primata com a sequência completa destes segmentos gênico disponível na base de dados do IMGT, além de ser a espécie mais próxima aos humanos depois dos grandes primatas (GIBBS et al., 2007). Como este teste leva em consideração a genealogia dos alelos de um gene, a adição de um grupo externo permite resolver quais mutações formam parte dos ramos internos ou externos das árvores. Quando não foi utilizado um grupo externo, os estatísticos *D** e *F** apresentaram valores negativos e significativos nos éxons *CH1* e *CH3*, e nos segmentos gênicos *IGHG1*, *IGHG2* e *IGHG3* da população GKW, em *IGHG3* em GRC e nos éxons *CH1* em JAP. Quando foi adicionado o grupo externo, a significância estatística se manteve apenas em *IGHG2*. Valores negativos de *D* e *F* indicam um excesso de mutações nos ramos externos, em comparação aos esperados sob evolução neutra.

FAY e WU (2000) sugerem que valores negativos do parâmetro *H* indicam excesso de mutações em alta frequência, produto de efeito carona com outra variação vantajosa, que está sendo favorecida pela seleção natural.

Isto pode ser verdade se levamos em consideração que estes segmentos gênicos se encontram ligados aos segmentos que determinam a região variável das imunoglobulinas. E como elas se encontram diretamente envolvidas na detecção e defesa contra patógenos, é de se esperar que existam fatores evolutivos atuando sobre essa região, afetando também a diversidade dos segmentos gênicos ligados a eles. Resultados similares foram observados, com valores negativos e significativos do parâmetro H de Fay e Wu nas populações Guarani e Kaingang no gene *MBL2*, junto com valores não significativos de D de Tajima, e D e F de Fu e Li. Naquele estudo, os resultados foram interpretados como produto de miscigenação com Neo-Brasileiros e não como resultado de efeito carona (BOLDT et al., 2006). Entretanto, o resultado do teste de Fay e Wu no presente estudo se manteve semelhante quando foi analisado na ausência dos haplótipos de Ig europeus (indivíduos possivelmente miscigenados). Adicionalmente, TANAKA e NEI (1989) encontraram uma taxa aumentada de mutações não sinônimas em relação a mutações sinônimas nos segmentos gênicos variáveis (*IGHV*), e sugerem que a seleção diversificadora ou sobredominante estaria atuando nestes segmentos.

Todos os testes de neutralidade considerados até o momento analisam a frequência de polimorfismos sem considerar se as substituições são sinônimas ou não sinônimas. Já o teste dN/dS compara as taxas de variações sinônimas e não sinônimas nas sequências. Este teste realizado dentro das populações para cada segmento gênico mostrou valores não significativos para a maioria das comparações (Tabela 26). Esse resultado indica que não há evidência de que substituições nucleotídicas que modifiquem a sequência polipeptídica estejam sendo favorecidas ou eliminadas pela seleção natural. Logo as substituições estariam sendo mantidas nas populações de forma aleatória.

Somente no caso de *IGHG2*, a hipótese alternativa de $dN < dS$ (seleção purificadora), apontou valores significativos (Tabela 26). No caso das populações de Guarani (com valores de $p = 0,05$), ao examinar a comparação entre pares de sequências, nenhuma manteve a significância. No caso das populações de Kaingang, foram as sequências dos alelos *IGHG2*03* e **n08* as que mostraram valores significativos. A diferença de sequência nucleotídica

entre elas acontece em três sítios, com substituições sinônimas, ou seja, em que as sequências de aminoácidos geradas pelos dois alelos é idêntica.

Finalmente, RAMOS-ONSINS e ROZAS (2002) sugerem que os testes F_s de Fu e R_2 são os testes com maior poder para detectar expansão populacional e são utilizados de acordo ao tamanho amostral: R_2 para populações pequenas ($n \approx 10$) e F_s para populações grandes ($n \approx 50$). Assim o estatístico que melhor se ajustaria aos nossos dados é o F_s . Entretanto realizamos os dois testes a fim de comparar os resultados. Tanto F_s quanto R_2 resultaram não significativos, exceto em Kaingang. Nesta população, R_2 foi 0,03, próximo a zero; e valores próximos a zero não rejeitam a hipótese nula de tamanho populacional constante.

Deste modo, como o estatístico H de Fay e Wu apresentou valores negativos e significativos em quase todos os testes realizados, sugere-se que há evidencia de seleção natural positiva nas regiões flangeadoras a *IGHG*, possivelmente nos segmentos gênicos que codificam as regiões variáveis da molécula de imunoglobulina as que atuam diretamente no reconhecimento de antígenos. Também é possível que os segmentos gênicos *IGHG* estejam sob a ação da seleção natural negativa, mas que este processo não seja forte o suficiente para ser capturados por testes mais robustos, como os realizados com a adição de um grupo externo às análises.

8 CONCLUSÕES

- Diversos sítios variáveis encontrados não estão presentes nos alelos descritos no banco de dados do IMGT.
- Foram encontrados 10 alelos novos nas populações de Ameríndios e 21 alelos novos na população de descendentes de Japoneses.
- Foi confirmada a sequência nucleotídica dos alelos novos: *IGHG2*n8*, **n10*, *IGHG3*n21* e **n22*. Os demais deverão ser ainda confirmado por PCR-SSP ou sequenciamento de fita única.
- As frequências alélicas de Ameríndios e descendentes de Japoneses mostram sinais de efeito fundador e gargalo de garrafa como consequência do processo de ocupação das Américas.
- Os alelos mais comuns nas populações, *IGHG1*02*, *IGHG2*03* e *IGHG3*14*, são provavelmente os alelos ancestrais que deram origem aos demais alelos.
- Os alótipos mais comuns em ameríndios foram Gm21,26,27,28;17,1;(..) e Gm21,26,27,28;17,1,2;(..).
- Os alótipos mais comuns nos descendentes de Japoneses foram Gm21;17,1;(..), Gm11,13,15,16;17,1;(..), Gm21;17,1,2;(..) e Gm5,11,13;3,1;23.
- A presença dos alótipos Gm5,10,11,13,14,26,27;3;(..) e Gm5,10,11,13,14,26,27;3;23 em indígenas são possivelmente devido à miscigenação com eurodescendentes.
- Entre populações ameríndias foram verificados níveis baixos a moderados de diferenciação genética, enquanto que comparações entre ameríndios e descendentes de Japoneses mostraram níveis altos de diferenciação.
- Há evidencia de que a diversidade observada em *IGHG* seja produto do efeito carona por seleção natural positiva nas regiões flaqueadoras do segmento gênico *IGHG*.

BIBLIOGRAFIA

ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. **Imunologia Celular e Molecular**. Elsevier, 2012.

ARNOLD, J. N.; WORMALD, M. R.; SIM, R. B.; RUDD, P. M.; DWEK, R. A. The impact of glycosylation on the biological function and structure of human immunoglobulins. **Annual review of immunology**, v. 25, p. 21–50, 2007.

ATHERTON, A.; ARMOUR, K. L.; BELL, S.; MINSON, A. C.; CLARK, M. R. The herpes simplex virus type 1 Fc receptor discriminates between IgG1 allotypes. **European Journal of Immunology**, v. 30, n. 9, p. 2540–2547, 2000. Disponível em: <[http://doi.wiley.com/10.1002/1521-4141\(200009\)30:9<2540::AID-IMMU2540>3.0.CO;2-S](http://doi.wiley.com/10.1002/1521-4141(200009)30:9<2540::AID-IMMU2540>3.0.CO;2-S)>.

AUTON, A.; ABECASIS, G. R.; ALTSHULER, D. M.; et al. A global reference for human genetic variation. **Nature**, v. 526, n. 7571, p. 68–74, 2015. Disponível em: <http://www.nature.com/nature/journal/v526/n7571/fig_tab/nature15393_SF1.html\n<http://dx.doi.org/10.1038/nature15393>\n<http://www.ncbi.nlm.nih.gov/pubmed/26432245>>.

AZEVEDO, M.; BRAND, A.; HECK, E.; MARQUES PEREIRA, L.; MELIÀ, B. Guarani Retã 2008. Povos Guarani na Fronteira Argentina, Brasil e Paraguai. , p. 28, 2008.

BANDELT, H. J.; FORSTER, P.; ROHL, A. Median-joining networks for inferring intraspecific phylogenies. **Molecular Biology and Evolution**, v. 16, n. 1, p. 37–48, 1999. Disponível em: <<http://mbe.oxfordjournals.org/cgi/doi/10.1093/oxfordjournals.molbev.a026036>>

BARRETT, J. C.; FRY, B.; MALLER, J.; DALY, M. J. Haploview: Analysis and visualization of LD and haplotype maps. **Bioinformatics**, v. 21, n. 2, p. 263–265, 2005.

BEALE, D.; FEINSTEIN, A. Structure and function of the constant regions of immunoglobulins. **Quarterly Reviews of Biophysics**, v. 9, n. 2, p. 135, 1976. Disponível em: <http://www.journals.cambridge.org/abstract_S0033583500002390>.

BOLDT, A. B. W.; CULPI, L.; TSUNETO, L. T.; et al. Diversity of the MBL2 Gene in Various Brazilian Populations and the Case of Selection at the Mannose-Binding Lectin Locus. **Human Immunology**, v. 67, n. 9, p. 722–734, 2006. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17002903>>. Acesso em: 24/9/2014.

BUCK, D.; ALBRECHT, E.; ASLAM, M.; et al. Genetic variants in the immunoglobulin heavy chain locus are associated with the IgG index in multiple sclerosis. **Annals of Neurology**, v. 73, n. 1, p. 86–94, 2013.

CROCE, C. M.; SHANDER, M.; MARTINIS, J.; et al. Chromosomal location of the genes for human immunoglobulin heavy chains. **Proceedings of the**

National Academy of Sciences of the United States of America, v. 76, n. 7, p. 3416–9, 1979. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=383836&tool=pmcentrez&rendertype=abstract>>.

DARD, P.; HUCK, S.; FRIPPIAT, J. P.; et al. The IGHG3 gene shows a structural polymorphism characterized by different hinge lengths: Sequence of a new 2-exon hinge gene. **Human Genetics**, v. 99, n. 1, p. 138–141, 1997.

DUDLEY, D. D.; CHAUDHURI, J.; BASSING, C. H.; ALT, F. W. Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. **Advances in immunology**, v. 86, n. D, p. 43–112, 2005. Disponível em: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&doctype=Citation&list_uids=15705419>.

EXCOFFIER, L.; LAVAL, G.; BALDING, D. Gametic phase estimation over large genomic regions using an adaptive window approach. **Human Genomics**, v. 1, n. 1, p. 7, 2003. Disponível em: <<http://humgenomics.biomedcentral.com/articles/10.1186/1479-7364-1-1-7>>.

EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. **Evol Bioinform Online**, v. 1, p. 47–50, 2005. Disponível em: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&doctype=Citation&list_uids=19325852>.

EXCOFFIER, L.; LISCHER, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. **Molecular Ecology Resources**, v. 10, n. 3, p. 564–567, 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21565059>>. Acesso em: 22/5/2013.

FAY, J. C.; WU, C. I. Hitchhiking under positive Darwinian selection. **Genetics**, v. 155, n. 3, p. 1405–1413, 2000.

FEERO, W. G.; GUTTMACHER, A. E.; MANOLIO, T. A. Genomewide Association Studies and Assessment of the Risk of Disease. **New England Journal of Medicine**, v. 363, n. 2, p. 166–176, 2010. Disponível em: <<http://www.nejm.org/doi/full/10.1056/NEJMra0905980>\n<http://www.nejm.org/doi/abs/10.1056/NEJMra0905980>\n<http://www.nejm.org.ezproxy.umassmed.edu/doi/full/10.1056/NEJMra0905980>>.

FELSENSTEIN, J. PHYLIP (Phylogeny Inference Package) version 3.6. **Department of Genome Sciences, University of Washington, Seattle**, 2004.

FU, Y. X.; LI, W. H. Statistical tests of neutrality of mutations. **Genetics**, v. 133, n. 3, p. 693–709, 1993. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1208208&tool=pmcentrez&rendertype=abstract>>.

GIBBS, R. A.; ROGERS, J.; KATZE, M. G.; et al. Evolutionary and Biomedical Insights from the Rhesus Macaque Genome. **Science**, v. 316, n. 5822, p. 222–

234, 2007. Disponível em:
<<http://science.sciencemag.org/content/316/5822/222.abstract>>.

GRANOFF, D. M.; BOIES, E.; SQUIRES, J.; et al. Interactive effect of genes associated with immunoglobulin allotypes and HLA specificities on susceptibility to Haemophilus influenzae disease. **Journal of immunogenetics**, v. 11, n. 3–4, p. 181–8, 1984. Disponível em:
<<http://www.ncbi.nlm.nih.gov/pubmed/6520406>>.

GRANOFF, D. M.; HOLMES, S. J. G2m(23) immunoglobulin allotype and immunity to Haemophilus influenzae type b. **The Journal of infectious diseases**, v. 165 Suppl, n. 23, p. S66-9, 1992. Disponível em:
<<http://www.ncbi.nlm.nih.gov/pubmed/1588180>>.

GRANTHAM, R. Amino acid difference formula to help explain protein evolution. **Science (New York, N.Y.)**, v. 185, n. 4154, p. 862–864, 1974.

GUO, S.; THOMPSON, E. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. **Biometrics**, v. 48, p. 361–72, 1992.

IBGE. Censo Demográfico 2010. Disponível em:
<<http://www.censo2010.ibge.gov.br>>.

INSTITUTO-SOCIOAMBIENTAL. Povos indígenas do Brasil. Disponível em:
<<http://pib.socioambiental.org/pt>>. Acesso em: 6/4/2016.

INTERNATIONAL HAPMAP CONSORTIUM. A haplotype map of the human genome. **Nature**, v. 437, n. 7063, p. 1299–320, 2005. Disponível em:
<<http://www.ncbi.nlm.nih.gov/pubmed/16255080>>.

JEFFERIS, R.; LEFRANC, M.-P. Human immunoglobulin allotypes: possible implications for immunogenicity. **mAbs**, v. 1, n. 4, p. 332–8, 2009. Disponível em:
<<http://www.ncbi.nlm.nih.gov/pubmed/1223557>>.

JOHNSON, W. E.; KOHN, P. H.; STEINBERG, A. G. Population genetics of the human allotypes Gm, Inv, and A2m: An analytical review. **Clinical Immunology and Immunopathology**, v. 7, n. 1, p. 97–113, 1977. Disponível em:
<<http://www.sciencedirect.com/science/article/pii/0090122977900344\papers3://publication/uuid/FCE0156D-A779-469E-8DB2-8E0370476D15>>.

JUNG, D.; GIALLOURAKIS, C.; MOSTOSLAVSKY, R.; ALT, F. W. Mechanism and Control of V(D)J Recombination at the Immunoglobulin Heavy Chain Locus. **Annual Review of Immunology**, v. 24, n. 1, p. 541–570, 2006. Disponível em:
<<http://www.annualreviews.org/doi/abs/10.1146/annurev.immunol.23.021704.115830>>.

KARSTEN, C. M.; KÖHL, J. The immunoglobulin, IgG Fc receptor and complement triangle in autoimmune diseases. **Immunobiology**, v. 217, n. 11, p. 1067–79, 2012. Disponível em:
<<http://www.ncbi.nlm.nih.gov/pubmed/22964232>>.

KOIDE, N.; NOSE, M.; MURAMATSU, T. Recognition of IgG by Fc receptor and

complement: effects of glycosidase digestion. **Biochemical and biophysical research communications**, v. 75, n. 4, p. 838–44, 1977. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/324473>>.

KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. **Molecular Biology and Evolution**, v. 33, n. 7, p. 1870–1874, 2016. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/27004904>>.

LAHIRI, D. K.; NURNBERGER, J. I. J. A rapid non-enzymatic method for the preparation of DNA from blood for RFLP studies. **Nucleic Acids Research**, v. 9, n. 19, p. 5444, 1991.

LEFRANC, M. P.; LEFRANC, G. Human Gm, Km, and Am allotypes and their molecular characterization: A remarkable demonstration of polymorphism. **Methods in Molecular Biology**, v. 882, p. 635–680, 2012. Disponível em: <<http://link.springer.com/10.1007/978-1-61779-842-9>>.

LEFRANC, M.-P.; GIUDICELLI, V.; DUROUX, P.; et al. IMGT®, the international ImMunoGeneTics information system® 25 years on. **Nucleic acids research**, v. 43, n. Database issue, p. D413-22, 2015. Disponível em: <<http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku1056>>.

LEFRANC, M.-P.; LEFRANC, G. **Immunoglobulin FactsBook**. Academic Press, 2001.

LIBRADO, P.; ROZAS, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. **Bioinformatics**, v. 25, n. 11, p. 1451–1452, 2009.

LOGHEM, E. VAN; NATVIG, J. B.; MATSUMOTO, H. Genetic markers of immunoglobulins in Japanese families Inheritance of associated markers belonging to one IgA and three IgG subclasses. **Annals of Human Genetics**, v. 33, n. 4, p. 351–359, 1970. Disponível em: <<http://doi.wiley.com/10.1111/j.1469-1809.1970.tb01661.x>>.

LÚCIA, M.; HAMEL, H.; SALZANO, F. M.; FREITAS, M. J. DE M. E. The Gm polymorphism and racial admixture in six Amazonian populations. **Journal of Human Evolution**, v. 13, n. 6, p. 517–529, 1984. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0047248484800056>>.

MCBRIDE, O. W.; BATTEY, J.; HOLLIS, G. F.; et al. Localization of human variable and constant region immunoglobulin heavy chain genes on subtelomeric band q32 of chromosome 14. **Nucleic acids research**, v. 10, n. 24, p. 8155–70, 1982. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/6819544>>.

MILLER-KITTRELL, M.; SPARER, T. E. Feeling manipulated: cytomegalovirus immune manipulation. **Virology journal**, v. 6, n. 1, p. 4, 2009. Disponível em: <<http://virologyj.biomedcentral.com/articles/10.1186/1743-422X-6-4>>.

Minitab 17 Statistical Software. ., 2010. Computer software: State College, PA. Disponível em: <www.minitab.com>.

MORAHAN, G.; BEREK, C.; MILLER, J. F. An idiotypic determinant formed by both immunoglobulin constant and variable regions. **Nature**, v. 301, n. 5902, p. 720–2, 1983. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/6186921>>.

MURPHY, K. **Janeway's Immunobiology**. Artmed, 2011.

NAMBOODIRI, A. M.; PANDEY, J. P. The human cytomegalovirus TRL11/IRL11-encoded FcγR binds differentially to allelic variants of immunoglobulin G1. **Archives of Virology**, v. 156, n. 5, p. 907–910, 2011. Disponível em: <<http://link.springer.com/10.1007/s00705-011-0937-8>>.

NEI, M. Genetic distance between populations. **Amer Naturalist**, , n. 106, p. 283–292, 1972.

PAGE, R. TREEVIEW: an application to display phylogenetic trees on personal computers. **Comput Appl Biosci**, , n. 12, p. 357–358, 1996.

PANDEY, J. P. Immunoglobulin GM and KM allotypes and vaccine immunity. **Vaccine**, v. 19, n. 6, p. 613–7, 2000. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/11090711>>.

PANDEY, J. P. Genomewide association studies and assessment of risk of disease. **The New England journal of medicine**, v. 363, n. 21, p. 2076–7; author reply 2077, 2010. Disponível em: <<http://www.nejm.org/doi/full/10.1056/NEJMc1010310>>.

PANDEY, J. P.; EBBESEN, P.; BOLOW, S.; SVENDSEN, L. B. Brief Communication in Familial Polyposis Coli. , p. 133–136, 1986.

PANDEY, J. P.; FUDENBERG, H. H.; VIRELLA, G.; et al. Association between immunoglobulin allotypes and immune responses to Haemophilus influenzae and Meningococcus polysaccharides. **Lancet (London, England)**, v. 1, n. 8109, p. 190–2, 1979. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/84209>>.

PANDEY, J. P.; GOUST, J. M.; SALIER, J. P.; FUDENBERG, H. H. Immunoglobulin G heavy chain (Gm) allotypes in multiple sclerosis. **Journal of Clinical Investigation**, v. 67, n. 6, p. 1797–1800, 1981.

PANDEY, J. P.; KISTNER-GRIFFIN, E.; IWASAKI, M.; et al. Genetic markers of immunoglobulin G and susceptibility to breast cancer. **Human Immunology**, v. 73, n. 11, p. 1155–1158, 2012. American Society for Histocompatibility and Immunogenetics. Disponível em: <<http://dx.doi.org/10.1016/j.humimm.2012.07.340>>.

PANDEY, J. P.; LI, Z. The forgotten tale of immunoglobulin allotypes in cancer risk and treatment. **Experimental hematology & oncology**, v. 2, n. 1, p. 6, 2013. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3598368&tool=pmc-entrez&rendertype=abstract>>.

PANDEY, J. P.; NAMBOODIRI, A. M.; ELSTON, R. C. Immunoglobulin G

genotypes and the risk of schizophrenia. **Human Genetics**, , n. Pandey 2014, 2016. Springer Berlin Heidelberg. Disponível em: <<http://link.springer.com/10.1007/s00439-016-1706-2>>.

PEAKALL, R.; SMOUSE, P. E. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. **Molecular Ecology Notes**, v. 6, n. 1, p. 288–295, 2006. Disponível em: <<http://doi.wiley.com/10.1111/j.1471-8286.2005.01155.x>>.

PETZL-ERLER, M. L.; LUZ, R.; SOTOMAIOR, V. S. The HLA polymorphism of two distinctive South-American Indian tribes: the Kaingang and the Guarani. **Tissue antigens**, v. 41, n. 5, p. 227–37, 1993. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/8236235>>.

PORTAL KAINGANG. Portal Kaingang. Disponível em: <<http://www.portalkaingang.org/>>. Acesso em: 25/7/2016.

PORTAL-AMAZÔNIA. No Title. Disponível em: <<http://www.portalamazonia.com/>>. Acesso em: 25/7/2016.

PORTER, R. R. The hydrolysis of rabbit γ -globulin and antibodies with crystalline papain. **The Biochemical journal**, v. 73, n. 1, p. 119–26, 1959. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/14434282>>.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. **American journal of human genetics**, v. 81, n. 3, p. 559–75, 2007. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17701901>>.

RAMOS-ONSINS, S. E.; ROZAS, J. Statistical properties of new neutrality tests against population growth. **Molecular biology and evolution**, v. 19, n. 12, p. 2092–100, 2002. Disponível em: <<http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/msl052>>.

RAYMOND, M.; ROUSSET, F. An exact test for population differentiation. **Evolution**, 1995.

REICH, D.; PATTERSON, N.; CAMPBELL, D.; et al. Reconstructing Native American population history. **Nature**, v. 488, n. 7411, p. 370–374, 2012. Disponível em: <<http://www.nature.com/doi/10.1038/nature11258>>.

REYNOLDS, J.; WEIR, B. S.; COCKERHAM, C. C. Estimation of the coancestry coefficient: basis for a short-term genetic distance. **Genetics**, v. 105, n. 3, p. 767–79, 1983. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17246175>>.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Mol Biol and Evol**, , n. 4, p. 406–425, 1987.

SALZANO, F. M.; STEINBERG, A. G. THE GM AND INV GROUPS OF INDIANS FROM SANTA CATARINA, BRAZIL. **American journal of human genetics**, v. 17, n. 3, p. 273–9, 1965. Disponível em:

<<http://www.ncbi.nlm.nih.gov/pubmed/14295495>>.

SAMBROOK, J.; FRITSCH, E. F.; MANIATIS, T. **Molecular cloning**. 1989.

SARVAS, H.; RAUTONEN, N.; MÄKELÄ, O. Allotype-associated differences in concentrations of human IgG subclasses. **Journal of clinical immunology**, v. 11, n. 1, p. 39–45, 1991. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/2022720>>.

SCHANFIELD, M. S. Population affinities of the Australian aborigines as reflected by the genetic markers of immunoglobulins. **Journal of Human Evolution**, v. 6, n. 4, p. 341–352, 1977. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0047248477800031>>.

SCHANFIELD, M. S. The Anthropological Usefulness of Highly Polymorphic Systems. **Current Developments in Anthropological Genetics**. p.65–85, 1980. Boston, MA: Springer US. Disponível em: <http://link.springer.com/10.1007/978-1-4613-3084-4_4>.

SCHANFIELD, M. S. Immunoglobulin allotypes (GM and KM) indicate multiple founding populations of Native Americans: evidence of at least four migrations to the New World. **Human biology**, v. 64, n. 3, p. 381–402, 1992. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/1607185>>.

SCHROEDER, H. W. J.; CAVACINI, L. Structure and Function of Immunoglobulins (author manuscript). **Journal of Allergy and Clinical Immunology**, v. 125, n. 125, p. S41–S52, 2010. Elsevier Ltd. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3670108&tool=pmc-entrez&rendertype=abstract>>.

SCHUR, P. H.; CHRISTIAN, G. D. THE ROLE OF DISULFIDE BONDS IN THE COMPLEMENT-FIXING AND PRECIPITATING PROPERTIES OF 7S RABBIT AND SHEEP ANTIBODIES. **The Journal of experimental medicine**, v. 120, n. 15, p. 531–45, 1964. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/14212117>>.

SELSING, E. Ig class switching: targeting the recombinational mechanism. **Current Opinion in Immunology**, v. 18, n. 3, p. 249–254, 2006.

SKOGLUND, P.; MALLICK, S.; BORTOLINI, M. C.; et al. Genetic evidence for two founding populations of the Americas. **Nature**, v. 525, n. 7567, p. 104–110, 2015. Disponível em: <<http://www.nature.com/doi/10.1038/nature14895>> <<http://www.ncbi.nlm.nih.gov/pubmed/26196601>>.

SLATKIN, M. A measure of population subdivision based on microsatellite allele frequencies. **Genetics**, v. 139, n. 1, p. 457–62, 1995. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/7705646>>.

STAVNEZER, J. Mechanism and regulation of class switch recombination. **Annual Review of Immunology**, v. 26, p. 261–92, 2008. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2707252/>>.

TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics**, v. 123, n. 3, p. 585–95, 1989. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1203831&tool=pmc-entrez&rendertype=abstract>>.

TAKAHASHI, N.; UEDA, S.; OBATA, M.; et al. Structure of human immunoglobulin gamma genes: implications for evolution of a gene family. **Cell**, v. 29, n. 2, p. 671–9, 1982. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/6811139>>.

TANAKA, T.; NEI, M. Positive darwinian selection observed at the variable-region genes of immunoglobulins. **Mol Biol Evol**, v. 6, n. 5, p. 447–459, 1989. Disponível em: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2796726>.

TAO, M. H.; CANFIELD, S. M.; MORRISON, S. L. The differential ability of human IgG1 and IgG4 to activate complement is determined by the COOH-terminal sequence of the CH2 domain. **The Journal of experimental medicine**, v. 173, n. April, p. 1025–1028, 1991.

TAO, M. H.; SMITH, R. I.; MORRISON, S. L. Structural features of human immunoglobulin G that determine isotype-specific differences in complement activation. **The Journal of experimental medicine**, v. 178, n. August, p. 661–7, 1993. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2191116&tool=pmc-entrez&rendertype=abstract>>.

TEAM, R. D. C. R: A language and environment for statistical computing. , 2008. Vienna, Austria: R Foundation for Statistical Computing. Disponível em: <<http://www.r-project.org>>.

TSUNETO, L. T.; PROBST, C. M.; HUTZ, M. H.; et al. HLA class II diversity in seven Amerindian populations. Clues about the origins of the Aché. **Tissue antigens**, v. 62, n. 6, p. 512–26, 2003. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/14617035>>.

VIDARSSON, G.; DEKKERS, G.; RISPENS, T. IgG subclasses and allotypes: From structure to effector functions. **Frontiers in Immunology**, v. 5, n. OCT, 2014.

WHITTINGHAM, S.; MATHEWS, J. D.; SCHANFIELD, M. S.; TAIT, B. D.; MACKAY, I. R. Interaction of HLA and Gm in autoimmune chronic active hepatitis. **Clinical and experimental immunology**, v. 43, n. 1, p. 80–6, 1981. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1537131&tool=pmc-entrez&rendertype=abstract>>.

WILLIAMS, R. C.; STEINBERG, A. G.; GERSHOWITZ, H.; et al. GM allotypes in Native Americans: Evidence for three distinct migrations across the Bering land bridge. **American Journal of Physical Anthropology**, v. 66, n. 1, p. 1–

19, 1985. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/3976868>>.

WRIGHT, S. **Evolution and the Genetics of Populations: Variability within and among natural populations**. Chicago, 1978.

XU, J.; TURNER, A.; LITTLE, J.; BLEECKER, E. R.; MEYERS, D. A. Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: Hint for genotyping error? **Human Genetics**, v. 111, n. 6, p. 573–574, 2002.

XU, Z.; ZAN, H.; PONE, E. J.; MAI, T.; CASALI, P. Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. **Nature Reviews Immunology**, v. 12, n. 7, p. 517–531, 2012. Nature Publishing Group. Disponível em: <<http://dx.doi.org/10.1038/nri3216>>.

ZELASCHI, D.; NEWBY, C.; PARSONS, M.; et al. Human immunoglobulin allotypes: previously unrecognized determinants and alleles defined with monoclonal antibodies. **Proceedings of the National Academy of Sciences**, v. 80, n. 12, p. 3762–3766, 1983. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.80.12.3762>>.

ANEXO 1

Tabela A1: SNPs para os quais foram extraídos dados de genótipos de indivíduos das populações PEL, BEB, FIN e YRI do projeto *1000 genomes*

Segmento gênico	Éxon	rs
<i>IGHG1</i>	<i>CH1</i>	<i>rs1071803</i>
	<i>CH2</i>	<i>rs377538050</i>
		<i>rs1043109</i>
	<i>CH3</i>	<i>rs1045853</i>
		<i>rs11621259</i>
<i>rs113804727</i>		
<i>IGHG2</i>	<i>CH1</i>	<i>rs587648672</i>
	<i>CH2</i>	<i>rs11627594</i>
		<i>rs8009156</i>
	<i>CH3</i>	<i>rs113678609</i>
<i>IGHG3</i>	<i>CH2</i>	<i>rs368359789</i>
		<i>rs145035200</i>
		<i>rs74093865</i>
		<i>rs60746425</i>
		<i>rs12890621</i>
		<i>rs141959627</i>
		<i>rs113169458</i>
		<i>rs77307099</i>
		<i>rs78376194</i>
		<i>rs587739524</i>
		<i>rs149653267</i>
		<i>rs139413052</i>
		<i>rs4042056</i>
	<i>rs1051112</i>	
<i>CH3</i>		

SNPs selecionados porque resultaram em substituições não sinônimas no presente estudo. Abreviatura e tamanho das populações: PEL: (Peruanos em Peru, n=85); FIN (Finlandeses em Finlândia, n=99), BEB (Bengalis em Bangladesh, n=86), YRI (Iorubás em Nigéria, n=108) (AUTON et al., 2015).

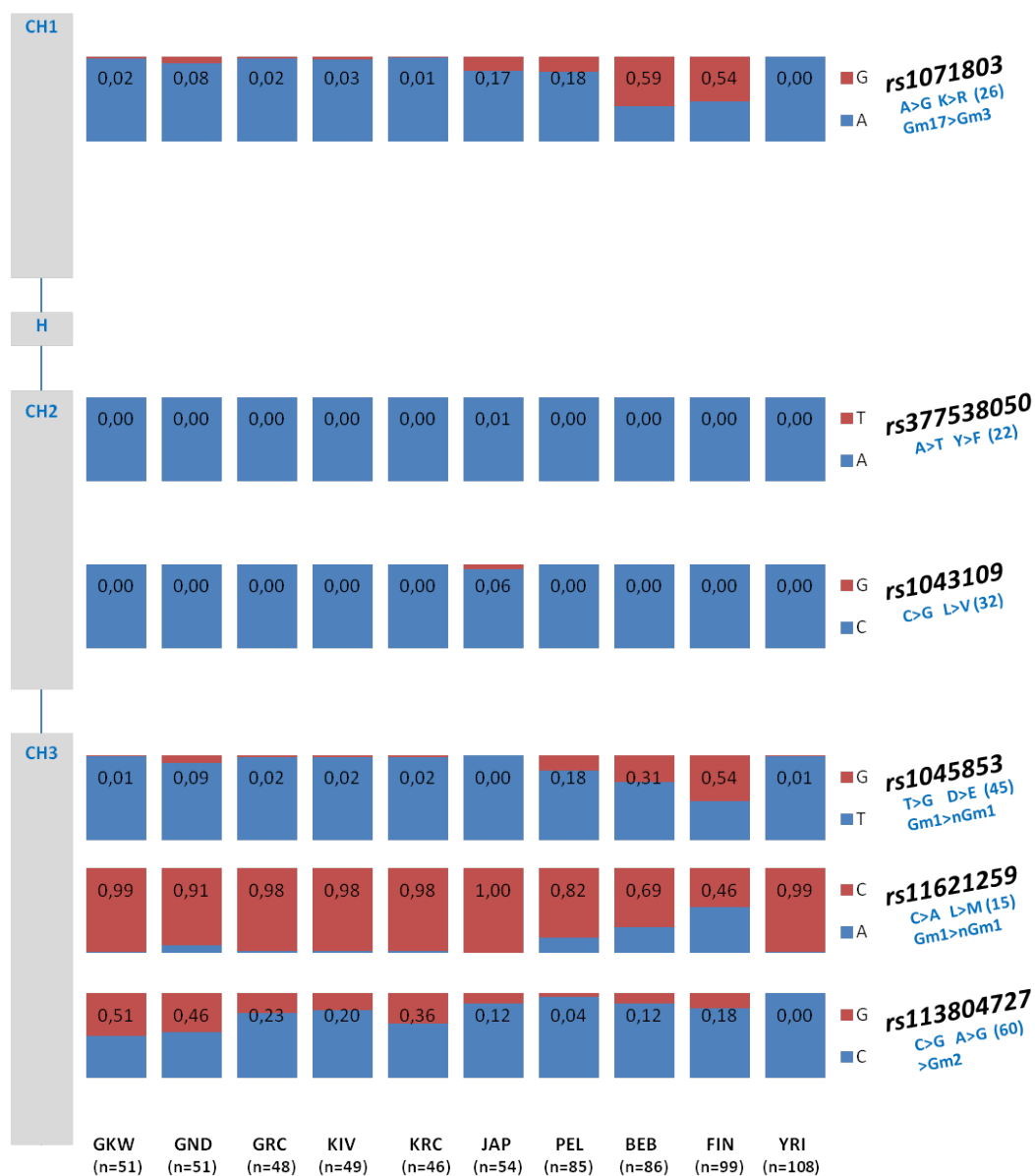


Figura A1: Frequências nucleotídica dos SNPs de *IGHG1* selecionados que resultaram em substituições não sinônimas no presente estudo. Fonte: A autora. Abreviaturas: GKW: Guarani Kaiowá; GND: Guarani Nãndeva; GRC: Guarani Mbyá; KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; JAP: Descendentes de Japoneses; PEL: Peruanos em Peru; FIN: Finlandeses em Finlândia; BEB: Bengalis em Bangladesh; YRI: Iorubás em Nigéria.

As frequências mostradas nos gráficos correspondem ao nucleotídeo indicado em vermelho. A substituição de aminoácidos resulta no determinante antigênico do alótipo (Gm) mencionado quando encontram-se após >, e na perda do alótipo quando encontram-se antes de >.

Entre parêntesis estão indicadas as distâncias evolutivas entre os aminoácidos envolvidos na substituição de acordo com (GRANTHAM, 1974). Esta escala pode ter valores desde 5 até 215 baseadas nas diferenças químico-físicas entre aminoácidos. Quanto a maior valor, maior diferença entre aminoácidos.

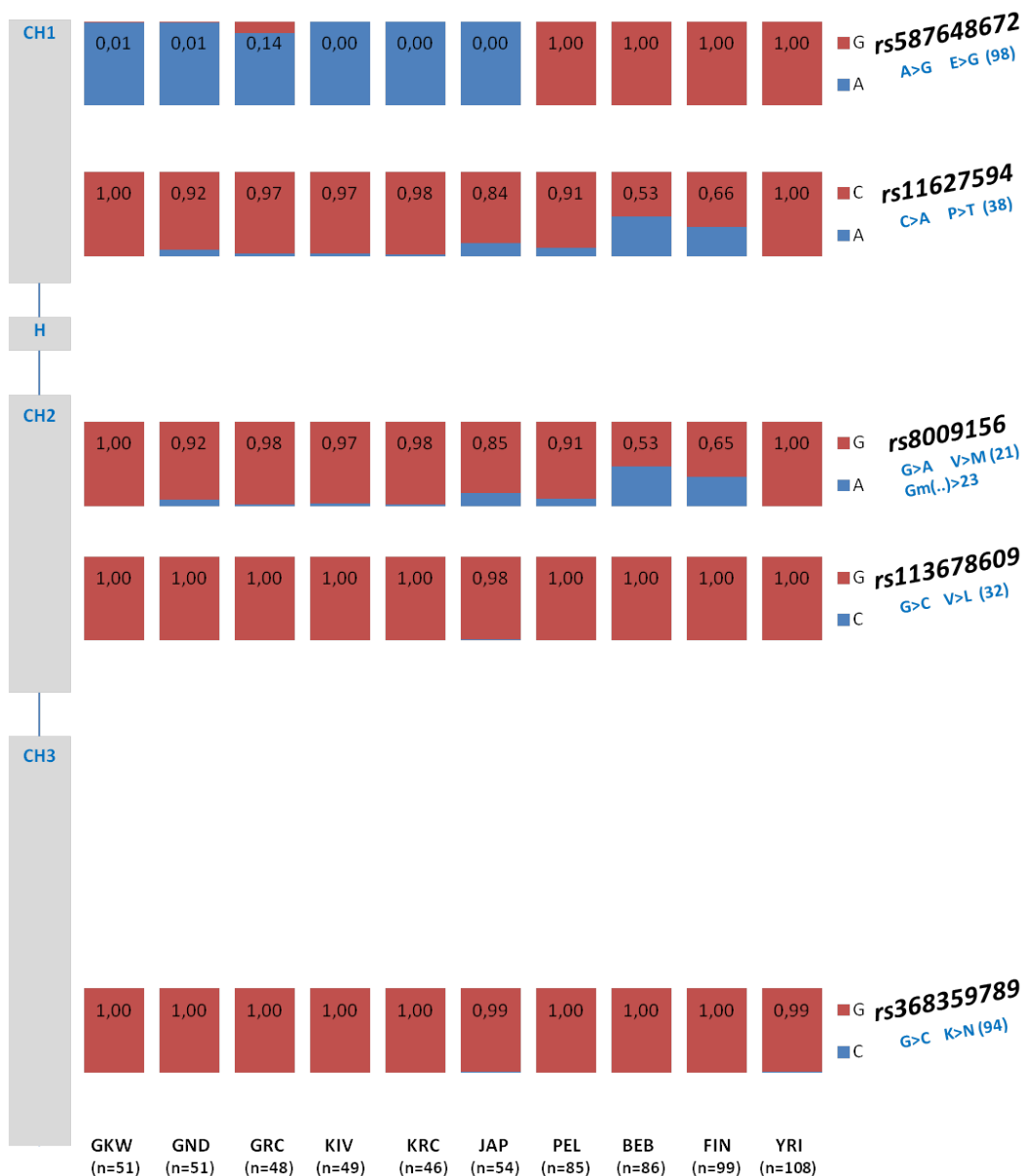


Figura A2: Frequências nucleotídica dos SNPs de *IGHG2* selecionados que resultaram em substituições não sinônimas no presente estudo. Fonte: A autora. Abreviaturas: GKW: Guarani Kaiowá; GND: Guarani Nãndeva; GRC: Guarani Mbyá; KIV: Kaingang de Ivaí; KRC: Kaingang de Rio das Cobras; JAP: Descendentes de Japoneses; PEL: Peruanos em Peru; FIN: Finlandeses em Finlândia; BEB: Bengalis em Bangladesh; YRI: Iorubás em Nigéria.

As frequências mostradas nos gráficos correspondem ao nucleotídeo indicado em vermelho. A substituição de aminoácidos resulta no determinante antigênico do alótipo (Gm) mencionado quando encontram-se após >, e na perda do alótipo quando encontram-se antes de >.

Entre parêntesis estão indicadas as distâncias evolutivas entre os aminoácidos envolvidos na substituição de acordo com (GRANTHAM, 1974). Esta escala pode ter valores desde 5 até 215 baseadas nas diferenças químico-físicas entre aminoácidos. Quanto a maior valor, maior diferença entre aminoácidos.

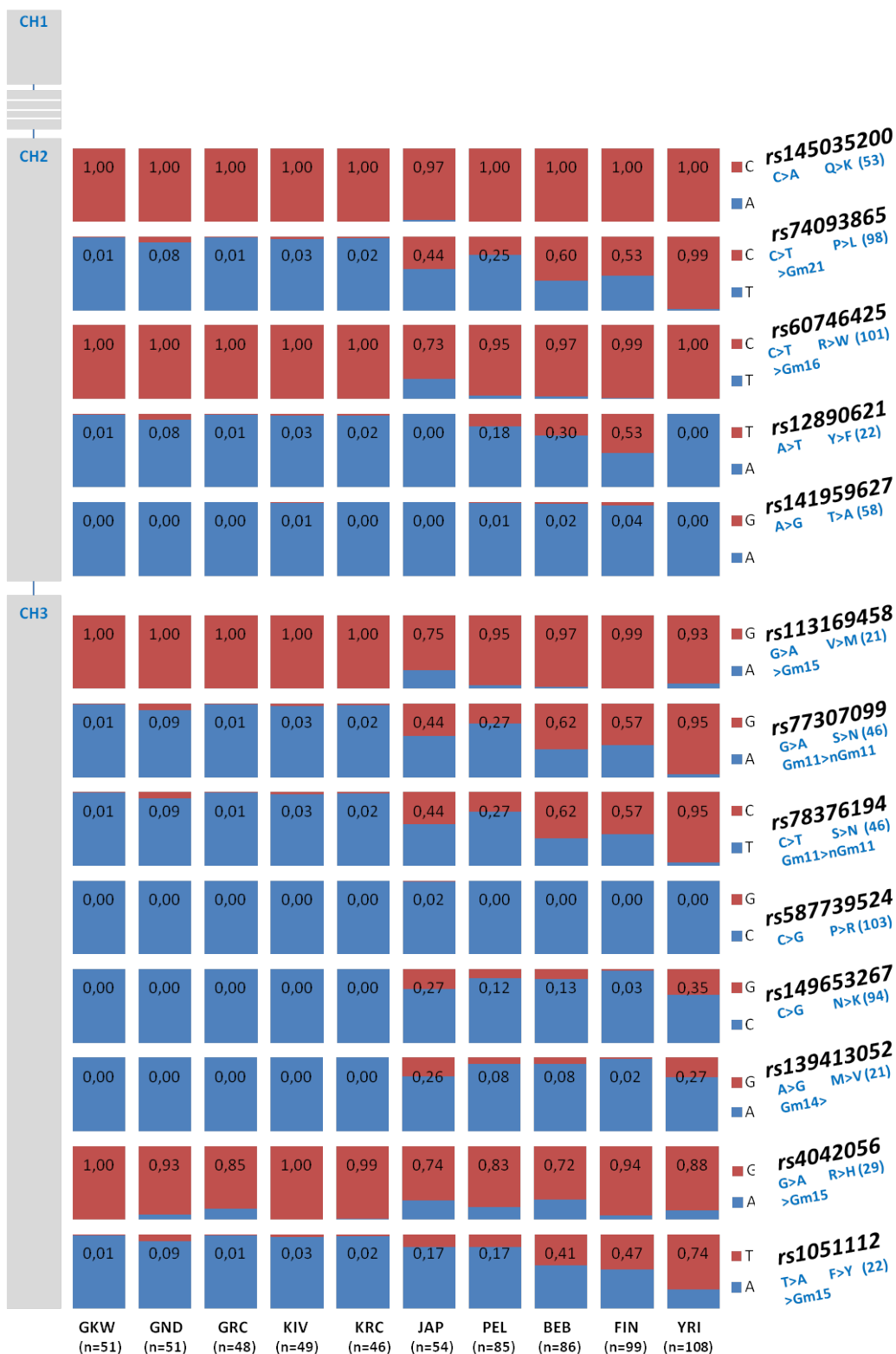


Figura A3: Frequências nucleotídica dos SNPs de *IGHG3* selecionados que resultaram em substituições não sinônimas no presente estudo. Fonte: A autora. Abreviaturas: GKW: Guarani Kaiowá; GND: Guarani Nãndeva; GRC: Guarani Mbyá; KIV: Kaingang de Ivai; KRC: Kaingang de Rio das Cobras; JAP: Descendentes de Japoneses; PEL: Peruanos em Peru; FIN: Finlandeses em Finlândia; BEB: Bengalis em Bangladesh; YRI: Iorubás em Nigéria.

As frequências mostradas nos gráficos correspondem ao nucleotídeo indicado em vermelho. A substituição de aminoácidos resulta no determinante antigênico do alótipo (Gm) mencionado quando encontram-se após >, e na perda do alótipo quando encontram-se antes de >.

Entre parêntesis estão indicadas as distâncias evolutivas entre os aminoácidos envolvidos na substituição de acordo com (GRANTHAM, 1974). Esta escala pode ter valores desde 5 até 215 baseadas nas diferenças químico-físicas entre aminoácidos. Quanto a maior valor, maior diferença entres aminoácidos.

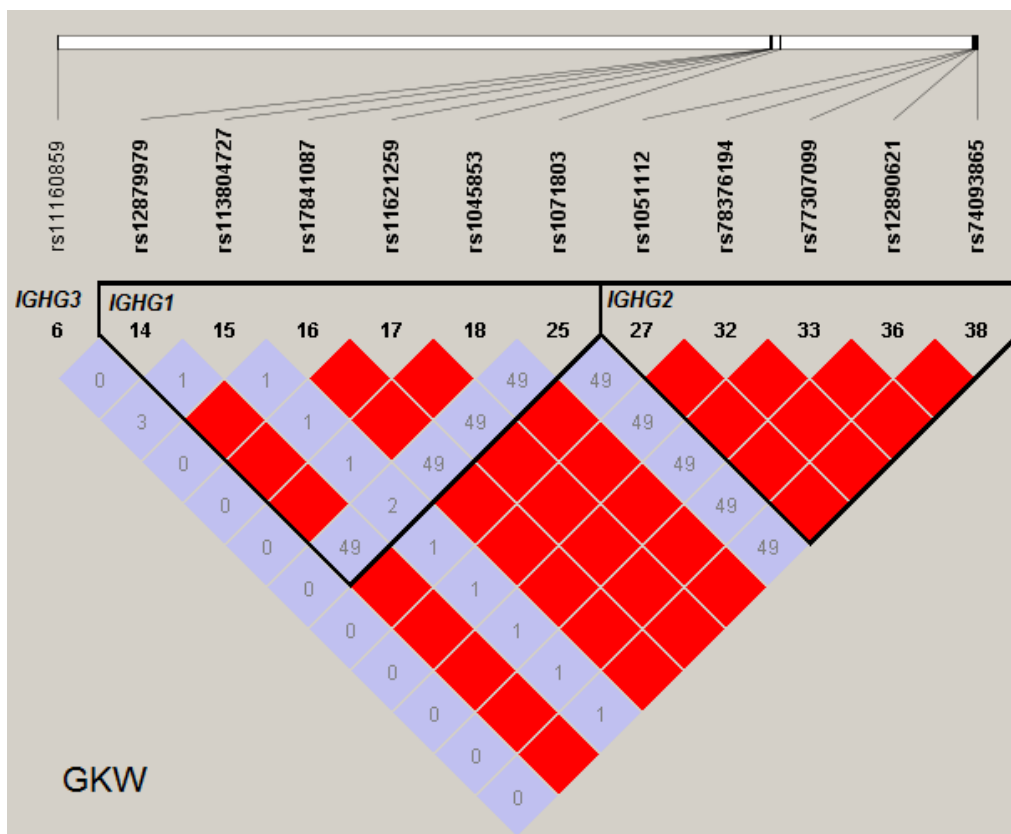


Figura A4: Desequilíbrio de ligação na população Guarani Kaiowa (GKW) entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* variáveis em todas as populações de Ameríndios. Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza. Dentro de cada losango estão representados os valores de $r^2(x100)$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).

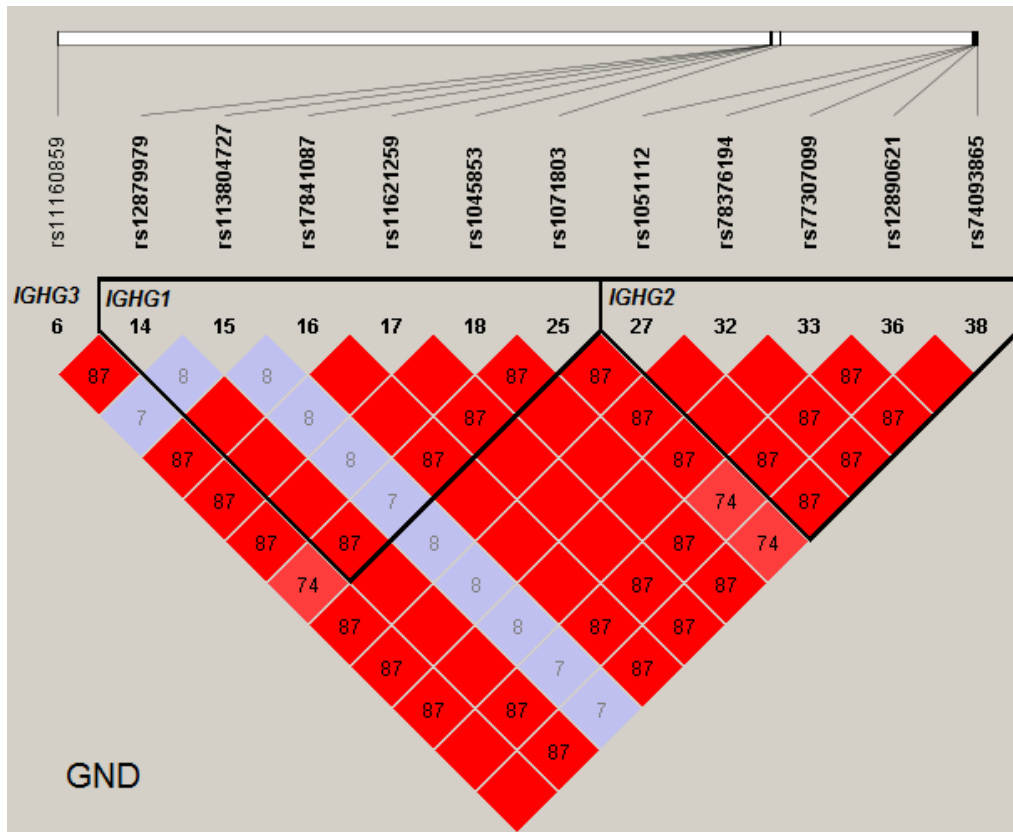


Figura A5: Desequilíbrio de ligação na população Guarani Nandeva (GND) entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* variáveis em todas as populações de Ameríndios. Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza. Dentro de cada losango estão representados os valores de $r^2(x100)$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).

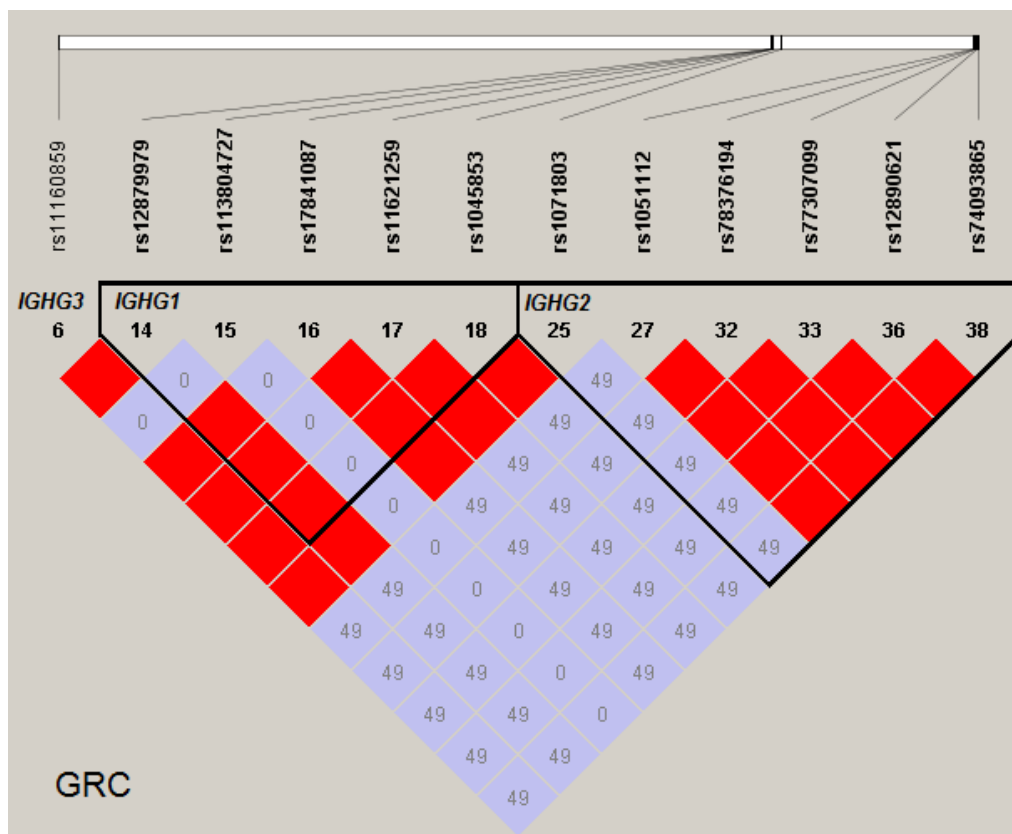


Figura A6: Desequilíbrio de ligação na população Guarani Mbya (GRC) entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* variáveis em todas as populações de Ameríndios. Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza. Dentro de cada losango estão representados os valores de $r^2(x100)$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).

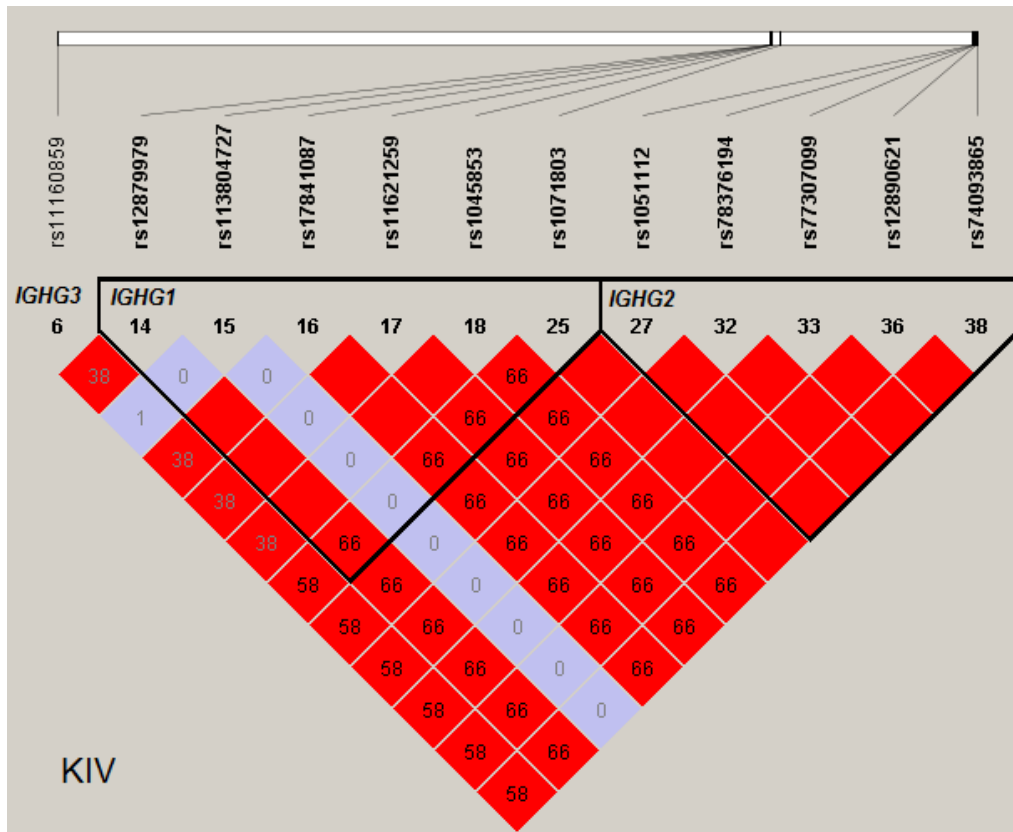


Figura A7: Desequilíbrio de ligação na população Kaingang de Ivai (KIV) entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* variáveis em todas as populações de Ameríndios. Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza. Dentro de cada losango estão representados os valores de $r^2(x100)$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).

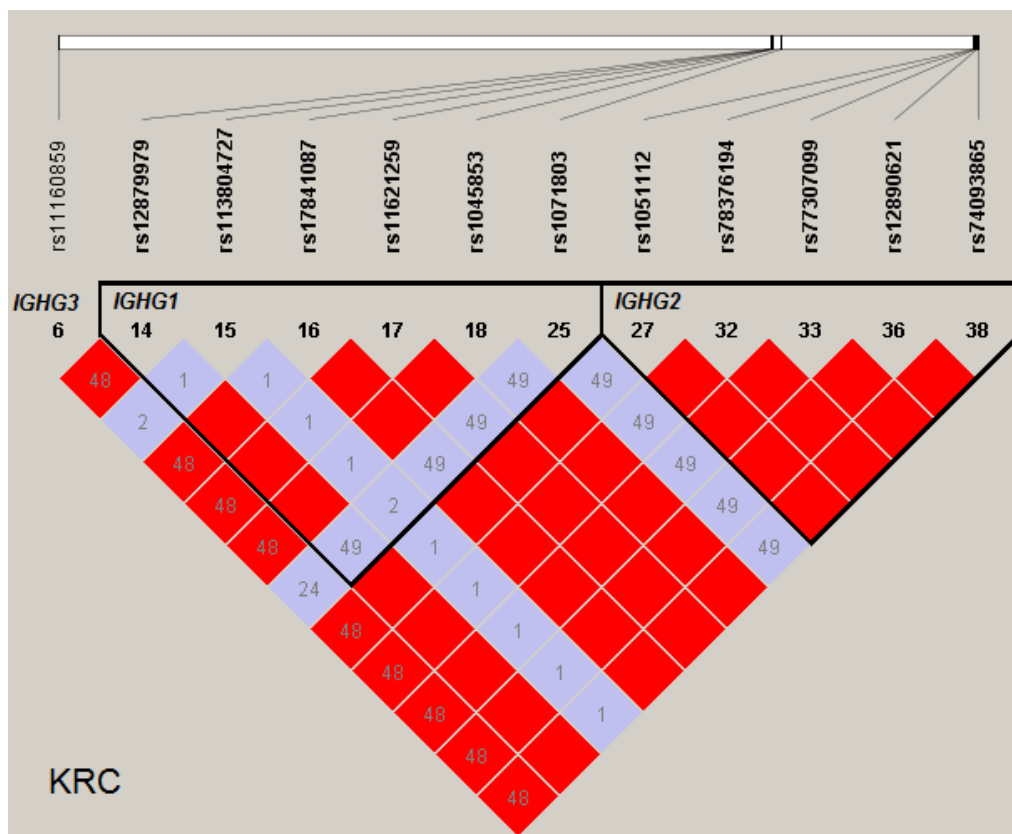


Figura A8: Desequilíbrio de ligação na população Kaingang de Rio das Cobras (KRC) entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* variáveis em todas as populações de Ameríndios. Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza. Dentro de cada losango estão representados os valores de $r^2(x100)$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto).

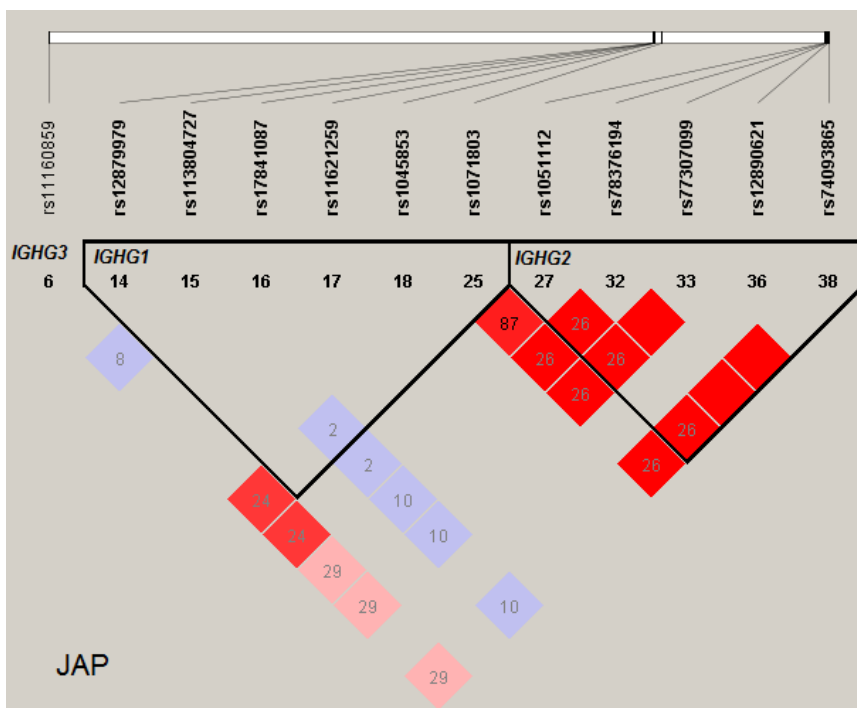


Figura A9: Desequilíbrio de ligação, na população descendentes de Japoneses (JAP), entre SNPs dos segmentos gênicos *IGHG3*, *IGHG1* e *IGHG2* variáveis em todas as populações de Ameríndios. Os valores máximos de D'/LOD estão representados pela cor vermelha, e os valores mínimos de cinza. Dentro de cada losango estão representados os valores de $r^2 \times 100$; quando está vazio indica que r^2 é 100 (desequilíbrio absoluto). Os lugares sem cor indicam que a população é monomórfica nesta posição e não é possível estimar o desequilíbrio de ligação.