

FERNANDA CARDOSO

**Divergências na distribuição dos genes V(D)J codificadores de anticorpos contra antígenos do *P. falciparum* nos genomas modernos da África e do Círculo Ártico, e nos genomas dos Neandertais**

São Paulo, 2023.

**FERNANDA CARDOSO**

**Divergências na distribuição dos genes V(D)J codificadores de anticorpos contra antígenos do *P. falciparum* nos genomas modernos da África e do Círculo Ártico, e nos genomas dos Neandertais**

Dissertação apresentada ao Departamento de Imunologia do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

**Área de concentração:** Imunologia

**Orientadora:** Profª. Dra. Maristela Martins de Camargo.

**Coorientador:** Prof. Dr. João Marcelo Pereira Alves.

Versão original.

São Paulo, 2023.

**À MINHA FAMÍLIA E MEU NAMORADO, FELIX P.,  
POR TODO O SUPORTE E AMOR INCONDICIONAL.**

***TO MY FAMILY AND MY BOYFRIEND, FELIX P., FOR ALL  
THE SUPPORT AND UNCONDITIONAL LOVE.***

## AGRADECIMENTOS

Ao longo da minha trajetória nesta pós-graduação, poucas pessoas desempenharam papéis fundamentais no meu crescimento pessoal e profissional. Por isso, é com profundo apreço que gostaria de expressar minha gratidão a elas.

À minha família, que sempre apoiou meus sonhos e me acolheu nos momentos em que mais precisei, dedico minha gratidão. O amor e a dedicação que me foram concedidos são inestimáveis. Especialmente à minha querida tia, **Luciana Cardoso**, que sempre confiou em mim e me apoiou em todos os momentos da minha vida. É a ela que dedico este trabalho, como um gesto de profundo reconhecimento.

*To my boyfriend, Felix P., I extend my gratitude for his partnership and love throughout this journey. Thank you for consistently believing in me, for bringing happiness to my days, and for supporting all my decisions.*

À minha orientadora, Professora Doutora Maristela Martins de Camargo, expresso minha sincera gratidão pela sua orientação acadêmica. Agradeço por me proporcionar a oportunidade de concluir meu mestrado, fornecendo recursos e ferramentas essenciais para esse feito, e por estar sempre disponível quando necessário.

Ao meu coorientador, Professor Doutor João Marcelo Pereira Alves, agradeço por aceitar a coorientação deste projeto de mestrado. Sou grata pela sólida orientação em bioinformática, pela excelência na co-supervisão acadêmica, paciência e sabedoria que compartilhou.

À Professora Doutora Alessandra Pontillo, expresso minha gratidão por todos os ensinamentos durante a disciplina PAE, que contribuíram significativamente para meu entendimento da imunologia. Além disso, sua simplicidade e gentileza tornaram essa jornada ainda mais enriquecedora.

Ao Professor Marcus Vinícius Chrysóstomo Baldo e aos membros do NePer (Núcleo de Estratégias em Planejamento Experimental e Reprodutibilidade), agradeço pela assistência valiosa na análise estatística deste projeto.

Aos professores que participaram da minha banca de qualificação, minha sincera gratidão por todas as sugestões e ensinamentos que foram essenciais para a conclusão bem-sucedida deste projeto de pesquisa.

Às alunas do Laboratório de Imuno Regulação Molecular, localizado no ICB IV, agradeço pelo auxílio durante a execução deste projeto.

Ao acolhimento generoso dos alunos do Laboratório 40 de Genômica e Bioinformática de Parasitas, localizado no ICB II, e em especial ao aluno de doutorado Davi Salles Xavier, agradeço pela amizade, carinho e confiança que encontrei durante meu último ano de mestrado.

À Secretaria de Pós-graduação e a Maria Eni do Sacramento Santos, agradeço pelo apoio inestimável durante o período da pós-graduação.

À Universidade de São Paulo (USP), expresso minha gratidão pelo suporte acadêmico que me proporcionou.

Aos respeitáveis professores e pós-doutores que gentilmente aceitaram o convite e dedicaram parte do seu tempo para a discussão do meu trabalho durante minha banca de defesa, agradeço profundamente por sua contribuição valiosa.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), estendo minha gratidão pelo suporte financeiro que viabilizou meus estudos.

**“Yo soy yo y mis circunstancias”**  
**José Ortega y Gasset**

## LISTA DE ABREVIACÕES

**Ab** - Anticorpo (do inglês *antibody*)  
**BCR** - Receptores de Células B (do inglês *B cell receptor*)  
**BLAST** - Do inglês *The Basic Local Alignment Search Tool*  
**C** - Segmento/Região constante  
**CDRs** - Regiões determinantes de complementaridade (do inglês *complementarity determining regions*)  
**CEPSH** - Comissão de Ética em Pesquisa com Seres Humanos  
**CEUA** - Comissão de Ética no Uso de Animais  
**chr** - Cromossomos (do inglês *chromosome*)  
**COVID-19** - Do inglês *Coronavirus Disease of 2019*  
**CSP** - Proteína Circunsporozoíta (do inglês *Circumsporozoite Protein*)  
**D** - Segmento gênico diversidade  
**DNA** - Ácido Desoxirribonucleico (do inglês *Deoxyribonucleic Acid*)  
**DNA-PKc** - Proteína quinase dependente de DNA catalítica subunidade (do inglês *DNA-dependent protein kinase*)  
**DSBs** - Quebras de fitas duplas (do inglês *Double-Strand Breaks*)  
**E** - Envelope  
**EBV** - Vírus Epstein-Barr (do inglês *Epstein-Barr Virus*)  
**ECA2** - Enzima Conversora de Angiotensina 2  
**Fabs** - Fragmento de ligação do antígeno (do inglês *Fragment antigen binding*)  
**Fc** - Fragmento cristalizável  
**GPI** - Glicosilfosfatidilinositol  
**gRNA** - RNA genômico  
**HCDR3** - CDR 3 da cadeia pesada  
**HGDP** - Do inglês *Human Genome Diversity Project*

**HMG1 e HMG2** - hidroximetilglutaril/coenzima A redutasehydroxymethylglutaryl-CoA redutase 1 e 2  
**HTS** - Sequenciamento de alto rendimento (do inglês *High-Throughput Sequencing*)  
**IG** - Imunoglobulina  
**IGH** - Cadeia pesada  
**IGL** - Cadeia leve  
**IGκ** - Cadeia kappa  
**IGλ** - Cadeia lambda  
**J** - Segmento gênico junção  
**kb** - Kilobases  
**LAG-3** - Gene 3 de ativação linfocitária  
**M** - Membrana  
**mAb** - Anticorpo monoclonal (do inglês *monoclonal antibody*)  
**mRNA** - RNA mensageiro  
**N** - Nucleocapsídeo  
**n** - Número  
**NANP** - *malária Asn-Ala-Asn-Pro*  
**NCBI** - do inglês *National Center for Biotechnology Information*  
**NHEJ** - Junção de extremidade não homóloga (do inglês *Non-homologous end joining*)  
**OMS** - Organização Mundial da Saúde  
**ORF** - Fases de leitura aberta (do inglês *Open Reading Frame*)  
**p** - Probabilidade de significância  
**P** - Pseudogenes  
***P. falciparum*** - *Plasmodium falciparum*  
**PAXX** - Parálogo de XRCC4 e XLF  
**pb** - Pares de base  
**PfCyRPA** - Antígeno protetor rico em cisteína do *P. falciparum* (do inglês *P. falciparum cysteine-rich protective antigen*)  
**PfRH5** - Proteína Homóloga de Ligação a Reticulócitos 5 do *P. falciparum* (do inglês *P. falciparum Reticulocyte-binding protein*)

*homologue 5)*

**PfRipr** - Proteína de interação com PfRH5

**Pfs25** - Proteína de superfície da fase sexuada do *P. falciparum* (do inglês *P. falciparum sexual-stage surface protein*)

**Proteína S** - do inglês *spike glycoprotein/protein*

**RAG** - Gene Ativador de Recombinação (do inglês *Recombination-Activating Genes*)

**RBD** - Domínio de Ligação ao Receptor (do inglês *Receptor-Binding Domain*)

**RNA** - ácido ribonucleico (do inglês *Ribonucleic Acid*)

**RSS** - Sequências de Sinais de Recombinação (do inglês *Recombination Signal Sequences*)

**SARS-CoV-2** - Síndrome Respiratória Aguda - Coronavírus 2

**SGDP** - Do inglês *Simons Genome Diversity Project*

**SR-NGS** - Sequenciamento de nova geração com leituras curtas (do inglês *Short-Read Next-Generation Sequencing*)

**TdT** - Desoxinucleotidil Transferase Terminal (do inglês *Terminal deoxynucleotidyl Transferase*)

**TIM-3** - Mucina e imunoglobulina de célula T

**TMPRSS2** - Protease Transmembrana Serina 2 (do inglês *Transmembrane Serine Protease 2*)

**UNICEF** - Fundo das Nações Unidas para a Infância (do inglês *United Nations International Children's Emergency Fund*)

**V** - Segmento gênico variável

**WGS** - Do inglês *Whole Genome Sequencing*

**XRCC4** - Proteína de complementação cruzada de reparo de raios X 4 (do inglês *DNA repair protein XRCC4* ou *X-ray repair cross*

*complementing protein 4)*

**βS** - Beta-globina S

**σ** - Desvio padrão



## SUMÁRIO

RESUMO .....	1
ABSTRACT .....	2
INTRODUÇÃO .....	3
Malária.....	3
Ciclo de vida .....	5
Os desafios para um tratamento eficaz contra a malária .....	6
O surgimento do <i>P. falciparum</i> e sua relação evolutiva com os seres humanos .....	7
SARS-COV-2.....	8
Ciclo de vida .....	8
SARS-COV-2 na África.....	9
SARS-COV-2 <i>versus</i> malária .....	11
Anticorpos.....	12
Recombinação V(D)J .....	13
A importância do estudo das sequências germinativas de imunoglobulinas.....	21
OBJETIVO GERAL.....	22
Objetivos específicos .....	22
MATERIAIS E MÉTODOS .....	23
Seleção de anticorpos neutralizantes e identificação dos segmentos V(D)J .....	23
Seleção dos genomas humanos.....	29
Análise dos genomas humanos.....	29
Análise estatística .....	30
RESULTADOS .....	31
DISCUSSÃO.....	42
CONCLUSÃO .....	48
LISTA DE FIGURAS.....	49
REFERÊNCIAS.....	52

## RESUMO

Os loci de imunoglobulina (IG) possuem variantes alélicas nos genes V(D)J, formando receptores de células B específicos para antígenos. Diferentes segmentos germinativos de IG influenciam respostas imunes, eficácia contra infecções e susceptibilidade a doenças. Por isso, identificar esses elementos é crucial para compreender doenças imunológicas. Este projeto teve como objetivo investigar possíveis diferenças no uso dos segmentos gênicos V(D)J relacionados à geração de anticorpos neutralizantes contra *P. falciparum* e SARS-CoV-2, entre populações que sofrem diferentes exposições ao *P. falciparum*. Para isso, sete anticorpos neutralizantes contra *P. falciparum* e SARS-CoV-2 foram selecionados manualmente a partir da literatura, levando em consideração a disponibilidade de suas estruturas cristalográficas. Os segmentos gênicos que compõem os anticorpos e foram analisados possuíam identidade igual ou acima de 75% com as suas sequências germinativas. O gene da insulina foi utilizado como controle para verificar a integridade dos genomas analisados. Foram selecionados 48 genomas humanos, com alta cobertura de sequenciamento, para a análise. Quatro grupos diferentes foram comparados: Antigo, África Subsaariana, Migrante e Círculo Polar Ártico. Foram identificados 17 segmentos gênicos, sendo três sequências comuns para a produção de anticorpos contra ambos os patógenos. Os genomas de hominídeos antigos apresentaram a menor presença dos segmentos que geram anticorpos contra os antígenos estudados. O grupo Círculo Ártico mostrou maior frequência normalizada para cinco segmentos gênicos em relação aos grupos África Subsaariana e Migrante. Enquanto esses últimos grupos apresentaram maior amplitude na distribuição dos segmentos VJ. No entanto, a exposição ao *P. falciparum* não parece ter afetado significativamente a frequência dos segmentos estudados. Os resultados sugerem uma diversidade genética influenciada por deriva genética, com segmentos gênicos presentes de forma estocástica em diferentes populações. O trabalho também destacou a necessidade de explorar mais genomas antigos para identificar possíveis diferenças entre humanos antigos e modernos em relação ao uso dos segmentos V(D)J. Nosso estudo utilizou a melhor técnica de sequenciamento disponível, mas apontou para a necessidade de novas tecnologias, como o sequenciamento de leituras longas, para compreender completamente a variação genética nos genes de imunoglobulina. Essas descobertas podem direcionar pesquisas futuras sobre segmentos V(D)J protetores contra infecções do *P. falciparum* e de outros agentes infecciosos. Ainda, saber quais segmentos gênicos estão mais presentes em uma determinada população pode auxiliar na predição da morbidade em relação a um determinado patógeno. A diversidade genética das populações humanas é fundamental para a resposta imunológica a patógenos, e a seleção de segmentos germinativos pode ter sido influenciada por interações com agentes infecciosos ao longo da história evolutiva.

**Palavras-chave:** anticorpos, *Plasmodium falciparum*, SARS-CoV-2, segmentos V(D)J.

## ***ABSTRACT***

The immunoglobulin (IG) loci harbor allelic variants of the V(D)J genes, forming specific B cell receptors for antigens. Different germline alleles of IG influence immune responses, efficacy against infections, and susceptibility to diseases. Therefore, identifying these elements is crucial for understanding immunological disorders. This project aimed to investigate possible differences in the usage of V(D)J gene segments related to the generation of neutralizing antibodies against *P. falciparum* and SARS-CoV-2 among populations with different exposures to *P. falciparum*. Seven neutralizing antibodies against *P. falciparum* and SARS-CoV-2 were manually selected from the literature, considering the availability of their crystallographic structures. The gene segments that make up the antibodies and were analyzed had identity equal to or above 75% with their germline sequences. The insulin gene was used as a control to verify the integrity of the analyzed genomes. For analysis, 48 human genomes with high sequencing coverage were selected. Four different groups were compared: Ancient, Sub-Saharan Africa, Migrant, and Arctic Circle. Seventeen gene segments were identified, with three common sequences for antibody production against both pathogens. Ancient genomes showed the lowest presence of segments generating antibodies against the studied antigens. The Arctic Circle group exhibited a higher normalized frequency for five gene segments compared to the Sub-Saharan Africa and Migrant groups. The latter two groups showed a broader distribution of VJ segments. However, exposure to *P. falciparum* does not appear to have significantly affected the frequency of the studied segments. The results suggest a genetic diversity influenced by genetic drift, with gene segments present stochastically in different populations. The work has also emphasized the need to further explore ancient genomes to identify potential differences between ancient and modern humans regarding the use of V(D)J segments. The study used the best available sequencing technique but pointed to the need for new technologies, such as long-read sequencing, to fully understand genetic variation in immunoglobulin genes. These findings can guide future research on protective V(D)J segments against *P. falciparum* infections and other infectious agents. Furthermore, knowing which gene segments are more prevalent in a particular population can assist in predicting morbidity concerning a specific pathogen. The genetic diversity of human populations is crucial for the immune response to pathogens, and the selection of germline segments may have been influenced by interactions with infectious agents throughout evolutionary history.

**Keywords:** antibodies, *Plasmodium falciparum*, SARS-Cov-2, V(D)J segments.

## INTRODUÇÃO

### Malária

A malária é uma das principais causas de morte na história humana (1). Transmitida por mosquitos do gênero *Anopheles*, a doença é causada pelos protozoários do gênero *Plasmodium spp.* (2) e exerce uma das mais importantes pressões seletivas na linhagem humana (3). As espécies mais comuns e que representam uma grande carga na saúde pública são o *Plasmodium falciparum* (*P. falciparum*) e o *P. vivax* (4). Segundo o Relatório de Estatísticas de Saúde Mundial 2023 (*World health statistics 2023*), divulgado pela Organização Mundial da Saúde (OMS), estima-se que houve 247 milhões de casos de malária em 84 países endêmicos em 2021, um aumento de 2 milhões em relação a 2020 (5). Este é um aumento menor em comparação com o salto de 232 milhões de casos em 2019 para 245 milhões em 2020, o que revelou o impacto massivo do primeiro ano da pandemia de COVID-19 (do Inglês *coronavirus disease of 2019*) nos esforços globais de controle da malária (5) (Figura 1a.). As mortes causadas pela malária em todo o mundo também aumentaram, e foram estimadas em 568.000 em 2019 e 625.000 em 2020, mas caíram para 619.000 em 2021 (5) (Figura 1b.). A região Africana respondeu por cerca de 95% dos casos e 96% das mortes globais causadas pela doença em 2021 (5). Além disso, a malária é uma das principais causas de mortalidade infantil no continente africano (6).



**Figura 1** Distribuição de (a) casos de malária e (b) mortes, por país, 2021. Extraída do Relatório de Estatísticas da Saúde Mundial 2023 da Organização Mundial da Saúde (OMS). Figura Modificada. Referência: LISTA DE FIGURAS.

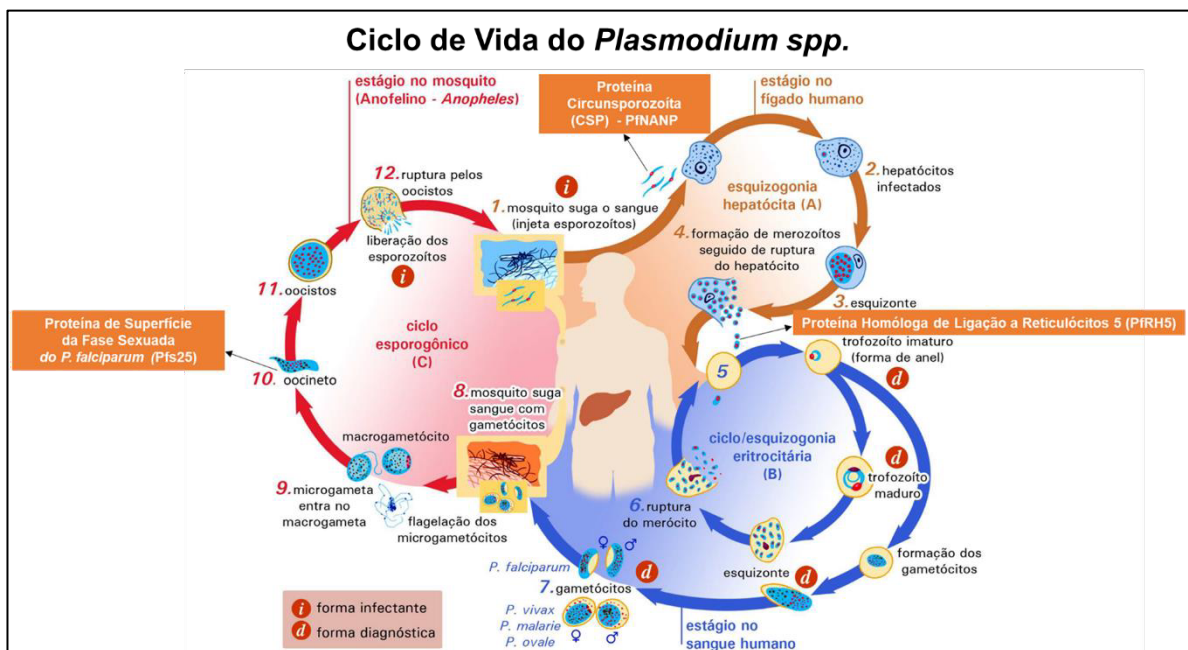
## Ciclo de vida

O *P. falciparum* é responsável pela forma mais grave da malária em seres humanos (7). A infecção ocorre quando um mosquito fêmea infectado pica uma pessoa e introduz entre 10 e 100 esporozoítos móveis na camada dérmica da pele (8). Estes esporozoítos possuem a proteína circunsporozoíta (CSP) (do inglês *Circumsporozoite Protein*), uma das principais proteínas sintetizadas pelo parasita, que forma um revestimento denso em sua superfície (9,10). A CSP desempenha um papel crucial na formação, migração e invasão dos esporozoítos, facilitando sua adesão e entrada em células-alvo (10–12). Uma característica importante da CSP é uma região repetitiva que contém de 25 a 42 cópias de quatro aminoácidos denominados NANP (*Asn-Ala-Asn-Pro*) (13). Os anticorpos contra essa porção repetitiva da CSP desempenham um papel vital no desenvolvimento da imunidade protetora (13). Após 30 a 60 minutos da inoculação inicial (14), os esporozoítos entram na corrente sanguínea e alcançam o fígado (15), escapando assim da imunidade do hospedeiro ou de serem retirados da circulação pelo sistema linfático (16), iniciando o ciclo hepático do parasita (15). Nessa fase, os esporozoítos invadem os hepatócitos, se multiplicam e se diferenciam em esquizontes, que produzem milhares de merozoítos hepáticos (15). Entre 6 e 10 dias (14), os merozoítos são liberados na corrente sanguínea, dando início ao ciclo sanguíneo da doença (8). Esse ciclo envolve a invasão dos eritrócitos pelos merozoítos por meio de interações específicas entre a Proteína Homóloga de Ligação a Reticulócitos 5 (PfRH5) (do inglês *reticulocyte-binding protein homologue 5*) do *P. falciparum* (*Pf*), que se liga ao receptor Basigin no eritrócito, marcando uma etapa crucial e irreversível no processo de invasão da hemácia (17). Durante a invasão, a PfRH5 se associa a outras proteínas de superfície do merozoíto, formando um complexo de invasão essencial, incluindo o antígeno protetor rico em cisteína (PfCyRPA) (do inglês *P. falciparum* *cysteine-rich protective antigen*), a proteína de interação com PfRH5 e o PfP113 ligado a glicosilfosfatidilinositol (GPI) (18).

Dentro da hemácia, o parasita se encontra envolvido pela membrana do vacúolo parasitóforo e remodela a célula hospedeira, causando modificações internas e externas que permitem sua sobrevivência e reprodução (8). Nessa etapa, o *P. falciparum* passa por reprodução assexuada, sofrendo mudanças morfológicas desde o estágio de anel até o estágio de trofozoíto e, finalmente, o estágio de esquizonte maduro (14). Este é o segundo ciclo de reprodução assexuada, e em 24 a 72 horas o parasita replica seu genoma haplóide (16,19). A forma esquizonte madura contém de 16 a 32 merozoítos (19) que são liberados quando a hemácia infectada se rompe, permitindo que invadam novas hemácias (20). Com ciclos repetidos de invasão e crescimento, o parasita estabelece infecções agudas e, eventualmente, crônicas (16). A fase sanguínea do ciclo de vida do *P. falciparum* é sintomática e está associada à destruição maciça de hemácias (8), causando sintomas como febre, dor de cabeça, vômitos, anemia, coma e, em casos graves, morte (20).

Alguns dos parasitas sanguíneos assexuados modificam seu desenvolvimento para a fase sexuada (16) diferenciando-se em gametócitos femininos e masculinos (20), permitindo a transmissão do parasita para o mosquito vetor (15). Quando um mosquito se alimenta do sangue de um hospedeiro infectado, os gametócitos passam por desenvolvimento sexual no interior do mosquito (19). Minutos após entrarem no intestino médio do vetor, os gametócitos masculinos e femininos se diferenciam em

gametas, que se fundem para formar um zigoto diplóide, que se desenvolverá em um oocineto móvel (14). Os oocinetos expressam em sua superfície a proteína de superfície da fase sexuada Pfs25 do *P. falciparum* (do inglês *Plasmodium falciparum sexual-stage surface protein*), que é importante para a sobrevivência do oocineto no intestino médio rico em proteases do mosquito, na invasão do epitélio do intestino médio e na maturação do oocisto (21). Os oocinetos viajam e conseguem penetrar na parede intestinal do mosquito para se desenvolver em oocistos, que produzem milhares de esporozoítos. O ciclo de vida é concluído quando os esporozoítos maduros, originados dos oocistos rompidos, migram para as glândulas salivares do mosquito e são injetados em um novo hospedeiro humano através da picada do inseto (14,20) (Figura 2).



**Figura 2** Malária, ciclo de vida do *Plasmodium* spp. no ser humano e mosquito. Figura Modificada. Referência: LISTA DE FIGURAS.

### Os desafios para um tratamento eficaz contra a malária

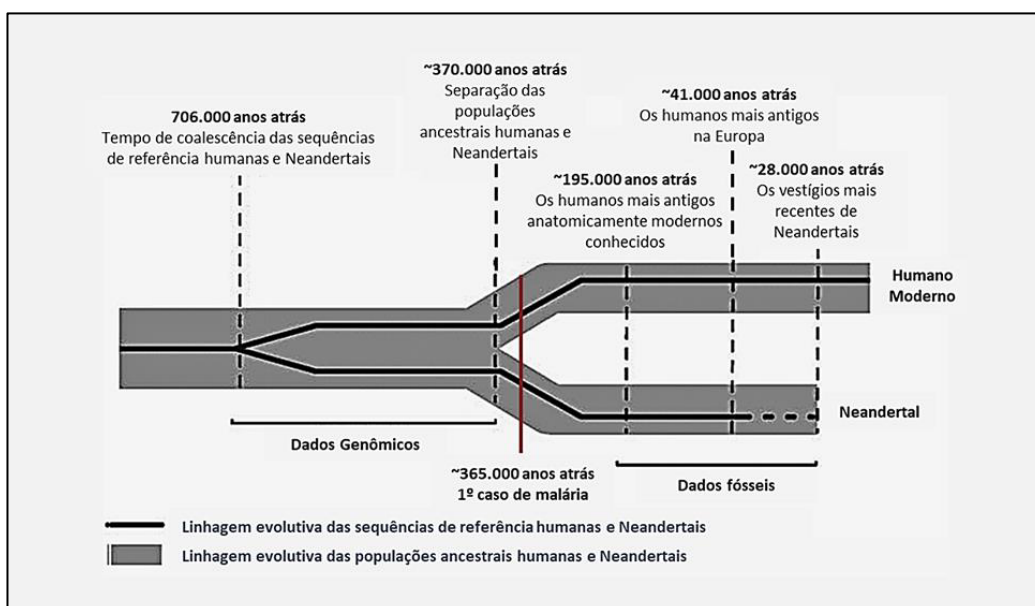
A resistência dos patógenos representa um desafio para os esforços de controle microbiano (22). No caso dos parasitas da malária, a resistência surgiu contra todos os antimaláricos conhecidos (23). Ao longo do último meio século, a implantação de cinco classes de medicamentos antimaláricos resultou em impactos seletivos bem definidos, levando à disseminação mundial da resistência (22). Isso impulsionou a presença de alelos de resistência aos medicamentos em várias linhagens de *P. falciparum* (22,24). A genômica do *P. falciparum* tem sido amplamente estudada em populações da África Ocidental, revelando forte seleção direcional em genes conhecidos por conferir resistência a medicamentos, bem como assinaturas de seleção em vários antígenos candidatos a vacinas (25).

Sabe-se que as imunoglobulinas, componentes centrais do sistema imunológico, desempenham um papel crítico na proteção contra vários patógenos, incluindo o *P. falciparum* (1). No entanto, o desenvolvimento de uma vacina contra a malária tem sido marcado por momentos de entusiasmo e desapontamento (26). As expectativas iniciais de controle vetorial inovador e cura radical com uma

única dose vacinal ainda não alcançaram o impacto necessário para atingir a meta de "malária zero" (27). Um dos principais obstáculos para o desenvolvimento de uma vacina contra a malária é o complexo ciclo de vida do *P. falciparum*, que ocorre tanto no hospedeiro humano quanto na fêmea do mosquito vetor *Anopheles* (21). Além disso, a imunidade natural e induzida por vacina para esse protozoário ainda não é totalmente compreendida (26). Nas áreas endêmicas, a imunidade ao *P. falciparum* se desenvolve lentamente após infecções repetidas, mas raramente é esterilizante (28). Estudos recentes têm destacado o papel das funções efetoras dos anticorpos mediados pelo fragmento cristalizável (Fc), responsável por determinar a função do anticorpo, na proteção contra a infecção pelo *P. falciparum* (24), e é amplamente reconhecido que uma vacina antimalárica eficaz, capaz de gerar imunidade adaptativa e duradoura, será fundamental para a erradicação dessa doença (18,26).

### O surgimento do *P. falciparum* e sua relação evolutiva com os seres humanos

Embora o período exato em que o *P. falciparum* infectou os ancestrais dos humanos modernos seja incerto, seu impacto na evolução humana é amplamente reconhecido (29). Até recentemente, acreditava-se que o *P. falciparum* havia coevoluído com os humanos ao longo de milhões de anos (30). No entanto, avanços recentes revelaram polimorfismos antigênicos no parasita (31), e evidências de que o *P. falciparum* evoluiu a partir de parasitas ancestrais de gorilas por meio de transferência entre espécies (29,30). No entanto, pesquisas mais atuais estimam que as primeiras infecções dos ancestrais dos humanos modernos pelo *P. falciparum* ocorreram há cerca de 112.000 a 1.036.000 anos atrás (com uma estimativa média de 365.000 anos atrás) (29) (Figura 3). Essa nova estimativa indica uma relação mais próxima entre a infecção por *P. falciparum* em humanos e o surgimento de polimorfismos em eritrócitos, considerados adaptações genéticas à malária (29). Além disso, esse intervalo de tempo coincide com o surgimento da espécie *Homo sapiens* (32), sugerindo uma possível troca de hospedeiros com adaptações específicas do *Plasmodium* para nossa espécie (29).



**Figura 3** Estimativas de divergência para sequências genômicas humanas e Neandertais e populações ancestrais humanas e Neandertais, mostradas em relação às datas de eventos críticos na evolução



dos seres humanos modernos e dos Neandertais. Também foi demonstrado a estimativa do primeiro caso de malária em ancestrais dos humanos modernos. Os comprimentos dos ramos são esquemáticos e não estão em escala. Figura Modificada. Referência: LISTA DE FIGURAS.

### **SARS-CoV-2**

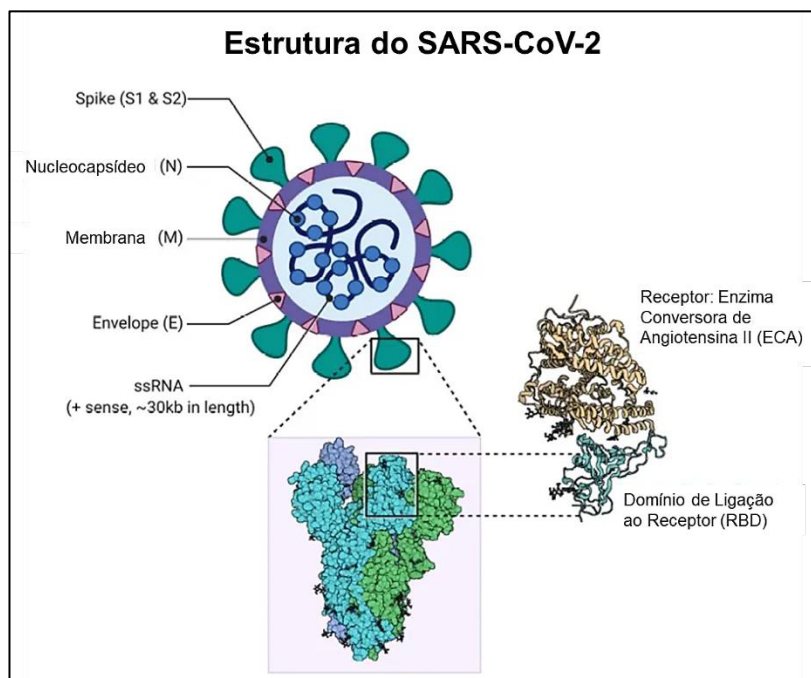
O Coronavírus-2, causador da Síndrome Respiratória Aguda Grave (SARS-CoV-2) é altamente transmissível e patogênico (33,34). Surgindo em 2019, causou a pandemia de doença respiratória aguda conhecida como COVID-19 (33). O SARS-CoV-2 teve um impacto sem precedentes na sociedade e economia global, marcando a terceira introdução zoonótica de um coronavírus altamente patogênico na população humana (35,36). Por isso, a doença levou a uma expansão extraordinária das pesquisas focadas no agente causador (37). Notavelmente, testemunhamos eventos evolutivos que anteriormente eram amplamente descritos por inferência indireta, incluindo a diversificação do SARS-CoV-2 em variantes com características fenotípicas distintas, incluindo transmissibilidade, gravidade e evasão imune (36). Segundo a OMS, até o momento, foram registrados 769.806.130 casos confirmados de COVID-19, com 6.955.497 mortes em âmbito global (38). Vacinas e tratamentos com anticorpos monoclonais foram desenvolvidos em um ano após o início da pandemia (39), e bilhões de pessoas foram vacinadas para se protegerem do vírus, salvando mais de 20 milhões de vidas (40).

### **Ciclo de vida**

O SARS-CoV-2 é transmitido por fômites e gotículas durante o contato social próximo e desprotegido entre infectados e não infectados (41), e pacientes sintomáticos e assintomáticos são as principais fontes de infecção (41). O vírus também pode se espalhar por contato indireto, quando as mãos contaminadas entram em contato com as mucosas da boca, nariz e olhos (41). O SARS-CoV-2 é um vírus envelopado com ácido ribonucleico (RNA) (do inglês *ribonucleic acid*) de fita simples positiva (35), e possui um genoma com aproximadamente 30 kilobases (kb) (37). É composto pelas proteínas do nucleocapsídeo (N), membrana (M), envelope (E) e a glicoproteína S (do inglês *spike*), que são proteínas estruturais (42) (Figura 4). A proteína S é montada como um homotrímero e é inserida em múltiplas cópias na membrana do vírion, conferindo-lhe uma aparência de coroa (34). A entrada do SARS-CoV-2 na célula depende da interação entre a glicoproteína S e o receptor da enzima conversora de angiotensina 2 (ECA2) (43) (Figura 4). Assim, a proteína S é clivada por proteínas convertases como a furina (34,37), e são separadas em duas subunidades, uma extracelular e outra transmembrana, nas células infectadas (42,44). Ou seja, a clivagem ocorre antes da liberação do vírus pela célula que o produz (42). Portanto, a proteína S no vírion maduro consiste em duas subunidades associadas não covalentemente: a subunidade S1 exposta à superfície e contendo o domínio de ligação ao receptor RBD (do inglês *receptor-binding domain*), que se liga especificamente ao receptor da célula hospedeira, a ECA2 (42). A subunidade S2 ancora a proteína S à membrana e inclui um peptídeo de fusão, que medeia a interação das membranas virais e celulares durante a infecção de uma nova célula (35,42).

A interação da ECA2 com a proteína S induz alterações conformacionais dramáticas em ambas as subunidades que unem as membranas viral e celular, criando um poro de fusão que permite ao

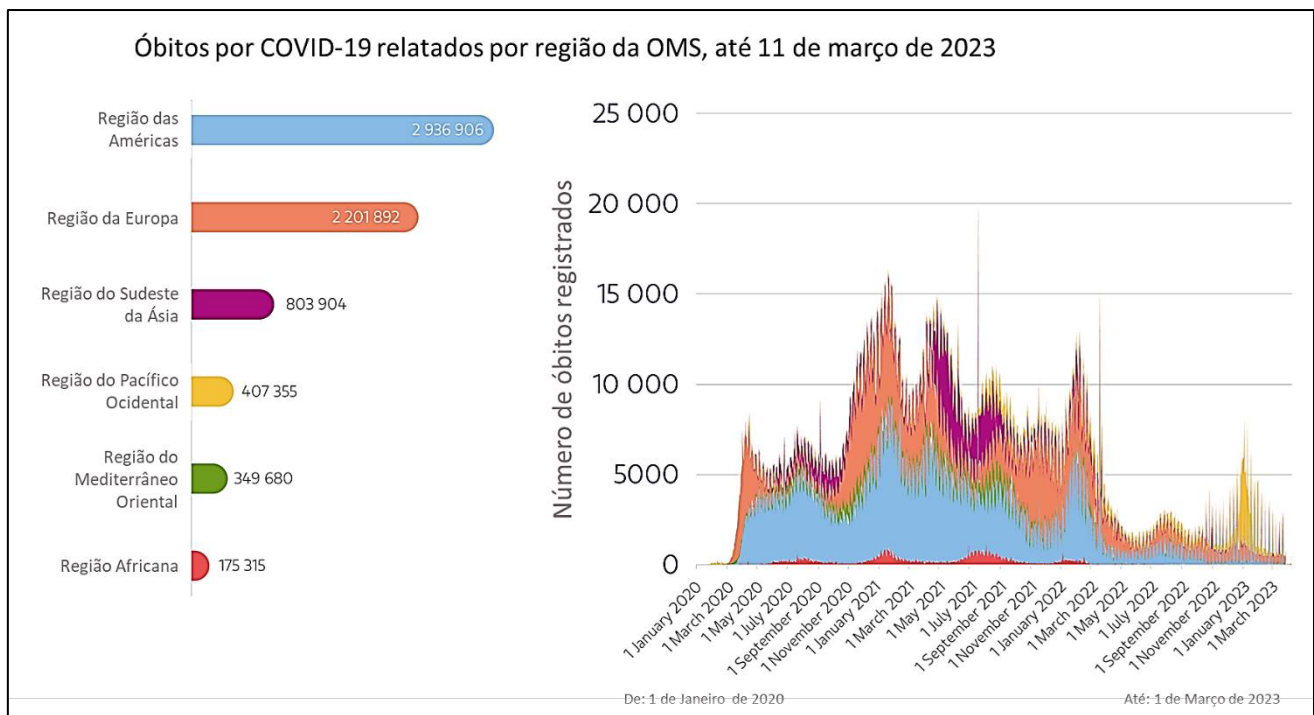
genoma viral alcançar o citoplasma celular (45). Uma dessas alterações é a clivagem de um sítio adicional interno à subunidade S2, denominado “sítio S2” (46). A interação da ECA2 com o vírus expõe o ‘sítio S2’, que é clivado pela protease transmembrana serina 2 (TMPRSS2) na superfície celular ou pela catépsina L no compartimento endossomal após endocitose mediada pela ECA2 (42). Em seguida, o peptídeo de fusão é liberado, iniciando a formação dos poros de fusão (42). Após a fusão das membranas, o RNA genômico (gRNA) do SARS-CoV-2 é liberado no citosol (37), e da mesma forma que todos os vírus de RNA de cadeia positiva em eucariotos, a replicação do SARS-CoV-2 ocorre inteiramente no citoplasma (37). Durante o ciclo de vida intracelular, o SARS-CoV-2 expressa e utiliza a maquinaria celular para replicar seu gRNA e produzir cópias completas que são incorporadas em partículas virais recém-produzidas (35). Durante a replicação, o gRNA é traduzido para proteínas não estruturais que rapidamente suprimem a tradução dos RNAs mensageiros (mRNA) do hospedeiro em favor das moléculas de RNA do vírus (47). Em seguida, o vírus transforma o retículo endoplasmático em vesículas de membrana dupla (47), que constituem o principal local de replicação do SARS-CoV-2 (41). Novos vírions foram recentemente relatados deixando a célula por meio do tráfego lisossômico, em vez da via secretora biossintética usada por muitos outros vírus envelopados (37). O SARS-CoV-2 replica-se abundantemente no epitélio respiratório superior, onde a ECA2 também é expressa, e, assim, o vírus é transmitido com eficiência (35).



**Figura 4** Estrutura do vírus SARS-CoV-2 e visualização do Domínio de Ligação ao Receptor (RBD).  
 Figura Modificada. Referência: LISTA DE FIGURAS.

### SARS-CoV-2 na África

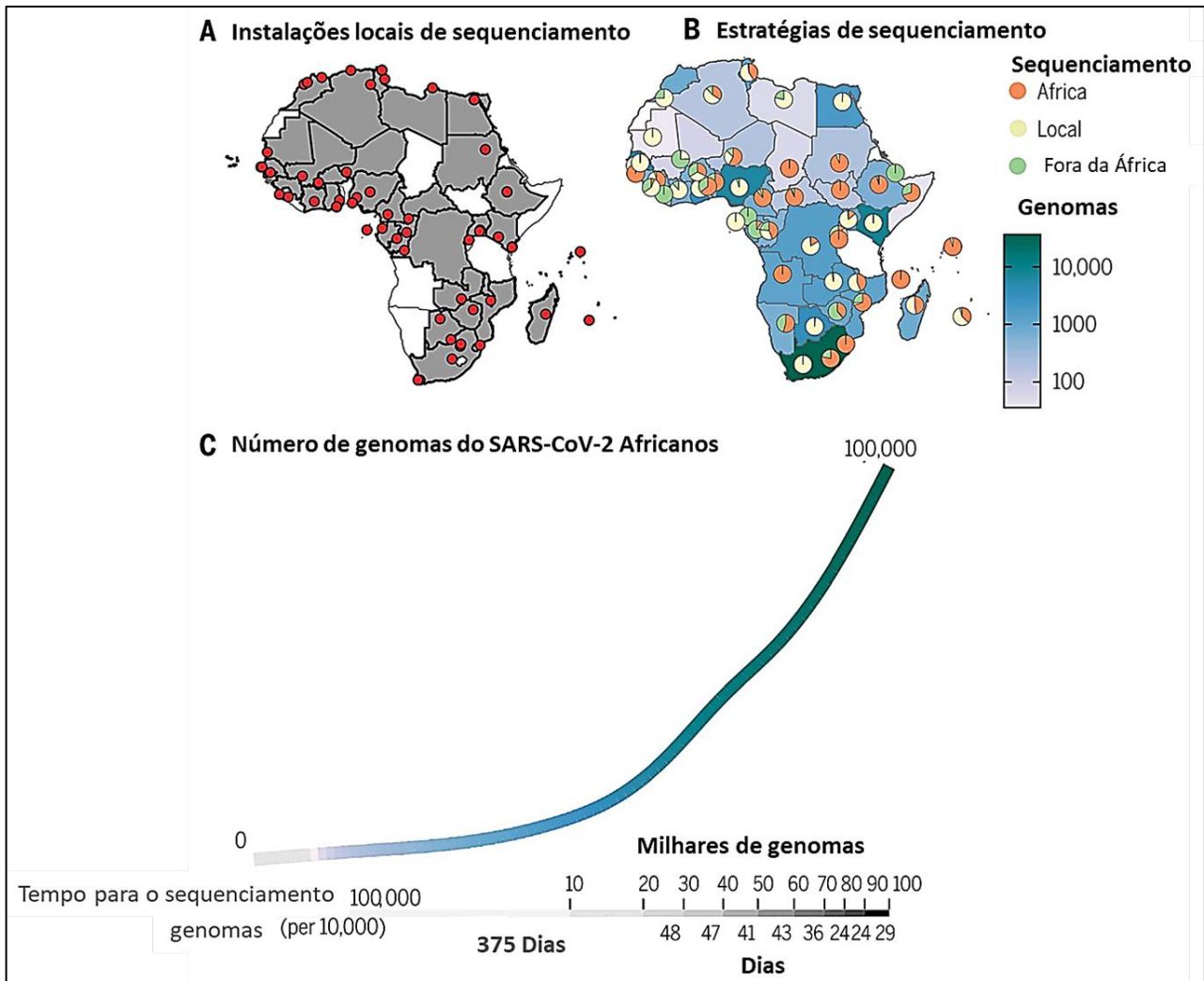
O impacto da COVID-19 na África tem sido substancialmente menor em comparação com os países das Américas, Europa e Ásia (48,49). A OMS relatou no continente africano 175.315 mortes até 11 de março de 2023 (5), o menor número entre todas as regiões da OMS, (Figura 5) (5).



**Figura 5** Óbitos por COVID-19 relatados por região da OMS, até 11 de março de 2023. COVID-19: doença do coronavírus de 2019, OMS: Organização Mundial da Saúde. Figura Modificada. Referência: LISTA DE FIGURAS.

A Plataforma “*Global Development Commons*” da UNICEF (Fundo das Nações Unidas para a Infância/ *United Nations International Children's Emergency Fund*) aponta cinco fatores relevantes pelas quais a COVID-19 tem sido menos letal na África do que em outras regiões (50). São eles: 1) Ação rápida. A maioria dos governos africanos tomou medidas drásticas para tentar retardar a propagação do vírus. 2) Apoio público ou adesão da comunidade aos protocolos de saúde pública. 3) População jovem - e poucos lares de terceira idade. 4) Clima favorável. 5) Bons sistemas de saúde comunitários. Várias nações da África Ocidental (que lutaram contra o pior surto de Ebola do mundo entre 2013 e 2016) se destacaram na implementação das medidas de saúde pública usadas para prevenir a COVID-19, incluindo isolar os infectados, rastrear seus contatos e colocá-los em quarentena enquanto eles eram testados (50).

Tegally e colaboradores (2021), mostraram como os esforços coordenados de talentosos cientistas africanos deram, em pouco tempo, grandes contribuições para a vigilância e coleta de dados da pandemia (51). Seus esforços e iniciativas forneceram um alerta precoce que provavelmente beneficiou os países mais ricos mais do que os seus próprios (51). As primeiras 100.000 sequências de SARS-CoV-2 da África ajudaram a monitorar a epidemia no continente, com a vigilância genômica se expandindo ao longo da mesma (52). Além disso, os métodos de sequenciamento foram adaptados para lidar com um vírus em evolução (52), (Figura 6). Por fim, foi também avaliado como as linhagens virais se espalharam pelo continente em uma estrutura filogeográfica para obter informações sobre a dinâmica de transmissão temporal e espacial para diversas variantes de preocupação (52).



**Figura 6** Expansão da capacidade de sequenciamento do SARS-CoV-2 na África. (A) Países Africanos (sombreados em cinza) e instituições (círculos vermelhos) com instalações de sequenciamento locais capazes de produzir genomas completos de SARS-CoV-2. (B) Número de genomas de SARS-CoV-2 produzidos por país e a proporção de genomas produzidos localmente, regionalmente na África, ou no exterior. (C) Diminuição do tempo de retorno da produção de sequenciamento na África para uma liberação quase em tempo real de dados genômicos. SARS-CoV-2: Síndrome Respiratória Aguda por Coronavírus 2. Figura Modificada. Referência: LISTA DE FIGURAS.

### SARS-CoV-2 versus malária

Osei e colaboradores (2022) investigaram a relação entre a letalidade ou proteção da co-infecção em pacientes diagnosticados com COVID-19 e malária (53). Através de uma revisão sistemática, incluindo artigos publicados a partir de dezembro de 2019 até maio de 2021, o grupo concluiu que a baixa incidência de COVID-19 em regiões endêmicas de malária apoia a hipótese de que o prognóstico ruim da primeira é prevenido pela segunda (53). Embora mais estudos sejam necessários para verificar essa hipótese, o aumento de receptores co-inibitórios, por exemplo mucina imunoglobulina de célula T (TIM-3) e o gene 3 de ativação linfocitária (LAG-3), e células B de memória atípicas que ocorrem na malária sugerem regulação imunológica entre essa e a COVID-19 (53). Em 2020, o sistema de

saúde italiano (um dos países europeus que mais sofreu com a pandemia de COVID-19 nesse período), identificou um número significativo de indivíduos negros africanos afetados pela COVID-19, porém assintomáticos, na região norte do país (54). Entre as suas hipóteses para explicar essa observação, estavam a faixa etária mais jovem dessa população e a presença de fatores genéticos que alterassem a suscetibilidade deles à doença (54). Entretanto, sabe-se que a população afro-americana sofreu com a severidade e letalidade da COVID-19 (55,56), não sendo possível explicar o fenômeno observado somente através de fatores genéticos relacionados ao sistema imunológico (54). Então, a última hipótese proposta fundamentou-se na possibilidade de que a imunidade anti-malária poderia exercer um papel relevante contra a COVID-19 nesses pacientes (54). As evidências levantadas apontam que em indivíduos infectados por malária existe a presença de anticorpos do tipo IgG contra o glicosilfosfatidilinositol (GPI), responsável por ancorar algumas proteínas de membrana do *Plasmodium spp.* (57,58). O GPI pode ativar leucócitos, induzir citocinas pró-inflamatórias e a expressão de moléculas de adesão via receptores do tipo Toll 2 e 4 (57). Esses anticorpos poderiam também reconhecer glicoproteínas de membrana presentes no SARS-CoV-2 e proteger os indivíduos, ou resultar em infecção leve (59).

Konozy e colaboradores (2022) revisaram mais de 50 artigos científicos para destacar as semelhanças entre infecções por malária e COVID-19 que poderiam influenciar a biologia desta última (49). Foi identificado que, dentre inúmeros fatores, a genética populacional parece desempenhar um papel significativo na dinâmica da COVID-19 (49). Isso é evidente quando análises recentes de triagem genômica da ECA2 e variantes associadas à malária identificaram seis genes que podem desempenhar um papel na incidência e gravidade de ambas as infecções (49). Além disso, semelhanças do quadro clínico e patológico entre as duas doenças causaram considerável confusão no diagnóstico e, posteriormente, interferiram com o levantamento preciso da incidência das enfermidades na África (49). Portanto, possíveis similaridades entre as doenças quanto ao quadro clínico, patológico, imunológico, e atribuição genética devem ser levados em consideração (49) ao analisar as duas infecções em conjunto.

### **Anticorpos**

Uma função central do sistema imunológico é desempenhada por imunoglobulinas (Igs) ou anticorpos (Abs) (*do inglês antibodies*) (60) que reconhecem e ligam-se ao antígeno resultando na ativação e/ou modulação da imunidade inata e adaptativa (61). Os anticorpos também são atualmente a classe mais importante de bioterapêuticos usados para tratar uma grande gama de doenças (62). As Igs são produzidas pelos linfócitos B, e são secretadas como anticorpos ou expressos na superfície celular como receptores de células B (BCR) (61). Os Abs típicos são compostos por duas cadeias pesadas (IgH) idênticas e duas cadeias leves (IgL) kappa (IgκL ou IgKL) ou lambda (IgλL ou IgLL) idênticas (61). Essas cadeias pareiam-se de forma a criar três domínios estruturais, dois Fabs (*do inglês fragment antigen binding*), responsáveis pela interação com o antígeno e um Fc (fragmento cristalizável), que designa a função do anticorpo (63). Esses domínios são ligados por uma região flexível de “dobradiça”, (64) e são mantidas unidas por pelo menos duas ligações dissulfeto intermoleculares (65); podendo ser facilmente clivados em fragmentos Fab e Fc por proteases (63). As

cadeias de IgH possuem cinco isotipos principais (IgM, IgA, IgG, IgD e IgE), com quatro subtipos de IgG (IgG1, IgG2, IgG3 e IgG4) e dois subtipos de IgA (IgA1 e IgA2) (66). IgAs, IgDs e IgGs possuem um domínio variável (V) e três domínios constantes (C) (67). IgEs e IgMs possuem um domínio variável e quatro constantes (67). Os isotipos IgA e IgM possuem uma cadeia J adicional, que permite a formação de dímeros e pentâmeros, respectivamente (67). O papel da cadeia pesada está principalmente relacionado ao reconhecimento do antígeno e a funções efetoras, enquanto a cadeia leve relaciona-se à interação com o antígeno e estabilidade estrutural (61,68).

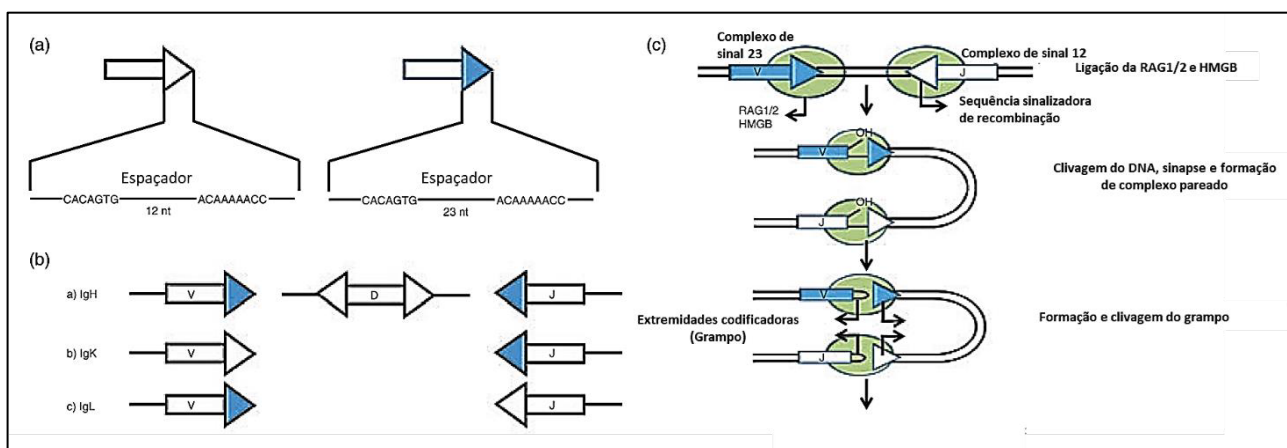
### **Recombinação V(D)J**

Para lidar com a enorme variedade de patógenos e reconhecer moléculas não próprias, as células B desenvolveram processos genéticos controlados em seus loci de imunoglobulina conhecidos coletivamente como “diversificação de imunoglobulina” (69). Os genomas dos vertebrados não são grandes o suficiente para codificar um sistema imunológico capaz de reconhecer todas as ameaças possíveis (70). A diversidade dos genes dos receptores de células B e T é gerada por rearranjos e mutações do ácido desoxirribonucleico (DNA, do inglês *deoxyribonucleic acid*), começando com a montagem desses genes a partir de conjuntos de partes intercambiáveis, um elegante mecanismo combinatório denominado “recombinação V(D)J” (70,71). Essa diversificação primária ocorre durante o desenvolvimento inicial das células B na medula óssea, por meio da montagem de um receptor de antígeno do tipo IgM (72) completo e exposto na superfície da célula B, o BCR (69). Os éxons que codificam as regiões variáveis (V), que de fato se ligam aos antígenos, são conhecidos como segmentos gênicos V (variável), D (diversidade) e J (junção). Esses segmentos V, D, J existem como múltiplas cópias no cromossomo (73), e o processo de recombinação V(D)J cliva e une esses segmentos não contíguos de imunoglobulina no genoma (71).

A recombinação V(D)J se inicia como uma etapa essencial na geração do sistema imune adaptativo de vertebrados (74). O complexo protéico codificado pelos genes ativadores de recombinação (RAG, do inglês *recombination-activating genes*) 1/2 (genes que codificam proteínas linfócito-específicas consistindo em RAG1 e RAG2) (66) catalisa duas reações sem dissociação: a clivagem e a formação de grampo no DNA (66) guiados por trechos curtos de DNA altamente conservados (75), chamados sequências de sinais de recombinação (RSS) (74). O RSS reside adjacente a cada éxon, presentes nos flancos dos segmentos dos genes IG, ou seja, *downstream* de V, *upstream* de J e em ambos os lados de D, governando o uso e, portanto, a seleção dos segmentos gênicos (75). O RSS consiste em um heptâmero palindrômico (CACAGTG) e um nonâmero rico em A/T (ACAAAACC) (76). Os três primeiros nucleotídeos do heptâmero são cruciais para a atividade de recombinação (76); este e o nonâmero são separados por uma sequência de DNA espaçador de 12 pares de base (pb) (12-RSS) ou 23 pb (23-RSS) (74,76). Geralmente, um espaçador 12-RSS recombina apenas com um espaçador 23-RSS e vice-versa, uma restrição denominada 'regra 12/23', o que evita rearranjos não produtivos (76). A clivagem do DNA mediado por RAG1/2 gera quebras de fitas duplas (DSBs, do inglês *double-strand breaks*) (77) precisamente na junção entre o segmento codificador e o RSS. Os RSSs que flanqueiam os segmentos individuais dos genes V, D e J, são distribuídos ao longo de cada locus do receptor de antígeno, e a formação de sinapses e clivagem podem ocorrer entre regiões

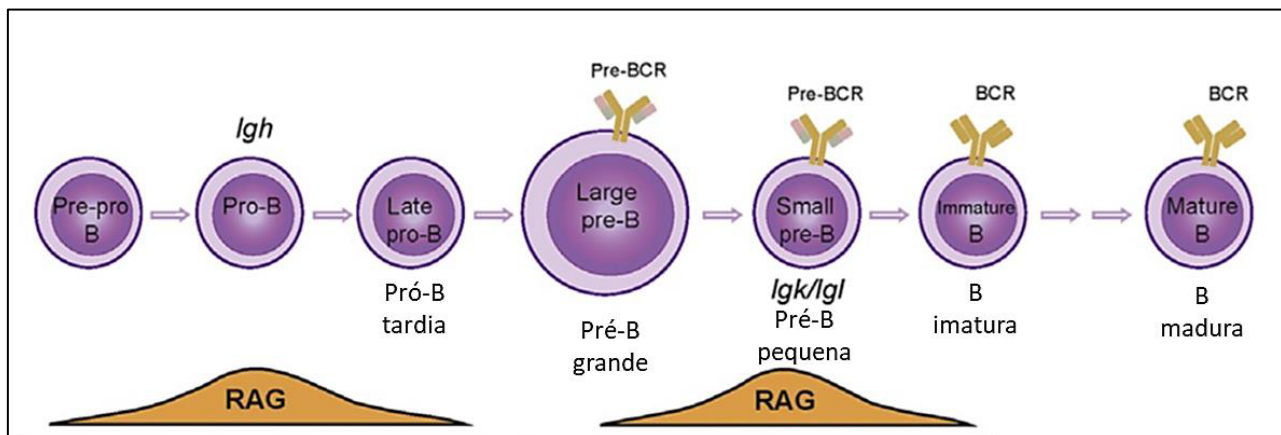
separadas por muitos DSBs, (Figura 7).

As quatro extremidades quebradas (duas extremidades codificadoras e os trechos codificadores contidos entre as RSSs em uso) são mantidas juntas em um complexo pós-clivagem (77). A 3'-hidroxila livre liberada na etapa de clivagem ataca a fita oposta por transesterificação na presença de  $Mg^{2+}$  (66), gerando um intermediário ligado covalentemente em forma de grampo, conhecido como extremidade de codificação em grampo (78). O complexo RAG pós-clivagem direciona o reparo através da junção de extremidade não homóloga (NHEJ, do inglês *non-homologous end joining*), muito importante para a manutenção da estabilidade do genoma (77). O rearranjo, ou recombinação V(D)J, é restrito pela linhagem celular: o rearranjo completo de segmentos dos genes IG ocorre apenas em células B, sendo ordenado por estágios dentro de uma determinada linhagem: a cadeia IGH é rearranjada no estágio pré-pró e pró de desenvolvimento de células B, antes do rearranjo das cadeias IGL ( $\kappa$  ou  $\lambda$ ) em células pré-B (77). Além disso, a recombinação DH-para-JH no locus IGH deve ocorrer em células pré-pró-B antes do rearranjo VH-para-DJH iniciar-se em células pró-B (77). A recombinação da cadeia leve ocorre em células pré-B após a conclusão bem-sucedida da recombinação da cadeia pesada (79), (Figura 8).



**Figura 7** Mecanismo da recombinação V(D)J. (a) Sequência de Sinal de Recombinação (RSS). Em 12-RSS (representado pelo triângulo aberto), o heptâmero está separado do nonâmero por 12 pares de base (pb), e em 23-RSS (representado por triângulo fechado) por 23 pb. (b) Distribuição de RSS nos loci dos receptores de antígeno em seres humanos e a regra 12/23. As regiões V, D e J dos segmentos de codificação são representadas como retângulos; o flanco 12-RSS é representado por um triângulo branco e 23-RSS por um triângulo azul. De acordo com a 'regra 12/23', a recombinação V(D)J é restrita a segmentos de genes que são flanqueados por sinais de diferentes comprimentos de espaçador. (c) Mecanismo de recombinação V(D)J. Modelo esquemático dos complexos proteína-DNA na recombinação V(D)J. O 12-RSS e o 23-RSS são representados como triângulos brancos e azuis, respectivamente. Segmentos de codificação são mostrados como retângulos e proteínas em formato oval sombreado. 'V' significa variável e 'J' para unir segmentos de genes. O complexo RAG 1/2 (RAGs, ovais verdes) se liga ao RSS e introduz um corte na fita simples do DNA, precisamente na fronteira entre o heptâmero do RSS e o segmento codificador. Essa clivagem gera um grupo 3'-OH livre na extremidade codificante que é então ligado covalentemente à ligação fosfodiéster oposta,

deixando uma ligação covalentemente fechada em forma de grampo. As extremidades permanecem associadas às RAGs constituindo uma estrutura transitória conhecida como “complexo pós-clivagem”. Em seguida, as extremidades codificantes são processadas e unidas pelo NHEJ para criar um éxon, que forma a região de ligação ao antígeno dos receptores de antígenos. D: diversidade, DNA: ácido desoxirribonucleico, HMGB: HMG-box Family, IgH: locus de cadeia pesada de imunoglobulina, Igk: imunoglobulina kappa, Igl: imunoglobulina lambda, J: junção, RAG: gene ativador de recombinação, V: variável. Figura modificada. Referência: LISTA DE FIGURAS.



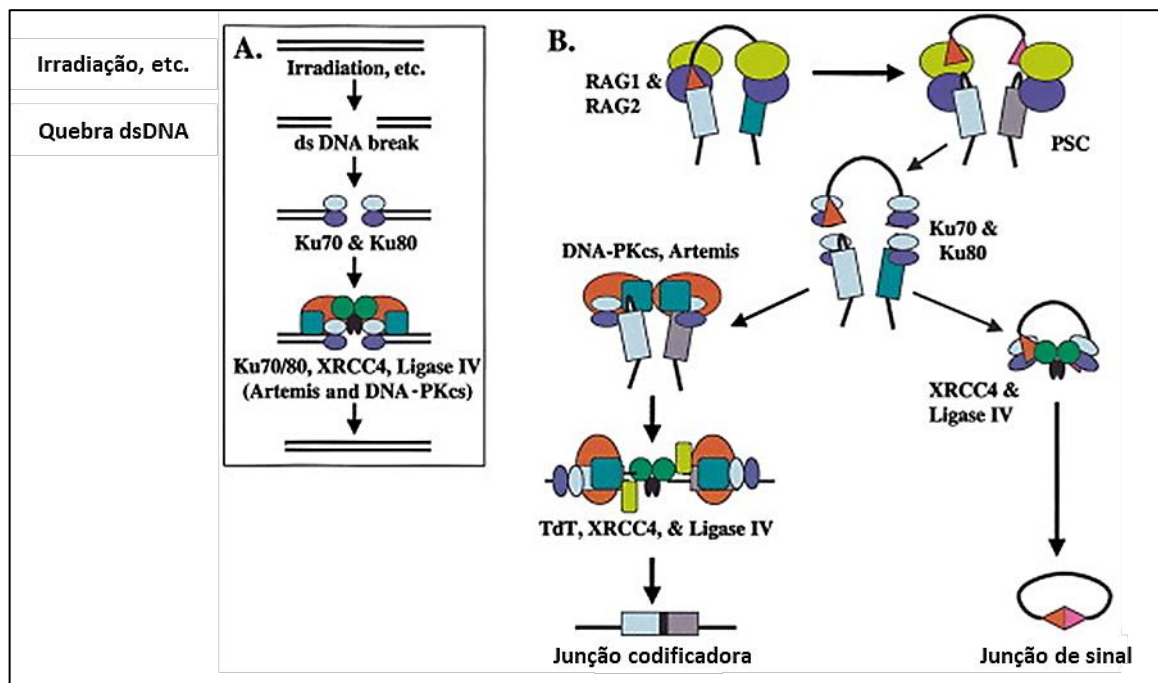
**Figura 8** Esquema mostrando os diferentes estágios de desenvolvimento das células B onde ocorre o rearranjo dos loci IG. BCR: receptores de células B, Igh: locus da cadeia pesada de imunoglobulina, Igk: locus da cadeia leve de imunoglobulina kappa, Igl: locus da cadeia leve de imunoglobulina lambda, RAG: gene ativador de recombinação. Figura modificada. Referência: LISTA DE FIGURAS.

Após a formação das quebras na dupla fita de DNA, os fatores participantes do NHEJ reabrem os grampos e unem as extremidades codificadoras (80). O NHEJ é a principal via para reparar quebras de fita dupla de DNA em células de mamíferos, e atua nos rearranjos das quebras de DNA introduzidas durante a recombinação V(D)J (81). Os componentes envolvidos nesta via são as proteínas Ku (Ku70 e Ku80, um heterodímero), a proteína quinase dependente de DNA subunidade catalítica (DNA-PKcs), a nuclease Artemis e o complexo de proteína de complementação cruzada de reparo de raios X 4 (XRCC4) -DNA ligase IV (80). Acredita-se que a Ku70/80 seja a primeira proteína a ligar-se ao DSB, com base em sua abundância e forte afinidade (66). A Ku70/80 protege as extremidades do DNA da digestão, e o complexo Ku:DNA também serve como um sítio no qual a nuclease, polimerases e ligase podem ser ligadas (66). Existe um complexo Ku:DNA em cada uma das duas extremidades que serão unidas, permitindo que cada extremidade do DNA seja modificada em preparação para suas junções (66). O complexo Ku:DNA recruta DNA-PKcs e Artemis, uma endonuclease ativada por DNA-PKcs, gerando a proteína quinase de DNA (DNA-PK) (82). O complexo DNA-PKcs sofre autofosforilação e ativa Artemis, que então ganha várias atividades de nuclease incluindo atividade de 5'-endonuclease, 3'-endonuclease, e atividade de abertura do grampo (66). A Artemis liga-se ao complexo DNA-PKcs e cliva as projeções 5' e 3' dos grampos gerados pelo complexo RAG1/2 (82). O complexo DNA-PK também tem um papel no recrutamento de outras proteínas da NHEJ que engloba ligase (Lig) IV,



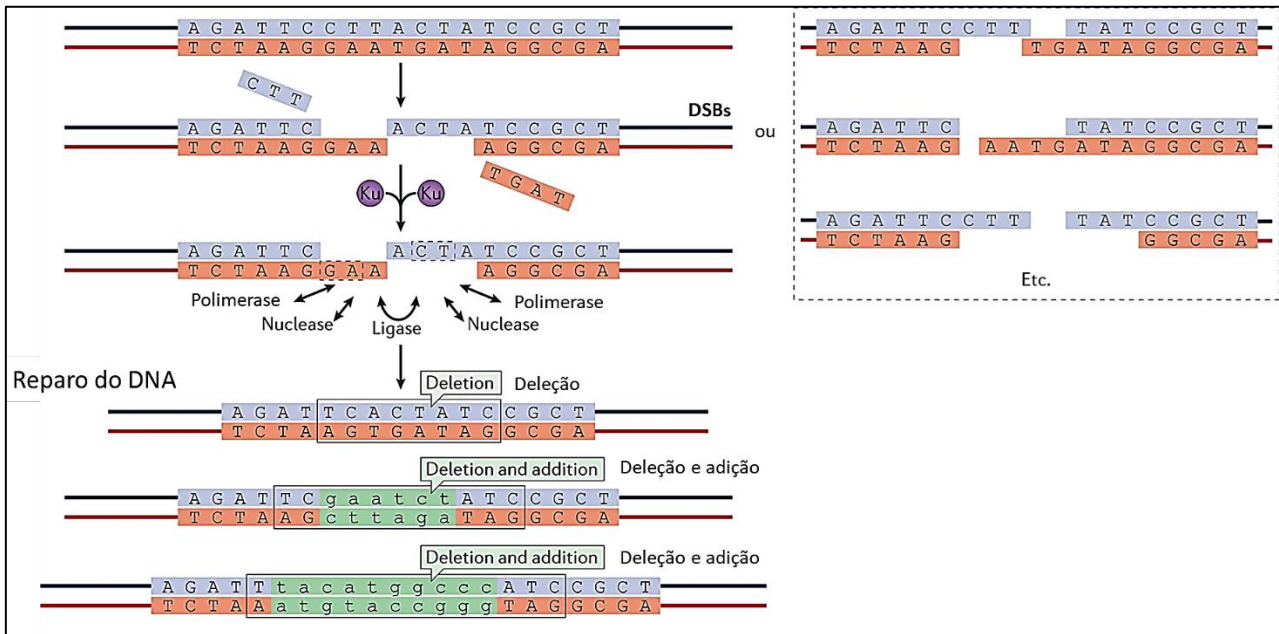
XRCC4 e XLF para as DSBs (66,82). O complexo Lig IV/XRCC4 liga as extremidades processadas com a ajuda de XLF (76), (Figura 9). O PAXX (parálogo de XRCC4 e XLF) é uma proteína recentemente descoberta com semelhança estrutural a XRCC4 e XLF e promove o reparo das DSBs interagindo com Ku (66).

A lacuna criada após o processamento é preenchida pela DNA polimerase  $\kappa$  recrutada por Ku (66). Nesta fase da junção dos segmentos gênicos V, D e J, várias modificações como inserções e deleções podem ocorrer nas junções, resultando em aumento adicional na diversidade do receptor de antígeno (76). A abertura do grampo nas extremidades codificadoras é assimétrica, devido à remoção de algumas bases da porção terminal resultar em uma fita de DNA mais longa do que a outra (76). A fita mais curta é estendida pela adição de nucleotídeos complementares à fita mais longa, resultando na inserção de nucleotídeos palindrômicos (P) nas junções codificadoras (83). A desoxinucleotidil transferase terminal (TdT) e a DNA Polimerase  $\mu$  diversificam ainda mais essas sequências juncionais catalisando a adição de nucleotídeos não modelados (N-nucleotídeos) nas extremidades codificadoras (84) (Figura 10). A diversidade juncional pode expandir a diversidade das cadeias em até  $10^{11}$  a partir das  $10^6$  anteriores oriundas da diversidade combinatória (76). Após a recombinação V(D)J dos genes IGH e IGL, as células B *naïves* resultantes transcrevem os genes IGH e IGL e o RNA transcrito sofre *splicing* alternativo (85). O *splicing* une cada gene IGH ou IGL a um gene constante (C), produzindo IgM pela fusão do éxon  $\mu$  ao éxon J, formando cadeias leves e pesadas completas (61,85), (Figura 11).

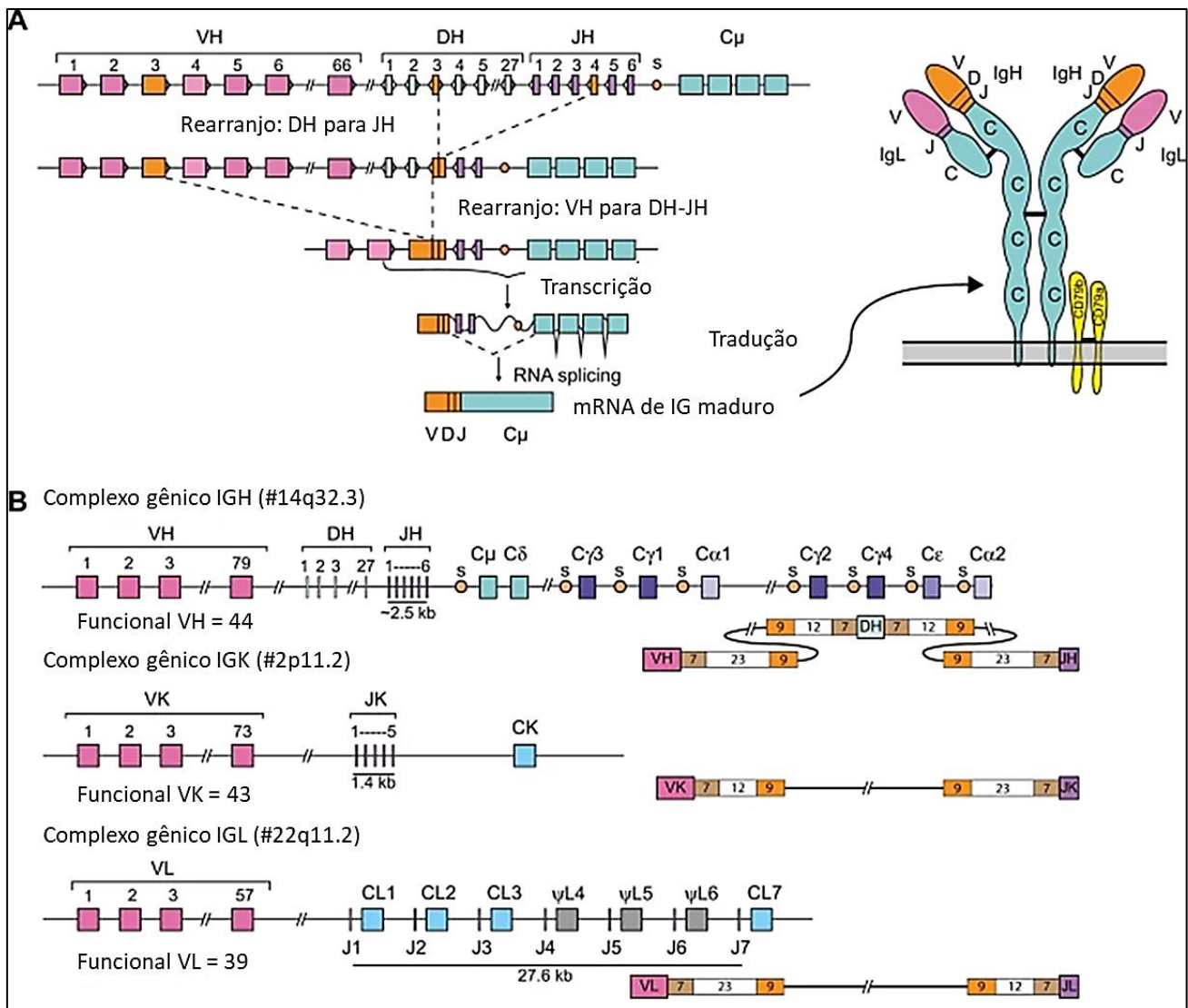


**Figura 9** Esquema com as proteínas NHEJ que reparam DSBs e unem as sequências codificadoras liberadas por RAG e as RSSs. (A) NHEJ pode reparar DSBs através das ações combinadas das proteínas Ku, DNA-PKcs, Artemis, XRCC4 e Ligase IV. Os parênteses indicam que diferentes tipos de células podem exibir dependência variável da Artemis e DNA-PKcs. (B) As proteínas Ku, XRCC4 e Lig IV são necessárias para a junção das sequências codificadoras (na figura está descrito como Junção codificadora) e RSS (na figura está descrito como Junção de sinal), enquanto DNA-PKcs e

Artemis são mais importantes para a junção das sequências codificadoras. As RAGs, no contexto do complexo de pós-clivagem, também são importantes para unir as extremidades codificadoras e os RSS terminais. DSBs: quebras de fita dupla de DNA (do Inglês Double-Strand Breaks), dsDNA: DNA dupla fita, Ku70/Ku80: proteínas Ku, DNA-PKcs: proteína quinase dependente da subunidade DNA catalítica, PSC: complexo de pós-clivagem, RAG: gene ativador de recombinação, XRCC4: proteína de complementação cruzada de reparo de raios X 4. Figura Modificada. Referência: LISTA DE FIGURAS.



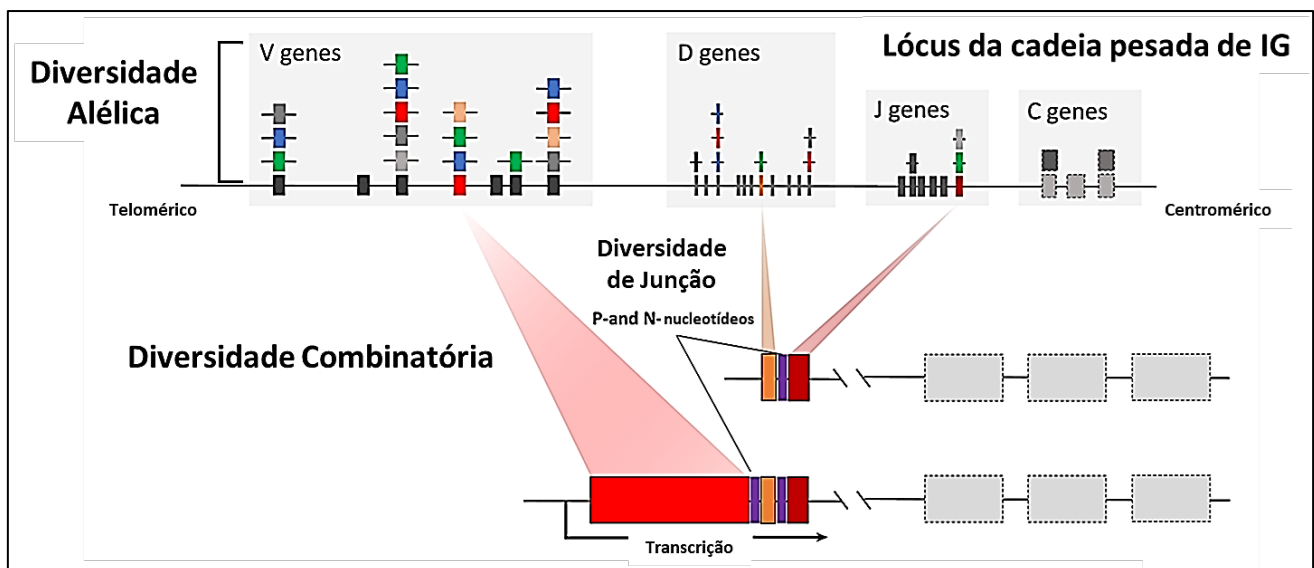
**Figura 10** Esquema de quebras de fita dupla de DNA (DSBs, do Inglês Double-Strand Breaks) e seu reparo por junção de extremidade não homóloga (NHEJ) (topo). O heterodímero Ku70-Ku80 se liga aos DSBs e melhora sua ligação subsequente pelos complexos NHEJ polimerase, nuclease e ligase. Essas enzimas podem atuar nos DSBs em qualquer ordem para retirar e adicionar nucleotídeos. Várias rodadas de remoção e adição são possíveis, e as atividades de nuclease e polimerase em cada uma das duas extremidades do DNA parecem ser independentes. A micro homologia entre as duas extremidades do DNA, que já está presente (caixas tracejadas) ou recém-criada quando as polimerases adicionam nucleotídeos de maneira independente do molde, é frequentemente usada para guiar a junção das extremidades. O processo é propenso a erros e pode resultar em diversas sequências de DNA na junção de reparo (abaixo). As adições de nucleotídeos são representadas em letras minúsculas verdes. Ku: proteína Ku. Figura Modificada. Referência: LISTA DE FIGURAS.



**Figura 11** Esquema demonstrando as etapas para a junção dos segmentos V(D)J no locus IGH. A) Na primeira etapa da recombinação V(D)J no locus IGH, um gene D é acoplado a um gene J. Posteriormente, um gene V é acoplado à junção DJ. O éxon VDJ é transcrito e unido ao éxon IGHM. Um processo análogo ocorre nos genes da cadeia leve de Ig. Quando uma proteína IgM funcional é transportada para a membrana plasmática com as moléculas âncoras CD79a e CD79b e se une a uma cadeia leve funcional da Ig, forma-se uma molécula completa de anticorpo. B) Visão geral esquemática dos três loci IG: IGH, IGK, IGL e a estrutura de suas Sequências Sinalizadoras de Recombinação (RSS) correspondentes. A posição genômica dos loci é indicada entre colchetes. Nos loci IG, cada retângulo representa um dos genes variável (V), diversidade (D), junção (J) e constante (C), e os círculos (atribuídos com “s”) representam regiões de troca. O número de genes funcionais conhecidos, conforme listado no IMGT, é indicado abaixo de cada esquema. Os esquemas de estrutura do RSS descrevem a posição dos heptâmeros (7), nonâmeros (9) e espaçadores (12/23) em relação aos genes V, D e J. Figura modificada. Referência: LISTA DE FIGURAS.

As imunoglobulinas neutralizam diretamente os patógenos e seus produtos (86). Além disso, recrutam efetores da imunidade celular para eliminar patógenos e células tumorais (82). Os anticorpos

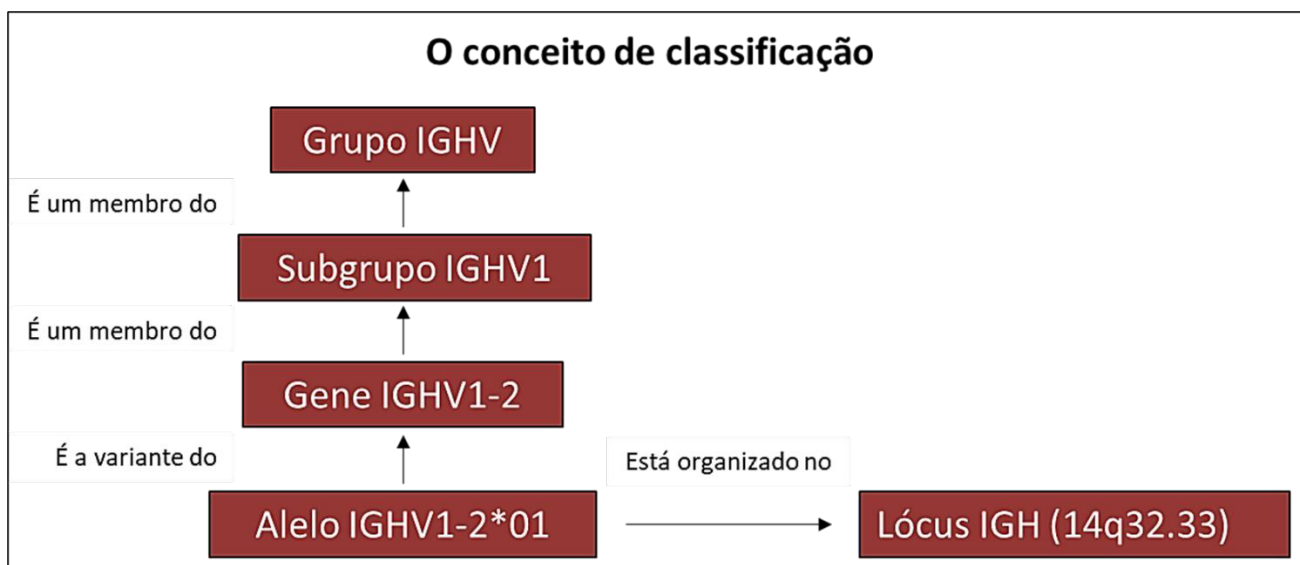
neutralizantes secretados possuem importância central para a proteção imunológica do hospedeiro (87), pois são capazes de reconhecer quase todos os tipos de antígenos devido à extrema diversidade de suas porções variáveis de ligação a antígenos (60,66). Cada domínio variável pode ser dividido em três regiões de variabilidade de sequência, denominadas regiões determinantes de complementaridade (CDRs, do inglês *complementarity determining regions*), e quatro regiões de sequência relativamente constante, denominadas regiões estruturais ou *frameworks*. Os *frameworks* de ambas as cadeias são predominantemente folhas  $\beta$  pregueadas e se unem para dar suporte às alças hipervariáveis das CDRs, que interagem entre si para conferir especificidade antigênica (88). As três regiões CDR da cadeia H são emparelhadas com as três regiões CDRs da cadeia L para formar o sítio de ligação do antígeno (89). Em comparação a outras regiões CDR, o comprimento variado e as propriedades bioquímicas do CDR 3 da cadeia pesada (HCDR3) contribuem para aumentar a diversidade das imunoglobulinas (90). O HCDR3 é a região de maior diversidade de sequências e é formada pela junção dos segmentos germinativos VH, DH e JH da cadeia pesada combinados com excisões aleatórias e adições de novos nucleotídeos entre esses segmentos (79). Em humanos, a sequência V(D)J completa e o HCDR3 incorporado são derivados da montagem aleatória sequencial de 1 em 56 segmentos gênicos VH, 1 em 23 DH e 1 em 6 JH disponíveis (90). Foi estimado que a diversidade teórica do HCDR3 excede  $10^{15}$  variantes, geradas a partir de sequências genômicas obtidas pelos mecanismos de diversidade combinatória e de junção (91), discutidos anteriormente. Isso fundamenta a imensa diversidade do repertório de anticorpos humanos (Figura 12).



**Figura 12** Uma visão geral dos principais elementos que contribuem para a diversidade do repertório de células B naïve. Um esquema simples do locus IGH da linhagem germinativa é mostrado (sem escala), consistindo em grupos de segmentos gênicos IGH V, D, J e C dispostos em sequência. Para um subconjunto desses segmentos, vários alelos são mostrados, representando a "diversidade alélica" a nível da população. Durante a formação inicial do repertório naïve, um único segmento gênico V, D e J em uma determinada célula B são somaticamente recombinados; em cada uma dessas etapas, os nucleotídeos P e N são adicionados nas junções D–J e V–D ("diversidade juncional"),

respectivamente. Este processo, conhecido como rearranjo  $V(D)J$ , é a base para a “diversidade combinatória”. Os segmentos recombinados  $V$  (vermelho),  $D$  (laranja) e  $J$  (marrom) serão então transcritos e, após o splicing, serão pareados com um éxon constante ( $C$ ) (cinza). O processo de recombinação somática também ocorre em um dos dois loci que codificam os segmentos gênicos da cadeia leve  $Ab$  [ $IG\kappa$  e  $IG\lambda$ ; porém, este envolve apenas os segmentos gênicos  $V$  (amarelo),  $J$  (marrom) e  $C$  (cinza claro)]. Duas cadeias pesadas idênticas e duas cadeias leves idênticas são finalmente emparelhadas e unidas através de pontes dissulfeto para formar um  $Ab$  funcional; assim, uma diversidade adicional no repertório  $Ab$  expresso vem do “emparelhamento de cadeia pesada e leve”. Juntos, os segmentos  $V$ ,  $D$  e  $J$  representados compreendem o domínio variável da cadeia pesada de um anticorpo funcional e, juntamente com o domínio variável da cadeia leve, codificado pelos segmentos  $V$  e  $J$ , são responsáveis pela ligação ao antígeno. Os domínios  $C$  das cadeias pesada e leve fornecem funções estruturais e/ou efetoras ao  $Ab$ .  $Ab$ : anticorpo,  $C$ : constante,  $D$ : diversidade,  $IGH$ : locus de cadeia pesada de imunoglobulina;  $IG\kappa$ : imunoglobulina kappa,  $IG\lambda$ : imunoglobulina lambda,  $J$ : junção,  $V$ : variável. Figura modificada. Referência: LISTA DE FIGURAS.

Os genes que codificam os anticorpos pertencem a famílias multigênicas altamente polimórficas (92). Os loci de imunoglobulina abrigam grandes famílias expandidas de segmentos de genes filogeneticamente relacionados, que se recombinam para codificar BCRs e anticorpos expressos por células B (93). Uma grande contribuição do *Immunogenetics Information System (IMGT)-Ontology* (94,95) foi estabelecer os princípios de sua classificação e propor uma nomenclatura padronizada baseada nos conceitos de “grupo”, “subgrupo”, “gene” e “alelo” (92) para as imunoglobulinas. O conceito de “grupo” classifica, para os anticorpos, um conjunto de genes do mesmo genótipo: variável ( $V$ ), diversidade ( $D$ ), junção ( $J$ ) ou constante ( $C$ ) (92). O conceito de “subgrupo” classifica um subconjunto de genes que pertencem a um mesmo grupo e que, em uma determinada espécie, compartilham pelo menos 75% de identidade no nível da sequência de nucleotídeos (e na configuração germinativa para os genes  $V$ ,  $D$  e  $J$ ) (92). O conceito de “gene” classifica uma unidade de sequência de DNA que pode ser potencialmente transcrita e/ou traduzida (nessa definição incluem os elementos reguladores em 5' e 3' e os íntrons, se presentes) (92). O nome de um gene é composto pelo nome da espécie e pelo símbolo do gene, por exemplo, *Homo sapiens* IGLV1-2 (92). O conceito de “alelo” classifica uma variante polimórfica de um gene (92). Os alelos identificados pelas mutações na sequência de nucleotídeos são classificados por referência ao alelo \*01 (92), (Figura 13).



**Figura 13** O “Conceito de classificação” organiza o conhecimento imunogenético útil para nomear e classificar os genes das imunoglobulinas. Figura modificada. Referência: LISTA DE FIGURAS.

Em humanos, os genes que codificam os BCRs e anticorpos estão localizados em loci de IGS encontrados em três cromossomos (chr) (96). O lócus da cadeia pesada, IGH está localizado no chr14 (região 14q32.33) (76), e abrange 1.250 kb deste cromossomo (97). O lócus IGH possui de 123 a 129 genes IGHV dependendo dos haplótipos (conjunto de variantes alélicas ao longo de um lócus IG que são co-herdados), 27 segmentos IGHD, pertencente a sete subgrupos (98), nove segmentos IGJ, e 11 segmentos gênicos IGHC (76). Oitenta e dois a 89 segmentos IGHV pertencem a sete subgrupos, enquanto 43 pseudogenes, que são muito divergentes para serem atribuídos a subgrupos, foram atribuídos a clãs (98). O lócus IGκ está localizado no chr2 (região 2p11.2) (76) e abrange 1.820 kb deste cromossomo (99). A cadeia IGκ consiste em 76 genes IGκV, pertencentes a sete subgrupos (98), cinco segmentos IGκJ e um segmento IGκC exclusivo (76). O lócus IGλ está localizado no chr22 (região 22q11.2) (76) e abrange ~ 920 kb deste cromossomo (61). A cadeia IGλ consiste em setenta e três a 74 segmentos IGλV, pertencentes a onze subgrupos (98), sete a 11 IGλJ e sete a 11 genes IGλC, dependendo dos haplótipos, e cada gene IGλC sendo precedido por um segmento IGλJ (76,96,98).

#### **A importância do estudo das sequências germinativas de imunoglobulinas**

A configuração do repertório do BCR após o encontro com o antígeno é orquestrada por células T (100). A interação das células B com as células T normalmente ocorre em estruturas linfóides organizadas, os centros germinativos, localizados por exemplo no baço e linfonodos (101). A comunicação cruzada entre células T e B afeta ambos os tipos de células; tanto as células T quanto as células B passarão por expansão clonal e algumas dessas células persistirão como células de memória. Um subconjunto de células B pós-centro germinativo se diferenciará em células efetoras (plasmablastos e células plasmáticas) que secretam anticorpos solúveis (101). Nos centros germinativos, os clones de células B com BCRs de maior afinidade serão selecionados em relação aos de menor afinidade, em um processo denominado maturação de afinidade. (102) Desta forma, a capacidade de um BCR de se ligar ao antígeno na configuração germinativa é essencial para o

desenvolvimento de uma resposta de células B, sendo altamente influenciado pela variação genética que afeta a sequência do seu receptor (102).

Os loci de imunoglobulina abrigam variantes alélicas dos genes V(D)J dos loci IGH, IG $\kappa$  e IG $\lambda$  que, juntos, formam a base genética dos receptores altamente diversos de células B específicos para antígenos (103). Essas variantes alélicas podem ser compartilhadas ou específicas em populações humanas (75). Estudos em gêmeos demonstraram que as características do repertório de anticorpos são hereditárias, e as variantes da linhagem germinativa IG impactam diretamente o uso de anticorpos e a especificidade ao antígeno (75,93). Ainda, diferentes segmentos germinativos de IG demonstram envolvimento em diferentes respostas e eficácia contra infecções em indivíduos (93), e susceptibilidade a doenças (104). Por exemplo, é conhecida a associação entre o alelo IGHV4-61\*02 e um alto risco de doença cardíaca reumática na região da Oceania (105,106). Ainda, a variação genética nos genes de IG provavelmente resulta da adaptação a patógenos (75). Como exemplo, foi identificado a importância do segmento do IGHV1-2\*02 ou IGHV1-2\*04 para a geração de anticorpos neutralizantes da classe VRC01 (anticorpo monoclonal do tipo IgG1) direcionados ao sítio de ligação CD4 do HIV-1 (102) e a propensão de alelos específicos do segmento IGHV1-69, variáveis em diferentes etnias, serem utilizados em anticorpos neutralizantes direcionados ao antígeno hemaglutinina (glicoproteína de superfície) do vírus da influenza (107). Recentemente, também foi identificado anticorpos neutralizantes contra o SARS-CoV-2 que dependem criticamente do uso de um alelo específico (107). Três anticorpos similares e isolados de diferentes pacientes convalescentes foram eficazes em neutralizar o vírus ao utilizar o alelo germinativo IGHV1-69\*20, mas foi perdida ao alterar a região V utilizando outros alelos do segmento IGHV1-69 (107). Portanto, é de suma importância identificar os elementos que contribuem para as variações nas respostas das imunoglobulinas, a fim de aprofundar nossa compreensão das doenças que envolvem componentes imunológicos e aprimorar a predição para essas enfermidades (103).

### **OBJETIVO GERAL**

Levando em consideração os recentes avanços nos conjuntos de dados de variação genômica de um número crescente de populações humanas (108–110), este projeto exploratório tem como objetivo investigar se a incidência de *P. falciparum* em quatro populações humanas geograficamente distintas resultou em diferenças ou similaridades dos segmentos gênicos que codificam os CDRs dos anticorpos neutralizantes contra o *P. falciparum* e o SARS-CoV-2.

### **Objetivos específicos**

**i.** Selecionar na literatura anticorpos com estrutura cristalográfica definida e responsáveis pela neutralização dos patógenos *P. falciparum* ou SARS-CoV-2.

**ii.** Obter a sequência germinativa dos anticorpos mencionados no objetivo **i**.

**iii.** Acessar os genomas completos de interesse a partir de bases de dados públicas.

**iv.** Identificar a presença dos segmentos gênicos V(D)J responsáveis pela estrutura dos anticorpos neutralizantes de interesse nos genomas mencionados no objetivo **iii**.

## MATERIAIS E MÉTODOS

Este projeto de pesquisa foi analisado pela CEUA (Comissão de Ética no Uso de Animais) e CEPESH-ICB (Comissão de Ética em Pesquisa com Seres Humanos) e foi dispensado de aprovação ética (protocolo nº 1248/2022).

### **Seleção de anticorpos neutralizantes e identificação dos segmentos V(D)J**

Foram selecionados manualmente da literatura sete anticorpos neutralizantes amplamente estudados contra *P. falciparum* e SARS-CoV-2 (Tabelas 1 e 2). Os quatro anticorpos selecionados contra o SARS-CoV-2 são específicos para o mesmo antígeno, RBD. Embora o antígeno seja o mesmo, os anticorpos que o reconhecem são diferentes. Portanto, para fins de análise, identificamos os antígenos como RBD 1, 2, 3 e 4 (Tabela 1). Um critério para inclusão das sequências selecionadas foi a disponibilidade da estrutura cristalográfica da imunoglobulina madura. As sequências proteicas de cada anticorpo foram obtidas no *National Center for Biotechnology Information* (NCBI). Em seguida, obtivemos as sequências de nucleotídeos dos sete anticorpos a partir de suas sequências de aminoácidos através da ferramenta IMGT/tBLASTn. Por fim, através do IMGT/V-QUEST, identificamos as sequências da linhagem germinativa V(D)J que codificam cada anticorpo, alinhando-os com as sequências humanas presentes no IMGT *reference directory*, que compreende as sequências de referência de todos os genes e alelos funcionais (F), todas as fases abertas de leitura (ORF, do inglês *Open Reading Frame*) e todos os pseudogenes (P). Esse é um banco de dados abrangente do genoma para genes de imunoglobulina (IG) e receptor de células T (TR) de humanos e camundongos (95,111). Apenas as sequências humanas V(D)J que possuíam identidade igual ou acima de 75% com as sequências germinativas dos anticorpos alvo foram incluídas na análise final (Figura 14). Por fim, também foi realizada a análise da presença do gene da insulina em todos os genomas selecionados.



## 1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

		Score	Identity
<a href="#">HM855336</a>	Homsap IGHV3-48*04 F	837	76.74% (221/288 nt)
<a href="#">M99675</a>	Homsap IGHV3-48*01 F	828	76.39% (220/288 nt)
<a href="#">AJ879484</a>	Homsap IGHV3-69-1*01 P	822	76.49% (218/285 nt)
<a href="#">M99652</a>	Homsap IGHV3-11*01 F	819	76.04% (219/288 nt)
<a href="#">HM855688</a>	Homsap IGHV3-21*04 F	819	76.04% (219/288 nt)

## 2. Alignment for D-GENE and allele identification

Closest D-REGIONS

		Score	Identity
<a href="#">X93616</a>	Homsap IGHD3-22*01 F	68	71.43% (20/28 nt)
<a href="#">X13972</a>	Homsap IGHD6-13*01 F	42	66.67% (14/21 nt)
<a href="#">X97051</a>	Homsap IGHD6-25*01 F	36	66.67% (12/18 nt)
<a href="#">X13972</a>	Homsap IGHD3-10*01 F	32	57.14% (16/28 nt)
<a href="#">X13972</a>	Homsap IGHD3-3*01 F	32	57.14% (16/28 nt)

## 3. Alignment for J-GENE and allele identification

Closest J-REGIONS

		Score	Identity
<a href="#">X86355</a>	Homsap IGHJ4*02 F	132	75.00% (36/48 nt)
<a href="#">X86355</a>	Homsap IGHJ5*02 F	129	72.55% (37/51 nt)
<a href="#">J00256</a>	Homsap IGHJ4*01 F	123	72.92% (35/48 nt)
<a href="#">J00256</a>	Homsap IGHJ2*01 F	121	69.81% (37/53 nt)
<a href="#">J00256</a>	Homsap IGHJ5*01 F	120	70.59% (36/51 nt)

**Figura 14** Exemplo de identificação das sequências V(D)J. As sequências destacadas em vermelho possuem similaridade igual ou acima de 75% com as sequências do anticorpo 6PH6\_A (cadeia pesada do anticorpo clone 2544 contra o antígeno Pfs25). Foi utilizada a ferramenta IMGT/V-QUEST para alinhar as sequências dos sete anticorpos neutralizantes às sequências humanas de referência presentes no IMGT reference directory. Referência: LISTA DE FIGURAS.

PATÓGENO	ANTÍGENO	ID ANTÍGENO	RELEVÂNCIA FISIOLÓGICA	PDB	CLONE
<i>P. falciparum</i>	PfNANP	PfNANP	Antígeno de superfície do esporozoítio de <i>P. falciparum</i> (28).	5BK0_A 5BK0_B (28).	663
<i>P. falciparum</i>	PfS25	PfS25	Maturação do oocisto e invasão do epitélio intestinal do mosquito (21).	6PHC_B 6PHC_A (21).	2544
<i>P. falciparum</i>	PfRH5	PfRH5	Formação do complexo de invasão (18).	6RCU_D 6RCU_E (18).	R5.016
SARS-CoV-2	RBD	1	Invasão da célula-alvo (112).	6W41_L 6W41_H (112).	CR3022
SARS-CoV-2	RBD	2	Invasão da célula-alvo.	7M7W_B 7M7W_A (113).	S2X259
SARS-CoV-2	RBD	3	Invasão da célula-alvo.	7M7W_D 7M7W_C (113).	S2H97
SARS-CoV-2	RBD	4	Invasão da célula-alvo.	7R6X_D 7R6X_C (113).	S2E12

**Tabela 1** Resumo dos antígenos, sua identificação e relevância fisiológica para os patógenos *P. falciparum* e SARS-CoV-2. Estão também presentes os códigos de acesso PDB e identificação dos clones para os anticorpos neutralizantes selecionados. Referência: *LISTA DE FIGURAS*.







## **Seleção dos genomas humanos**

Quarenta e oito genomas completos com sequenciamento de alta cobertura ou WGS (do inglês *Whole Genome Sequencing*), ou seja, cada um com mais de um bilhão de pares de bases foram selecionados da base de dados *1000 Genomes* (116) e provenientes de dois estudos: *Human Genome Diversity Project (HGDP)* (117) e *Simons Genome Diversity Project (SGDP)* (110). As células utilizadas para os sequenciamentos do estudo HGDP foram linhagens celulares linfoblásticas (LCLs, do inglês *lymphoblastoid B-cell lines*) (118). Para o estudo SGDP, não foram encontradas informações relacionadas ao tipo celular utilizado para a obtenção de material genético. De acordo com Khatri e colaboradores (2021), ~75% dos genomas (1.941 genomas) oriundos do *1000 Genomes* são derivados de culturas de células B transformadas pelo vírus Epstein-Barr (EBV, do inglês *Epstein-Barr virus*), enquanto os outros ~25% dos genomas são oriundos das células sanguíneas, bem como de fontes não declaradas (75). Entretanto, esse mesmo grupo testou todos os genomas presentes no *1000 Genomes* para as novas variantes de genes encontradas e não detectaram influências por mutações somáticas relacionadas à transfecção de EBV em células B maduras pós-centro germinativo (75).

Os genomas selecionados foram divididos em grupos que diferiram em sua coevolução com o *P. falciparum*; indivíduos do Círculo Polar Ártico (compostos pelos grupos étnicos Adygei, Chukchi, Eskimo e Mansi da Rússia e Saami da Finlândia; n = 16), que teoricamente nunca tiveram contato com *P. falciparum*, foram comparados com Africanos Subsaarianos (compostos pelos grupos étnicos Yoruba da Nigéria, Luhya e Luo do Quênia e Mandenka do Senegal; n = 16) que coevoluíram com o *P. falciparum* por milhares de anos. Também selecionamos um grupo de Sul-Africanos, o grupo Migrante (composto pelo grupo étnico Bantu da Botswana, Namíbia e África do Sul; n = 09) que, há cerca de 3.000 a 5.000 anos, emigrou de áreas endêmicas do *P. falciparum* na África Ocidental para áreas livres de malária na África Austral (119). Nosso quarto grupo, o grupo Antigo, é composto por sete genomas extraídos de fósseis ósseos de Neandertais (composto por espécimes antigos da Croácia, Alemanha, Espanha e Rússia; n = 7) (120–124), que viveram entre 130.000 e 40.000 anos atrás (121,125) e não sofreram pressão seletiva pelo *P. falciparum* por serem habitantes do território europeu, onde não há evidências da presença de *P. falciparum* naquela época.

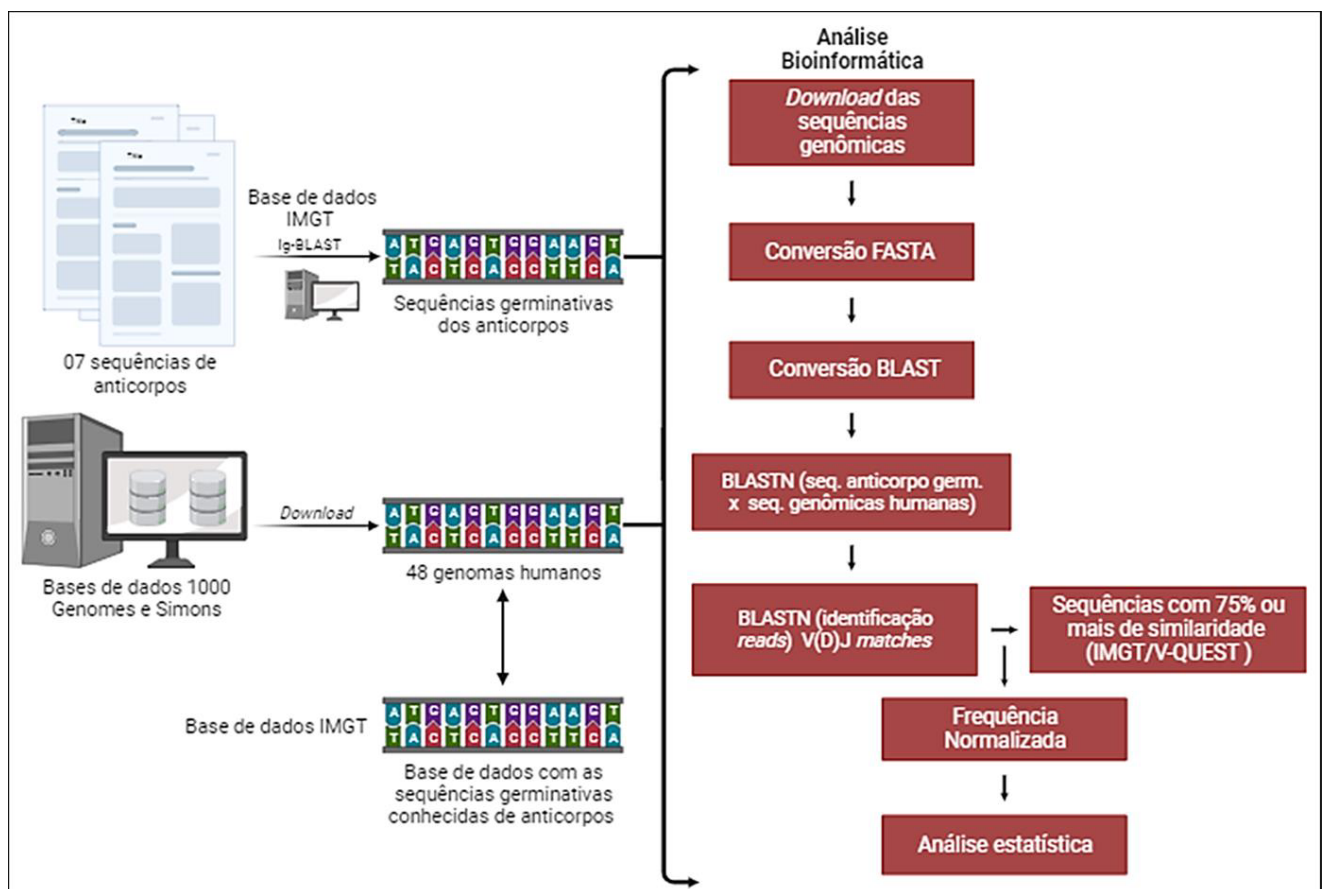
## **Análise dos genomas humanos**

Usando a linha de comando bash para Linux, foi realizado o download das sequências dos genomas selecionados e convertidas para o formato FASTA, que representa as sequências de nucleotídeos. Após a conversão, as sequências foram transformadas para o formato do banco de dados BLAST (do inglês *The Basic Local Alignment Search Tool*), um banco de dados que utiliza um algoritmo de busca permitindo comparar e encontrar regiões de similaridade entre um DNA desconhecido com outras sequências conhecidas (126). O BLAST possui uma ferramenta, BLASTN, específica para procura de sequências de nucleotídeos em uma base de dados, e o próximo passo foi o alinhamento das sequências de interesse. Então, as sete sequências germinativas dos anticorpos previamente selecionados (sequências de cadeia pesada e leve) foram alinhadas com todas as sequências de nucleotídeos dos genomas, enviando como resultado apenas as sequências de nucleotídeos V(D)J com mais de 80% de similaridade. Um segundo BLASTN foi realizado para

alinhar as sequências V(D)J encontradas com as sequências da linhagem germinativa dos anticorpos do banco de dados IMGT GENE/DB (Figura 15). Esta etapa foi realizada para se certificar que as sequências V(D)J finais com mais de 80% de similaridade seriam as que possuem maior semelhança com as sequências *germinativas* presentes no IMGT GENE/DB. As sequências finais analisadas foram selecionadas por comparação com as sequências humanas V(D)J que mostraram identidade igual ou acima de 75% com as sequências *germinativas* dos anticorpos (seção Seleção de anticorpos neutralizantes e identificação dos segmentos V(D)J). A frequência normalizada para cada segmento V(D)J foi calculada de acordo com a fórmula:

$$\text{Frequência Normalizada} = (\text{Matches}/\text{Reads}) \times 10^9$$

A contagem de *matches* representa o número de ocorrências de uma sequência V(D)J específica em um genoma analisado. Já o termo *read* ou "leitura" se refere a um fragmento de sequência de DNA, ou seja, uma pequena porção do DNA. No sequenciamento WGS, os genomas são fragmentados em *reads*, e o número total de *reads* indica se o DNA foi sequenciado com uma alta cobertura.



**Figura 15** Ilustração representativa dos passos realizados durante a análise bioinformática descrita. Referência: LISTA DE FIGURAS.

### Análise estatística

Todas as análises estatísticas foram realizadas utilizando o software *Graphpad Prism* (*Graphpad Software Inc.*), 5ª versão. A fim de harmonizar a distribuição dos valores, a frequência normalizada de

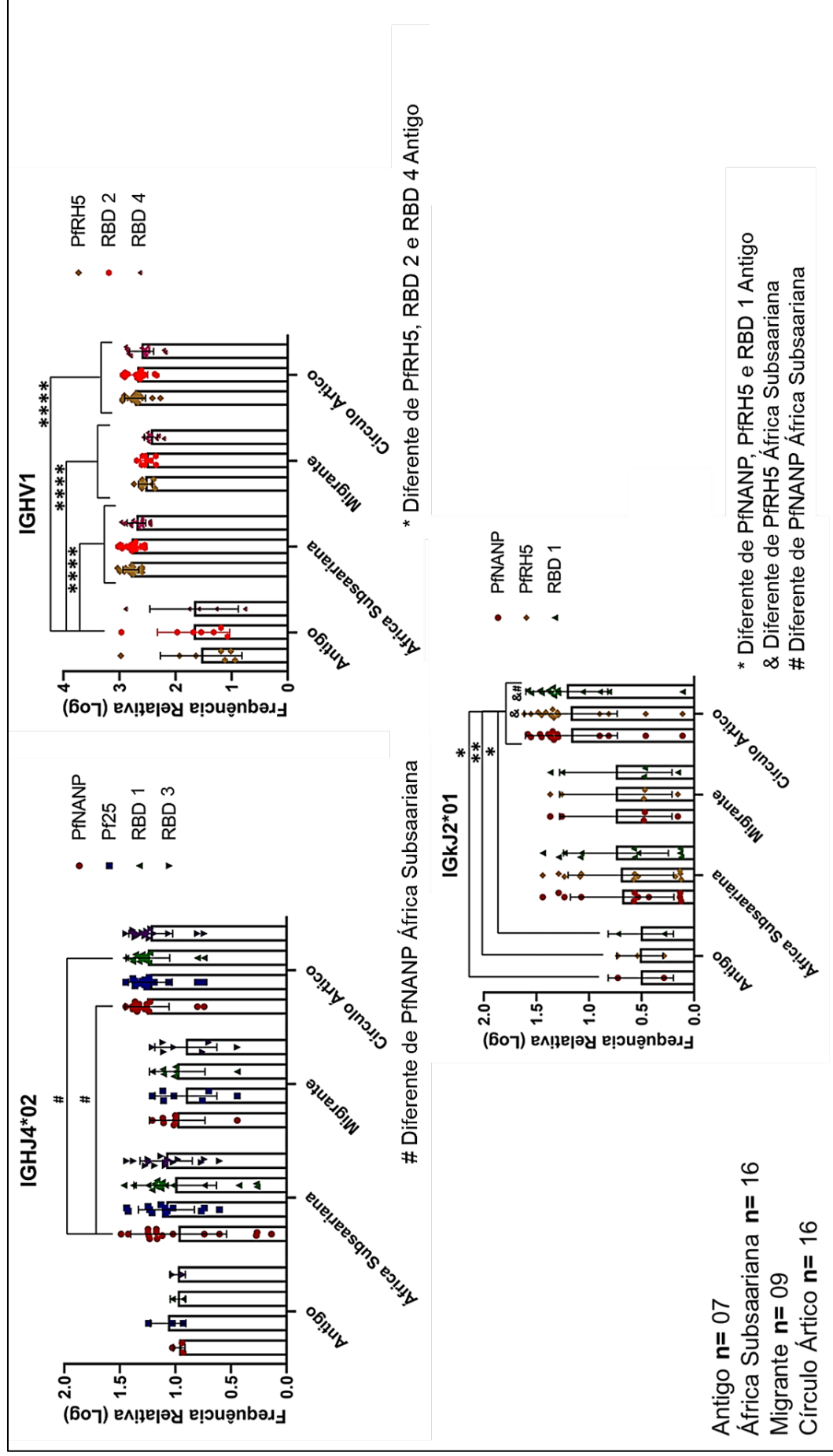
cada segmento gênico foi transformada por  $Y = \text{Log}(Y)$ . Para analisar os dados, foram utilizados ANOVA *one-way* e *two-way*, bem como o teste de *Tukey*, considerando as variáveis de grupo e grupo/alelo, respectivamente. O valor de  $p$  foi considerado estatisticamente significativo quando menor que 0,05.

## RESULTADOS

**Segmentos gênicos V(D)J específicos, que codificam os CDRs das cadeias leves e pesadas dos anticorpos neutralizantes direcionados aos antígenos do *P. falciparum* e SARS-CoV-2, apresentaram diferença na frequência normalizada no grupo Círculo Ártico quando comparado ao grupo África Subsaariana e distribuição mais ampla nos grupos África Subsaariana e Migrante.**

A Figura 16 apresenta a frequência normalizada logarítmica ( $\log$ ) e o desvio padrão ( $\sigma$ ) dos segmentos V(D)J responsáveis pelas regiões CDR das cadeias leves (IGk ou IGL) e pesadas (IGH) dos anticorpos neutralizantes específicos para os antígenos do *P. falciparum* (PfNANP, Pf25 e PfRH5) e do SARS-CoV-2 (RBD 1, RBD 2, RBD 3 e RBD 4), presentes nos genomas dos grupos Antigo, África Subsaariana, Migrante e Círculo Ártico.





**Figura 16** Gráficos de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) dos genes *V(D)J* que codificam as cadeias pesadas (*IGH*) e leve (*IGk* ou *IGL*) de anticorpos neutralizantes específicos para os antígenos do *P. falciparum* e SARS-CoV-2 em diferentes grupos. Antígenos alvo: P1NANP, P125, P1RH5 e SARS-CoV-2 RBD (identificados como 1, 2, 3 e 4). n = 48 indivíduos. Pf: *P. falciparum*. Referência: LISTA DE FIGURAS.

A sequência IGHJ4\*02 (Figura 16) apresentou o maior número de ocorrências ou *matches* nos genomas do grupo Círculo Ártico quando comparado aos genomas do grupo África Subsaariana, apresentando variação estatisticamente significativa da frequência normalizada quando esta sequência codifica os CDRs dos anticorpos específicos para o antígeno PfNANP ( $p = 0,0358$ ), demonstrando diferença no uso deste segmento gênico entre populações diferentemente expostas ao *P. falciparum*. Também foi observado diferença estatística significativa para o segmento IGHJ4\*02 quando utilizado para codificar os CDRs dos anticorpos direcionados ao antígeno RBD 1 no grupo Círculo Ártico em relação ao antígeno PfNANP no grupo África Subsaariana ( $p = 0,047$ ). Entretanto, o grupo África Subsaariana apresentou o maior desvio padrão para a sequência IGHJ4\*02 para os antígenos PfNANP e RBD 1 ( $\sigma = 0,431$  e  $0,372$ , respectivamente) vs. PfNANP (Antigo  $\sigma = 0,053$ ; Migrante  $\sigma = 0,250$  e Círculo Ártico  $\sigma = 0,198$ ) e RBD 1 (Antigo  $\sigma = 0,068$ ; Migrante  $\sigma = 0,250$  e Círculo Ártico  $\sigma = 0,195$ ), e o grupo Migrante apresentou o maior desvio padrão desta sequência para os antígenos Pf25 ( $\sigma = 0,278$ ) vs. Pf25 (Antigo  $\sigma = 0,160$ ; África Subsaariana  $\sigma = 0,252$  e Círculo Ártico  $\sigma = 0,199$ ) e RBD 3 ( $\sigma = 0,278$ ) vs. (Antigo  $\sigma = 0,068$ ; África Subsaariana  $\sigma = 0,235$  e Círculo Ártico  $\sigma = 0,198$ ). Esses resultados sugerem uma distribuição mais ampla do segmento IGHJ4\*02 nos genomas de origem africana.

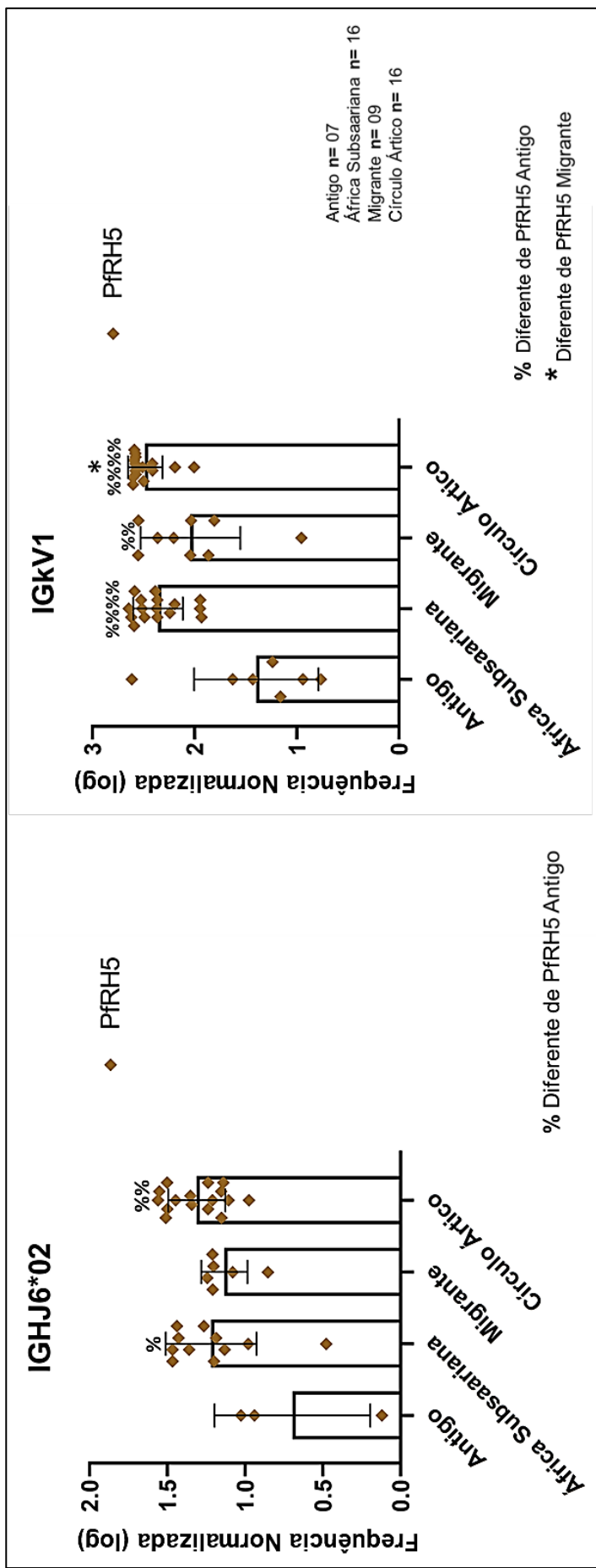
Em relação à sequência IGHV1 (Figura 16), o grupo Antigo apresentou a menor frequência normalizada deste segmento gênico para a codificação dos CDRs dos anticorpos específicos para os antígenos PfrH5, RBD 2 e RBD 4, com variação estatisticamente significativa ( $p = <0,0001$ ) em relação aos demais grupos. O grupo Antigo também apresentou o maior desvio padrão para o segmento IGHV1 (Antigo  $\sigma = 0,789$  vs. África Subsaariana  $\sigma = 0,167$ ; Migrante  $\sigma = 0,116$  e Círculo Ártico  $\sigma = 0,215$ ). Não foi observada diferença estatística significativa na frequência normalizada da sequência IGHV1 entre os grupos África Subsaariana, Migrante e Círculo Ártico.

Quanto a sequência IGkJ2\*01 (Figura 16), o grupo Antigo apresentou a menor frequência normalizada desse segmento gênico quando utilizado para codificar os CDRs dos anticorpos direcionados ao antígeno PfNANP em comparação ao grupo Círculo Ártico para os antígenos PfNANP ( $p = 0,0258$ ), PfrH5 ( $p = 0,0244$ ) e RBD 1 ( $p = 0,0163$ ); PfrH5 em comparação aos antígenos PfNANP ( $p = 0,084$ ), PfrH5 ( $p = 0,0078$ ) e RBD 1 ( $p = 0,0046$ ); e RBD 1 em comparação aos antígenos PfNANP ( $p = 0,0258$ ), PfrH5 ( $p = 0,0244$ ) e RBD 1 ( $p = 0,0163$ ). Além disso, o segmento gênico IGkJ2\*01 (Figura 16) apresentou diferença estatística significativa da frequência normalizada para a codificação dos CDRs dos anticorpos específicos para o antígeno PfrH5 no grupo África Subsaariana quando comparado ao antígeno PfrH5 no grupo Círculo Ártico ( $p = 0,048$ ), demonstrando diferenças no uso da sequência IGkJ2\*01 entre populações diferentemente expostas ao *P. falciparum*. Também houve variação estatisticamente significativa na frequência normalizada no uso desta sequência para codificar os CDRs dos anticorpos direcionados ao antígeno RBD 1 no grupo Círculo Ártico em relação ao antígeno PfNANP ( $p = 0,0257$ ) e PfrH5 ( $p = 0,0225$ ) no grupo África Subsaariana. Os grupos África Subsaariana e

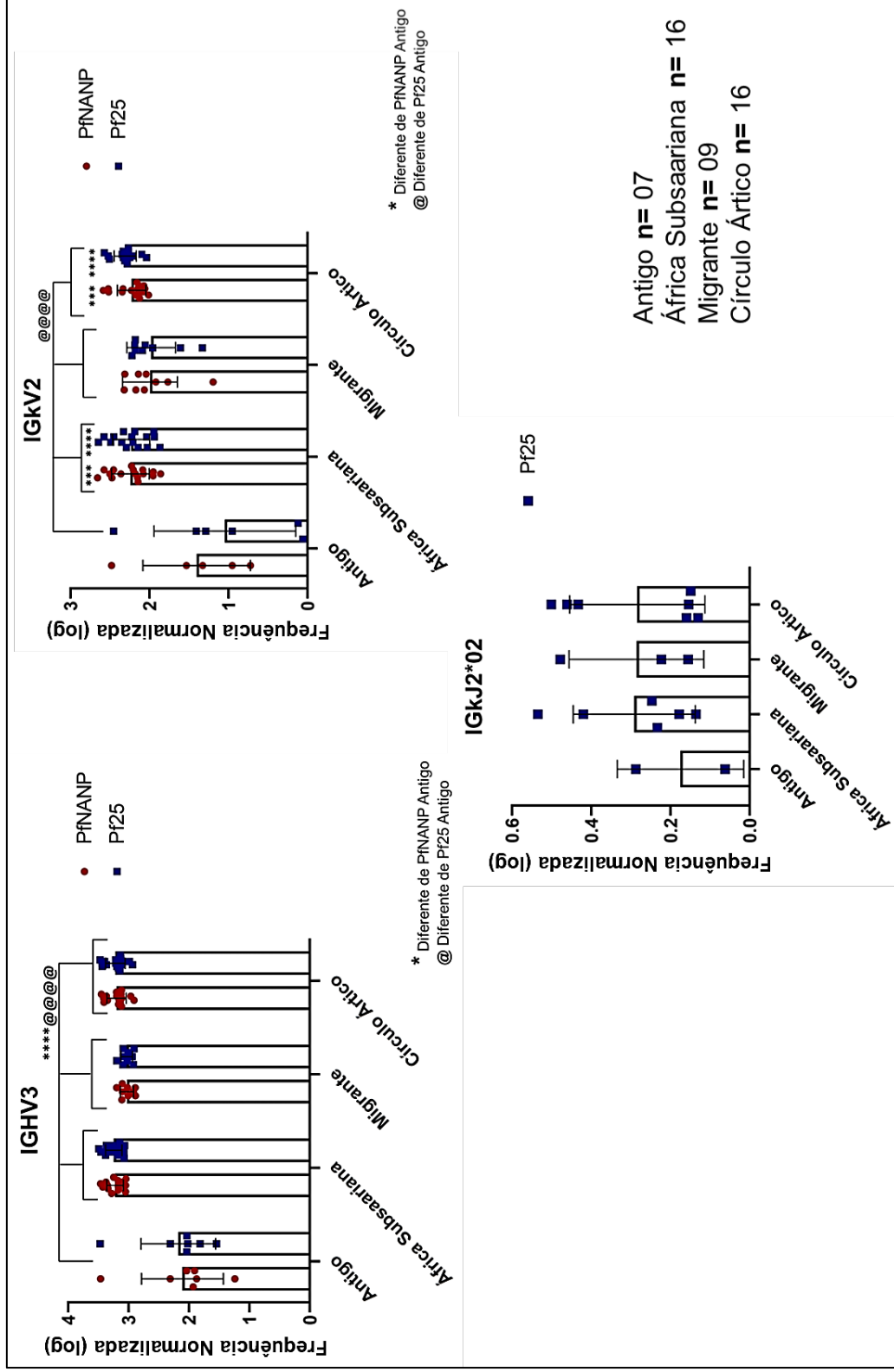
Migrante apresentaram os maiores desvios padrão para o segmento IGkJ2\*01, indicando uma distribuição mais ampla desse segmento gênico nos dois grupos de origem africana PfNANP ( $\sigma = 0,492$ ;  $0,535$ , respectivamente) vs. PfNANP (Antigo  $\sigma = 0,310$ ; Círculo Ártico  $\sigma = 0,437$ ), PfrH5 ( $\sigma = 0,502$ ;  $0,535$ , respectivamente) vs. PfrH5 (Antigo  $\sigma = 0,220$ ; Círculo Ártico  $\sigma = 0,441$ ), RBD 1 ( $0,498$ ;  $0,535$ , respectivamente) vs. RBD 1 (Antigo  $\sigma = 0,310$ ; Círculo Ártico  $\sigma = 0,389$ ).

**Com exceção do segmento gênico IGkV1 as sequências V(D)J que codificam os CDRs das cadeias leves e pesadas dos anticorpos neutralizantes direcionados aos antígenos do *P. falciparum*, são encontrados nas populações estudadas independentemente da exposição ao patógeno.**

As Figuras 17 e 18 apresentam a frequência normalizada logarítmica (log) e o desvio padrão ( $\sigma$ ) dos segmentos V(D)J responsáveis pelas regiões CDR das cadeias leves (IGk ou IGL) e pesadas (IGH) dos anticorpos neutralizantes específicos para os antígenos do *P. falciparum* (PfNANP, Pf25 e PfrH5), presentes nos genomas dos grupos Antigo, África Subsaariana, Migrante e Círculo Ártico.



**Figura 17** Gráficos de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) dos genes *V(D)J* que codificam as cadeias pesadas (*IGH*) e leve (*IGk* ou *IGL*) de anticorpos neutralizantes específicos para os antígenos do *P. falciparum* em diferentes grupos. Antígenos alvo: PfNANP, Pfs25, PfRH5. n = 48 indivíduos. Pf: *P. falciparum*. Referência: LISTA DE FIGURAS.



**Figura 18** Gráficos de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) dos genes *V(D)J* que codificam as cadeias pesadas (*IGH*) e leve (*IGk* ou *IGL*) de anticorpos neutralizantes específicos para os antígenos do *P. falciparum* em diferentes grupos. Antígenos alvo: PfNANP, Pfs25, PFRH5. n = 48 indivíduos. Pf: *P. falciparum*. Referência: LISTA DE FIGURAS.

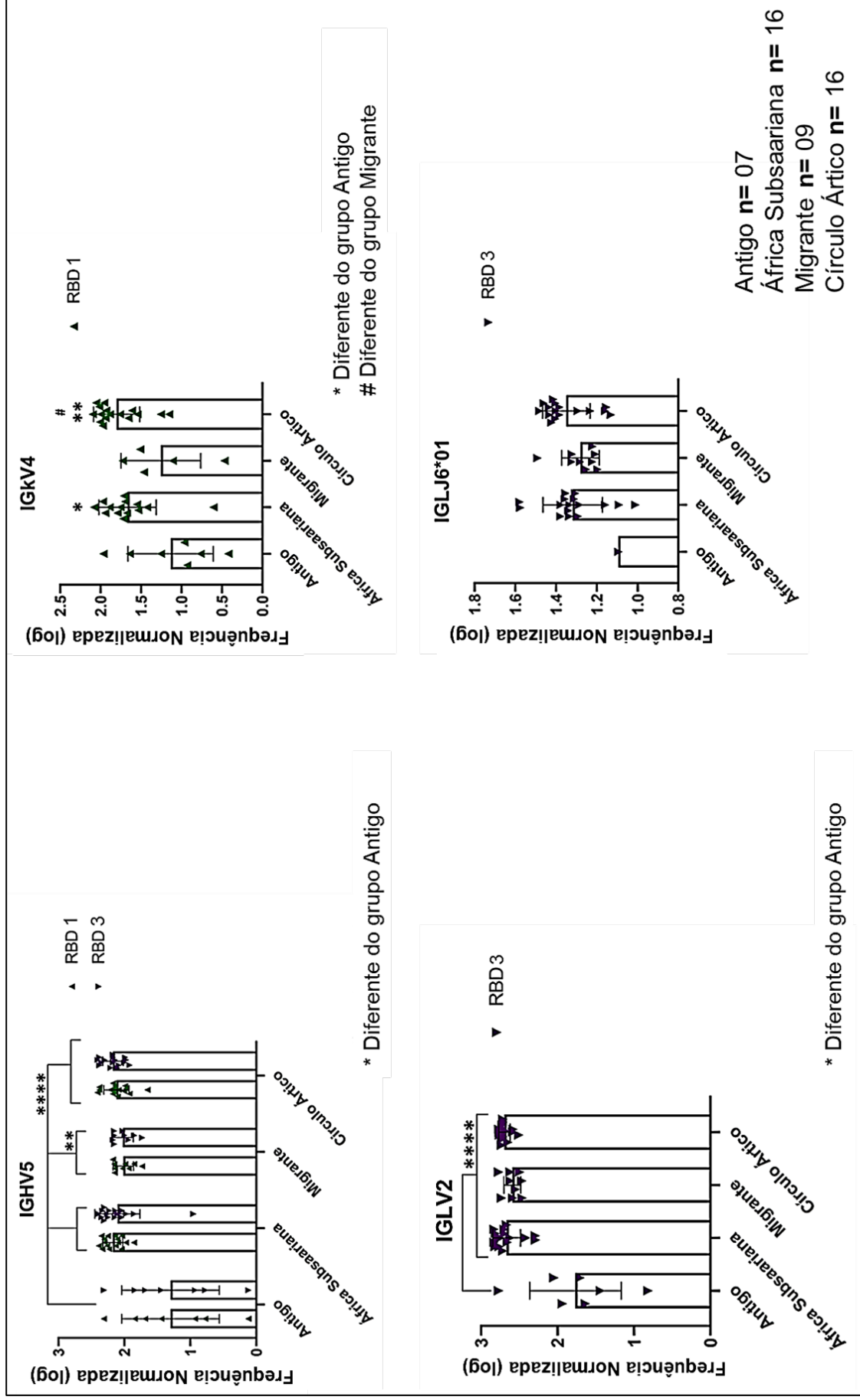
A sequência IGHJ6\*02 (Figura 17), que codifica os CDRs dos anticorpos específicos para o antígeno PfRH5, apresentou seu maior desvio padrão ( $\sigma = 0,501$ ) e menor frequência normalizada no grupo Antigo, com diferença estatisticamente significativa em comparação aos grupos África Subsaariana ( $\sigma = 0,292$ ;  $p = 0,0143$ ) e Círculo Ártico ( $\sigma = 0,184$ ;  $p = 0,0022$ ). Para o segmento IGkV1 (Figura 17), que também codifica os CDRs dos anticorpos direcionados ao antígeno PfRH5, o grupo Antigo mais uma vez apresentou o maior desvio padrão ( $\sigma = 0,610$ ) e a menor frequência normalizada, com diferença estatisticamente significativa em relação aos grupos África Subsaariana ( $\sigma = 0,243$ ;  $p < 0,0001$ ), Migrante ( $\sigma = 0,489$ ;  $p = 0,0039$ ) e Círculo Ártico ( $\sigma = 0,168$ ;  $p < 0,0001$ ). Além disso, o grupo Migrante apresentou diferença estatisticamente significativa em relação ao grupo Círculo Ártico ( $p = 0,0213$ ) para o segmento IGkV1, demonstrando diferença no uso deste segmento gênico entre populações diferentemente expostas ao *P. falciparum*.

Os segmentos gênicos IGHV3 e IGkV2 (Figura 18), que codificam os CDRs dos anticorpos específicos para os antígenos PfNANP e Pf25, também apresentaram os maiores desvios padrão e as menores frequências normalizadas no grupo Antigo. A sequência IGHV3 (Figura 18) apresentou diferença estatisticamente significativa em relação aos demais grupos analisados para ambos os antígenos PfNANP Antigo ( $\sigma = 0,679$ ;  $p < 0,0001$ ) vs. PfNANP África Subsaariana ( $\sigma = 0,135$ ;  $p < 0,0001$ ), Migrante ( $\sigma = 0,110$ ;  $p < 0,0001$ ) e Círculo Ártico ( $\sigma = 0,159$ ;  $p < 0,0001$ ), PfNANP Antigo ( $\sigma = 0,679$ ;  $p < 0,0001$ ) vs. Pf25 África Subsaariana ( $\sigma = 0,135$ ;  $p < 0,0001$ ), Migrante ( $\sigma = 0,096$ ;  $p < 0,0001$ ) e Círculo Ártico ( $\sigma = 0,161$ ;  $p < 0,0001$ ) e Pf25 Antigo ( $\sigma = 0,618$ ) vs. Pf25 África Subsaariana ( $p < 0,0001$ ), Migrante ( $p < 0,0001$ ) e Círculo Ártico ( $p < 0,0001$ ). E a sequência IGkV2 do grupo Antigo só não apresentou diferença estatística significativa para o antígeno PfNANP em comparação ao grupo Migrante. Para os demais grupos temos, PfNANP Antigo ( $\sigma = 0,680$ ) vs. PfNANP África Subsaariana ( $\sigma = 0,239$ ;  $p < 0,0002$ ) e Círculo Ártico ( $\sigma = 0,180$ ;  $p = 0,0003$ ), PfNANP Antigo ( $\sigma = 0,680$ ) vs. Pf25 África Subsaariana ( $\sigma = 0,234$ ;  $p < 0,0001$ ), e Círculo Ártico ( $\sigma = 0,140$ ;  $p < 0,0001$ ); Pf25 Antigo vs. Pf25 África Subsaariana ( $p < 0,0001$ ), Migrante ( $p < 0,0001$ ) e Círculo Ártico ( $p < 0,0001$ ). Por fim, não foram observadas diferenças estatisticamente significativas na análise do segmento IGkJ2\*02 (Figura 18) que codifica os CDRs dos anticorpos direcionados ao antígeno Pf25. É relevante destacar que não foram observadas diferenças estatisticamente significativas na frequência normalizada para a maioria dos segmentos gênicos analisados entre os grupos África Subsaariana, Migrante e Círculo Ártico. Isso sugere que a presença das sequências IGHJ6\*02, IGHV3, IGkV2 e IGkJ2\*02 ocorre de forma independente da exposição ao *P. falciparum*.

**Com exceção dos segmentos gênicos IGkV4 e IGLJ3\*02, as sequências V(D)J que codificam os CDRs das cadeias leves e pesadas dos anticorpos neutralizantes direcionados aos antígenos do SARS-CoV-2, ocorrem de forma independente da exposição ao patógeno.**

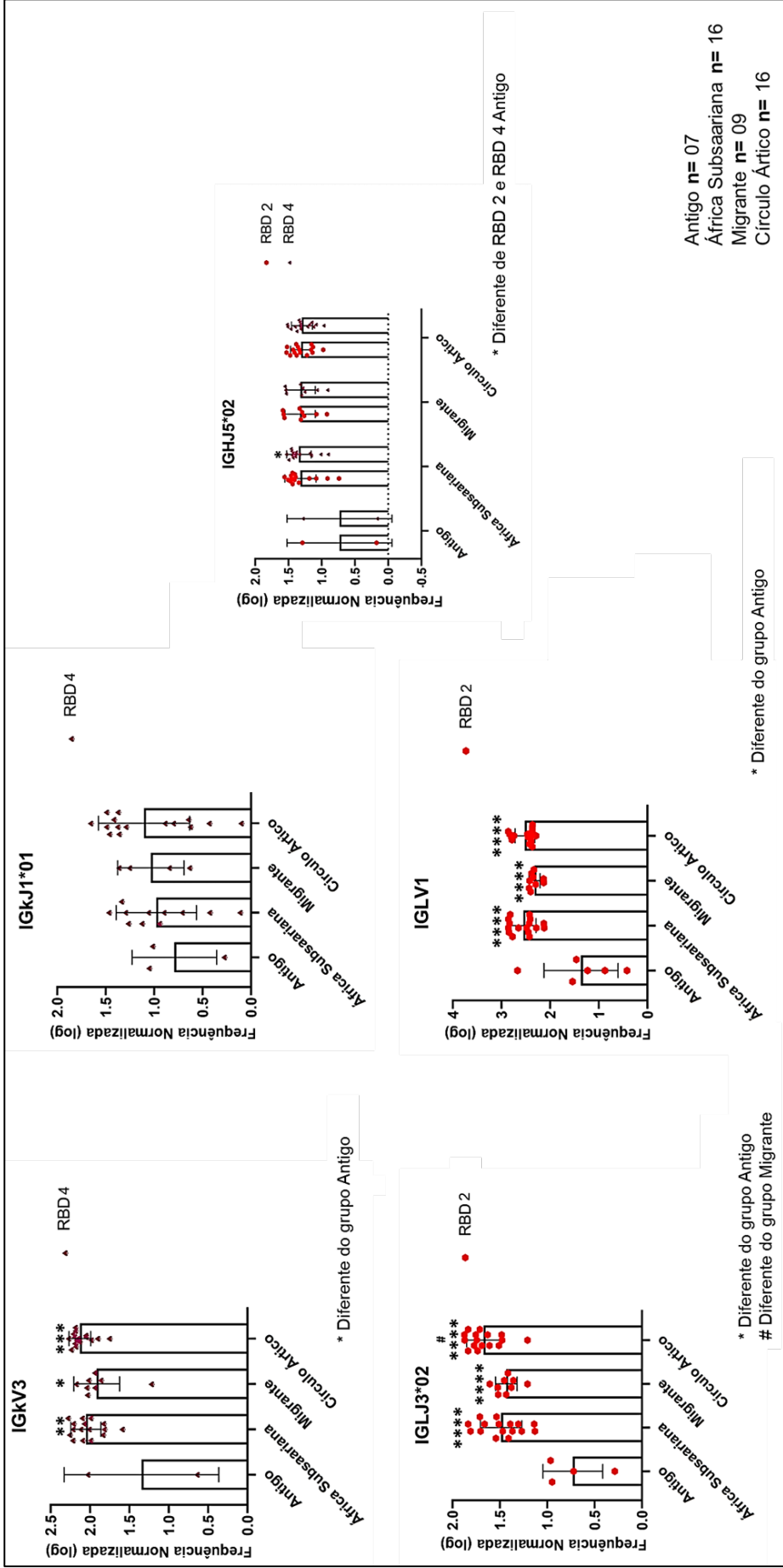
Os gráficos das Figuras 19 e 20 apresentam a frequência normalizada logarítmica (log) e o desvio padrão ( $\sigma$ ) dos segmentos V(D)J responsáveis pelas regiões CDR das cadeias leves (IGk ou IGL) e pesadas (IGH) dos anticorpos neutralizantes específicos para os antígenos do SARS-CoV-2 (RBD 1, RBD 2, RBD 3 e RBD 4), presentes nos genomas dos grupos Antigo, África Subsaariana, Migrante e

Círculo Ártico.



**Figura 19** Gráficos de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) dos genes *V(D)J* que codificam as cadeias pesadas (*IGH*) e leve (*IGk* ou *IGL*) de anticorpos neutralizantes específicos para os antígenos do SARS-CoV-2 em diferentes grupos. Antígenos alvo: RBD 1, 2, 3 e 4.  $n = 48$  indivíduos. Referência: LISTA DE FIGURAS.



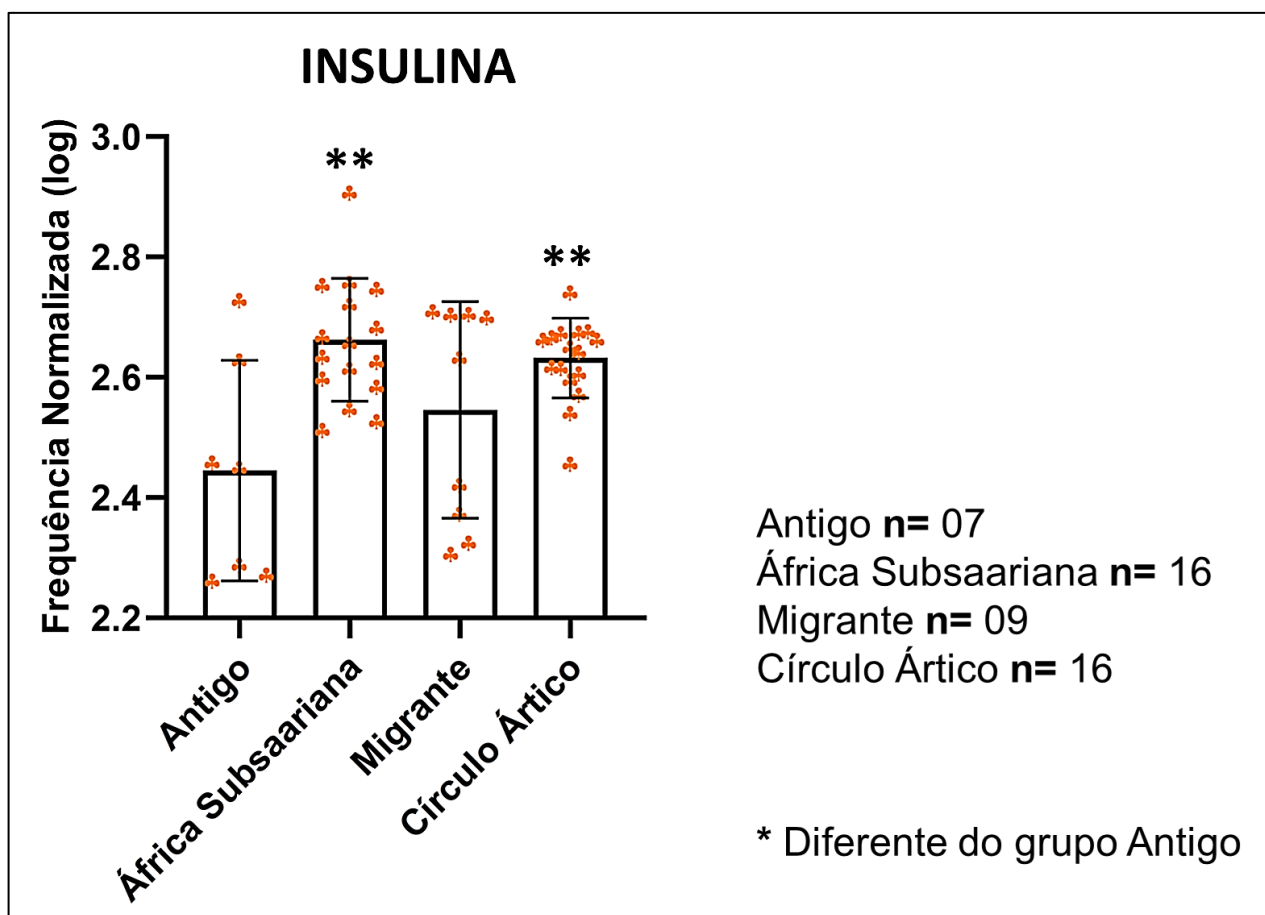


**Figura 20** Gráficos de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) dos genes *V(D)J* que codificam as cadeias pesadas (*IGH*) e leve (*IGk* ou *IGL*) de anticorpos neutralizantes específicos para os antígenos do SARS-CoV-2 em diferentes grupos. Antígenos alvo: RBD 1, 2, 3 e 4.  $n = 48$  indivíduos. Referência: LISTA DE FIGURAS.

Para a sequência IGHV5 (Figura 19), que codifica os CDRs dos anticorpos específicos para os antígenos RBD 1 e RBD 3, o grupo Antigo apresentou o maior desvio padrão ( $\sigma = 0,741$ ) e a menor frequência normalizada em comparação aos demais grupos estudados, com diferença estatisticamente significativa, RBD 1 África Subsaariana ( $\sigma = 0,151$ ;  $p = <0,0001$ ), Migrante ( $\sigma = 0,156$ ;  $p = 0,0015$ ) e Círculo Ártico ( $\sigma = 0,191$ ;  $p = <0,0001$ ) e RBD 3 África Subsaariana ( $\sigma = 0,342$ ;  $p = <0,0001$ ), Migrante ( $\sigma = 0,157$ ;  $p = 0,0013$ ) e Círculo Ártico ( $\sigma = 0,156$ ;  $p = <0,0001$ ). Esse resultado também foi observado para os segmentos gênicos IGLV2 (Figura 19), IGkV3 (Figura 20), IGLV1 (Figura 20) e IGLJ3\*02 (Figura 20), com a diferença que este último segmento também apresentou diferença estatística significativa ao codificar os CDRs dos anticorpos específicos para o antígeno RBD 2 no grupo Migrante ( $\sigma = 0,115$ ) quando comparado ao grupo Círculo Ártico ( $\sigma = 0,180$ ;  $p <0,0298$ ). Esse resultado aponta diferença no uso da sequência IGLJ3\*02 entre populações teoricamente igualmente expostas ao SARS-CoV-2. Para a sequência IGHJ5\*02 (Figura 20), que codifica os CDRs dos anticorpos direcionados aos antígenos RBD 2 e RBD 4 houve diferença estatística significativa entre o grupo Antigo em relação ao antígeno RBD 4 no grupo África Subsaariana ( $p = 0,0163$ ). O segmento gênico IGkV4 (Figura 19), que codifica os CDRs dos anticorpos específicos para o antígeno RBD 1, apresentou o maior desvio padrão ( $\sigma = 0,528$ ) e a menor frequência normalizada no grupo Antigo, com diferença estatística significativa entre este e os grupos África Subsaariana ( $\sigma = 0,355$ ;  $p <0,022$ ) e Círculo Ártico ( $\sigma = 0,285$ ;  $p <0,0022$ ). E o grupo Migrante ( $\sigma = 0,493$ ) quando comparado ao grupo Círculo Ártico ( $\sigma = 0,285$ ;  $p <0,0382$ ), demonstrando diferença no uso do segmento gênico IGkV4 entre populações teoricamente igualmente expostas ao SARS-CoV-2. Por outro lado, não foram encontradas diferenças estatisticamente significativas para as sequências IGLJ6\*01 (Figura 19) e IGkJ1\*01 (Figura 20). É importante ressaltar que não foram observadas variações estatisticamente significativas na frequência normalizada dos segmentos gênicos analisados entre os grupos África Subsaariana, Migrante e Círculo Ártico para os segmentos gênicos IGHV5, IGLV2, IGLJ6\*01, IGkV3, IGkJ1\*01, IGLV1 e IGkJ5\*02, sugerindo que eles ocorrem de forma independente da exposição ao SARS-CoV-2.

**A análise da presença do gene da insulina nos genomas revelou variações estatisticamente significativas no grupo Antigo.**

A Figura 21 apresenta a frequência normalizada em escala logarítmica (log) e o desvio padrão ( $\sigma$ ) do gene da insulina humana nos genomas dos grupos Antigo, África Subsaariana, Migrante e Círculo Ártico.



**Figura 21** Gráfico de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) do gene da insulina em diferentes grupos.  $n = 48$  indivíduos. Referência: LISTA DE FIGURAS.

O grupo Antigo apresentou diferença estatisticamente significativa na frequência normalizada deste gene em comparação aos grupos África Subsaariana ( $\sigma = 0,102$ ;  $p = 0,0021$ ) e Círculo Ártico ( $\sigma = 0,033$ ;  $p = 0,0071$ ). No entanto, não foram encontradas diferenças significativas em relação ao grupo Migrante ( $\sigma = 0,180$ ;  $p = 0,3257$ ).

Neste trabalho, não encontramos nenhum número de ocorrências ou *matches* para as sequências D em nenhum dos segmentos gênicos V(D)J analisados. Também, não foi observado variação estatística significativa da frequência normalizada intragrupos para os segmentos gênicos responsáveis pela codificação das regiões CDR dos anticorpos específicos para os antígenos do *P. falciparum* e SARS-CoV-2. Além disso, não foi observada diferença estatística significativa entre os grupos África Subsaariana e Migrante para nenhum segmento gênico.

## DISCUSSÃO

Esse projeto foi delineado com intuito de investigar possíveis diferenças no uso dos segmentos gênicos V(D)J envolvidos na geração dos CDRs de anticorpos neutralizantes e específicos contra *P. falciparum* e SARS-CoV-2, entre populações sob diferentes influências do *P. falciparum*.

Neste estudo, foram identificados dezessete segmentos gênicos diferentes, somente três deles envolvidos na formação de CDRs contra os antígenos do *P. falciparum* e SARS-CoV-2 concomitantemente. São eles os segmentos gênicos IGHJ4\*02, IGHV1 e IGkJ2\*01. Ainda, cinco segmentos gênicos codificam CDRs para os anticorpos exclusivamente contra os antígenos do *P. falciparum* (IGHJ6\*02, IGkV1, IGHV3, IGkV2 e IGkJ2\*02) e nove, a grande parte dos segmentos gênicos estudados, codificam apenas CDRs contra os antígenos do SARS-CoV-2 (IGHV5, IGkV4, IGLV2, IGLJ6\*01, IGkV3, IGkJ1\*01, IGLJ3\*02, IGLV1 e IGHJ5\*02). Como esperado, todos os segmentos gênicos apareceram em todos os grupos de genomas estudados. Não foram identificados os segmentos gênicos D para nenhuma sequência V(D)J codificadora. Os segmentos D são muito curtos e provavelmente não apresentaram mais de 80% de similaridade com as sequências germinativas dos anticorpos estudados, por isso, foram automaticamente excluídos da análise.

Nossos resultados demonstraram que os genomas de Neandertais (grupo Antigo) possuem uma frequência normalizada muito baixa para os segmentos gênicos germinativos VJ necessários para a geração de anticorpos neutralizantes contra antígenos do *P. falciparum*, apesar de seu genoma ser 99,7% idêntico ao genoma humano moderno (122,127). Os Neandertais eram hominídeos antigos mais próximos em relação aos humanos modernos (120), e viveram na Europa e no sudoeste da Ásia central (128) entre 40.000 e 400.000 anos atrás (121,125,128). De acordo com Baron e colaboradores (2011), a primeira infecção de um ancestral humano por *P. falciparum* aconteceu por volta de 112.000 a 1.036.000 anos atrás (estimativa média 365.000 anos atrás) (29). Conseqüentemente, os Neandertais coexistiram com o *P. falciparum* embora não tenham sofrido sua influência por serem habitantes de territórios europeus, onde não há registros deste patógeno naquele período.

Com o objetivo de identificar se possíveis danos físico-químicos nos genomas Neandertais pudessem justificar as frequências normalizadas mais baixas, foi conduzido um estudo comparativo utilizando o gene da insulina, que é muito conservado dentre os vertebrados. Se todos os genomas estudados fossem equivalentes em qualidade do sequenciamento, não se esperariam variações significativas na frequência normalizada do gene da insulina entre os grupos analisados. Entretanto, o grupo Antigo apresentou variação estatisticamente significativa em relação aos grupos África Subsaariana e Círculo Ártico. Notadamente, o grupo Migrante, que compartilha genomas provenientes dos mesmos estudos dos demais grupos modernos e é presumido possuir a mesma qualidade de sequenciamento, não demonstrou diferença estatística em relação ao grupo Antigo. Assim como este, o grupo Migrante também apresenta um tamanho amostral menor quando comparado aos outros grupos de genomas modernos. Portanto, é possível que o aumento do tamanho amostral do grupo Antigo pudesse eliminar a diferença estatística e o elevado  $\sigma$  encontrados neste estudo comparativo. Infelizmente, devido à escassez de estudos que realizam e divulgam o sequenciamento de genomas antigos, foi inviável incluir e analisar mais genomas Neandertais neste projeto. A falta de disponibilidade de dados genômicos antigos pode ser atribuída às dificuldades técnicas e logísticas envolvidas em estudos dessa magnitude. Para mitigar a ausência de genomas Neandertais a serem analisados, uma alternativa seria buscar o gene da insulina humana contra os *reads* ou leituras de genomas de chimpanzé ou gorila e verificar como os números se comportam. Dado a alta conservação

deste gene, é provável que esta estratégia não geraria falsos positivos para a nossa análise.

Isso nos leva à segunda hipótese de que o resultado estatístico observado pode estar relacionado à degradação do DNA do grupo Antigo ao longo de dezenas de milhares de anos (120), o que afeta o subsequente sequenciamento do genoma e, conseqüentemente, a qualidade dos dados utilizados em nosso estudo. Além disso, é importante considerar possíveis variações ou diferenças genéticas entre os Neandertais e os humanos modernos. Estudos genéticos comparativos entre essas duas populações revelaram algumas variações específicas dos Neandertais (129), incluindo genes relacionados ao metabolismo e à resposta imune (130,131). Portanto, embora seja provável que o gene da insulina estivesse presente nos Neandertais, bem como os segmentos gênicos V(D)J, algumas diferenças ou variações genéticas em comparação com as versões humanas modernas podem ter ocorrido. Diante disso, visto que o nosso estudo possui um caráter exploratório, decidimos manter os resultados relacionados aos genomas antigos, com a premissa de que estudos futuros mais aprofundados podem revelar diferenças nas sequências VDJ mais raras ou que ainda não foram identificadas entre os humanos modernos e antigos. Essas descobertas poderiam fornecer novas evidências sobre a evolução desses genes ao longo da história humana e, como as possíveis variações podem conferir resistência ou suscetibilidade aos novos desafios imunológicos.

Os segmentos gênicos IGHJ4\*02 e IGkJ2\*01 demonstraram variações na frequência normalizada de acordo com a exposição dos genomas ao *P. falciparum* e ao SARS-CoV-2. De maneira interessante, no grupo do Círculo Ártico, que teoricamente não teve contato prévio com o *P. falciparum*, as frequências normalizadas dessas sequências foram mais elevadas em comparação ao grupo da África Subsaariana, especificamente nos segmentos que codificam os CDRs dos anticorpos direcionados aos antígenos deste patógeno (PfNANP e PfRH5) e ao SARS-CoV-2 RBD 1. Bin Shi e colaboradores (2020) investigaram as características da composição dos repertórios BCR para as cadeias pesadas por meio do sequenciamento de alto rendimento HTS (do inglês *high-throughput sequencing*) e concluíram que o gene IGHJ4 humano possui a mais alta frequência e consistência de uso entre os segmentos IGHJ, independentemente das condições fisiológicas ou patológicas, implicando que sua seleção não é aleatória (132). Em contrapartida, Alanine e colaboradores (2019) identificaram o segmento gênico IGkJ2\*01 na região variável de três anticorpos monoclonais (mAbs) anti-PfRH5, os quais foram isolados de voluntários previamente imunizados (18). O anticorpo catalogado como 6RCU (clone R5.016) está entre esses três mAbs e se destaca por ser altamente eficaz em bloquear a entrada da forma merozoíta do *P. falciparum* nas hemácias (18). No entanto, era esperado que a sequência IGHJ4 não apresentasse diferença estatística entre os grupos estudados, enquanto a sequência IGkJ2\*01 deveria exibir uma frequência normalizada maior nos grupos de origem africana, que foram expostos ao *P. falciparum*.

Os segmentos gênicos IGkV1, IGkV4 e IGLJ3\*02 também revelaram variações na frequência normalizada de acordo com a exposição dos genomas ao *P. falciparum* e ao SARS-CoV-2. Mais uma vez, no grupo do Círculo Ártico, que teoricamente não teve contato prévio com o *P. falciparum*, as frequências normalizadas destas sequências foram maiores em comparação ao grupo Migrante, mas não em comparação ao grupo África Subsaariana. Esse padrão foi observado nos segmentos que

codificam os CDRs dos anticorpos direcionados aos antígenos PfrH5, SARS-CoV-2 RBD 1 e SARS-CoV-2 RBD 2, respectivamente. Vários autores identificaram o grupo de alelos do segmento IGkV1 presente em anticorpos potentes contra o antígeno PfrH5 (114,115,133,134). No entanto, era esperado que esta sequência, assim como a sequência IGkJ2\*01, apresentasse uma frequência normalizada maior nos grupos de origem africana, a fim de que nossos resultados pudessem validar a seleção desses segmentos gênicos no contexto de proteção dos indivíduos expostos ao *P.falciparum*. Ainda, Coelho e colaboradores (2020) investigaram as sequências peptídicas do fragmento Fab humano da IgG plasmática de 79 indivíduos adultos (incluindo vacinados e um grupo controle) que receberam 4 doses da vacina Pfs25-EPA/Alhydrogel. Entre as sequências de cadeia leve, o gene IGkV4 foi o mais frequentemente encontrado entre os indivíduos imunizados (135). Em nossa pesquisa, a sequência IGkV4 não demonstrou um aumento significativo na frequência normalizada ao codificar CDRs contra o antígeno Pfs25 do *P. falciparum*, mas foi mais prevalente no grupo do Círculo Ártico ao codificar apenas os CDRs dos anticorpos direcionados ao antígeno SARS-CoV-2 RBD 1.

Adicionalmente, He e colaboradores (2021) realizaram uma análise da diversidade e frequência das cadeias pesadas e leves das imunoglobulinas naturais emparelhadas em cinco pacientes convalescentes da COVID-19 (136). Para isso, eles empregaram single-cell RNA-seq e single-cell VDJ-seq. De forma semelhante, a sequência IGkV4-1 foi identificada como um dos cinco segmentos gênicos mais frequentes nos BCRs desses indivíduos, quando pareada com a cadeia pesada IGHV4-59 (136). Assim como evidenciado no estudo de Coelho e colaboradores (2020) (135), esses achados sugerem uma tendência de preferência por segmentos gênicos de cadeia leve e pesada específicos de imunoglobulina em pacientes convalescentes da COVID-19. Isso indica que as células B podem ter passado por rearranjos V(D)J únicos e não estocásticos (ou seja, não aleatórios) nesses pacientes (136). Logo, esses estudos demonstram uma preferência pela seleção do gene IGkV4-1 em situações patológicas. O nosso trabalho parcialmente corrobora este achado, uma vez que a sequência IGkV4 está exclusivamente presente ao codificar os CDRs de anticorpos específicos para o SARS-CoV-2, antígeno RBD 1. Entretanto não esperávamos encontrar nenhuma diferença relacionada aos segmentos envolvidos na imunidade contra a COVID-19 entre os grupos estudados. Até o momento, não existem dados na literatura que apoiem o uso da sequência IGLJ3\*02 na proteção contra os dois patógenos abordados neste projeto, nem evidências de uma preferência pela seleção desse segmento gênico em situações fisiológicas ou patológicas.

Os efeitos funcionais da variação na linhagem germinativa dos genes de IG devem ser encontrados ao estudar as respostas imunes a patógenos (137). Existem evidências de que a diversidade dos genes de anticorpos pode ser grandemente influenciada por seleção (93). Como descrito anteriormente, descobertas que corroboram esse fato incluem a associação entre variantes dos genes de imunoglobulina e suscetibilidade a doenças, a hereditariedade do repertório BCR *naïve* (93), etc. Entretanto, ao analisar as variações do repertório germinativo dos loci de IG, deve-se considerar razoável a hipótese de que a maioria das variações interindividuais a nível germinativo é resultado de processos estocásticos (aleatórios) (93). Os nossos resultados apontam mais para uma variação germinativa entre genomas europeus e africanos do que devido a exposição ao *P. falciparum*. Os cinco

segmentos gênicos que demonstraram diferença estatística significativa da frequência normalizada estão mais presentes no grupo do Círculo Ártico do que nos genomas africanos, sem variações significativas entre os grupos África Subsaariana e Migrante e independente da sua participação na codificação de anticorpos já conhecidos, contra o *P. falciparum* e SARS-CoV-2.

Embora seja amplamente reconhecido que o processo de maturação da afinidade das imunoglobulinas seja uma forma de evolução estrita, os imunologistas dedicaram relativamente pouca atenção à genética evolutiva dos alelos e genes germinativos dessas moléculas (93). Devido à compreensão limitada sobre a diversidade populacional no que diz respeito ao repertório dos anticorpos (138), ainda não foi estabelecido se a natureza da variação nos loci de IG é predominantemente resultado de mutações estocásticas, deriva genética, seleção direcional ou seleção balanceada, uma questão que já vem sendo elucidada em estudos de genes como os HLA/MHC, por exemplo (93,139,140). Além disso, para os genes e alelos das imunoglobulinas, não existem evidências de varreduras seletivas, nas quais um novo alelo se torna predominante na população devido à seleção natural (93). Também não existem dados científicos que indiquem a ocorrência de adaptação local, na qual um alelo ou combinação de alelos é vantajoso em um ambiente específico, mas não globalmente (93). Com o intuito de endereçar esse último tópico, nossos resultados sugerem que a seleção não é a principal força genética que atua na utilização de segmentos V(D)J específicos em resposta a um determinado patógeno. A deriva genética, um princípio evolutivo fundamental que descreve mudanças aleatórias nas frequências alélicas (141), é o que melhor explica os nossos achados neste estudo. Se a seleção for fraca, é possível que a variação nos genes de IG seja mantida pelo equilíbrio entre mutação e deriva genética, no qual novos alelos são introduzidos na população devido a mutações (amplamente definidas, incluindo duplicação de genes e rearranjos genômicos) e são perdidos, simplesmente devido ao acaso (93). Como resultado, as diferenças na frequência dos segmentos gênicos de IG encontradas nas populações estudadas podem simplesmente refletir padrões históricos de migração e/ou isolamento (93).

As populações humanas no continente africano possuem a maior diversidade genética no mundo moderno (119,142). Essa diversidade é influenciada por diversos fatores, como eventos demográficos, migração, mistura de grupos populacionais e forças genéticas, como mutação, recombinação e seleção natural (143). Essa complexidade explica os variados padrões de distribuição das sequências VJ nos grupos África Subsaariana e Migrante. De fato, nossos resultados mostram uma tendência de distribuição mais dispersa nestes genomas em comparação ao grupo do Círculo Ártico, que é mais homogêneo. Esse fenômeno pode ser observado para os segmentos gênicos IGHJ4\*02 e IGkJ2\*01 onde os desvios padrão se mostraram maiores nos grupos da África Subsaariana e Migrante e, se não considerarmos os desvios padrões do grupo Antigo, a distribuição dos demais três segmentos que demonstraram diferença significativa em sua frequência normalizada (IGkV1, IGkV4 e IGLJ3\*02) estão também mais dispersos nos grupos de origem Africana. Os grupos África Subsaariana e Migrante foram selecionados para diferenciar possíveis variações no uso dos segmentos VJ devido à influência primária do *P. falciparum*. O grupo Migrante, que se originou das mesmas populações que compõem o grupo África Subsaariana, deixou de sofrer a influência desse parasita ao migrar para regiões da

África Austral em um momento em que o *P. falciparum* não estava presente nesta parte do continente. Apesar disso, não foram identificadas variações na frequência normalizada de nenhum segmento gênico entre os dois grupos. Isso pode sugerir que o tempo estimado para a migração da população Bantu, há cerca de 3.000 a 5.000 anos, não foi suficiente para permitir a seleção das sequências de imunoglobulina de forma independente à pressão seletiva do *P. falciparum*.

A seleção é mais observada em patógenos que têm uma longa relação com o *Homo sapiens*, incluindo os agentes etiológicos da malária (137). A infecção por *P. falciparum* exerceu pressão seletiva significativa nos genomas humanos, levando ao desenvolvimento de mecanismos de resistência contra muitas doenças maláricas (144). Um exemplo desses fatores genéticos do hospedeiro é a anemia falciforme (144), causada por uma mutação pontual no gene da beta-globina ( $\beta S$ ) em homozigose (145), que ainda está selecionando pessoas para sobreviverem à malária na África Central (144) e é um dos exemplos mais conhecidos de seleção balanceada (130). Essa mutação, surgiu aproximadamente há 7.300 anos nas regiões endêmicas de malária, incluindo a África Subsaariana (146). Acredita-se que as expansões Bantu tenham disseminado um dos haplótipos do alelo  $\beta S$  para diversas regiões africanas, incluindo a África do Sul (146,147). Apesar disso, a incidência de anemia falciforme na África do Sul é muito baixa, e a maioria dos casos (que estão aumentando na região) são resultado das migrações de outras regiões africanas, incluindo a África Central, onde a anemia falciforme é prevalente (147). Até onde se sabe, não existem estudos que correlacionam a menor quantidade de casos de anemia falciforme com a baixa incidência de *P. falciparum* na África do Sul, o que seria um indício da seleção gênica do gene  $\beta S$  independente da influência deste patógeno.

Este trabalho utilizou genomas sequenciados com a melhor técnica largamente disponível nos dias atuais, o sequenciamento de nova geração com leituras curtas, o SR-NGS (do inglês, *short-read next-generation sequencing*). No entanto, as informações disponíveis sobre a composição e a variação sequencial nos loci de IG humanos permanecem incompletas (75). Isso se deve aos genes serem altamente similares e, por vezes, duplicados (75). Mais alelos V(D)J estão sendo descobertos através do advento de novas tecnologias, como o desenvolvimento de ferramentas de software especializadas e o sequenciamento de leituras longas (LRS, do inglês *long-read sequencing*). A principal diferença entre essas duas abordagens é o significativo aumento no comprimento das leituras. Enquanto as leituras curtas têm um tamanho de 150-300 pb, o LRS tem a capacidade de sequenciar, em média, mais de 10 kb em uma única leitura, necessitando assim de menos leituras para cobrir o mesmo gene e possibilitando a descoberta de regiões repetitivas (148). A literatura atual menciona que, embora haja uma grande quantidade de dados de sequenciamento de genoma humano de leituras curtas disponíveis, as abordagens atuais não são capazes de resolver completamente as variantes em todos os loci dos genes V(D)J. Isso se deve à alta densidade de duplicações segmentares, o que provavelmente limita a precisão das abordagens padrão de sequenciamento de leituras curtas para detectar o espectro completo de variação genômica dentro desses genes (61,68,97,103,148). No entanto, a tecnologia LRS é nova, de alto custo e atualmente há poucos dados relacionados ao sequenciamento dos segmentos V(D)J. Por essa razão, não foi possível obter LRS para os genomas que compõem este projeto.

Até o presente momento, o banco de dados IMGT é amplamente utilizado em pesquisas devido



à sua ampla disponibilidade, experiência de longa data e estrutura mais completa (75). O número de variantes alélicas do repertório do receptor imune adaptativo aumentou rapidamente no banco de dados IMGT/GENE-DB nos últimos anos (75). Entretanto, seus dados de sequência são, pelo menos em parte, derivados de células B pós-centro germinativo maduras, ou seja, que já sofreram maturação de afinidade nos centros germinativos (75). Para mitigar essa interferência, foi analisado apenas os segmentos VJ genômicos que possuíam igual ou mais de 75% de similaridade com as suas sequências germinativas, para que os segmentos analisados fossem os mais próximos possíveis da sequência pré-maturação de afinidade. Vale ressaltar que, de acordo com os seus respectivos autores, cinco dos sete anticorpos selecionados para o nosso estudo possuem uma conformação tridimensional próxima da sua potencial conformação germinativa, o que também valida a aproximação dos nossos resultados aos eventos ocorridos antes da maturação de afinidade sofrida pelas células B produtoras desses anticorpos. Os anticorpos 7M7W (clones S2X259 e S2H97) e 7R6X (clone S2E12), contra os antígenos do SARS-CoV-2, foram os únicos que sofreram modificações em suas regiões variáveis na etapa de hipermutação somática. Também, foi considerado para a análise somente os *matches* ou ocorrências com mais de 80% de similaridade com os segmentos V(D)J dos anticorpos de interesse.

## CONCLUSÃO

Este trabalho de caráter exploratório identificou uma extensa variação na frequência normalizada dos segmentos VJ, que codificam os CDRs das imunoglobulinas neutralizantes direcionados aos antígenos do *P. falciparum* e SARS-CoV-2, nos genomas antigos em relação aos genomas modernos. Será importante explorar mais genomas antigos com características semelhantes para determinar se nossas descobertas realmente apontam para diferenças nos segmentos gênicos utilizados por humanos antigos em relação aos humanos modernos. A identificação de segmentos VJ que apresentaram variações em sua frequência normalizada entre diferentes populações é uma descoberta relevante, uma vez que sugere o uso estocástico desses segmentos gênicos e pode refletir, por exemplo, os padrões históricos de migração das populações estudadas. Além disso, nosso estudo corrobora a vasta literatura que suporta a grande diversidade genética dos genomas africanos. Com base nos dados científicos disponíveis, este é o primeiro estudo a avaliar as diferenças na frequência e distribuição das sequências gênicas das imunoglobulinas de acordo com a exposição ao *P. falciparum*. Nossos resultados podem estimular pesquisas sobre o uso de sequências germinativas em diferentes populações, com o objetivo de investigar sua natureza protetora e/ou risco de morbidade contra infecções causadas pelo *P. falciparum* e outros patógenos de interesse local ou global.

## LISTA DE FIGURAS

- Figura 1** Distribuição de (a) casos de malária e (b) mortes, por país, 2021. World health statistics 2023: monitoring health for the SDGs, Sustainable Development Goals. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO. ....4
- Figura 2** Malária, ciclo de vida do *Plasmodium spp.* no ser humano e mosquito. Muñoz Suzana Segura, Fernandes Ana Paula Moraes. Principais doenças causadas por protozoários. Vol. 5, Licenciatura em Ciências - USP/UNIVESP. [cited 2023 Oct 19]. p. 1–22. Available from: [https://midia.atp.usp.br/plc/plc0501/impresos/plc0501\\_06.pdf](https://midia.atp.usp.br/plc/plc0501/impresos/plc0501_06.pdf). ....6
- Figura 3** Estimativas de divergência para sequências genômicas e populações ancestrais humanas e neandertais. Também foi demonstrado a estimativa do primeiro caso de malária em ancestrais dos humanos modernos. James P. Noonan et al. Sequencing and Analysis of Neanderthal Genomic DNA. Science 314, 1113 (2006). ....7
- Figura 4** Estrutura do vírus SARS-CoV-2 e visualização do Domínio de Ligação ao Receptor (RBD). Assessoria de comunicação do ICB. Estudo aponta alta probabilidade de surgirem novas variantes mais perigosas da covid-19 nos próximos meses. JORNAL DA USP. 2022 Apr 26 [cited 2023 Oct 19]; Available from: <https://jornal.usp.br/ciencias/estudo-aponta-alta-probabilidade-de-surgirem-novas-variantes-mais-perigosas-da-covid-19-nos-proximos-meses/>. ....9
- Figura 5** Óbitos por COVID-19 relatados por região da OMS, até 11 de março de 2023. World health statistics 2023: monitoring health for the SDGs, Sustainable Development Goals. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO. ....10
- Figura 6** Expansão da capacidade de sequenciamento do SARS-CoV-2 na África. Adams, Janica , et al. The Conundrum of Low COVID-19 Mortality Burden in sub-Saharan Africa: Myth or Reality? Global Health: Science and Practice. 30 de September de 2021, pp. 433-443. ....11
- Figura 7** Mecanismo da recombinação V(D)J. Nishana, Mayilaadumveetil e Raghavan, Sathees C . 2012. Role of recombination activating genes in the generation of antigen receptor diversity and beyond. Immunology. 20 de November de 2012, pp. 271-281. ....14
- Figura 8** Esquema mostrando os diferentes estágios de desenvolvimento das células B onde ocorre o rearranjo dos loci IG. Proudhon, Charlotte , et al. 2015. Chapter Four - Long-Range Regulation of V(D)J Recombination. Advances in Immunology. 20 de August de 2015, pp. 123-182. ....15
- Figura 9** Esquema com as proteínas NHEJ que reparam DSBs e unem as sequências codificadoras liberadas por RAG e as RSSs. Bassing, Craig H, Swat, Wojciech e Al, Frederick W . 2002. The Mechanism and Regulation of Chromosomal V(D)J Recombination. Cell. 2019 de April de 2002, pp. 45-55. ....16
- Figura 10** Esquema de quebras de fita dupla de DNA (DSBs) e seu reparo por junção de extremidade não homóloga (NHEJ). Chang, Howard H. Y. , et al. 2017. Non-homologous DNA end joining and

alternative pathways to double-strand break repair. Nature Reviews Molecular Cell Biology. 17 de May de 2017, pp. 495-506. ....	17
<b>Figura 11</b> Esquema demonstrando as etapas para a junção dos segmentos V(D)J no locus IGH. Population matched (pm) germline allelic variants of immunoglobulin (IG) loci: Relevance in infectious diseases and vaccination studies in human populations. Genes & Immunity. 12 de June de 2021, pp. 172-186. ....	18
<b>Figura 12</b> Uma visão geral dos principais elementos que contribuem para a diversidade de repertório de células B naive. Watson, Corey T. , Glanville, Jacob e Marasco, Wayne A. 2017. The Individual and Population Genetics of Antibody Immunity. Trend in Immunology. 20 de May de 2017, pp. 459-470.....	19
<b>Figura 13</b> “O conceito de classificação” organiza o conhecimento imunogenético útil para nomear e classificar os genes das imunoglobulinas. Lefranc, Marie Paule e Lefranc, Gérard. 2001. The Immunoglobulin Facts Book. San Diego : Academic Press, 2001. ....	21
<b>Figura 14</b> Exemplo de identificação das sequências V(D)J. ....	24
<b>Figura 15</b> Ilustração representativa dos passos realizados durante a análise bioinformática descrita. ....	30
<b>Figura 16</b> Gráficos de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) dos genes V(D)J que codificam as cadeias pesadas (IGH) e leve (IGk ou IGL) de anticorpos neutralizantes contra os antígenos do <i>P. falciparum</i> ou SARS-CoV-2 em diferentes grupos. ....	32
<b>Figura 17</b> Gráficos de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) dos genes V(D)J que codificam as cadeias pesadas (IGH) e leve (IGk ou IGL) de anticorpos neutralizantes contra os antígenos do <i>P. falciparum</i> em diferentes grupos. ...	35
<b>Figura 18</b> Gráficos de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) dos genes V(D)J que codificam as cadeias pesadas (IGH) e leve (IGk ou IGL) de anticorpos neutralizantes contra os antígenos do <i>P. falciparum</i> em diferentes grupos. ...	36
<b>Figura 19</b> Gráficos de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) dos genes V(D)J que codificam as cadeias pesadas (IGH) e leve (IGk ou IGL) de anticorpos neutralizantes contra os antígenos do SARS-CoV-2 em diferentes grupos. ...	39
<b>Figura 20</b> Gráficos de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) dos genes V(D)J que codificam as cadeias pesadas (IGH) e leve (IGk ou IGL) de anticorpos neutralizantes contra os antígenos do SARS-CoV-2 em diferentes grupos. ...	40
<b>Figura 21</b> Gráfico de dispersão intercalado com barras mostrando as frequências normalizadas em escala log e desvios padrão ( $\sigma$ ) do gene da insulina em diferentes grupos. ....	42

<b>Tabela 1</b> Resumo dos antígenos, sua identificação e relevância fisiológica para os patógenos <i>P. falciparum</i> e SARS-CoV-2. ....	<b>25</b>
<b>Tabela 2</b> Sequências germinativas somente das regiões determinantes de complementaridade (CDRs) e a sequência do gene da insulina. ....	<b>28</b>

## REFERÊNCIAS

1. Ji C, Shen H, Su C, Li Y, Chen S, Sharp TH, et al. Plasmodium falciparum has evolved multiple mechanisms to hijack human immunoglobulin M. *Nat Commun.* 2023 May 8;14(1):2650. Available from: <https://www.nature.com/articles/s41467-023-38320-z>
2. Pimenta PFP, Orfano AS, Bahia AC, Duarte APM, Ríos-Velásquez CM, Melo FF, et al. An overview of malaria transmission from the perspective of amazon anopheles vectors. *Mem Inst Oswaldo Cruz.* 2015;110(1):23–47.
3. Tennessen JA, Duraisingh MT. Three hallmarks of malaria-induced selection in human genomes. Available from: <https://doi.org/10.1101/2020.07.07.170092>
4. Phillips MA, Burrows JN, Manyando C, Van Huijsduijnen RH, Van Voorhis WC, Wells TNC. *Malaria. Nat Rev Dis Primers.* 2017 Aug 3;3.
5. Who. World Health Statistics 2023 Monitoring health for the SDGs Sustainable Development Goals HEALTH FOR ALL. 2023. Available from: <https://www.who.int/publications/book-orders>.
6. Leffler EM, Band G, Busby GBJ, Kivinen K, Le QS, Clarke GM, et al. Resistance to malaria through structural variation of red blood cell invasion receptors. *Science (1979).* 2017 Jun 16;356(6343):1140–52.
7. Triglia T, Scally SW, Seager BA, Pasternak M, Dagley LF, Cowman AF. Plasmepsin X activates the PCRCR complex of Plasmodium falciparum by processing PfRh5 for erythrocyte invasion. *Nat Commun.* 2023 Apr 19;14(1):2219. Available from: <https://www.nature.com/articles/s41467-023-37890-2>
8. Molina-Franky J, Patarroyo ME, Kalkum M, Patarroyo MA. The Cellular and Molecular Interaction Between Erythrocytes and Plasmodium falciparum Merozoites. Vol. 12, *Frontiers in Cellular and Infection Microbiology.* Frontiers Media S.A.; 2022.
9. Robert Ménard AASCCRAMR van DCJJAPWRSN& VN. Circumsporozoite protein is required for development of malaria sporozoites in mosquitoes. *Nature.* 1997 Jan 23;(385):336–40.
10. Seaton KE, Spreng RL, Abraha M, Reichartz M, Rojas M, Feely F, et al. Subclass and avidity of circumsporozoite protein specific antibodies associate with protection status against malaria infection. *NPJ Vaccines.* 2021 Dec 1;6(1).
11. Coppi A, Pinzon-Ortiz C, Hutter C, Sinnis P. The Plasmodium circumsporozoite protein is proteolytically processed during cell invasion. *Journal of Experimental Medicine.* 2005 Jan 3;201(1):27–33.
12. Huang WC, Mabrouk MT, Zhou L, Baba M, Tachibana M, Torii M, et al. Vaccine co-display of CSP and Pfs230 on liposomes targeting two Plasmodium falciparum differentiation stages. *Commun Biol.* 2022 Dec 1;5(1).
13. Evert BJ, Chen S, McConville R, Steel RWJ, Healer J, Boddey JA, et al. Epitope-coated polymer particles elicit neutralising antibodies against Plasmodium falciparum sporozoites. *NPJ Vaccines.* 2021 Dec 1;6(1).
14. Chahine Z, Le Roch KG. Decrypting the complexity of the human malaria parasite biology through systems biology approaches. *Frontiers in Systems Biology.* 2022 Sep 16;2.
15. Soulard V, Bosson-Vanga H, Lorthiois A, Roucher C, Franetich JF, Zanghi G, et al. Plasmodium falciparum full life cycle and Plasmodium ovale liver stages in humanized mice. *Nat Commun.* 2015 Jul 24;6.
16. Venugopal K, Hentzschel F, Valkiūnas G, Marti M. Plasmodium asexual growth and sexual development in the haematopoietic niche of the host. Vol. 18, *Nature Reviews Microbiology.* Nature Research; 2020. p. 177–89.
17. Mangou K, Moore AJ, Thiam LG, Ba A, Orfanó A, Desamours I, et al. Structure-guided insights into potential function of novel genetic variants in the malaria vaccine candidate PfRh5. *Sci Rep.* 2022 Dec 1;12(1).
18. Alanine DGW, Quinkert D, Kumarasingha R, Mehmood S, Donnellan FR, Minkah NK, et al. Human Antibodies that Slow Erythrocyte Invasion Potentiate Malaria-Neutralizing Antibodies. *Cell.* 2019 Jun 27;178(1):216–228.e21.
19. Su XZ, Zhang C, Joy DA. Host-Malaria Parasite Interactions and Impacts on Mutual Evolution. Vol. 10, *Frontiers in Cellular and Infection Microbiology.* Frontiers Media S.A.; 2020.
20. Siciliano G, Alano P. Enlightening the malaria parasite life cycle: Bioluminescent Plasmodium in

- fundamental and applied research. Vol. 6, *Frontiers in Microbiology*. Frontiers Media S.A.; 2015.
21. McLeod B, Miura K, Scally SW, Bosch A, Nguyen N, Shin H, et al. Potent antibody lineage against malaria transmission elicited by human vaccination with Pfs25. *Nat Commun*. 2019 Dec 1;10(1).
  22. Amambua-Ngwa A, Button-Simons KA, Li X, Kumar S, Brennenman KV, Ferrari M, et al. Chloroquine resistance evolution in *Plasmodium falciparum* is mediated by the putative amino acid transporter AAT1. *Nat Microbiol*. 2023 May 11; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/37169919>
  23. Wang S, Huang F, Yan H, Yin J, Xia Z. A review of malaria molecular markers for drug resistance in *Plasmodium falciparum* and *Plasmodium vivax* in China. Vol. 13, *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A.; 2023.
  24. Haldar K, Bhattacharjee S, Safeukui I. Drug resistance in *Plasmodium*. Vol. 16, *Nature Reviews Microbiology*. Nature Publishing Group; 2018. p. 156–70.
  25. Abera D, Kibet CK, Degefa T, Amenga-Etego L, Bargul JL, Golassa L. Genomic analysis reveals independent evolution of *Plasmodium falciparum* populations in Ethiopia. *Malar J*. 2021 Dec 1;20(1).
  26. Nkumama IN, Osier FHA. Malaria vaccine roller coaster. Vol. 6, *Nature Microbiology*. Nature Research; 2021. p. 1345–6.
  27. Kaslow DC. Malaria vaccine research & innovation: the intersection of IA2030 and zero malaria. Vol. 5, *npj Vaccines*. Nature Research; 2020.
  28. Triller G, Scally SW, Costa G, Pissarev M, Kreschel C, Bosch A, et al. Natural Parasite Exposure Induces Protective Human Anti-Malarial Antibodies. *Immunity*. 2017 Dec 19;47(6):1197-1209.e10.
  29. Baron JM, Higgins JM, Dzik WH. A revised timeline for the origin of *plasmodium falciparum* as a human pathogen. *J Mol Evol*. 2011 Dec 1;73(5–6):297–304.
  30. Loy DE, Liu W, Li Y, Learn GH, Plenderleith LJ, Sundararaman SA, et al. Out of Africa: origins and evolution of the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. Vol. 47, *International Journal for Parasitology*. Elsevier Ltd; 2017. p. 87–97.
  31. Rich SM, Ayala FJ. Population structure and recent evolution of *Plasmodium falciparum* The Malaria Plague and Control Efforts. Available from: [www.pnas.org](http://www.pnas.org)
  32. Oldest *Homo sapiens* fossil claim rewrites our species' history.
  33. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. Vol. 19, *Nature Reviews Microbiology*. Nature Research; 2021. p. 141–54.
  34. Harrison AG, Lin T, Wang P. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. Vol. 41, *Trends in Immunology*. Elsevier Ltd; 2020. p. 1100–15.
  35. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. Vol. 19, *Nature Reviews Microbiology*. Nature Research; 2021. p. 155–70.
  36. Markov P V., Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of SARS-CoV-2. *Nature Reviews Microbiology*. Nature Research; 2023.
  37. Malone B, Urakova N, Snijder EJ, Campbell EA. Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. Vol. 23, *Nature Reviews Molecular Cell Biology*. Nature Research; 2022. p. 21–39.
  38. WHO COVID-19 Dashboard. 2023. WHO COVID-19 Dashboard. Geneva: World Health Organization, 2020. Available online: <https://covid19.who.int/> (last cited: 16.05.2023).
  39. Toussi SS, Hammond JL, Gerstenberger BS, Anderson AS. Therapeutics for COVID-19. *Nat Microbiol*. 2023 May 4;8(5):771–86. Available from: <https://www.nature.com/articles/s41564-023-01356-4>
  40. Callaway E. THE NEXT GENERATION OF CORONAVIRUS VACCINES. *Nature*. 2023 Feb 2;614:22–5.
  41. Wang MY, Zhao R, Gao LJ, Gao XF, Wang DP, Cao JM. SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. Vol. 10, *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A.; 2020.
  42. Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. Vol. 23, *Nature Reviews Molecular Cell Biology*. Nature Research; 2022. p. 3–20.
  43. Praharaj MR, Garg P, Kesarwani V, Topno NA, Khan RIN, Sharma S, et al. SARS-CoV-2 Spike Glycoprotein and ACE2 Interaction Reveals Modulation of Viral Entry in Wild and Domestic Animals. *Front Med (Lausanne)*. 2022 Mar 11;8.
  44. Cohen FS. How Viruses Invade Cells. Vol. 110, *Biophysical Journal*. Biophysical Society; 2016. p. 1028–32.
  45. Saponaro F, Rutigliano G, Sestito S, Bandini L, Storti B, Bizzarri R, et al. ACE2 in the Era of SARS-

- CoV-2: Controversies and Novel Perspectives. Vol. 7, *Frontiers in Molecular Biosciences*. Frontiers Media S.A.; 2020.
46. Li Q, Liu Y, Zhang L. Cytoplasmic tail determines the membrane trafficking and localization of SARS-CoV-2 spike protein. Vol. 9, *Frontiers in Molecular Biosciences*. Frontiers Media S.A.; 2022.
  47. Scudellari M. HOW THE CORONAVIRUS INFECTS OUR CELLS Scientists are unpicking SARS-CoV-2's life cycle. Vol. 595, | *Nature* |. 2021.
  48. Adams J, Mackenzie MJ, Amegah AK, Ezech A, Gadanya MA, Omigbodun A, et al. The Conundrum of Low COVID-19 Mortality Burden in sub-Saharan Africa: Myth or Reality? Available from: [www.ghspjournal.org](http://www.ghspjournal.org)
  49. Konozy EHE, Osman ME fadil M, Ghartey-Kwansah G, Abushama HM. The striking mimics between COVID-19 and malaria: A review. Vol. 13, *Frontiers in Immunology*. Frontiers Media S.A.; 2022.
  50. Anne Soy. Coronavirus in Africa: Five reasons why Covid-19 has been less deadly than elsewhere. 2020 [cited 2023 Jun 5]. Available from: <https://gdc.unicef.org/resource/coronavirus-africa-five-reasons-why-covid-19-has-been-less-deadly-elsewhere>
  51. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021 Apr 15;592(7854):438–43.
  52. Tegally H, San JE, Cotten M, Moir M, Tegomoh B, Mboowa G, et al. The evolving SARS-CoV-2 epidemic in Africa: Insights from rapidly expanding genomic surveillance. *Science* (1979). 2022 Oct 7;378(6615).
  53. Osei SA, Biney RP, Anning AS, Nortey LN, Ghartey-Kwansah G. Low incidence of COVID-19 case severity and mortality in Africa; Could malaria co-infection provide the missing link? *BMC Infect Dis*. 2022 Dec 1;22(1).
  54. Parodi A, Cozzani E. Coronavirus disease 2019 (COVID 19) and Malaria: Have anti glycoprotein antibodies a role? Vol. 143, *Medical Hypotheses*. Churchill Livingstone; 2020.
  55. Millett GA, Jones AT, Benkeser D, Baral S, Mercer L, Beyrer C, et al. Assessing differential impacts of COVID-19 on black communities. *Ann Epidemiol*. 2020 Jul 1;47:37–44.
  56. Yancy CW. COVID-19 and African Americans. Vol. 323, *JAMA - Journal of the American Medical Association*. American Medical Association; 2020. p. 1891–2.
  57. Dunst J, Azzouz N, Liu X, Tsukita S, Seeberger PH, Kamena F. Interaction between Plasmodium glycosylphosphatidylinositol and the host protein moesin has no implication in malaria pathology. *Front Cell Infect Microbiol*. 2017 May 16;7(MAY).
  58. Naik RS, Branch OH, Woods AS, Vijaykumar M, Perkins DJ, Nahlen BL, et al. Glycosylphosphatidylinositol Anchors of Plasmodium falciparum : Molecular Characterization and Naturally Elicited Antibody Response That May Provide Immunity to Malaria Pathogenesis. Vol. 192, *The Journal of Experimental Medicine* •. 2000. Available from: <http://www.jem.org/cgi/content/full/192/11/1563>
  59. Gomes LR, Martins YC, Ferreira-Da-Cruz MF, Daniel-Ribeiro CT. Autoimmunity, phospholipid-reacting antibodies and malaria immunity. In: *Lupus*. SAGE Publications Ltd; 2014. p. 1295–8.
  60. Tonegawa S. Somatic generation of antibody diversity. 1983.
  61. Gibson WS, Rodriguez OL, Shields K, Silver CA, Dorgham A, Emery M, et al. Characterization of the immunoglobulin lambda chain locus from diverse populations reveals extensive genetic variation. *Genes Immun*. 2023 Feb 1;24(1):21–31.
  62. Hummer AM, Abanades B, Deane CM. Advances in computational structure-based antibody design. Vol. 74, *Current Opinion in Structural Biology*. Elsevier Ltd; 2022.
  63. Stanfield RL, Wilson IA, Crowe JE. *Antibody Structure*. 2014; Available from: [www.rcsb.org](http://www.rcsb.org)
  64. Rom /, Blake CCF. Antibody structure and antigen binding. Vol. 253, *Nature*. 1975.
  65. Horx P, Geyer A. Comparing the Hinge-Type Mobility of Natural and Designed Intermolecular Bisdulfide Domains. *Front Chem*. 2020 Jan 28;8.
  66. Chi X, Li Y, Qiu X. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. Vol. 160, *Immunology*. Blackwell Publishing Ltd; 2020. p. 233–47.
  67. Chiu ML, Goulet DR, Teplyakov A, Gilliland GL. Antibody structure and function: The basis for engineering therapeutics. Vol. 8, *Antibodies*. MDPI; 2019.
  68. Collins AM, Watson CT. Immunoglobulin light chain gene rearrangements, receptor editing and the development of a self-tolerant antibody repertoire. Vol. 9, *Frontiers in Immunology*. Frontiers Media

- S.A.; 2018.
69. Nguyen TV, Pawlikowska P, Firlej V, Rosselli F, Aoufouchi S. V(D)J recombination process and the Pre-B to immature B-cells transition are altered in Fanca <sup>-/-</sup> mice. *Sci Rep.* 2016 Nov 24;6.
  70. Dyda F, Rice PA. A new twist on V(D)J recombination. *Nat Struct Mol Biol.* 2018 Aug 1;25(8):648–9.
  71. Ru H, Chambers MG, Fu TM, Tong AB, Liao M, Wu H. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell.* 2015 Nov 19;163(5):1138–52.
  72. Maity PC, Datta M, Nicolò A, Jumaa H. Isotype Specific Assembly of B Cell Antigen Receptors and Synergism With Chemokine Receptor CXCR4. Vol. 9, *Frontiers in Immunology.* Frontiers Media S.A.; 2018.
  73. Chi X, Li Y, Qiu X. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. Vol. 160, *Immunology.* Blackwell Publishing Ltd; 2020. p. 233–47.
  74. Kim MS, Chuenchor W, Chen X, Cui Y, Zhang X, Zhou ZH, et al. Cracking the DNA Code for V(D)J Recombination. *Mol Cell.* 2018 Apr 19;70(2):358–370.e4.
  75. Khatri I, Berkowska MA, van den Akker EB, Teodosio C, Reinders MJT, van Dongen JJM. Population matched (pm) germline allelic variants of immunoglobulin (IG) loci: Relevance in infectious diseases and vaccination studies in human populations. *Genes Immun.* 2021 Jul 1;22(3):172–86.
  76. Nishana M, Raghavan SC. Role of recombination activating genes in the generation of antigen receptor diversity and beyond. *Immunology.* 2012 Dec;137(4):271–81.
  77. Proudhon C, Hao B, Raviram R, Chaumeil J, Skok JA. Long-Range Regulation of V(D)J Recombination. In: *Advances in Immunology.* Academic Press Inc.; 2015. p. 123–82.
  78. Mansilla-Soto J, Cortes P. VDJ recombination: Artemis and its in vivo role in hairpin opening. Vol. 197, *Journal of Experimental Medicine.* 2003. p. 543–7.
  79. Sankar K, Hoi KH, Hötzel I. Dynamics of heavy chain junctional length biases in antibody repertoires. *Commun Biol.* 2020 Dec 1;3(1).
  80. Niewolik D, Pannicke U, Lu H, Ma Y, Wang LCV, Kulesza P, et al. DNA-PKcs dependence of artemis endonucleolytic activity, differences between hairpins and 5' or 3' overhangs. *Journal of Biological Chemistry.* 2006 Nov 10;281(45):33900–9.
  81. Dai Y, Kysela B, Hanakahi LA, Manolis K, Riballo E, Stumm M, et al. Nonhomologous end joining and V(D)J recombination require an additional factor. 2002. Available from: [www.pnas.org](http://www.pnas.org)
  82. Gullickson P, Xu YW, Niedernhofer LJ, Thompson EL, Yousefzadeh MJ. The Role of DNA Repair in Immunological Diversity: From Molecular Mechanisms to Clinical Ramifications. Vol. 13, *Frontiers in Immunology.* Frontiers Media S.A.; 2022.
  83. Lafaille JJ, Decloux A, Bonneville M, Takagaki Y, Howard ST. Junctional Sequences of T Cell Receptor  $\gamma 6$  Genes: Implications for  $\gamma 6$  T Cell Lineages and for a Novel Intermediate of V-(D)-J Joining. Vol. 59, *Cell.* 1989.
  84. Schreffler I, Desiderio S V, Yancopoulost GD, Paskind M, Thomas E, Boss MA, et al. 13. Bernards, R. <1 a. Vol. 17, *Topics Microbiol. Immun.*
  85. Bashford-Rogers RJM, Smith KGC, Thomas DC. Antibody repertoire analysis in polygenic autoimmune diseases. Vol. 155, *Immunology.* Blackwell Publishing Ltd; 2018. p. 3–17.
  86. Dauba A, Khamlichi AA. Long-Range Control of Class Switch Recombination by Transcriptional Regulatory Elements. Vol. 12, *Frontiers in Immunology.* Frontiers Media S.A.; 2021.
  87. Kurosaki T, Kometani K, Ise W. Memory B cells. *Nat Rev Immunol.* 2015 Mar 27;15(3):149–59.
  88. Su CTT, Ling WL, Lua WH, Poh JJ, Gan SKE. The role of Antibody V $\kappa$  Framework 3 region towards Antigen binding: Effects on recombinant production and Protein L binding. *Sci Rep.* 2017 Dec 1;7(1).
  89. Schroeder HW, Cavacini L. Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology.* 2010 Feb;125(2 SUPPL. 2).
  90. D'Angelo S, Ferrara F, Naranjo L, Erasmus MF, Hrabec P, Bradbury ARM. Many routes to an antibody heavy-chain CDR3: Necessary, yet insufficient, for specific binding. *Front Immunol.* 2018 Mar 8;9(MAR).
  91. Elhanati Y, Sethna Z, Marcou Q, Callan CG, Mora T, Walczak AM. Inferring processes underlying B-cell repertoire diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences.* 2015 Sep 5;370(1676).
  92. Lefranc MP. Antibody nomenclature: From IMGT-ONTOLOGY to INN definition. Vol. 3, *mAbs.* 2011. p. 1–2.



93. Pennell M, Rodriguez OL, Watson CT, Greiff V. The evolutionary and functional significance of germline immunoglobulin gene variation. Vol. 44, Trends in Immunology. Elsevier Ltd; 2023. p. 7–21.
94. Langerak Editor AW. Immuno-genetics Methods and Protocols. Available from: <http://www.springer.com/series/7651>
95. Lefranc MP, Lefranc G. Igt® and 30 years of immunoinformatics insight in antibody v and c domain structure and function. *Antibodies*. 2019 Jun 1;8(2).
96. Mikocziova I, Greiff V, Sollid LM. Immunoglobulin germline gene variation and its impact on human disease. Vol. 22, Genes and Immunity. Springer Nature; 2021. p. 205–17.
97. Watson CT, Steinberg KM, Huddleston J, Warren RL, Malig M, Schein J, et al. Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *Am J Hum Genet*. 2013 Apr 4;92(4):530–46.
98. Lefranc MP, Lefranc G. Immunoglobulins or antibodies: IGT® bridging genes, structures and functions. Vol. 8, Biomedicines. MDPI AG; 2020.
99. Lefranc MP, Lefranc G. IGT® Homo sapiens IG and TR Loci, Gene Order, CNV and Haplotypes: New Concepts as a Paradigm for Jawed Vertebrates Genome Assemblies. *Biomolecules*. 2022 Mar 1;12(3).
100. Crotty S. A brief history of T cell help to B cells. *Nat Rev Immunol*. 2015 Mar 27;15(3):185–9.
101. Mikocziova I, Greiff V, Sollid LM. Immunoglobulin germline gene variation and its impact on human disease. Vol. 22, Genes and Immunity. Springer Nature; 2021. p. 205–17.
102. Lee JH, Toy L, Kos JT, Safonova Y, Schief WR, Havenar-Daughton C, et al. Vaccine genetics of IGHV1-2 VRC01-class broadly neutralizing antibody precursor naïve human B cells. *NPJ Vaccines*. 2021 Dec 1;6(1).
103. Rodriguez OL, Gibson WS, Parks T, Emery M, Powell J, Strahl M, et al. A Novel Framework for Characterizing Genomic Haplotype Diversity in the Human Immunoglobulin Heavy Chain Locus. *Front Immunol*. 2020 Sep 23;11.
104. Collins AM, Yaari G, Shepherd AJ, Lees W, Watson CT. Germline immunoglobulin genes: Disease susceptibility genes hidden in plain sight? Vol. 24, Current Opinion in Systems Biology. Elsevier Ltd; 2020. p. 100–8.
105. Parks T, Mirabel MM, Kado J, Auckland K, Nowak J, Rautanen A, et al. Association between a common immunoglobulin heavy chain allele and rheumatic heart disease risk in Oceania. *Nat Commun*. 2017;8.
106. Mikocziova I, Greiff V, Sollid LM. Immunoglobulin germline gene variation and its impact on human disease. Vol. 22, Genes and Immunity. Springer Nature; 2021. p. 205–17.
107. Pushparaj P, Nicoletto A, Sheward DJ, Das H, Castro Dopico X, Perez Vidakovics L, et al. Immunoglobulin germline gene polymorphisms influence the function of SARS-CoV-2 neutralizing antibodies. *Immunity*. 2023 Jan 10;56(1):193-206.e7.
108. Quintana-Murci L, Clark AG. Population genetic tools for dissecting innate immunity in humans. Vol. 13, Nature Reviews Immunology. 2013. p. 280–93.
109. Manso T, Folch G, Giudicelli V, Jabado-Michaloud J, Kushwaha A, Nguefack Ngoune V, et al. IGT® databases, related tools and web resources through three main axes of research and development. *Nucleic Acids Res*. 2022 Jan 7;50(D1):D1262–72.
110. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*. 2016;538(7624):201–6.
111. Lefranc MP, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, et al. IGT R, the international ImMunoGeneTics information system R 25 years on. *Nucleic Acids Res*. 2015 Jan 28;43(D1):D413–22.
112. Yuan M, Wu NC, Zhu X, Lee CCD, So RTY, Lv H, et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV.
113. Starr TN, Czudnochowski N, Liu Z, Zatta F, Park YJ, Addetia A, et al. SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature*. 2021 Sep 2;597(7874):97–102.
114. Imkeller K, Scally SW, Bosch A, Martí GP, Costa G, Triller G, et al. Antihomotypic affinity maturation improves human B cell responses against a repetitive epitope. Available from: <http://science.sciencemag.org/>
115. Murugan R, Buchauer L, Triller G, Kreschel C, Costa G, Martí GP, et al. Clonal selection drives protective memory B cell responses in controlled human malaria infection. Vol. 3, Sci. Immunol. 2018. Available from: <http://immunology.sciencemag.org/>

116. Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012 Nov 1;491(7422):56–65.
117. Cavalli-Sforza LL. The Human Genome Diversity Project: past, present and future. *Nat Rev Genet*. 2005;6(4):333–40.
118. Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science (1979)*. 2020 Mar 20;367(6484).
119. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature*. 2015 Jan 15;517(7534):327–32.
120. Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, et al. Sequencing and Analysis of Neanderthal Genomic DNA. Available from: [www.sciencemag.org](http://www.sciencemag.org)
121. Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, et al. Targeted retrieval and analysis of five neandertal mtDNA genomes. *Science (1979)*. 2009 Jul 17;325(5938):318–21.
122. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the neandertal genome. *Science (1979)*. 2010 May 7;328(5979):710–22.
123. Prüfer K, De Filippo C, Grote S, Mafessoni F, Korlević P, Hajdinjak M, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science (1979)*. 2017 Nov 3;358(6363):655–8.
124. Mafessoni F, Grote S, De Filippo C, Slon V, Kolobova B □ KA, Viola B, et al. A high-coverage Neandertal genome from Chagyrskaya Cave. Available from: <https://www.pnas.org>
125. Harvati K. 3 Neanderthals and Their Contemporaries.
126. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, et al. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res*. 2013;41(Web Server issue).
127. Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, et al. Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. Available from: [www.sciencemag.org/cgi/content/full/328/5979/723/DC1](http://www.sciencemag.org/cgi/content/full/328/5979/723/DC1)
128. Pollen AA, Kilik U, Lowe CB, Camp JG. Human-specific genetics: new tools to explore the molecular and cellular basis of human evolution. *Nature Reviews Genetics*. *Nature Research*; 2023.
129. Hsieh PH, Dang V, Vollger MR, Mao Y, Huang TH, Dishuck PC, et al. Evidence for opposing selective forces operating on human-specific duplicated TCAF genes in Neanderthals and humans. *Nat Commun*. 2021 Dec 1;12(1).
130. Aqil A, Speidel L, Pavlidis P, Gokcumen O. Balancing selection on genomic deletion polymorphisms in humans. *Elife*. 2023 Jan 1;12.
131. Viscardi LH, Paixão-Côrtes VR, Comas D, Salzano FM, Rovaris D, Bau CD, et al. Searching for ancient balanced polymorphisms shared between Neanderthals and modern humans. *Genet Mol Biol*. 2018 Jan 1;41(1):67–81.
132. Shi B, Dong X, Ma Q, Sun S, Ma L, Yu J, et al. The Usage of Human IGHJ Genes Follows a Particular Non-random Selection: The Recombination Signal Sequence May Affect the Usage of Human IGHJ Genes. *Front Genet*. 2020 Dec 8;11.
133. Oyen D, Torres JL, Wille-Reece U, Ockenhouse CF, Emerling D, Glanville J, et al. Structural basis for antibody recognition of the NANP repeats in *Plasmodium falciparum* circumsporozoite protein. *Proc Natl Acad Sci U S A*. 2017 Nov 28;114(48):E10438–45.
134. Murugan R, Scally SW, Costa G, Mustafa G, Thai E, Decker T, et al. Evolution of protective human antibodies against *Plasmodium falciparum* circumsporozoite protein repeat motifs. *Nat Med*. 2020 Jul 1;26(7):1135–45.
135. Coelho CH, Nadakal ST, Hurtado PG, Morrison R, Galson JD, Neal J, et al. Antimalarial antibody repertoire defined by plasma IG proteomics and single B cell IG sequencing. *JCI Insight*. 2020 Nov 19;5(22).
136. He B, Liu S, Wang Y, Xu M, Cai W, Liu J, et al. Rapid isolation and immune profiling of SARS-CoV-2 specific memory B cell in convalescent COVID-19 patients via LIBRA-seq. *Signal Transduct Target Ther*. 2021 Dec 1;6(1).
137. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Vol. 15, *Nature Reviews Genetics*. Nature Publishing Group; 2014. p. 379–93.
138. Watson CT, Glanville J, Marasco WA. The Individual and Population Genetics of Antibody Immunity. Vol. 38, *Trends in Immunology*. Elsevier Ltd; 2017. p. 459–70.
139. Liston A, Humblet-Baron S, Duffy D, Goris A. Human immune diversity: from evolution to modernity.

- Vol. 22, *Nature Immunology*. Nature Research; 2021. p. 1479–89.
140. Meyer D, Vitor VR, Bitarello BD, Débora DY, Nunes K. A genomic perspective on HLA evolution. Vol. 70, *Immunogenetics*. Springer Verlag; 2018. p. 5–27.
  141. Lenart P, Bienertová-Vášků J, Berec L. Learning mitigates genetic drift. *Sci Rep*. 2022 Dec 1;12(1).
  142. Choudhury A, Ramsay M, Hazelhurst S, Aron S, Bardien S, Botha G, et al. Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. *Nat Commun*. 2017 Dec 1;8(1).
  143. Campbell MC, Tishkoff SA. African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. Vol. 9, *Annual Review of Genomics and Human Genetics*. 2008. p. 403–33.
  144. Ravenhall M, Campino S, Sepúlveda N, Manjurano A, Nadjm B, Mtove G, et al. Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. *PLoS Genet*. 2018 Jan 1;14(1).
  145. Cruz PRS, Ananina G, Gil-da-Silva-Lopes VL, Simioni M, Mena F, Bezerra MAC, et al. Genetic comparison of sickle cell anaemia cohorts from Brazil and the United States reveals high levels of divergence. *Sci Rep*. 2019 Dec 1;9(1).
  146. Shriner D, Rotimi CN. Whole-Genome-Sequence-Based Haplotypes Reveal Single Origin of the Sickle Allele during the Holocene Wet Phase. *Am J Hum Genet*. 2018 Apr 5;102(4):547–56.
  147. Wonkam A, Ponde C, Nicholson N, Fieggen K, Ramesar R, Davidson A. The burden of sickle cell disease in Cape Town. *South African Medical Journal*. 2012;102(9):752–4.
  148. Mantere T, Kersten S, Hoischen A. Long-read sequencing emerging in medical genetics. Vol. 10, *Frontiers in Genetics*. Frontiers Media S.A.; 2019.