UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Departamento de Bioquímica e Imunologia

Yala Sampaio

ANÁLISE MOLECULAR DO REPERTÓRIO DE ANTICORPOS DE CAVALO APÓS IMUNIZAÇÃO COM A PROTEÍNA *SPIKE* DE SARS-COV-2

Belo Horizonte,
2024

Yala Sampaio

ANÁLISE MOLECULAR DO REPERTÓRIO DE ANTICORPOS DE CAVALO APÓS IMUNIZAÇÃO COM A PROTEÍNA S*PIKE* DE SARS-COV-2

Dissertação apresentada ao Programa de Pós-graduação em Bioquímica e Imunologia da Universidade Federal de Minas Gerais, como requisito parcial para à obtenção do título de Mestre em Bioquímica e Imunologia.

Orientadora: Profa. Dra. Liza Felicori
Co-orientadora: Dra. Carlena Navas

Belo Horizonte,
2024

043 Sampaio, Yala.
Análise molecular do repertório de anticorpos de cavalo após imunização com a proteína spike de SARS-COV-2 [manuscrito] / Yala Sampaio. – 2024.
95 f. : il. ; 29,5 cm.

Orientadora: Profa. Dra. Liza Felicori. Co-orientadora: Dra. Carlena Navas.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Bioquímica e Imunologia.

1. Bioquímica e imunologia. 2. Anticorpos. 3. Imunoglobulinas. 4. Betacoronavirus. 5. Cavalos. I. Vilela, Liza Figueiredo Felicori. II. Reyes, Carlena Tahina Navas de. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 577.1

Ficha catalografica elaborada pela bibliotecária Fabiane C M Reis - CRB 6 2680

Universidade Federal de Minas Gerais
Curso de Pós-Graduação em Bioquímica e Imunologia ICB/UFMG
Av. Antônio Carlos, 6627 – Pampulha
31270-901 – Belo Horizonte – MG
e-mail: pg-biq@icb.ufmg.br (31)3409-2615

ATA DA DEFESA DA DISSERTAÇÃO DE MESTRADO DE YALA SAMPAIO. Aos vinte e três dias do mês de fevereiro de 2024 às 13:00 horas, reuniu-se no Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, a Comissão Examinadora da dissertação de Mestrado, indicada *ad referendum* do Colegiado do Curso, para julgar, em exame final, o trabalho intitulado "Análise molecular do repertório de anticorpos de cavalo após imunização com a proteína spike de SARS-CoV-2 ", requisito final para a obtenção do grau de Mestre em Bioquímica e Imunologia, área de concentração: Bioquímica. Abrindo a sessão, a Presidente da Comissão, Profª. Liza Figueiredo Felicori Vilela, da Universidade Federal de Minas Gerais, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores, com a respectiva defesa da candidata. Logo após a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações: Dra. Glória Regina Franco (Universidade Federal de Minas Gerais), aprovada; Dr. Diogo Magnani ((UMAss - Boston), aprovada; Dra. Carlena Tahina Navas de Reyes - Coorientadora (Universidade Federal de Minas Gerais), aprovada; Dra. Liza Figueiredo Felicori Vilela - Orientadora (Universidade Federal de Minas Gerais), aprovada. Pelas indicações a candidata foi considerada:

(x) APROVADA
( ) REPROVADA

O resultado final foi comunicado publicamente à candidata pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente da Comissão encerrou a reunião e lavrou a presente Ata que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 23 de fevereiro de 2024.

Dra. Glória Regina Franco (UFMG) gov.br Documento assinado digitalmente GLORIA REGINA FRANCO Data: 05/03/2024 11:43:16-0300 Verifique em https://validar.iti.gov.br

Dr. Diogo Magnani ((UMAss - Boston)

Dra. Carlena Tahina Navas de Reyes - Coorientadora (UFMG)

Dra. Liza Figueiredo Felicori Vilela - Orientadora (UFMG) gov.br Documento assinado digitalmente LIZA FIGUEIREDO FELICORI VILELA Data: 26/02/2024 09:35:02-0300 Verifique em https://validar.iti.gov.br

Dedico esse trabalho à Dona Mãe, por sempre acreditar em mim.

AGRADECIMENTOS

À CNPq, expresso minha sincera gratidão pela concessão da bolsa de estudos e pelo valioso financiamento que viabilizaram a realização deste trabalho.

À CAPES, FAPEMIG e INOVATOX, pelo apoio essencial nesse projeto.

Ao Instituto de Ciências Biológicas e ao Programa de Pós-Graduação em Bioquímica e Imunologia, agradeço pelo programa de excelência que contribuiu para minha formação.

À minha orientadora Dra. Liza Felicori, expresso minha profunda gratidão pela sua animação e entusiasmo pela ciência, bem como pela orientação valiosa ao longo deste trabalho.

À minha co-orientadora Dra. Carlena Navas, agradeço pela paciência e pela parceria essencial para o desenvolvimento deste trabalho.

Ao Doutorando João Gervásio, meu agradecimento por estar sempre disponível para me auxiliar e ensinar sobre bioinformática.

Aos colegas de laboratório, especialmente Glória, Luciana, Manuela, Marcele e Marcella, agradeço pelo trabalho conjunto e dedicação ao desenvolvimento deste projeto. Aos demais membros, Camila, Cristina, Igor, Juan, Juliana, Laís, Lígia, Raniele e Regina minha gratidão pela companhia diária, pelas conversas e pela troca constante de ideias.

À minha família, expresso minha profunda gratidão pelo apoio ao meu caminho científico. Em especial, à Dona Mãe, agradeço por acreditar diariamente na minha capacidade e por ser uma fonte constante de motivação.

Aos meus amigos, meu sincero agradecimento por estarem sempre ao meu lado, compartilhando as alegrias e desafios desta jornada científica.

RESUMO

Os anticorpos derivados de cavalos desempenham um papel crucial no tratamento de diversas doenças há mais de um século. No Brasil, o Instituto Vital Brazil e o Instituto Butantan têm desempenhado um papel significativo na produção de anticorpos terapêuticos. Dentre estes, estudos tem utilizado o plasma de cavalos como estratégia no enfrentamento da COVID-19. Foi mostrado que os plasmas desses animais podem ser 150 vezes mais neutralizantes do que os plasmas de pacientes convalescentes. No entanto, nossa compreensão do sistema imunológico equino, incluindo os anticorpos equinos, ainda é muito limitada.

Diante da elevada neutralização do plasma equino, isolar anticorpos do mesmo, poderia levar a uma alternativa terapêutica de combate à COVID-19. Por isso, este trabalho tem como objetivo aprofundar a compreensão dos anticorpos transcritos e circulantes contra a proteína spike de SARS-CoV-2 de equinos com o objetivo de desenvolver novos imunoterapêuticos.

Neste estudo, adotamos uma abordagem integrada, empregando o sequenciamento dos transcritos dos receptores de células B (Rep-seq) para criar um banco de sequências. Essas sequências foram então usadas para mapear os peptídeos dos anticorpos específicos anti-spike presentes no plasma (Ig-Seq), permitindo a identificação de sequências completas dos anticorpos circulantes.

O sequenciamento NGS dos transcritos de 4 cavalos resultou em uma média de 57.000 clones para a cadeia pesada (IGH), 71.000 para a cadeia leve lambda (IGL) e 8.000 para a cadeia leve kappa (IGK). Através da sorologia, identificamos 106 sequências específicas para IGH, 49 para IGL e 19 para IGK. Além disso, conseguimos determinar que as características dos anticorpos *spike*-específicos são muito similares àquelas do repertório transcrito total.

As cadeias pesadas (19) e leves (2) do cavalo que apresentou o maior título neutralizante foram analisadas e selecionadas para produção de anticorpos. Assim, como perspectiva, estes anticorpos serão produzidos e testados.

Este estudo é o primeiro a revelar a composição singular dos anticorpos equinos circulantes anti-spike, destacando características distintas em comparação com os anticorpos previamente caracterizados contra o SARS-CoV-2. Essa

descoberta abre promissoras perspectivas para a criação de anticorpos equinos terapêuticos destinados ao combate da COVID-19.

Palavras-chaves: anticorpos; imunoglobulinas; Sars-CoV-2; cavalo.

ABSTRACT

Antibodies derived from horses have played a crucial role in the treatment of various diseases for over a century. In Brazil, the Vital Brazil Institute and the Butantan Institute have played a significant role in the production of therapeutic antibodies. Among these, studies have used horse plasma as a strategy to fight COVID-19. It has been shown that plasma from these animals can be 150 times more neutralizing than plasma from convalescent patients. However, our understanding of the equine immune system, including equine antibodies, is still very limited.

Given the high neutralization of equine plasma, isolating antibodies from it could lead to a therapeutic alternative to combat COVID-19. Therefore, this work aims to deepen the understanding of transcribed and circulating antibodies against the SARS-CoV-2 spike protein in horses with the aim of developing new immunotherapeutics.

In this study, we adopted an integrated approach, employing the sequencing of B-cell receptor transcripts (Rep-seq) to create a bank of sequences. These sequences were then used to map the peptides of the specific anti-spike antibodies present in plasma (Ig-Seq), allowing the identification of complete sequences of circulating antibodies.

NGS sequencing of transcripts from 4 horses resulted in an average of 57,000 clones for the heavy chain (IGH), 71,000 for the lambda light chain (IGL) and 8,000 for the kappa light chain (IGK). Through serology, we identified 106 specific sequences for IGH, 49 for IGL and 19 for IGK. In addition, we were able to determine that the characteristics of the spike-specific antibodies are very similar to those of the total transcript repertoire.

The heavy (19) and light (2) chains of the horse that showed the highest neutralizing title were analyzed and selected for antibody production. As a prospect, these antibodies will be produced and tested.

This study is the first to reveal the unique composition of circulating equine anti-spike antibodies, highlighting distinct characteristics compared to previously characterized antibodies against SARS-CoV-2. This discovery opens up promising prospects for the creation of therapeutic equine antibodies to combat COVID-19.

Keywords: antibodies; immunoglobulins; Sars-CoV-2; horse.

LISTA DE FIGURAS

Figura 1 - Estrutura dos anticorpos. 18
Figura 2 – Método de análise do repertório de anticorpos. 20
Figura 3 – Representação esquemática do *locus* de IGH do cavalo no cromossomo 24 25
Figura 4 – Representação esquemática do *locus* de IGK de cavalo no cromossomo 15 27
Figura 5 – Desenho dos primers *forward* e *reverse* da cadeia pesada, com um exemplo ilustrativo do processo de desenho.....32
Figura 6 – Representação gráfica da amplificação e sequenciamento da região variável de IG de cavalos com primers específicos.....35
Figura 7 – Workflow para análise do repertório de anticorpos por NGS e LC-MS/MS 37
Figura 8 – Identificação de peptídeos de anticorpos específicos contra a proteína *spike* 42
Figura 9 – Representação do fluxo de trabalho utilizado para a análise do Rep-Seq e Ig-Seq.....43
Figura 10 – Frequência relativa média dos genes IGH anotados mais frequentes ..49
Figura 11 – Frequência relativa média dos genes IGL anotados mais frequentes ...50
Figura 12 – Frequência relativa média dos genes IGK anotados mais frequentes...51
Figura 13 – Comparação da frequência média da hipermutação somática por cadeia em transcritos de anticorpos e anticorpos circulantes. 52
Figura 14 – Frequência relativa média do tamanho (em aminoácidos) do CDR3 da cadeia pesada. 53
Figura 15 – Frequência relativa média da composição dos aminoácidos no CDR3-H. 54
Figura 16 – Frequência relativa média do tamanho (em aminoácidos) do CDR3 da cadeia leve *lambda* 55
Figura 17 – Frequência relativa média da composição dos aminoácidos no CDR3-L 55
Figura 18 – Frequência relativa média do tamanho (em aminoácidos) do CDR3 da cadeia leve *kappa* 56

Figura 19 – Frequência relativa média da composição dos aminoácidos no CDR3-K. 57
Figura 20 – Distribuição de resíduos de aminoácidos por posição das 19 sequências da cadeia pesada encontradas no cavalo 843-I. 60
Figura 21 – Distribuição de resíduos de aminoácidos por posição das 2 sequências da cadeia leve *lambda* encontradas no cavalo 843-I.. 60
Figura 22 – Uso gênico da cadeia pesada dos anticorpos do cavalo 843-I.....61
Figura 23 – Alinhamento estrutural dos 38 possíveis anticorpos específicos, alinhados conforme as regiões de FR.. 62
Figura 24 – Alinhamento das 19 sequencias IGH utilizando Clustal Omega.....63

LISTA DE TABELAS

Tabela 1 – Primers *forward* e *reverse* (5'-3') utilizados na PCR Multiplex para amplificação das cadeias dos anticorpos equinos.....33
Tabela 2 – Resumo dos resultados do sequenciamento da cadeia pesada dos quatro cavalos imunizados com a proteína *spike* do SARS-CoV-2. 45
Tabela 3 – Resumo dos resultados do sequenciamento da cadeia leve (*kappa* e *lambda*) dos quatro cavalos imunizados com a proteína *spike* do SARS-CoV-2. 46
Tabela 4 – Número de peptídeos, CDR3 e sequências (IGH, IGK e IGL) identificadas específicas contra a proteína *spike**.....47
Tabela 5 – Título de neutralização dos plasmas dos cavalos. 57
Tabela 6 – Peptídeos específicos (IGH e IGL) encontrados no cavalo 843-I. 59

SUMÁRIO

1 INTRODUÇÃO 16

1.1 Coronavírus 16

1.2 Função e estrutura dos anticorpos 16

1.3 Repertório de anticorpos 19

1.4 Metodologias para análise do repertório de anticorpos 19

1.4.1 Análise do repertório transcrito de anticorpos 19

1.4.2 Análise do repertório de anticorpos circulantes 21

1.5 Anticorpos para terapia contra COVID-19 22

1.6 Anticorpos equinos para terapia contra a COVID-19 23

2 JUSTIFICATIVA 29

2.1 Objetivo Geral 29

2.2 Objetivos Específicos 30

3 MATERIAIS E MÉTODOS 31

3.1 Obtenção de PBMCs e plasma de cavalos 31

3.2 Sequenciamento do Repertório dos Receptores de Células B, Rep-Seq 31

3.2.1 Extração de RNA 31

3.2.2 Síntese de cDNA 32

3.2.3 Amplificação das cadeias leves e pesadas de anticorpos de cavalo por PCR Multiplex 32

3.2.4 Preparação da biblioteca para sequenciamento 34

3.2.5 Sequenciamento 35

3.3 Sequenciamento das Imunoglobulinas Presente no Plasma dos Cavalos Imunizados, IG-Seq 35

3.3.1 Isolamento das Imunoglobulinas G (IgG) 35

3.3.2 Obtenção dos Fragmentos F(ab')^2 36

3.3.3 Seleção dos anticorpos específicos 36
3.3.4 Preparo para espectrometria 36
3.3.5 Sequenciamento dos peptídeos dos anticorpos específicos 36
3.4 Análise bioinformática do repertório de receptores de células B (Rep-Seq) 37
3.4.1 Pré-processamento das *reads* 38
3.4.2 Anotação dos segmentos V(D)J 39
3.4.2.1 Cadeia Pesada 39
3.4.2.2 Cadeia Leve 39
3.4.3 Clonotipagem 40
3.4.3.1 Cadeia Pesada 40
3.4.3.2 Cadeia Leve 40
3.5 Análise bioinformática do repertório das Imunoglobulinas específicas circulantes (Ig-Seq) 41
3.5.1 Identificação de peptídeos oriundos da LC-MS/MS no banco de sequências de anticorpos anotados 41
3.5.2 Mapeamento dos peptídeos nos CDR3 41
3.6 Seleção dos anticorpos específicos contra proteína *spike* para expressão *in vitro* 44
3.6.1 Modelagem e clusterização estrutural 44
3.7 Desenho dos genes que codificam os anticorpos 44
4 RESULTADOS 45
4.1 Análise dos dados do sequenciamento 45
4.1.1 Anotação e obtenção de clones da cadeia pesada de anticorpos equinos 45
4.1.2 Anotação e obtenção de clones da cadeia leve de anticorpos equinos 46

4.2 Análise dos dados Ig-Seq para identificação de peptídeos específicos anti-*spike* 47
4.3 Análise comparativa entre o repertório dos transcritos (Rep-Seq) e do repertório sorológico (Ig-Seq) 48
4.3.1 Análise dos segmentos gênicos mais frequentes em anticorpos equinos 48
4.3.2 Análise da frequência da hipermutação somática (SHM) nos anticorpos equinos 51
4.3.3 Análise do tamanho e composição dos CDR3 dos anticorpos equinos 52
4.4 Estudo de Caso (Amostra 843-I) 57
4.4.1 Uso gênico dos anticorpos específicos 61
4.4.2 Modelagem estrutural de sequências de anticorpos específicas para a proteína *spike* 61
4.4.3 Seleção das sequências para síntese 62
5 DISCUSSÃO 64
6 LIMITAÇÃO DO ESTUDO 70
7 CONCLUSÃO 71
8 PERSPECTIVAS 72
REFERÊNCIAS 73
ANEXO I 84
ANEXO II 86
ANEXO III 88
ANEXO IV 90

1 INTRODUÇÃO

1.1 Coronavírus

O coronavírus humano (hCoVs) foi identificado pela primeira vez em 1960, quando seu agente infeccioso foi encontrado no trato respiratório de humanos **(Kahn and McIntosh, 2005)**. Recentemente, o Coronavírus da Síndrome Respiratória Aguda Grave 2 (também conhecido como SARS-CoV-2) desencadeou uma pandemia global.

Atualmente a vacinação é a principal estratégia de prevenção contra as formas graves da doença causada pelo coronavírus, enquanto as terapias com anticorpos são consideradas uma alternativa terapêutica com grande potencial promissor **(Widyasari and Kim, 2023)**.

Ao contrário da imunidade adquirida por meio da vacinação, que se desenvolve gradualmente ao longo do tempo, a imunidade gerada pela administração de anticorpos é imediata e passiva, com um grande potencial para reduzir imediatamente a progressão da doença, hospitalizações e mortes. Hoje, a imunidade passiva envolve a infusão de anticorpos monoclonais ou policlonais antígeno-específico derivados de pacientes convalescentes ou soro de animais **(Taylor et al., 2021)**.

1.2 Função e estrutura dos anticorpos

Os anticorpos estão sendo amplamente utilizados para diferentes terapias devido ao papel fundamental que desempenham no sistema imunológico adaptativo, fornecendo uma defesa eficaz contra infecções causadas por vírus, bactérias, diversos patógenos e antígenos diferentes. Essas proteínas são produzidas pelas células B, sendo secretadas em diversas formas, cada uma com sequências de aminoácidos distintas e com diferentes afinidades aos antígenos **(Alberts et al., 2002)**. Os anticorpos podem desempenhar um papel crucial na prevenção da ligação ou fusão de vírus com as células hospedeiras, um processo conhecido como neutralização. Além disso, facilitam a opsonização dos antígenos, tornando-os mais facilmente reconhecidos e fagocitados por células do sistema imunológico **(Snapkov et al., 2022)**.

Em humanos e na maioria dos mamíferos a estrutura básica dos anticorpos possui forma de "Y" **(Figura 1)**, sendo composto por quatro cadeias polipeptídicas, duas cadeias pesadas (H) e duas cadeias leves (L). Essas cadeias são conectadas por pontes dissulfeto, formando a característica configuração em "Y". As regiões superiores dos braços do "Y" são altamente variáveis e são as principais responsáveis

pelo reconhecimento a antígenos, como patógenos e outras substâncias estranhas ao organismo. Já a região constante, confere atividade funcional à molécula e é responsável pela ligação em moléculas efetoras que desencadeiam vias de sinalização para eliminar o antígeno que foi reconhecido **(Murphy, 2014)**. A manutenção da estrutura é assegurada pelas interações não covalentes e covalentes (pontes dissulfeto**)(Alberts et al., 2002)**.

Ambas as cadeias e suas respectivas regiões variáveis, região variável da cadeia leve (V_L) e variável da cadeia pesada (V_H) combinam-se e formam a região de interação com o antígeno. A região variável da cadeia pesada (IGH) é composta pela fusão dos segmentos gênicos *variable* (IGHV), *diversity* (IGHD) e *joining* (IGHJ). Já as cadeias leves podem ser do tipo cadeia leve *lambda* (IGL) ou cadeia leve *kappa* (IGK), que são expressas em diferentes proporções dependendo da espécie. As regiões variáveis das cadeias leves são compostas pela fusão apenas pelos segmentos gênicos *variable* (IGLV ou IGKV) e pelo segmento *joining* (IGLJ ou IGKJ)**(Murphy, 2014).**

As regiões variáveis **(Figura 1)** são formados por regiões muito variáveis entre os diferentes anticorpos, conhecidas como hipervariáveis ou regiões determinantes de complementariedade (CDR, do inglês *Complementary Determining Regions)* e por regiões conservadas, que são as regiões estruturais chamadas de *Framework* (FR). As regiões de CDRs, 3 na cadeia pesada e 3 na cadeia leve, são mais comumente envolvidas na interação ao antígeno, do que as regiões conservadas de FR, que desempenham um papel de arcabouço estrutural nos anticorpos. O CDR-H3 (CDR 3 da cadeia pesada), possui uma maior diversidade, uma vez que é resultado da combinação dos segmentos V, D e J **(Snapkov et al., 2022)**.

Vale mencionar que, em algumas espécies, encontramos variações notáveis nas estruturas dos anticorpos. Os camelos, por exemplo, possuem anticorpos compostos exclusivamente por cadeias pesadas. Já os bovinos se destacam por suas regiões do CDR-H3, que são notavelmente longas, formando um pequeno domínio independente **(Muyldermans and Smider, 2016).**

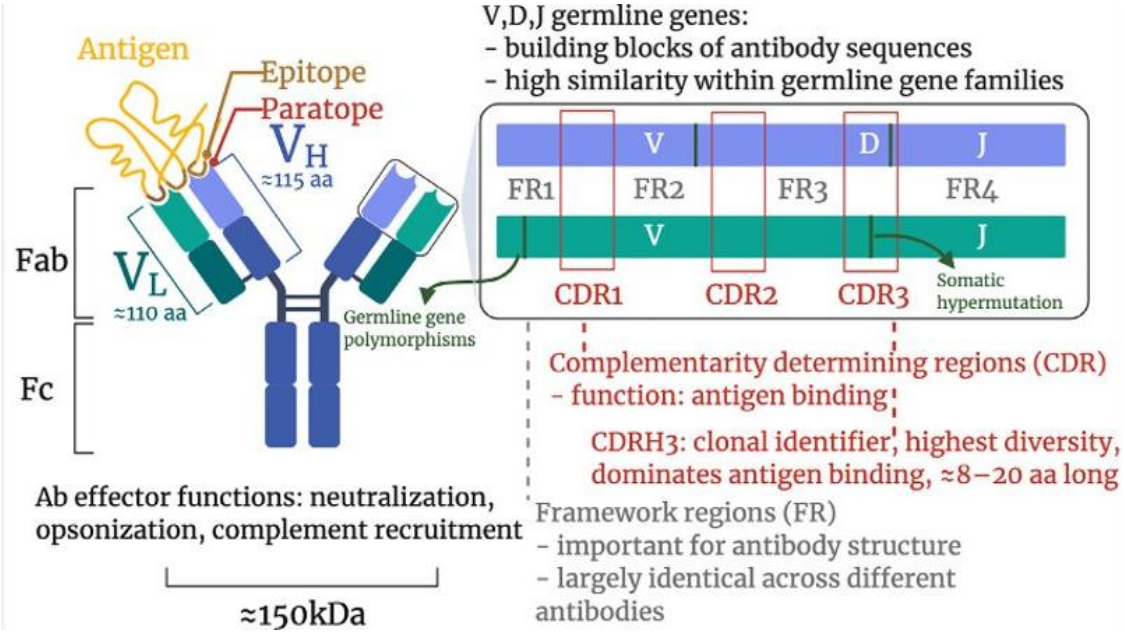

Figura 1 - Estrutura dos anticorpos. A unidade básica é composta por duas cadeias leves idênticas (em verde) e duas cadeias pesadas idênticas (em azul), que são mantidas juntas por ligações dissulfeto que possui forma de "Y". Cada cadeia é composta por uma região variável (V) e uma região constante (C). A região variável é composta pela recombinação dos segmentos gênicos V(D)J, que são formados por regiões hipervariáveis, conhecidas como CDRs, mais comumente implicadas na interação ao antígeno e são separadas por regiões conservadas, conhecidas como *framework* (FR), com função de manter a estrutura do anticorpo **(Snapkov et al., 2022).**

Uma característica fundamental para que os anticorpos possam exercer sua função é a diversidade. Dentre os diversos mecanismos para gerar a diversidade de imunoglobulinas podemos citar o mecanismo primário que envolve a diversidade combinatória dos segmentos gênicos V(D)J e a diversidade juncional, onde há inserções ou exclusões de nucleotídeos entre os segmentos gênicos **(Tonegawa, 1983)**. Além disso, existem mecanismos secundários, como a Hipermutação Somática (SMH, do inglês *Somatic Hypermutation*) e a troca de isotipos, que podem aumentar ainda mais a diversidade **(Walther et al., 2015)**. Em humano, esses mecanismos de diversificação podem produzir em teoria 10^{16} à 10^{18} anticorpos diferentes **(Briney et al., 2019),** número teórico que excede o número total de linfócitos B no corpo humano (~2-4x10^{9}) **(Georgiou et al., 2014)**.

O conjunto diversificado de anticorpos que um organismo pode originar é denominado de repertório de anticorpos.

1.3 Repertório de anticorpos

O repertório de anticorpos é caracterizado por sua heterogeneidade e especificidade. Cada anticorpo pode reconhecer e se ligar a um antígeno específico, desencadeando uma resposta imunológica adaptativa. O repertório é moldado ao longo da vida do organismo por meio de processos como a recombinação genética, mutação somática e seleção clonal, permitindo a geração de uma vasta gama de anticorpos com diferentes especificidades.

O estudo do repertório dos anticorpos, que tem como objetivo compreender essa rica diversidade é fundamental para desvendar os mecanismos da resposta imune e abrir portas para avanços na medicina e na pesquisa básica, como: estudar o desenvolvimento das células B e mapear sua jornada desde a maturação até a produção de anticorpos específicos; gerar anticorpos recombinantes com alto potencial terapêutico, combatendo doenças com precisão e eficiência; desenvolver vacinas mais eficazes, induzindo uma resposta imune robusta e direcionada contra patógenos específicos; descobrir antígenos e epítopos de interesse, expandindo o conhecimento sobre doenças e abrindo caminho para novas abordagens terapêuticas **(Calis and Rosenberg, 2014; Georgiou et al., 2014; Lee et al., 2016; Mathonet and Ullman, 2013; Yaari and Kleinstein, 2015).**

1.4 Metodologias para análise do repertório de anticorpos

1.4.1 Análise do repertório transcrito de anticorpos

Diferentes metodologias de amplificação podem ser aplicadas para o estudo do repertório dos anticorpos. No entanto, a maioria se baseia no isolamento de células B, provenientes de órgãos linfóides ou sangue periférico para a purificação do DNA genômico (gDNA) ou mRNA. Posteriormente, o repertório é amplificado por meio de *primers* específicos para enriquecer as amostras, seguido de sequenciamento dos *amplicons* **(Figura 2)**.

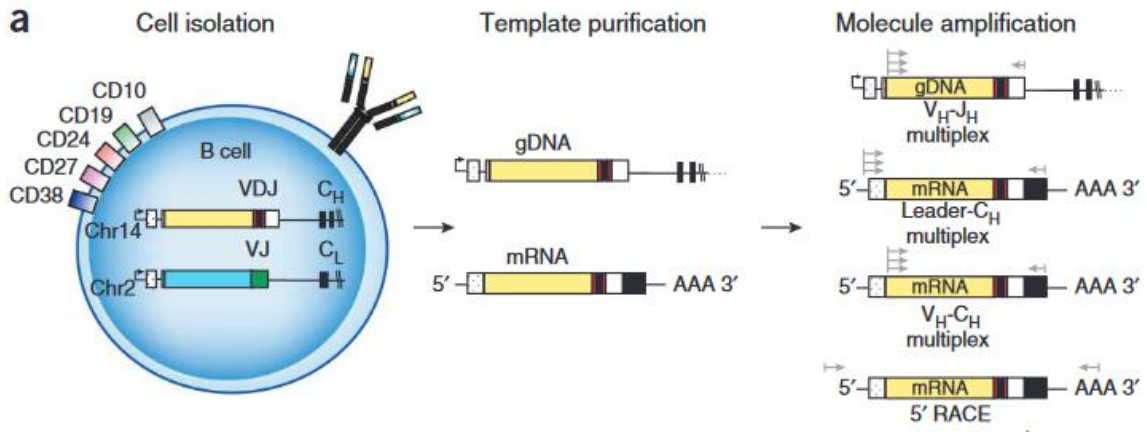

Figura 2 – Método de análise do repertório de anticorpos. Inicialmente é realizado o isolamento das células B totais ou de uma subpopulação de interesse, seguido por amplificação do gDNA ou do mRNA utilizando primers específicos. Figura adaptada **(Georgiou et al., 2014).**

Na década de 90 e início dos anos 2000, sequenciamento por Sanger permitiu a determinação dos segmentos V(D)J das cadeias IGH e IGL a partir de mRNA ou DNA **(Georgiou et al., 2014; Robinson, 2015)** em células únicas em camundongos **(Ehlich et al., 1994)** e em humanos **(Küppers et al., 1993)**. Essa metodologia permitiu a análise em baixa profundidade do repertório, uma vez que é possível sequenciar apenas algumas centenas de clones.

Apesar da baixa profundidade, essa técnica permitiu que fossem identificados anticorpos neutralizantes contra diversos patógenos, incluindo HIV **(Scheid et al., 2009)** e influenza **(Corti et al., 2011)**. No entanto, esse tipo estudo do repertório das células B não fornece uma visão ampla do repertório, devido à alta diversidade do repertório de imunoglobulinas.

Posteriormente, com a disseminação do NGS, novas abordagens em larga escala foram desenvolvidas. A técnica de sequenciamento em larga escala do repertório de transcritos de anticorpos, conhecida como Rep-Seq, tem se beneficiado do NGS, permitindo a geração de milhões a bilhões de pares de bases em uma única análise

Diversos estudos tem aplicado essa metodologia para analisar o repertório de anticorpos em uma variedade de espécies, incluindo humanos **(DeKosky et al., 2016; Lavinder et al., 2014; McDaniel et al., 2016)**, cavalos **(Manso et al., 2019; Navas et al., 2022)**, galinha **(Wu et al., 2012)**, vacas **(de los Rios et al., 2015)**, cachorro **(Steiniger et al., 2014)** e gato **(Steiniger et al., 2017)**.

Contudo, apesar dos estudos dos transcritos de anticorpos serem extremamente relevantes, eles não refletem as características das imunoglobulinas secretadas encontradas no repertório circulante, visto que apenas parte dos transcritos de imunoglobulinas serão secretadas.

1.4.2 Análise do repertório de anticorpos circulantes

A partir de 2008, começou a ser utilizado espectrometria de massas acoplada com cromatografia liquida (LC-MS/MS do inglês *Liquid Chromatography – Mass Spectrometry*) na tentativa de detectar anticorpos circulantes. No entanto, os fragmentos trípticos de imunoglobulinas encontrados nesse estudo não forneciam informações o suficiente para reunir as sequências completas dos anticorpos **(Obermeier et al., 2008)**.

Em 2012, Sato e Cheung utilizaram pela primeira vez a abordagem proteômica em conjunto com o sequenciamento de nova geração de trascriptoma para a descoberta de anticorpos sorológicos antígeno-específico **(Cheung et al., 2012; Sato et al., 2012)**. Sendo assim, novos estudos baseados na imunoproteogenômica, tem utilizado uma abordagem que combina a transcriptômica com proteômica para determinar repertório de anticorpos em plasma, soro, líquido cefalorraquidiano, biopsias e outros **(Bonissone et al., 2020)**.

É importante mencionar que à medida que as tecnologias de sequenciamento continuaram a melhorar, passaram a produzir conjuntos de dados cada vez maiores, com dezenas a centenas de milhões de sequências. Esses dados exigem que *pipelines* especializados de bioinformática. Vários métodos e ferramentas foram desenvolvidos para lidar com diferentes etapas da análise, e conjuntos de software foram integrados. No entanto, ainda não há consenso na utilização dos *pipelines* para o processamento, anotação e análise de dados. Também faltam formatos de arquivo comuns para compartilhamento de dados **(Yaari and Kleinstein, 2015)**.

Outro ponto importante é que, embora essas abordagens acima mencionadas proporcionem uma visão abrangente dos transcritos dos receptores de células B e de cadeias de anticorpos circulantes, elas apresentam uma limitação importante: não é possível encontrar o par correspondente da cadeia pesada e leve que formará o anticorpo, o que dificulta a síntese e teste destas moléculas. Os dados gerados para cada cadeia são independentes.

1.5 Anticorpos para terapia contra COVID-19

Anticorpos tem sido utilizado como terapia para o combate à COVID-19, e grandes progressos foram alcançados na pesquisa e desenvolvimento de anticorpos monoclonais (mAbs) direcionados ao SARS-CoV-2. Em sua última atualização em junho de 2023, o CoV-AbDab (do inglês, *Coronavirus Antibody Database)*, um banco de dados de anticorpos anti-coronavírus, revela que há 12.536 anticorpos disponibilizado em seu banco de dados **(Raybould et al., 2021).**

Desses anticorpos, diversos ensaios clínicos foram conduzidos, inicialmente para o tratamento de pacientes hospitalizados, todos apresentando eficácia moderada e boa segurança **(Focosi et al., 2022)**. Nos Estados Unidos, o FDA (*Food and Drug Administration*) aprovou o uso dos anticorpos monoclonais Bamlanivimab e Etesevimab (LY-CoV555 e LY-CoV016), seguido pela aprovação do Casirivimab e Imdevimabe (REGN-COV2), Sotrovimab (VIR-7831), e Bebtelovimab (Tixagevimab e Cilgavimab). No Brasil, a ANVISA (Agência Nacional de Vigilância Sanitária) também aprovou o uso da combinação de Cassirodovimabe e Imdevimabe (REGN-COV2) e Sotrovimab (VIR-7831), além da combinação de Tixagevimab e Cilgavimab (AZD7442).

Entretanto, com o surgimento das novas variantes em 2022, tanto o FDA quanto a OMS (Organização Mundial da Saúde) suspenderam o uso da maioria dos anticorpos monoclonais, devido à constatação de que não ofereciam uma resposta eficaz contra as novas variantes do vírus, especialmente a Omicrôn, que estava em circulação na época **(Kumari et al., 2022; Tuekprakhon et al., 2022)**.

Sendo assim, torna-se um grande desafio a busca de um anticorpo com capacidade ampla de neutralização, que tem como objetivo os epítopos compartilhados por todas ou pela maioria das variantes. Ye e colaboradores (2020) demonstraram a eficácia da terapia utilizando plasmas de pacientes convalescentes de COVID-19, se tornando uma terapia promissora **(Ye et al., 2020)**. No entanto, em dezembro de 2021, a OMS passou a não recomendar o uso de plasmas de pacientes convalescentes, uma vez que um estudo de revisão sistemática demonstrou que há evidencias que o plasma reduza a mortalidade **(Iannizzi et al., 2023).** Desta maneira, a busca por terapêuticos eficazes contra a COVID-19 continua.

1.6 Anticorpos equinos para terapia contra a COVID-19

Os anticorpos provenientes de cavalos têm desempenhado um papel fundamental no tratamento de diversas doenças e toxinas há mais de um século. A prática da soroterapia, estabelecida em 1880 por Emil von Behring, demonstrou que anticorpos presentes no sangue de animais imunizados poderiam neutralizar doses letais em outros animais **(Behring, 1890)**. Esse marco possibilitou a Behring receber o primeiro Prêmio Nobel em Fisiologia e Medicina em 1901 pelo desenvolvimento da terapia contra difteria.

Ao longo dos anos, a imunização de diferentes espécies tornou-se essencial na produção de anticorpos destinados à terapia ou diagnóstico. Embora animais como coelhos, camundongos e porquinhos-da-índia sejam comumente utilizados devido à praticidade no laboratório, para a produção de anticorpos terapêuticos contra venenos, toxinas e vírus, o soro de cavalos ou ovelhas é preferido devido à capacidade desses animais de produzirem grandes volumes de soro **(Muyldermans and Smider, 2016)**. No Brasil, o Instituto Butantan, o Instituto Vital Brazil, o Centro de Produção e Pesquisa de Imunobiológicos e a Fundação Ezequiel Dias são as principais instituições produtoras de soro equinos abrangendo soro antibotrópicos, antiescorpiônicos, antiaracnídicos, antidiftéricos e antitetânicos, entre outros, todos produzidos a partir do plasma de cavalos (disponíveis em https://butantan.gov.br/; https://lvb.com.br/; https://www.saude.pr.gov.br/Pagina/CPPI-Centro-de-Producao-e-Pesquisa-de-Imunobiologicos; http://www.funed.mg.gov.br/).

No Brasil, o Instituto Vital Brazil e o Instituto Butantan estão contribuindo para a produção de anticorpos terapêuticos, utilizando plasma de cavalo contra a COVID-19 **(Botosso et al., 2022; Cunha et al., 2021)**. Estudos realizados pelo Instituto Vital Brazil demonstraram a eficácia do plasma desses animais, sendo cerca de 150 vezes mais neutralizantes do que os plasmas de pacientes convalescentes analisados **(Cunha et al., 2021)**.

Essas descobertas destacam o potencial do estudo do plasma equino como uma alternativa promissora para a identificação de anticorpos neutralizantes. No entanto, apesar da importância desses anticorpos, nosso conhecimento sobre os aspectos moleculares associados aos anticorpos equinos ainda é limitado.

1.7 *Loci* de imunoglobulinas de cavalo

O cavalo possui três cadeias de imunoglobulinas: cadeia pesada (IGH), leve *lambda* (IGL) e leve *kappa* (IGK). Os segmentos gênicos de cada tipo, na maioria dos vertebrados, estão localizados em três *loci* diferentes, ou seja, em três cromossomos **(Murphy, 2014)**.

Importante mencionar que os segmentos gênicos V(D)J são classificados pelo IMGT como funcionais, pseudogenes ou ORF (quadro de leitura aberto, do inglês *Open Reading Frame*). Os segmentos gênicos funcionais são comumente rearranjados e dão origem a anticorpos funcionais. Já alguns segmentos gênicos são considerados não funcionais, sendo divididos em pseudogenes ou ORFs. São classificados como pseudogene quando o segmento gênico e/ou sua região codificadora apresentar códons de parada e/ou adição ou exclusão não múltipla de 3, que leva a alteração do *frame* de leitura da sequência, ou ORF, onde apresentam alterações em seu sítio de *splicing* e/ou nos elementos de regulação e/ou alteração nos aminoácidos conservados que podem levar ao enovelamento incorreto **(Lefranc, 2014, 2001)**.

A anotação de seus *loci* foi realizada inicialmente Sun **(2010)** e revisto por Walter **(2015)**, utilizando o genoma de referência o *EquCab2.0*. Em 2018, um novo sequenciamento do genoma de cavalo foi feito, o *EquCab3.0* **(Kalbfleisch et al., 2018)** e com isso nosso grupo de pesquisa, junto com o IMGT (do inglês, *international ImMunoGeneTics information* system) estão trabalhando na nova anotação.

O *locus* da IGH está localizado no cromossomo 24 **(Figura 3)**, onde possui os segmentos gênicos dispostos no seguinte padrão: $(V_H)n$-$(D_H)n$-$(J_H)n$-$(C_H)n$ **(Sun et al., 2010; Walther et al., 2015)**. De acordo com as nomenclaturas do IMGT **(Lefranc, 2001)** e com a anotação feita pelo IMGT em colaboração com o nosso grupo e baseada no genoma EquCab3.0, atualizado em setembro de 2023 (Disponível em **IMGT: Locus description: horse (Equus caballus) IGH**), os cavalos possuem 104 segmentos gênicos IGHV (subdividido em 7 grupos, dos quais 21 são funcionais), 44 segmentos gênicos IGHD, 9 segmentos gênicos IGHJ e 11 IGHC (IgM, IgD, IgE, IgA e 7 subclasses de IgG) dispostos na orientação *reverse*. É interessante mencionar que o cavalo é um dos mamíferos que possui maior número de IGHD (apesar de 16 deles serem classificados como ORF), o que possibilita o aumento da diversidade da cadeia pesada **(Walther et al., 2015)**.

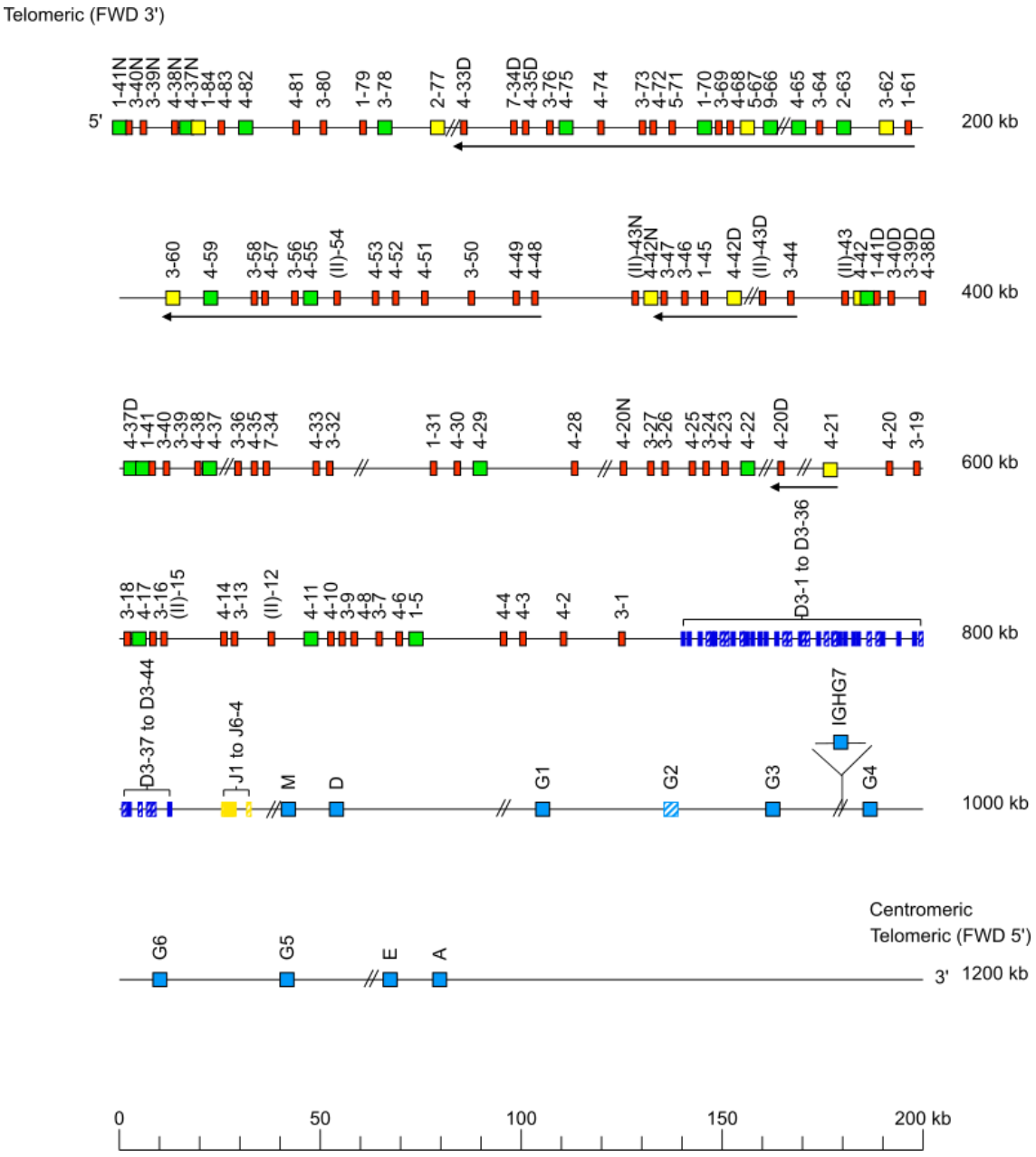

Figura 3 – Representação esquemática do *locus* de IGH do cavalo no cromossomo 24 (Equus Caballus 3.0). Em verde, observam-se os segmentos gênicos funcionais do IGHV, em amarelo os ORFs e em vermelho são os pseudogenes. Em azul escuro, observa-se os genes funcionais de IGHD, seguidos pelos segmentos gênicos funcionais IGHJ em amarelo. Em azul claro, observam-se o IGHC funcionais e em azul claro listrado verifica os pseudogenes IGHC. Figura retirada do site do IMGT em Locus representation: horse (Equus caballus) IGH).

Em humanos e camundongos a cadeia leve *kappa* é predominante, mas ao contrário desses animais, o plasma dos cavalos possui mais de 95% dos anticorpos

circulantes compostos pela cadeia leve *lambda* **(Tallmadge et al., 2014; Walther et al., 2015).**

O *locus* da IGK está localizado no cromossomo 15 **(Figura 4)**, onde possui os segmentos gênicos dispostos no seguinte padrão: (V_k)n-(J_k)n-(C_k) **(Sun et al., 2010; Walther et al., 2015)**. Até o presente momento contém 66 segmentos gênicos IGKV (divididos em 11 subgrupos, 17 funcionais, 47 pseudogenes e 2 ORFs), 5 IGKJ e 1 IGKC dispostos na orientação *forward* de acordo com a mais nova atualização do IMGT com base no novo genoma de referência o EquCab3.0, datada de setembro de 2023 (Disponível **Locus description: horse (Equus caballus) IGK)**

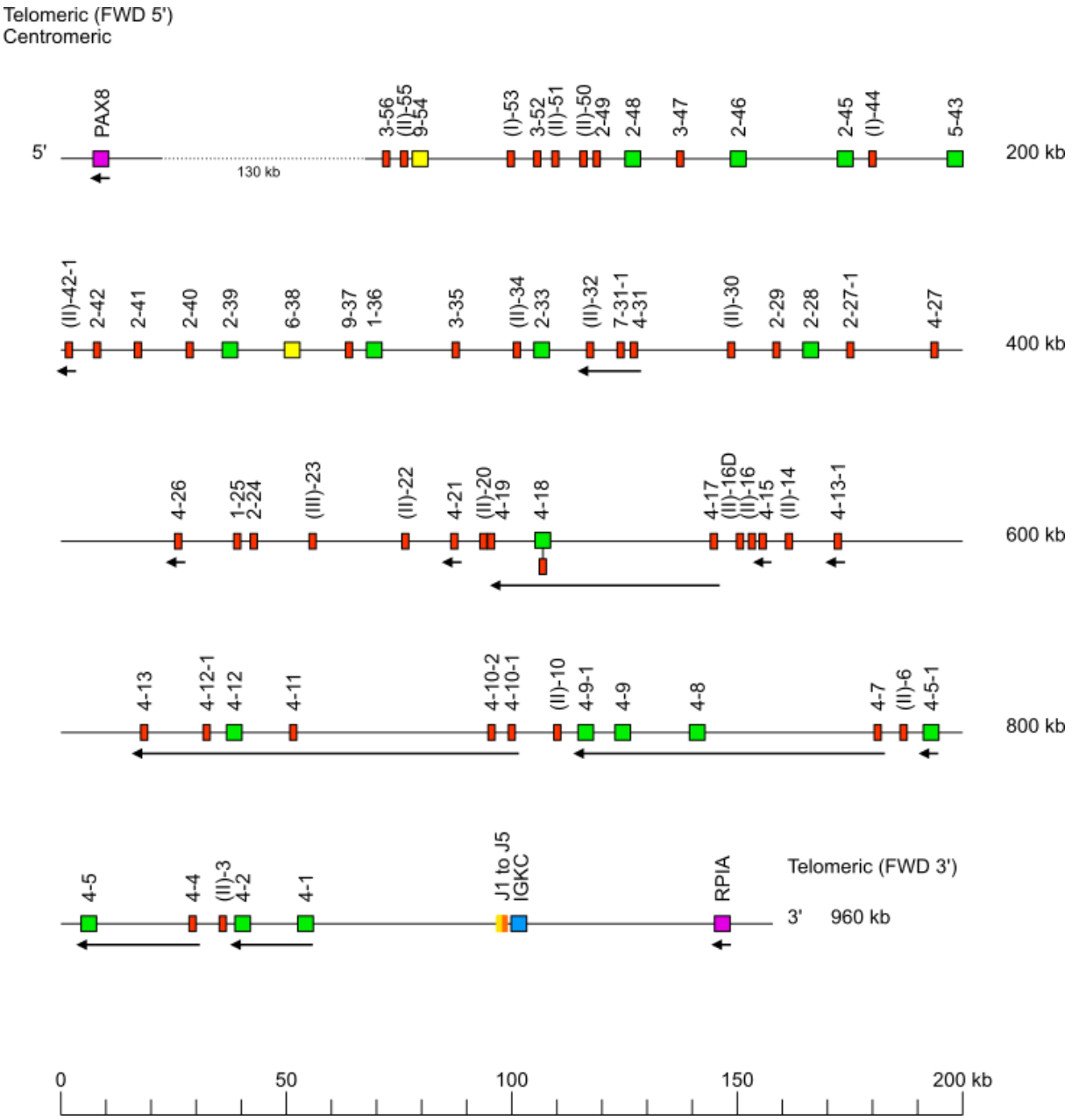

Figura 4 – Representação esquemática do *locus* de IGK de cavalo no cromossomo 15 (EquCab3.0). Em verde, observam-se os segmentos gênicos funcionais do IGKV, em amarelo os ORFs e em vermelho são os pseudogenes. Em amarelo, observa-se os segmentos gênicos funcionais IGHJ em amarelo. Em azul claro, observa-se o IGHC funcional. Figura retirada do IMGT Locus representation: horse (Equus caballus) IGK

E por fim, o *locus* da IGL está localizado no cromossomo 8, onde contém 144 segmentos gênicos IGLV (das quais 27 são genes funcionais, 112 pseudogenes e 5 ORFs), 7 IGLJ e 7 IGLC, de acordo com a revisão de Walther com base no genoma de referência do EquCab2.0 **(Sun et al., 2010; Walther et al., 2015)**. Diferente dos

demais *loci* (pesado e leve *kappa*), a cadeia leve *lambda* do cavalo possui um padrão gênico: (V_H)n-[(J_H)-(C_H)]n-(V_H)n. Onde há dois clusters dos segmentos gênicos IGLV cada um disposto em uma orientação, seguido por cada um dos segmentos IGLJ que são procedidos por um segmento IGLC e outro cluster dos segmentos gênicos IGLV **(Walther et al., 2015)**. Há também diversos *gaps* no sequenciamento do *locus* da cadeia *lambda* do cavalo, sendo assim, o IMGT não incorporou essas informações em seu banco de dados. Diante dessa lacuna, nosso grupo analisou de forma detalhada do *locus* no genoma EquCab3.0, visando identificar novos segmentos gênicos. Nesse processo, constatamos a presença de 187 segmentos gênicos IGLV (33 funcionais, 147 pseudogenes e 6 OFRs), 10 IGLJ (6 funcionais e 4 pseudogenes) e 10 IGLC **(Silva, 2024, em preparação).**

2 JUSTIFICATIVA

Embora o uso do cavalo para a produção do anticorpo remonte a mais de um século **(Behring, 1890)**, nosso conhecimento sobre o sistema adaptativo dessa espécie permanece limitado, especialmente em nível molecular. Ao compreendermos a resposta imunológica dos equinos a patógenos, poderemos aprimorar o desenvolvimento de vacinas e terapias tanto para esses animais, quanto para seres humanos.

Na revisão de Horohov **(2015)**, o cavalo é destacado como um bom modelo imunológico para estudar diversas doenças humanas, como doenças alérgicas, efeito de envelhecimento, efeito da obesidade, imunologia da gravidez e uso de células tronco, uma vez que esses processos ocorrem de forma similar aos descritos em humanos **(Horohov, 2015)**. Neste contexto, nosso grupo pioneiramente caracterizou a composição do repertório de anticorpos de cavalo em larga escala, obtendo 45.000 e 30.000 clones de IGH e IGL, respectivamente **(Manso et al., 2019).** Também analisamos o papel da diversidade juncional no repertório desses animais **(Navas et al., 2022).** Mais recentemente imunoglobulinas anti-SARS-CoV2 F(ab')2 oriundas do soro de cavalos estão sendo avaliadas para soroterapia, tanto pelo Instituto Vital Brazil (imunizados com a proteína *spike* do SARS-CoV-2) e quanto pelo Instituto Butantan (imunizados com o vírus inativado) **(Botosso et al., 2022; Cunha et al., 2021)**.

Neste sentido, o grupo do Dr. Jerson Lima, do Instituto Vital Brazil, imunizou cavalos com a proteína da espícula de SARS-CoV-2 e obteve plasma 150 vezes mais neutralizante do que plasma de doadores humanos convalescentes. Esse plasma foi neutralizante contra as variantes P.1 e P2 (Cunha et al., 2021). Tendo em vista o alto potencial neutralizante deste plasma equino, partimos do pressuposto de que conseguiremos isolar anticorpos altamente neutralizante deste plasma, visto a necessidade de novos terapêuticos contra o SAS-CoV-2.

Este trabalho é o primeiro passo para tal, visto que ele visa aprofundar a compreensão dos anticorpos equinos transcritos e circulantes. O foco está nas características dos anticorpos anti-*spike* equinos, visando identificar alternativas para o desenvolvimento de terapias contra a COVID-19.

2.1 Objetivo Geral

Analisar as características do repertório sorológico e celular de anticorpos produzidos contra a proteína *spike* em cavalos hiperminuizados.

2.2 Objetivos Específicos

- Obter sequências de cadeias pesadas e leves de anticorpos equinos e seus clones;
- Identificar anticorpos circulantes *spike*-específicos;
- Comparar as características moleculares dos transcritos de anticorpos e de anticorpos circulantes;
- Selecionar sequências de anticorpos equinos anti-*spike*.

3 MATERIAIS E MÉTODOS

3.1 Obtenção de PBMCs e plasma de cavalos

Amostras de sangue periférico foram obtidas de quatro cavalos adultos saudáveis em colaboração com o Instituto Vital Brazil, no estado do Rio de Janeiro. As amostras foram coletadas de cavalos sem raça definida, em dois momentos distintos: antes da primeira imunização, referida como amostra pré-imunização, e sete dias após a sexta imunização com a proteína *spike* do SARS-CoV-2, designada como amostra pós-imunização. Cada cavalo foi submetido a seis imunizações subcutâneas nos dias 0, 7, 14, 21, 28 e 35, sendo administrados 200 mg da proteína *spike* (variante Wuhan) em cada imunização, conforme detalhado no protocolo descrito por Cunha e colaboradores **(2021).**

As amostras de sangue total foram coletadas em tubos contendo EDTA e em seguida foram processadas para separar as células mononucleadas (PBMCs) e o plasma através de gradiente de centrifugação com o Ficoll-Paque da Sigma Aldrich. Posteriormente, as células e plasma foram divididas em microtubos para serem enviadas à UFMG.

As amostras foram devidamente rotuladas e armazenadas a -80ºC na UFMG. Para o desenvolvimento desse trabalho, foram utilizadas somente as amostras pós-imunização.

3.2 Sequenciamento do Repertório dos Receptores de Células B, Rep-Seq

Esta etapa, brevemente descrita a seguir, foi realizada pela Doutora Carlena Navas, durante seu doutorado no programa de Pós-Graduação em Bioquímica e Imunologia da UFMG, de forma semelhante descrito em sua tese **(Navas, 2022)**.

3.2.1 Extração de RNA

A partir das PBMCs de cada um dos quatro cavalos, foi feita a extração do mRNA total utilizando o método TRIzol **(Rio et al., 2010)**. A quantificação das concentrações de RNA obtidas durante o processo de extração foi realizada utilizando o kit Qubit RNA BR (Thermo Fisher Scientific).

3.2.2 Síntese de cDNA

Apenas amostras que obtiveram 500 ng de RNA total foram utilizadas para realizar a síntese de cDNA, seguindo as instruções do fabricante da enzima SuperScript IV, fornecida pela Thermo Fisher Scientific.

3.2.3 Amplificação das cadeias leves e pesadas de anticorpos de cavalo por PCR Multiplex

A amplificação dos segmentos gênicos das cadeias leve (*lambda* e *kappa*) e pesada foi realizada por PCR Multiplex utilizando um conjunto de *primers* específicos **(Tabela 1)**.

Para amplificação das cadeias pesadas e leves dos anticorpos equinos, utilizamos os *primers forward* e *reverse* previamente publicados e patenteados **(Manso, 2019)** pelo nosso grupo **(Manso et al., 2019; Navas et al., 2022)**. Os *primers forward* de todas as cadeias foram desenhados com base nas sequências do segmento *Leader* (L) e do *Framework 1* (FR1), enquanto os *primers reverse* foram desenhados para se alinhar com a região constante **(Figura 5)**.

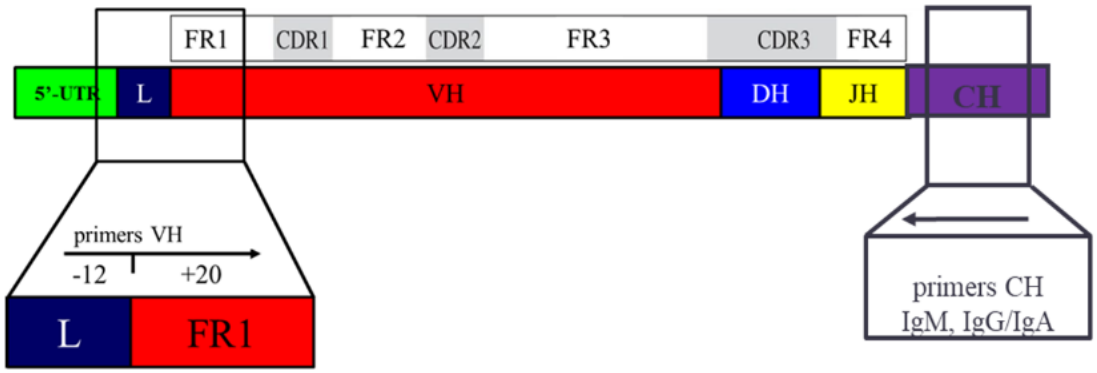

Figura 5 – Desenho dos primers *forward* e *reverse* da cadeia pesada, com um exemplo ilustrativo do processo de desenho. No caso do primer *forward*, a estratégia envolveu a incorporação de sequências provenientes do segmento *Leader* e do *Framework* 1, enquanto o *primer reverse* foi especificamente projetado para abranger a região constante mesma abordagem foi realizada para o desenho de primers da cadeia leve. Figura retirada da tese de Taciana Manso **(Manso, 2018).**

Tabela 1 – Primers *forward* e *reverse* (5'-3') utilizados na PCR Multiplex para amplificação das cadeias dos anticorpos equinos

Cadeia		Primers Forward (FOR)*		Primers Reverse (REV)**
IGH	VH1	GTGTCCTGTCCCAGGTGCARC	IgM	ATGACGTTGGGTAGGAAGTCCCG
	VH2	GWGTCCTSTCCCAGGTGMAGC	IgG	CCACCGTGGMGTCAGAYGTG
	VH3	GTGTCTCCTCCGAGGKYCAGC	IgA	AGCTGCACTGTTTCTGAACACACTCC
	VH4	TGTCTTTCGCACAGTAATACATGGC		
	VH5	AGCTCACAGCTCTCTCCTAAGACGAG		
	VH6	AGGCTGAGGACCCTCTCGTGC		
	VH7	CTCTGTCCCAGATCAGCCTGC		
	VH8_22	GTTGTGTGAAGTTMAGCTGGTGGA		
	VH9_22	TGTCCAGKGTGAGGTGCAG		
	VH10_22	AGRTCCAGCTGGYGCAGWCT		
IGK	hoKV1S1_6f	GAGCCTGTGGGGACRTCATG	IgK	GAAGCCATACACCAAGCATACGACAG
	hoKV1S2_7f	CCTGTGGGGACMTCGTGATG		
	hoKV1S3-5f	CCTGTGGGGACATCGTGWTGAC		
	hoKV2_5f	GKGASATTGTGCTGACCCAGTCT		
	hoKV3S1-2f	CCAGTGGAGATATTGTGATGACCCAGAC		
	hoKV3f	GSGGATGTTGTGWTGACCCAGAC		
	hoKV1S8f	CCTTGGGGGACATCATGATGACC		
	hoKV6S1f	CCAGGGCAGAAATAACAGTCACACAGTC		
	hoKV3S6f	CACTGGGGACACAGTTTTGACCC		
	hoKV4S1f	CCAGGTGTGAGATCCAGATGACCC		
	hoKV7S1f	GGGGGCAGATTATCCTGACTCAGTC		
IGL	hoLV8f	CCCAGTCTSTGACBCAGCCC	IgL1	TTGTTGYTGCYRAGCTCCTCA
	hoLV5f	TCTGTGKCCYCTTCTGAGCTKACTCA	IgL2	GCTGAGCTCCTCAGAGGAGG
	hoLV2_4f	CYGTGGCYTCTTCTAWGCTGAC		
	hoLV1_4f	TCGGAGSTGACTCAGCCA		
	hoLV4f	TCTCTTCTGMRGTGACTCAGCCA		
	hoLV6f	CAGTCTGCCCTGRYTCAGCC		
	hoLV9f	AGCCTGTGCTGACCCAG		
	hoLVxf	CRGATTCTCAGKCTGTGGTG		
	hoLV3S1f	TCTGTGACCACCTATGACTTGACG		
	hoLV7S1f	ACAGGATGTGGTGATTCAGG		

* Para as sequências FOR foi incorporado o adaptador TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG, na região 5'

** Para as sequências REV foi incorporado o adaptador GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG na região 5'.

Também é interessante ressaltar que os *primers* foram estrategicamente projetados com *overhangs*, ou seja, sequências adaptadoras que permite a ligação ao index e posteriormente a ligação na *flow cell* (processo fundamental para realizar o sequenciamento) **(Figura 6B)**.

A amplificação dos segmentos gênicos V(D)J da região variável foi conduzida por meio da técnica de PCR Multiplex, utilizando os *primers* previamente mencionados, através de duas reações: uma para a cadeia pesada e outra para as cadeias leves. Preparada com os seguintes componentes: tampão High Fidelity (10X), 50 mM de MgSO4, 10 mM de dNTPs, 0,5 µM de cada *primer* F, 0,5 µM de cada *primer* R e a enzima Taq DNA Polymerase Platinum High Fidelity da ThermoFisher Scientific. O protocolo de ciclagem estabelecido consistiu nas seguintes etapas: inicialmente, uma desnaturação a 94 °C por 2 minutos; seguida de 4 ciclos de 94 °C por 1 minuto, 50 °C por 1 minuto e 72 °C por 1 minuto; depois, 4 ciclos de 94 °C por 1 minuto, 55 °C por 1 minuto e 72 °C por 1 minuto; finalmente, 26 ciclos de 94 °C por 1 minuto, 63 °C por 1 minuto e 72 °C por 1 minuto, seguido por uma extensão a 72 °C por 7 minutos.

Os *amplicons* gerados a partir das amostras foram subsequentemente submetidos à análise por eletroforese em gel de agarose a 1%. As bandas específicas foram identificadas e isoladas para purificação, conforme as orientações do fabricante, utilizando o kit GFX PCR DNA e Gel Band Purification da GE Healthcare.

3.2.4 Preparação da biblioteca para sequenciamento

As amostras de DNA purificado passaram por quantificação usando o kit Qubit DNA High Sensitivity da Thermo Fisher Scientific. Em seguida, o DNA de cada amostra foi empregado na preparação das bibliotecas de sequenciamento, seguindo as instruções do fabricante do kit Nextera XT DNA Library Prep da Illumina. Durante esse processo, os index P5 e P7 de sequenciamento foram incorporados aos *amplicons* por meio da ligação com adaptadores *overhang* adicionados aos *primers* **(Figura 6C)**. A concentração das bibliotecas foi novamente confirmada com o kit Qubit DNA High Sensitivity da Thermo Fisher Scientific, e o tamanho dos *amplicons* foi verificado utilizando Bioanalyzer da Agilent.

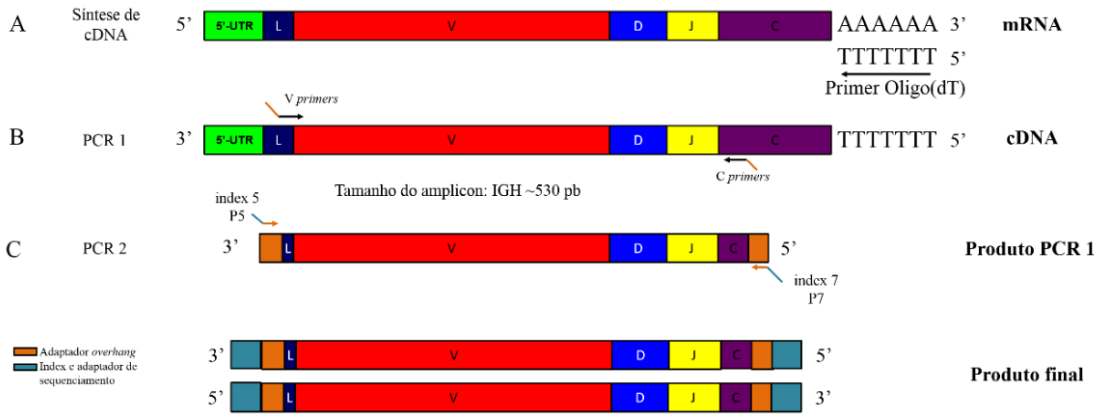

Figura 6 – Representação gráfica da amplificação e sequenciamento da região variável de IG de cavalos com primers específicos. A) O mRNA é convertido em cDNA através de *primers* Oligo(dT). B) O cDNA é utilizado como molde para amplificação através de *primers forward* e *reverse*. Aos *primers*, foram adicionadas sequências dos adaptadores (Illumina). C) Uma segunda PCR é realizada para incorporar os index P5 e P7 para o sequenciamento na plataforma Illumina MiSeq. Figura retirada da Tese de Taciana Manso **(Manso, 2018).**

3.2.5 Sequenciamento

O sequenciamento das cadeias leve e pesada dos anticorpos obtidos de cavalos imunizados com a proteína *spike* do SARS-CoV-2 foi conduzido na plataforma Illumina MiSeq 2x300 pb, de acordo com as diretrizes do fabricante. Esse processo de sequenciamento foi efetuado nas instalações do Centro de Laboratórios Multiusuário do ICB-UFMG, utilizando o sistema MiSeq da Illumina (Instrumento no. M02832), com a realização de 301 ciclos de sequenciamento para cada *read*, tanto no sentido *forward* quanto no *reverse*.

3.3 Sequenciamento das Imunoglobulinas Presente no Plasma dos Cavalos Imunizados, IG-Seq

A metodologia descria a seguir foi realizada pela então Mestranda Manuela Ferreira no programa de Bioquímica e Imunologia da UFMG e está descrita brevemente abaixo **(Ferreira, 2022, 2019)**.

3.3.1 Isolamento das Imunoglobulinas G (IgG)

As IgGs no plasma dos 4 cavalos foram purificadas por meio da cromatografia de afinidade com proteína G, utilizando a resina Pierce™ Protein G Agarose conforme descrito na monografia e dissertação de Manuela Ferreira **(Ferreira, 2022, 2019)**. As concentrações de IgG resultantes foram estimadas pelo Nanodrop.

3.3.2 Obtenção dos Fragmentos F(ab')^2

Com o intuito de obter os fragmentos F(ab')^2, o eluato da coluna de afinidade com a proteína G foi submetido a uma digestão com pepsina (Sigma). Utilizou-se uma proporção de 1:45 de enzima para cada miligrama de IgG dosado, e o pH foi devidamente ajustado para o pH ótimo da enzima, de 3,5. O processo de digestão ocorreu durante 7 horas a uma temperatura de 37ºC. A confirmação da clivagem das imunoglobulinas foi realizada por meio de eletroforese em gel de poliacrilamida, com a adição de dodecil sulfato de sódio (SDS-PAGE) 10%.

3.3.3 Seleção dos anticorpos específicos

Para selecionar os anticorpos específicos contra o antígeno, foi utilizado uma nova cromatografia de afinidade, desta vez, a resina de agarose Pierce™ NHS-Activated Agarose, Dry (Thermo) imobilizada seguido o protocolo do fabricante com 2mg da proteína *spike* (variante Wuhan) **(Alvim et al., 2021)** cedida pela colaboradora e Professora Leda Castilho da Universidade Federal do Rio de Janeiro.

Resumidamente, foram aplicados 2mg do fragmento F(ab')^2 na resina, que foi lavada com PBS. Os fragmentos F(ab')^2 específicos contra a proteína *spike* foram eluídos e em seguida, foi verificado a fração que apresentava maior reatividade através de ELISA frente à proteína *spike*.

3.3.4 Preparo para espectrometria

As frações reativas à proteína *spike* dos fragmentos F(ab')^2 de cada cavalo foram combinadas e concentradas no speedvac. Posteriormente, as amostras foram submetidas à desnaturação, redução, alquilação e, em seguida, clivadas por tripsina, seguindo a estratégia "*bottom up*" descrita por **(Lavinder et al., 2014)**.

3.3.5 Sequenciamento dos peptídeos dos anticorpos específicos

Os peptídeos foram submetidos a cromatografia líquida de alta performance (HPLC) e foram direcionadas para o equipamento Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher) para sequenciamento em colaboração com o Prof. Dr. Martin. R Larsen do Departamento de Bioquímica e Biologia Molecular da *University of Southern Denmark*.

De forma sumarizada, neste trabalho, foram usados os dados experimentais previamente obtidos utilizando as metodologias descritas acima e sumarizada na **Figura 7**. Ambas as técnicas têm sido empregadas de maneira sinérgica com sucesso

na avaliação da composição dos anticorpos policlonais presentes em plasmas após a imunização, metodologia utilizada para analisar o repertório imunológico de humanos, coelhos e camundongos **(Sato et al., 2012; Wine et al., 2013; Ionov and Lee, 2022)** e mais recentemente, nosso grupo vem empregando parte desta metodologia para estudar o repertório de anticorpos dos equinos **(Manso et al., 2019; Navas et al., 2022)**.

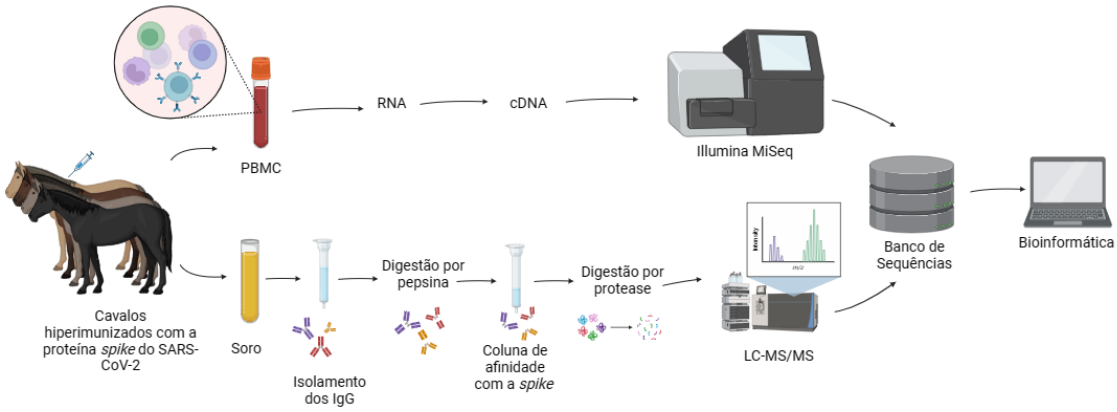

Figura 7 – Workflow para análise do repertório de anticorpos por NGS e LC-MS/MS. Na parte superior da figura, encontra-se a primeira abordagem, que envolve a transcrição reversa dos receptores das células B, utilizando *primers* específicos para as regiões variáveis e constantes desses receptores, seguido por sequenciamento de nova geração (NGS). Esse método é conhecido como Sequenciamento do Repertório dos Receptores de Células B, Rep-Seq. Na parte inferior da figura, destacamos a segunda metodologia: o Sequenciamento de Imunoglobulinas presentes no plasma, referido como Ig-Seq. Nesse processo, anticorpos específicos são selecionados a partir do plasma por meio de uma coluna de afinidade previamente imobilizada com o antígeno de interesse. Em seguida, o eluato obtido passa por uma etapa de digestão e é submetido ao sequenciamento utilizando LC-MS/MS. Após a obtenção dos dados brutos por meio dessas técnicas, uma série de ferramentas de bioinformática é empregada para análise. Isso inclui a identificação e mapeamento dos peptídeos dos anticorpos específicos, bem como a determinação da sequência completa dos anticorpos em questão.

3.4 Análise bioinformática do repertório de receptores de células B (Rep-Seq)

Em 2015, foram estabelecidas diretrizes que delineiam as principais etapas para a análise do repertório de anticorpos: (I) pré-processamento dos dados brutos, (II) anotação da linhagem germinativa do anticorpo (*germline)*, (III) determinação da

população clonal e, por fim, (IV) análise das características dos anticorpos **(Yaari and Kleinstein, 2015)**.

3.4.1 Pré-processamento das *reads*

As reads provenientes do sequenciamento Illumina MiSeq passaram por um processo de pré-processamento utilizando o *software* pRESTO **(Vander Heiden et al., 2014)**, integrado à plataforma *Immcantation*, que incluiu as seguintes etapas: (I) montagem das *reads*, (II) eliminação das sequências de baixa qualidade, cujo *Phred score* fosse inferior a 30, (III) identificação dos *primers forward* e *reverse*, e por fim, (IV) remoção das *reads* montadas com um tamanho inferior a 100 nucleotídeos.

A montagem das *reads paired-end* foi realizada por meio do alinhamento das extremidades 3' na região onde ocorre o *overlap* das mesmas através do comando *AssemblePairs.py align*, permitindo a montagem das *reads* em uma sequência única.

Na fase pós-montagem das *reads*, um *Phred score* (Q), métrica que fornece a acurácia de cada base sequenciada, de 30 (Q30: probabilidade de erro equivalente a 1 em 1.000) foi aplicado, proporcionando uma acurácia de 99,9% através do comando *FilterSeq.py quality*.

Posteriormente, os *primers* foram devidamente identificados por meio de um processo de alinhamento através do comando *MaskPrimers.py align.* Para isso, foram criados dois arquivos no formato *.FASTA para a cadeia leve (*lambda* e *kappa* no mesmo arquivo) e dois arquivos para a cadeia pesada, contendo as sequências dos *primers forward* e *reverse*. Durante esse procedimento, foi estabelecido um limite de erro máximo de 10% para garantir a identificação dos *primers*, mas ao mesmo tempo permitir pequenas alterações nas sequências. Sequências que não tiveram os primers identificados foram excluídas da análise.

Além disso, foi implementado um filtro de qualidade que tem como critério a exclusão de todas as sequências cujo tamanho seja menor a 100 nucleotídeos, através do comando *FilterSeq.py length*. Esse filtro foi aplicado tendo em vista que o tamanho esperado das sequências é superior a 350 nucleotídeos para a cadeia pesada e 300 nucleotídeos para as cadeias leves.

A linha de comando empregada para executar o pré-processamento no pRESTO encontra-se detalhada no Anexo I.

Após o pré-processamento das cadeias leves, realizamos um passo adicional para a separação das sequências de *lambda* e *kappa*. Isso se tornou necessário devido

ao sequenciamento conjunto das amostras. Para esse propósito, desenvolvemos um pequeno *script* (Anexo II).

3.4.2 Anotação dos segmentos V(D)J

3.4.2.1 Cadeia Pesada

Para realizar a anotação, ou inferência da linhagem germinativa dos anticorpos, as sequências após o pré-processamento, no formato *.FASTA, foram submetidas a plataforma IMGT High-VQuest **(Brochet et al., 2008)**. A ferramenta, disponível de forma online, permite a identificação dos segmentos V(D)J das sequências de imunoglobulina já rearranjadas, através de alinhamento com a *germline* do genoma de referência disponibilizado pelo diretório, além da identificação das regiões estruturais (FR) e regiões de complementariedade (CDR), com as sequencias devidamente numeradas de acordo o IMGT. Como *output,* a plataforma oferece onze arquivos em formado *.CSV*.* Para a anotação da cadeia pesada, foi utilizado como referência o genoma *Equus caballus 3.0*.

3.4.2.2 Cadeia Leve

Para as cadeias leves, notamos que há informações disponíveis no IMGT apenas cadeia leve *kappa* dos cavalos, mas não para a cadeia leve *lambda*. Considerando que cerca de 95% dos anticorpos circulantes em cavalos são compostos pela cadeia leve *lambda* **(Tallmadge et al., 2014; Walther et al., 2015),** optamos por realizar a anotação de ambas as cadeias leves no IgBlast.

O IgBlast **(Ye et al., 2013)** é um *software* de anotação que permite a utilização de linhagens germinativas disponíveis no IMGT ou a incorporação de um banco de dados de linhagens germinativas personalizado, possibilitando, assim, a anotação por meio de alinhamento local.

Desta forma, criamos um banco de sequências da linhagem germinativa da cadeia *lambda*. Para criar esse banco, criamos arquivos *.FASTA contendo os genes V e J da *germline* dos anticorpos de cavalos, utilizando os genomas de referência disponíveis: EquCab 3.0 e EquCab 2.0 e com base nas coordenadas dos segmentos gênicos disponibilizadas na literatura **(Sun et al., 2010; Walther et al., 2015)**. A escolha de incorporar as sequências de ambos os genomas de referência foi feita após observar um baixo percentual de anotação das sequências *lambda* utilizando apenas o genoma EquCab 2.0.

Já para a cadeia leve *kappa*, foi realizado também a anotação pelo IgBlast, utilizando apenas o genoma de referência do EquCab 3.0, já disponibilizado pelo IMGT.

Adicionalmente, foram criados arquivos auxiliares para cada uma das cadeias leves, contendo informações sobre o frame de leitura, o tipo da cadeia e a posição de término do CDR3 para cada um dos segmentos do gene J (Anexo III). É importante mencionar que a posição de término do CDR3 é determinada pelo primeiro nucleotídeo do *motif* (F/W-G-X-G), onde X pode representar qualquer aminoácido.

3.4.3 Clonotipagem

3.4.3.1 Cadeia Pesada

Para realizar a clonotipagem da cadeia pesada, foram utilizados os arquivos de saída do IMGT 1_Summary, 2_IMGT-gapped, 3_Nt-sequences e 6_Junction como entrada, empregando o comando MakeDb.py do pacote *Change-O* da plataforma *Immcantation* para a criação de um documento único com todas as informações necessárias para a clonotipagem. Em seguida, o documento foi submetido ao *ShaZam* utilizando a função "*distToNearest*", para calcular o limite apropriado (cut-off) do agrupamento hierárquico **(Gupta et al., 2015)**. Após, o arquivo foi submetido, à ferramenta de clonotipagem, YClon, desenvolvida pelo nosso grupo **(Gervásio et al., 2023).**

3.4.3.2 Cadeia Leve

Já para as cadeias leve, utilizou-se diretamente o arquivo de saída do *IgBlast* como *input* no *YClon*, pois o arquivo já possui todas as informações em um único arquivo. Também não foi necessário determinar o limite ideal para a atribuição clonal, uma vez que as plataformas do *Change-O* foram criadas especificamente para a cadeia pesada. Sendo assim, utiliza-se o *threshold default* de 0,9 para tal.

3.4.4 Análise das características do repertório de anticorpos

Em seguida, utilizando os dados dos clones de anticorpos de cada cavalo é analisado o uso gênico, hipermutação somática, tamanho e composição do CDR3 através do *script* feitos *in house* em R Studio. O *script* foi desenvolvido por meio das bibliotecas `data.table`, `tidyverse`, `dplyr`, `openxlsx`, `biostrings`, e `stringr`, onde realiza um processo de análise das sequências provenientes de amostras clonotipadas. Inicialmente, ele define o diretório das amostras e cria um *dataframe*

contendo os arquivos de entrada, sendo eles os relatórios e os resultados da clonotipagem. O script então entra em um loop, onde para cada amostra, importa e organiza os dados. Após a junção desses dados, o script realiza análises detalhadas, incluindo a frequência gênica dos segmentos gênicos V e J, a hipermutação somática, a distribuição do tamanho do CDR3, e a composição do CDR3 em termos de aminoácidos. Finalmente, os resultados são exportados para um arquivo no Excel para facilitar a manipulação dos dados (Anexo IV). Para comparar a diferença entre o Rep-Seq e o Ig-Seq, utilizamos a ANOVA 2-way no GraphPad Prism 5.00 (GraphPad Software, Inc., San Diego Califórnia EUA).

3.5 Análise bioinformática do repertório das Imunoglobulinas específicas circulantes (Ig-Seq)

3.5.1 Identificação de peptídeos oriundos da LC-MS/MS no banco de sequências de anticorpos anotados

A partir dos transcritos dos anticorpos devidamente anotados e traduzidos, cria-se um banco de sequências das regiões variáveis dos anticorpos. A esse banco de sequências foi também adicionado de forma manual sequências contaminantes comuns da proteômica e sequências equinas (como por exemplo, queratina, albumina, tripsina, *spike,* etc). Esse banco de sequências foi então, confrontado com as sequências dos peptídeos encontrados na espectrometria de massa pelo *software* Proteome Discoverer (Thermo Scientific), utilizando a ferramenta de busca SEQUEST HT.

3.5.2 Mapeamento dos peptídeos nos CDR3

Os resultados exportados do Proteome Discoverer contêm informações sobre todos os peptídeos cujas sequências foram identificadas no banco de dados completo personalizado, o que engloba possíveis contaminantes, proteínas de cavalos ou proteína *spike*.

No entanto, devido a características específicas, um mesmo peptídeo poderia ser potencialmente associado a diversas sequências distintas, uma vez que essas moléculas apresentam regiões conservadas. Para superar essa particularidade, foi estabelecido que apenas os peptídeos mapeados à região CDR3 dos anticorpos seriam considerados para a identificação das sequências, uma vez que essa região é altamente variável **(Figura 8)**. Sendo assim, para a avaliação das sequências de

anticorpos específicos foi necessária a criação de um *script* personalizado cedido pelo grupo de pesquisa do professor e colaborador Dr. George Georgiou da *University of Texas at Austin* e posteriormente personalizados pelo nosso grupo, para filtrar os peptídeos mapeados na região CDR3. Esses peptídeos trípticos informativos, quando confrontados com o banco de sequências criado, permitiram a recuperação completa das sequências dos anticorpos.

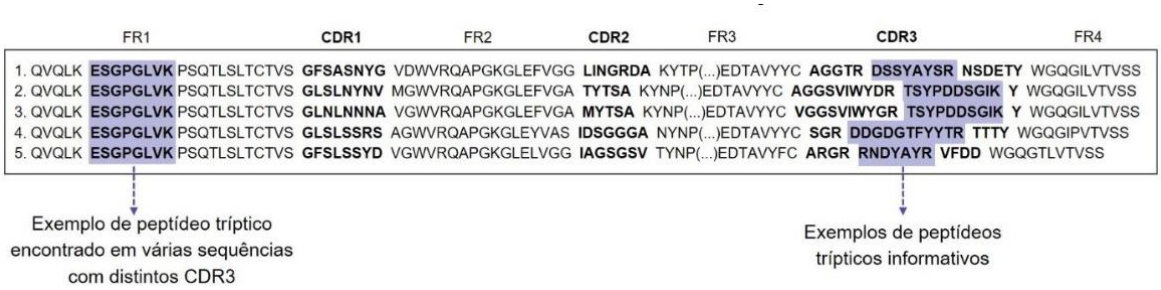

Figura 8 – Identificação de peptídeos de anticorpos específicos contra a proteína *spike*. Observa-se que na região de FR1, os peptídeos trípticos encontrados podem ser referentes a qualquer uma das cinco sequências. Enquanto, a região CDR3 mostra 5 peptídeos trípticos diferentes, sendo informativos para identificação para identificação da sequência completa ao serem confrontados com o banco de sequências do Rep-Seq. Figura retirada da dissertação da Manuela Ferreira **(Ferreira, 2022)**

Em seguida, o uso gênico, hipermutação somática, tamanho e composição do CDR3 também foram analisados através do *script* feitos *in house* em R Studio e comparados com os resultados das sequencias dos transcritos (Anexo IV), conforme descrito no tópico 3.4.4.

O workflow completo da análise bioinformática do sequenciamento dos transcritos de imunoglobulina e dos peptídeos derivados dos anticorpos pode ser visualizado na **Figura 8**.

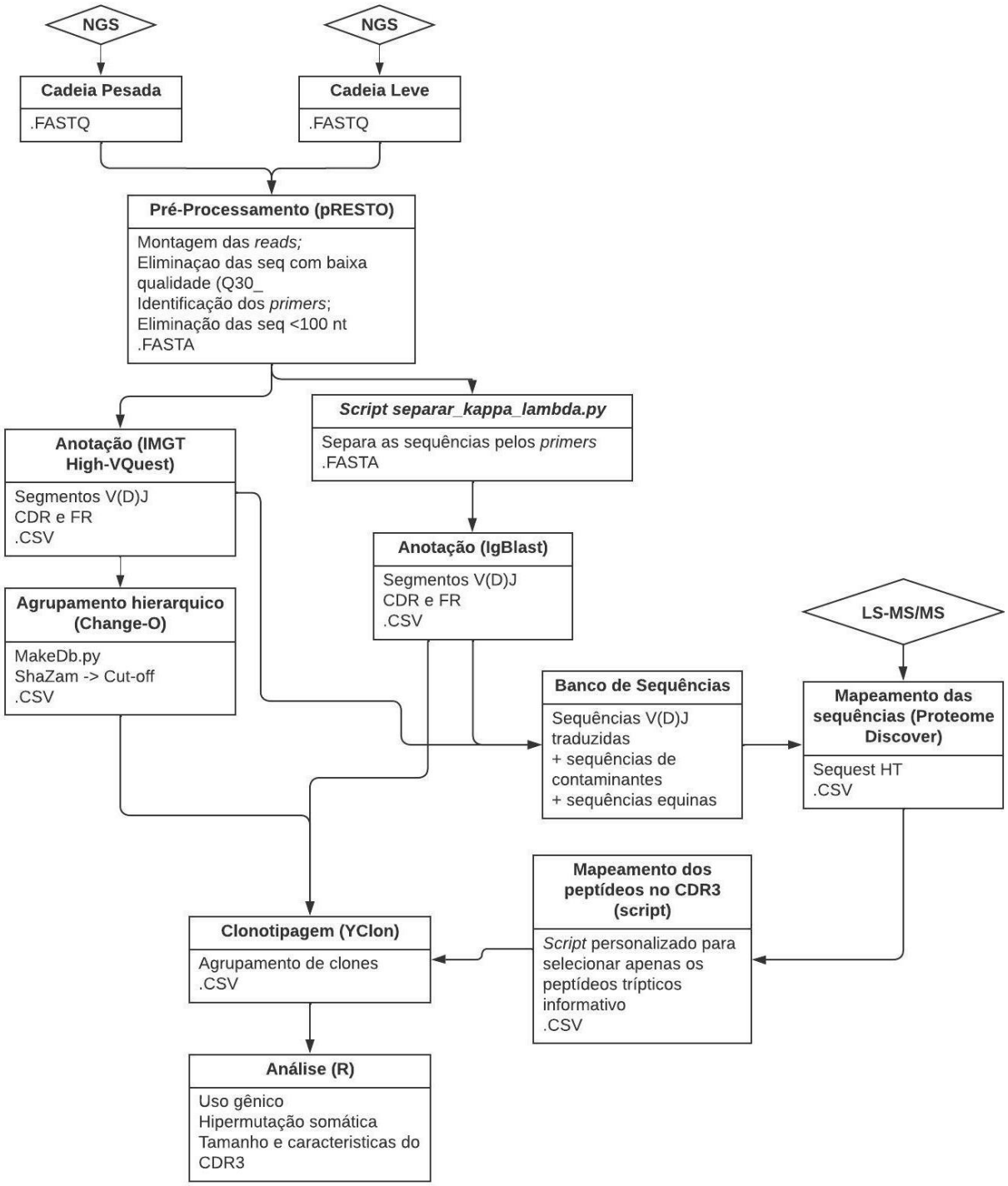

Figura 9 – Representação do fluxo de trabalho utilizado para a análise do Rep-Seq e Ig-Seq. Inicialmente, as sequências NGS passam por uma etapa de pré-processamento utilizando o software pRESTO. Em seguida, as sequências são submetidas à anotação, onde as cadeias pesadas são anotadas utilizando o sistema IMGT, enquanto as cadeias leves são anotadas através do IgBlast. Posteriormente, as cadeias são clonotipadas utilizando o YClon, seguido pela análise no ambiente R. Os resultados da clonotipagem também alimentam a construção de um banco de sequências, utilizado para mapear peptídeos específicos e

circulantes identificados por LC-MS/MS. Estes peptídeos são comparados com o banco de sequências, e o resultado é processado por um *script* para selecionar apenas peptídeos trípticos informativos. Por fim, os resultados são novamente analisados no ambiente R. Representação gráfica desenvolvida pela plataforma Lucid.

3.6 Seleção dos anticorpos específicos contra proteína *spike* para expressão *in vitro*

Para a seleção de anticorpos específicos equinos contra a *spike*, demos prioridade à escolha do cavalo que demonstrou a maior capacidade de neutralização. Esses dados de neutralização foram fornecidos pelo Dr. Herbert Guedes, da UFRJ. Como não possuímos dados de cadeia leve e pesada pareadas selecionamos cadeias leves e pesadas independentes. Em relação à cadeia pesada, selecionamos sequências que apresentaram uma abundância do peptídeo detectado na espectrometria de massa e que eram presentes em clones expandidos. No caso da cadeia leve, devido ao baixo número de sequências obtidas, selecionamos todas as sequencias especificas encontradas no plasma do cavalo.

3.6.1 Modelagem e clusterização estrutural

Para auxiliar na seleção dos anticorpos prioritários a serem caracterizados em uma próxima fase deste trabalho, realizamos uma modelagem molecular de todos os possíveis anticorpos do cavalo 4, o que apresentou anticorpos mais neutralizantes, utilizando o *software* ImmuneBuilder **(Abanades et al., 2023)**. O software utiliza modelos de aprendizado de máquina treinados para prever com alta precisão a estrutura de anticorpos, nanocorpos e receptores de células T.

As sequências dos anticorpos foram submetidas a plataforma (disponível em https://github.com/oxpig/ImmuneBuilder) que disponibiliza ferramentas para realizar a modelagem estrutural. Posteriormente, os modelos foram submetidos ao *software* SPACE2 **(Spoendlin et al., 2023)**, que realiza uma clusterização estrutural dos anticorpos, considerando as regiões de complementaridade (CDR1, CDR2 e CDR3) de ambas as cadeias.

3.7 Desenho dos genes que codificam os anticorpos

As sequências de DNA dupla fita selecionadas para a expressão foram otimizadas para expressão em célula de mamífero utilizando a ferramenta *Codon Optimization Tool*, da IDT. Foram adicionadas, às extremidades dos genes de cadeia

pesada e leve, sequências idênticas as sequências do plasmídeo pcDNA 3.4 para posterior clonagem utilizando a metodologia de *Gibson Assembly* **(Gibson et al., 2009)**.

É importante ressaltar que o plasmídeo pcDNA 3.4, disponível em nosso laboratório, já possui as regiões constantes humanas de IgG1, permitindo a expressão do anticorpo quimérico, na próxima etapa deste trabalho.

4 RESULTADOS

4.1 Análise dos dados do sequenciamento

4.1.1 Anotação e obtenção de clones da cadeia pesada de anticorpos equinos

Utilizando a plataforma Illumina MiSeq com sequenciamento de 2x300 pares de bases, observamos uma heterogeneidade no número de reads das amostras analisadas. O número mínimo de *reads* foi registrado na amostra 834-I, totalizando 364.671 *reads*, enquanto a amostra 843-I apresentou o máximo de *reads*, atingindo um total de 470.017 *reads* **(Tabela 2).**

Em seguida, realizamos o processamento das sequências utilizando o *software* pRESTO **(Vander Heiden et al., 2014)**, onde realizamos a montagem das *reads* e posteriormente eliminamos aquelas de baixa qualidade (q30) e de tamanho inferior a 100 nucleotídeos. Aproximadamente 98% dessas sequências foram completamente anotadas, ou seja, os segmentos V(D)J e o CDR3 foram determinadas para cada sequência pela plataforma HighV-Quest do IMGT **(Brochet et al., 2008)**. Das sequencias anotadas, realizamos a clonotipagem das sequencias, onde obtivemos um mínimo de 41.603 clones na amostra 839-I e um máximo de 78.229 clones na amostra 843-I, utilizando o YClon **(Gervásio et al., 2023)**.

Tabela 2 – Resumo dos resultados do sequenciamento da cadeia pesada dos quatro cavalos imunizados com a proteína *spike* do SARS-CoV-2.

Amostra	Dados Brutos	Pré-Processamento	Sequências Anotadas	Clones Únicos
834-I	364.671	346.376	332.838	67.391
839-I	456.247	424.459	416.712	41.603
843-I	470.017	432.560	422.548	78.229
844-I	634.499	600.688	591.310	46.790
Média	**481.359**	**451.021**	**440.852**	**58.503**
SD	**76.570**	**74.834**	**75.229**	**14.307**

4.1.2 Anotação e obtenção de clones da cadeia leve de anticorpos equinos

Na análise das cadeias leves, também foi observado uma heterogeneidade nos dados brutos, com a menor quantidade de *reads* registrada na amostra 843-I, totalizando 500.783 *reads*, e a maior quantidade na amostra 839-I, com um total de 683.548 *reads*.

Em seguida, as *reads* foram submetidas ao pré-processamento no *software* pRESTO para montagem das sequências e garantia de qualidade. Após essa etapa de pré-processamento, as sequências foram classificadas em cadeia *lambda* ou cadeia *kappa* com base nos *primers* utilizados através de um *script* feito *in house* (Anexo II). Em média, 42%±12 das sequências foram identificadas como cadeia *lambda*, enquanto as 58%±12 restantes foram identificadas como cadeia *kappa*.

Posteriormente, as sequências foram submetidas à anotação usando o IgBlast **(Ye et al., 2013)**. Constatou-se que em média 89,6%± 2,4 das sequências da cadeia *lambda* foram devidamente anotadas, incluindo os segmentos V/J e o CDR3, enquanto 99,3%±0,24 das sequências da cadeia *kappa* foram devidamente anotadas.

Após a anotação das sequências, procedeu-se à clonotipagem das mesmas utilizando o YClon, revelando uma considerável heterogeneidade. No caso das sequências da cadeia *kappa*, foi observada uma contagem mínima de 4.302 clones únicos na amostra 839-I e uma contagem máxima na amostra 834-I, com 119.700 clones. No caso das sequências da cadeia *lambda*, foi identificado um mínimo de 54.998 clones na amostra 834-I e um máximo de 106.933 na amostra 839-I.

Tabela 3 – Resumo dos resultados do sequenciamento da cadeia leve (*kappa* e *lambda*) dos quatro cavalos imunizados com a proteína *spike* do SARS-CoV-2.

Amostra	Dados brutos	Pré-processamento	Sequências anotadas		Clones únicos	
			Kappa	Lambda	Kappa	Lambda
834-I	513.027	496.466	369.077	119.700	22.742	54.998
839-I	683.548	663.197	210.361	397.292	4.302	106.933
843-I	500.783	488.579	282.462	184.741	8.730	67.062
844-I	675.293	658.609	335.583	283.891	6.850	75.746
Média	**593.163**	**576.713**	**299.371**	**246.406**	**10.656**	**76.185**
SD	**86.258**	**84.190**	**52.959**	**94.186**	**6.043**	**15.374**

4.2 Análise dos dados Ig-Seq para identificação de peptídeos específicos anti-*spike*

Para a detecção dos anticorpso circulantes *spike*-específicos no plasma equino, os espectros dos peptídeos trípticos obtidos foram comparados com o banco de sequências IGH, IGL e IGK obtidos na etapa anterior.

A análise considerou apenas as sequências que estavam presentes nas réplicas experimentais e que apresentavam uma sobreposição de pelo menos quatro resíduos entre o peptídeo e o CDR3 das sequências no banco de sequências. Durante esse procedimento, um mesmo peptídeo poderia estar associado a diferentes CDR3, e um mesmo CDR3 poderia ser encontrado em mais de uma sequência na região variável do banco de sequências de transcritos. Os números de peptídeos, CDR3 e sequências resultantes desse processo estão resumidos na tabela abaixo **(Tabela 4)**.

Sendo assim, encontramos 106 sequencias especificas para IGH, 49 para IGL e 19 para IGK.

Tabela 4 – Número de peptídeos, CDR3 e sequências (IGH, IGK e IGL) identificadas específicas contra a proteína *spike.**

	Pesada			Kappa			Lambda		
Amostra	**Pept**	**CDR3**	**Seq**	**Pept**	**CDR3**	**Seq**	**Pept**	**CDR3**	**Seq**
834-I	20	20	20	16	16	16	17	17	17
839-I	52	48	48	1	1	1	30	30	30
843-I	19	18	19	0	0	0	2	2	2
844-I	19	19	19	2	2	2	0	0	0

*A coluna de peptídeos (chamada de Pept) refere-se às sequencias especificas que foram identificadas pela espectrometria de massas. A coluna CDR3, representa o número de correspondência do CDR3 encontrados no banco de sequências criado pelo NGS. Por fim, a coluna sequência indica o total de sequências de região variável de anticorpo abrangidas pelo CDR3. Para esclarecimento, tomemos o exemplo do cavalo 839-I na cadeia pesada. Foram identificados 52 peptídeos, os quais foram encontrados em 48 CDR3. Isso indica que alguns desses peptídeos compartilham o mesmo CDR3. Finalmente, esses 48 CDR3 foram atribuídos à 48 sequências distintas.

4.3 Análise comparativa entre o repertório dos transcritos (Rep-Seq) e do repertório sorológico (Ig-Seq)

Conduzimos uma análise comparativa entre os clones dos transcritos obtidos por meio do Rep-Seq e o repertório sorológico que continha os anticorpos específicos direcionados à proteína *spike*, obtidos pelo Ig-Seq. Nesse processo, analisamos a frequência do uso gênico e a hipermutação somática desses segmentos utilizados nos anticorpos, bem como o tamanho e as características do CDR3.

4.3.1 Análise dos segmentos gênicos mais frequentes em anticorpos equinos

A análise inicial dos repertórios de anticorpos equinos se concentrou na frequência de segmentos gênicos V(D)J presentes nos anticorpos, derivados tanto dos transcritos das células B totais quanto daqueles específicos para a proteína *spike*, conhecidos como Rep-Seq e Ig-Seq, respectivamente. Os resultados revelaram uma marcante preferência nos cavalos pela seleção de segmentos gênicos na produção de anticorpos, destacando uma notável concordância entre os transcritos de anticorpos (Rep-Seq) e os anticorpos circulantes (Ig-Seq).

Na cadeia pesada, a família IGHV4 se destaca em mais de 99,9% das sequências, notadamente os segmentos IGHV4-21*01 ORF, IGHV4-22*01 F, IGHV4-29*01 F e IGHV4S1*01 F, que compõem, em média, 88,9% e 89,8% das sequências nos conjuntos Rep-Seq e Ig-Seq, respectivamente. Isso implica que cerca de 90% dos transcritos e anticorpos circulantes compartilham um desses quatro segmentos IGHV, dos 21 segmentos gênicos funcionais disponíveis. Embora haja poucas alterações nas frequências gênicas da cadeia pesada entre Rep-Seq e Ig-Seq, observou-se que a frequência do segmento gênico IGHV4-22*01 F é significativamente maior no Ig-Seq em comparação com o repertório celular ($p < 0,05$), sugerindo uma preferência específica por esse segmento nos anticorpos circulantes.

Quanto aos segmentos IGHD, foram identificados 42 e 33 segmentos no Rep-Seq e Ig-Seq, respectivamente, com destaque para o segmento IGHD2-26*01 F, que apresenta uma frequência média de quase 10% em ambos os repertórios. Na preferência por segmentos gênicos J, destaca-se o IGHJ6-1*01 F, presente em 46% das sequências **(Figura 10)**.

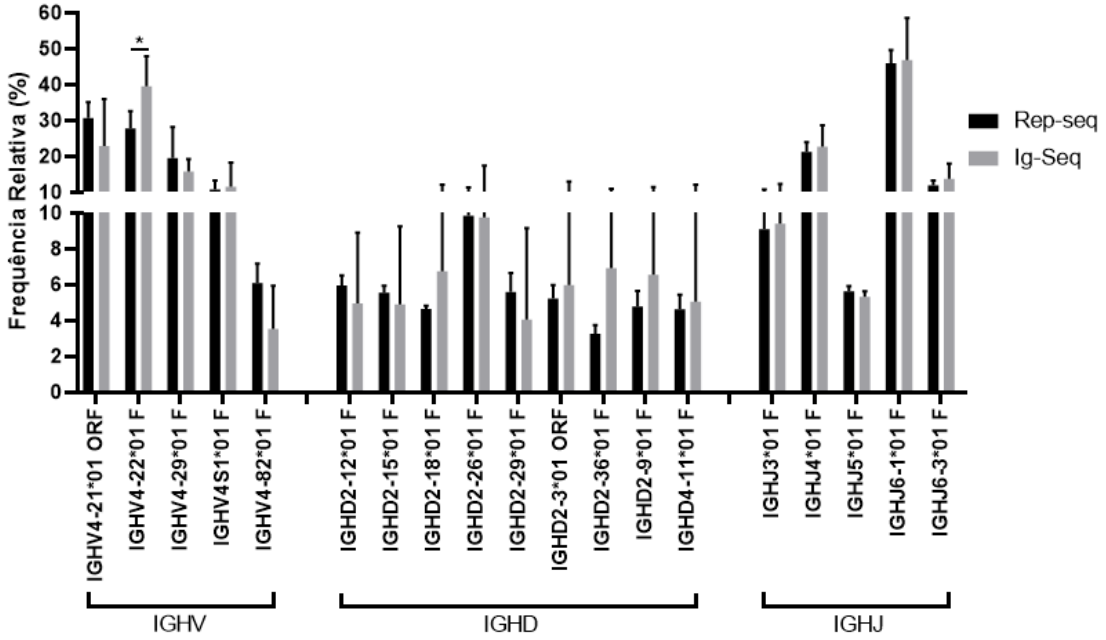

Figura 10 – Frequência relativa média dos genes IGH anotados mais frequentes. Em cinza, a frequência relativa do Ig-Seq, enquanto em preto, temos a frequência relativa do Rep-Seq. A plotagem está restrita aos segmentos gênicos que exibiram uma frequência relativa média igual ou superior a 5% tanto no Rep-Seq quanto no Ig-Seq. A diferença nas médias foi avaliada por meio de um teste de ANOVA de dois fatores (ANOVA 2way). Os asteriscos indicam os seguintes níveis de significância: * $p < 0,05$.

No que diz respeito à cadeia leve *lambda*, observa-se uma destacada prevalência da família de segmentos gênicos IGLV8 no segmento IGLV, presente em 86,8% e 89,7% dos segmentos no Rep-seq e no Ig-seq, respectivamente. Ao comparar essas frequências, evidencia-se um aumento estatisticamente significante na preferência por IGLV8S1, no Ig-seq quando comparado com Req-seq com $p < 0,001$. Em relação aos segmentos IGLJ, o destaque vai para o segmento IGLJ4S1P, com uma frequência média de 66,9% no Rep-seq e 62,9% no Ig-Seq **(Figura 11)**.

É relevante mencionar que a média dessas frequências gênicas do Ig-Seq foi calculada com base em dados de três cavalos, pois no cavalo 844-I não foram identificadas sequências *lambda* específicas.

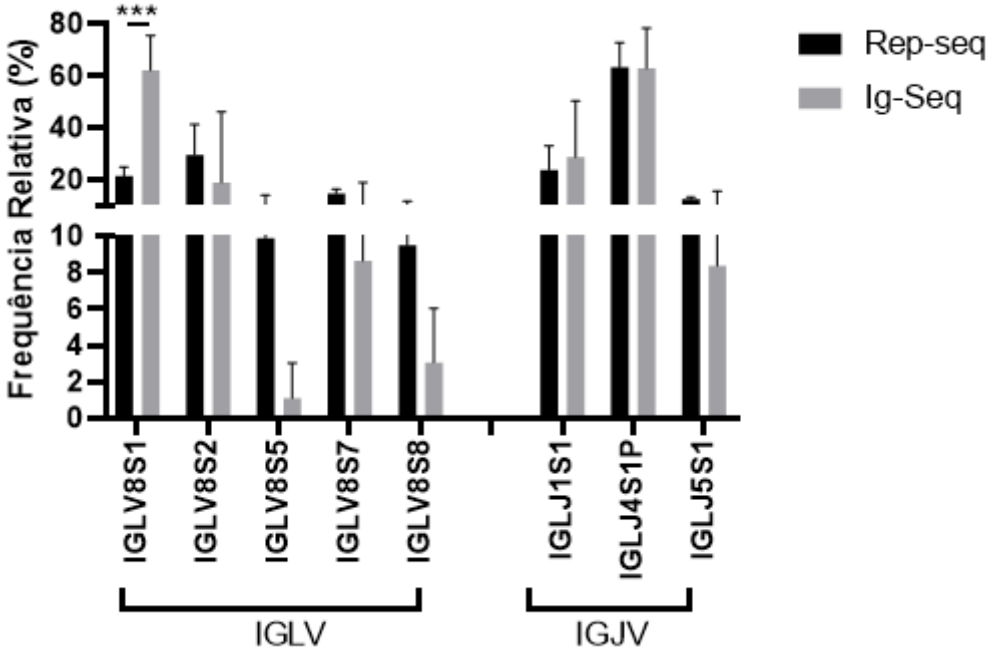

Figura 11 – Frequência relativa média dos genes IGL anotados mais frequentes. Em cinza, a frequência relativa do Ig-Seq, enquanto em preto, temos a frequência relativa do Rep-Seq. A plotagem está restrita aos segmentos gênicos que exibiram uma frequência relativa média igual ou superior a 5% tanto no Rep-Seq quanto no Ig-Seq. Importante mencionar que a média da frequência relativa do Ig-Seq foi realizada com 3 amostras, enquanto a frequência média do Rep-Seq foi realizada com 4 amostras. A diferença nas médias foi avaliada por meio de um teste de ANOVA de dois fatores (ANOVA 2way). Os asteriscos indicam os seguintes níveis de significância: *** $p < 0,001$.

Assim como nas cadeias pesada e leve *lambda*, a cadeia leve *kappa* também manifesta preferências distintas por determinados segmentos gênicos **(Figura 12)**. No segmento V, observa-se uma forte predileção pelos genes da família IGKV4, presentes em 96,2% das sequências no Rep-seq e 66,7% no Ig-Seq. Quanto aos genes J, destacam-se os segmentos IGKJ4*01 e IGKJ2*02, que, combinados, representam 92,1% e 98% das sequências no Rep-seq e no Ig-seq, respectivamente, nos anticorpos que possuem cadeia *kappa*.

Embora existam discrepâncias nas frequências, não foi observada diferença significativa entre a presença dos segmentos gênicos nos transcritos e nos anticorpos circulantes. É essencial salientar que a média da frequência dos segmentos gênicos encontrados no Ig-Seq da cadeia *kappa* foi calculada com dados de três cavalos, uma vez que não foram identificadas sequências da cadeia *kappa* no cavalo 843-I.

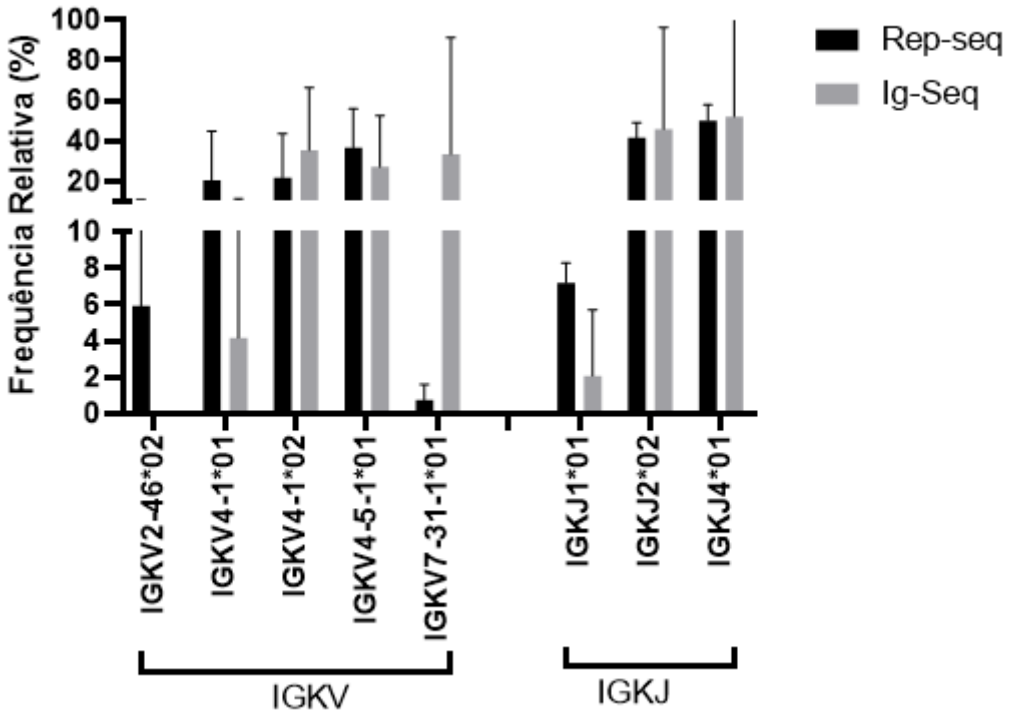

Figura 12 – Frequência relativa média dos genes IGK anotados mais frequentes. Em cinza, a frequência relativa do Ig-Seq, enquanto em preto, temos a frequência relativa do Rep-Seq. A plotagem está restrita aos segmentos gênicos que exibiram uma frequência relativa média igual ou superior a 5% tanto no Rep-Seq quanto no Ig-Seq. Importante mencionar que a média da frequência relativa do Ig-Seq foi realizada com 3 amostras, enquanto a frequência média do Rep-Seq foi realizada com 4 amostras. Não houve diferença estatística da frequência dos segmentos gênicos, por meio de um teste de ANOVA de dois fatores.

4.3.2 Análise da frequência da hipermutação somática (SHM) nos anticorpos equinos

A hipermutação somática é um processo biológico que ocorre nas células B ativadas durante uma resposta imune, principalmente durante a formação de uma resposta imune adaptativa. Este processo é crucial para a geração de anticorpos com alta especificidade e afinidade para um antígeno específico e determina o grau de maturação do repertório.

A comparação entre os transcritos e os anticorpos circulantes específicos direcionados à proteína *spike* do SARS-CoV-2 revela uma ausência de diferença significativa na frequência de mutações nos segmentos V e J das cadeias pesada e leve. Na cadeia pesada, a taxa de mutação somática (SHM) para o segmento IGHV é

de 8,9% ± 0,3 para o Rep-Seq e 8,8% ± 1,1 para o Ig-Seq. O segmento IGHJ exibe uma SHM mais elevada, com média de 15,4% ± 0,4 para o Rep-Seq e 14,9% ± 1,1 para o Ig-Seq.

Para a cadeia leve *lambda*, a SHM média do segmento IGLV é de 8,9% ± 0,3 no Rep-Seq e 12,2% ± 5,5 no Ig-Seq, enquanto o segmento IGLJ apresenta uma média de 2,8% ± 0,2 no Rep-Seq e 4,1% ± 2,2 no Ig-Seq. Quanto à cadeia leve *kappa*, a média da SHM para o segmento IGKV é de 7,3% ± 0,6 no Rep-Seq e 14,1% ± 10,3 no Ig-Seq, enquanto o segmento IGKJ tem uma média de 1,2% ± 0,2 no Rep-Seq e 1% ± 1,4 no Ig-Seq. É importante mencionar que a análise das cadeias leves (*lambda* e *kappa*) baseia-se em apenas três amostras.

Observa-se uma tendência interessante na cadeia pesada, que apresenta uma maior taxa de SHM no segmento gênico J de aproximadamente 15%, enquanto os segmentos J das cadeias leves exibem uma SHM menor **(Figura 13)**.

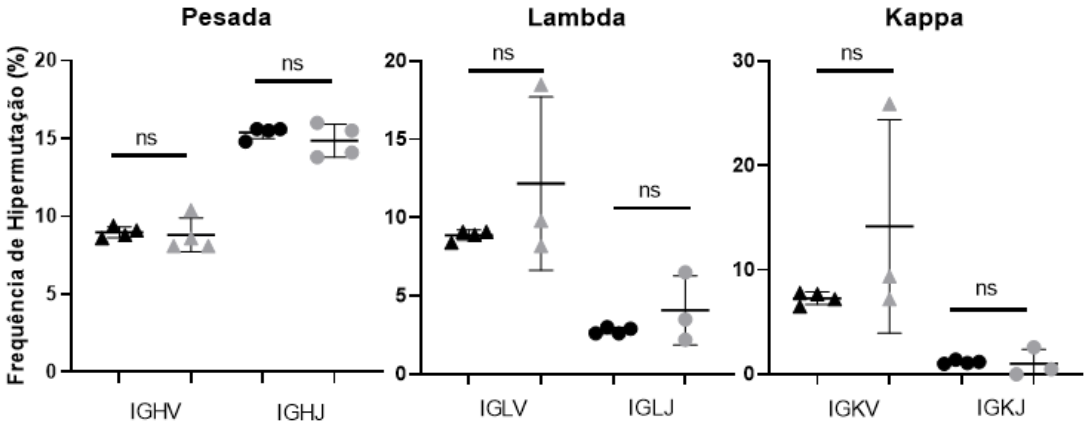

Figura 13 – Comparação da frequência média da hipermutação somática por cadeia em transcritos de anticorpos e anticorpos circulantes. Os segmentos gênicos V são indicados no lado esquerdo em formato de triângulo, enquanto os segmentos gênicos J são exibidos no lado direito em formato circular. A coloração cinza destaca a frequência relativa do Ig-Seq, enquanto a coloração preta representa a frequência relativa do Rep-Seq.

4.3.3 Análise do tamanho e composição dos CDR3 dos anticorpos equinos

Das CDRs que compõem um anticorpo, a mais estudada é a CDR3-H, pois esta é a mais variável, pois é formada pelos 3 segmentos gênicos IGHV, IGHD e IGHJ, e também se acredita que possui o maior contato com os antígenos. Ainda que não tenha muito conhecimento do CDR3 da cadeia leve, formada pelos segmentos

IGLV/IGKV e IGLJ/IGKJ, nos decidimos analisar o tamanho e composição com a finalidade de compreender melhor as características dos anticorpos dos equinos.

O CDR3 da cadeia pesada (CDR3-H) possui uma diversidade de tamanhos **(Figura 14)**. No Rep-seq, 99,8% das sequências possuem CDR3-H com 4 a 26 resíduos de aminoácidos e quase metade das sequencias possuem entre 15 a 19 aminoácidos. Já no Ig-Seq, há uma preferência significativa pelos CDR3 de tamanho 22 para os anticorpos circulantes, $p < 0,05$. A média das amostras apresentam 15,1 aminoácidos nos transcritos e média de 15,9 aminoácido para os anticorpos circulantes.

É notável a presença frequente dos aminoácidos G (glicina), Y (tirosina) e A (alanina) no CDR3-H tanto no repertório dos transcritos quanto no repertório dos anticorpos circulantes. Ao comparar os dois repertórios, destaca-se o aumento do resíduo de aminoácido R (arginina) no repertório sorológico, **** $p < 0,0001$ **(Figura 15)**.

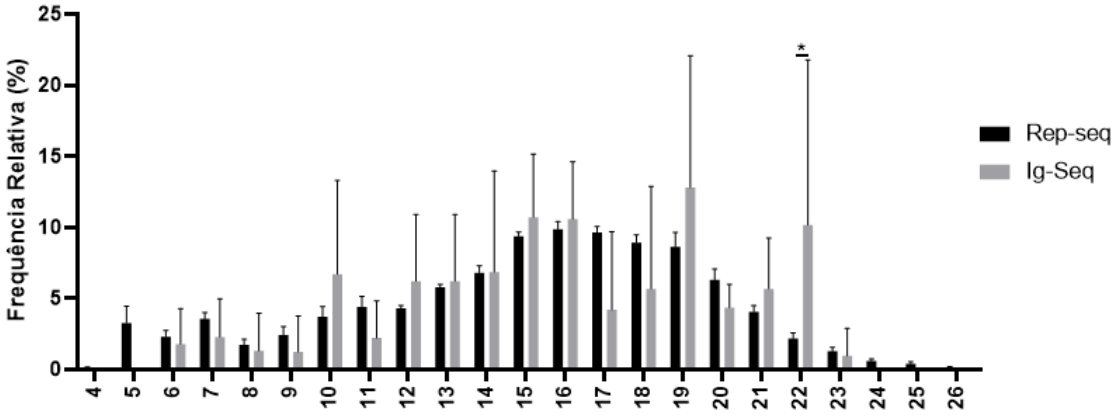

Figura 14 – Frequência relativa média do tamanho (em aminoácidos) do CDR3 da cadeia pesada. Em cinza, a frequência relativa do Ig-Seq, enquanto em preto, temos a frequência relativa do Rep-Seq. A plotagem está restrita aos segmentos gênicos que exibiram uma frequência relativa média igual ou superior a 5% tanto no Rep-Seq quanto no Ig-Seq. A diferença nas médias foi avaliada por meio de um teste de ANOVA de dois fatores (ANOVA 2way). Os asteriscos indicam os seguintes níveis de significância: * $p < 0,05$.

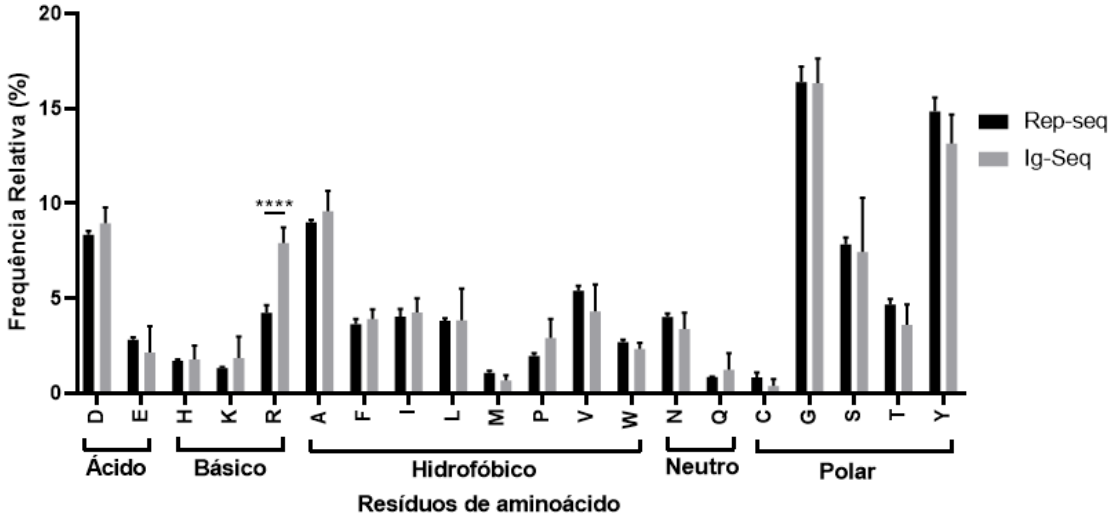

Figura 15 – Frequência relativa média da composição dos aminoácidos no CDR3-H. Em cinza, a frequência relativa do Ig-Seq, enquanto em preto, temos a frequência relativa do Rep-Seq. Os aminoácidos estão agrupados de acordo com a carga do radical, segundo **(Crooks et al., 2004)**. A diferença nas médias foi avaliada por meio de um teste de ANOVA de dois fatores (ANOVA 2way). Os asteriscos indicam os seguintes níveis de significância: **** $p < 0,0001$.

A cadeia leve *lambda*, por sua vez, apresenta variação no tamanho do CDR3-L, compreendendo entre 5 e 17 resíduos de aminoácidos, com uma média de 10,7 no Rep-seq. Na sorologia, há uma alteração na frequência, evidenciada pela diminuição da frequência do tamanho 10 aminoácidos ($p < 0,01$) e um aumento significativo no CDR3 com 11 resíduos de aminoácidos ($p < 0,0001$) **(Figura 16)**.

A composição do CDR3-L revela uma predominância de aminoácidos S (serina), G (glicina) e D (ácido aspártico) nos transcritos. Em contraste, no repertório sorológico, o CDR3 é composto principalmente por S (serina), A (alanina) e L (leucina). Ao realizar uma comparação entre ambos os repertórios, observa-se um aumento estatisticamente significativo de R (arginina) (** $p < 0,01$) e L ($p < 0,0001$) no Ig-Seq, enquanto há uma diminuição de G (** $p < 0,01$) **(Figura 17)**.

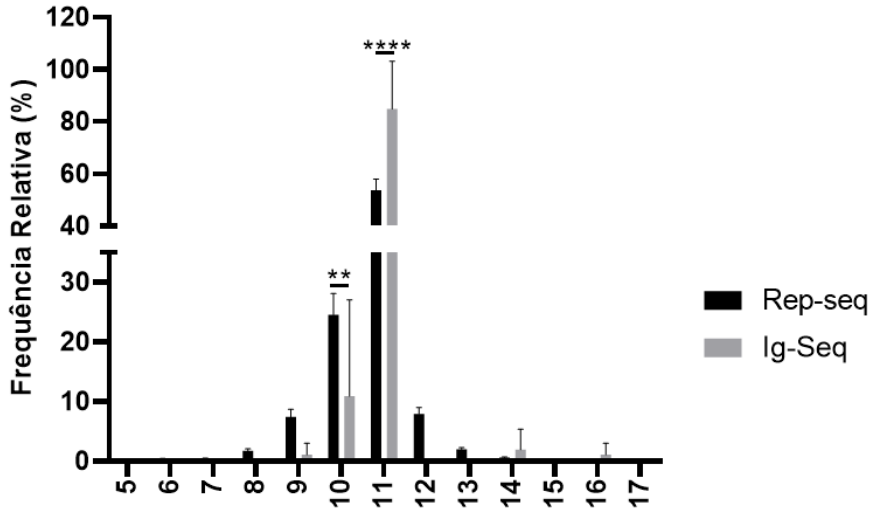

Figura 16 – Frequência relativa média do tamanho (em aminoácidos) do CDR3 da cadeia leve *lambda*. Em cinza, a frequência relativa do Ig-Seq, enquanto em preto, temos a frequência relativa do Rep-Seq. A plotagem está restrita aos segmentos gênicos que exibiram uma frequência relativa média igual ou superior a 5% tanto no Rep-Seq quanto no Ig-Seq. A diferença nas médias foi avaliada por meio de um teste de ANOVA de dois fatores (ANOVA 2way). Os asteriscos indicam os seguintes níveis de significância: ** p < 0,01 e **** p < 0,0001.

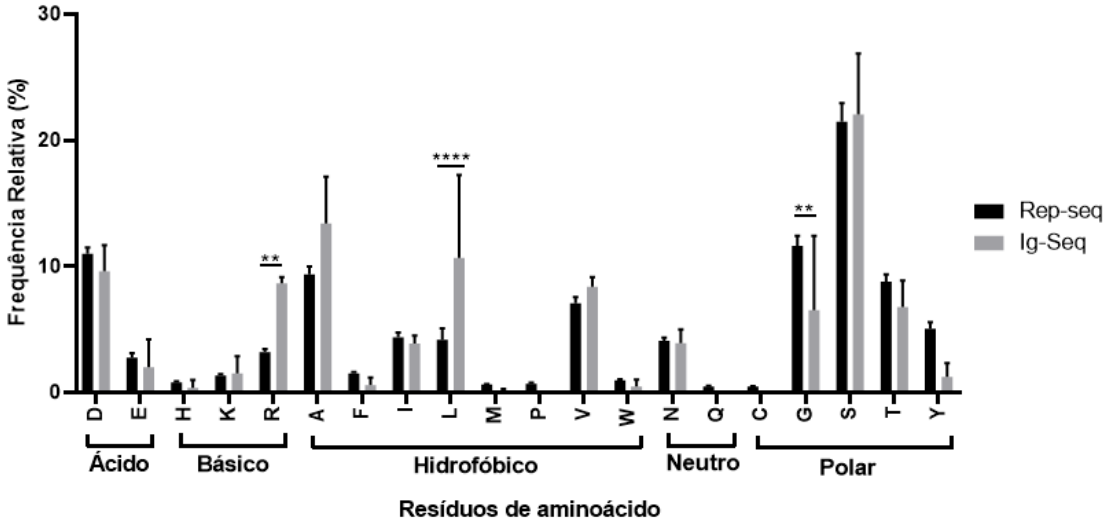

Figura 17 – Frequência relativa média da composição dos aminoácidos no CDR3-L. Em cinza, a frequência relativa do Ig-Seq, enquanto em preto, temos a frequência relativa do Rep-Seq. Os aminoácidos estão agrupados de acordo com a carga do radical, segundo **(Crooks et al., 2004)**. A diferença nas médias foi avaliada por meio de um teste de ANOVA de dois fatores (ANOVA 2way). Os asteriscos indicam os seguintes níveis de significância: ** p < 0,01, e **** p < 0,0001.

No contexto da cadeia leve *kappa*, observa-se que no Rep-seq, 91% dos CDR3-K apresentam um tamanho de 9 resíduos de aminoácidos. Intrigantemente, no Ig-seq, esse número aumenta significativamente para 100% ($p < 0,0001$) **(Figura 18)**.

Quanto aos resíduos de aminoácidos presentes no CDR3-K, merece destaque a presença frequente de Q (glutamina), T (treonina), Y (tirosina) e P (prolina). Interessantemente, não se observam diferenças significativas entre o repertório dos transcritos e o repertório sorológico **(Figura 19)**.

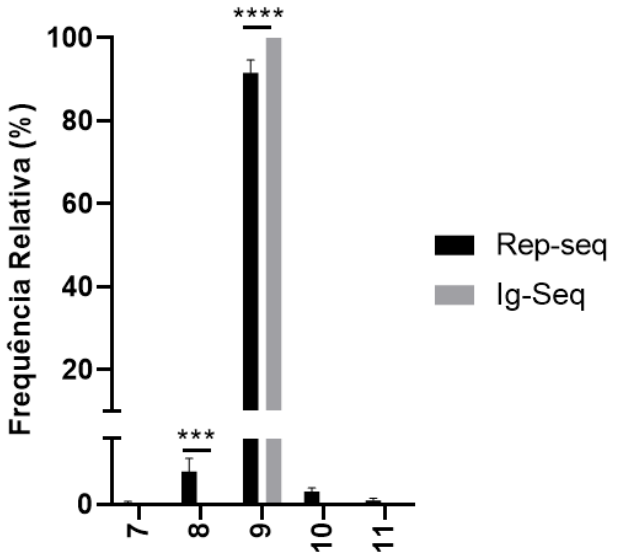

Figura 18 – Frequência relativa média do tamanho (em aminoácidos) do CDR3 da cadeia leve *kappa*. Em cinza, a frequência relativa do Ig-Seq, enquanto em preto, temos a frequência relativa do Rep-Seq. A plotagem está restrita aos segmentos gênicos que exibiram uma frequência relativa média igual ou superior a 5% tanto no Rep-Seq quanto no Ig-Seq. A diferença nas médias foi avaliada por meio de um teste de ANOVA de dois fatores (ANOVA 2way). Os asteriscos indicam os seguintes níveis de significância: *** $p < 0,001$ e **** $p < 0,0001$.

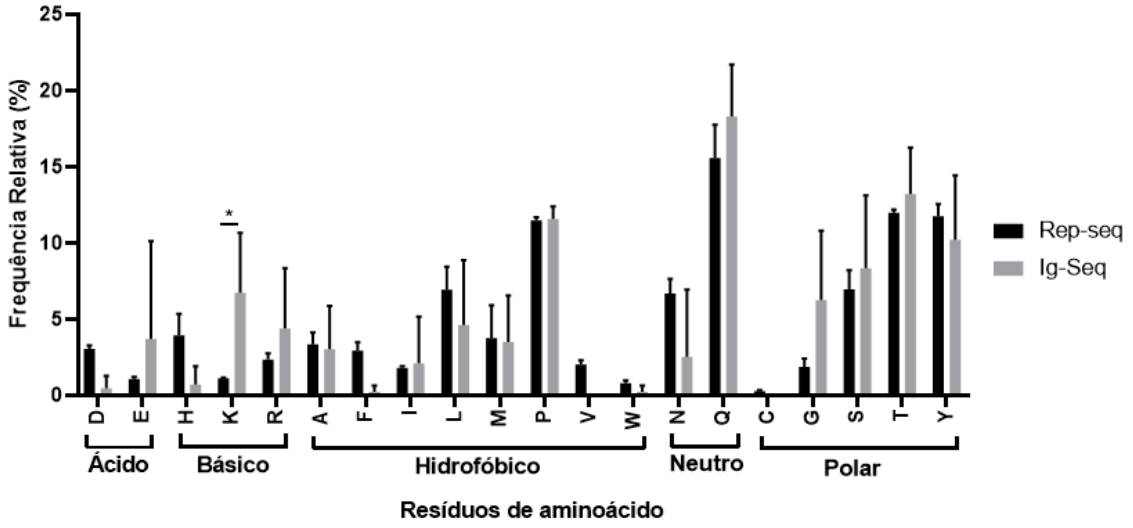

Figura 19 – Frequência relativa média da composição dos aminoácidos no CDR3-K. Em cinza, a frequência relativa do Ig-Seq, enquanto em preto, temos a frequência relativa do Rep-Seq. Os aminoácidos estão agrupados de acordo com a carga do radical, segundo **(Crooks et al., 2004)**. A diferença nas médias foi avaliada por meio de um teste de ANOVA de dois fatores (ANOVA 2way). Os asteriscos indicam os seguintes níveis de significância: * $p < 0,05$.

4.4 Estudo de Caso (Amostra 843-I)

Em parceria com a Universidade Federal do Rio de Janeiro (UFRJ), foi realizado ensaio de neutralização com os plasmas dos quatro cavalos. Observamos, que o cavalo 843-I apresentou altos títulos de neutralização, sendo capaz de neutralizar 90% da infecção *in vitro* em uma diluição de 1:32.000 **(Tabela 5)**.

Tabela 5 – Título de neutralização dos plasmas dos cavalos.

Amostra	PRNT50	PRNT90
834-I	2.048	512
839-I	2.048	1.024
843-I	32.000	32.000
844-I	8.000	4.000

Dessa forma, ao analisar as sequências circulantes específicas do cavalo 843-I, foi possível confrontar os peptídeos trípticos do CDR3 com a biblioteca de sequências criada pelo NGS do mesmo animal. Assim, foram identificadas 19

sequências específicas contra a proteína *spike* para a cadeia pesada, 2 sequências para a cadeia leve *lambda* e nenhuma para a cadeia leve *kappa*. A tabela a seguir apresenta em detalhes os peptídeos trípticos identificados no cavalo 843-I, incluindo o uso gênico dos segmentos V(D)J, rank da clonalidade e o ID de cada uma das sequências **(Tabela 6)**.

As sequências completas dos anticorpos circulantes podem ser verificadas através do alinhamento e plotagem pelo WebLogo, destacando a região variável IGH **(Figura 20)** e IGL **(Figura 21).**

Tabela 6 – Peptídeos específicos (IGH e IGL) encontrados no cavalo 843-I.

Cadeia	ID	Clones*	Rank	Segmento V	Segmento D	Segmento J	Sequência do Peptídeo (CDR3)	Área
IGH	A	260	65	IGHV4-22*01 F	IGHD2-26*01 F	IGHJ6-1*01 F	**TQNGDGLAFYPR**	572.939.808
IGH	B	33	2268	IGHV4-21*01 ORF	IGHD3-7*01 ORF	IGHJ6-3*01 F	**DYGSGVFR**	214.061.176
IGH	C	51	1178	IGHV4-22*01 F	IGHD2-26*01 F	IGHJ3*01 F	SQVYLTLNSLTGEDTAVYYC**AGGAYTA**WGQGILVTVSSASTTAPK	183.690.448
IGH	D	260	65	IGHV4-22*01 F	IGHD2-26*01 F	IGHJ6-1*01 F	**TQDGDGLAFYPR**	131.745.324
IGH	E	163	155	IGHV4-21*01 ORF	IGHD7-33*01 ORF	IGHJ3*01 F	**SYAGSVGD**WGQGILVTVSSASTTAPK	100.855.724
IGH	F	369	43	IGHV4-29*02 F	IGHD2-39*01 ORF	IGHJ6-1*01 F	**SSDSSNTYQYID**WGQGIR	91.900.512
IGH	G	13	6073	IGHV4S1*01 F	IGHD3-7*01 ORF	IGHJ6-3*01 F	**VGDIDY**WGQGVLVTVSSASTTAPK	77.421.528
IGH	H	201	108	IGHV4-21*01 ORF	IGHD7-33*01 ORF	IGHJ4*01 F	**SYAGSVGD**WGQGTLVTVSSASTTAPK	74.538.848
IGH	I	6	12165	IGHV4-29*01 F	IGHD2-36*01 F	IGHJ5*01 F	**YGYDGDIER**	64.394.396
IGH	J	1	73806	IGHV4S1*01 F	IGHD2-26*01 F	IGHJ6-1*01 F	**SLGYGYGGR**	54.889.792
IGH	K	13	6302	IGHV4-22*01 F	IGHD3-1*01 ORF	IGHJ6-1*01 F	**TQNSDGVVTYPR**	43.871.786
IGH	L	271	61	IGHV4-22*01 F	IGHD2-15*01 F	IGHJ6-1*01 F	**SIAHGSWDSSNR**	42.265.325
IGH	M	1	59344	IGHV4-29*01 F	IGHD2-36*01 F	IGHJ6-1*01 F	**SGGQQPSVIDYR**	35.601.312
IGH	N	567	23	IGHV4-22*01 F	IGHD4-25*01 ORF	IGHJ6-1*01 F	**SSSGAASAHYGYR**	20.272.628
IGH	O	21	3972	IGHV4-29*01 F	IGHD7-43*01 ORF	IGHJ4*01 F	**DWPFGY**WGQGTLVTVSSASTTAPK	17.967.288
IGH	P	38	1887	IGHV4-82*01 F	IGHD2-23*01 F	IGHJ6-1*01 F	**LYGLSTDY**WGQGILVTVSSASTTAPK	17.081.149
IGH	Q	271	61	IGHV4-22*01 F	IGHD2-15*01 F	IGHJ6-1*01 F	**SIAHGSWDSR**	13.324.385
IGH	R	1	38762	IGHV4-22*01 F	IGHD2-12*01 F	IGHJ4*01 F	**FSSHYTVIGNR**	5.339.247
IGH	S	62	819	IGHV4-22*01 F	IGHD2-21*01 ORF	IGHJ6-1*01 F	**FPFGINY**WGQGILVTVSSASTTAPK	3.945.454
IGL	A	37	282	IGLV8S2	NA	IGLJ4S1P*01	SGNTATLTITGLQAEDEADYYC**SSLDSNLR**	253.119.088
IGL	B	223	13	IGLV8S1	NA	IGLJ1S1	SGNTAILTISGVQAEDEADYYC**SAADESLR**	229.241.624

*Clones: Número de sequências que se encontram no clone.

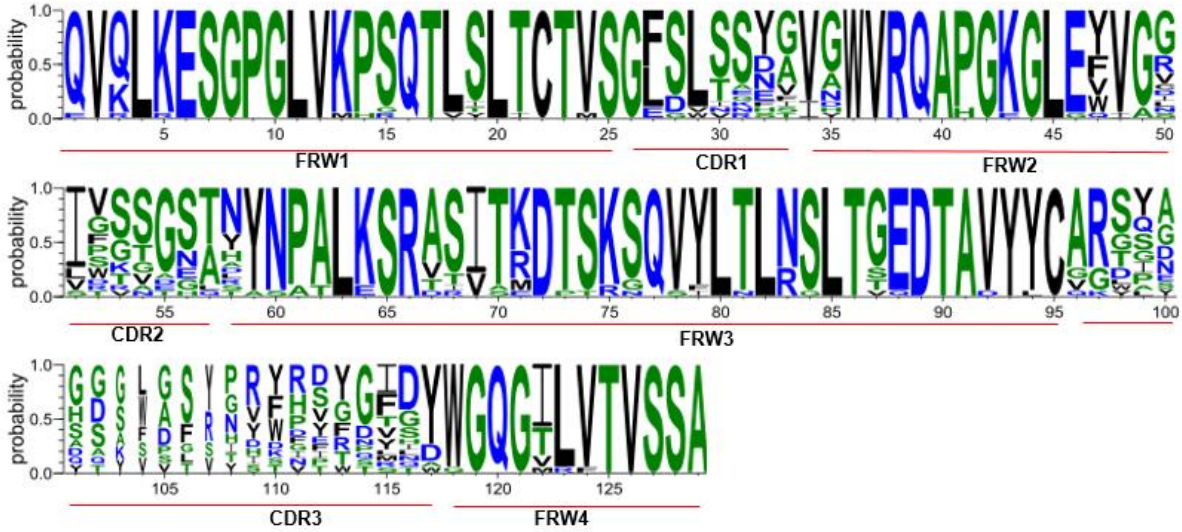

Figura 20 – Distribuição de resíduos de aminoácidos por posição das 19 sequências da cadeia pesada encontradas no cavalo 843-I. As sequências da região variável do IGH foram alinhadas pelo ClustalOmega, em seguida, foi construído o WebLogo para verificar os resíduos mais frequentes em cada posição **(Crooks et al., 2004)**.

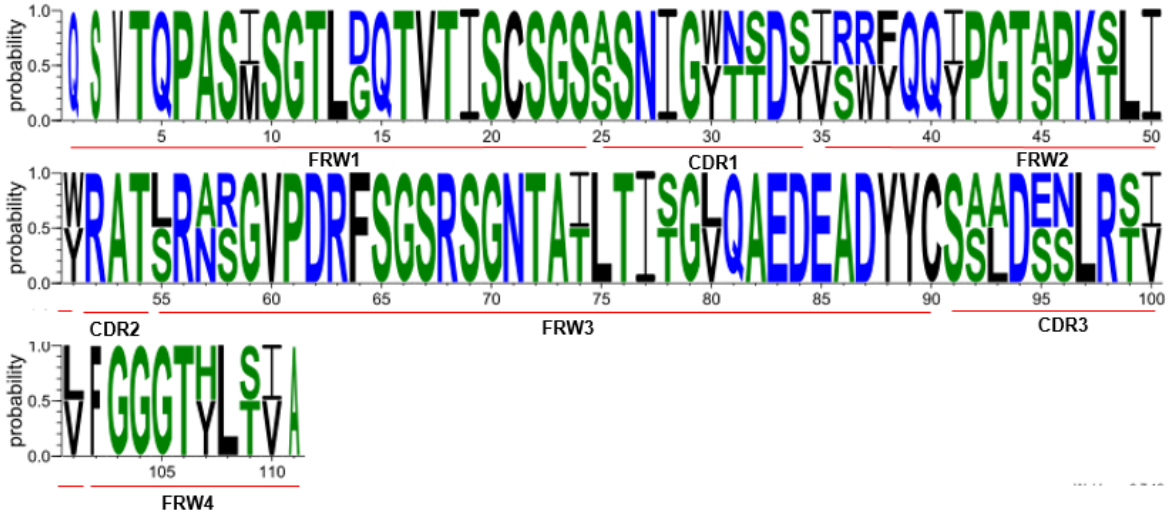

Figura 21 – Distribuição de resíduos de aminoácidos por posição das 2 sequências da cadeia leve *lambda* encontradas no cavalo 843-I. As sequências da região variável do IGH foram alinhadas pelo ClustalOmega, em seguida, foi construído o WebLogo para verificar os resíduos mais frequentes em cada posição **(Crooks et al., 2004).**

Ao confrontar as 21 sequências (19 pesadas e 2 *lambda*) identificadas na proteômica com um banco de dados Cov-Abdab contendo anticorpos anti-SARS-CoV-2 previamente conhecidos (totalizando 12.536 anticorpos), observa-se uma similaridade máxima de 69% das cadeias de anticorpos já descritos com os anticorpos equinos encontrados neste trabalho.

4.4.1 Uso gênico dos anticorpos específicos

Ao observar os anticorpos específicos do cavalo 843-I, verificamos que também apresentam uma clara preferência para os segmentos da família IGHV4 e IGHJ6 para a cadeia pesada **(Figura 22)**. Já para a cadeia leve *lambda*, o cavalo apresenta apenas duas sequências, sendo constituídas pelo segmento gênico IGLV8S1 e IGLV8S2 com o segmento IGLJ1S1 e IGLJ4S1P* (dados não mostrados).

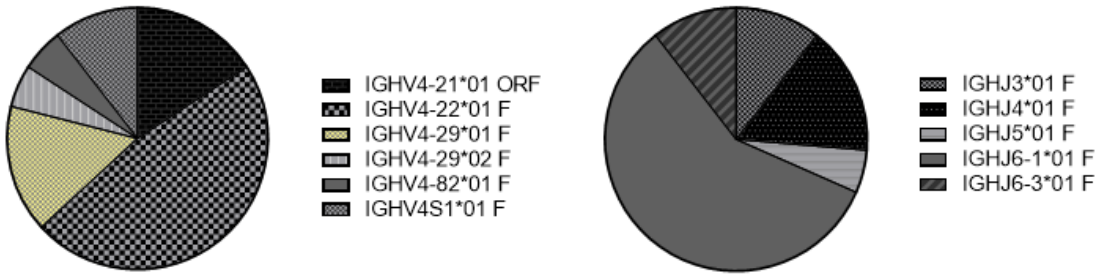

Figura 22 – Uso gênico da cadeia pesada dos anticorpos do cavalo 843-I. Na direita, observa-se o uso gênico dos segmentos V. Na esquerda, observa-se o uso gênico dos segmentos J.

4.4.2 Modelagem estrutural de sequências de anticorpos específicas para a proteína *spike*

Com o objetivo de auxiliar na escolha das sequencias para a expressão dos anticorpos, realizamos a modelagem por meio do *software* ImmunoBuilder, onde foi possível modelar os 38 (2 sequencias da cadeia pesada *versus* 19 sequencias da cadeia pesada), uma vez que não pretendemos sintetizar todas as sequencias encontradas no cavalo 843-I.

Em um passo subsequente, utilizando o SPACE2, uma ferramenta que realiza a clusterização estrutural dos anticorpos com base nas características das seis regiões de loops de complementaridade (CDR) e seus tamanhos, observamos que dentre os 38 anticorpos possíveis, ocorreu a formação de 34

clusters distintos (dados não mostrados). Isso evidencia uma considerável diversidade estrutural entre os anticorpos, indicando variações significativas em suas conformações e propriedades nas regiões de interação com antígenos, em especial no CDRs **(Figura 23)**.

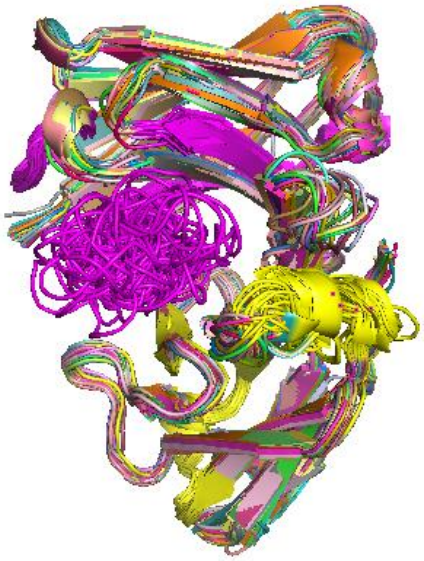

Figura 23 – Alinhamento estrutural dos 38 possíveis anticorpos específicos, alinhados conforme as regiões de FR. Em rosa, se destaca as regiões de CDR da cadeia pesada, em amarelo, destaca-se as regiões de CDR da cadeia leve.

4.4.3 Seleção das sequências para síntese

Optamos por sintetizar ambas as sequências da cadeia leve, uma vez que o cavalo 843-I apresentava apenas duas sequências disponíveis para essa cadeia.

Na seleção das sequências da cadeia pesada, adotamos uma estratégia fundamentada no ranking de clonalidade, que aponta para a presença de expansão clonal nessas sequências. Adicionalmente, empregamos a análise da área sob a curva obtida por meio da cromatografia líquida acoplada à espectrometria de massa (LS/MS-MS). Essa abordagem possibilitou a inferência da abundância relativa do peptídeo na amostra, com as sequências sendo ordenadas com base na área sob a curva **(Tabela 6)**.

Nota-se que as sequências IGH denominadas A e D compartilham o mesmo clone, composto por 260 sequências, e ocupam a 65ª posição no ranking.

Da mesma forma, as sequências L e Q, pertencentes ao mesmo clone composto por 271 sequências, são classificadas na 61ª posição.

Paralelamente, conduzimos um alinhamento com as 19 sequências pesadas utilizando o Clustal Omega **(Figura 24)**. O alinhamento revela que as sequências IGH-A e D, marcadas em verde, assim como as sequências IGH-L e IGH-Q, destacadas em laranja, apresentam similaridades entre si.

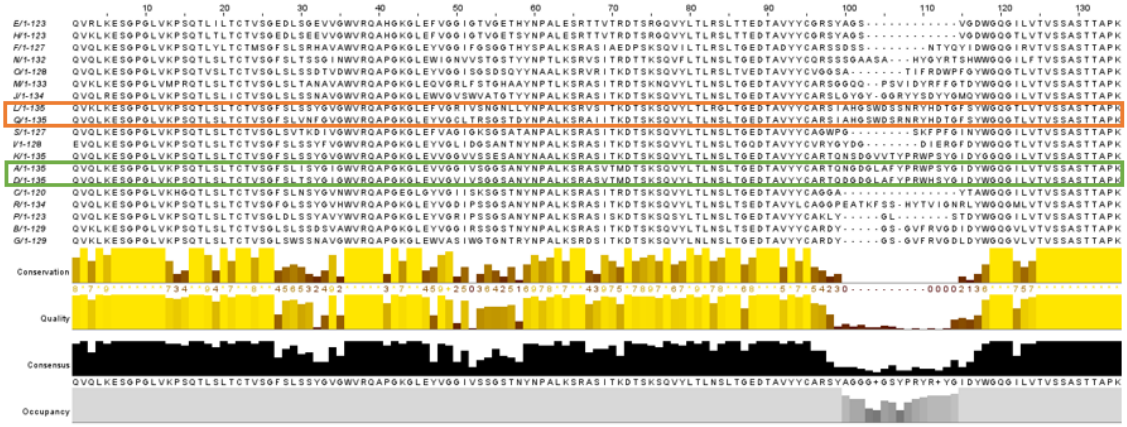

Figura 24 – Alinhamento das 19 sequencias IGH utilizando Clustal Omega. Observa-se a sequências que estão no mesmo clone: IGH-L e IGH-Q, em laranja e as sequências IGH-A e IGH-D, em verde.

Assim, escolhemos as seguintes sequências de cadeia pesadas:

- **IGH-A**: Selecionada devido à maior área sob a curva, indicando maior abundância, e também por possuir um ranking elevado entre os clones na amostra (65º clone mais abundante). Além disso, apresenta outra sequência no mesmo clone (IGH-D).
- **IGH-L**: Escolhida por possuir um ranking elevado entre os clones da amostra (61º clone mais abundante) e também por ter outra sequência presente no mesmo clone (IGH-Q).
- **IGH-N**: Selecionada devido ao ranking mais elevado encontrado (23º clone mais abundante), indicando uma alta expansão do clone.

Uma vez selecionada as sequências, a clonagem será feita combinando as duas cadeias leves com as três pesadas (Tabela 6, sequências escolhidas destacadas de amarelo), possibilitando a expressão de 6 anticorpos.

5 DISCUSSÃO

Atualmente existem 12.536 anticorpos disponibilizados no CoV-AbDab, banco de dados de anticorpos anti-coronavírus. Destes, 133 anticorpos estão em alguma etapa de fase clínica (disponível em: https://www.clinicaltrials.gov/). Alguns anticorpos já foram aprovados em alguns países como o caso do Bamlanivimanb e Etesevimab, Casirivimab e Imdevimab, Sotrovimab e Bebtelovimab aprovados pelo FDA, no Brasil a ANVISA aprovou o Cassirodovimabe e Imdevimabe, Tixagevimab e Cilgavimab e Sotrovimab.

Entretanto, com o surgimento das novas variantes em 2022, tanto o FDA quanto a OMS (Organização Mundial da Saúde) suspenderam o uso da maioria dos anticorpos monoclonais que previamente haviam sidos autorizados, devido à constatação de que não ofereciam uma resposta eficaz contra as novas variantes do vírus, especialmente a Omicrôn **(Cox et al., 2023; Kumari et al., 2022; Tuekprakhon et al., 2022).** Novos tratamentos com anticorpos tem sido estudado, incluindo a pesquisa de pan-anticorpos contra betacoronavírus, **(Johnson et al., 2024)** e a utilização de plasma de equinos como potencial terapia **(Cunha et al., 2021; Zylberman et al., 2020).**

Baseado nisso, este estudo, busca caracterizar o repertório de anticorpos equinos imunizados com a proteína *spike* a fim de encontrar anticorpos neutralizantes. Para isso, foram analisadas as características moleculares dos transcritos dos receptores de células B, juntamente com as imunoglobulinas presentes no plasma de equinos hiperimunizados com a proteína *spike* do SARS-CoV-2. O ensaio de neutralização demonstrou que os plasmas desses animais são capazes de neutralizar a infecção *in vitro* em grandes diluições.

Pelo nosso conhecimento, esse é o primeiro trabalho que estuda o repertório dos anticorpos dos equinos também por uma abordagem sorológica, o que possibilita a análise dos anticorpos específicos contra o antígeno de interesse. Esse também é, o primeiro trabalho a estudar os segmentos gênicos e CDR3 do IGK dos equinos, que compõe apenas 5% dos anticorpos circulantes. Vale ressaltar que a maior parte dos anticorpos em fase clínica e aprovados para diversos tratamentos são compostos pela cadeia leve *kappa* **(Raybould et al., 2024)**.

Através do NSG dos transcritos de PBMC de quatro cavalos hiperimunizados com a proteína *spike,* 99,34%±0,24% das sequências da cadeia

kappa foram devidamente anotadas, enquanto 89,58% ±2,4% das sequências da cadeia *lambda* receberam anotações adequadas. Esses dados demonstram que a combinação do EquCab2.0 com o EquCab3.0 revelou-se uma alternativa promissora para aumentar o número de anotações na cadeia *lambda*, quando comparado ao trabalho de Manso, onde apenas 50% das sequências puderam ser devidamente anotadas **(Manso et al., 2019)**. Isso pode ser devido ao fato de que como ambos os sequenciamentos dos genomas de referência foram realizados por tecnologia de short reads, ainda existam lacunas em seu genoma. Portanto, utilizar os 2 genomas para gerar a construção da biblioteca de *germiline* personalizada permitiu uma maior e melhor identificação dos segmentos gênicos. Além disso, é interessante mencionar que nesse trabalho utilizamos *primers reverse* desenhados de forma diferente ao empregado no estudo anterior de Manso, esses primers anelam-se na região constante da cadeia leve e não no segmento gênico IGLJ.

Na clonotipagem, obtivemos em média 58.000 clones únicos por amostra para a cadeia pesada, 76.000 para a cadeia *lambda* e 10.000 para a cadeia *kappa.* observados em outros trabalhos do nosso grupo de pesquisa, um número semelhante de clones pode ser obtido pelas amostras de PBMC equino **(Manso et al., 2019; Navas et al., 2022)**.

Para a análise dos anticorpos circulantes e específicos, os peptídeos dos CDR3 foram comparados com o banco de sequências obtido por meio do NGS do mesmo indivíduo. Esse método possibilitou a identificação completa das sequências das cadeias dos anticorpos. Como resultado, conseguimos identificar 106 sequências específicas para a cadeia pesada, 49 para a cadeia leve *lambda* e 19 para a cadeia leve *kappa*.

Embora os números identificados possam parecer modestos, eles estão em conformidade com outros estudos que empregaram a mesma metodologia. Em uma pesquisa similar, Lavinder (2014) identificou aproximadamente 100 sequências de anticorpos humanos após vacinação contra tétano **(Lavinder et al., 2014)**. Além disso, em um estudo conduzido por Ferreira (2022), utilizando a mesma abordagem, foram buscadas sequências de anticorpos da cadeia pesada de cavalos em resposta a um pool de venenos de *Loxosceles*, resultando na identificação de 132 sequências **(Ferreira, 2022)**. Já Voss (2021), observou que em uma amostra de paciente que teve COVID-9, foram identificadas 4 linhagens

de anticorpos que compõem 93% dos anticorpos presente no plasma **(Voss et al., 2020)**.

O uso dos genes em IGH e IGL em anticorpos equinos identificados nesse estudo foram similares aos descritos em outros trabalhos em cavalos não imunizados utilizando sequenciamento de Sanger **(Almagro et al., 2006; Sun et al., 2010; Tallmadge et al., 2014, 2013)** e anteriormente identificados pelo nosso grupo também por NGS **(Manso et al., 2019; Navas et al., 2022)** e por outros grupos **(Wibmer and Mashilo, 2022)**. O repertório sorológico estudado nesse trabalho, apesar das poucas sequências encontradas, apresenta o mesmo padrão.

É interessante observar que em todas as cadeias há uma clara preferência de certos segmentos gênicos para a produção dos anticorpos: Por exemplo, na cadeia pesada, 90% dos anticorpos transcritos e circulantes são compostos 4 segmentos gênicos IGHV4 (IGHV4-21*01 ORF, IGHV4-22*01 F, IGHV4-29*01 F e IGHV4S1*01 F), dos 21 segmentos gênicos funcionais disponiveis. Estes são combinado metade das vezes com o segmento IGHJ6-1*01 F**.**

Já para a cadeia *lambda*, há uma preferência de quase 90% para os segmentos gênicos da família IGLV8 que são combinando, em mais de 60% das vezes com o segmento gênico IGLJ4S1P. É interessante observar que, contrariamente à observação anterior de apenas 5% das sequências serem anotadas com IGLJ4S1P **(Manso et al., 2019)**, este trabalho revela uma mudança significativa, indicando que mais de 60% das sequências agora utilizam esse gene catalogado como pseudogene. Uma explicação plausível para essa alteração de frequência pode ser atribuída ao desenho de novos primers e à utilização da nova biblioteca empregados neste estudo. Outra consideração é que, de acordo com Sun, esse segmento gênico é classificado como pseudogene devido à presença de uma citosina extra que altera o *frameshift* **(Sun et al., 2010)**. No entanto, com a inclusão do genoma EquCab 3.0 na biblioteca do *germline*, vimos que esse segmento gênico nesse genoma de referência não apresenta a citosina extra, portanto, não havendo *frameshift*. Assim, seria interessante revisar a classificação desse gene, especialmente considerando sua alta expressão.

E finalmente, a cadeia leve *kappa* também apresentou preferência gênica, onde se destaca os segmentos da família IGKV4 que compõe mais da metade dos segmentos gênicos utilizados para a produção dos anticorpos e que, em 90% das vezes serão combinados com o segmento IGKJ4*01 ou IGKJ2*02.

Apesar dessa ampla variedade de segmentos gênicos V(D)J que podem compor os anticorpos equinos, observa-se que os anticorpos equinos usam preferencialmente alguns segmentos específicos, o que torna esse repertório relativamente pouco diversos pela recombinação V(D)J. De acordo com Navas e colaboradores (2022), a diversidade dos anticorpos equinos é atribuída, em grande parte, à inserção de nucleotídeos juncionais, especialmente na cadeia pesada entre os segmentos V-D e D-J **(Navas et al., 2022)**. Além disso, o cavalo é um dos mamíferos que apresenta uma maior quantidade de segmentos gênicos IGHD **(Walther et al., 2015)**. Apesar da maioria desses segmentos serem classificados como ORF (por apresentaram *gaps* nas extremidades 5’ dos segmentos gênicos), devido à inexatidão no sequenciamento algum desses ORF podem ser funcionais. Um exemplo é o segmento IGHD2-3*01 ORF, presente em cerca de 6% dos anticorpos tanto no transcrito quanto nos anticorpos circulantes.

Navas também levanta a hipótese de que a diversidade dos anticorpos nos equinos decorre da hipermutação somática, uma vez que está relacionada com a maturação da afinidade dos anitocorpos. Portanto, seria razoável esperar uma maior taxa de hipermutação nos anticorpos circulantes anti-*spike*. No entanto, não encontramos diferenças significativas entre o repertório total e o repertório dos anticorpos circulantes anti-*spike*. Essa observação sugere a necessidade de realizar novas investigações, especialmente considerando a limitação deste trabalho devido à quantidade reduzida de sequências específicas encontradas no plasma.

Na literatura, a cadeia leve, devido à sua menor diversidade, é menos estudada do que a cadeia pesada. No entanto, a estrutura do CDR3 da cadeia leve impacta diretamente a estrutura do CDR3 da cadeia pesada, e vice-versa **(Kuroda et al., 2009)**. Embora estudos mais abrangentes sobre a cadeia leve sejam limitados, a interação e influência recíproca entre as duas cadeias destacam a complexidade dos anticorpos. Wibmer (2022) hipotetiza que apesar da baixa diversidade dos anticorpos equinos resultante do uso de poucos

segmentos gênicos, os equinos são capazes de reconhecer inúmeros antígenos, através da diversidade de conformação e composição loops dos CDR3s **(Wibmer and Mashilo, 2022)**.

O CDR3 da cadeia pesada é particularmente importante para as moléculas de anticorpos, pois contém a maior diversidade e também costuma realizar o maior número de contato com o antígeno **(Smider and Smider, 2020)**. O tamanho da região de CDR3 nos anticorpos varia entre as espécies. A maioria dos anticorpos de mamíferos tem loops CDR3-H de 8-16 resíduos **(Zhang et al., 2013)**, enquanto alguns anticorpos humanos têm loops CDR3-H mais longos envolvidos na neutralização de vírus , como no caso de HIV que possui anticorpos com mais de 25 resíduos **(Burton et al., 2012; Burton and Hangartner, 2016)**, influenza **(Kwong and Wilson, 2009)** e poliomielite **(Puligedda et al., 2017)**.

Em humanos, o CDR3-H é composto de sequências que variam de 2 a 26 resíduos de aminoácidos, já os camundongos possuem um tamanho do CDR3-H mais restrito, com pico de 8 à 16 **(Ma et al., 2021)**. Já os anticorpos de vacas possuem CDR3 ultralongos, variando de 40 à 67 resíduos **(Smider and Smider, 2020)** Nesse estudos, identificamos CDR3-H com 4 à 26 resíduos de aminoácidos, similar ao descritos em outros trabalhos **(Almagro et al., 2006; Manso et al., 2019; Tallmadge et al., 2013)**. Essa variação do tamanho do CDR3 na cadeia pesada contribui diretamente para a especificidade dos anticorpos.

O CDR3-H, composto principalmente por resíduos polares e hidrofóbicos, manifesta uma preferência pelos aminoácidos como glicina, tirosina e alanina. Essa preferência é compartilhada com cavalos não imunizados **(Navas et al., 2022; Wibmer and Mashilo, 2022)** e com cavalos imunizados com outros antígenos **(Ferreira, 2022)**. Essa mesma composição é uma característica também observada em humanos e camundongos, indicando uma notável similaridade na composição de aminoácidos entre essas espécies **(Manso et al., 2019)**. Essa similaridade sugere que o cavalo, assim como humanos e camundongos, utiliza estratégias semelhantes na resposta imunológica, ressaltando a importância funcional do CDR3 na interação específica com antígenos.

A manutenção dos resíduos de tirosina e glicina, que juntos compõem mais de 30% do CDR3, é provavelmente essencial para a geração, migração e maturação de células B, como demonstrado em camundongos **(Ippolito et al.,**

2006). Já os resíduos de aminoácidos aromáticos, especialmente a tirosina, que compõe aproximadamente 15% do CDR3-H, são em grande parte determinantes para a especificidade do complexo antígeno-anticorpo **(Peng et al., 2022)**.

Neste estudo, observamos uma variabilidade no tamanho do CDR3-L, varia de 5 a 17 resíduos de aminoácidos. Destaca-se uma preferência notável nos anticorpos específicos por comprimentos entre 10 e 11 resíduos. Esses resultados estão alinhados com descobertas anteriores **(Tallmadge et al., 2014; Wibmer and Mashilo, 2022)**. Em humanos, também é evidenciada uma preferência por CDR3-L com tamanhos de 10 e 11 resíduos, além de 9 resíduos, enquanto em camundongos, a preferência é por 9 resíduos **(Kuroda et al., 2009)**. No que diz respeito ao CDR3-K, observamos uma clara preferência por anticorpos contendo 9 resíduos de aminoácidos, alinhando-se com achados em humanos e camundongos **(Kuroda et al., 2009)**. A literatura também sugere que, além de serem mais longos, os CDR3-L são mais hidrofóbicos **(Townsend et al., 2016)**.

Sobre a composição, no CDR3-L notamos uma predominância de resíduos polares, como serina, glicina e aspartato, nos transcritos. Entretanto, nos anticorpos específicos, essa composição sofre alterações significativas, com um aumento de arginina e leucina, e uma redução de glicina. A presença de aminoácidos pequenos e flexíveis, como glicina, alanina e serina, tanto no CDR3-H quanto no CDR3-L, provavelmente contribui para o aumento da conformação e flexibilidade dos loops. Para o CDR3-K, observou-se uma preferência por resíduos de glutamina, treonina, tirosina e prolina.

Em suma, esta pesquisa analisou as características do repertório de anticorpos circulantes anti-*spike* e comparou com o repertório total de transcritos de anticorpos. Curiosamente, as características dos anticorpos circulantes específicos são muito similares aos transcritos totais de anticorpos equinos.

6 LIMITAÇÃO DO ESTUDO

Estima-se que apenas 2% dos 1-2x10^{11} linfócitos B nos humanos estejam presentes na circulação **(Georgiou et al., 2014)**. Sendo assim, a análise realizada neste estudo utiliza apenas uma pequena parte do sangue periférico destes animais, o que leva a uma limitação para a análise do repertório global dos animais.

Neste estudo, realizamos uma comparação entre milhares de sequências obtidas por meio de Rep-Seq, provenientes de técnicas de NGS, e um número reduzido de sequências identificadas no Ig-Seq. Como resultado, surge uma limitação relacionada à quantidade limitada de dados disponíveis no Ig-Seq, o que pode impactar a robustez estatística do estudo. A disparidade no número de sequências entre as duas abordagens pode influenciar a representatividade da amostra, potencialmente afetando os resultados.

7 CONCLUSÃO

O desenvolvimento deste trabalho permitiu uma análise abrangente das características do repertório sorológico e celular dos anticorpos específicos contra a proteína spike em cavalos hiperimunizados. Vale ressaltar que este foi o primeiro estudo a abordar as cadeias leves, especialmente a cadeia kappa, por meio de uma combinação de abordagens do transcriptoma (Rep-Seq) e sorológica (Ig-Seq).

Nossa investigação revelou sequências específicas equinas anti-spike que apresentam potencial neutralizante. Essas descobertas não apenas contribuem para o entendimento mais completo da resposta imunológica em cavalos, mas também podem ter implicações significativas no desenvolvimento de terapias e vacinas voltadas para a prevenção e tratamento de patógenos associados à proteína spike, como, por exemplo, em cenários de infecções virais.

8 PERSPECTIVAS

- Clonar as sequências escolhidas;
- Expressar os 6 anticorpos em célula de mamífero (Expi 293);
- Realizar testes de reconhecimento antigênico e de neutralização dos aanticorpos *in vitro* contra diferentes variantes da proteína *spike*.

REFERÊNCIAS

Abanades, B., Wong, W.K., Boyles, F., Georges, G., Bujotzek, A., Deane, C.M., 2023. ImmuneBuilder: Deep-Learning models for predicting the structures of immune proteins. Commun Biol 6, 1–8. https://doi.org/10.1038/s42003-023-04927-7

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2002. B Cells and Antibodies, in: Molecular Biology of the Cell. 4th Edition. Garland Science.

Almagro, J.C., Martinez, L., Smith, S.L., Alagon, A., Estevez, J., Paniagua, J., 2006. Analysis of the horse VH repertoire and comparison with the human IGHV germline genes, and sheep, cattle and pig VH sequences. Molecular Immunology 43, 1836–1845. https://doi.org/10.1016/j.molimm.2005.10.017

Alvim, R.G.F., Lima, T.M., Rodrigues, D.A.S., Marsili, F.F., Bozza, V.B.T., Higa, L.M., Monteiro, F.L., Abreu, D.P.B., Leitão, I.C., Carvalho, R.S., Galliez, R.M., Castineiras, T.M.P.P., Travassos, L.H., Nobrega, A., Tanuri, A., Ferreira, O.C., Castilho, L.R., Vale, A.M., 2021. Development and large-scale validation of a highly accurate SARS-COV-2 serological test using regular test strips for autonomous and affordable finger-prick sample collection, transportation, and storage. https://doi.org/10.1101/2020.07.13.20152884

Behring, E. von, 1890. Untersuchungen über das Zustandekommen der Diphtherie-Immunität bei Thieren.

Bonissone, S.R., Lima, T., Harris, K., Davison, L., Avanzino, B., Trinklein, N., Castellana, N., Patel, A., 2020. Serum proteomics expands on high-affinity antibodies in immunized rabbits than deep B-cell repertoire sequencing alone. https://doi.org/10.1101/833871

Botosso, V.F., Jorge, S.A.C., Astray, R.M., de Sá Guimarães, A.M., Mathor, M.B., de Carneiro, P. dos S., Durigon, E.L., Covas, D., de Oliveira, D.B.L., das Neves Oliveira, R., Maria, D.A., Eto, S.F., Gallina, N.M.F., Pidde, G., Squaiella-Baptistão, C.C., Silva, D.T., Villas-Boas, I.M., Fernandes, D.C., Auada, A.V.V., Banari, A.C., de Souza Filho, A.F., Bianconi, C., de Agostini Utescher, C.L., Oliveira, D.C.A., Mariano, D.O.C., Barbosa, F.F., Rondon, G., Kapronezai, J., da Silva, J.G., Goldfeder, M.B., Comone, P., Junior, R.E.C., Pereira, T.T.S., Wen, F.H., Tambourgi, D.V., Chudzinski-Tavassi, A.M., 2022. Anti-SARS-CoV-2

equine F (Ab′)2 immunoglobulin as a possible therapy for COVID-19. Sci Rep 12, 3890. https://doi.org/10.1038/s41598-022-07793-1

Briney, B., Inderbitzin, A., Joyce, C., Burton, D.R., 2019. Commonality despite exceptional diversity in the baseline human antibody repertoire. Nature 566, 393–397. https://doi.org/10.1038/s41586-019-0879-y

Brochet, X., Lefranc, M.-P., Giudicelli, V., 2008. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res 36, W503-508. https://doi.org/10.1093/nar/gkn316

Burton, D.R., Hangartner, L., 2016. Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. Annu Rev Immunol 34, 635–659. https://doi.org/10.1146/annurev-immunol-041015-055515

Burton, D.R., Poignard, P., Stanfield, R.L., Wilson, I.A., 2012. Broadly neutralizing antibodies suggest new prospects to counter highly antigenically diverse viruses. Science 337, 183–186. https://doi.org/10.1126/science.1225416

Calis, J.J.A., Rosenberg, B.R., 2014. Characterizing immune repertoires by high throughput sequencing: strategies and applications. Trends in Immunology 35, 581–590. https://doi.org/10.1016/j.it.2014.09.004

Cheung, W.C., Beausoleil, S.A., Zhang, X., Sato, S., Schieferl, S.M., Wieler, J.S., Beaudet, J.G., Ramenani, R.K., Popova, L., Comb, M.J., Rush, J., Polakiewicz, R.D., 2012. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. Nat Biotechnol 30, 447–452. https://doi.org/10.1038/nbt.2167

Corti, D., Voss, J., Gamblin, S.J., Codoni, G., Macagno, A., Jarrossay, D., Vachieri, S.G., Pinna, D., Minola, A., Vanzetta, F., Silacci, C., Fernandez-Rodriguez, B.M., Agatic, G., Bianchi, S., Giacchetto-Sasselli, I., Calder, L., Sallusto, F., Collins, P., Haire, L.F., Temperton, N., Langedijk, J.P.M., Skehel, J.J., Lanzavecchia, A., 2011. A Neutralizing Antibody Selected from Plasma Cells That Binds to Group 1 and Group 2 Influenza A Hemagglutinins. Science 333, 850–856. https://doi.org/10.1126/science.1205669

Cox, M., Peacock, T.P., Harvey, W.T., Hughes, J., Wright, D.W., Willett, B.J., Thomson, E., Gupta, R.K., Peacock, S.J., Robertson, D.L., Carabelli, A.M., 2023. SARS-CoV-2 variant evasion of monoclonal antibodies based on in vitro

studies. Nat Rev Microbiol 21, 112–124. https://doi.org/10.1038/s41579-022-00809-7

Crooks, G.E., Hon, G., Chandonia, J.-M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res 14, 1188–1190. https://doi.org/10.1101/gr.849004

Cunha, L.E.R., Stolet, A.A., Strauch, M.A., Pereira, V.A.R., Dumard, C.H., Gomes, A.M.O., Monteiro, F.L., Higa, L.M., Souza, P.N.C., Fonseca, J.G., Pontes, F.E., Meirelles, L.G.R., Albuquerque, J.W.M., Sacramento, C.Q., Fintelman-Rodrigues, N., Lima, T.M., Alvim, R.G.F., Marsili, F.F., Caldeira, M.M., Zingali, R.B., de Oliveira, G.A.P., Souza, T.M.L., Silva, A.S., Muller, R., Rodrigues, D. del R.F., Jesus da Costa, L., Alves, A.D.R., Pinto, M.A., Oliveira, A.C., Guedes, H.L.M., Tanuri, A., Castilho, L.R., Silva, J.L., 2021. Polyclonal F(ab')2 fragments of equine antibodies raised against the spike protein neutralize SARS-CoV-2 variants with high potency. iScience 24, 103315. https://doi.org/10.1016/j.isci.2021.103315

de los Rios, M., Criscitiello, M.F., Smider, V.V., 2015. Structural and genetic diversity in antibody repertoires from diverse species. Current Opinion in Structural Biology 33, 27–41. https://doi.org/10.1016/j.sbi.2015.06.002

DeKosky, B.J., Lungu, O.I., Park, D., Johnson, E.L., Charab, W., Chrysostomou, C., Kuroda, D., Ellington, A.D., Ippolito, G.C., Gray, J.J., Georgiou, G., 2016. Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. Proceedings of the National Academy of Sciences 113, E2636–E2645. https://doi.org/10.1073/pnas.1525510113

Ehlich, A., Martin, V., Müller, W., Rajewsky, K., 1994. Analysis of the B-cell progenitor compartment at the level of single cells. Current Biology 4, 573–583. https://doi.org/10.1016/S0960-9822(00)00129-9

Ferreira, M., 2022. ESTUDO DE REPERTÓRIO SOROLÓGICO DE ANTICORPOS DE CAVALO CONTRA VENENO DE ARANHAS DO GÊNERO LOXOSCELES DE IMPORTÂNCIA MÉDICA NO BRASIL.

Ferreira, M.C.E., 2019. PADRONIZAÇÃO DA METODOLOGIA PARA ANÁLISE PROTEÔMICA DE ANTICORPOS EQUINOS ANTI-LOXOSCELES.

Focosi, D., McConnell, S., Casadevall, A., Cappello, E., Valdiserra, G., Tuccori, M., 2022. Monoclonal antibody therapies against SARS-CoV-2. The

Lancet Infectious Diseases 22, e311–e326. https://doi.org/10.1016/S1473-3099(22)00311-5

Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H., Quake, S.R., 2014. The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat Biotechnol 32, 158–168. https://doi.org/10.1038/nbt.2782

Gervásio, J., Ferreira, A., Felicori, L.F., 2023. Yclon: Ultrafast clustering of B cell clones from high-throughput immunoglobulin repertoire sequencing data. https://doi.org/10.1101/2022.02.17.480909

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318

Gupta, N.T., Vander Heiden, J.A., Uduman, M., Gadala-Maria, D., Yaari, G., Kleinstein, S.H., 2015. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. Bioinformatics 31, 3356–3358. https://doi.org/10.1093/bioinformatics/btv359

Horohov, D.W., 2015. The equine immune responses to infectious and allergic disease: A model for humans? Molecular Immunology, Non-rodent animal models for immunology research 66, 89–96. https://doi.org/10.1016/j.molimm.2014.09.020

Iannizzi, C., Chai, K.L., Piechotta, V., Valk, S.J., Kimber, C., Monsef, I., Wood, E.M., Lamikanra, A.A., Roberts, D.J., McQuilten, Z., So-Osman, C., Jindal, A., Cryns, N., Estcourt, L.J., Kreuzberger, N., Skoetz, N., 2023. Convalescent plasma for people with COVID-19: a living systematic review. Cochrane Database Syst Rev 2, CD013600. https://doi.org/10.1002/14651858.CD013600.pub5

Ionov, S., Lee, J., 2022. An Immunoproteomic Survey of the Antibody Landscape: Insights and Opportunities Revealed by Serological Repertoire Profiling. Front Immunol 13, 832533. https://doi.org/10.3389/fimmu.2022.832533

Ippolito, G.C., Schelonka, R.L., Zemlin, M., Ivanov, I.I., Kobayashi, R., Zemlin, C., Gartland, G.L., Nitschke, L., Pelkonen, J., Fujihashi, K., Rajewsky, K., Harry W. Schroeder, J., 2006. Forced usage of positively charged amino acids in immunoglobulin CDR-H3 impairs B cell development and antibody production.

The Journal of Experimental Medicine 203, 1567. https://doi.org/10.1084/jem.20052217

Johnson, N.V., Wall, S.C., Kramer, K.J., Holt, C.M., Periasamy, S., Richardson, S., Suryadevara, N., Andreano, E., Paciello, I., Pierleoni, G., Piccini, G., Huang, Y., Ge, P., Allen, J.D., Uno, N., Shiakolas, A.R., Pilewski, K.A., Nargi, R.S., Sutton, R.E., Abu-Shmais, A.A., Parks, R., Haynes, B.F., Carnahan, R.H., Crowe, J.E., Montomoli, E., Rappuoli, R., Bukreyev, A., Ross, T.M., Sautto, G.A., McLellan, J.S., Georgiev, I.S., 2024. Discovery and Characterization of a Pan-betacoronavirus S2-binding antibody. https://doi.org/10.1101/2024.01.15.575741

Kahn, J.S., McIntosh, K., 2005. History and recent advances in coronavirus discovery. Pediatr Infect Dis J 24, S223-227, discussion S226. https://doi.org/10.1097/01.inf.0000188166.17324.60

Kalbfleisch, T.S., Rice, E.S., DePriest, M.S., Walenz, B.P., Hestand, M.S., Vermeesch, J.R., O′Connell, B.L., Fiddes, I.T., Vershinina, A.O., Saremi, N.F., Petersen, J.L., Finno, C.J., Bellone, R.R., McCue, M.E., Brooks, S.A., Bailey, E., Orlando, L., Green, R.E., Miller, D.C., Antczak, D.F., MacLeod, J.N., 2018. Improved reference genome for the domestic horse increases assembly contiguity and composition. Commun Biol 1, 1–8. https://doi.org/10.1038/s42003-018-0199-z

Kumari, M., Lu, R.-M., Li, M.-C., Huang, J.-L., Hsu, F.-F., Ko, S.-H., Ke, F.-Y., Su, S.-C., Liang, K.-H., Yuan, J.P.-Y., Chiang, H.-L., Sun, C.-P., Lee, I.-J., Li, W.-S., Hsieh, H.-P., Tao, M.-H., Wu, H.-C., 2022. A critical overview of current progress for COVID-19: development of vaccines, antiviral drugs, and therapeutic antibodies. J Biomed Sci 29, 68. https://doi.org/10.1186/s12929-022-00852-9

Küppers, R., Zhao, M., Hansmann, M.L., Rajewsky, K., 1993. Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. EMBO J 12, 4955–4967.

Kuroda, D., Shirai, H., Kobori, M., Nakamura, H., 2009. Systematic classification of CDR-L3 in antibodies: Implications of the light chain subtypes and the VL–VH interface. Proteins: Structure, Function, and Bioinformatics 75, 139–146. https://doi.org/10.1002/prot.22230

Kwong, P.D., Wilson, I.A., 2009. HIV-1 and influenza antibodies: seeing antigens in new ways. Nat Immunol 10, 573–578. https://doi.org/10.1038/ni.1746

Lavinder, J.J., Wine, Y., Giesecke, C., Ippolito, G.C., Horton, A.P., Lungu, O.I., Hoi, K.H., DeKosky, B.J., Murrin, E.M., Wirth, M.M., Ellington, A.D., Dörner, T., Marcotte, E.M., Boutz, D.R., Georgiou, G., 2014. Identification and characterization of the constituent human serum antibodies elicited by vaccination. Proceedings of the National Academy of Sciences 111, 2259–2264. https://doi.org/10.1073/pnas.1317793111

Lee, J., Boutz, D.R., Chromikova, V., Joyce, M.G., Vollmers, C., Leung, K., Horton, A.P., DeKosky, B.J., Lee, C.-H., Lavinder, J.J., Murrin, E.M., Chrysostomou, C., Hoi, K.H., Tsybovsky, Y., Thomas, P.V., Druz, A., Zhang, B., Zhang, Y., Wang, L., Kong, W.-P., Park, D., Popova, L.I., Dekker, C.L., Davis, M.M., Carter, C.E., Ross, T.M., Ellington, A.D., Wilson, P.C., Marcotte, E.M., Mascola, J.R., Ippolito, G.C., Krammer, F., Quake, S.R., Kwong, P.D., Georgiou, G., 2016. Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. Nat Med 22, 1456–1464. https://doi.org/10.1038/nm.4224

Lefranc, M.-P., 2014. Immunoglobulin and T Cell Receptor Genes: IMGT(®) and the Birth and Rise of Immunoinformatics. Front Immunol 5, 22. https://doi.org/10.3389/fimmu.2014.00022

Lefranc, M.P., 2001. Nomenclature of the human immunoglobulin lambda (IGL) genes. Exp Clin Immunogenet 18, 242–254. https://doi.org/10.1159/000049203

Ma, Lina, Tao, X., He, X., Wang, P., Ma, Long, Shi, B., Yao, X., 2021. Analysis of the heterogeneity of the BCR H-CDR3 repertoire in the bone marrow and spleen of 3-, 12-, and 20-month old mice. Immunity & Ageing 18, 17. https://doi.org/10.1186/s12979-021-00231-2

Manso, T., 2019. KIT DE OLIGONUCLEOTÍDEOS ESPECÍFICOS PARA AMPLIFICAÇÃO DA REGIÃO VARIÁVEL DE IMUNOGLOBULINAS DE CAVALO E USO. BR 10 2019 004709 7.

Manso, T., 2018. Sequenciamento de Nova Geração na Análise do Repertório de Anticorpos de Cavalo [WWW Document]. URL https://drive.google.com/drive/folders/1CcQ_PpgfZ0qBNU89qQ1odHN4EuWmzsCX (accessed 2.28.23).

Manso, T.C., Groenner-Penna, M., Minozzo, J.C., Antunes, B.C., Ippolito, G.C., Molina, F., Felicori, L.F., 2019. Next-generation sequencing reveals new

insights about gene usage and CDR-H3 composition in the horse antibody repertoire. Molecular Immunology 105, 251–259. https://doi.org/10.1016/j.molimm.2018.11.017

Mathonet, P., Ullman, C., 2013. The Application of Next Generation Sequencing to the Understanding of Antibody Repertoires. Frontiers in Immunology 4.

McDaniel, J.R., DeKosky, B.J., Tanno, H., Ellington, A.D., Georgiou, G., 2016. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. Nat Protoc 11, 429–442. https://doi.org/10.1038/nprot.2016.024

Murphy, K.M., 2014. Imunobiologia de Janeway, 8ª edição. ed. Artmed.

Muyldermans, S., Smider, V.V., 2016. DISTINCT ANTIBODY SPECIES: STRUCTURAL DIFFERENCES CREATING THERAPEUTIC OPPORTUNITIES. Curr Opin Immunol 40, 7–13. https://doi.org/10.1016/j.coi.2016.02.003

Navas, C., 2022. Análise do repertório de anticorpos de cavalo.

Navas, C., Manso, T., Martins, F., Minto, L., Moreira, R., Minozzo, J., Antunes, B., Vale, A., McDaniel, J.R., Ippolito, G.C., Felicori, L.F., 2022. The major role of junctional diversity in the horse antibody repertoire. Molecular Immunology 151, 231–241. https://doi.org/10.1016/j.molimm.2022.09.011

Obermeier, B., Mentele, R., Malotka, J., Kellermann, J., Kümpfel, T., Wekerle, H., Lottspeich, F., Hohlfeld, R., Dornmair, K., 2008. Matching of oligoclonal immunoglobulin transcriptomes and proteomes of cerebrospinal fluid in multiple sclerosis. Nat Med 14, 688–693. https://doi.org/10.1038/nm1714

Peng, H.-P., Hsu, H.-J., Yu, C.-M., Hung, F.-H., Tung, C.-P., Huang, Y.-C., Chen, C.-Y., Tsai, P.-H., Yang, A.-S., 2022. Antibody CDR amino acids underlying the functionality of antibody repertoires in recognizing diverse protein antigens. Sci Rep 12, 12555. https://doi.org/10.1038/s41598-022-16841-9

Puligedda, R.D., Kouiavskaia, D., Al-Saleem, F.H., Kattala, C.D., Nabi, U., Yaqoob, H., Bhagavathula, V.S., Sharma, R., Chumakov, K., Dessain, S.K., 2017. Characterization of human monoclonal antibodies that neutralize multiple poliovirus serotypes. Vaccine 35, 5455–5462. https://doi.org/10.1016/j.vaccine.2017.03.038

Raybould, M.I.J., Kovaltsuk, A., Marks, C., Deane, C.M., 2021. CoV-AbDab: the coronavirus antibody database. Bioinformatics 37, 734–735. https://doi.org/10.1093/bioinformatics/btaa739

Raybould, M.I.J., Turnbull, O.M., Suter, A., Guloglu, B., Deane, C.M., 2024. Contextualising the developability risk of antibodies with lambda light chains using enhanced therapeutic antibody profiling. Commun Biol 7, 1–13. https://doi.org/10.1038/s42003-023-05744-8

Rio, D.C., Ares, M., Hannon, G.J., Nilsen, T.W., 2010. Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc 2010, pdb.prot5439. https://doi.org/10.1101/pdb.prot5439

Robinson, W.H., 2015. Sequencing the functional antibody repertoire—diagnostic and therapeutic discovery. Nat Rev Rheumatol 11, 171–182. https://doi.org/10.1038/nrrheum.2014.220

Sato, S., Beausoleil, S.A., Popova, L., Beaudet, J.G., Ramenani, R.K., Zhang, X., Wieler, J.S., Schieferl, S.M., Cheung, W.C., Polakiewicz, R.D., 2012. Proteomics-directed cloning of circulating antiviral human monoclonal antibodies. Nat Biotechnol 30, 1039–1043. https://doi.org/10.1038/nbt.2406

Scheid, J.F., Mouquet, H., Feldhahn, N., Seaman, M.S., Velinzon, K., Pietzsch, J., Ott, R.G., Anthony, R.M., Zebroski, H., Hurley, A., Phogat, A., Chakrabarti, B., Li, Y., Connors, M., Pereyra, F., Walker, B.D., Wardemann, H., Ho, D., Wyatt, R.T., Mascola, J.R., Ravetch, J.V., Nussenzweig, M.C., 2009. Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. Nature 458, 636–640. https://doi.org/10.1038/nature07930

Smider, B.A., Smider, V.V., 2020. Formation of ultralong DH regions through genomic rearrangement. BMC Immunol 21, 30. https://doi.org/10.1186/s12865-020-00359-8

Snapkov, I., Chernigovskaya, M., Sinitcyn, P., Quý, K.L., Nyman, T.A., Greiff, V., 2022. Progress and challenges in mass spectrometry-based analysis of antibody repertoires. Trends in Biotechnology 40, 463–481. https://doi.org/10.1016/j.tibtech.2021.08.006

Spoendlin, F.C., Abanades, B., Raybould, M.I.J., Wong, W.K., Georges, G., Deane, C.M., 2023. Improved computational epitope profiling using structural models identifies a broader diversity of antibodies that bind the same epitope. https://doi.org/10.1101/2023.06.09.543890

Steiniger, S.C.J., Dunkle, W.E., Bammert, G.F., Wilson, T.L., Krishnan, A., Dunham, S.A., Ippolito, G.C., Bainbridge, G., 2014. Fundamental characteristics of the expressed immunoglobulin VH and VL repertoire in different canine breeds in comparison with those of humans and mice. Molecular Immunology 59, 71–78. https://doi.org/10.1016/j.molimm.2014.01.010

Steiniger, S.C.J., Glanville, J., Harris, D.W., Wilson, T.L., Ippolito, G.C., Dunham, S.A., 2017. Comparative analysis of the feline immunoglobulin repertoire. Biologicals 46, 81–87. https://doi.org/10.1016/j.biologicals.2017.01.004

Sun, Y., Wang, C., Wang, Y., Zhang, T., Ren, L., Hu, X., Zhang, R., Meng, Q., Guo, Y., Fei, J., Li, N., Zhao, Y., 2010. A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. Developmental & Comparative Immunology 34, 1009–1020. https://doi.org/10.1016/j.dci.2010.05.003

Tallmadge, R.L., Tseng, C.T., Felippe, M.J.B., 2014. Diversity of immunoglobulin lambda light chain gene usage over developmental stages in the horse. Developmental & Comparative Immunology 46, 171–179. https://doi.org/10.1016/j.dci.2014.04.001

Tallmadge, R.L., Tseng, C.T., King, R.A., Felippe, M.J.B., 2013. Developmental progression of equine immunoglobulin heavy chain variable region diversity. Developmental & Comparative Immunology 41, 33–43. https://doi.org/10.1016/j.dci.2013.03.020

Taylor, P.C., Adams, A.C., Hufford, M.M., de la Torre, I., Winthrop, K., Gottlieb, R.L., 2021. Neutralizing monoclonal antibodies for treatment of COVID-19. Nat Rev Immunol 21, 382–393. https://doi.org/10.1038/s41577-021-00542-x

Tonegawa, S., 1983. Somatic generation of antibody diversity. Nature 302, 575–581. https://doi.org/10.1038/302575a0

Townsend, C.L., Laffy, J.M.J., Wu, Y.-C.B., Silva O'Hare, J., Martin, V., Kipling, D., Fraternali, F., Dunn-Walters, D.K., 2016. Significant Differences in Physicochemical Properties of Human Immunoglobulin Kappa and Lambda CDR3 Regions. Frontiers in Immunology 7.

Tuekprakhon, A., Nutalai, R., Dijokaite-Guraliuc, A., Zhou, D., Ginn, H.M., Selvaraj, M., Liu, C., Mentzer, A.J., Supasa, P., Duyvesteyn, H.M.E., Das, R., Skelly, D., Ritter, T.G., Amini, A., Bibi, S., Adele, S., Johnson, S.A.,

Constantinides, B., Webster, H., Temperton, N., Klenerman, P., Barnes, E., Dunachie, S.J., Crook, D., Pollard, A.J., Lambe, T., Goulder, P., Paterson, N.G., Williams, M.A., Hall, D.R., Fry, E.E., Huo, J., Mongkolsapaya, J., Ren, J., Stuart, D.I., Screaton, G.R., Conlon, C., Deeks, A., Frater, J., Frending, L., Gardiner, S., Jämsén, A., Jeffery, K., Malone, T., Phillips, E., Rothwell, L., Stafford, L., 2022. Antibody escape of SARS-CoV-2 Omicron BA.4 and BA.5 from vaccine and BA.1 serum. Cell 185, 2422-2433.e13. https://doi.org/10.1016/j.cell.2022.06.005

Vander Heiden, J.A., Yaari, G., Uduman, M., Stern, J.N.H., O'Connor, K.C., Hafler, D.A., Vigneault, F., Kleinstein, S.H., 2014. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics 30, 1930–1932. https://doi.org/10.1093/bioinformatics/btu138

Voss, W.N., Hou, Y.J., Johnson, N.V., Kim, J.E., Delidakis, G., Horton, A.P., Bartzoka, F., Paresi, C.J., Tanno, Y., Abbasi, S.A., Pickens, W., George, K., Boutz, D.R., Towers, D.M., McDaniel, J.R., Billick, D., Goike, J., Rowe, L., Batra, D., Pohl, J., Lee, J., Gangappa, S., Sambhara, S., Gadush, M., Wang, N., Person, M.D., Iverson, B.L., Gollihar, J.D., Dye, J., Herbert, A., Baric, R.S., McLellan, J.S., Georgiou, G., Lavinder, J.J., Ippolito, G.C., 2020. Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes in COVID-19 convalescent plasma. bioRxiv 2020.12.20.423708. https://doi.org/10.1101/2020.12.20.423708

Walther, S., Rusitzka, T.V., Diesterbeck, U.S., Czerny, C.-P., 2015. Equine immunoglobulins and organization of immunoglobulin genes. Developmental & Comparative Immunology 53, 303–319. https://doi.org/10.1016/j.dci.2015.07.017

Wibmer, C.K., Mashilo, P., 2022. Exploiting V-Gene Bias for Rapid, High-Throughput Monoclonal Antibody Isolation from Horses. Viruses 14, 2172. https://doi.org/10.3390/v14102172

Widyasari, K., Kim, J., 2023. A Review of the Currently Available Antibody Therapy for the Treatment of Coronavirus Disease 2019 (COVID-19). Antibodies (Basel) 12, 5. https://doi.org/10.3390/antib12010005

Wine, Y., Boutz, D.R., Lavinder, J.J., Miklos, A.E., Hughes, R.A., Hoi, K.H., Jung, S.T., Horton, A.P., Murrin, E.M., Ellington, A.D., Marcotte, E.M., Georgiou, G., 2013. Molecular deconvolution of the monoclonal antibodies that comprise

the polyclonal serum response. Proceedings of the National Academy of Sciences 110, 2993–2998. https://doi.org/10.1073/pnas.1213737110

Wu, L., Oficjalska, K., Lambert, M., Fennell, B.J., Darmanin-Sheehan, A., Ní Shúilleabháin, D., Autin, B., Cummins, E., Tchistiakova, L., Bloom, L., Paulsen, J., Gill, D., Cunningham, O., Finlay, W.J.J., 2012. Fundamental Characteristics of the Immunoglobulin VH Repertoire of Chickens in Comparison with Those of Humans, Mice, and Camelids. The Journal of Immunology 188, 322–333. https://doi.org/10.4049/jimmunol.1102466

Yaari, G., Kleinstein, S.H., 2015. Practical guidelines for B-cell receptor repertoire sequencing analysis. Genome Medicine 7, 121. https://doi.org/10.1186/s13073-015-0243-2

Ye, J., Ma, N., Madden, T.L., Ostell, J.M., 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res 41, W34-40. https://doi.org/10.1093/nar/gkt382

Ye, M., Fu, D., Ren, Y., Wang, F., Wang, D., Zhang, F., Xia, X., Lv, T., 2020. Treatment with convalescent plasma for COVID-19 patients in Wuhan, China. J Med Virol 92, 1890–1901. https://doi.org/10.1002/jmv.25882

Zhang, Y., Wang, D., de Lichtervelde, L., Sun, S.B., Smider, V.V., Schultz, P.G., Wang, F., 2013. Functional Antibody CDR3 Fusion Proteins with Enhanced Pharmacological Properties. Angew Chem Int Ed Engl 52, 8295–8298. https://doi.org/10.1002/anie.201303656

Zylberman, V., Sanguineti, S., Pontoriero, A.V., Higa, S.V., Cerutti, M.L., Morrone Seijo, S.M., Pardo, R., Muñoz, L., Acuña Intrieri, M.E., Alzogaray, V.A., Avaro, M.M., Benedetti, E., Berguer, P.M., Bocanera, L., Bukata, L., Bustelo, M.S., Campos, A.M., Colonna, M., Correa, E., Cragnaz, L., Dattero, M.E., Dellafiore, M., Foscaldi, S., González, J.V., Guerra, L.L., Klinke, S., Labanda, M.S., Lauché, C., López, J.C., Martínez, A.M., Otero, L.H., Peyric, E.H., Ponziani, P.F., Ramondino, R., Rinaldi, J., Rodríguez, S., Russo, J.E., Russo, M.L., Saavedra, S.L., Seigelchifer, M., Sosa, S., Vilariño, C., López Biscayart, P., Corley, E., Spatz, L., Baumeister, E.G., Goldbaum, F.A., 2020. Development of a hyperimmune equine serum therapy for COVID-19 in Argentina. Medicina (B Aires) 80 Suppl 3, 1–6.

ANEXO I

Linha de comando utilizada para realizar o pré-processamento das *reads* através do pacote pRESTO da plataforma *Immcantation.*

```
#!/bin/bash

echo "**********  Montando as reads  **********"
/DIRETÓRIO/presto-0.6.2/bin/AssemblePairs.py align -1 AMOSTRA_R1.fastq
-2 AMOSTRA_R2.fastq --coord illumina --rc tail --outname AMOSTRA --log
AMOSTRA_AP.log --faile

echo "**********  Cortar sequencias de baixa qualidade q30  **********"
/DIRETÓRIO/presto-0.6.2/bin/FilterSeq.py quality -s AMOSTRA_assemble-
pass.fastq -q 30 --outname AMOSTRA_q30 --log AMOSTRA_q30.log --failed

echo "**********  Identificacao dos primers **********"
#FWD
/DIRETÓRIO/presto-0.6.2/bin/MaskPrimers.py align -s
AMOSTRA_q30_quality-pass.fastq -p PRIMERS_FWD.fasta --mode tag --
outname AMOSTRA_FWD --log AMOSTRA_MPV.log --failed --maxerror 0.1
#REV
/DIRETÓRIO/presto-0.6.2/bin/MaskPrimers.py align -s
AMOSTRA_FWD_primers-pass.fastq -p PRIMERS_REV.fasta --mode tag --
outname AMOSTRA_REV --log AMOSTRA_REV_MPJ.log --failed --revpr --
maxerror 0.1

echo "**********  filtrar por tamanho de read  **********"
/DIRETÓRIO/presto-0.6.2/bin/FilterSeq.py length -s
AMOSTRA_REV_primers-pass_reheader.fastq -n 100 --outname
AMOSTRA_trim --log AMOSTRA_trim_FS_length.log --failed

#As sequências devem estar no formato fasta para análise de germline. A
conversão dos arquivos fastq em fasta é feita pela linha de comando abaixo
```

```
cat AMOSTRA_trim_length-pass.fastq | paste - - - - | sed 's/^@/>/g'| cut -f1-2 |
tr '\t' '\n' > AMOSTRA-FINAL.fasta
```

ANEXO II

Programa desenvolvido pelo doutorando João Gervásio para a segregação de sequências lambda e kappa por meio dos *primer*. Este programa, implementado em Python, é projetado para operar com arquivos *.fasta, gerados durante a fase de pré-processamento. Ao aplicar o *primer*, o programa cria dois novos arquivos fasta que contêm exclusivamente as sequências lambda e kappa presentes na amostra.

```
#!/usr/bin/env python3
import sys

file = sys.argv[1]

f = open(file,'r')

lambda_chain = open(file.replace(".fasta","_")+"lambda_chain.fasta",'w')
kappa_chain = open(file.replace(".fasta","_")+"kappa_chain.fasta",'w')
l = False
k = False

for x in f:
        if x.find(">") != -1:
                if x.find("IgL") != -1:
                        lambda_chain.write(x)
                        l = True
                        k = False
                else:
                        kappa_chain.write(x)
                        l = False
                        k = True
        else:
                if l == True:
                        lambda_chain.write(x)
```

```
        else:
            kappa_chain.write(x)
```

ANEXO III

Arquivo auxiliar da cadeia leve *lambda* e da cadeia leve *kappa.* Esses arquivos contém as informações dos genes J da *germiline*: nome do gene J, frame de leitura, tipo da cadeia, posição de parada do CDR3. A posição de parada do CDR3 do gene J é dada pelo primeiro nucleotídeo da trinca do motif (F/W-G-X-G), onde X é qualquer aminoácido.

OBS: As posições dos frames de leitura são ‘0-based’, ou seja, se for 1º frame de leitura, deve ser colocado 0. Se for 2º frame de leitura, deve colocar 1.

```
# The chain type, first coding frame start position, chain type, CDR3 stop.
#positions are 0-based



IGLJ1S1          1      JL     6      2
IGLJ2S1P  1      JL     6      0
IGLJ3S1P  1      JL     4      0
IGLJ4S1P*02      0      JL     8      0
IGLJ4S1P*01      0      JL     8      0
IGLJ5S1*02  1    JL     6      1
IGLJ5S1*01  1    JL     6      1
IGLJ6S1*02  1    JL     6      1
IGLJ6S1*01  1    JL     6      1
IGLJ7S1          1      JL     6      1

# The chain type, first coding frame start position, chain type, CDR3 stop.
#positions are 0-based

IGKJ1*01    1    JK     6      1
IGKJ2*01    2    JK     7      1
IGKJ2*02    2    JK     7      1
IGKJ3*01    1    JK     6      1
IGKJ4*01    1    JK     6      1
IGKJ5*01    1    JK     6      1
```

IGKJ5*02	1	JK	6	1

ANEXO IV

O script em R abaixo foi desenvolvido para análise da clonotipagem, focando na cadeia leve *kappa*. Vale ressaltar que, para a análise da proteomica, foram realizadas pequenas modificações, mantendo, contudo, a mesma estrutura. O script começa definindo o diretório das amostras e criando um *dataframe* que contém os arquivos de entrada, ou seja, os relatórios e resultados da clonotipagem. Em seguida, um loop é utilizado para processar cada amostra individualmente. O script importa e organiza os dados dos relatórios e clonotipagem, realiza junções de dados, e conduz análises detalhadas, tais como a frequência genética de genes V e J, a hipermutação somática, a distribuição do tamanho do CDR3, e a composição do CDR3 em termos de aminoácidos. Finalmente, os resultados são exportados para um arquivo Excel.

```
library(data.table)
library(tidyverse)
library(dplyr)
library(openxlsx)
library(Biostrings)
library(stringr)
library(data.table)

#Definindo o caminho das minhas amostras, foi definido pela grnade pasta
meus_arquivos <- "/home/yala/Documentos/YALA/5.CLONOTIPAGEM-
ANALISE"

#Cria um dataframe contendo os arquivos de INPUT que sairam do YClon de
cada amostra
INPUT <- data.frame(report = sort(c(list.files(path = meus_arquivos,
                                    recursive = TRUE,
                                    all.files = TRUE,
                                    pattern = "clonotyped_report.tsv.gz"))), #todos os
arquivos com esse padrao entra na coluna de report
                    clonotipado = sort(c(list.files(path = meus_arquivos,
```

```
                                    recursive = TRUE,
                                    all.files = TRUE,
                                    pattern = "clonotyped.tsv.gz")))) #todos os
arquivos com esse padrao entra na coluna de clonotipado

#Cria um loop onde todos as análises sao feitas das amostras.
for(y in 1:nrow(INPUT)){
 print(INPUT[y,])

  #Cria uma variavel REPORT com a coluna report do df INPUT
  REPORT <- fread(paste(meus_arquivos,
                        INPUT$report[y],
                        sep="/"),
                  header = FALSE, sep = "\t")
  # REPORT <- fread(paste(meus_arquivos,
  #                       "PESADO/834-I-H_db-
pass_YClon_clonotyped_report.tsv.gz",
  #                       sep="/"),
  #                 header = FALSE, sep = "\t")
  #Renomeando os nomes das colunas do report
  colnames(REPORT) <- c("sequence_id",
"seq_count","most_common_cdr3","clone_id")
  # Remover tudo que aparece depois do "|" na coluna sequence_id
  REPORT$sequence_id <- sub("\\|.*", "", REPORT$sequence_id)

  #Importar arquivo da Clonotipagem e organizar a tabela
  CLONOTIPAGEM <- fread(paste(meus_arquivos,
                              INPUT$clonotipado[y],
                              sep="/"),
                       header = TRUE, sep = "\t")
  # CLONOTIPAGEM <- fread(paste(meus_arquivos,
  #                             "PESADO/834-I-H_db-pass_YClon_clonotyped.tsv.gz",
  #                             sep="/"),
```

```
  #                   header = TRUE, sep = "\t")
  CLONOTIPAGEM <- CLONOTIPAGEM %>% select(sequence_id, v_call,
v_identity, j_call, j_identity, cdr3, cdr3_aa, clone_id, clone_seq_count)
  # Ordenar o data frame em ordem decrescente pela coluna
  CLONOTIPAGEM <- arrange(CLONOTIPAGEM, desc(clone_seq_count))
#nesse momento nao e necessario
  # Remover tudo que aparece depois do "|" na coluna sequence_id no
merged_data
  CLONOTIPAGEM$sequence_id <- sub("\\|.*", "",
CLONOTIPAGEM$sequence_id)
  #Manter na coluna apenas o 1o anotação e deletar tudo depois da vírgula
  CLONOTIPAGEM$v_call <- c(sub(",(.*)", "",CLONOTIPAGEM$v_call))
#exclui tudo depois da virgula, estuda os alelos
  #CLONOTIPAGEM$v_call <- c(sub("-(.*)", "",CLONOTIPAGEM$v_call))
#exclui tudo depois do -, só estuda familia genica
  CLONOTIPAGEM$j_call <- c(sub(",(.*)", "",CLONOTIPAGEM$j_call))
#exclui tudo depois da virgula, estuda os alelos
  #CLONOTIPAGEM$j_call <- sub("\\*.*", "", CLONOTIPAGEM$j_call) #exclui
tudo depois do *, só estuda familia genica

  # Junta os df atraves da coluna "sequence_id" que está presente em ambas
  #REPORT <- left_join(REPORT[, names(REPORT) != "clone_id"],
CLONOTIPAGEM, by = "sequence_id") #não está funcionado
  REPORT <- merge(REPORT, CLONOTIPAGEM, by = "sequence_id", all.x =
TRUE) #CORRIGIR, PQ TEM A MESMA COLUNA 2x

  #Daqui para baixo, é apenas análise dos dados

  ##Análise da frequência genica
  #V
  ValorVabsoluto <- table(REPORT$v_call)
  ValorVrelativo <- prop.table(table(REPORT$v_call))
  TabelaGeneV <- data.frame (ValorVabsoluto, ValorVrelativo)
```

```
  #J
  ValorJabsoluto <- table(REPORT$j_call)
  ValorJrelativo <- prop.table(table(REPORT$j_call))
  TabelaGeneJ <- data.frame (ValorJabsoluto, ValorJrelativo)

  #Hipermutacao somatica
  #Colocar as sequencias de forma rankeada pela coluna do
clone_seq_count
  #V
  hiperV <- REPORT %>%
    mutate(v_identity = 100 - v_identity) %>%
    group_by(v_call) %>%
    summarise(hipermutacao_v = mean(v_identity))
  #J
  hiperJ <- REPORT %>%
    mutate(j_identity = 100 - j_identity) %>%
    group_by(j_call) %>%
    summarise(hipermutacao_j = mean(j_identity))

  #Análise do CDR3
  #Tamanho
  REPORT$cdr3_tam <- nchar(REPORT$cdr3_aa)
  #Cria uma tabela de frequencia dos tamanhos do cdr3_aa
  CDR3absoluto <- table(REPORT$cdr3_tam)
  CDR3relativo <- prop.table(table(REPORT$cdr3_tam))
  TAM_CDR3 <- data.frame (CDR3absoluto, CDR3relativo)
  # Criar o histograma da coluna "cdr3_tam"
  #hist(REPORT$cdr3_tam, main = "Distribuição de Frequência do tamanho
do CDR3", xlab = "Tamanho CDR3", ylab = "Frequência")

  #Análise da composição do CDR3
  # Cria novo data frame apenas com a coluna cdr3_aa da REPORT
  Analise_CDR3 <- data.frame(cdr3_aa = REPORT$cdr3_aa)
```

```
  #cria 20 colunas no data frame, e conta quantos caracteres tem na
sequencia de aa na coluna cdr3_aa
  amino_acidos <- c("A", "C", "D", "E", "F", "G", "H", "I", "K", "L", "M", "N", "P",
"Q", "R", "S", "T", "V", "W", "Y")
  #Conta a quantidade de determinados caracteres/aminoacidos que estao
nas sequencias do CDR3
  for (aa in amino_acidos) {
    Analise_CDR3[[aa]] <- str_count(REPORT$cdr3_aa, aa)
  }
   COMP_CDR3 <- list()
    for (x in amino_acidos) {
    frequencia <- sum(Analise_CDR3[x]) / sum(REPORT$cdr3_tam)
    COMP_CDR3[[x]] <- frequencia
  }
  #Cria um df com as informações da freq dos aminoacidos
  COMP_CDR3 <- data.frame(amino_acido = names(COMP_CDR3),
frequencia = unlist(COMP_CDR3))

    # Exportando! Criando arquivos finais
  # Uma unica tabela do excel com 6 abas
   workbook <- createWorkbook()
   addWorksheet(workbook, sheetName = "Freq_V")
   writeData(workbook, sheet = 1, x = TabelaGeneV)
   addWorksheet(workbook, sheetName = "Freq_J")
   writeData(workbook, sheet = 2, x = TabelaGeneJ)
   addWorksheet(workbook, sheetName = "Hiper_V")
   writeData(workbook, sheet = 3, x = hiperV)
   addWorksheet(workbook, sheetName = "Hiper_J")
   writeData(workbook, sheet = 4, x = hiperJ)
   addWorksheet(workbook, sheetName = "Tamanho_CDR3")
   writeData(workbook, sheet = 5, x = TAM_CDR3)
   addWorksheet(workbook, sheetName = "Composicao_CDR3")
   writeData(workbook, sheet = 6, x = COMP_CDR3)
```

```
  saveWorkbook(workbook, paste(meus_arquivos,
                        str_replace(INPUT$report[y],
                                    "report.tsv.gz",
                                    "analise_resultado.xls"),
                        sep = "/"))

} #fechando loop inicial
```