

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOQUÍMICA E IMUNOLOGIA

Carlena Tahina Navas de Reyes

Análise do repertório de anticorpos de cavalo

Belo Horizonte

2022

Carlena Tahina Navas de Reyes

Análise do repertório de anticorpos de cavalo

Tese apresentada à Universidade Federal de Minas Gerais, como parte das exigências do Programa Interunidades de Pós-graduação em bioquímica e imunologia, para a obtenção do título de Doutora em bioquímica e imunologia.

Orientadora: Prof.^a Dra. Liza Figueiredo Felicori Vilela

Belo Horizonte

2022

043

Reyes, Carlena Tahina Navas de.

Análise do repertório de anticorpos de cavalo [manuscrito] / Carlena Tahina Navas de Reyes. – 2022.

174 f. : il. ; 29,5 cm.

Orientador: Liza Figueiredo Felicori Vilela.

Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Bioquímica e Imunologia.

1. Bioquímica e imunologia. 2. Anticorpo. 3. Imunização Passiva. 4. Aranha Marrom Reclusa. 5. Toxina Tetânica. I. Vilela, Liza Figueiredo Felicori. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 612



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 TESE DE DOUTORADO DE CARLENA TAHINA NAVAS DE

REYES. Aos vinte e cinco dias do mês de agosto de 2022 às 14:00 horas, reuniu-se de forma “on line” utilizando a plataforma “Zoom” e de forma presencial na sala 301 CAD1, na Universidade Federal de Minas Gerais, a Comissão Examinadora da tese de Doutorado, indicada *ad referendum* do Colegiado do Curso, para julgar, em exame final, o trabalho intitulado "Análise do repertório de anticorpos de cavalos", requisito final para a obtenção do grau de Doutor em Ciências: Biologia Molecular. 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. Silvia Boscardin (Universidade de São Paulo), aprovada; Dr. André Vale (Universidade Federal de Rio de Janeiro), aprovada; Dra. Glória Regina Franco (Universidade Federal de Minas Gerais), aprovada; Dra. Ana Maria Caetano de Faria (Universidade Federal de Minas Gerais), aprovada; Dra. Liza Figueiredo Felicori Vilela - Orientadora (Universidade Federal de Minas Gerais), aprovada. Pelas indicações a candidata foi considerada:

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 presente Ata que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 25 de agosto de 2022.

Silvia B. Boscardin
 Dra. Silvia Boscardin (USP)

André Vale
 Dr. André Vale (UFRJ)

Glória Regina Franco
 Dra. Glória Regina Franco (UFMG)

Ana Maria Caetano de Faria
 Dra. Ana Maria Caetano de Faria (UFMG)

Liza Felicori
 Dra. Liza Figueiredo Felicori Vilela - Orientadora (UFMG)

AGRADECIMENTOS

Agradeço, primeiramente, a Deus por sempre me mostrar o caminho a seguir.

Agradeço, à Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos e pelo financiamento deste trabalho.

Agradeço aos meus pais, Pablo e Carmen Elena, por sempre acreditarem no meu potencial e por sacrificarem muito para permitir que eu seguisse estudando e desenvolvendo minha pesquisa.

Agradeço ao meu marido Arnaldo e aos meus filhos Camila e Maximiliano pelo apoio incondicional e paciência ao longo desta jornada.

Agradeço à minha orientadora, Prof. Dra. Liza Felicori, pela confiança e oportunidade de ser sua aluna e acreditar na minha capacidade, obrigada pela paciência e pelo apoio nessa jornada, sou muito grata a tudo o que aprendi nesse período.

Agradeço a Dra. Taciana Manso que me ensinou muito e me apoiou mesmo à distância

Agradeço aos meus companheiros de laboratório, em especial a Marcella, Marcele e Milena, que acompanharam de perto esse trabalho e me ajudaram de infinitas maneiras a desenvolver a pesquisa científica.

Agradeço também aos amigos que conheci durante o doutorado. Laís, Chris, Manu, Raniele, Maura, Igor, Fabio e Luciana. Obrigado por sempre estarem comigo, por ouvirem as lamentações de um resultado ruim, por celebrarem um resultado esperado.

E a todos que direta ou indiretamente contribuíram para a realização deste trabalho, que aqui não foram citados, os meus sinceros agradecimentos.

RESUMO

Em 1901, o primeiro prêmio Nobel de medicina foi concedido pela investigação de soro imune contra a toxina tetânica produzidas pela imunização de cavalos. Depois disso, anticorpos policlonais de cavalo foram produzidos e utilizados para tratamento e profilaxia da difteria, tuberculose, tétano e pneumonia. Hoje, a soroterapia no Brasil é o tratamento específico oferecido pelo SUS para acidentes de aranha do gênero *Loxosceles* e em alguns países chegou a ser usado como tratamento para SARS-CoV-2. Mais de 120 anos se passaram desde a primeira vez que o cavalo foi usado para produzir anticorpos contra diferentes doenças e o que se sabe sobre o sistema imunológico adaptativo desta espécie é muito pouco, principalmente a nível molecular. É por isso que apresentamos uma caracterização do repertório de cadeia pesada (IGH) de anticorpos de cavalo não imunizados e imunizados com veneno de aranha do gênero *Loxosceles* usando a tecnologia HTS. No estudo em cavalos não imunizados, obtivemos uma média de 248.169 clones IgM e 66.141 clones IgG únicos de quatro cavalos adultos domésticos. Observamos que apesar do cavalo usar todos os segmentos gênicos funcionais do IGHV, cerca de 80% de seus anticorpos utilizam apenas três segmentos gênicos, e cerca de 55% utilizam apenas um segmento gênico do IGHJ. Essa diversidade limitada de VJ parece ser compensada pela diversidade juncional desses anticorpos. Neste estudo foi visto que a diversidade juncional em anticorpos de cavalos é muito frequente, presente em mais de 90% dos anticorpos de cavalos. Além disso, o comprimento dessa região parece ser maior em anticorpos de cavalos do que em outras espécies. A adição de nucleotídeos N1 e N2 varia de 0 a 111 nucleotídeos. Esse mecanismo de diversidade pode ser um dos mais importantes para fornecer variabilidade ao repertório de anticorpos equinos. No caso dos cavalos imunizados, descrevemos uma média de 114.933 clones para IgG, com cobertura determinada pela rarefação de 80%. Utilizando os clones dos repertórios antes, após a imunização, e aqueles expandidos, observamos semelhanças entre as características do repertório, mas deve-se notar que no grupo de clones expandidos há três segmentos gênicos IGHV mais utilizados, além de menor hidrofobicidade na região de CDRH3. Este estudo aponta a necessidade de caracterização de outras moléculas envolvidas na geração da diversidade de anticorpos equinos, como a TdT, para melhor entendimento das características do repertório de anticorpos de cavalos. Em suma, nossas análises fornecem novas informações sobre a composição de anticorpos de cavalo, geração de diversidade e particularidades em comparação com outras espécies, como a frequência e o comprimento e adição de N nucleotídeos e pode ser aplicável ao desenho racional de anticorpos sintéticos destinados a uso terapêutico.

Palavras chave

Anticorpo, repertório cavalo, célula b, loxosceles, aranha marrom

ABSTRACT

In 1901, the first Nobel Prize in medicine was awarded for investigating immune serum against tetanus toxin produced by immunized horses. After that, polyclonal horse antibodies were produced and used for treating and prophylaxis diphtheria, tuberculosis, tetanus, and pneumonia. Today, in Brazil, it represents the specific treatment for spider accidents of the genus *Loxosceles*, and in some countries, it has even been used as a treatment for SARS-CoV-2. More than 120 years have passed since the horse was first used to produce antibodies against different diseases. However, very little is known about the adaptive immune system of this species, especially at the molecular level. For that reason, we characterize horses' heavy chain repertoire (IGH) from non-immunized, and *Loxosceles* spider immunized horses using HTS technology. In the non-immunized horses, we obtained an average of 248,169 IgM clones and 66,141 unique IgG clones from four domestic adult horses. Although the horse uses all the functional gene segments of the IGHV, about 80% of its antibodies use only three gene segments, and about 55% use only one gene segment of the IGHJ. This limited diversity of VJ appears to be offset by the junctional diversity of these antibodies. In this study, it was seen that junctional diversity in horse antibodies is very frequent, present in more than 90% of horse antibodies. Furthermore, the length of this region appears to be longer in horse antibodies than in antibodies from other species. The addition of N1 and N2 nucleotides ranges from 0 to 111 nucleotides. This diversity mechanism may be one of the most important in providing variability to the equine antibody repertoire. In the case of immunized horses, we have found an average of 114,933 clones for IgG, with coverage determined by the rarefaction of 80%. Using the clones from the repertoires before, after immunization, and those expanded, we observed similar characteristics among them, except for the use of three IGHV gene segments and a lower hydrophobicity in the CDR-H3 region in the expanded clones. This study points out the need to characterize the molecules involved in the generation of equine antibody diversity, such as TdT, to better understand the characteristics of the horse antibody repertoire. In summary, our analyzes provide new insights on horse antibody composition, generation of diversity, and particularities compared to other species, such as frequency and length, and addition of N nucleotides, and may apply to the rational design of synthetic antibodies intended for therapeutic use.

Key words

Antibody, horse repertoire, b cell, loxosceles, brown spider

LISTA DE FIGURAS

Figura 1: Representação da estrutura de anticorpo.....	17
Figura 2: Representação gráfica das subdivisões da região variável da cadeia pesada de imunoglobulinas.....	19
Figura 3: Eventos moleculares de um rearranjo entre um segmento gênico IGHD e um segmento gênico IGHJ no locus IGH que geram a junção de diversidade N.....	23
Figura 4: Diagrama dos processos que ocorrem no mecanismo de diversidade secundária de anticorpos.....	25
Figura 5: Representação esquemática do locus da cadeia pesada de imunoglobulina de <i>Equus caballus</i>	33
Figura 6: Representação esquemática do locus da cadeia leve <i>Kappa</i> da imunoglobulina de <i>Equus caballus</i>	34
Figura 7: Exemplo de um pipeline para estudar o repertório de anticorpos de células B.....	40
Figura 8: Esquema de imunização de cavalo com vírus inativado SARS-CoV-2.....	41
Figura 9: Organização dos genes para cadeia pesada de anticorpo de cavalo.....	51
Figura 10: Metodologia utilizada para o estudo do repertório de anticorpos da cadeia pesada de cavalos não-imunes e imunes.....	56
Figura 11: Fluxograma da análise de dados do repertório de anticorpos de cavalo.....	57
Figura 12: Exemplo de arquivo IMGT/HighV-QUEST de <i>hotspots</i> localizados em IGHV.....	63
Figura 13: Gel de agarose (1%) dos <i>amplicons</i> de cadeias pesadas de anticorpos equinos.....	66
Figura 14: Frequência do uso dos segmentos gênicos da região variável da cadeia pesada em anticorpos equinos.....	69
Figura 15: Frequência do uso dos segmentos gênicos IGHD em anticorpos equinos.....	70
Figura 16: Estimativas da riqueza e diversidade do repertório da cadeia pesada de anticorpos para IgM e IgG nos quatro cavalos não imunizados.....	71
Figura 17: Caracterização dos clones e do CDR-H3 dos repertórios da região variável de cadeia pesada de cavalo.....	72
Figura 18: Características de Hipermutação Somática (SHM) da região variável de cadeia pesada de Cavalo de isotipos IgG e IgM.....	74

Figura 19: Características de hipermutação somática (SHM) de nucleotídeos da região variável de cadeia pesada.....	76
Figura 20: Análise da junção VDJ.....	78
Figura 21: Alinhamento de sequência múltipla de TdT de diferentes espécies.....	80
Figura 22: Índice de hidrofobicidade CDR-H3 repertório IgM e IgG.....	82
Figura 23: Preferência da fase de leitura (RF) para o segmento gênico IGHD em cavalos não imunizados.....	83
Figura 24: Frequência de Subclasses de IgG.....	86
Figura 25: Gel de agarose 1% dos amplicons de cadeias pesadas de anticorpos provenientes de PBMCs de cavalos imunizados com veneno de <i>Loxosceles</i>	87
Figura 26: Comparação da rarefação e diversidade de IGH entre cavalos não imunes e imunizados com veneno de aranha do gênero <i>Loxosceles</i>	90
Figura 27: Comparação da frequência gênica de anticorpos de cavalos não imunes, imunes e de clones expandidos.....	92
Figura 28: Características do CDR-H3.....	94
Figura 29: Avaliação da frequência de hipermutação somática (SHM) nos cavalos não imunes, imunizados com o <i>pool</i> de venenos das aranhas do gênero <i>Loxosceles</i> e de clones expandidos.....	95
Figura 30: Perfil de mutação de nucleotídeos na região IGHV.....	96
Figura 31: Comparação da diversidade juncional entre os anticorpos de cavalos não imunes, imunes e de clones expandidos.....	99
Figura 32: Comparação de subclasses de IgG entre os anticorpos de cavalos não imunes, imunes e de clones expandidos.....	100

LISTA DE TABELAS

Tabela 1: Resumo das principais características do repertório de anticorpos em mamíferos.....	30
Tabela 2: Sequências dos <i>primers</i> utilizados para amplificar os genes IGH.....	50
Tabela 3: Sequências de indexes Illumina.....	53
Tabela 4: Concentração da biblioteca de <i>amplicons</i> obtida por Qubit em ng/ μ L e mM.....	67
Tabela 5: Visão geral dos resultados de sequenciamento da região variável da cadeia pesada de IgM e IgG de quatro cavalos não imunizados.....	68
Tabela 6: Motivos de AID (RGYW/WRCY) e frequências de mutação nas regiões de CDR e FR.....	75
Tabela 7: Análise de adições de nucleotídeos em anticorpos de cavalo.....	77
Tabela 8: Porcentagem de nucleotídeos presentes nas junções N1 e N2 de anticorpos IgM e IgG de cavalo.....	77
Tabela 9: Percentual da composição de aminoácidos por fase de leitura do segmento gênico IGHD.....	84
Tabela 10: Composição de Aminoácidos por fase de leitura de segmentos gênicos funcionais IGHD em anticorpos de cavalo do IMGT.....	85
Tabela 11: Concentração da biblioteca de sequenciamento obtida por qubit em ng/ μ L e mM.....	88
Tabela 12: Resumo dos dados de sequenciamento MiSeq. Para IGH de cavalos imunizados.....	89
Tabela 13: Características de rarefação e diversidade do repertório de anticorpo de cavalos antes e depois da imunização com veneno de aranha do gênero <i>loxosceles</i>	91
Tabela 14: Combinações de segmentos de genes mais frequentes para cada grupo estudado.....	93
Tabela 15: Comparação dos motivos da AID nos três grupos estudados.....	97
Tabela 16: Frequência dos diferentes tipos de junção nos repertórios estudados.....	98

LISTA DE ABREVIATURAS

A	Base nitrogenada adenina
A ou Ala	Alanina
aa	Aminoácido
AID	<i>Activation Induced Desaminase</i>
Asp	<i>Activation Induced Desaminase</i>
BCR	Receptor de Célula B
BER	<i>Base-excision repair</i>
BRCT	proteína de suscetibilidade ao câncer de mama
C	Base nitrogenada Citocina
cDNA	DNA Complementar
CDR	Região Determinante de Complementariedade, variando de 1 a 3
CEUA	Comissão de Ética no Uso de Animais da Universidade Federal de Minas Gerais
CPPI	Centro de Produção e Pesquisa de Imunobiológicos
CSR	<i>Class-switch Recombination</i>
CSV	Valores Separados por Vírgula
Cys	Cisteína
D	Aspartato
D-PBS	Salina tamponada com fosfato Dulbecco
DMSO	Dimetilsulfóxido
DNA	Ácido desoxirribonucleico
dNTP	Desoxirribonucleotídeos fosfatados
DP	Desvio padrão
DSBR	Maquinaria de reparo de quebra de dupla fita (Double-Strand Break Repair)
EDTA	Ácido etilenodiamino tetra-acético
F	<i>Forward</i>
F	ou Phe Fenilalanina
FASTA	Formato baseado em texto para sequências de nucleotídeo
FASTQ	Formato baseado em texto para sequências de nucleotídeo e a qualidade associada a cada base
FBS	Soro fetal bovino (Fetal Bovine Serum)

FR	Região de Framework, variando de 1 a 4
G	Base nitrogenada Guanina
G ou Gly	Glicina
HTS	High-Throughput Sequencing
HhH	domínios hélice-gancho-hélice
IG	Imunoglobulina
IgA – α	Isotipo imunoglobulina A
IgD – δ	Isotipo imunoglobulina D
IgE – ϵ	Isotipo imunoglobulina E
IgG – γ	Isotipo imunoglobulina G
IGH	Cadeia pesada de imunoglobulina
IGHC	Segmento gênico C da cadeia pesada de imunoglobulina
IGHD	Segmento gênico D da cadeia pesada de imunoglobulina
IGHJ	Segmento gênico J da cadeia pesada de imunoglobulina
IGHV	Segmento gênico V da cadeia pesada de imunoglobulina
IGL	Cadeia leve de imunoglobulina
IGLJ	Segmento gênico J da cadeia leve de imunoglobulina
IGLV	Segmento gênico V da cadeia leve de imunoglobulina
IgM	Imunoglobulina M
IMGT	<i>International Immunogenetics Information System</i>
IVB	Instituto Vital Brasil
kDa	Quilodaltons
MMR	<i>Mismatch repair</i>
N	Nucleotídeos aleatório
NGS	<i>Next Generation Sequencing</i> (Sequenciamento de Nova Geração)
NTC	Controle negativo
ORF	<i>Out of Reading Frame</i>
P	Nucleotídeos Palindrômicos
PBMC	Células mononucleares do sangue periférico
PCR	<i>Polymerase Chain Reaction</i> (Reação em Cadeia da Polimerase)
pRESTO	<i>Repertorie Sequencing Toolkit</i>
PRKDC	protein kinase, DNA-activated, catalytic subunit
Q	<i>Phred Score</i>

R	<i>Reverse</i>
RAG	<i>Recombination-Activating Genes</i> , variando de 1 a 2
RepSeq	<i>Repertoire Sequencing</i> (Sequenciamento de Repertório)
RNA	Ácido ribonucleico
RSS	<i>Recombination Signal Sequence</i>
S ou Ser	Serina
SHM	Hipermutação Somática
SUS	Sistema Único de Saúde
T	Base nitrogenada Timina
TdT	Desoxinucleotidil Transferase Terminal
Trp	Triptofano
XRCC5	X-Ray Repair Cross Complementing 5
XRCC6	X-Ray Repair Cross Complementing 6
Y ou Tyr	Tirosina

SUMÁRIO

1. INTRODUÇÃO	16
1.1. ESTRUTURA E FUNÇÃO DOS ANTICORPOS	16
1.2. ESTRUTURAÇÃO E NUMERAÇÃO DA REGIÃO VARIÁVEL DE ANTICORPOS.....	18
1.3. DIVERSIDADE DE REPERTÓRIO DE ANTICORPOS	20
1.3.1. Mecanismos de diversificação primária	20
1.3.2. Mecanismos de diversificação secundária.....	24
1.4. ESTUDO DO REPERTÓRIO DE ANTICORPOS EM DIFERENTES ESPÉCIES	26
1.4.1 REPERTÓRIO DE ANTICORPOS EM CAMUNDONGOS.....	26
1.4.2. Repertório de anticorpos em cães.....	27
1.4.3. Repertório de anticorpos em bovinos	28
1.4.4. Repertório de anticorpos em Macacos	29
1.4.5. Repertório de anticorpos em Porcos.....	30
1.5. LOCUS DE IMUNOGLOBULINAS DE CAVALOS.....	31
1.6. REPERTÓRIO DE ANTICORPOS PARA CAVALOS	35
1.7. ANÁLISES DAS CARACTERÍSTICAS DO REPERTÓRIO DE ANTICORPOS	36
1.7.1. Pré-processamento das reads	36
1.7.2. Anotação dos segmentos gênicos	36
1.7.3. Análise de clonalidade das sequências	37
1.7.4. Caracterização do Repertório de Anticorpos.....	37
1.8. SOROTERAPIA	40
1.8.1. Soroterapia no combate a picada por animais peçonhentos	42
1.8.1.1. Loxoscelismo.....	43
1.8.1.2. Tratamento do loxoscelismo: antivenenos	44
2. OBJETIVO GERAL	46
2.1. OBJETIVOS ESPECÍFICOS.....	46
3. MATERIAL E MÉTODOS	47
3.1. OBTENÇÃO DE AMOSTRAS DE CAVALO.....	47
3.2. EXTRAÇÃO DE RNA.....	48
3.3. SÍNTESE DE cDNA E PCR MULTIPLEX.....	49
3.4. PREPARAÇÃO DA BIBLIOTECA PARA SEQUENCIAMENTO.....	52

3.5.	PURIFICAÇÃO DA BIBLIOTECA COM ADAPTADORES UTILIZANDO BEADS MAGNÉTICAS	54
3.6.	SEQUENCIAMENTO	55
3.7.	ANÁLISES BIOINFORMÁTICAS DO REPERTÓRIO DE CAVALO.....	56
3.7.1.	Pré-Processamento	57
3.7.2.	Anotação	58
3.7.3.	Análise de clonalidade.....	59
3.8.	ANÁLISE DE REPERTÓRIO.....	60
3.8.1.	Análise de rarefação e construção de curva de riqueza de espécies.....	60
3.8.2.	Determinação do repertório expandido	61
3.8.3.	Determinação da fase de leitura do Gene IGHD	61
3.8.4.	Hipermutação somática	62
3.8.5.	Análise de Junção	64
3.8.6.	Determinação das subclasses IgG	64
3.8.7.	Análises estatísticas	64
4.	RESULTADOS	66
4.1.	CARACTERIZAÇÃO DE REPERTÓRIO NATURAL DE ANTICORPOS DE CAVALO.....	66
4.1.1.	Obtenção de amplicons das cadeias pesadas de anticorpos equinos	66
4.1.2.	Análise das bibliotecas de amplicons de IGH para o sequenciamento do repertório de anticorpos equinos	67
4.1.3.	Uso restrito do gene VDJ em anticorpos de cavalos	67
4.1.4.	Repertório público de anticorpos de cavalo análise de CDR-H3	71
4.1.5.	Caracterização da frequência e padrão de hipermutação somática (SHM) encontrada em anticorpos de cavalos	73
4.1.6.	Caracterização da Diversidade Juncional de Anticorpos de Cavalo	76
4.1.7.	Contribuição das adições de N na hidrofobicidade de CDR-H3	81
4.1.8.	Análise do segmento gênico IGHD	82
4.1.9.	Subclasses de IgG.....	86
4.2.	COMPARAÇÃO DO REPERTÓRIO DE ANTICORPOS DE EQUINOS ANTES E APÓS IMUNIZAÇÃO COM VENENO DE ARANHA LOXOSCELES.....	87
4.2.1.	Obtenção das cadeias pesadas de anticorpos equinos após imunização com veneno de Loxosceles	87

4.2.2. Análise da biblioteca de amplicons de IGH para o sequenciamento do repertório de anticorpos equinos imunizados com veneno de <i>Loxosceles</i>	88
4.2.3. Análise dos sequenciamentos de IGH de cavalos imunizados	88
4.2.4. Comparação do uso gênico de segmentos gênicos V (D) e J entre cavalos não imunes e imunes	91
4.2.5. Tamanho e conteúdo de aminoácidos na região de CDR-H3.....	93
4.2.6. Análise de Hipermutação somática	94
4.2.7. Comparação de diversidade juncional	97
4.2.8. Comparação de subclasses de IgG entre os grupos estudados	99
5. DISCUSSÃO	101
5.1. REPERTÓRIO DE ANTICORPOS NATURAL DE CAVALOS	101
5.2. COMPARAÇÃO DO REPERTÓRIO DE ANTICORPOS ANTES E APÓS A IMUNIZAÇÃO COM VENENO DE ARANHA <i>LOXOSCELES</i>	104
6. CONCLUSÃO.....	107
7. PERSPECTIVAS.....	108
8. REFERÊNCIAS BIBLIOGRÁFICAS	109
ANEXO I.....	128
ANEXO II.....	129
ANEXO III.....	130
ANEXO IV	131
ANEXO V.....	174

1. INTRODUÇÃO

O sistema imune adaptativo possui moléculas capazes de reconhecer patógenos externos, as quais constitui o mecanismo chave de defesa dos vertebrados, sendo essas moléculas chamadas anticorpos (Schroeder, 2015; Sirisinha, 2014; Sun et al., 2012). Os anticorpos têm a capacidade de se ligar de forma específica a antígenos, podendo mediar a neutralização dessas moléculas ou induzir mecanismos que culminam na morte dos patógenos (Litman et al., 1999). Dessa forma, o sistema imune adaptativo desenvolve uma memória imunológica aos patógenos e pode combater uma segunda invasão ou eliminar rapidamente o invasor recorrente, mobilizando uma resposta imune mais rápida e eficiente (Beutler, 2004).

Os principais elementos do nosso sistema imunológico adaptativo são dois tipos principais de linfócitos que podem reconhecer e responder especificamente a determinantes antigênicos de patógenos e toxinas potencialmente perigosas, que são gerados no timo (linfócito T) e na medula óssea ou na Bursa aviária de Fabricius (linfócito B) (Cooper & Alder, 2006). Os anticorpos são produzidos pelos linfócitos B, como receptores de membrana, BCR (receptor de célula B, do inglês *B Cell Receptor*) ou na forma secretada (Hoffman et al., 2016).

Após o reconhecimento do antígeno, as células B geradas na medula óssea se diferenciam e se diversificam. Essas células migram para os órgãos linfoides secundários, onde ocorre expansão clonal, seguida de hipermutação somática. De uma maneira geral, as células com alta afinidade ao antígeno se diferenciam em: células B de memória ou células secretoras de anticorpos (plasmócitos) (Murphy, 2014).

1.1. Estrutura e função dos anticorpos

A unidade estrutural básica de uma molécula de anticorpo consiste em quatro cadeias polipeptídicas, duas cadeias leves (IGL) idênticas (cada uma contendo cerca de 220 aminoácidos) e duas cadeias pesadas (IGH) idênticas (cada uma geralmente contendo cerca de 440 aminoácidos). As quatro cadeias são mantidas juntas por uma combinação de ligações não covalentes e covalentes (dissulfeto). As cadeias leve e pesada geralmente cooperam para formar a superfície de ligação ao antígeno (Alberts et al., 2002) (Figura 1). Algumas espécies, porém, apresentam anticorpos formados por apenas um par de cadeias pesadas, como é o caso de camelos, lhamas e tubarões (Könning et al., 2017).

anticorpo predominante que é produzido nos estágios iniciais da resposta imune (Koshland, 1985);

- Imunoglobulinas G (IgG): É a principal classe de anticorpos no soro humano normal e forma aproximadamente 70% do total de imunoglobulinas. A IgG é um monômero, e é uma imunoglobulina relacionada a fases mais crônicas da infecção e participa efetivamente na memória imunológica dos indivíduos, sendo permanentemente detectável no soro (Lefranc & Lefranc, 2020).

No caso de cadeias leves, nos vertebrados superiores, existem dois tipos de cadeia leve, *kappa* (κ) e *lambda* (λ). Os genes que codificam essas duas cadeias estão em cromossomos separados e a proporção dessas cadeias como parte das imunoglobulinas varia entre as espécies: há predominância de cadeias leves *lambda* em cavalos (Wagner, 2006) e de cadeias leves *kappa* em camundongos (Lavinder et al., 2014).

1.2. Estruturação e numeração da região variável de anticorpos

A região variável da cadeia pesada (IGH) é formada por três segmentos gênicos: *variable* (IGHV), *diversity* (IGHD) e *joining* (IGHJ); e o domínio variável das cadeias leves (IGL) é composto por dois segmentos gênicos: *variable* (IGLV) e *joining* (IGLJ). O domínio variável está formado por as regiões hipervariáveis ou regiões determinantes de complementaridade (CDR, do inglês *Complementarity Determining Regions*), as quais são separadas por regiões mais conservadas, os *frameworks* (FR) que servem de arcabouço para a estrutura dos anticorpos (Georgiou et al., 2014). Os CDRs são as principais regiões responsáveis pelo reconhecimento ao antígeno. Dentre os seis CDRs existentes na molécula do anticorpo, a CDR-H3 é onde está concentrada a maior parte da diversidade em anticorpos, devido ao fato de ser a região formada pela união dos três segmentos gênicos V(D)J na IGH (Georgiou et al., 2014) (Figura 2). Essa região possui variabilidade em seu comprimento, estrutura e sequência (Kuroda et al., 2008).

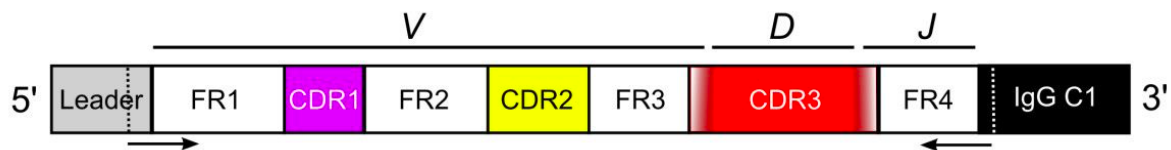


Figura 2. Representação gráfica das subdivisões da região variável da cadeia pesada de imunoglobulinas. O segmento IGHV ou V que é composto por FR1, CDR1, FR2, CDR2 e FR3. O segmento IGHD ou D corresponde ao CDR-H3 (formado por parte do segmento IGHV, todo o segmento IGHD e parte do segmento IGHJ). O segmento IGHJ ou J corresponde ao FR4 e delimita o fim da região. O gene da região constante está representado em preto. Fonte: (Larsen & Smith, 2012).

Para poder comparar as diferentes imunoglobulinas, com tamanho de sequências tão variáveis, é necessária a identificação de resíduos de aminoácidos conservados ou estruturalmente equivalentes (Jung et al., 2001). Sendo assim, diversos grupos sugeriram esquemas de numeração para essas sequências, por meio de alinhamentos múltiplos entre *reads* da família das imunoglobulinas (Chothia & Lesk, 1987). Como os estudos apresentavam divergências em relação à definição de cadeia leve e pesada, Lefranc e colaboradores propuseram um esquema de numeração unificado que levava em conta não só as sequências, mas a estrutura dessas moléculas (Lefranc, 2003; Lefranc et al., 2015). Esse esquema é utilizado hoje como padrão-ouro e implementado pelo *International Immunogenetics Information System* (IMGT), um dos maiores bancos de dados mundiais de imunoglobulinas.

Sendo assim, as subdivisões da região variável em FR e CDR são definidas por:

- Framework 1 (FR1): posições 1 a 26 (Cys23);
- CDR1: posições 27 a 38;
- Framework 2 (FR2): posições 39 a 55 (Trp41);
- CDR2: posições 56 a 65;
- Framework 3 (FR3): posições 66 a 104 (Cys104);
- CDR-H3: posições 105 a 117;
- Framework 4 (FR4): posições 118 a 129 (Trp/Phe118).

* os resíduos de aminoácidos em parêntesis indicam os aminoácidos altamente conservados nas imunoglobulinas.

Também, o IMGT classificou os segmentos gênicos como funcionais e não-funcionais. Os genes funcionais são aqueles que podem ser rearranjados e produzirem anticorpos (Brochet et al., 2008). Já os genes não-funcionais podem ser classificados como: pseudogenes (genes que possuem códon de parada ou mutações que alterem a fase de leitura) (Giudicelli et al., 2004) e fase de leitura aberto (ORF, do inglês *Open Reading Frame*). O IMGT considera como ORF

genes que possuam alterações em seu sítio de *splicing*, mutações em motivos conservados entre as imunoglobulinas, deleção de um ou mais aminoácidos nas regiões de CDR ou FR e sequências que possuam tamanhos inesperados (>500 nucleotídeos).

1.3. Diversidade de repertório de anticorpos

Existem dois mecanismos para gerar diversidade em imunoglobulinas que são, o mecanismo de diversidade primária e secundária.

1.3.1. Mecanismos de diversificação primária

Os mecanismos de diversidade primária são aqueles que ocorrem na medula óssea. O primeiro deles é a diversidade alélica, uma vez que um subconjunto de segmentos gênicos apresenta múltiplos alelos, isso foi demonstrado pelo sequenciamento genômico, que indica que os loci IG, especificamente IGH, podem estar entre os mais polimórficos do genoma humano (Watson et al., 2015, 2017). Outro mecanismo que produz diversidade do repertório de anticorpos é o rearranjo gênico, ou a recombinação gênica de V (D) J que ocorre no DNA da linhagem germinal e reúne os segmentos descontínuos dos genes V, (D) e J em uma região variável durante o desenvolvimento das células B (Bassing et al., 2002; Gellert, 2002; Schlissel, 1998). Esses segmentos gênicos são flanqueados por sequências de sinais de recombinação (RSSs), que compreendem dois motivos altamente preservados: um heptâmero palindrômico e um nonâmero rico em “A” ou “T”, separados por um espaçador não conservado de 12 ± 1 ou 23 ± 1 nucleotídeos. Eles são reconhecidos por enzimas recombinases (RAG1 e RAG2 do inglês *Recombination-Activating Genes 1 and 2*) (Schatz et al., 1989). Rearranjos eficientes ocorrem entre RSSs de diferentes comprimentos, ou seja, um com espaçador de 12 ± 1 e outro com espaçador de 23 ± 1 (regra de junção 12/23) (Early et al., 1980). O repertório potencial resultante da diversidade combinatória depende do número de segmentos gênicos IGHV, IGHD e IGHJ da espécie (Lefranc & Lefranc, 2020) (Figura 3).

O terceiro mecanismo primário de diversidade é a diversidade de junção, a qual é representada pela diversidade observada essencialmente nas junções V(D)J de cadeias pesadas de IG, (e, em menor grau, a junção VJ da cadeia leve), e representando a principal fonte de diversidade CDR-H3. Esse tipo de diversidade é resultado da clivagem de nucleotídeos por uma exonuclease nas extremidades dos segmentos gênicos IGHV, IGHD e IGHJ durante o rearranjo

(extremidade 3' da região-V, extremidade 5' da região-J e/ou ambas as extremidades da região-D), seguido pela adição aleatória de nucleotídeos (N) via DNA nucleotidil exotransferase (DNNT, terminal deoxynucleotidyl transferase TdT) (Landau et al., 1984). Esta adição de nucleotídeos envolve preferencialmente nucleotídeos 'G' e é usualmente independente do molde. Além da adição de N nucleotídeos, também pode ocorrer a adição de nucleotídeos Palindrômicos (P). (Lafaille et al., 1989). Os P nucleotídeos são sequências curtas de 1 a 3 nucleotídeos além da extremidade do segmento gênico codificado. Os nucleotídeos P resultam de abertura assimétrica do grampo formado nas extremidades das regiões de codificação durante os rearranjos VJ ou V(D)J (Lewis, 1994) (Figura 3).

TdT pertence à família pol X de polimerases (Ito & Braithwaite, 1991), juntamente com pol μ , pol β e pol λ , um subgrupo de uma família de nucleotidiltransferases que é definida por semelhanças nos domínios de ligação de nucleotídeos e motivos de sítios ativos (Holm & Sander, 1995). Esta enzima é caracterizada por sua ação de polimerase independente de molde, mas há alguns anos foi demonstrado que a TdT pode realizar a incorporação de nucleotídeo dependente de molde na presença de DNA em excesso (Loc'h & Delarue, 2018).

O núcleo catalítico de TdT adota uma conformação fechada, mediada por uma interação da porção N-terminal da proteína com o C-terminal. O sítio ativo de TdT contém três resíduos de aspartato conservados (D343, D345 e D434), consistentes com um estudo prévio de mutagênese (Yangs et al., 1994), e se assemelha ao de pol β , sugerindo um mecanismo dependente de íon metálico bivalente idêntico para catálise (Steitz & Steitz, 1993).

A homologia entre os membros da família pol X não se limita ao seu centro catalítico. Domínios adicionais compartilhados por várias dessas polimerases incluem os domínios de proteína de suscetibilidade ao câncer de mama 1 (BRCA-1 do inglês *breast cancer susceptibility protein*) C-terminal, bem como os domínios hélice-gancho-hélice (HhH). Um desses domínios BRCT também foi identificado nas regiões N-terminais de TdT, pol μ e pol λ (Aoufouchi et al., 2000). Os domínios HhH são motivos de ligação ao DNA não específicos, contato com o DNA por meio de interações dos átomos de nitrogênio do esqueleto do peptídeo com grupos fosfato de DNA (Doherty et al., 1996).

Em algumas polimerases da família pol X existe um inter-domínio não enzimático entre BRCA e HhH, no caso da pol λ este inter-domínio é rico em prolina (contém múltiplos resíduos de serina, treonina e prolina) cuja função é suprimir funcionalmente a atividade da polimerase de pol λ e limita a síntese da fita de DNA (Fiala et al., 2006).

Foi relatada uma estrutura cristalográfica de TdT que confirma muitas das previsões apresentadas anteriormente (Delarue et al., 2002). A porção TdT cristalizada (aa 130-510) exclui o sinal de localização nuclear putativo (NLS do inglês *Nuclear Localization Signal*) e o domínio BRCT, mas retém a atividade catalítica (Sukumar et al., 2000). A estrutura da TdT assemelha-se muito à de outras DNA polimerases (Delarue et al., 2002), com uma forma semelhante à da mão direita, consistindo de um domínio N-terminal de dedo (finger), uma palma central contendo o sítio ativo (palm) e um domínio polegar (thumb). TdT exhibe semelhança estrutural particularmente alta com pol μ (Loch & Delarue, 2018).



Figura 3. Eventos moleculares de um rearranjo entre um segmento gênico IGHD e um segmento gênico IGHJ no locus IGH que geram a junção de diversidade N. (A) O complexo proteico enzimático RAG1/RAG2 reconhece os sítios de sinal de recombinação (RSS) e corta o DNA da linhagem germinativa entre o heptâmero e a região de codificação. (B) Um grampo é formado em cada extremidade de codificação, aqui as extremidades 3'D e 5'J, por transesterificação. (C) O Grampo é clivado pelo complexo proteico (XRCC6, PRKDC, XRCC5, Ku80, DNA-PKcs, DNA-PK, e DCLRE1C, Artemis). (D) Uma exonuclease remove nucleotídeos nas extremidades codificantes. (E) DNA nucleotidilexotransferase (DNTT, TdT, desoxinucleotidil transferase terminal) adiciona N nucleotídeos, sem molde. (F) Finalmente, o

DNA rearranjado é reparado e as extremidades são ligadas pelo complexo ligase. Fonte (Lefranc & Lefranc, 2020).

1.3.2. Mecanismos de diversificação secundária

As regiões variáveis recombinadas nas células B do centro germinativo sofrem hipermutações somáticas (SHM do inglês *Somatic hypermutation*) em formas de substituições, inserções e deleções de nucleotídeos (Peled et al., 2008). Embora toda a região variável seja alvo de SHM, as regiões CDRs são mais propensas a sofrer mutações do que as regiões estruturais (FRs) (Cowell et al., 1999).

De acordo com o modelo atual, o processo de SHM é iniciado pela enzima citidina desaminase induzida pela ativação (AID do inglês *Activation Induced cytidine Deaminase*), que desamina as citosinas em uracilas no DNA de fita simples formado durante a transcrição dos genes imunoglobulina (IG) (Muramatsu et al., 2000). Em seguida, a fita de DNA onde estão as uracilas geradas poderá ser replicada, resultando nas transições de C para U (transições de G para A na cadeia oposta) ou ser reparados pela reparação por excisão de base (BER, do inglês *Base-excision repair*), que pode ser de um ou mais nucleotídeos, ou reparo de *Mismatch* (MMR, do inglês *Mismatch repair*). As duas últimas vias de reparo são propensas a erros, levando a um grande número de mutações nas bases “A” e “T” (Peled et al., 2008). (Figura 4A).

Outro processo que garante a diversidade dos anticorpos é a conversão gênica. Ainda que considerados como não-funcionais, graças a este processo, os pseudogenes podem sofrer recombinação e gerar anticorpos funcionais (Bolland et al., 2016) e isso pode explicar porque algumas espécies, como camundongos, gados, aves e cavalos, possuem um grande número de pseudogenes em seu DNA.

A conversão gênica ocorre quando parte das sequências do segmento IGHV recombinado são substituídas por sequências de um pseudogene IGHV próximo à esta região no mesmo cromossomo sem troca recíproca de genes, ou seja, a sequência do pseudogene permanece inalterada (Flajnik, 2002) (Figura 4B).

Após a recombinação dos segmentos e a inserção de mutações na região variável por SHM ou conversão gênica, toda a progênie dessa célula B irá expressar os mesmos segmentos gênicos V(D)J. Contudo, vários isotipos diferentes da região constata podem ser expressos no processo de maturação e proliferação dos linfócitos B durante uma resposta imune. O último evento de diversificação secundária do repertório é a recombinação por troca de classe (CSR,

do inglês *Class-switch Recombination*) permitindo a troca da região efetora do anticorpo (região C) (Murphy, 2014) (Figura 4C).

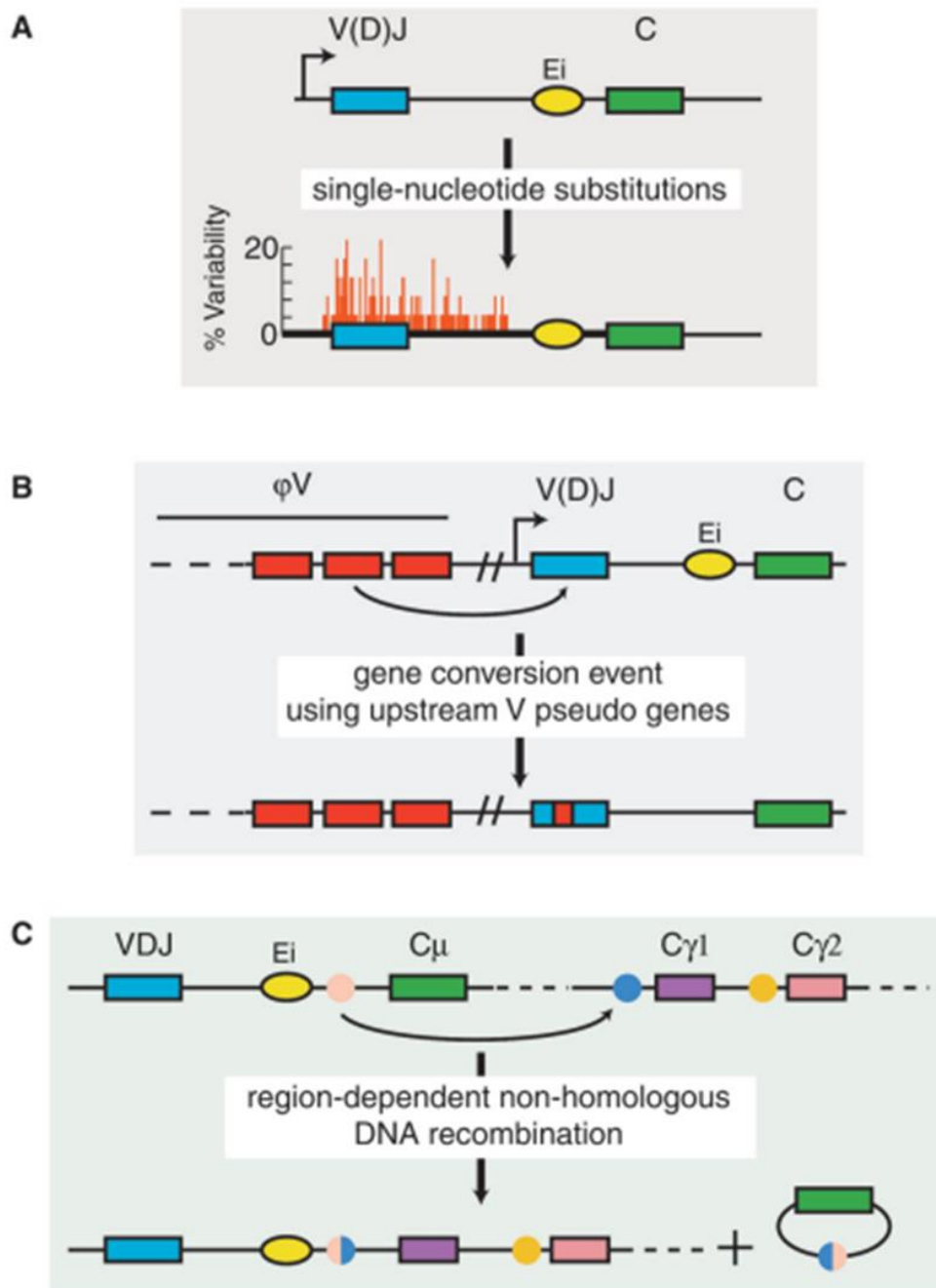


Figura 4. Diagrama dos processos que ocorrem no mecanismo de diversidade secundária de anticorpos. (A) Hipermutação somática: V(D)J recombinação está sujeito a mutações pontuais cujas posições e frequências são indicadas por linhas vermelhas. (B) Conversão gênica: A sequência de DNA de V(D)J recombinação é alterado por meio de um mecanismo de conversão gênica, no qual os segmentos de pseudogene doam trechos curtos de sua sequência. (C) Recombinação por troca de classe: A região de troca de IgM de um dos loci de Ig expressos recombina-se com uma região de troca através de uma via de recombinação de DNA não

homóloga, resultando em um loop e a região constante IgM e eliminada. Fonte (Petersen-Mahrt, 2005).

Todos esses fatores contribuem para a grande diversidade do repertório de anticorpos, estimada em 10^{16} - 10^{18} , em um repertório sem estimulação antigênica (Briney et al., 2019). O estudo das proteínas receptoras de antígenos localizadas nas membranas das células T e B é chamado de repertório. O sequenciamento do repertório (Rep-seq) revolucionou os estudos de diversidade de receptores de antígenos, permitindo uma análise profunda e quantitativa para decifrar a função da imunidade adaptativa na saúde e na doença (Georgiou et al., 2014). Os repertórios imunológicos são um alvo importante da pesquisa imunológica e clínica, pois contêm informações sobre as respostas imunes passadas e em andamento (Murphy, 2014).

1.4. Estudo do repertório de anticorpos em diferentes espécies

O estudo de repertórios de anticorpos em mamíferos é de grande importância não só para a terapêutica veterinária, mas também devido à semelhança do sistema imunológico de alguns animais com humanos. O conhecimento a nível molecular da resposta imune adaptativa de algumas espécies de animais pode ajudar a entender os processos realizados pelo sistema imunológico tanto na saúde quanto nas doenças em humanos. Um resumo das características do repertório de anticorpos das espécies de mamíferos mais estudadas é apresentado na Tabela 1. É importante destacar que existe um padrão canônico de mamíferos de isotipos de cadeia pesada. Todos os mamíferos possuem genes que codificam IgM, IgD, IgG, IgE e IgA (Butler J., 1998).

1.4.1 Repertório de anticorpos em camundongos

O banco de dados IMGT lista 107 segmentos gênicos IGHV funcionais em camundongos, com base em uma análise da sequência de referência do genoma do camundongo e a associação desses genes com sequências de sinal de recombinação intactas (RSS) (Johnston et al., 2006). IMGT também reconhece 31 segmentos gênicos IGHD, incluindo 25 genes funcionais, 1 pseudogene e 5 ORF (<http://www.imgt.org>)

Quatro segmentos gênicos IGHJ funcionais são encontrados em camundongos (<http://www.imgt.org>). Os segmentos gênicos IGHJ1, IGHJ2, IGHJ3 e IGHJ4 de camundongo são ortólogos dos segmentos gênicos humanos IGHJ2, IGHJ4, IGHJ5 e IGHJ6. O

sequenciamento genômico de camundongos inferiu a existência de 93 segmentos gênicos funcionais de IGKV. Estes incluem dez sequências definidas pelo IMGT como ORFs. A estrutura do locus *lambda* de camundongo é diferente de todos os outros loci de camundongos e humanos. Ele contém apenas dois pequenos grupos de genes, cada um contendo IGLV, IGLJ e segmentos gênicos de região constante (<http://www.imgt.org>).

Em relação ao tamanho do CDR-H3, os anticorpos de camundongo possuem entre um e 23 aminoácidos de comprimento, com média de 11 aminoácidos (Rettig et al., 2018).

Outro parâmetro analisado no repertório de anticorpos do camundongo é a hipermutação somática, no que se observou que cerca de 2,4 a 4,6% dos anticorpos sofre SMH, dependendo da espécie.

Estudos também investigaram a diversidade juncional dos anticorpos de camundongo. Até o momento sabe-se que a adição de nucleotídeos P é comum em rearranjos de anticorpos de camundongos, embora as adições de mais de quatro nucleotídeos P sejam muito raras e a maioria das adições tenham apenas um ou dois nucleotídeos de comprimento (Collins et al., 2015), com média de $0,5 \pm 0,7$ nucleotídeos (Shi et al., 2014). A adição de N nucleotídeo não codificado por molde é um importante fator de diversidade de repertório, em camundongos, foi visto, em média, 4,0 e 2,9 N nucleotídeos adicionados à junção V-D e à junção D-J, respectivamente. (Collins et al., 2015).

Os clonotipos de células B públicas são geralmente definidos como rearranjos V(D)J que são compartilhados por diferentes indivíduos, que usam segmentos gênicos IGHV e IGHJ idênticos na linhagem germinativa e que compartilham sequências de aminoácidos na região de CDR-H3 (Collins et al., 2015). Um estudo de sequenciamento de alto rendimento (HTS do inglês *High-Throughput Sequencing*) envolvendo amostras relativamente pequenas de repertórios de anticorpos de camundongos C57BL/6 e BALB/c, descobriu que entre 4% e 6% das sequências possuíam CDR-H3 compartilhados por dois ou mais camundongos da mesma linhagem (Collins et al., 2015).

1.4.2. Repertório de anticorpos em cães

O estudo do repertório de anticorpos em cães tornou-se importante devido ao aumento da demanda por anticorpos monoclonais terapêuticos para uso veterinário no tratamento de doenças. O cão tem 89 segmentos gênicos IGHV em 3 subgrupos IGHV, dos quais 36 são funcionais, 2 são ORF e 51 pseudogenes, 6 segmentos gênicos IGHD e 6 segmentos gênicos

IGHJ, dos quais 5 são funcionais e 1 ORF (Martin et al., 2015). Assim, a diversidade combinatória dos anticorpos de cães tende a ser menor do que a de humanos e camundongos.(Steiniger et al., 2014). Nesta espécie o IMGT relata 4 subtipos de IgG. Os anticorpos de cães possuem CDR-H3 de dois a 25 aminoácidos., com a média de $13,6 \pm 3,74$ aminoácidos (Hwang et al., 2018).

Também no cão, a diversificação de imunoglobulinas pode ser aumentada pela inserção ou remoção de nucleotídeos em pontos de junção durante a recombinação. No caso de inserções de N nucleotídeos entre as junções V-D, bem como entre as D-J foi observada uma variação de um a 23 nucleotídeos de comprimento, evidenciando uma forte correlação entre o comprimento de CDR-H3 e o número de N nucleotídeos. A maioria das adições da região P representa apenas um ou dois nucleotídeos. Deleções de nucleotídeos em segmentos IGHD foram encontradas nas extremidades 5' ou 3' de cães possuem CDR-H3 de dois a 25 aminoácidos., com a média de $13,6 \pm 3,74$ aminoácidos (Hwang et al., 2018).

Hwang et al. em 2018 analisaram o repertório público em cães e encontraram 0,03% (159 sequências) de clones compartilhados estudando 3 indivíduos, que possuíam como características sequências de CDR-H3 de dois a 22 aminoácidos (média de $10,5 \pm 3,04$ aminoácidos) (Hwang, 2018).

1.4.3. Repertório de anticorpos em bovinos

O atual locus genético de IgH da linha germinativa bovina de referência compreende 12 segmentos gênicos funcionais IGHV e 35 pseudogenes, 20 IGHD mais três ORF, e 12 IGHJ dos quais três são funcionais, sete ORF e dois pseudogenes (Liljavirta et al., 2014).

As sequências de CDR-H3 de anticorpos bovinos têm um comprimento médio de 22 resíduos de aminoácidos (Deiss et al., 2019; Ma et al., 2016), sendo que um subconjunto destes possuem regiões de CDR-H3 ultra-longas que podem atingir mais de 70 resíduos de aminoácidos de comprimento (Saini et al., 1999). Estes são derivados da recombinação dos segmentos gênicos IGHV1-7 e IGHD8-2 (não adição de nucleotídeos na junção) (Deiss et al., 2019; Koti et al., 2010).

Estudos identificaram a presença de um grande número de resíduos de cisteína dentro de domínios CDR-H3 ultralongos. Esses resíduos de cisteína geram minidomínios por meio de ligações dissulfeto que se projetam do anticorpo na forma de grandes *loops*, conforme revelado

por dados de cristalografia, o que pode facilitar estruturalmente o acesso a epítomos que de outra forma não estariam disponíveis (F. Wang et al., 2013).

Em relação à diversidade juncional, o número médio de N nucleotídeos foi de 1 nas junções VD e 2 nas junções DJ. No total, 90% das sequências de anticorpos continham N nucleotídeos. Também foram observados nucleotídeos P, com variação de 0 a 6 na junção VD e 0 a 3 na junção DJ.

A frequência de SHM de anticorpos bovinos no segmento gênico IGHV do anticorpo é em média de 5,22% (Deiss et al., 2019). Ainda, a carga mutacional é de 5 a 9 mutações por quilobase em amostras fetais da placa de Peyer ileal e do baço, 26 por quilobase, na placa de Peyer ileal do bezerro de 51 dias e 80 por quilobase no linfonodo do bezerro de 2 anos. A razão de mutações CDR para FR é de 5,6 em fetos (média de todas as amostras de íleo e baço), 6,7 no bezerro de 51 dias e 3,4 no bezerro. A razão de substituição sobre mutações não sinônimas foi de 3,3 nos fetos, 3,1 no bezerro de 51 dias e 2,1 no bezerro imunizado.(Liljavirta et al., 2014).

1.4.4. Repertório de anticorpos em Macacos

Devido à sua semelhança genômica com os humanos, os macacos rhesus têm sido amplamente utilizados como modelo animal para vários estudos imunológicos e clínicos (Fu et al., 2017).

Atualmente, 228 genes da linhagem germinativa de IGHV para macacos rhesus estão listados no banco de dados do IMGT (<http://www.imgt.org>), dos quais 86 são funcionais, 2 são ORF e 140 são pseudogenes, com relação à os segmento gênicos IGHD, 45 genes são relatados, dos quais 44 são funcionais e 1 ORF e o segmento gênico IGHJ possui 7 genes, todos funcionais (<http://www.imgt.org>). A região CDR-H3 em anticorpos de macaco rhesus tem um comprimento entre 10 e 20 aminoácidos com uma média de 15 aminoácidos (Robins, 2013) .

Em relação à diversidade juncional, determinou-se as frequências dos quatro tipos de nucleotídeos inserido nos locais de junção V-(D)-J e são tendenciosos para “C” e “G”, no repertório IGH (aproximadamente 27% C e 39% G) (Fu et al., 2017).

Nesta espécie, a SHM também é de grande importância na geração de diversidade. Pesquisas determinaram que o percentual de SHM em anticorpos de macacos rhesus não imunizados é de 6,43%, encontrando uma diminuição significativa na SHM do isotipo IgM quando comparado ao isotipo IgG (Guo et al., 2015).

1.4.5. Repertório de anticorpos em Porcos

De acordo com o IMGT, os Porcos possuem 21 segmentos gênicos IGHV, dos quais 13 são funcionais e 8 pseudogenes. No caso dos segmentos gênicos IGHD e IGHJ todos são funcionais, sendo 4 genes para o IGHD e 5 para o IGHJ. (<http://www.imgt.org>).

O comprimento do CDR-H3 de porco varia de 6 a 28 aminoácidos, com média de $16,7 \pm 4$ aminoácidos (ALMAGRO et al., 2006).

As pesquisas geralmente analisam a composição de aminoácidos do CDR-H3, principalmente no conteúdo de resíduos de cisteína e glicina, uma vez que esses dois aminoácidos são críticos para a estrutura geral da alça. No estudo realizado por Almagro et al em 2006, foi determinado que as alças CDR-H3 de porco têm 43 resíduos de cisteína de um total de 1000 aminoácidos analisados (4,3%). Em relação à proporção de glicina foi de 12,0%. (ALMAGRO et al., 2006).

No que diz respeito à diversidade juncional, sabe-se até o momento que as adições de nucleotídeos são encontradas em anticorpos de porcos de todas as idades, sendo encontradas em uma faixa de 0 a 30 nucleotídeos e com uma média de 8,6 adições de nucleotídeos em todas as regiões de junção. Raramente foram observadas adições de nucleotídeos 5' ou 3' de IGHD (Šinkora et al., 2003).

Notavelmente, no repertório pré-imune de porco, SHM permanece baixo e está concentrado na região CDR de rearranjos de cadeia pesada. (Butler et al., 2009, 2011; Wertz et al., 2013)

Tabela 1. Resumo das principais características do repertório de anticorpos em mamíferos

Espécies	Número IGHV Funcional	Número IGHD Funcional	Número IGHJ Funcional	Tamanho (aminoácidos)	CDR-H3	Tamanho Junção (nucleotídeos)	SMH (%)
Camundongos	113	25	4	11		P= 0,5 ± 0,7 N= 4,0 e 2,9	2,4 a 4,6
Cães	36	5	5	13,6 ± 3,74		P= 1 a 2 N=entre 0 e 23	-
Bovinos	12	20	3	22		P= 0 a 6 N=1 a 2	5,22
Macacos	86	44	7	15		-	6,43
Porcos	13	4	5	16,7 ± 4		N= 8,6	-

P- Palindrômico

N- Aleatório

1.5. Locus de imunoglobulinas de cavalos

Anticorpos policlonais de equinos (*Equus caballus*) foram produzidos e utilizados desde o século XIX para tratamento e profilaxia de diferentes doenças, (ANDERSON, 1955; Cole & Moore, 1917; Glatman-Freedman & Casadevall, 1998; Lang et al., 2000) e ainda é usado nos dias atuais. Apesar da facilidade de produção e da utilidade médica, o conhecimento de como o sistema imunológico humoral equino gera as imunoglobulinas é pequeno, especialmente a nível molecular.

Similar a outros vertebrados, o cavalo possui três tipos de cadeias de imunoglobulinas: leve *lambda* (IGL), leve *kappa* (IGK) e pesada (IGH). Na anotação dos segmentos gênicos de cavalo feito por SUN et al. (2010) e revisto por Walther et al. (2015), que usou como referência o Genoma do *Equus Caballus* 2.0, a cadeia pesada IGH, presente no cromossomo 24, pode ser formada por 11 diferentes regiões constantes do anticorpo (IGHM, IGHD, IGHA, IGHE e sete subclasses IGHG) (Wagner, 2006). Nesta anotação o locus IGH de cavalo tem 56 segmentos gênicos IGHV (16 funcionais, 34 pseudogenes e seis ORF) 40 segmentos gênicos IGHD e oito segmentos gênicos IGHJ, todos funcionais (Sun et al., 2010; Walther et al., 2015). Todos os segmentos desse locus apresentam orientação *sense* nas *scaffolds* analisados.

O cavalo é uma das espécies que mais possuem segmentos gênicos IGHD anotados no genoma (Walther et al., 2015), aumentando a diversidade do locus IGH. Geralmente para todas as espécies os segmentos gênicos IGHV funcionais podem ser classificados em subgrupos que compartilham ao menos 75% de identidade da sequência de nucleotídeos entre seus membros (Eronique Giudicelli & Lefranc, 1999), no cavalo existem 7 subgrupos. Utilizando essa mesma abordagem, os segmentos gênicos IGHD foram divididos em 28 subgrupos e os segmentos gênicos IGHJ em dois subgrupos (Walther et al., 2015).

O locus IGK, presente no cromossomo 15 (NW_001867379), contém 60 segmentos gênicos (19 funcionais, 39 pseudogenes e dois ORFs) divididos em sete subgrupos, cinco segmentos gênicos IGKJ que compõem cinco subgrupos (quatro funcionais e um pseudogene) e um gene IGKC. Nesse locus, parte dos segmentos IGKV se encontram na orientação *sense* e outra parte na orientação *anti-sense*.

O locus IGL, que está no cromossomo 8 (NW_001867428), apresenta 144 segmentos gênicos (27 funcionais, 112 pseudogenes e cinco ORFs) agrupados em 11 subgrupos, sete segmentos gênicos IGLJ (quatro funcionais e três pseudogenes) classificados em sete subgrupos diferentes, e sete genes IGLC (um pseudogene), cada um associado a um segmento gênico J

distinto (Sun et al., 2010; Walther et al., 2015). Similar ao locus IGK, parte dos segmentos IGLV se encontram na orientação *sense* e outra parte na orientação *anti-sense*.

Nos equinos há uma predominância de expressão da cadeia leve *lambda*, compondo cerca de 95% das imunoglobulinas circulantes, semelhante a ovelhas, gado, cães e gatos (Arun et al., 1996; Foley, 1992), em contraste com humanos que têm uma predominância do isotipo da cadeia leve *kappa* em seus anticorpos (Weiss et al., 1984).

Em 2018, foi sequenciado um novo genoma de cavalo (*Equus caballus* 3.0) e com isso nosso grupo de pesquisa junto com a IMGT está trabalhando em uma nova anotação.

Até agora, o IGH foi anotado, e o IMGT mostra que o segmento gênico IGHV possui 8 subgrupos compostos por 21 genes funcionais, 9 ORFs e 74 pseudogenes, no caso do segmento gênico IGHD possui 6 subgrupos, compostos por 17 são funcionais e 28 ORF, e em relação ao segmento gênico IGHJ, possui 6 subgrupos compostos por 6 são funcionais e 3 ORF (Figura 5).

Até agora, não foi possível fazer a anotação da cadeia leve *lambda*. Será necessário um novo sequenciamento do genoma do locus da imunoglobulina, trabalho que está em andamento.

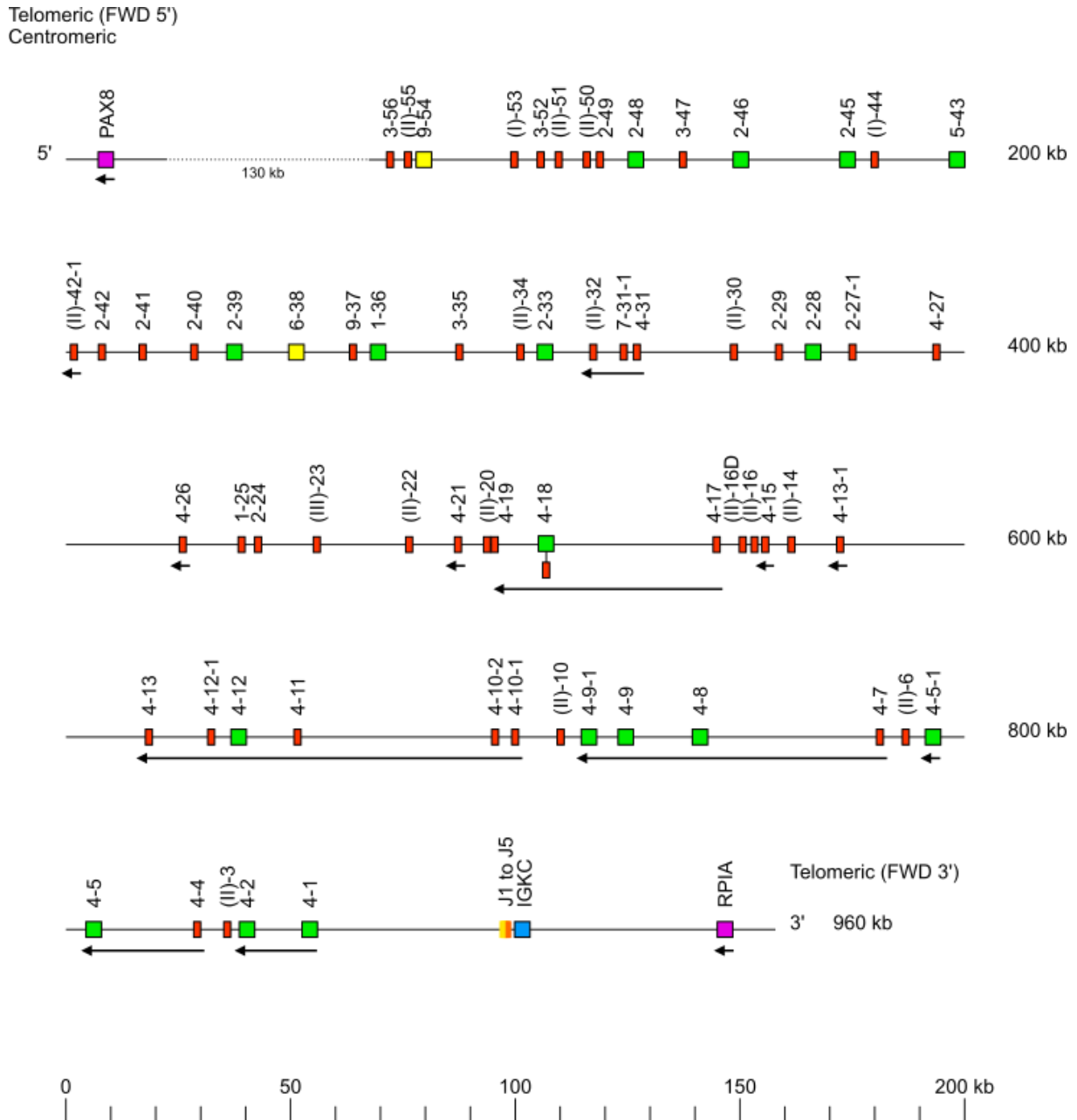


Figura 6. Representação esquemática do locus da cadeia leve *Kappa* da imunoglobulina de *Equus caballus*. Locus anotado no genoma de cavalo 3.0. As caixas verdes representam os genes funcionais do IGK, as caixas amarelas os ORFs do IGK, as caixas vermelhas os pseudogenes do IGK. Os amarelos os funcionais do IGKJ, o quadrado laranja o pseudogene do o IGKJ. O azul claro corresponde ao IGKC, e quadrado roxo correspondem a genes funcionais não relacionados. Fonte: (<http://www.imgt.org>).

Nota-se nesses animais uma grande quantidade de pseudogenes (o dobro, quando comparados aos genes funcionais), o que nos leva a pensar que a conversão gênica deve ter

papel importante na formação da diversidade do seu repertório de imunoglobulinas (Walther et al., 2015).

1.6. Repertório de anticorpos para cavalos

Até o momento, análises do repertório de anticorpos equinos foram realizadas por diferentes metodologias de amplificação, porém, com sequenciamento dos *amplicons* por Sanger, constituindo análises de baixa profundidade (ALMAGRO et al., 2006; Tallmadge et al., 2013, 2014).

Em 2019, nosso grupo de pesquisa foi o primeiro a publicar um estudo do repertório de anticorpos de cavalos usando a tecnologia de nova geração (NGS) com a plataforma Illumina, com um aumento notável na profundidade do repertório. Foram observados pela primeira vez em estudo com cavalos, cerca de 45.000 clones de IGH e 30.000 clones de IGL, assim como novos transcritos gênicos (IGHV6S1 e IGLV4S2), a composição dos aminoácidos por posição de CDR-H3 (Manso et al., 2019).

Neste estudo, observou-se que o comprimento de aminoácidos da CDR-H3 em equinos varia de 4 a 26 resíduos de aminoácidos, seguindo um modelo de distribuição bimodal com um pico claro em cinco e seis resíduos e um pico mais extenso variando entre 10 e 26 aminoácidos. O comprimento médio foi de 14,5 resíduos, e os cinco aminoácidos mais utilizados no CDR-H3 foram Gly, Tyr, Ala, Asp e Ser. (Manso et al., 2019).

Além do tipo de sequenciamento, o outro fator importante no estudo do repertório de anticorpos é a análise bioinformática. A grande quantidade de dados de sequências de DNA geradas pelas tecnologias NGS aumentou a necessidade de programas computacionais para a análise massiva desses dados, e com o Rep-Seq não foi diferente.

1.7. Análises das características do repertório de anticorpos

Os dados gerados em estudos de Rep-Seq requerem pipelines de bioinformática especializados para análises eficazes e robustas dessas moléculas com estrutura gênica tão peculiar (Yaari & Kleinstein, 2015).

Diversos programas foram sugeridos, contudo não há um consenso de um pipeline padrão para o processamento e análise de dados de Rep-Seq. A falta de padronização e compartilhamento dos programas disponíveis dificulta a análise e comparação de dados publicados por diferentes laboratórios (Georgiou et al., 2014).

Pacotes de ferramentas foram desenvolvidos para facilitar essa análise. Esse é o caso do MiXCR para análise de dados brutos e análise do repertório (Bolotin et al., 2015), do pRESTO para análise de dados brutos (vander Heiden et al., 2014), do Change-O (Gupta et al., 2015) e VDJtools (Shugay et al., 2015) para análise do repertório. Em geral, as principais análises dos dados de Rep-Seq serão descritas abaixo.

1.7.1. Pré-processamento das *reads*

Através do pré-processamento é possível transformar os dados brutos em sequências IG prontas para anotação e análise do repertório. Esta etapa inclui: (I) retirada de adaptadores de sequenciamento, (II) montagem das *reads*, (III) exclusão de bases de qualidade baixa, (IV) identificação de primers e (V) contagem das *reads* (Manso et al., 2019). Existe uma vasta gama de abordagens que podem ser consideradas como correção de erros computacionais.

O mais simples seria filtrar conjuntos de dados, antes de qualquer anotação V(D)J, para *reads* de baixa qualidade (por exemplo, pontuação Phred) usando pacotes de software, como pRESTO ou FastQC (Smith & de Sena Brandine, 2021; vander Heiden et al., 2014).

1.7.2. Anotação dos segmentos gênicos

A atribuição da linhagem germinativa IGHV, IGHD, e IGHJ é uma das etapas mais importantes no processamento de dados de IG. O objetivo desta etapa é inferir os alelos corretos da linhagem germinativa que se recombinaram para produzir cada anticorpo. Uma boa inferência da linhagem germinativa é crítica para identificar corretamente mutações somáticas

para cada sequência, definir os grupos clonais e ter uma aproximação justa da diversidade (Ralph & Matsen, 2016).

Geralmente, essa inferência é feita aplicando um algoritmo para escolher a melhor correspondência entre um conjunto de segmentos potenciais da linhagem germinativa a partir de um banco de dados de alelos de segmentos conhecidos (López-Santibáñez-Jácome et al., 2019). Diferentes programas são usados para anotação de segmentos gênicos, alguns deles são: IMGT/HighV-QUEST (Alamyar et al., 2012), IgBLAST (Ye et al., 2013) e MiGEC (Kuchenbecker et al., 2015) que usam o banco de dados do IMGT, MiXCR (Bolotin et al., 2015), LymAnalyzer (Yu et al., 2015), entre outros.

1.7.3. Análise de clonalidade das sequências

Após a anotação dos segmentos gênicos, as sequências precisam ser agrupadas para formar clusters, identificando as sequências que sofreram expansão clonal e SHM, mas tiveram origem de uma mesma célula B não mutada (Robinson, 2015; Yaari & Kleinstein, 2015). A clusterização das sequências em grupos clonais, ou clonotipos, pode fornecer informações sobre a evolução da resposta específica ao antígeno daquela linhagem (Wine et al., 2013).

Existem diferentes definições para agrupar os clonotipos. Uma estratégia amplamente utilizada, por exemplo, é considerar todas as sequências com a mesma identificação da linha germinativa IGHV e IGHJ e alguma semelhança nas regiões CDR-H3 como relacionadas clonalmente (Hershberg & Luning Prak, 2015). Entre os programas mais utilizados para esse tipo de análise estão, IMGT/StatClonotype (Aouinti et al., 2016), CD-HIT (Fu et al., 2012), MIXCR (Bolotin et al., 2015), Yclon (Gervásio et al., 2022) e USEARCH (Edgar, 2010).

1.7.4. Caracterização do Repertório de Anticorpos

Após a definição dos clonotipos, diversas características do repertório de anticorpos podem ser analisadas. Entre as principais características estão:

Análise de rarefação: É a maneira mais simples de descrever a diversidade de anticorpos, mas é muito sensível à profundidade da amostragem. A riqueza de espécies é o número total de espécies únicas em uma comunidade, não leva em consideração a abundância de espécies. Maior profundidade de amostragem resulta na captura de cada vez mais espécies raras, resultando em maior riqueza de espécies. A raridade pode ter um impacto significativo

nesta medida, uma vez que muitas vezes se perdem espécies sub-representadas. Para explicar o problema das espécies raras, várias medidas foram elaboradas para prever a riqueza real de espécies com base em dados amostrados, incluindo iNEXT (Hsieh et al., 2016) e Recon (Kaplinsky & Arnaout, 2016).

Frequências de uso do segmento V(D)J: Um repertório de IG pode ser descrito em termos das frequências com as quais ele usa os segmentos gênicos, particularmente o segmento IGHV. As frequências do segmento gênico IGHV, ou frequências das combinações IGHV-IGHJ, têm sido usadas para comparar estágios de respostas imunes, em doenças específicas, indivíduos com diferentes faixas etárias, entre outros (Chaudhary & Wesemann, 2018; Martin et al., 2015).

Propriedades da Região Determinante de Complementaridade 3 da cadeia pesada (CDR-H3): O comprimento do CDR-H3 e as propriedades dos aminoácidos são utilizados para caracterizar um repertório (Chaudhary & Wesemann, 2018). Comparações de comprimentos de CDR-H3 entre grupos de células que expressam diferentes isotipos de IgH mostram que a IgM tem CDR-H3s mais longos em comparação com todos os outros isotipos examinados (Kitaura et al., 2017). Os repertórios de células B que já foram expostos a algum tipo de antígeno parecem ter uma região CDR-H3 mais exposta, e com aminoácidos carregados (Mcdaniel et al., 2016).

A exposição ao antígeno também parece estar associada a uma diminuição no comprimento da CDR-H3 (Galson et al., 2015), entre outros estudos realizados na área (Kitaura et al., 2017). Compreender o surgimento de CDR-H3s pode ajudar a entender o processo de maturação por afinidade e desenvolvimento de anticorpos (Hoehn et al., 2016).

Análise de Mutação: Anticorpos específicos para antígenos são o resultado de eventos mutacionais que se acumulam durante a maturação das células B. Isso inclui inserções e deleções durante a recombinação V (D) J e SHM. Em geral, as mutações são analisadas como grau de divergência em relação às sequências da linhagem germinativa e fornecem informações sobre o processo biológico de SHM e a maturação por afinidade (Kitaura et al., 2017). O objetivo da análise de mutações é ter uma ideia do processo de maturação que as células B sofreram durante o curso de uma resposta imune e seu encontro com antígenos (de Wildt et al., 1999; Yaari & Kleinstein, 2015).

Análise da Diversidade Juncional: A diversidade juncional é formada pela adição e/ou deleção de nucleotídeos nas junções dos genes V(D)J o que confere ainda mais diversidade ao repertório. Em algumas sequências de anticorpos, a hiperatividade da exonuclease pode

remover completamente os nucleotídeos da sequência IGHD original ou movê-los para uma fase de leitura diferente, como mostrado por Darsley & Rees em 1985, gerando sequências inteiramente novas na região central do CDR-H3. Berek & Milstein estimaram que o fator de diversidade devido à diversidade das junções N e P nas cadeias pesadas é 10 vezes mais do que se não tivesse esse tipo de diversidade. (Darsley & Rees, 1985)

Análise de clones públicos: A análise de convergência refere-se a um fenômeno que ocorre quando sequências de receptores imunes idênticos (ou altamente semelhantes) são compartilhadas por dois ou mais indivíduos (H Stern et al., n.d.; Jackson et al., 2014). Por exemplo, foi demonstrado que a similaridade de segmentos de genes em repertórios produtivos de IgH de irmãos gêmeos é maior em genes de Ig não mutados do que mutados (B. Wang et al., 2015).

Análise do isotipo: A análise do repertório de imunoglobulinas pode fornecer informações sobre a biologia dos isotipos de IgH. Cada isotipo tem funções biológicas distintas governadas pelo domínio da região C. As sequências em um repertório podem ser categorizadas em seus respectivos isotipos se o desenho experimental acomodar a sequência da região C na biblioteca. A análise isotípica incluiu a categorização de características, funções ou condições do repertório de IG para grupos isotípicos, entre outras funcionalidades (Kitaura et al., 2017).

Expansão clonal e diversidade: A quantificação da expansão clonal e da diversidade de repertórios representa um objetivo importante nas análises do repertório, porque fornece informações sobre o status imunológico atual de um hospedeiro (Attaf et al., 2015), o que é particularmente importante para o perfil de doenças e vacinas. Os fundamentos matemáticos da avaliação da diversidade biológica foram desenvolvidos décadas atrás para pesquisa ecológica (MAGURRAN, 1983). Já existem vários pacotes no R *study* dedicados para cálculos de o índice de diversidade (Cortina-Ceballos et al., 2015; Dixon, 2003; Gupta et al., 2015).

A Figura 7 mostra um exemplo de *pipeline* para estudar o repertório de anticorpos.

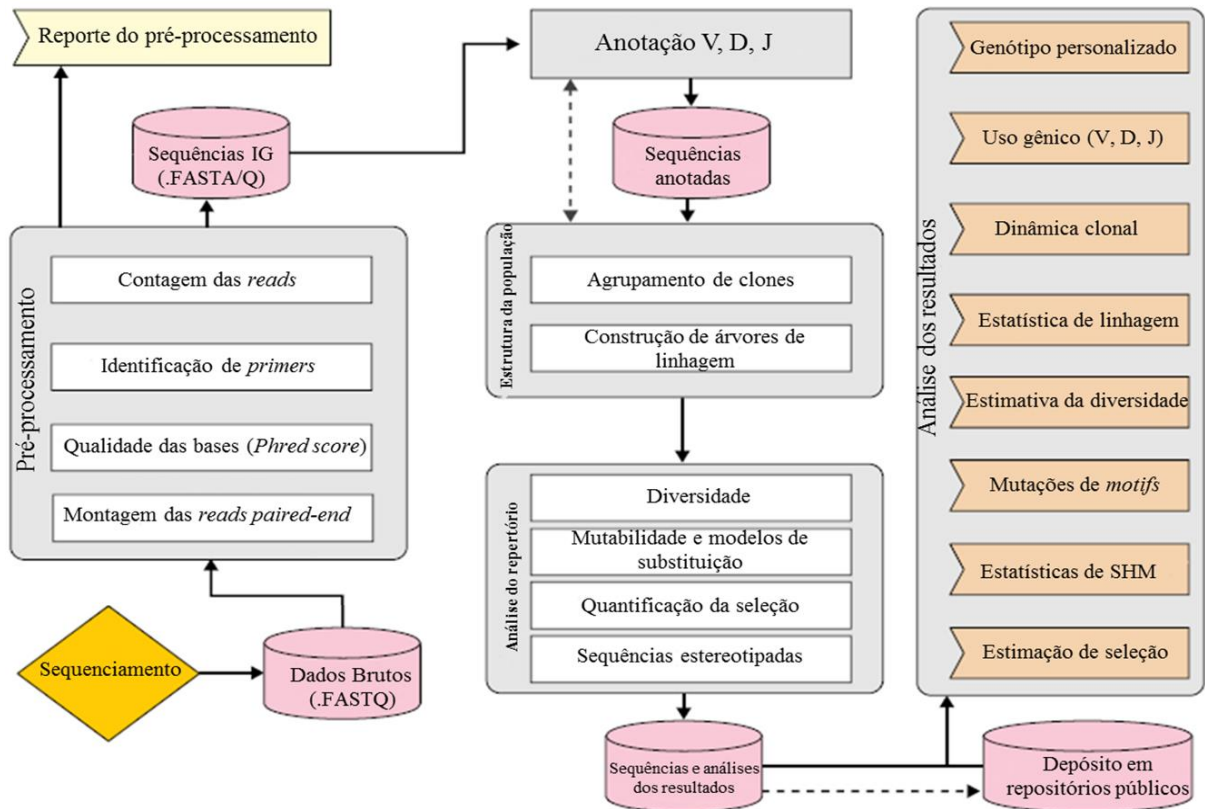


Figura 7. Exemplo de um pipeline para estudar o repertório de anticorpos de células B. A análise de sequenciamento de repertório (Rep-seq) pode ser dividida em pré-processamento; anotação e clonotipagem; e análise detalhada do repertório. O pré-processamento transforma leituras de sequenciamento de nova geração em sequências de BCR corrigidas que são então alinhadas para identificar genes V(D)J da linhagem germinativa. A seguir, infere-se a estrutura populacional dinâmica do repertório BCR. Finalmente, são calculadas as características quantitativas do repertório de células B como hipermutação somática (SHM). Fonte: (Yaari & Kleinstein, 2015).

1.8. Soroterapia

A análise do repertório de cavalos é muito importante não só para compreender melhor a função da imunidade adaptativa desta espécie, mas também a nível biotecnológico.

O soro de cavalo está inclusive sendo usado na atual pandemia de COVID-19 como tratamento em alguns países (Cunha et al., 2020; Zylberman et al., 2020). Alguns grupos de pesquisa utilizam o vírus atenuado para imunizar cavalos. O vírus é isolado de um paciente com a doença, cultivado em células específicas e, em seguida, inativado. O curso de imunização consiste em imunização primária com partículas de vírus inativadas misturadas com adjuvante, enquanto as imunizações de reforço subsequentes são administradas nos dias 15, 29, 35 e 45.

Os anticorpos gerados alcançam um título de anticorpo específico de antígeno satisfatório (Cunha et al., 2020).

Zylberman e colaboradores em 2020 também criaram um soro imune em cavalos contra o SARS-CoV-2, mas em vez de usar vírus inativado usaram o domínio de ligação ao receptor da glicoproteína viral Spike para a imunização, obtendo resultados bastante favoráveis usando este soro como terapia em pacientes hospitalizados (Zylberman et al., 2020) (Figura 8).

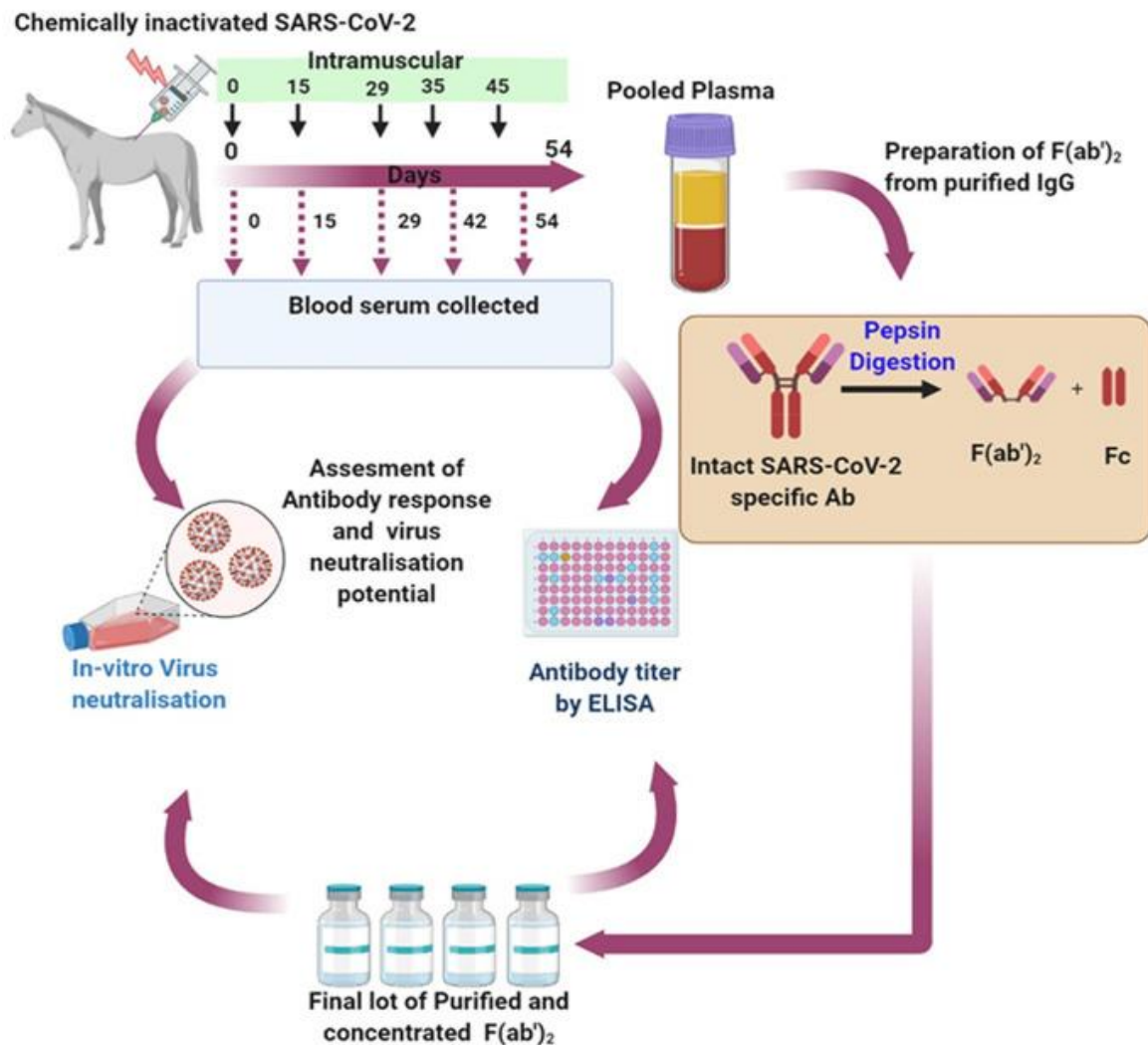


Figura 8. Esquema de imunização de cavalo com vírus inativado SARS-CoV-2. Partículas inativadas de SARS-CoV-2 foram injetadas por via intramuscular em equinos. A imunização foi repetida vários dias como mostrado no esquema. O plasma coletado de animais imunizados foi combinado, testado quanto à resposta de anticorpos e o título de neutralização do vírus quantificado por ensaio de microneutralização. Subsequentemente, a IgG foi purificada a partir do plasma combinado, digerida com pepsina e o fragmento F(ab')₂ purificado. Os títulos de neutralização destes fragmentos purificados e concentrados também foram testados. Fonte: (Cunha et al., 2020).

No Brasil, diferentes grupos de pesquisa também estão realizando soro imune contra SARS-CoV-2. O Instituto Vital Brasil (IVB) está trabalhando nesse tipo de terapia, imunizando cavalos com uma proteína S recombinante do SARS-CoV-2 produzida na Coppe/UFRJ. Após 70 dias, plasmas de animais apresentam anticorpos 20 a 50 vezes mais potentes contra COVID-19 do que aqueles presentes no plasma sanguíneo de pessoas que tiveram COVID-19 (Cunha et al., 2020).

O Instituto Butantã também fez um soro imune contra SARS-CoV-2, mas neste caso os cavalos foram inoculados com vírus inativado por via subcutânea, por quatro vezes em intervalos de 1 semana (dias 0, 7, 14 e 21). Após a purificação das imunoglobulinas, foram realizados os ensaios de neutralização. Entre os resultados obtidos está que estudos de segurança pré-clínica realizados de acordo com as diretrizes da Agência Reguladora Brasileira com camundongos e coelhos não mostraram efeitos tóxicos da imunoglobulina F(ab')₂ anti-SARS-CoV-2 equina (Botosso et al., 2022).

1.8.1. Soroterapia no combate a picada por animais peçonhentos

Acidentes com animais peçonhentos representam um problema mundial e de grande importância em saúde pública devido ao grande número de casos existentes. O atendimento rápido é essencial para a sobrevivência das vítimas e redução de possíveis sequelas (Cupo, 2015). A exposição das pessoas a esses tipos de animais foi aumentada devido ao processo de urbanização, que aumentou a disponibilidade de alimentos e esconderijos, como os entulhos e materiais de construção, além do acúmulo de lixo e à presença de baratas, moscas e pequenos roedores. (Gonçalves et al., 2007).

De acordo com dados do Ministério da Saúde, em 2021 foram notificados cerca de 240.294 casos de acidentes envolvendo animais peçonhentos no Brasil, resultando em 496 mortes nesse período. Dentre os principais causadores desses acidentes estão os escorpiões (62%), as serpentes (12%) e as aranhas (11%). No que diz respeito aos acidentes ocasionados pela picada de aranhas, a principal espécie envolvida pertence ao gênero *Loxosceles*, a qual em 2021 foi responsável por 6.094 casos notificados (MINISTÉRIO DA SAÚDE, 2022). As aranhas do gênero *Loxosceles* são popularmente conhecidas como aranhas-marrom por apresentarem a coloração típica (Futrell, 1992). Essas aranhas apresentam ampla distribuição mundial (P. H. da Silva et al., 2004; Futrell, 1992), sendo predominantemente encontradas no

continente americano e africano, em função do clima tropical característico dessas regiões (Swanson & Vetter, 2006).

No Brasil, existem atualmente 19 espécies descritas de aranhas do gênero *Loxosceles*, sendo as espécies *L. gaucho*, *L. intermedia* e *L. laeta* responsáveis pela maioria dos acidentes notificados por esse gênero, sendo estes concentrados na região Sul e Sudeste (Miranda et al., 2021).

1.8.1.1. Loxoscelismo

O Loxoscelismo se refere ao termo usado para designar acidentes envolvendo aranhas do gênero *Loxosceles* (Schenone Fernández, 1998). A principal característica gerada pela inoculação do veneno durante a picada por esse tipo de aranha, é a presença de uma ferida dermonecrótica, podendo seguir com efeitos sistêmicos, evoluindo em alguns casos à falência renal e possível morte do indivíduo (Sezerino et al., 1998). A lesão na pele tem desenvolvimento lento, assim como sua cicatrização pode levar de dois a quatro meses devido à perda do tecido. Áreas com alta concentração de tecido adiposo, como abdômen, nádega e coxa, tendem a evoluir com lesões mais extensas e profundas (S. T. da Silva et al., 2005). A picada das aranhas marrons é indolor e os primeiros sinais e sintomas se manifestam algumas horas após o acidente. Raramente a vítima é capaz de capturar ou identificar o animal, dificultando o diagnóstico do loxoscelismo e contribuindo para que os casos não sejam corretamente notificados pelas autoridades (P. H. da Silva et al., 2004; Futrell, 1992). O veneno neste tipo de aranha é composto por uma variedade de componentes tóxicos com massa molecular geralmente entre 5 a 40 kDa (Veiga et al., 2000). No caso de aranhas de espécies distintas, o veneno possui uma composição similar, apresentando semelhante atividade biológica e imunoquímica (Barbaro et al., 2005).

Grande parte de os constituintes do veneno são proteínas com atividade enzimática, tais como hialuronidases, fosfatases alcalinas, proteases, esterases, hidrolases, ATPases, esfingomielinases D/Fosfolipases D, lipases e insetoxinas (P. H. da Silva et al., 2004; Gremski et al., 2010). Proteínas pertencentes à família de fosfolipases D, são conhecidas como toxinas dermonecróticas, visto que podem produzir efeitos deletérios que contribuem significativamente para a resposta típica após envenenamento (Gremski et al., 2010). A esfingomielinase D (SMases D) é uma fosfolipase que ocasiona os efeitos tóxicos do loxoscelismo tais como dermonecrose, hemólise e agregação plaquetária.

1.8.1.2. Tratamento do loxoscelismo: antivenenos

Diversos protocolos de tratamento para o envenenamento por animais peçonhentos foram propostos e testados. Atualmente, o tratamento dos efeitos do envenenamento é realizado de modo inespecífico e específico. De forma inespecífica são utilizados corticosteróides para o tratamento de feridas, e antibióticos como a tetraciclina no uso tópico contra a dermonecrose (Paixão-Cavalcante et al., 2007).

De modo específico, os anticorpos são os principais componentes dos antivenenos utilizados para neutralizar a atividade tóxica dos venenos de animais peçonhentos de qualquer origem, o que inclui serpentes, escorpiões e aranhas (Chippaux & Goyffon, 2008; de Almeida et al., 2008; Hogan et al., 2004; Theakston et al., 2003), quando administrado nas primeiras horas do envenenamento (Futrell, 1992; Hogan et al., 2004).

Dois antivenenos podem ser utilizados no tratamento do loxoscelismo. A produção do soro antiloxoscélico (SALox) é realizada pelo Centro de Produção e Pesquisa de Imunobiológicos (CPPI), no Estado do Paraná, e distribuído para todo o Brasil pelo Ministério da Saúde. O SALox é poliespecífico contra o veneno das três principais espécies de aranha marrom: *L. laeta*, *L. gaucho* e *L. intermedia* (Acácia Maria Lourenço Francisco Nasr et al., 2017).

O Instituto Butantã, em São Paulo, produz o soro antiaracnídico, a partir do veneno das aranhas *L. gaucho* e *Phoneutria nigriventer* e do escorpião *Tityus serrulatus* (Butantã, 2020). Na produção do antiveneno, as imunoglobulinas presentes no plasma dos animais imunizados são precipitadas e digeridas pela enzima pepsina, que digere os fragmentos Fc das moléculas de IgG resultando em fragmentos F(ab')₂. Dessa forma obtém-se anticorpos policlonais específico contra o veneno, ou seja, uma mistura de anticorpos secretados contra um antígeno específico, além de anticorpos não específicos. Dado que a maioria dos tratamentos são baseados em anticorpos, o seu estudo é muito importante, especialmente a nível biotecnológico.

Em 1901, o primeiro prêmio Nobel de medicina foi concedido pela investigação de soro imune contra a toxina tetânica produzidas pela imunização de cavalos. Depois disso, anticorpos policlonais de cavalo foram produzidos e utilizados para tratamento e profilaxia da difteria, tuberculose, tétano e pneumonia. O soro imune de cavalo em alguns países chegou a ser usado como tratamento para SARS-CoV-2 e representa no Brasil o tratamento específico oferecido pelo Sistema Único de Saúde (SUS) como terapia para acidentes de aranha do gênero *Loxosceles*. Mais de 120 anos se passaram desde a primeira vez que o cavalo foi usado para

produzir anticorpos contra diferentes doenças e o que se sabe sobre o sistema imunológico adaptativo desta espécie é muito pouco, principalmente a nível molecular.

Por isto, este trabalho visou compreender um pouco mais sobre os aspectos básicos do repertório de anticorpos de cavalo. Através do emprego de método de sequenciamento de nova geração (NGS), foi possível entender o nível de variabilidade do repertório e as características principais de anticorpos naturais de cavalos, e desta forma entender melhor como a geração de anticorpos nesta espécie. Também daqueles anticorpos induzidos após imunização de cavalos com veneno de aranhas do gênero *Loxosceles*.

Dessa maneira, além do conhecimento básico adquirido neste trabalho, ele também nos fornece ferramentas e abre perspectivas para obtenção de anticorpos recombinantes de cavalo que reconheçam o veneno de aranhas marrom e que possam ter propriedades neutralizantes, mas sem os efeitos adversos produzidos pelo tratamento atual como o choque anafilático, além de minimizar o uso de animais para a produção do soro policlonal.

2. OBJETIVO GERAL

Caracterizar o repertório natural de anticorpos de cavalos e também o de animais imunizados com veneno de aranha do gênero *loxosceles*, sequenciados por plataforma de sequenciamento de nova geração.

2.1. Objetivos específicos

- Obter o repertório natural (não imune) de anticorpos de quatro cavalos utilizando sequenciamento de nova geração;
- Caracterizar o repertório natural de anticorpos de cavalos quanto às suas principais características como clonalidade, uso gênico, tamanho e composição de aminoácidos da CDR-H3, SHM e diversidade juncional;
- Obter o repertório de anticorpos de quatro cavalos após imunização com veneno de *L. intermedia*, *L. laeta* e *L. gaucho*, utilizando sequenciamento de nova geração;
- Comparar as características do repertório de anticorpos de cavalos antes e após a imunização com veneno de *Loxosceles*.

3. MATERIAL E MÉTODOS

3.1. Obtenção de amostras de cavalo

As amostras de sangue periférico de cavalos machos, adultos, com idade entre 5 e 9 anos, saudáveis e sem raça definida foram obtidas em parceria com o Centro de Produção e Pesquisa de Imunobiológicos (CPPI) do Estado do Paraná. O delineamento experimental foi aprovado pela Comissão de Ética no Uso de Animais da Universidade Federal de Minas Gerais (CEUA - UFMG) sob o número de protocolo 190/2018.

Inicialmente, amostras de sangue periférico dos quatro animais foram obtidas em dois tempos distintos: antes do processo de imunização (animais não imunes) e após o processo de imunização (animais imunes) com o *pool* de venenos de aranhas do gênero *Loxosceles*.

A imunização dos cavalos foi realizada da seguinte maneira:

- Dia 0 (1ª imunização): 300 µg de veneno (divididos igualmente entre *L. intermedia*, *L. gaucho* e *L. laeta*) + 1200 µl de solução salina + 1500 µl de Adjuvante Completo de Freund;
- Dia 10: (2ª imunização): 450 µg de veneno (divididos igualmente entre *L. intermedia*, *L. gaucho* e *L. laeta*) + 1050 µl de solução salina + 1500 µl de Adjuvante Incompleto de Freund;
- Dia 21 (3ª imunização): 750 µg de veneno (divididos igualmente entre *L. intermedia*, *L. gaucho* e *L. laeta*) + 2250 µl de solução salina;
- Dia 30 (4ª imunização): como descrito na imunização anterior;
- Dia 37: coleta de sangue sete dias após a última imunização.

Em ambas as coletas (antes e depois da imunização), cerca de 60 ml de sangue periférico foram obtidos de cada animal utilizando tubos BD Vacutainer® contendo o anticoagulante EDTA. As células mononucleares (PBMCs) foram isoladas por gradiente de centrifugação utilizando Ficoll-Paque (Sigma-Aldrich). A centrifugação foi conduzida a 400 x g durante 30 minutos a 16°C, removendo o freio da centrífuga. A camada de PBMC foi aspirada e lavada duas vezes com D-PBS 1X. As células resultantes foram coradas com azul de Tripán para contagem das células na câmara Neubauer. Após a contagem, cerca de 1×10^7 células foram aliqüotadas em microtubos, adicionando-se 1 ml de TRizol (Life Technologies) para extração de RNA. O restante das células foi criopreservada em FBS 90% e DMSO 10%. Todas as amostras foram armazenadas a -80°C até o uso.

3.2. Extração de RNA

A extração do RNA total foi realizada pelo método do TRIzol (Rio et al., 2010). Os tubos contendo 1 ml de solução de PBMCs (1×10^7 células) em TRIzol foram descongelados em gelo. Em seguida, em uma capela de exaustão 200 μ l de clorofórmio foram adicionados a cada tubo. Os tubos foram homogeneizados em vórtex por 15 segundos, e depois incubados temperatura ambiente por 5 minutos.

Para conseguir a separação das fases, o tubo foi centrifugado a 12.000 x g por 15 minutos a 4 °C. Ao fim da centrifugação, a solução apresenta três fases: uma fase inferior de cor vermelha, contendo fenol, clorofórmio e proteínas; uma interfase contendo DNA; e uma fase superior aquosa e incolor, contendo o RNA. O sobrenadante, fase aquosa, foi então transferido cautelosa e lentamente, sem tocar ou aspirar material da fase intermediária para outro tubo de 1,5 ml. Em seguida, foi adicionado isopropanol na mesma quantidade do RNA e as soluções foram misturadas por inversão. Os tubos foram incubados overnight à -20 °C e então, os tubos foram centrifugados a 12.000 \times g por 15 minutos, a 4 °C. Após a centrifugação, o sobrenadante de isopropanol foi descartado e no fundo do tubo ficou o RNA. Em seguida, iniciou-se a lavagem do RNA com 1 ml de etanol 80% em água ultrapura e adicionado ao tubo que foi mantido no gelo durante toda a etapa de lavagem. O procedimento foi realizado na cabine de segurança biológica. Tomou-se o cuidado de misturar bem, por pipetagem, até o desaparecimento de todo o sedimento de RNA visível. Os tubos foram então centrifugados a 12.000 g \times por 10 min a 4 °C. O sobrenadante de etanol foi descartado, com cautela para manter o sedimento de RNA fixado ao tubo, e um novo ciclo de adição de etanol, mistura e centrifugação, idêntico ao anterior foi realizado.

Após novo descarte do sobrenadante mantendo o sedimento intacto, foi feita uma nova centrifugação nas mesmas condições, no tempo de 1 minuto. O excedente de etanol foi então removido por aspiração cautelosa com pipeta, e o tubo foi deixado aberto, para secagem do etanol residual presente no sedimento de RNA, por 1 hora, dentro da cabine de segurança biológica.

O sedimento de RNA foi então eluído em 15 μ l de água ultrapura tratada com DEPC (ThermoFisher Scientific). A amostra foi então mantida em gelo por 30 minutos, para completa solubilização do RNA, e em seguida, a qualidade e concentração do RNA extraído foram quantificadas usando o espectrofotômetro Nanodrop (ThermoFisher Scientific). As amostras

que apresentaram razão A260/A280 igual ou superior à 1,7 foram encaminhados para a etapa de transcrição reversa.

3.3. Síntese de cDNA e PCR Multiplex

Cerca de 500 ng de RNA foram utilizados para a síntese de cDNA utilizando a enzima SuperScript IV First-Strand (ThermoFisher Scientific) de acordo com instruções do fabricante.

Os reagentes do kit, com exceção da SuperScript™ IV Reverse Transcriptase (200 U/μl), foram descongelados mediante incubação em gelo. Foram adicionados 1 μl da solução 10 mM dNTP mix do kit, 1 μl de Oligo d(T) e 4 μl de amostra de cDNA e foi levado a um volume final de 13 μl com água ultrapura tratada com DEPC (ThermoFisher Scientific).

Os tubos foram então incubados em termociclador SimpliAmp™ Thermal Cycler (ThermoFisher Scientific) a 65 °C por 5 minutos. Foi então preparada uma solução "mix", contendo 4 μl do 5× SSIV Buffer, 1 μl da solução 100 mM DTT, 1 μl de Inibidor de Ribonuclease (40 U/μl) e 1 μl da SuperScript™ IV Reverse Transcriptase (200 U/μl) e incubado no termociclador a 50 °C por 10 minutos, seguido de 80 °C por 10 minutos. Por fim, 1 μl de RNase H do kit foi adicionado a cada tubo, e foi feita uma incubação final em termociclador à 37 °C por 20 min, para degradação do RNA total remanescente.

A amplificação dos segmentos gênicos IGHV da região variável e a região constante foi realizada por meio da técnica de PCR multiplex e utilizando um conjunto de *primers* específicos (Tabela 2). Este desenho experimental foi baseado na metodologia descrita para amplificação do repertório de anticorpos em humano (Dekosky et al., 2013; Mcdaniel et al., 2016) e camundongo (Reddy et al., 2010).

Tabela 2. Sequências dos *primers* utilizados para amplificar os genes IGH

IGHV <i>primers forward</i>	
hoHV2_4f	GTG TCC TGT CCC AGG TGC ARC
hoHV4f	GWG TCC TST CCC AGG TGM AGC
hoHV1f	GTG TCT CCT CCG AGG KYC AGC
hoHV1S6f	TGT CTT TCG CAC AGT AAT ACA TGG C
hoHV4S17f	AGC TCA CAG CTC TCT CCT AAG ACG AG
hoHV6S1f	AGG CTG AGG ACC CTC TCG TGC
hoHV7S1f	CTC TGT CCC AGA TCA GCC TGC
IGHV <i>primers reverse</i>	
IgM	ATGACGTTGGGTAGGAAGTCCCG
IgG1	
IgG2	
IgG3	
IgG4	CCACCGTGGMGTCAGAYGTG
IgG5	
IgG6	
IgG7	
IgA	AGCTGCAGTGTTTCTGAACACACTCC

Os *primers forward* (F) foram desenhados e testados em um trabalho anterior do nosso grupo de pesquisa (Manso et al., 2019). Esses primers compreendem sequências dos segmentos gênicos funcionais IGHV (16 genes), os quais se anelam na parte inicial desses, entre o *Leader* (L) e o FR1 (Figura 9). Foram usados novos primers específicos reversos (R) para a região constante da cadeia pesada desenhados neste estudo.

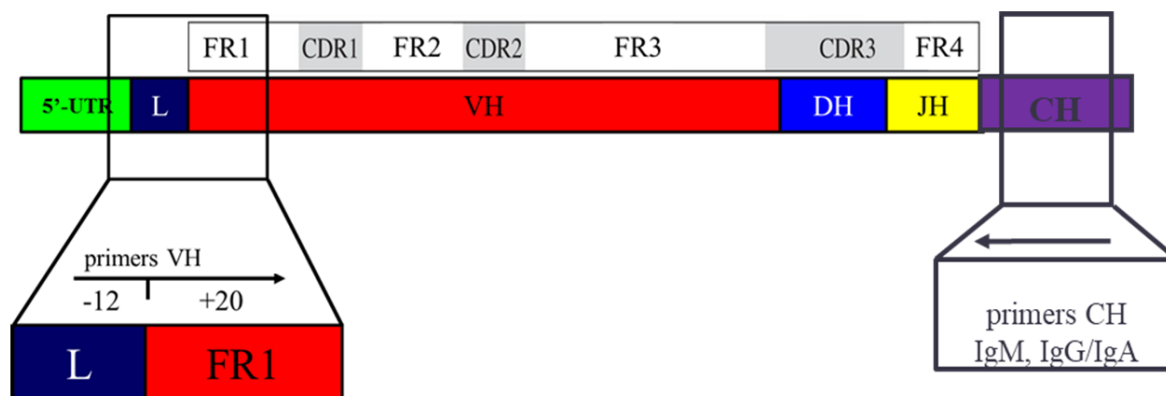


Figura 9. Organização dos genes para cadeia pesada de anticorpo de cavalo. Fonte: (MANSO, 2018).

Todos os *primers* incorporaram a sequência de adaptadores da Illumina para preparar a biblioteca. A necessidade de diferentes *primers forward* é devido à grande diversidade encontrada nos segmentos gênicos IGHV, e, portanto, para possibilitar a captura de todos os segmentos gênicos conhecidos e que podem estar sendo expressos nas amostras de cavalo.

As reações de PCR multiplex foram conduzidas para obter *amplicons* (produtos da PCR) IGH dos animais estudados. Todas as reações foram preparadas contendo 10X tampão High Fidelity, 50 mM MgSO₄, 10 mM dNTPs, 0,5 μM de cada *primer F*, 0,5 μM de cada *primer R* e Taq DNA polymerase Platinum High Fidelity (ThermoFisher Scientific). Os parâmetros de ciclagem foram 94 °C por 2 min; 4 ciclos de 94 °C por 1 min, 50 °C por 1 min e 72 °C por 1 min; 4 ciclos de 94 °C por 1 min, 55°C por 1 min e 72°C por 1 min; 26 ciclos de 94°C por 1 min, 63°C por 1 min e 72 °C por 1 min, e 72 °C por 7 min.

Os *amplicons* dos animais obtidos com os *primers* IGHV + IgA/IgG e IGHV + IgM, foram aplicados em gel de agarose 1% para avaliação eletroforética do tamanho do produto obtido. A corrida deu-se por cerca de 40 minutos a 100V, tempo suficiente para que a banda de 500 pares de base (pb), correspondente ao *amplicon* esperado e ficasse bem separada para excisão. As bandas específicas foram cortadas com bisturi e purificadas do gel utilizando o kit GFX™ PCR DNA e Gel Band Purification (GE Healthcare) conforme as instruções do fabricante.

Cada fragmento do gel contendo a banda foi pesado para que fosse ajustada a quantidade de tampão NTI (componente do kit Gel Band Purification) pudesse ser calculada. Após a adição deste tampão, o tubo foi colocado por 5 a 10 min a 60 °C, depois homogeneizado em vórtex por 1 a 3 min ou até que a banda estivesse completamente diluída no tampão. O volume máximo

de amostra que pode ser adicionado à coluna é um volume de 700 µl. Depois de centrifugado a 11.000 x g por 30 segundos, a solução que ficou no tubo foi descartada e a coluna lavada com etanol 80%, centrifugada novamente a 11.000 x g por 30 segundos. Esta etapa foi repetida mais uma vez para certificar que não há contaminantes, então a coluna foi centrifugada com um novo tubo a 11.000 x g por 1 min, a coluna foi deixada para secar por aproximadamente 5 min e então adicionou 15 µl de água ultrapura. A coluna foi centrifugada novamente a 11.000 x g por 1 min, e a amostra foi recuperada e então colocada no gelo por 30 min para quantificação. Uma vez purificado, o *amplicon* foi utilizado para a preparação da biblioteca necessária para realização do sequenciamento dos anticorpos dos animais.

3.4. Preparação da biblioteca para sequenciamento

As amostras do DNA purificado foram quantificadas por *Qubit* DNA High Sensitivity kit (ThermoFisher Scientific). Em seguida, o DNA de cada amostra de animal foi utilizado para a preparação das bibliotecas de sequenciamento utilizando o kit Nextera XT DNA *Library* Prep (Illumina) de acordo com as instruções do fabricante.

A PCR de indexação foi realizada para a adição dos identificadores, nas pontas dos *amplicons*, sequências necessárias para o sequenciamento Illumina. O protocolo e reagentes utilizados para essa PCR são do kit Platinum™ Taq DNA Polymerase High Fidelity, da ThermoFisher (ThermoFisher Scientific). Cada uma das amostras de cada cavalo, e cada isotipo (IgM e IgG) foram marcados com uma combinação distinta combinatorial dual indexes utilizados na PCR. As amostras e suas respectivas combinações de indexes estão na Tabela 3.

Tabela 3. Sequências de indexes Illumina.

Amostra	Isotipo	Index i7		Index i5	
		Código	Sequência	Código	Sequência
Cavalo 1	IgM	H703	AGGCAGAA	S503	TATCCTCT
	IgG/IgA	H704	TCCTGAGC	S503	TATCCTCT
Cavalo 2	IgM	H703	AGGCAGAA	S517	GCGTAAGA
	IgG/IgA	H704	TCCTGAGC	S517	GCGTAAGA
Cavalo 3	IgM	H702	CGTACTAG	S504	AGAGTAGA
	IgG/IgA	H703	AGGCAGAA	S504	AGAGTAGA
Cavalo 4	IgM	H706	TAGGCATG	S503	TATCCTCT
	IgG/IgA	H705	GGACTCCT	S503	TATCCTCT
Cavalo 1 H	IgM	H703	AGGCAGAA	S502	CTCTCTAT
	IgG/IgA	H704	TCCTGAGC	S502	CTCTCTAT
Cavalo 2H	IgM	H702	AGGCAGAA	S503	TATCCTCT
	IgG/IgA	H704	TCCTGAGC	S504	AGAGTAGA
Cavalo 3 H	IgM	H702	CGTACTAG	S502	CTCTCTAT
	IgG/IgA	H702	AGGCAGAA	S517	GCGTAAGA
Cavalo 4 H	IgM	H706	TAGGCATG	S502	CTCTCTAT
	IgG/IgA	H705	GGACTCCT	S502	CTCTCTAT

Para cada amostra, foi preparada uma reação contendo: 5 µl do Nextera CD Index 1 (i7) Primer (H7XX); 5 µl do Nextera CD Index 2 (i5) Primer (H5XX); 5 µl da solução 10X High Fidelity PCR Buffer do kit; 1 µl da solução 10 mM dNTP mix (Promega); 100 nmol da solução de MgSO₄ 50 mM do kit; 1 U da solução de Platinum® Taq DNA High Fidelity Polymerase (5 U/µl) do kit; 5 µl (caso a concentração estivesse acima de 10 ng/µl ou mas) ou 10 µl (caso a concentração estivesse entre 5 e 10 ng/µl) do *amplicon* purificado da PCR multiplex de cada amostra; e por fim, o volume da reação foi completado para 50 µl utilizando água ultrapura livre de nucleases. Os tubos foram então postos no termociclador, cujo programa foi: 5 minutos a 95 °C; 8 ciclos de 95 °C por 30s, 60 °C por 40s e 72 °C por 40s; 72 °C por 5 minutos; e 4 °C de incubação final.

3.5. Purificação da biblioteca com adaptadores utilizando *beads* magnéticas

A purificação do amplificado da PCR contendo os adaptadores (biblioteca) foi então realizada. O intuito dessa purificação foi à exclusão de *primers* livres e de dímeros e quimeras remanescentes dessa PCR (que podem competir pelos *clusters* do sequenciador contra os *amplicons* de sequências íntegras de transcritos de IGH).

As soluções do kit de purificação foram retiradas da geladeira para a temperatura ambiente, e homogeneizadas por vórtex. O volume final da PCR foi medido através de pipeta, e foi adicionado ao tubo da PCR a solução ProNex Chemistry (Promega), no volume de 1,1 vezes o volume da solução da PCR. A mistura foi homogeneizada por pipetagem, e incubada por 10 minutos a temperatura ambiente.

O tubo contendo a mistura foi posto na rack magnética por 2 minutos, e ainda com o tubo na rack, o sobrenadante foi extraído cuidadosamente com uma pipeta, sem desfazer o sedimento magnetizado. O sobrenadante foi descartado. Ainda com o tubo na rack magnética contendo o sedimento de *beads* foram adicionados 200 µl do tampão de lavagem, e o tubo foi incubado por 2 minutos. Após o descarte do sobrenadante por pipeta, sem tocar o sedimento, foi realizado um novo ciclo idêntico de lavagem e incubação com o tampão de lavagem.

Após o descarte do sobrenadante de tampão de lavagem, o tubo foi aberto para a secagem do sedimento de *beads* (já que o tampão de lavagem contém etanol, que pode acabar inibindo reações subsequentes). Após a secagem, o tubo foi removido com cautela da rack magnética, e foi feita a adição de 25 µl do tampão de eluição ao tubo, para ressuspensão do sedimento de *beads* com a pipeta. O tubo foi então incubado à temperatura ambiente por 5 minutos, para que o DNA se desprendesse das *beads*. O tubo foi então retornado à rack magnética, e incubado por 1 minuto. Após a incubação, o sobrenadante contendo o produto da PCR index de interesse foi aspirado e transferido para um novo tubo.

A concentração da biblioteca foi quantificada usando o *Qubit DNA High Sensitivity* kit (ThermoFisher Scientific) e o tamanho do *amplicon* foi corroborado com o Bioanalyzer 2100 G2939AA (Agilent), presente no Laboratório de Equipamentos Multiusuário do Departamento de Bioquímica e Imunologia do ICB-UFMG. Nos casos em que o Bioanalyzer mostrou a presença de pequenos fragmentos menores ao esperado, ou seja, houve remoção insuficiente dos *amplicons* de etapa de purificação, a amostra foi repurificada, seguindo o mesmo procedimento descrito nesta seção.

3.6. Sequenciamento

Após a preparação e quantificação da biblioteca, foi realizado o sequenciamento da cadeia pesada dos anticorpos de cavalos não-ímmunes e ímmunes. Esse procedimento foi feito na plataforma Illumina MiSeq 2x300 pb. As distintas amostras de cada cavalo foram unidas equimolarmente para que, somadas, apresentassem a concentração final de 18 picomolar. Essa solução também recebeu volume suficiente de PhiX Sequencing Control v3 (Illumina) para que esse tivesse a concentração de 5%. O sequenciamento foi realizado no MiSeq System (Illumina Instrument no. M02832) do Centro de Laboratórios Multiusuário do ICB-UFMG, com 301 ciclos de sequenciamento para cada *Read*, *forward* e *reverse*.

Os arquivos de saída do sequenciador foram então processados pelo aplicativo FASTQ Generation (Versão 1.0.0), disponível no BaseSpace Sequence Hub, serviço em nuvem da Illumina. Através desse software, os dados de leitura dos *clusters* do sequenciamento foram convertidos ao formato ".fastq", e as entradas do sequenciamento (isto é, os dados do par *Read* 1- 2 advindo de cada cluster) foram divididas, agrupando em um mesmo arquivo as entradas marcadas com o mesmo par de indexes *i5* e *i7*, isto é, advindas da mesma amostra de sequenciamento. Para a execução do FASTQ *Generation*, foi desativada a opção padrão de remoção (*trimming*) de bases advindas dos adaptadores Illumina. Um esquema da metodologia utilizada para a preparação das amostras é mostrado na Figura 10.

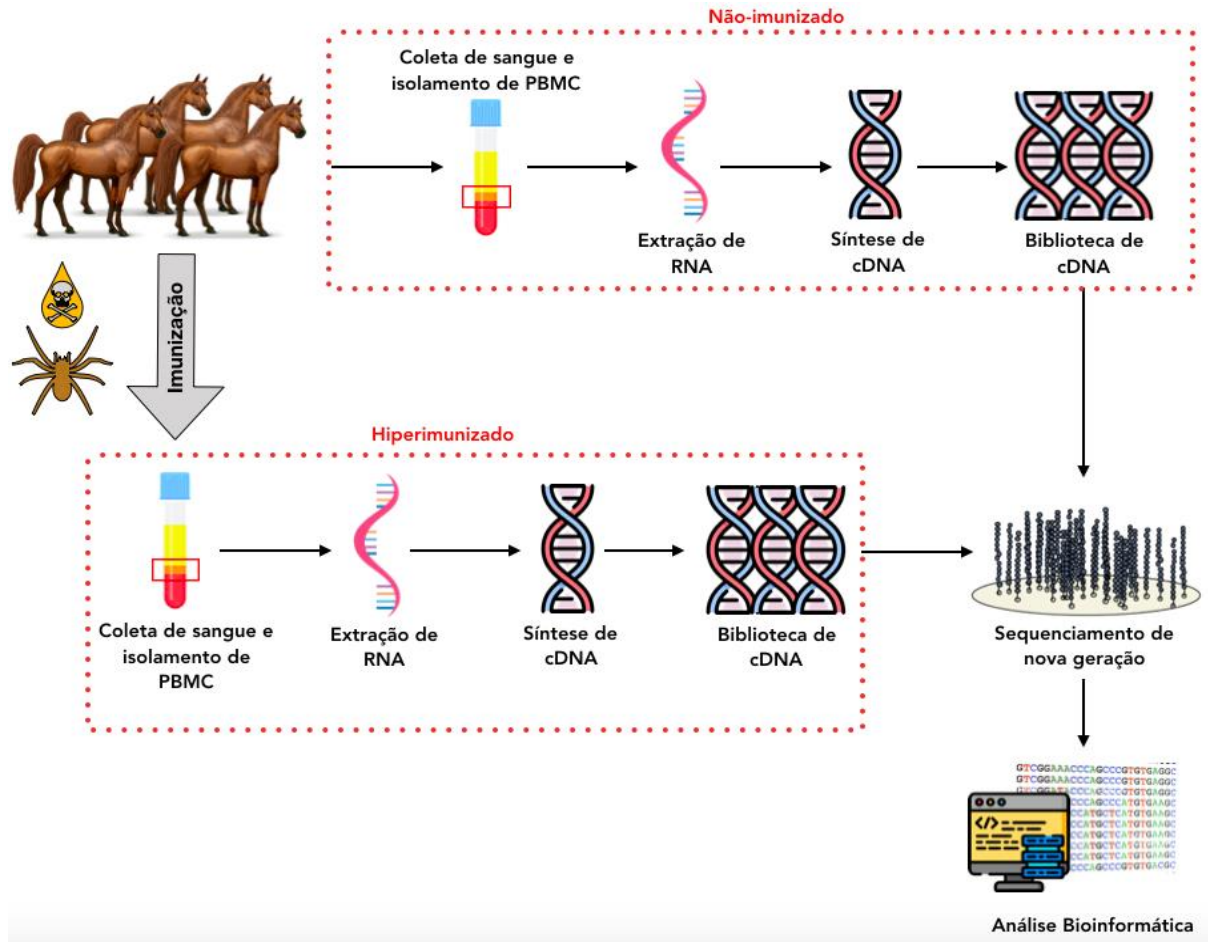


Figura 10. Metodologia utilizada para o estudo do repertório de anticorpos da cadeia pesada de cavalos não-imunes e imunes, onde é mostrada a extração de sangue dos cavalos, então a separação das PBMCs, extração de RNA, sínteses de cDNA, preparação de biblioteca, sequenciamento, e análise bioinformática.

3.7. Análises bioinformáticas do repertório de cavalo

Para a análise do nosso conjunto de dados foram necessárias adaptações nos programas já existentes e desenvolvimento de scripts específicos. A Figura 11 apresenta um esquema do pipeline seguindo neste trabalho e que será descrito posteriormente.

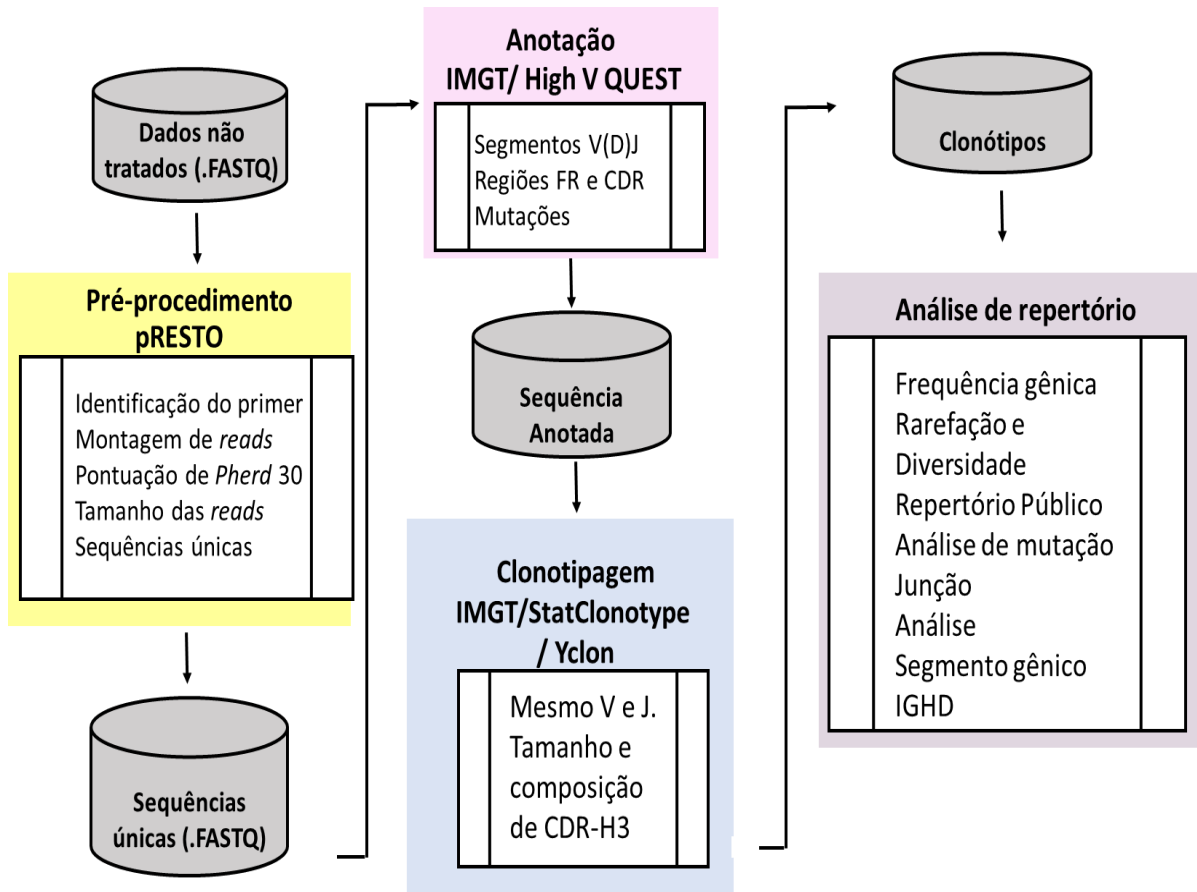


Figura 11. Fluxograma da análise de dados do repertório de anticorpos de cavalo. Os dados obtidos do sequenciamento estão no formato .Fastq. O primeiro passo é o pré-processamento que foi realizado com pRESTO, depois com a saída e formato .Fasta, os dados foram anotados e processados com IMGT/ High V Quest. Depois, foram agrupados usando o programa IMGT/StatClonotype, Yclone para comparação de animais não imunes e imunes, e então as diferentes características do repertório foram analisadas.

3.7.1. Pré-Processamento

As sequências foram processadas pelo pipeline pRESTO (VANDER HEIDEN et al., 2014) sendo conduzidas: (I) a montagem, (II) a eliminação de bases com *Phred score* abaixo de 30, (III) a identificação dos *primers*, (IV) eliminação de *reads* montadas com tamanho inferior a 100 nucleotídeos, (V) e a contagem das *reads*. Essas etapas serão detalhadas a seguir.

A montagem das *reads* foi realizada por alinhamento de suas extremidades 3', uma vez que o sequenciamento das *reads* foi realizado em ambos os sentidos (sequenciamento *paired-end*).

Após a montagem das *reads*, a qualidade das mesmas foi verificada através do índice *Phred Score*. O *Phred Score* (Q) é utilizado para analisar a acurácia das bases sequenciadas a

partir da probabilidade destas terem sido erroneamente inseridas na sequência (Richterich, n.d.). O *score* Q é definido como uma propriedade que é logaritmicamente relacionada à probabilidade (P) de inserções erradas (Ewing & Green, n.d.) e é obtido através da fórmula $Q = -10 \log_{10} P$.

Neste trabalho, utilizou-se um *Phred Score* de 30 (Q30) que corresponde a uma acurácia (probabilidade de uma inserção correta) de 99,9%. Sendo assim, a probabilidade de uma base ter sido adicionada erroneamente na sequência é 1 em 1000. Sequências que não atenderam ao critério de qualidade foram excluídas da análise.

Posteriormente, os *primers* utilizados para amplificação das sequências de IGH foram identificados por alinhamento em cada sequência. Para isso, um arquivo .Fasta contendo a sequência dos *primers* utilizados foi submetido ao pRESTO para que a identificação dos fragmentos fosse realizada, e as sequências em que os *primers* não foram identificados foram excluídas das análises. Finalmente, as sequências com sequências iguais foram contadas e a redundância do conjunto de dados removida, deixando apenas uma sequência representativa de cada sequências.

3.7.2. Anotação

Um arquivo no formato .Fasta, com sequências únicas, foi gerado para cada isotipo de imunoglobulina de cada cavalo (arquivos separados de IgM e IgG/IgA) para que fossem analisados isoladamente. Adotou-se essa abordagem, pois a amplificação do DNA de cada uma dessas classes foi feita em reações separadas e, no momento do sequenciamento, ocuparam poços distintos no mesmo chip.

Após o pré-processamento das sequências, e os genes IG foram anotados usando IMGT/HighV-QUEST (Alamyar et al., 2012), que identifica os genes variável (IGHV), diversidade (IGHD) e joining (IGHJ) em sequências de imunoglobulinas rearranjadas e, as mutações de nucleotídeos (nt) e alterações de aminoácidos (AA) resultantes de alterações de hipermutações somáticas em comparação com os conjuntos de diretórios de referência. Os conjuntos de diretórios de referência incluem sequências de referência presentes no IMGT (uma por alelo) de genes e alelos funcionais (F), fase de leitura aberta (ORF) e alelos de pseudogenes (P). A ferramenta integra IMGT/*Junction analysis* para a caracterização detalhada das junções V(D)J para uma anotação de sequência completa.

A saída IMGT/HighV-QUEST é fornecida em onze arquivos de resultados em formato .CSV. Nesta anotação, foi utilizado como referência o genoma *Equus caballus* 3.0.

3.7.3. Análise de clonalidade

Os rearranjos improdutivos V(D)J foram eliminados do conjunto de dados, bem como as sequências produtivas contendo inserções, deleções (*indels*) ou códon de parada nos segmentos dos genes V e J. As sequências foram agrupadas usando o IMGT/StatClonotype (Aouinti et al., 2016). Um clonotipo IMGT (AA) é definido por:

- Apresentarem os mesmos genes (e alelos) V e J (nt);
- Apresentarem um CDR-H3 idêntico (AA);
- Apresentarem aminoácidos âncoras conservadas 104, 118 (C104, W118 ou F118).

Os resultados do clonotipo IMGT (AA e nt) por locus são fornecidos em 10 seções (páginas HTML):

- Clonotipos IMGT (AA) por número;
- Clonotipos IMGT (AA) por número com clonotipos detalhados por nucleotídeos (nt);
- Clonotipos IMGT (AA) por gene IGHV;
- Clonotipos IMGT (AA) por gene IGHV com clonotipos detalhados (nt);
- Clonotipos IMGT (AA) por comprimento CDR3;
- Clonotipos IMGT (AA) por comprimento CDR3 com clonotipos detalhados (nt);
- Clonotipos IMGT (AA) por sequência CDR3 em ordem alfabética com clonotipos detalhados (nt);
- Diversidade do clonotipo IMGT (AA) e histogramas de expressão: por gene IGHV, IGHD, IGHJ e por comprimento CDR3;
- Tabelas de diversidade e expressão de clonotipo IMGT (AA): por gene IGHV, IGHD, IGHJ e por comprimento CDR3;
- Gene IGHV e tabela de alelos: Rearranjos, número de sequências e clonotipos de IMGT (AA) por gene IGHV e alelo.

Para análises de clonotipo foi usado como arquivo de entrada o arquivo de saída de IMGT/HighV-QUEST.

Para comparar os animais antes e depois da imunização, a clusterização foi realizada pelo Yclon, programa desenvolvido pelo nosso grupo de pesquisa, onde os clones são definidos como sequências com os mesmos segmentos gênicos IGHV e IGHJ, o mesmo tamanho que CDR-H3 e similaridade na sequência CDR-H3. Essas similaridades são determinadas pela vetorização da sequência CDR-H3 e, em seguida, criando uma matriz de similaridade para comparar cada vetor. Após isso, as sequências são agrupadas usando um *threshold* de 0,09 (Gervásio et al., 2022).

3.8. Análise de repertório

Após o processamento das sequências, análises da diversidade dos anticorpos foram conduzidas, avaliando as frequências do uso dos segmentos gênicos e dos subgrupos gênicos e a combinação do uso dos genes V(D)J em cada animal. As sequências de aminoácidos das CDR-H3 dos animais foram analisadas quanto ao tamanho e a composição de aminoácidos através do programa R *studio*. A região de CDR-H3 corresponde às posições 105 a 117 de acordo com a numeração do IMGT (Lefranc, 2003).

O R *studio* também foi usado para obter as análises de repertório público de anticorpos dos cavalos não imunes. Os clones públicos foram definidos como aqueles presentes em diferentes cavalos e contendo anticorpos com os mesmos segmentos gênicos IGHV, IGHJ e CDR-H3 idênticos.

3.8.1. Análise de rarefação e construção de curva de riqueza de espécies

Usamos o programa iNEXT (Hsieh et al., 2016) para subamostrar populações de clonotipos de cadeias pesadas de imunoglobulinas dos quatro cavalos com base na frequência de sua ocorrência. Este programa realiza um *resample* das amostras (cada 10%) e calcula a abundância e diversidade de clones para um determinado número de sequências. O iNEXT também foi usado para extrapolar além do número de sequências produtivas observadas experimentalmente, o que poderíamos esperar com sequenciamento adicional. Este pacote calcula as estimativas de diversidade, cobertura da amostra e tamanhos de amostra entre o tamanho 1 e o ponto final. Cada nó representa um tamanho de amostra particular para o qual as

estimativas de diversidade serão calculadas. Por padrão, o ponto final é igual ao dobro do tamanho da amostra de referência (tamanho total da amostra para dados de abundância).

Recon (Kaplinsky & Arnaout, 2016) foi usado para estimar o número de clonotipos ausentes nos conjuntos de dados de cadeia pesada de imunoglobulina. O Recon assume que cada clone no indivíduo contribui com células para a população amostrada de acordo com uma distribuição de Poisson. Isso será verdade se (i) os clones estiverem bem misturados no sangue ou distribuídos uniformemente no tecido que está sendo amostrado, (ii) a população parental for suficientemente grande para que a estimativa de Poisson da probabilidade de, por exemplo, um *singleton* contribuir >1 célula seja insignificante e (iii) nenhum clone único é uma grande fração (~30% ou mais) da população parental.

3.8.2. Determinação do repertório expandido

Para determinar quais clones produtores de anticorpos foram expandidos após a imunização com veneno de aranha do gênero *Loxosceles*, utilizamos como referência a definição de clones expandidos utilizada por Wang et al no 2018. Assim, os clones expandidos apresentam duas características: a primeira é que são definidos como todos os clones com expansão de frequência de mais de 3 vezes no animal imune quando comparados com o mesmo clone antes da imunização; a segunda característica é que aqueles que estão em alta frequência (>0,05%) no repertório imune, também são considerados clones expandidos. Os clones que apresentam uma ou as duas características são considerados expandidos (L. Wang et al., 2018).

3.8.3. Determinação da fase de leitura do Gene IGHD

Para determinar a fase de leitura, ou RF (do inglês *reading frame*) do segmento gênico IGHD, primeiro determinamos o índice de hidrofobicidade de cada uma das 6 possíveis RF, os três por exclusão (RF1, RF2, RF3), e três por inversão iRF (iRF1, iRF2, iRF3) utilizando o pacote Peptídeo do R *studio* (Osorio et al., 2015). O pacote usa as definições e valores de hidrofobicidades da escala de Kyte-Doolittle (Kyte & Doolittle, 1982).

Assim, a RF1 é aquela mais hidrofílica, a RF2 a mais hidrofóbica e a RF3, aquela hidrofóbica com códon de terminação, no caso de fases de leitura invertidas, iRF1 foi

determinado como aquele com aminoácidos carregados, iRF2 como o mais hidrofóbico e iRF3 como o hidrofóbico com códon de terminação. (Ivanov I, Link J., Ippolito G.C., 2002).

3.8.4. Hipermutação somática

A hipermutação somática é calculada pelo IMGT/HighV-QUEST. Para tal, as sequências de entrada são comparadas com àquelas da linha germinativa do segmento IGHV mais similar à sequência de entrada. As mutações são descritas de acordo com a numeração do IMGT para IGHV. A mutação do nucleotídeo é descrita, seguida por uma vírgula e a mudança de aminoácido da linha germinativa do segmento gênico IGHV mais semelhante. Um exemplo é se "T" foi mutado para "C" na posição 88, é reportado "t88>c" e se o aminoácido "F" na posição 30 foi alterado para "L" se reporta "F30>L".

Números de mutações e alterações de aminoácidos são mostrados para toda a região IGHV: o FR1-IMGT, CDR1-IMGT, FR2-IMGT, CDR2-IMGT, FR3-IMGT e a parte do CDR3-IMGT proveniente do IGHV. Os motivos, as posições e a localização dos FR-IMGT ou CDR-IMGT de *hotspots*, ou pontos chave, dentro de IGHV da linhagem germinativa mais próxima são listados. Um exemplo de como IMGT/HighV-QUEST relata os *hotspots* pode ser visto na Figura 12.

Hotspots motifs and localizations in germline V-REGION

<u>(a/t)a</u> wa		<u>t(a/t)</u> tw		<u>(a/g)g(c/t)(a/t)</u> rgyw		<u>(a/t)(a/g)c(c/t)</u> wrcy	
Motif	Localization	Motif	Localization	Motif	Localization	Motif	Localization
aa	58-59 (FR1)	tt	34-35 (FR1)	agct	8-11 (FR1)	agct	8-11 (FR1)
aa	59-60 (FR1)	tt	82-83 (CDR1)	ggct	31-34 (FR1)	agcc	41-44 (FR1)
ta	114-115 (CDR1)	tt	88-89 (CDR1)	ggtt	80-83 (CDR1)	aact	59-62 (FR1)
aa	142-143 (FR2)	ta	114-115 (CDR1)	ggct	106-109 (CDR1)	agcc	72-75 (FR1)
aa	143-144 (FR2)	tt	135-136 (FR2)	ggct	132-135 (FR2)	tgct	111-114 (CDR1)
ta	165-166 (FR2-CDR2)	tt	158-159 (FR2)	ggct	146-149 (FR2)	agct	177-180 (CDR2)
ta	168-169 (CDR2)	ta	165-166 (FR2-CDR2)	ggtt	156-159 (FR2)	tgct	201-204 (FR3)
aa	171-172 (CDR2)	tt	167-168 (CDR2)	agca	172-175 (CDR2)	agcc	277-280 (FR3)
aa	175-176 (CDR2)	ta	168-169 (CDR2)	agct	177-180 (CDR2)	aacc	285-288 (FR3)
aa	176-177 (CDR2)	ta	180-181 (CDR2)	agtt	184-187 (CDR2)	tact	307-310 (FR3)
ta	180-181 (CDR2)	tt	186-187 (CDR2)	agca	195-198 (CDR2)	tact	312-315 (FR3)
aa	181-182 (CDR2)	ta	187-188 (CDR2)	ggca	220-223 (FR3)		
ta	187-188 (CDR2)	ta	199-200 (FR3)	ggtt	224-227 (FR3)		
ta	199-200 (FR3)	tt	226-227 (FR3)				
aa	214-215 (FR3)	tt	246-247 (FR3)				
aa	215-216 (FR3)	ta	262-263 (FR3)				
aa	249-250 (FR3)	ta	304-305 (FR3)				
aa	250-251 (FR3)	tt	306-307 (FR3)				
aa	253-254 (FR3)	ta	307-308 (FR3)				
ta	262-263 (FR3)	ta	312-313 (FR3)				
aa	269-270 (FR3)	ta	315-316 (CDR3)				
aa	270-271 (FR3)						
aa	274-275 (FR3)						
aa	283-284 (FR3)						
aa	284-285 (FR3)						
aa	285-286 (FR3)						
ta	304-305 (FR3)						
ta	307-308 (FR3)						
ta	312-313 (FR3)						
ta	315-316 (CDR3)						

Figura 12. Exemplo de arquivo IMGT/HighV-QUEST de *hotspots* localizados em IGHV. A primeira linha mostra o tipo de motivo analisado, depois a primeira coluna mostra o motivo e a próxima coluna mostra a posição desse motivo. Por exemplo, se for um motivo de 4 nucleotídeos (agtc), primeiro exibe a localização em número (8 a 11) e em seguida, especifica a região em que está localizado (FR1).

3.8.5. Análise de Junção

A análise da junção é realizada também pelo IMGT/HighV-QUEST. Nesta, cada nucleotídeo, ou conjunto de nucleotídeos, tem que ser atribuído à região mais provável, de acordo com a ordem de eventos biológicos. A ordem da pesquisa, da primeira região à última é a seguinte: região IGHV, região IGHJ, região IGHD, regiões P (de nucleotídeos palindrômicos) e regiões N (de nucleotídeos aleatórios). Uma vez delimitada uma região, a seguinte é pesquisada, excluindo a região anterior (Monod et al., 2004).

No caso das regiões IGHV, IGHD e IGHJ, o software compara cada nucleotídeo com os da linhagem germinativa correspondente (é necessário anotar o gene específico). Feito isso, inicia-se a análise das regiões P a partir do último nucleotídeo da região IGHV, do primeiro e último nucleotídeo da região IGHD e do primeiro nucleotídeo da região IGHJ, onde é buscado se o nucleotídeo vizinho é complementar. Em caso afirmativo, busca-se também pelo nucleotídeo antes do último de cada região IGHV, IGHD ou IGHJ, e se este é complementar ao segundo nucleotídeo na junção, e assim por diante. As regiões P são geralmente muito curtas (1–3 nt).

As regiões N são as partes das sequências, que permanecem após as delimitações das regiões IGHV, IGHD e IGHJ e as identificações das regiões P (Monod et al., 2004).

3.8.6. Determinação das subclasses IgG

A determinação de subclasses de IgG foi realizada com um programa python desenvolvido pelo nosso grupo de pesquisa que alinha as sequências com a sequência de nucleotídeos dos segmentos gênicos (subclasse de IgG) encontrada na base de dados do IMGT.

3.8.7. Análises estatísticas

Para realizar a análise estatística, primeiramente foi verificada a normalidade dos dados por meio do teste de Shapiro-Wilk, obtendo-se como resultado dados não paramétricos

Diferenças na comparação das medianas entre os grupos foram analisados pelo teste não paramétrico de Mann-Whitney.

A comparação dos indivíduos antes da imunização, após a imunização e os clones expandidos foi realizada pelo teste de Fisher. A expressão percentual média foi calculada a partir de as médias dos cavalos individuais para cada grupo.

4. RESULTADOS

4.1. Caracterização de Repertório Natural de Anticorpos de Cavalo

4.1.1. Obtenção de *amplicons* das cadeias pesadas de anticorpos equinos

Com o objetivo de conhecer a composição e diversidade de anticorpos equinos, a primeira etapa deste projeto visou a amplificação das cadeias pesadas de anticorpos de cavalo. Essa amplificação foi realizada por PCR multiplex, separadamente para IgM e para IgG/IgA. A Figura 13 mostra o resultado dessa amplificação em algumas das amostras estudadas., onde é possível observar uma banda de 500 pb conforme esperado para o *amplicon* da cadeia pesada (IGH), para os 4 cavalos analisados.

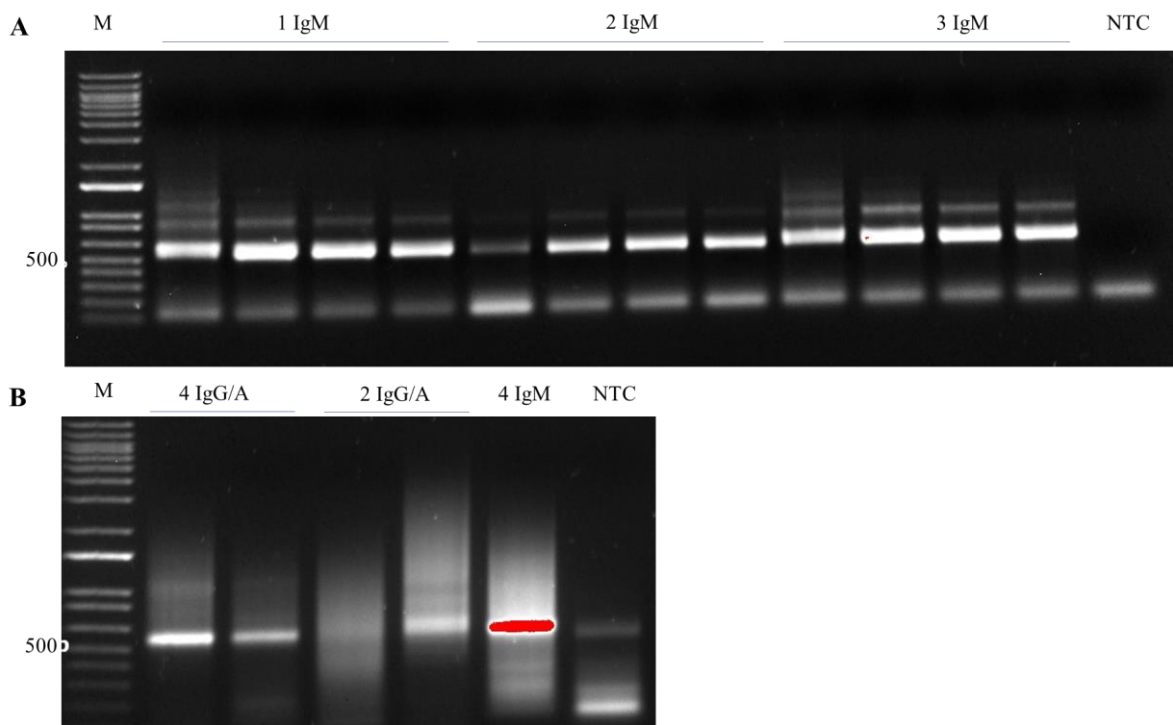


Figura 13. Gel de agarose (1%) dos *amplicons* de cadeias pesadas de anticorpos equinos. Amostras representam a amplificação das cadeias pesadas dos anticorpos presentes nas amostras de PBMCs. NTC corresponde ao controle negativo conduzido com todos os primers F (*forward*) e sem o cDNA de cavalo. (A) Reações individuais com os *primers* IgM da cadeia pesada para amostra de cavalo 1, 2, e 3. (B) Reações individuais com os primers IgG/A da cadeia pesada para as amostras de cavalo 4 e 2 e do *primers* IgM para o cavalo 4. Todas as reações foram realizadas em quadruplicata para se obter uma maior concentração de DNA. Após a amplificação do repertório de imunoglobulinas de cavalos, os *amplicons* foram

excisados do gel de agarose, purificados e quantificados para a preparação da biblioteca de sequenciamento. O marcador molecular é 1kb plus Invitrogen.

4.1.2. Análise das bibliotecas de amplicons de IGH para o sequenciamento do repertório de anticorpos equinos

A biblioteca de *amplicons* de cadeia pesada de imunoglobulina preparada para o sequenciamento foi quantificada por *Qubit* (concentração dada em ng/ μ L e convertida em mM) (Tabela 4). Para converter em mM é necessário o tamanho do *amplicon*, o que foi verificado usando o Bioanalyzer, onde é considerado o tamanho da região variável da cadeia pesada do anticorpo de aproximadamente 450 pb mais o tamanho dos *primers* juntamente com os adaptadores Illumina. Tendo isso em vista, o tamanho do *amplicon* resultou em aproximadamente 590 pb. Nesta análise feita para os 4 cavalos não imunes, para anticorpos IgM e IgG/A, obtivemos uma concentração de 3,7 a 24,5 ng/ μ L, variando entre 9,5 a 63,5 mM (Tabela 4).

Tabela 4. Concentração da biblioteca de amplicons obtida por *Qubit* em ng/ μ L e mM.

Amostra	Qubit (ng/ μ L)	mM
1N-IgM	20,2	51,1
1N-IgG/A	3,7	9,5
2N-IgM	5,4	13,9
2N-IgG/A	9,7	24,9
3N-IgM	15,4	24,7
3N-IgG/A	24,5	63,5
4N-IgM	23,8	60,1
4N-IgG/A	4,2	10,8

4.1.3. Uso restrito do gene VDJ em anticorpos de cavalos

Não foram encontradas sequencias para o subtipo IgA, portanto todos os resultados obtidos nas análises são exclusivamente para IgG e IgM. Sequenciamos a região variável das cadeias pesadas de IgG e IgM de quatro cavalos individuais. No geral, a média de *reads* brutas

por amostras de IgM foi de 1.082.148 (354.598-1.558.035) e 347.300 (210.964-481.141) para amostras de IgG. Obtivemos 31-64% de sequências produtivas e entre 40.018 a 328.300 clonotipos de anticorpos de cavalo (Tabela 5).

Tabela 5. Visão geral dos resultados de sequenciamento da região variável da cadeia pesada de IgM e IgG de quatro cavalos não imunizados

Amostras	Isotipo	Dado bruto	Reads Preprocessadas	Reads anotadas	Reads produtivas	Clones
1	IgM	1.558.035	779.236	526.221	484.236	294.600
2	IgM	1.499.997	900.306	891.250	790.308	328.300
3	IgM	915.963	690.915	529.584	485.888	283.600
4	IgM	354.598	268.020	203.623	193.644	86.177
1	IgG	210.964	187.229	182.918	175.686	43.861
2	IgG	288.611	233.147	219.402	209.248	40.018
3	IgG	408.483	306.844	304,394	271,204	91.136
4	IgG	481.141	390.341	382,438	361,320	89.551

O genoma do cavalo EquCab3.0 inclui 21 segmentos gênicos funcionais IGHV, 17 IGHD e 6 IGHJ, levando, somente com essas combinações a 2.016 possíveis anticorpos. No entanto, em nosso estudo, observamos uma forte preferência por segmentos gênicos do subgrupo IGHV4, onde o IGHV4-21, IGHV4-29 e IGHV4-22 são usados por 80% dos anticorpos de cavalo nos isotipos IgM e IgG (Figura 14A).

Da mesma forma, IGHJ4 e IGHJ6 são os segmentos gênicos IGHJ mais usados pelos isotipos IgM e IgG (Figura 14B). Curiosamente, IGHJ6 está presente em quase 60% de todos os clones dos cavalos, mostrando um uso restrito de segmentos gênicos IGHV e IGHJ. Todos os 17 genes funcionais de IGHD parecem ser usados por anticorpos de cavalos.

É importante notar que além dos genes funcionais, outros genes foram encontrados nos repertórios estudados, por exemplo, os genes ORFs IGHV4-21 e IGHV3-62, além dos IGHJ1 e IGHJ2. O mesmo aconteceu com os genes IGHD onde foram encontrados todos os genes ORFs IGHD1-32, IGHD1-42, IGHD2-21, IGHD2-3, IGHD2-39, IGHD3-1, IGHD3-10, IGHD3-13, IGHD3-16, IGHD3-19, IGHD3-22, IGHD3-24, IGHD3-27, IGHD3-30, IGHD3-

37, IGHD3-4, IGHD3-40, IGHD3-7, IGHD4-14, IGHD4-17, IGHD4-20, IGHD4-25, IGHD4-28, IGHD4-41, IGHD4-8, IGHD7-33 e IGHD7-43 (Figura 15).

As combinações de V(D)J mais frequentes utilizadas por anticorpos de cavalo foram IGHV4-21, IGHD2-26 e IGHJ6-1, encontradas em 2,8% ($\pm 0,9$) dos isotipos IgM e 2,6% ($\pm 0,6$) do IgG (Figura 14C). Nossos resultados sugerem que o repertório de IGHV de cavalo parece ser derivado de famílias de genes de linhagem germinativa limitadas.

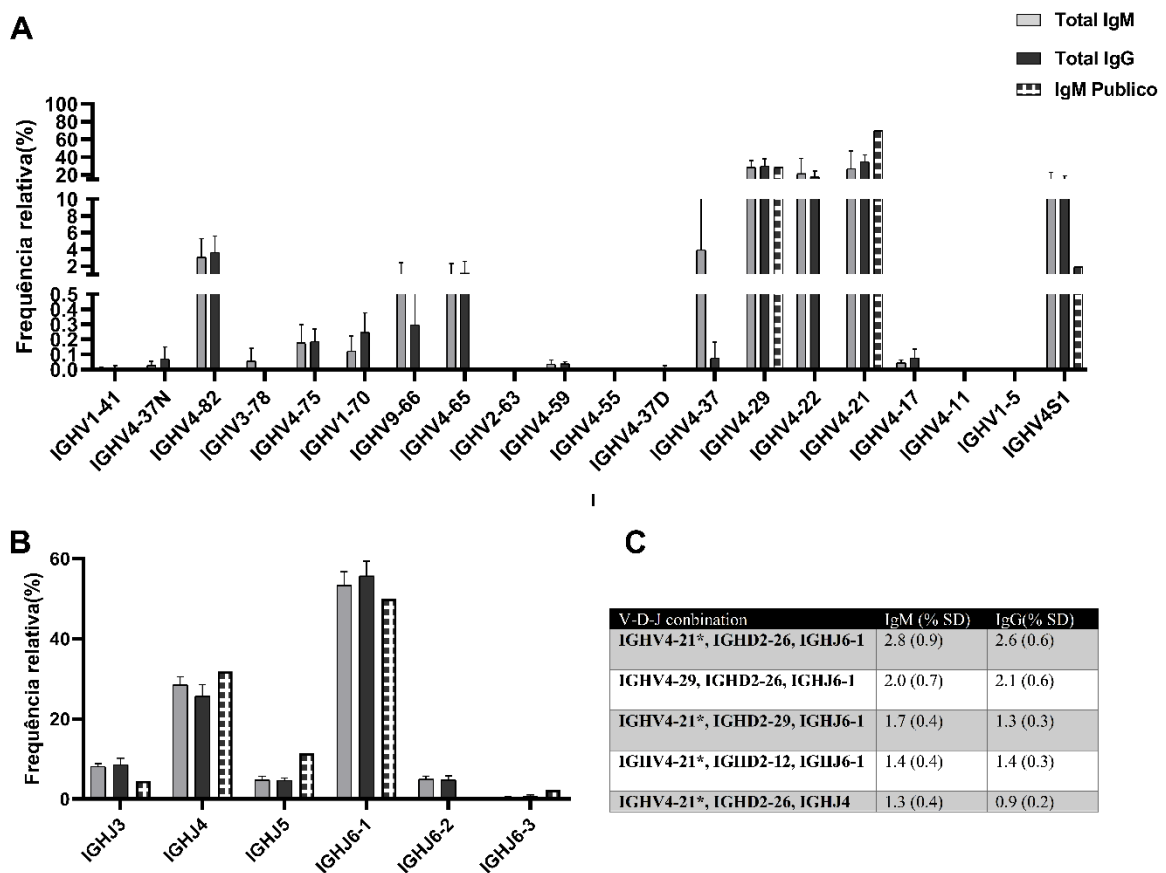


Figura 14. Frequência do uso dos segmentos gênicos da região variável da cadeia pesada em anticorpos equinos. (A) Mediana da frequência relativa (%) de IGHV, (B) IGHJ e (C) combinação V(D)J mais frequente no isotipo IgM e IgG e no repertório público de quatro cavalos não imunizados. Os genes são organizados na ordem em que aparecem no genoma do EquCab 3.0 (5'-3').

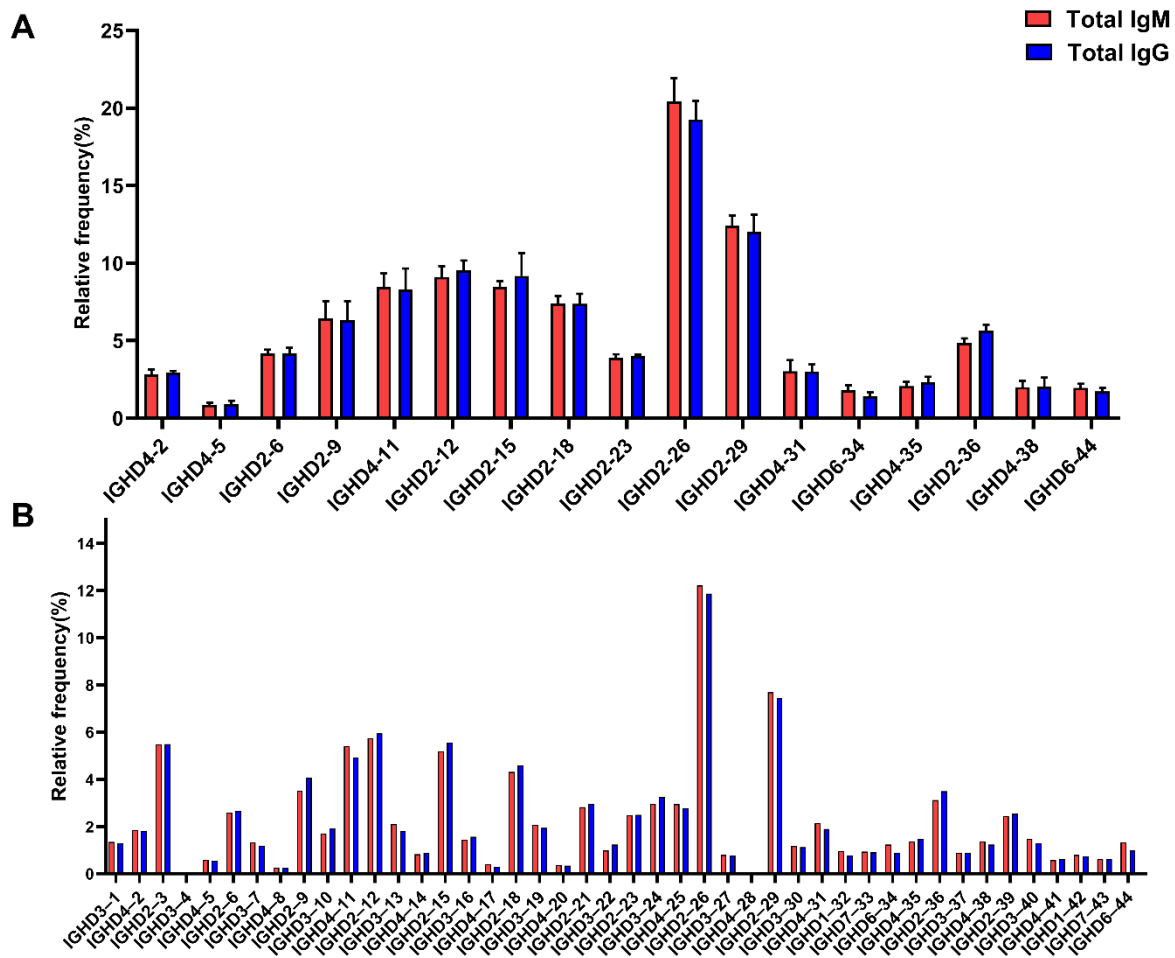


Figura 15. Frequência do uso dos segmentos gênicos IGHD em anticorpos equinos. (A) Frequência relativa de segmentos gênicos funcionais de IGHD. (B) Frequência relativa dos segmentos gênicos IGHD total, funcional e ORF.

Nossa análise também mostrou que mais clones raros foram observados em amostras de IgM do que em IgG, uma vez que as curvas de rarefação não começam a atingir o platô, indicando que é improvável que capturemos toda a diversidade dessa população (Figuras 16A e 16B). No entanto, conseguimos capturar entre 52 a 66% e entre 62 a 82% de todas as IgM e IgG, respectivamente (Figura 16C). Ainda, não observamos diferença estatística significativa entre a diversidade dos dois isotipos de cavalo quando foi comparada a diversidade de Shannon (Figura 16D) e Simpson (Figura 16E).

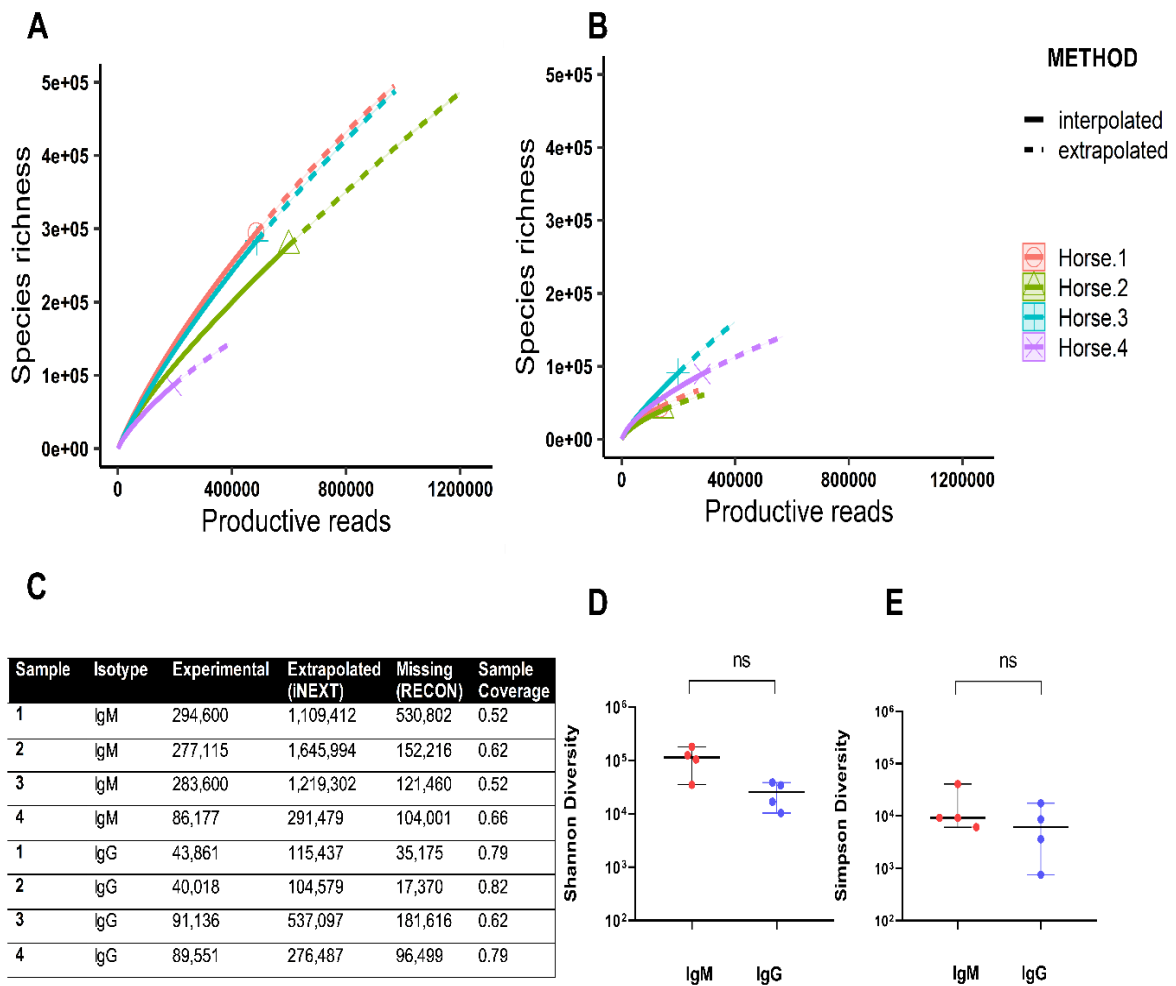


Figura 16. Estimativas da riqueza e diversidade do repertório da cadeia pesada de anticorpos para IgM e IgG nos quatro cavalos não imunizados. A interpolação e extrapolação da riqueza de espécies foram obtidas usando iNEXT para IgM (A) e IgG (B). As linhas contínuas correspondem à interpolação (com base em dados experimentais), e as linhas tracejadas pertencem aos dados extrapolados. (C) Resumo das estimativas de tamanho do repertório, incluindo clones ausentes. (D) A comparação da diversidade de Shannon e (E) Simpson entre os isotipos IgM e IgG ($p < 0,05$).

4.1.4. Repertório público de anticorpos de cavalo análise de CDR-H3

Neste estudo, observamos que os quatro cavalos compartilhavam (repertório público) apenas 0,05% (44 clones) de seu repertório de IgM e 0,0099% (4 clones) do repertório de IgG (Figuras 17A e 17B, respectivamente). Para o repertório público de IgM, a maioria dos clones apresenta o segmento gênico IGHV4-21 (77%), enquanto no repertório total, esse segmento gênico representa aproximadamente 32% (Figura 14A). No caso dos genes IGHJ, observamos

um aumento da utilização dos segmentos gênicos IGHJ4 (de 4% para 9%) e IGHJ5 (de 29 para 32%) no repertório público de IgM (Figura 14B).

Interessante notar que mais de 90% das CDR-H3 encontradas no repertório público de anticorpos IgM apresentavam apenas cinco aminoácidos de comprimento (Figura 17C). Em geral, a distribuição de tamanho de CDR-H3 de cavalos possui tamanhos variando de 4 a 51 resíduos de aminoácidos com um comprimento médio de 16 resíduos para ambos os isotipos IgM e IgG (Figura 17C).

Curiosamente, aminoácidos polares como glicina (G) e tirosina (Y) estão aumentados no repertório público IgM, diferentemente dos aminoácidos carregados negativamente ácidos aspártico (D) e glutâmico (E) e dos aminoácidos hidrofóbicos fenilalanina (F), triptofano (W) e alanina (A) que diminuem (Figura 17D).

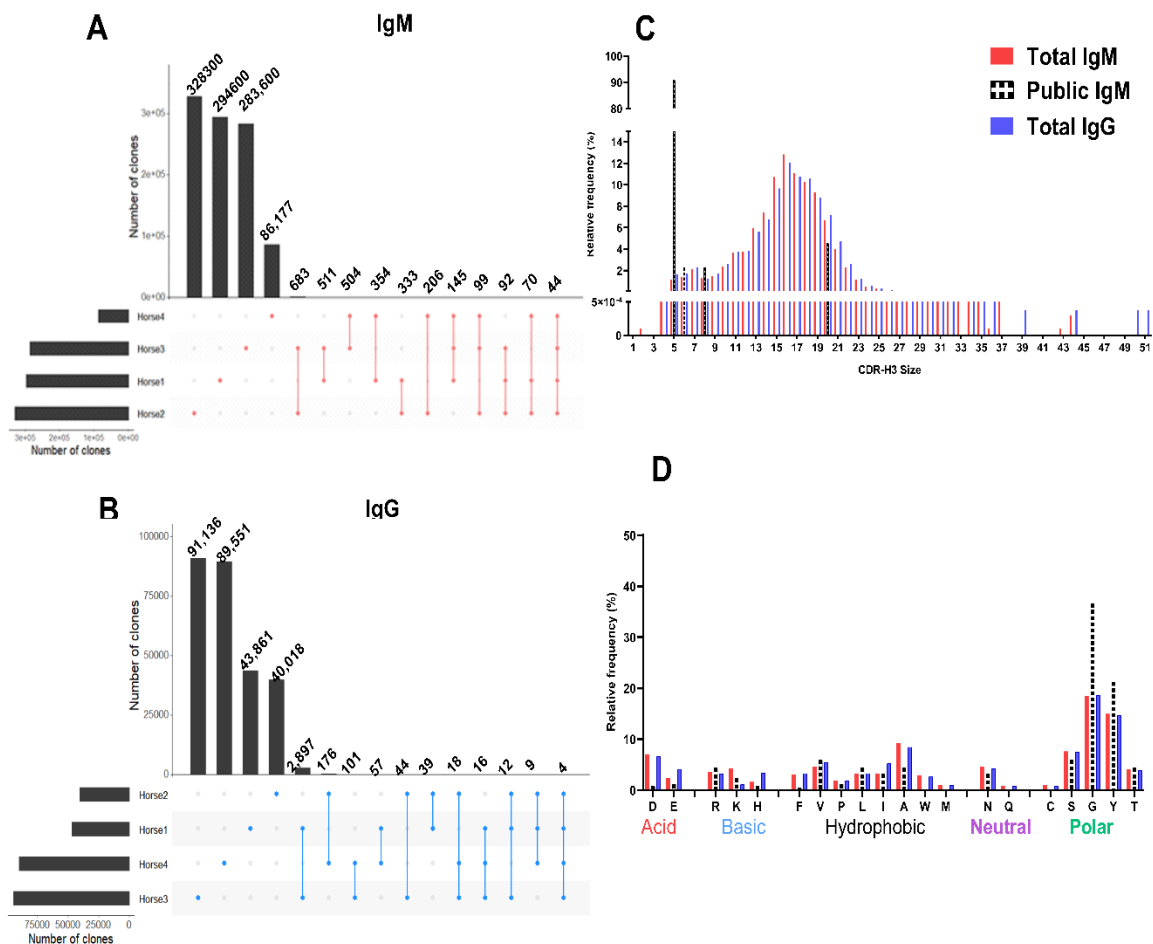


Figura 17. Caracterização dos clones e do CDR-H3 dos repertórios da região variável de cadeia pesada de cavalo. O número de clones de anticorpos apresentados pelos diferentes cavalos e o número compartilhado de clones entre 2, 3 ou 4 cavalos neste estudo para (A) IgM e (B) IgG. (C) Comparação do comprimento do CDR-H3 e (D) sua composição de aminoácidos entre o repertório de IgM Total, IgG Total e IgM público.

4.1.5. Caracterização da frequência e padrão de hipermutação somática (SHM) encontrada em anticorpos de cavalos

Com base em nossos resultados anteriores, hipotetizamos que a maior diversidade de imunoglobulinas de cavalo vem da hipermutação somática e/ou da diversidade juncional.

Aqui, observamos que a frequência de mutações no isotipo IgG (média: 7,22%) foi semelhante à IgM (média: 6,46%) em comparação com sua linhagem germinativa mapeada no genoma EquCab3.0 (Figura 18A). A maioria das mutações dos aminoácidos na região de IGHV foi encontrada nas regiões de CDR, especialmente nas posições 32 (CDR1), 50, 52 e 58 de CDR2, mas também no FR3, posição 88 para ambos os isotipos IgM e IgG (Figura 18B). Uma tabela com a variação percentual dos três aminoácidos principais e cada pico de mutação é mostrada no anexo (Anexo I). O perfil de mutação para cada cavalo estudado para ambos os isotipos pode ser visto no Anexo II.

Também observamos que uma média de 16,79% dos nucleotídeos estão mutados nas regiões de CDRs da região IGHV, sendo que a maioria deles (45 a 51%) está presente nos motivos de reconhecimento da AID (RGYW e motivos complementares de nucleotídeos WRCY) (Tabela 6).

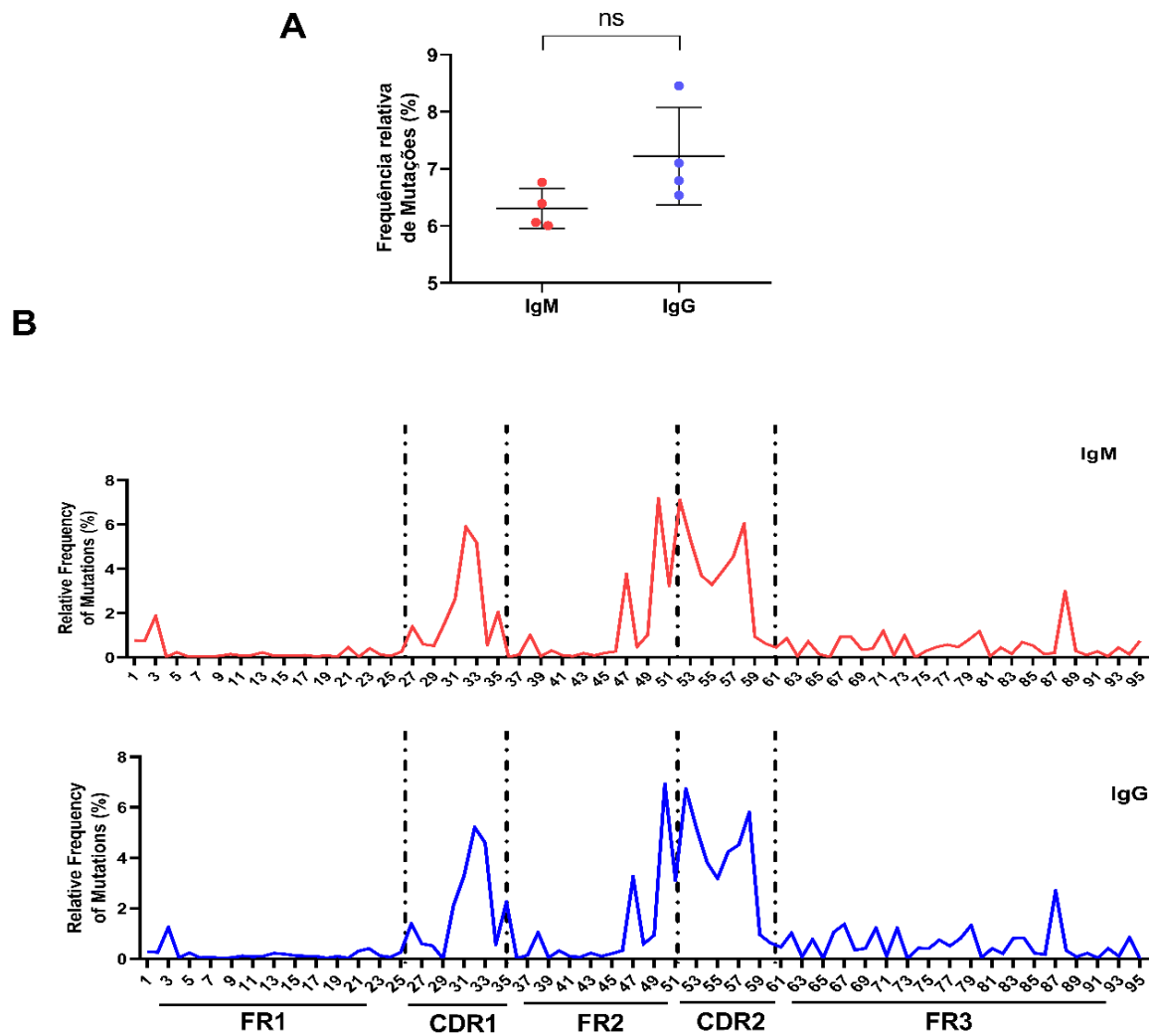


Figura 18. Características de Hipermutação Somática (SHM) da região variável de cadeia pesada de cavalo de isotipos IgG e IgM. (A) Média da frequência SHM (%) no segmento gênico IGHV dos isotipos IgM e IgG ($p < 0,05$). (B) O número de mutações por posição de aminoácidos no segmento do gene IGHV da cadeia pesada dos anticorpos de cavalos (de acordo com a numeração IMGT, sem *gaps*). As linhas pontilhadas delimitam as regiões FR e CDR. IgM é mostrado em vermelho e IgG em azul.

Tabela 6. Motivos de AID (RGYW/WRCY) e frequências de mutação nas regiões de CDR e FR

	IgM	IgG
Número de motivos RGYW/WRCY por segmento IGHV	19 (1,18)	20 (1,17)
% de nucleotídeos mutados nos CDRs	16,79	18,81
% de mutações nos CDR presentes em motivos RGYW/WRCY	51,61	45,85
% de nucleotídeos mutados nos FR	4,24	4,71
% de mutações nos FR presentes em motivos RGYW/WRCY	16,92	18,76
% de todos os nucleotídeos mutados	6,37	7,12
% de todas as mutações presentes nos motivos RGYW/WRCY	24,20	25,59

Também foram analisadas as mutações de nucleotídeos por posição presente na região IGHV para os dois isotipos estudados (Figura 19). Pode ser visto um pico na posição 9 que em sua maioria representa a mudança do nucleotídeo “A” por “G”, porém esse mesmo pico não foi observado na análise de aminoácidos, uma vez que essa alteração de nucleotídeos não causou alteração de aminoácidos, ou seja, é uma mutação sinônima.

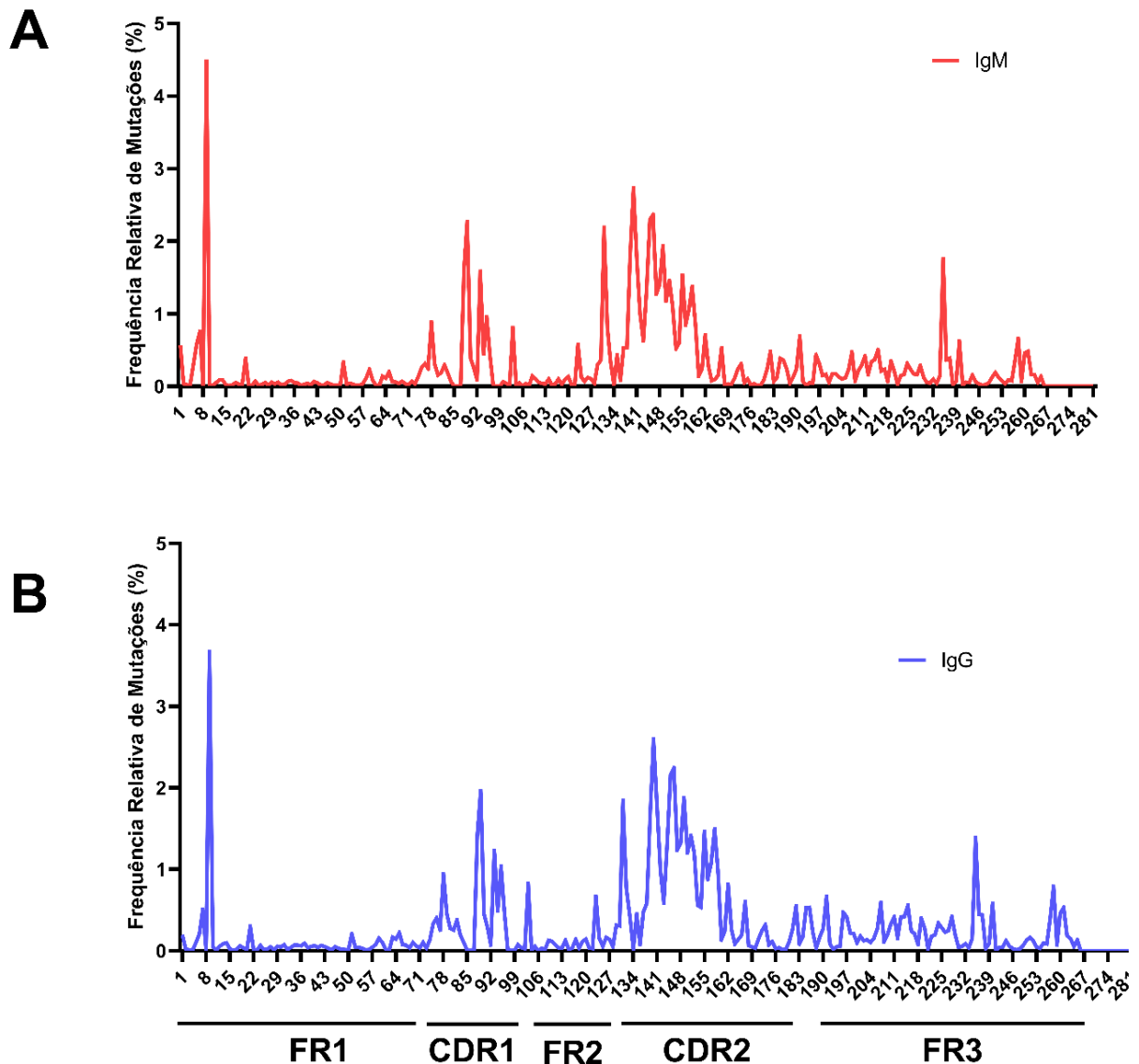


Figura 19. Características de hipermutação somática (SHM) de nucleotídeos da região variável de cadeia pesada. (A) Número de mutações por posição de nucleotídeo na região da cadeia pesada dos isotipos IgM e (B) IgG de quatro cavalos não imunizados (de acordo com a numeração IMGT, sem *gaps*).

4.1.6. Caracterização da Diversidade Juncional de Anticorpos de Cavalo

Uma fonte essencial de diversificação de anticorpos é a adição e deleção de nucleotídeos entre as junções V(D)J. Portanto, analisamos a ocorrência da adição de nucleotídeos P/N e corte de exonuclease para anticorpos IgM e IgG de cavalos. Observamos características muito semelhantes em todas as regiões juncionais dos dois isotipos estudados (Figuras 20A e 20B).

As médias de nucleotídeos nas regiões de junção N1 e N2 variam de 8,7 a 9,5 presentes em cerca de 92% dos clones de IgM e IgG (n = 992.677 clones para IgM, n = 264.566 clones para IgG) (Figura 22A e 22B, Tabela 7). Curiosamente, 43-44% dos anticorpos possuem junções N1 (variando de 10 a 111 nt) e N2 (variando de 10 a 104 nt), sendo que 7,4 a 10% dessas regiões possuem 22 ou mais nucleotídeos (Tabela 7). Também foi possível observar que metade dos 10 maiores CDR-H3 apresentam cisteínas. Também notamos que a região N1 é altamente enriquecida em guanina (G) (35,59%), e a região N2 é enriquecida em G e T (30%) para ambos os isotipos (Tabela 8). A composição de nucleotídeos nas outras regiões da junção é mostrada no Anexo III.

Tabela 7. Análise de adições de nucleotídeos em anticorpos de cavalo

	Número médio de nucleotídeos		Adição de clones (%)		
	Número de clones	N1 (intervalo)	N2 (intervalo)	N1 maiores que 10 nucleotídeos/ e maiores que 22 nt	N2 maiores que 10 nucleotídeos/ e maiores que 22 nt
IgM	998.756	9,54 (0-78)	9,71 (0-62)	43,95/7,4	43,74/8,6
IgG	264.566	8,72(0-111)	9,24 (0-104)	43,67/7,6	45,7/10,0

Tabela 8. Porcentagem de nucleotídeos presentes nas junções N1 e N2 de anticorpos IgM e IgG de cavalo

	%A	%T	%G	%C	
IgM	24,82	20,21	35,59	19,37	REGIÃO N1
IgG	24,28	21,47	34,90	19,35	
IgM	21,29	28,20	30,32	20,19	REGIÃO N2
IgG	21,57	29,60	28,87	19,96	

Da mesma forma, o corte de exonuclease foi observado em cerca de 70% a 97% dos clones de IG, com o maior número de nucleotídeos cortados nas extremidades da região 5' do segmento gênico IGHJ (média de 10 para IgM e 11 para IgG) (Figura 20A e 20B). Ao analisar os componentes que contribuem para o comprimento da CDR-H3, observamos que uma média

de 15 nucleotídeos do segmento gênico IGHD contribuem para o comprimento das CDR-H3s das IgM e IgG que contêm uma média de 54 nucleotídeos (em torno de 26-27% da contribuição do IGHD para o CDR-H3). Surpreendentemente, quando analisadas as cinco maiores CDR-H3 dos clones de IgG, observamos uma contribuição de 12 a 22 nucleotídeos de genes IGHD que representam apenas 9% a 12% dos CDR-H3. A maior contribuição nestes casos veio da adição de N1 que pode contribuir com 111 nucleotídeos de uma região de CDR-H3 com 150 nucleotídeos (74%) ou da adição de N2 que pode contribuir com 91 nucleotídeos de uma CDR-H3 contendo 129 nucleotídeos (70%) (Figura 20C).

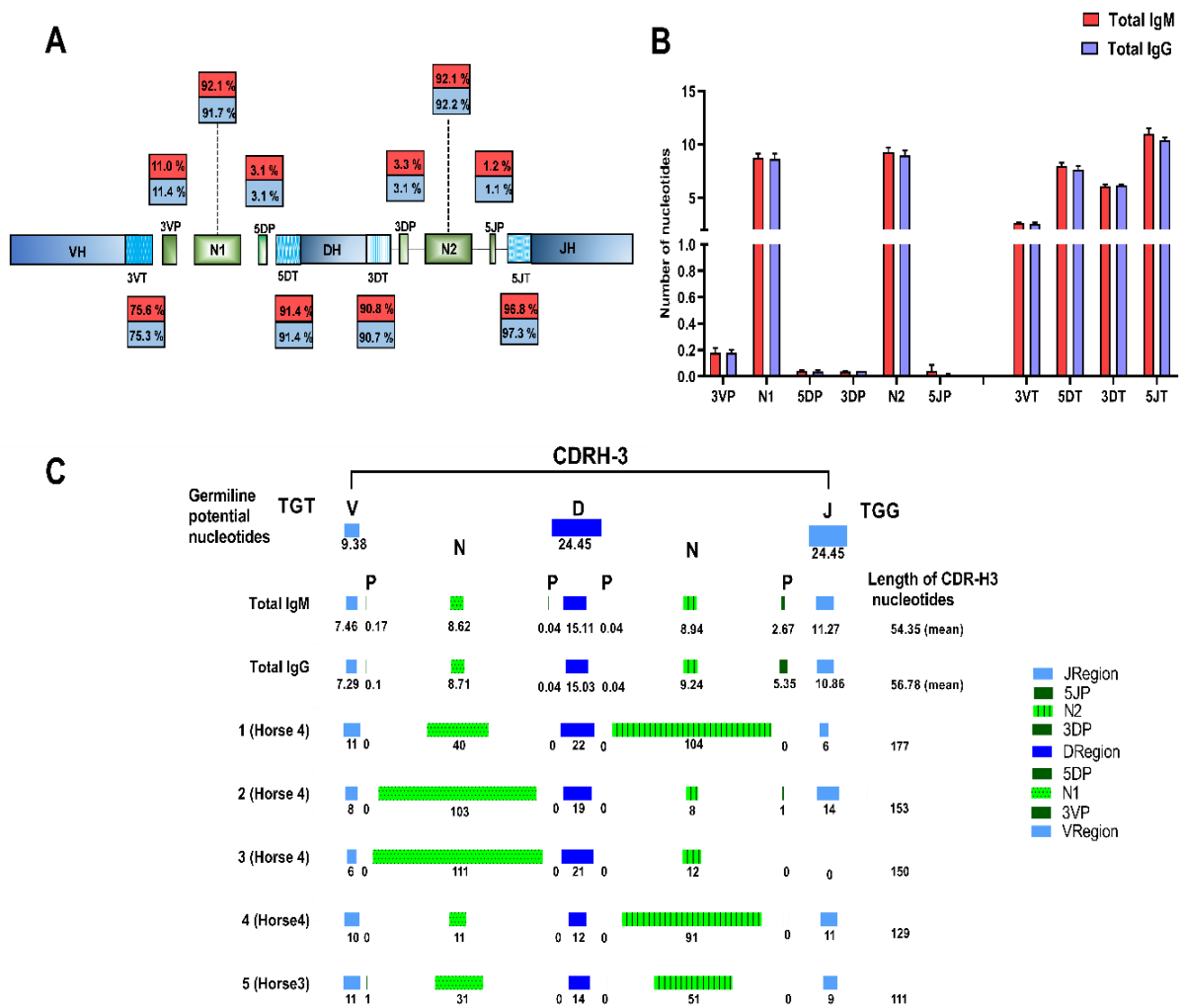


Figura 20. Análise da junção VDJ. (A) Esquema de diferentes tipos de alterações juncionais durante o rearranjo VDJ, mostrando suas localizações e ocorrência. Nos quadros estão a frequência de mutação (%) de cada segmento juncional em todos os clones de anticorpos analisados: 3VP e 3VT para a região 3'V, 5DP e 5DT para os genes 5'D, 3DP e 3DT para a região 3'D e 5JP e 5JT para genes 5'J, onde P significa adições de nucleotídeos palindrômicos e T significa trimming pela exonuclease; N1 as adições de nucleotídeos aleatórios sem molde nos genes 3'V e 5'D; N2 para adições de N nos genes 3'D e 5'J. (B) O número médio de nucleotídeos por região de junção adicionados ou removidos (C) Desconstrução dos componentes que contribuem para o comprimento do CDR-H3 em relação ao total dos clones

IgM e IgG, bem como os 5 maiores CDR-H3 dos clones de IgG. A média de nucleotídeos da sequência da linha germinativa do segmento gênico IGHV, junções P e N, o segmento gênico IGHD e o segmento gênico IGHJ para o comprimento CDR-H3 é ilustrada.

Visto a alta frequência e comprimento de adição de nucleotídeos N, resolvemos investigar a similaridade da enzima desoxinucleotidil-transferase terminal (TdT) de cavalos em comparação com outras espécies. Por isso, realizamos o alinhamento da sequência de TdT isoforma 1 e 2 de cavalo com TdT humana, de camundongo, porco e bovino. Essas enzimas são compostas principalmente de duas regiões, um núcleo catalítico composto pelos domínios dedo (finger), palma (palm) e polegar (thumb) no C-terminal e um domínio, na região N-terminal composto pela região C-terminal de BRCA1, denominado BRCT. Observamos a conservação dos ácidos aspárticos catalíticos e da alça substrato-específica 1 (Figura 21). A maior diferença observada foi na região interdomínio, região rica em prolinas.

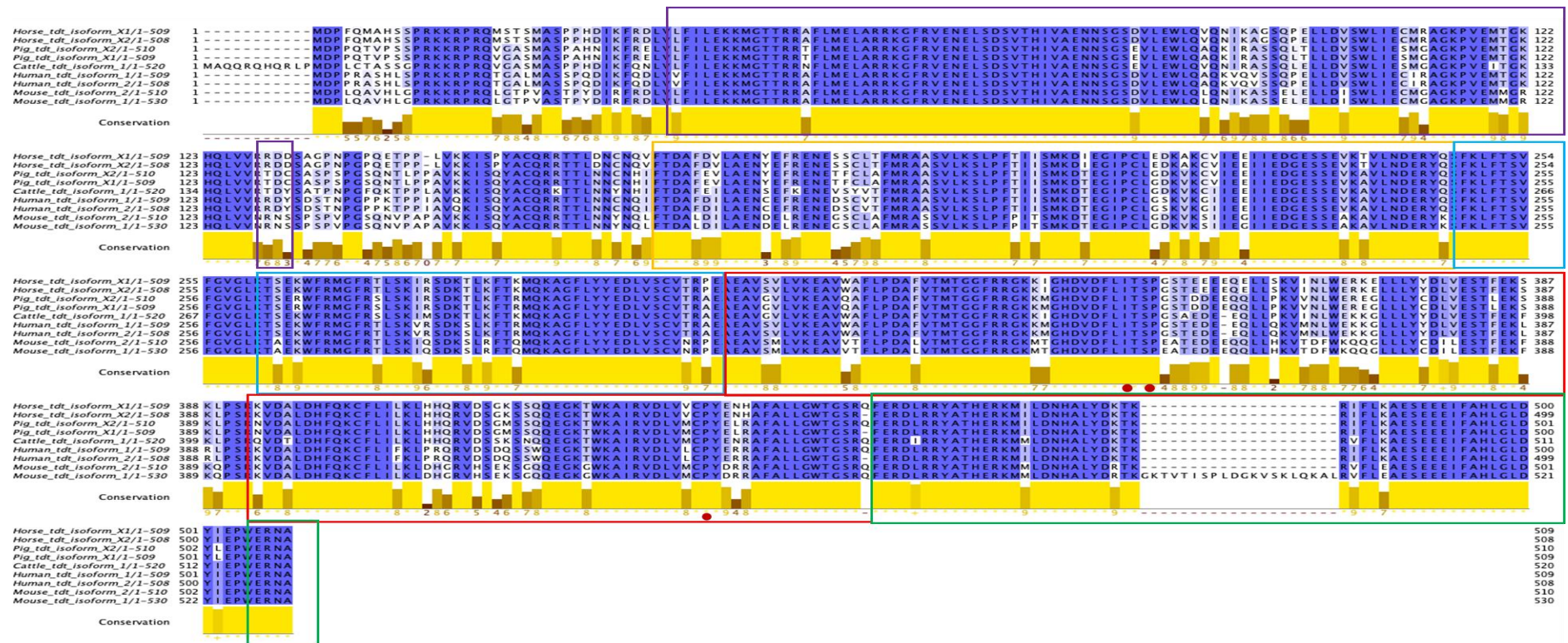


Figure 21. Alinhamento de sequência múltipla de TdT de diferentes espécies. Um alinhamento da sequência de aminoácidos das duas isoformas de TdT equina com outras TdT de vertebrados foi feito usando o programa Clustal Omega (Sievers et al., 2011) presente no pacote Jalview (Waterhouse et al., 2009) que também forneceu análise de conservação. As diferentes cores das caixas representam domínios TdT, domínio BRCT anotado por CDD (Lu et al., 2020) e núcleo catalítico TdT conforme descrito Delarue e colaboradores em 2002. : Roxo: domínio BRCA1 C Terminus (BRCT) (CL0459), amarelo: domínio hélice-gancho-hélice (HHH_8), azul: domínio dedos da DNA polimerase lambda, vermelho: DNA polimerase beta palma, destacando os 3 ácidos aspárticos catalíticos (vermelho círculo) e loop1, Verde: DNA polimerase beta polegar. Sequências de Referência NCBI: Horse_tdt_isoform_X1 (XP_005602408.1); Horse_tdt_isoform_X2 (XP_001501812.3); Pig_tdt_isoform_X2 (XP_003133204.1); Pig_tdt_isoform_X1 (XP_005671421.1); Cattle_tdt_isoform_1 (NP_803461.1); Human_tdt_isoform_1 (NP_004079.3); Human_tdt_isoform_2 (NP_001017520.1); Mouse_tdt_isoform_2 (NP_001036693.1); Mouse_tdt_isoform_1 (NP_033371.2)

4.1.7. Contribuição das adições de N na hidrofobicidade de CDR-H3

A hidrofobicidade do CDR-H3 é causada pelos aminoácidos que o compõem, sendo que esses aminoácidos são provenientes de parte do segmento gênico IGHV, todo o segmento gênico IGHD, parte do segmento gênico IGHJ e as adições P e N. A Figura 22A mostra a hidrofobicidade para os CDR-H3 total do repertório de IgG e IgM, calculado pelo índice Kyte Doolittle. Também comparamos os CDR-H3 que tinham menos de 10 nucleotídeos adicionados nas regiões N1 e N2 com aqueles que tinham N1 e N2 maiores que 10 nucleotídeos e a 22 nucleotídeos (Figura 22B e 22C) nas mesmas regiões, para avaliar se as adições da junção causaram alterações na hidrofobicidade do CDR-H3 para os isotipos IgM e IgG. Embora, observou-se que não houve diferença significativa entre os grupos analisados, percebe-se que no caso das adições de N1 e N2, os CDR-H3 do grupo IgG tendem a ser menos hidrofílicos que o grupo de IgM.

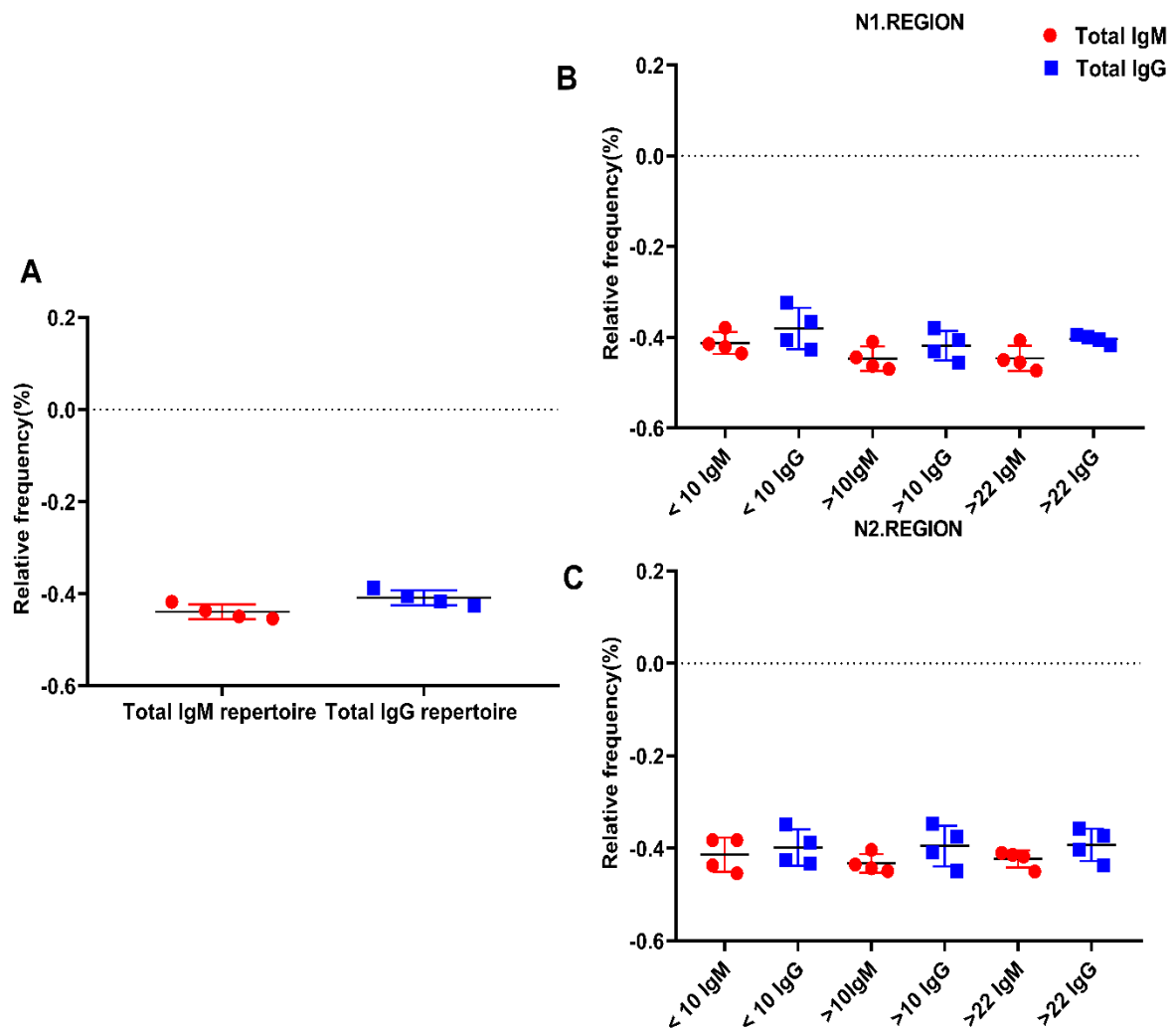


Figura 22. Índice de hidrofobicidade CDR-H3 repertório IgM e IgG. (A) Índice de hidrofobicidade CDR-H3 do repertório total de IgM (vermelho) e IgG (azul) calculado pelo índice Kyte Doolittle. (B) Comparação do índice de hidrofobicidade de CDR-H3 do repertório de IgM e IgG com adições de N1 inferiores a 10 nucleotídeos (< 10IgM/G), igual ou maior a 10 nucleotídeos (\Rightarrow 10IgM/G) e igual ou maior a 22 nucleotídeos (\Rightarrow 22IgM/G). (C) Comparação do índice de hidrofobicidade de CDR-H3 do repertório de IgM e IgG com adições de N2 inferiores a 10 nucleotídeos (< 10IgM/G), igual ou maior a 10 nucleotídeos (\Rightarrow 10IgM/G) e igual ou maior a 22 nucleotídeos (\Rightarrow 22IgM/G).

4.1.8. Análise do segmento gênico IGHD

Ao analisar o uso do segmento gênico IGHD, observamos uma preferência pelo uso da RF1 em mais de 80% dos anticorpos de cavalos (Figura 23), enriquecidos em aminoácidos polares como tirosina e glicina (Tabela 9). Dos 102 conjuntos possíveis de sequências de aminoácidos IGHD ($17 \times 6 = 102$), 36 (35,3%) incluem uma ou mais tirosinas, enquanto apenas 13 (12,7%) possuem uma ou duas argininas (Tabela 10).

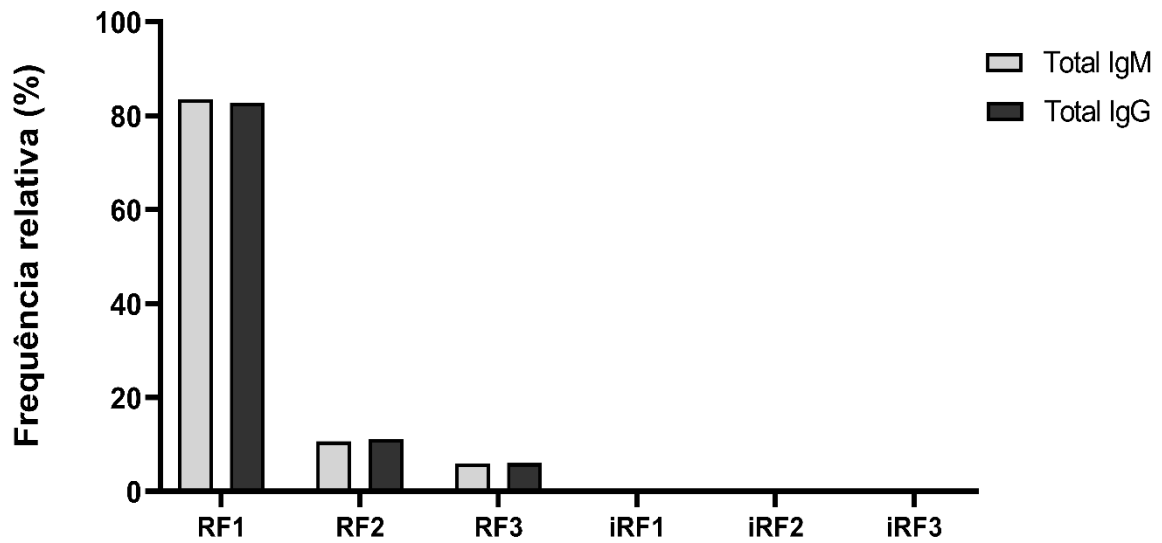


Figura 23. Preferência da fase de leitura (RF) para o segmento gênico IGHD em cavalos não imunizados. Frequência relativa das seis possíveis fases de leitura aberta para o segmento do gene IGHD. RF1, RF2 e RF3 são gerados por exclusão, enquanto iRF1, iRF2 e iRF3 são gerados por inversão.

Tabela 9. Percentual da composição de aminoácidos por fase de leitura do segmento gênico IGHD.

	RF1	RF2	RF3	iRF1	iRF2	iRF3	
D	5,22	0,75	0	0	0	0	Ácido
E	0,74	0	2,90	0	0	3,00	
R	0,74	1,50	4,34	5,97	0,74	0	Básico
K	0,74	0	0	0,74	0	0	
H	0	0	0,72	21,64	1,49	0	
F	0	0	0,72	0	0	0,75	Hidrofóbico
P	0	2,25	0	2,24	4,48	17,29	
V	0	27,82	0,72	0,74	29,10	3,00	
L	0	5,26	42,75	1,49	1,49	18,04	
I	2,23	6,77	1,44	0,74	18,66	3,00	
A	5,97	3,00	1,45	3,54	5,26	2,93	
W	3,73	0	13,76	0	0	0	
M	0	18,04	0	0	0	0,75	
N	5,97	6,66	0,72	15,67	2,99	0	Neutral
Q	0	0	2,90	1,49	0	4,51	
C	0,74	0	5,80	5,97	0	2,94	Polar
S	14,92	0	0	23,88	4,48	6,77	
G	18,65	4,5	1,45	0	5,97	1,50	
Y	39,55	0	2,90	11,94	0,74	2,25	
T	0,74	29,32	0	0,74	26,12	1,50	
*	0	0	17,39	4,47	0,74	32,33	
AA Media de RF	7,88	7,82	8,11	7,88	7,88	7,82	

Tabela 10. Composição de Aminoácidos por fase de leitura de segmentos gênicos funcionais IGHD em anticorpos de cavalo do IMGT

Grupo	Gene	RF1	RF2	RF3	iRF1	iRF2	iRF3
	IGHD2-6	GYYSRSCYT	MVTIVGVAI	WLL**ELLY	YSNSYYSNH	GIATPTIVT	V*QLLL**P
	IGHD2-9	GYyasgydy	MVTMLVVMtt	WLLC*WL*L	CSNH*HSNH	V VITTSIVT	*S*PLA**P
	IGHD2-12	YSYGSYYA	TIVMVVTM	L*LW*LLC	HSNYHNYS	GIVTTITI	A*LP*L*
	IGHD2-15	GYyGSySSyYA	MVTMVVTTVVTM	WLLW*LLQ*LLC	HSNYCSNYHSNH	GIVTTVVTTIVT	A*LL**LP**P
IGHD2	IGHD2-18	GYyAGSYyA	MVTMLVVTM	WLLCW*LLC	HSNYQHSNH	GIVTTSIVT	A*LPA**P
	IGHD2-23	DYYGISDSY	MITMVLVTPT	*LLWY**LL	CRSH*YHSNH	VGVTNTIVI	*ESLIP**S
	IGHD2-26	YGYGGAYYY	TMVMVVLTTT	LWLWWCLLL	CSSKHHNHS	VVVSTTITI	***APP**P*
	IGHD2-29	SYyGGSSWYS	TVTMMVVVPGT	QLLWW*FLVL	STRNYHHSNC	GVPGTTTIVT	EYQELPP**L
	IGHD2-36	DYYGAIDYI	MIIMVLLTT*	*LLWCY*LHN	LCSQ*HHNNH	YVVNSTIII	VM*SIAP**S
	IGHD4-2	YYGwGN	TMAGV	LLWLG*	YPSHS	VTPAIV	LPQP**
	IGHD4-5	YYNYNY	TTTTT	LLQL*L	SYSCS	VVIVVV	*L*L*
IGHD4	IGHD4-11	NYGYGYA	TTVMVML	*LRLWLCY	VA*P*P*L	*HNHNRS	SITITVV
	IGHD4-31	YDDGYYN	TMTDTT	LR*RILQ	CSIRHR	VVVSIV	L*YPSS*
	IGHD4-35	NYGSYNY	TMAPIIT	*LWLL*LL	SNYRSHS	VVIIGAIV	**L*EP*L
	IGHD4-38	EKSWSN	RRVGV	GEELE*	YSNSS	VTPTLL	LLQLFS
IGHD6	IGHD6-34	YGSGW	TVAVG	LR*RLA	ANRYR	PTATV	GQPLP*
	IGHD6-44	YGSGW	TVVVG	IR*WLA	ANHYR	PTTTV	GQPLPY

4.1.9. Subclasses de IgG

A IgG desempenha todas as funções de uma molécula de imunoglobulina típica, como ligação ao antígeno, fixação do complemento e ligação a várias células, como células fagocíticas, linfócitos, plaquetas, mastócitos e basófilos. Existem sete subclasses de região constante de IgG equina que constitui o repertório atual de IgG. Destes IGHG4 e IGHG7 possuem aproximadamente 96% de identidade, portanto neste estudo serão analisados em conjunto. (WAGNER et al., 2006).

Neste trabalho, analisamos as subclasses de IgG e observamos uma maior frequência de IgG1, em cerca de 54% no repertório avaliado, que foi seguida pela IgG3 com pouco mais de 24%. As outras subclasses foram encontradas em menos de 10%, exceto a IgG5 que não foi observada em nenhuma sequência (Figura 24).

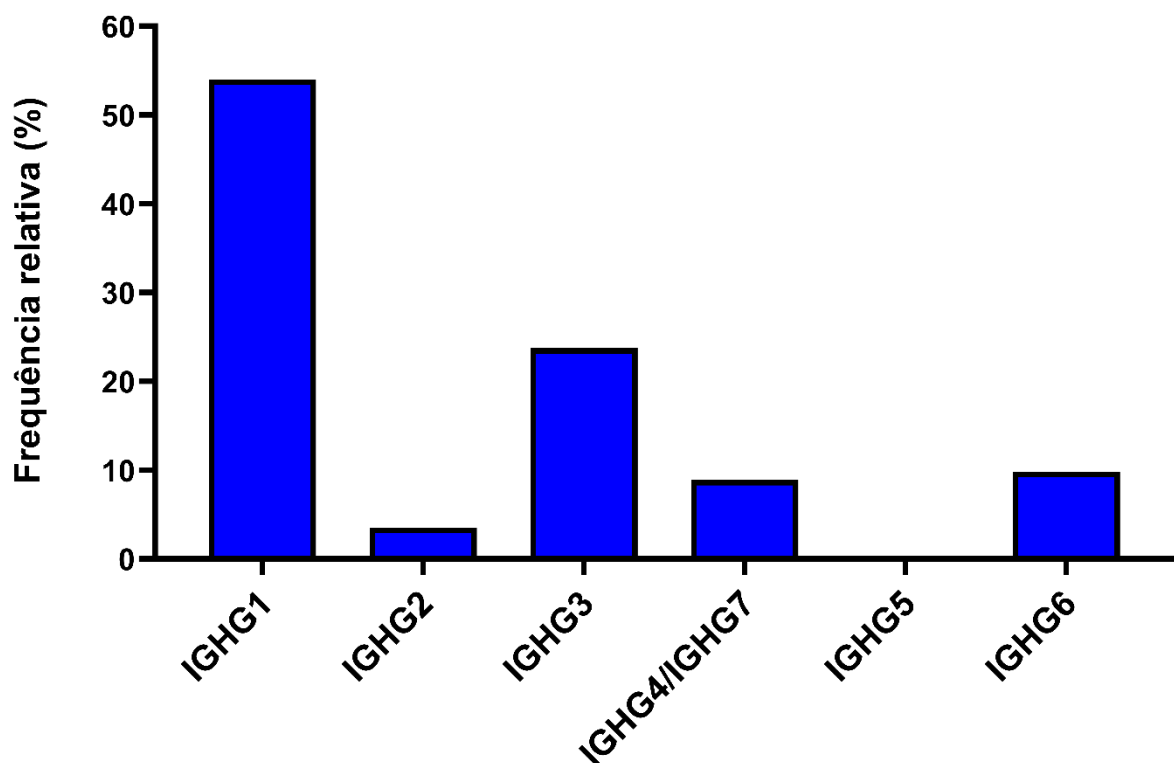


Figura 24. Frequência de Subclasses de IgG. A média da frequência relativa das 7 diferentes subclasses de IgG para os 4 cavalos analisados.

Os dados apresentados até este momento, sobre a análises do repertório natural de cavalos não imunizados foram submetidos à publicação na revista *Molecular Immunology*. O manuscrito do artigo encontra-se no ANEXO IV.

4.2. Comparação do repertório de anticorpos de equinos antes e após imunização com veneno de aranha *Loxosceles*.

4.2.1. Obtenção das cadeias pesadas de anticorpos equinos após imunização com veneno de *Loxosceles*

Com o objetivo de se conhecer a composição e diversidade de anticorpos equinos após a imunização de cavalos com veneno de aranhas do gênero *Loxosceles*, a primeira etapa deste projeto visou a amplificação das cadeias pesadas de anticorpos.

A amplificação das cadeias pesadas dos anticorpos de cavalo foi realizada por PCR multiplex. A Figura 25 mostra o resultado dessa amplificação, onde é possível observar uma banda de 500 pb conforme esperado para o *amplicon* da cadeia pesada (IGH).

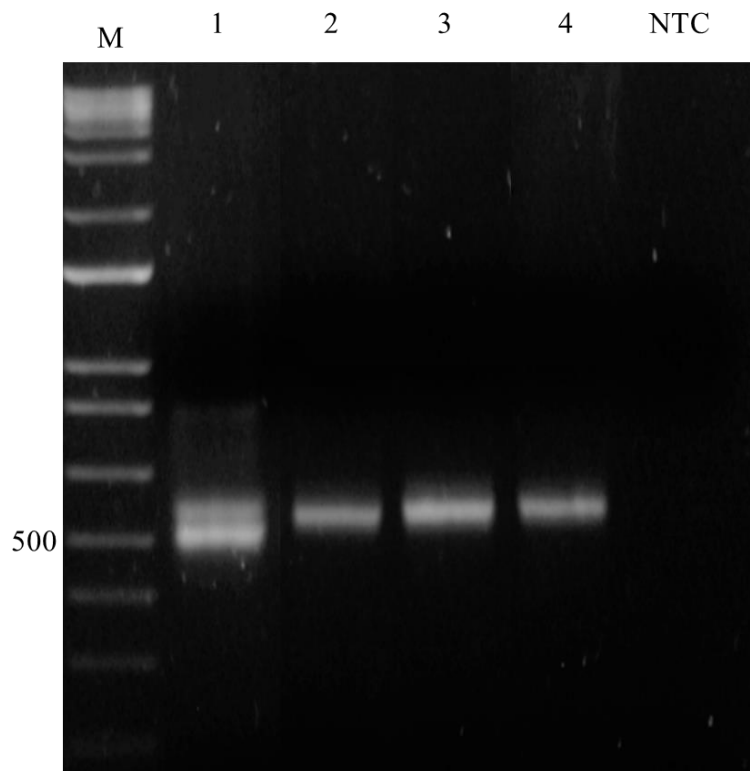


Figura 25. Gel de agarose 1% dos amplicons de cadeias pesadas de anticorpos provenientes de PBMCs de cavalos imunizados com veneno de *Loxosceles*. Amostras representam a amplificação das cadeias pesadas dos anticorpos presentes em PBMCs de cavalos imunizados

com o *pool* de venenos das aranhas *L. Intermedia*, *L. gaucho* e *L. laeta*. NTC corresponde ao controle negativo realizado com todos os *primers* F e R, e reagentes da PCR, mas sem o cDNA de PBMCs dos cavalos. Reações individuais com os *primers reverse* IgG e IgA (IgG/IgA) da cadeia pesada para amostra de cavalo 1 imune (1), 2 imune (2), 3 imune (3) e 4 imune (4).

4.2.2. Análise da biblioteca de amplicons de IGH para o sequenciamento do repertório de anticorpos equinos imunizados com veneno de *Loxosceles*

A biblioteca de *amplicons* de cadeia pesada de imunoglobulina preparada para o sequenciamento foi quantificada por *qubit* (concentração dada em ng/ μ L e convertida em mM). Observamos uma variação de 1,9 a 41,9 ng/ μ L e de 5,3 a 115 mM (Tabela 11), concentrações suficientes para realização do sequenciamento.

Tabela 11. Concentração da biblioteca de sequenciamento obtida por *qubit* em ng/ μ L e mM.

Amostra	Qubit (ng/uL)	mM
1H IgG/A	41,9	115,0
2H IgG/A	9,7	24,9
3H IgG/A	1,9	5,3
4H IgG/A	10,4	26,7

4.2.3. Análise dos sequenciamentos de IGH de cavalos imunizados

Para conhecer as características dos anticorpos de cavalos após imunização com veneno de aranhas do gênero *Loxosceles* realizamos o sequenciamento da cadeia pesada de seus anticorpos. Os dados brutos foram processados pelo *pipeline* pRESTO obtendo-se entre 437.667 a 1.445.092 sequências únicas de IGH. Aproximadamente 68% destas foram completamente anotadas, ou seja, os segmentos V(D)J foram identificados nas sequências. As demais sequências apresentaram anotação parcial dos segmentos, principalmente na região J. Assim, apenas as sequências completas foram utilizadas nas análises posteriores.

De forma a retirar a redundância dos clones de células B do conjunto de dados, uma análise de clonalidade foi realizada, utilizando a ferramenta Yclon, nas sequências completamente anotadas. Obtivemos então, entre 85.450 e 151.948 clones únicos neste conjunto de dados (Tabela 12).

Tabela 12. Resumo dos dados de sequenciamento MiSeq. Para IGH de cavalos imunizados

Amostra	Dado Bruto	Sequencias			
		Pré-processadas	Anotadas	Produtivas	Clones
1	437.667	394.526	345.049	317.894	90.415
2	1.445.092	829.657	714.791	653.022	131.922
3	525.890	394.399	374.399	340.916	85.450
4	1.163.364	708.126	708.106	612.728	151.948

Quando analisamos a presença de clones raros no sequenciamento de IGH de cavalos imunes, através da curva de rarefação, observamos a presença de clones raros nestas amostras, de uma forma semelhante àquela observada em animais não imunes (Figuras 26A e 26B), visto que as curvas de rarefação não começaram a se estabilizar (atingir o platô), indicando que era improvável que capturássemos toda a diversidade dessa população.

No entanto, conseguimos capturar aproximadamente 80% do repertório de anticorpos em ambos os casos (Tabela 13). Importante ressaltar que os anticorpos dos cavalos não imunes foram clusterizados da mesma forma que os anticorpos dos cavalos imunes. Ainda, nos cavalos imunizados foram calculados o número de clones expandidos, assim, o número de clones que obtivemos para cada cavalo, antes e após a imunização, bem como os clones expandidos são mostrados na Tabela 12. Ainda, não observamos nenhuma diferença estatística entre a diversidade de anticorpos de cavalos não imunes e imunes ao comparar os índices de diversidade de Shannon (Figura 26C) e de Simpson (Figura 26D).

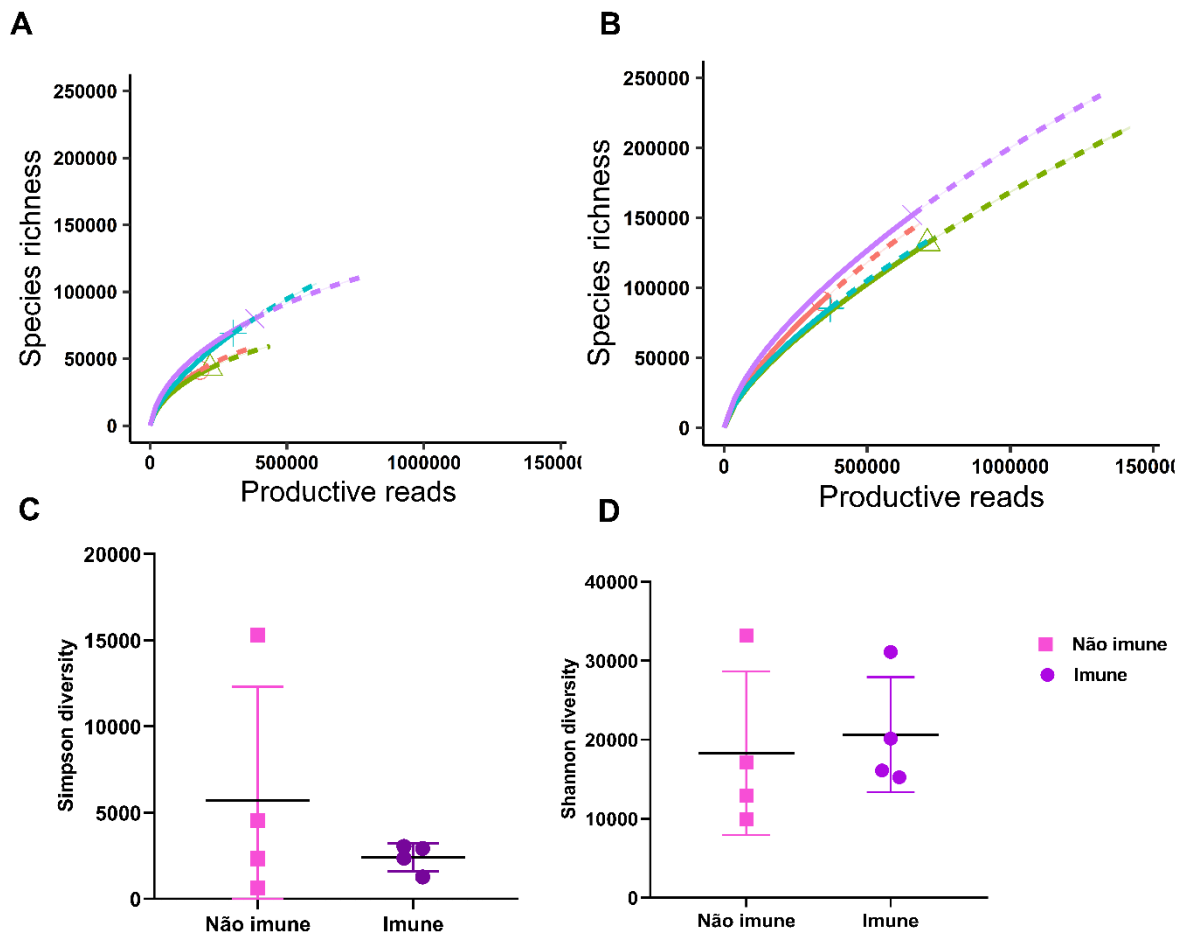


Figura 26. Comparação da rarefação e diversidade de IGH entre cavalos não imunes e imunizados com veneno de aranha do gênero *Loxosceles*. (A) Análise de rarefação de equinos não imunizados. (B) Análise de rarefação em equinos após imunização com veneno de aranha *loxosceles*. (C) Comparação da diversidade de anticorpos utilizando o índice de Simpson entre cavalos não imunes e cavalos imunizados com veneno de aranha do gênero *Loxosceles*. (D) Comparação da diversidade de anticorpos utilizando o índice de Shannon entre cavalos não imunes e cavalos imunizados com veneno de aranha do gênero *Loxosceles*.

Tabela 13. Características de rarefação e diversidade do repertório de anticorpo de cavalos antes e depois da imunização com veneno de aranha do gênero *loxosceles*

Amostra	Tipo	Experimental	Extrapolado	Expandido	Cobertura
1	Imune	90.415	301.211	907	0,8117
2	Imune	131.922	489.226	131	0,8672
3	Imune	85.450	289.945	236	0,8392
4	Imune	151.948	452.919	213	0,8468
1	Não imune	41.354	78.288	-	0,8810
2	Não imune	42.923	78.926	-	0,8994
3	Não imune	69.112	183.070	-	0,8525
4	Não imune	79.944	149.072	-	0,8944

4.2.4. Comparação do uso gênico de segmentos gênicos V (D) e J entre cavalos não imunes e imunes

A frequência de uso do segmento gênico V (D) e J dos cavalos não imunes foi comparada com a frequência gênica correspondente dos cavalos imunes e dos clones expandidos dos cavalos imunes. Esses resultados são mostrados na Figura 27, onde pode-se observar que a frequência gênica dos anticorpos dos animais não imunes, imunes e dos clones expandidos é muito semelhante. Apenas os segmentos gênicos IGHV4-37N, IGHV4-65 e IGHV4-37 são mais frequentes nos clones expandidos quando comparados aos demais grupos estudados (Figura 27A). Observamos também que o segmento gênico IGHD2-26 é o segmento gênico mais utilizado nos anticorpos dos 3 grupos analisados (Figura 27B), assim como o IGHJJ6-1 (Figura 27C).

Combinações V(D)J únicas foram analisadas para os animais não imunes, imunes e expandidos. Os três mais frequentes segmentos recombinados foram similares para os animais não imune, imunes e expandido, IGHV4-21, IGHD2-26 e IGHJ6-1, com frequência de 1,8% para não imune, 2,35% para imunes e 3,55% para os expandidos. As três combinações de segmentos de genes mais frequentes para cada grupo estão representadas na Tabela 14.

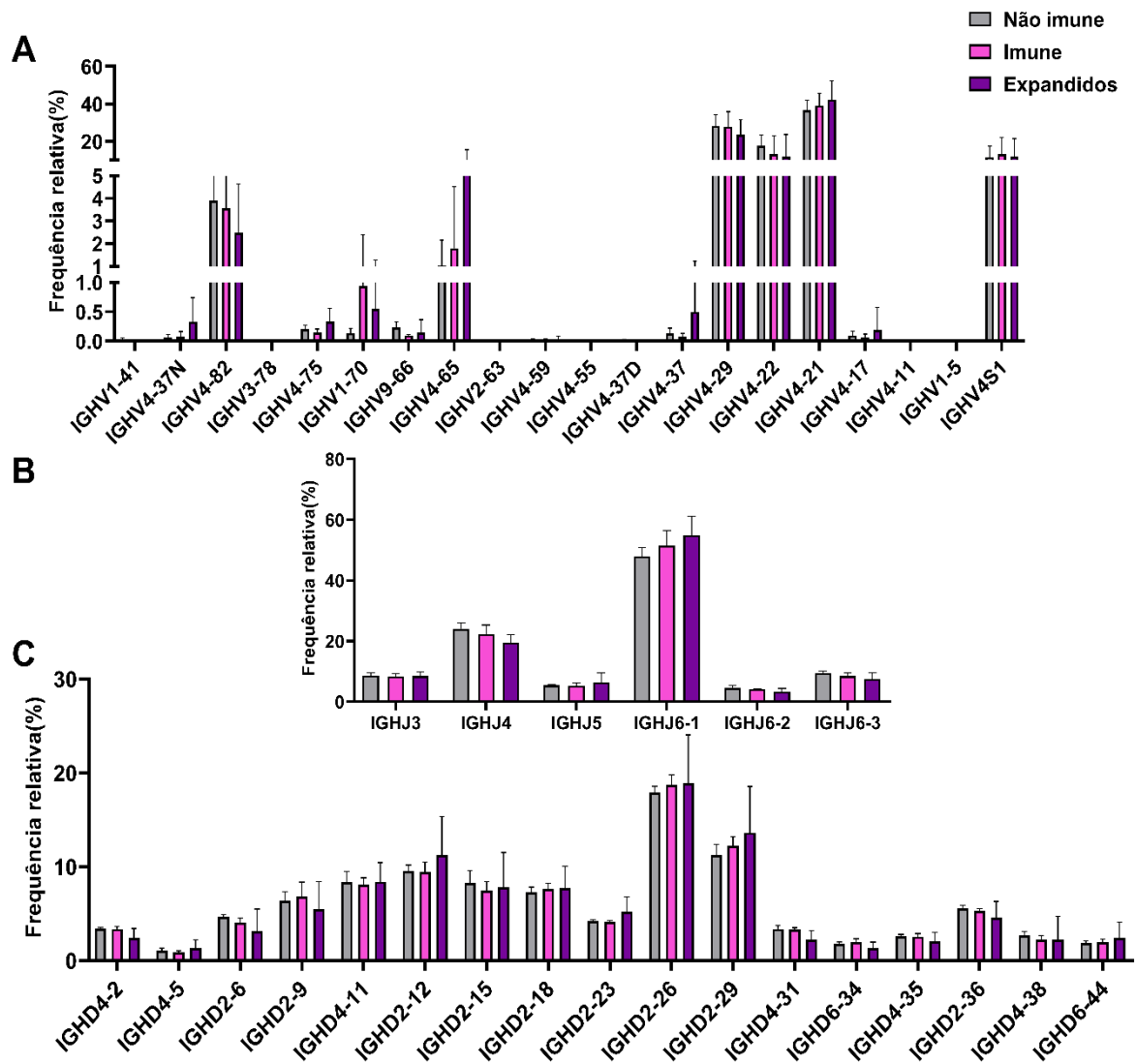


Figura 27: Comparação da frequência gênica de anticorpos de cavalos não imunes, imunes e de clones expandidos. (A) Comparação da frequência do segmento gênico IGHV em cavalos não imunes, imunes e clones expandidos. (B) Comparação da frequência do segmento gênico IGHJ. (C) Comparação da frequência do segmento gênico IGHD.

Tabela 14. Combinações de segmentos de genes mais frequentes para cada grupo estudado

Grupo de Amostras	Combinação 1	% ± SD	Combinação 2	% ± SD	Combinação 3	% ± SD
Não imunes	IGHV4-21	1,8 (0,31)	IGHV4-29	1,52	IGHV4-21	1,24
	IGHD2-26		IGHD2-26	(0,17)	IGHD2-29	(0,14)
	IGHJ6-1		IGHJ6-1		IGHJ6-1	
Imunes	IGHV4-21	2,35	IGHV4-21	1,52	IGHV4-29	1,52
	IGHD2-26	(0,45)	IGHD2-29	(0,30)	IGHD2-29	(0,23)
	IGHJ6-1		IGHJ6-1		IGHJ6-1	
Expandidos	IGHV4-21	3,55	IGHV4-29	3,11	IGHV4-21	2,29
	IGHD2-26	(1,81)	IGHD2-26	(1,93)	IGHD2-9	(0,78)
	IGHJ6-1		IGHJ6-1		IGHJ6-1	

Porcentagem e (desvio padrão)

4.2.5. Tamanho e conteúdo de aminoácidos na região de CDR-H3

Uma comparação da distribuição do tamanho da região de CDR-H3 de cavalos não imunes, imunes e dos clones expandidos mostrou que nos 3 grupos estudados, o tamanho médio da CDR-H3 foi semelhante, sendo 15 aminoácidos para o grupo de imunes e expandidos e de 14 para não imunes. É importante destacar que os clones expandidos apresentaram maior percentual de CDR-H3 com 5 aminoácidos (7,27%) quando comparado aos não imunes e imunes (3,2 e 2,4% respectivamente). Da mesma forma, ocorre com CDR-H3 de 24 aminoácidos onde o percentual de expandidos (1,40%) representa aproximadamente 2,6 vezes mais que os demais grupos estudados (média 0,48%) (Figura 28A).

A composição de aminoácidos da região CDR-H3 foi similar nos grupos estudados. Os cinco aminoácidos mais frequentes nos cavalos foram Tirosina (Y), Glicina (G), Alanina (A) e Serina (S) (Figura 28B), apresentando uma frequência em torno de 17,8%, 15,2%, 9,5% e 8,2% de Y, G, A e S, respectivamente.

A hidrofobicidade média do CDR-H3 também foi analisada para cada grupo estudado (Figura 28C), e pode-se observar que o grupo expandido apresenta anticorpos com média de índice de hidrofobicidade estatisticamente inferior (-0,4464) quando comparado o grupo imune (-0,3864) ($p < 0,05$), os demais grupos estudados não apresentaram diferença estatística (Figura 28C).

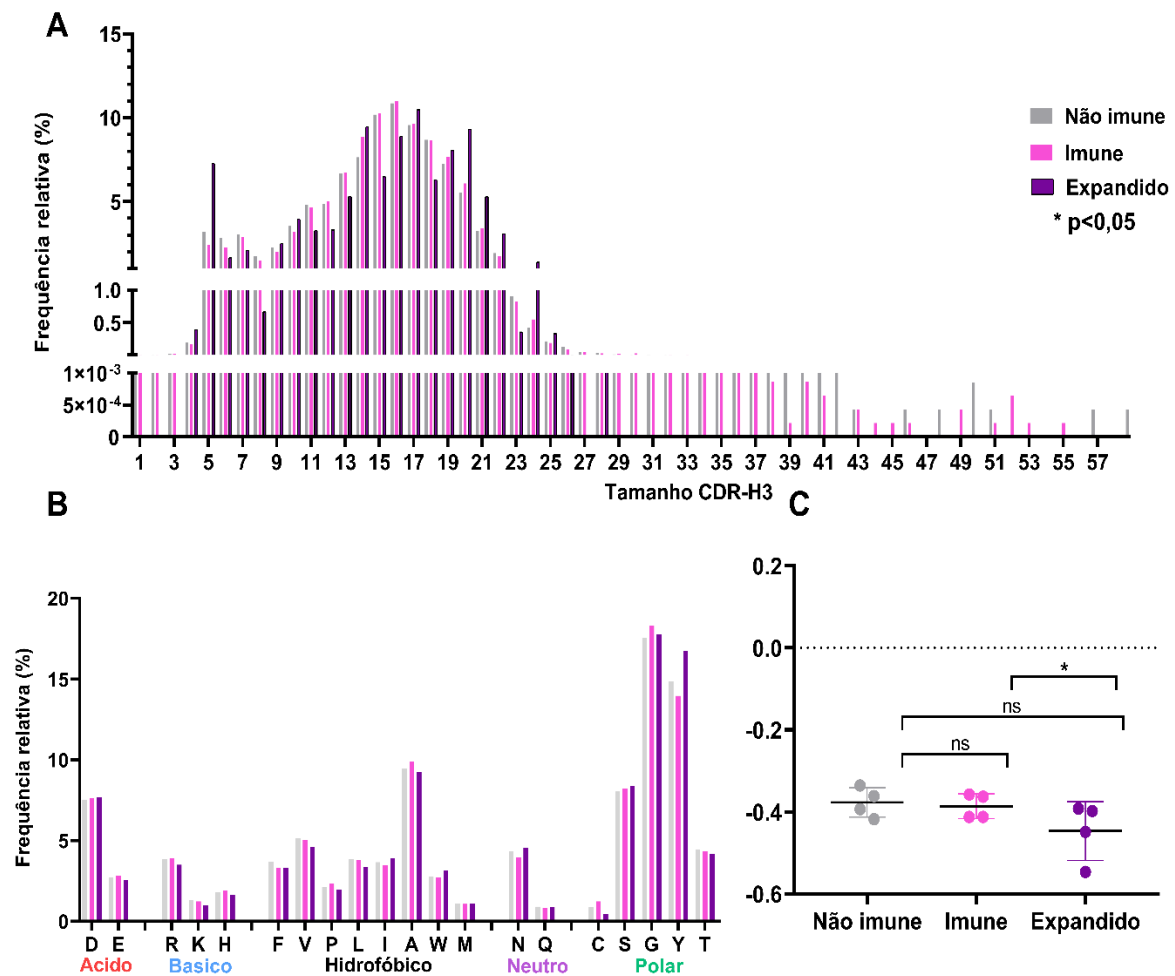


Figura 28. Características do CDR-H3. (A) Comparação do tamanho da região CDR-H3 dos repertórios dos avalos não imunes, imunes e de clones expandidos. (B) Composição dos aminoácidos da região CDR-H3 nas posições 105-117. (C) Comparação do índice de Hidrofobicidade CDR-H3. Os valores de foram calculados usando o teste de Fisher. Foram encontrados valores de p significativos para a comparação do grupo hiperimune com o grupo expandido ($p = 0,04$), para o restante das comparações não houve diferença significativa.

4.2.6. Análise de Hipermutação somática

Foram comparadas as hipermutações somáticas dos anticorpos dos cavalos antes e após a imunização com veneno de *Loxosceles*, assim como o dos clones expandidos, onde se observa que não houve diferença significativa na frequência de mutação dos anticorpos nos grupos estudados (Figura 29 A). Ainda, o perfil de mutação de aminoácidos ao longo da região IGHV desses grupos também mostrou um perfil semelhante entre os grupos, com picos de mutação nas regiões de CDR (Figura 29 B, C e D).

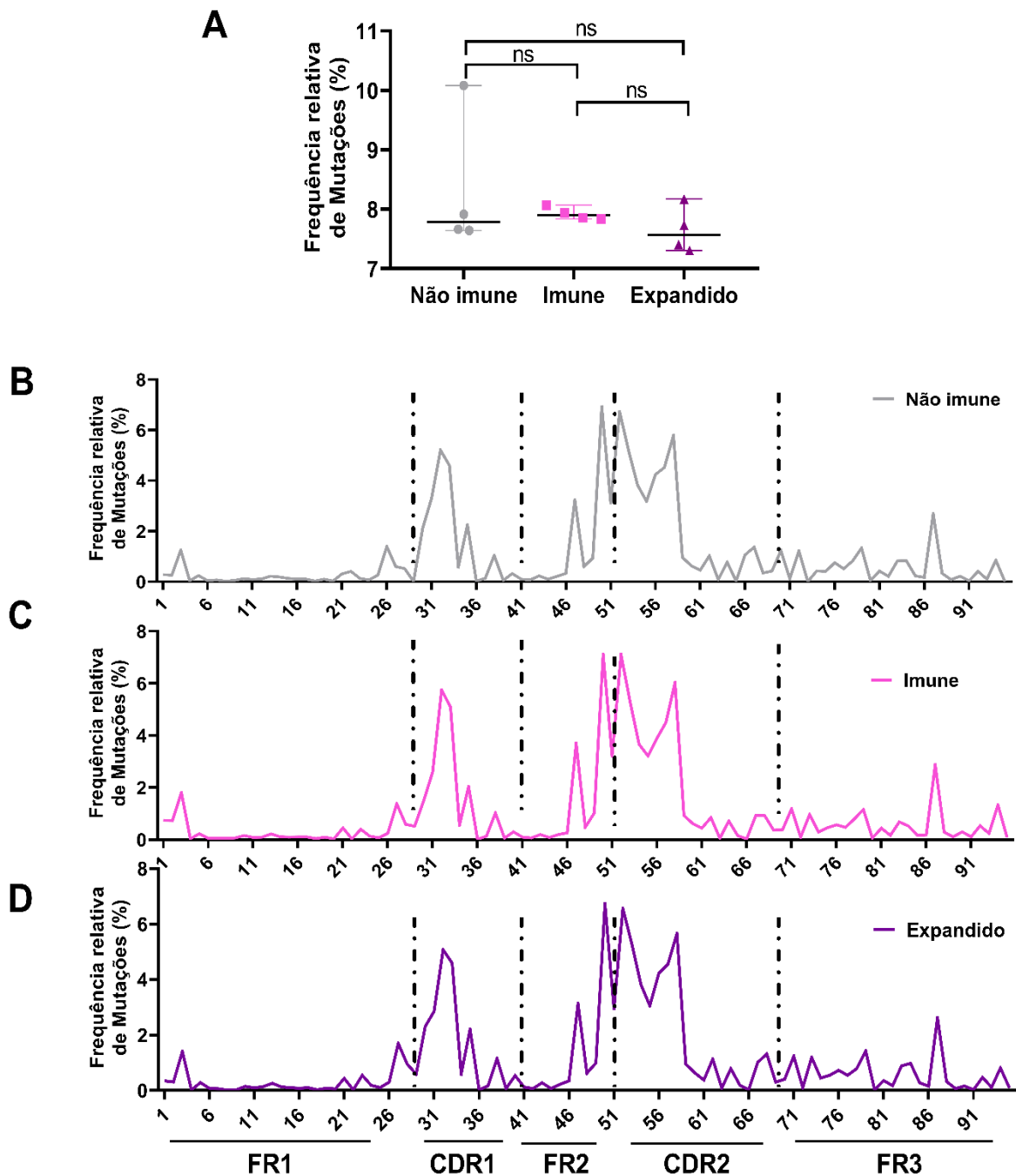


Figura 29. Avaliação da frequência de hipermutação somática (SHM) nos cavalos não imunes, imunizados com o *pool* de venenos das aranhas do gênero *Loxosceles* e de clones expandidos. (A) Comparação da média de SHM dos anticorpos de cavalos antes e após a imunização e de clones expandidos. (B) Perfil de mutação de aminoácidos na região IGHV de anticorpos em cavalos não imunizados. (C) Perfil de mutação de aminoácidos na região IGHV de anticorpos em cavalos imunes. (D) Perfil de mutação de aminoácidos na região IGHV de clones expandidos.

Além de analisar o perfil de mutação de aminoácidos na região IGHV, também analisamos o perfil de mutação de nucleotídeos nesta região nos três grupos estudados, e embora o perfil entre eles seja semelhante, encontrando mutações mais frequentes no CDR do que no FR, é importante destacar uma alta frequência de mutação na posição 9 que corresponde principalmente à mutação de “A” por “G”, e uma mutação sinônima, pois não é observada no perfil de mutação de aminoácidos (Figura 30).

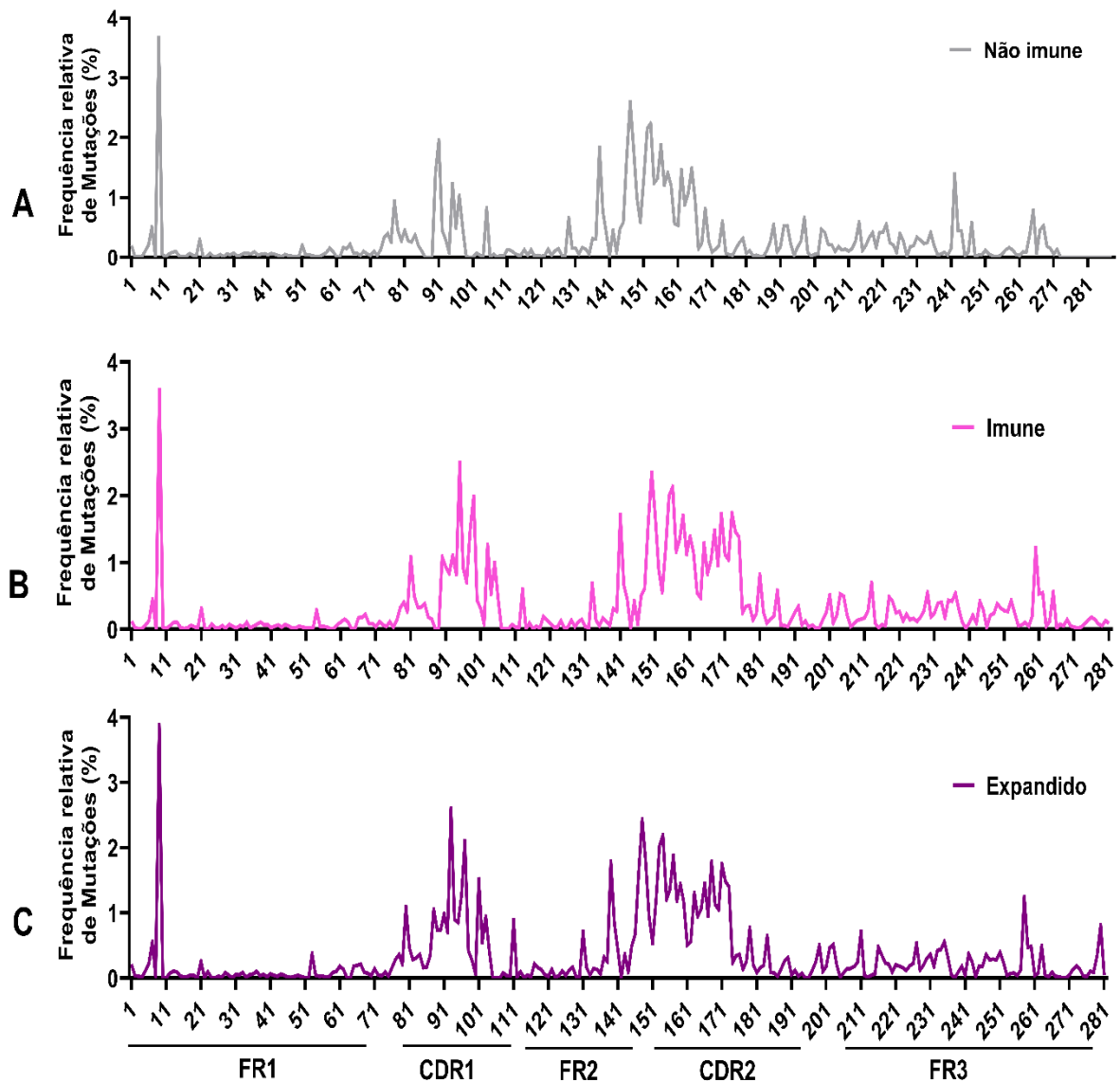


Figura 30. Perfil de mutação de nucleotídeos na região IGHV. (A) Perfil de mutação de nucleotídeos na região IGHV de cavalos não imunizados. (B) Perfil de mutação nucleotídica na região IGHV de cavalos imunizados. (C) Perfil de mutação de nucleotídeos na região IGHV de clones expandidos.

Outra análise importante é a comparação dos motivos de AID nos anticorpos dos três grupos estudados (Tabela 15) onde se observa que a média desses motivos é um pouco menor nos equinos imunes e expandidos (19) do que nos não imunes (20). Interessante notar que a porcentagem de mutações que ocorrem nos motivos da AID nos CDRs é bem maior nos grupos imune (61,64%) e expandido (58,90) do que nos não imunes (45,85).

Tabela 15. Comparação dos motivos da AID nos três grupos estudados.

	Não Imune	Imune	Expandido
Número de motivos RGYW/ WRCY por segmento IGHV	20 (1,17)	19 (1,17)	19 (1,30)
% de nucleotídeos mutados no CDR	18,81	20,56	19,45
% de mutações CDR presentes em motivos RGYW/WRCY	45,85	61,64	58,90
% de nucleotídeos mutados no FR	4,71	5,34	5,06
% de mutações FR presentes em motivos RGYW/WRCY	18,76	21,24	20,05
% de todos os nucleotídeos mutados	7,12	7,94	7,52
% de todas as mutações presentes nos motivos RGYW/WRCY	25,59	29,92	28,33

4.2.7. Comparação de diversidade juncional

Em relação à diversidade juncional, foi analisada a frequência com que esse tipo de diversidade apareceu nas diferentes sequências de cada grupo estudado. Pudemos observar que as regiões 3DP, 5VP e 3VP aparecem com maior frequência nos clones expandidos (Tabela 15). Em relação às regiões N1, N2, 5DP e 5JP nos três grupos, a frequência é semelhante.

Quando analisamos a atividade da exonuclease nos diferentes segmentos gênicos, observamos que esta é bastante frequente nos três grupos estudados (Tabela 16).

Tabela 16. Frequência dos diferentes tipos de junção nos repertórios estudados

	Não imune	Imune	Expandido
3VP	10,18 (1,02)	10,76 (0,98)	15,70 (3,05)
N1	89,83 (1,50)	90,90 (2,09)	89,56 (3,58)
5DP	2,65 (0,40)	2,44 (0,51)	1,95 (2,16)
3DP	2,90 (0,15)	3,01 (0,12)	3,78 (1,15)
N2	90,84 (1,31)	92,01 (1,63)	91,39 (1,80)
5JP	0,83 (0,22)	0,73 (0,88)	0,94 (0,19)
3VT	76,80 (1,65)	75,60 (1,62)	71,07 (2,39)
5DT	91,36 (0,42)	91,69 (0,46)	92,39 (3,22)
3DT	90,85 (0,58)	90,77 (0,16)	90,35 (1,99)
5JT	97,72 (0,42)	97,99 (0,22)	97,96 (0,81)

Também analisamos o número de nucleotídeos em cada uma das regiões da junção e observamos que este é semelhante nos três grupos estudados, estando entre 8,61 e 9,68 para N1 e de 9,56 a 10,28 para N2, representando estes as maiores adições nas diferentes regiões da junção. Em relação à deleção de nucleotídeos, pudemos observar que o tamanho é semelhante nos três grupos estudados, e neste caso, a região 5J foi a que apresentou maior corte de nucleotídeos, em torno, de 10,85 a 11,70 (Figura 31A).

Realizamos uma desconstrução do CDR-H3 para determinar a contribuição média dos nucleotídeos de cada região e compará-la entre os três grupos estudados (Figura 31B). Pudemos observar que a maior contribuição no CDR-H3 vem do segmento gênico IGHD e que isso é semelhante nos três grupos. Ainda, foi observado que os clones expandidos possuem maior número de nucleotídeos provenientes do segmento gênico IGHJ do que nos outros dois grupos estudados e que a contribuição de N2 é um pouco maior nos cavalos imunes.

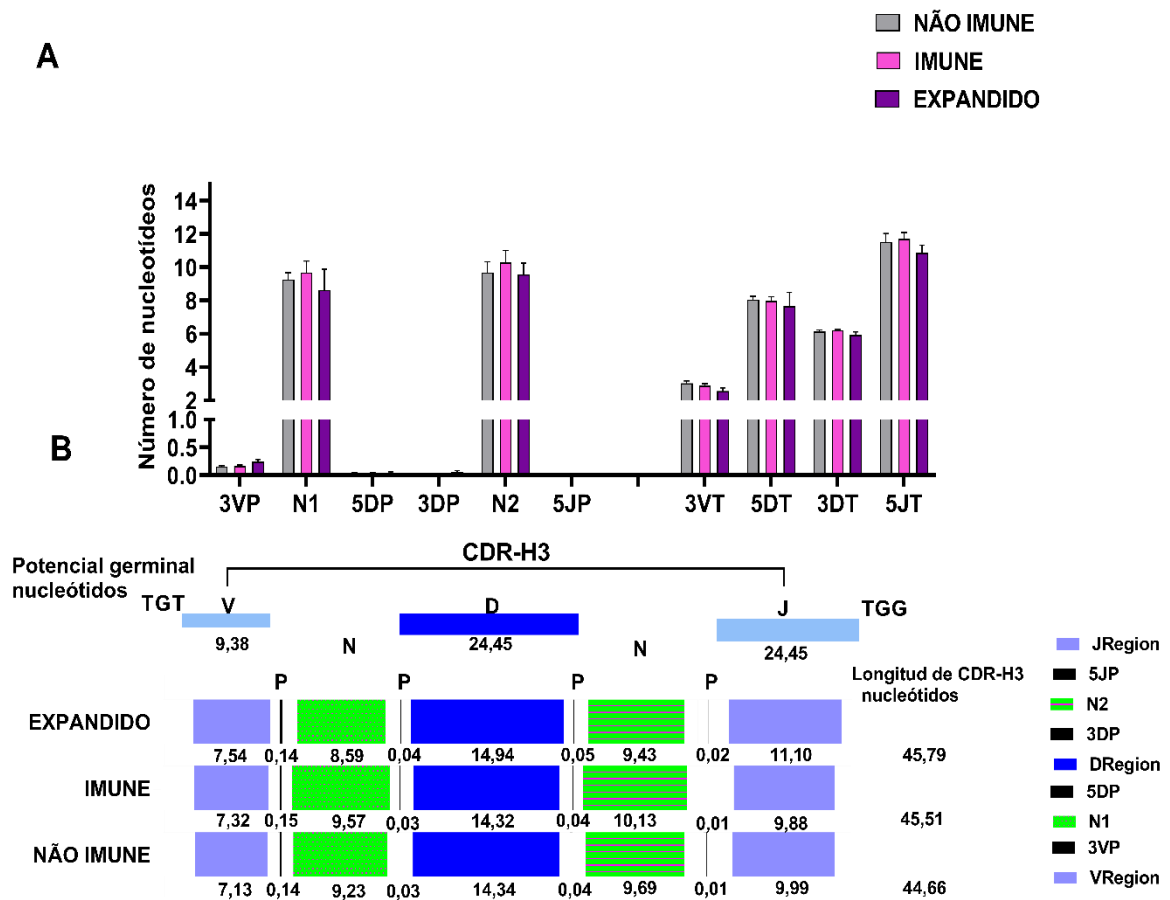


Figura 31. Comparação da diversidade juncional entre os anticorpos de cavalos não imunes, imunes e de clones expandidos. (A) Comparação do número de nucleotídeos adicionados e cortados em cada região da junção. (B) Desconstrução de CDR-H3 nos três grupos estudados.

4.2.8. Comparação de subclasses de IgG entre os grupos estudados

Por fim, comparamos as subclasses de IgG entre os três grupos estudados, o resultado desta comparação amostral está na Figura 32, onde pode-se observar que no caso de IGHG4/IGHG7 a frequência do grupo expandido é maior (19,8%) quando comparado ao grupo não imune (8,9%), no caso do IGHG3 é o contrário, o grupo não imune apresenta maior frequência relativa (23,7%) quando comparado ao grupo imune (11,0%) e expandido (3,3%), assim como na subclasse IGHG6 cuja frequência é maior nos não imunes do que nos demais grupos estudados, nas demais subclasses a frequência foi semelhante.

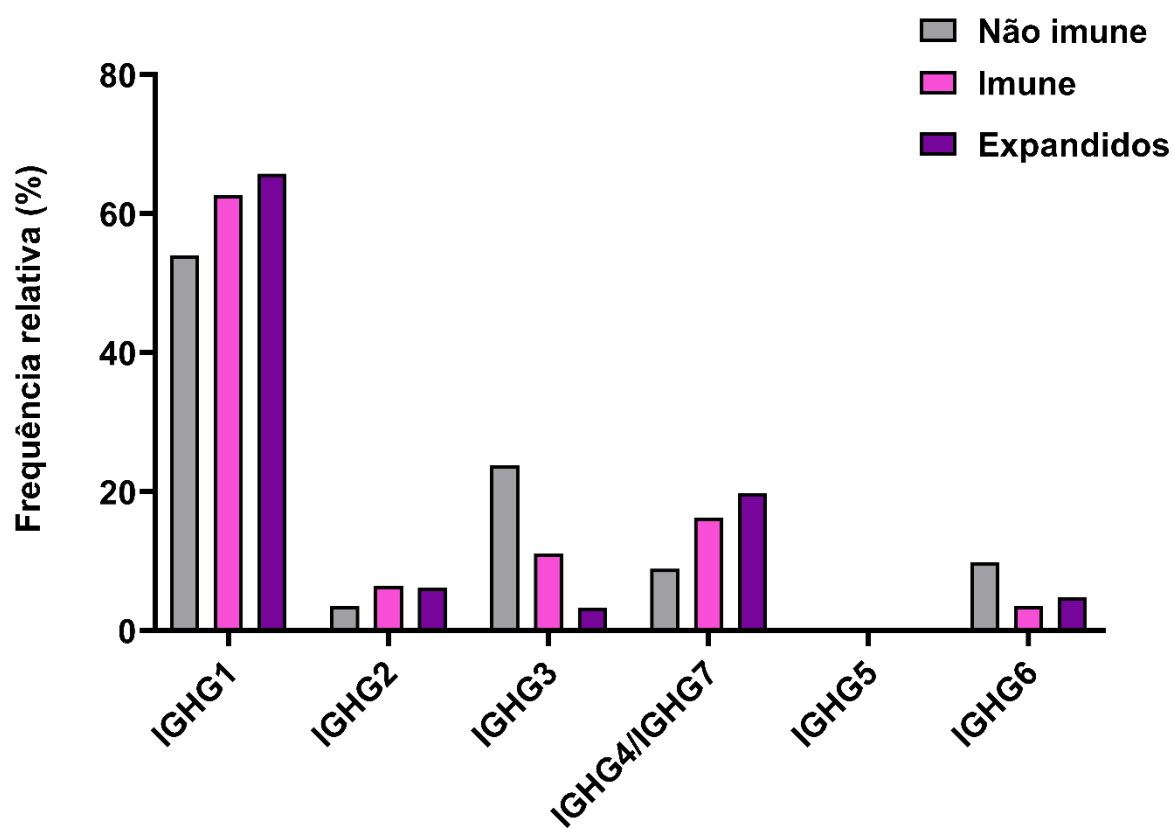


Figura 32. Comparação de subclasses de IgG entre os anticorpos de cavalos não imunes, imunes e de clones expandidos.

5. DISCUSSÃO

Neste trabalho, investigamos o repertório de cadeia pesada de anticorpos de quatro cavalos diferentes, apresentando a maior coleção de sequências de receptores imunes adaptativos descritos até o momento para cavalos.

5.1. Repertório de anticorpos natural de cavalos

Foram analisados 40.018 a 328.300 clones de IgG ou IgM de cavalo, com boa cobertura, de 52 a 82% do repertório. Semelhante a este trabalho, observou-se diferença de profundidade para IgM (36%) quando comparado com IgG (64%) para anticorpos humanos (Galson et al., 2015). Embora não tenha sido observada muita diferença entre os isotipos IgM e IgG, este é o primeiro estudo de sequenciamento de alto rendimento que caracteriza os repertórios eqüinos de ambos os isotipos.

Curiosamente, encontramos que aproximadamente 80% dos anticorpos IgM e IgG apresentam o grupo IGHV4 como um segmento gênico utilizado em seus anticorpos, resultados semelhantes também foram observados por Tallmadge e colaboradores (2013) utilizando 5' RACE, no qual eles encontraram uma forte preferência (80%) pelo uso de IGHV4-29 (anteriormente chamado de IGHV2S3) e IGHJ6-1 (55%) (anteriormente chamado de IGHJ1S5) em anticorpos de cavalos adultos (Chaudhary & Wesemann, 2018; Manso et al., 2019; Sun et al., 2010; Tallmadge et al., 2013). Da mesma forma, os anticorpos humanos também têm preferência pela família IGHV4 (Arnaout et al., 2011), diferenciando-se de outros organismos como bovinos, cães e camundongos que apresentam predominância de genes IGHV1 em seus anticorpos e gatos com preferência de genes (Pasman et al., 2017; Rettig et al., 2018; Steiniger et al., 2014, 2017). Também observamos uma predominância de IGHJ6 em anticorpos de cavalos, enquanto os anticorpos de cães e gatos são compostos principalmente por IGHJ4 e camundongos pelo grupo IGHJ1 (Arnaout et al., 2011; Steiniger et al., 2014, 2017). É importante destacar que foi realizada uma análise de dissimilaridade comparando os clones do repertório IgM e IgG de cada um dos animais estudados e observamos que apesar de ser o mesmo animal, o repertório IgM é diferente do IgG (Anexo V).

É interessante notar que o repertório de anticorpos de cavalo é altamente dominado por apenas alguns genes IGHV e IGHJ, mesmo que eles possam usar todas as suas combinações teóricas de linhagem germinativa. Observamos uma alta frequência de anticorpos ($2,8 \pm 0,9\%$)

contendo a combinação de segmentos gênicos IGHV4-21, IGHD2-26 e IGHJ6-1, também identificados em trabalhos anteriores em cavalos não imunizados (Manso et al., 2019; Tallmadge et al., 2013). Este resultado não é diferente dos repertórios de anticorpos humanos (Arnaout et al., 2011), onde 0,1% a 2,7% das sequências têm as mesmas combinações V(D)J.

Além disso, mesmo para alguns clones, observamos a presença de um repertório público de anticorpos de cavalo na ausência de qualquer estimulação imunológica específica. Encontramos mais clones no repertório público IgM (0,05%) do que no repertório público IgG (0,009%). O pequeno número de clones de anticorpos públicos compartilhados pode ser devido à alta diversidade de anticorpos de cavalos, mas também pode ser devido a um artefato do método de clonotipagem usado neste trabalho que considera o mesmo clone apenas anticorpos com região CDR-H3 idêntica (IMGT/HighVQuest). A observação de mais IgM pública (1,4%) do que IgG (0,3%) e IgA (0,5%) também foi observada no anticorpo humano (Galson et al., 2015).

Curiosamente, mesmo com um número menor de anticorpos públicos, encontramos diferenças entre o comprimento de CDR-H3 do público e todo o repertório, observando-se maior percentual de CDR-H3 curto no repertório público, também observado em humanos (Briney et al., 2019; Galson et al., 2014; Soto et al., 2019). Supõe-se que as células B que expressam receptores com CDR-H3 curtos são selecionadas porque aumentam sua afinidade pelo antígeno, fazem expansão clonal e se diferenciam em plasmócitos ou células B de memória (Rosner et al., 2001).

Neste trabalho, também avaliamos os SHMs no segmento gênico IGHV dos anticorpos de cavalos (regiões FR1, CDR1, FR2, CDR2 e FR3). As sequências de IgG apresentaram frequência de mutação semelhante às sequências de IgM, provavelmente devido à estimulação limitada do patógeno, uma vez que não são imunizadas. Para entender melhor como as mutações são distribuídas ao longo da região IGHV, avaliamos o número de mutações presentes em cada posição para ambos os isotipos de IG estudados. Observamos um perfil de mutação semelhante entre IgG e IgM, com mais mutações encontradas nas regiões CDR, mesmo nesses animais não imunizados. Isso corrobora estudos anteriores em cavalos adultos (Tallmadge et al., 2013) e pacientes saudáveis com HIV (Bowers et al., 2014).

O mecanismo de variabilidade que produz SHM é realizado pela enzima citidina desaminase (AID) induzida por ativação, por desaminação da base de citosina, criando um descasamento U:G. A AID tem como alvo mutações SHM em "hotspots", motivos de nucleotídeos RGYW e WRCY complementares (Spencer & Dunn-Walters, 2005). Este

trabalho observou que cerca de 19% das sequências de IGHV apresentam motivos AID, semelhantes ao IGHV humano, apresentando 17,8% desses motivos (Bowers et al., 2014). Além disso, assim como nos repertórios humanos, encontramos maior percentual de motivos da AID nos CDRs (média de 13,5%) do que nos FRs (média de 5,2%), bem como um maior número de nucleotídeos mutados nesta região (Bowers et al., 2014).

Além da recombinação e da SHM, a diversidade juncional tem um papel muito importante na diversidade dos anticorpos. No entanto, muito pouco tem sido descrito sobre as características da diversidade juncional em anticorpos de cavalos. Aqui, observamos adições de nucleotídeos N1 e N2 na maioria dos clones IgM e IgG, e também observamos adições de nucleotídeos N excepcionalmente longas para IgG N1 e N2 (10-111 pb) em cerca de 44% dos anticorpos. Adições não-modelo aos genes IGH foram relatadas em humanos e camundongos (Shi et al., 2014), porcos (Šinkora et al., 2003) e bovinos (Liljavirta et al., 2014). O número médio de adição de nucleotídeos em humanos é $6,6 \pm 4,3$ em N1 e $6,4 \pm 4,6$ em N2, enquanto em camundongos é $2,4 \pm 2,2$ e $2,1 \pm 1,8$ em N1 e N2, respectivamente (Shi et al., 2014). O número médio de adições de nucleotídeos em bovinos que apresentam CDR-H3s ultralongos não é particularmente alto (2,5 em N1 e 2,6 em N2), refletindo a alta frequência (35%) de junções com zero adições (Liljavirta et al., 2014). Tal frequência e quantidade de adições de N nucleotídeos não foram relatadas em outras espécies, sugerindo que este mecanismo de diversidade é essencial para gerar variabilidade em imunoglobulinas equinas.

O grande número de deleções de genes IGHD em anticorpos de cavalos (valor médio de 6 e 7,9 nucleotídeos para a junção 3'D e 5'D, respectivamente) observado não é muito diferente dos anticorpos de gado (aparamento de 5 a 6 nucleotídeos) (Liljavirta et al., 2014). É interessante notar que o maior número de deleções em anticorpos de cavalos foi observado no gene IGHI, variando entre 10 e 11 nucleotídeos, enquanto em outras espécies, como bovinos, humanos e camundongos, possuem respectivamente 2, 6 e 4 nucleotídeos aparados nesta região (Liljavirta et al., 2014; Shi et al., 2014). Em nossos dados, entre 70% a 97% dos clones de anticorpos tiveram nucleotídeos deletados em qualquer parte da junção, semelhante a outras espécies (Liljavirta et al., 2014; Shi et al., 2014).

Esta impressionante frequência e comprimento de adição de nucleotídeos N pode ser devido a diferenças na atividade da enzima desoxinucleotidil transferase terminal (TdT) em cavalos em comparação com outras espécies.

Essas enzimas são compostas principalmente de duas regiões, um núcleo catalítico composto pelos domínios do dedo, palma e polegar em sua região C-terminal e um domínio N-

terminal composto pelo C-terminal de BRCA1, denominado BRCT. Ao comparar a sequência de cavalo de TdT isorforma 1 e 2 com TdT humana, de camundongo, de suíno e de bovino, observamos a conservação dos ácidos aspárticos catalíticos e da alça substrato-específica 1. Curiosamente, a região do domínio palma, entre os primeiros ácidos aspárticos e a alça1, é uma das regiões mais diferentes entre o TdT de outras espécies.

Além desta região, também podemos observar uma região muito diferente entre o domínio BRCT e o núcleo catalítico de TdT. Até onde sabemos, não está claro como esse domínio não enzimático contribui para a função biológica única da TdT. Interessante notar que esta região interdomínio é enriquecida em aminoácidos prolina. Parece que para a DNA polimerase lambda, que apresenta um domínio rico em prolina maior nesta região em comparação com a TdT, esta região pode impactar a fidelidade da DNA polimerase e com o domínio BRCT pode atuar cooperativamente para promover o realinhamento do primer/template entre as fitas de DNA de homologia de sequência limitada (Fiala et al., 2006; Taggart et al., 2014). Uma vez que o template TdT (Loc'h et al., 2016) é proposto como essencial para a diversidade de anticorpos, estudos futuros precisam investigar o papel da região interdomínio nessas atividades, bem como o papel da região no domínio de palma entre ácidos aspárticos e alça1.

Este estudo também observou que mais de 80% dos anticorpos de cavalo usam a fase de leitura 1 (RF1) para os isotipos IgM e IgG. Várias espécies, como humanos, camundongos e tubarões, produzem anticorpos usando a fase de leitura 1 do IGHD (RF1) (Schroeder et al., 2010). No entanto, à semelhança de outras espécies, o RF1 utilizado para anticorpos de cavalos é fortemente enriquecido em tirosina, representando 39,55% dos aminoácidos IGHD. Sabemos que a tirosina é o aminoácido que normalmente faz a contribuição mais significativa para a afinidade de ligação nas interfaces ligante-receptor da proteína (Bogan & Thorn, 1998). Isso sugere que a seleção natural estava operando em segmentos gênicos de diversidade de imunoglobulinas para restringir e controlar sua evolução de forma a influenciar a composição e variedade de diversidade de sítios de ligação de antígeno de imunoglobulina (Burnet, 1976).

5.2. Comparação do repertório de anticorpos antes e após a imunização com veneno de aranha *loxosceles*

A soroterapia é o tratamento mais indicado para picadas de animais peçonhentos (Alvarenga et al., 2014). Ainda assim, não se conhece quais moléculas presentes nos soros dos

animais imunizados são responsáveis pela neutralização dos efeitos tóxicos dos venenos. Estima-se que apenas 20% dos anticorpos do soro policlonal de cavalo são específicos contra as toxinas do veneno, onde apenas 5% possuem capacidade neutralizante (Theakston et al., 2003).

Neste trabalho verificamos a frequência gênica para os segmentos gênicos IGHV, IGHD e IGHI em anticorpos de cavalos antes e após a imunização com veneno de aranhas do gênero *Loxosceles*, bem como em clones expandidos. Observamos grande semelhança no uso dos segmentos gênicos, com excessão dos genes IGHV4-37N, IGHV4-65 e IGHV4-37 onde a frequência foi maior nos clones expandidos que nos outros grupos estudados. Este é o primeiro estudo que compara o uso gênico em anticorpos de cavalos antes e após a imunização com o veneno de aranha do gênero *Loxosceles*. No entanto, Tallmadge e colaboradores em 2014, ao comparar segmentos gênicos de anticorpos de cavalos em diferentes faixas etárias, encontrando também maior abundância nos segmentos gênicos da família IGHV4.

Em humanos, estudos de repertório antes ou após um estímulo antigênico são mais frequentes. Wolf e colaboradores em 2021 estudaram o repertório de cadeia pesada de anticorpos antes e depois do reforço antitetânico, e descobriram que três segmentos gênicos eram mais frequentes nos anticorpos após duas semanas de vacinação em comparação com antes da vacinação e 1 semana depois, sugerindo uma expansão seletiva (Wolf et al., 2021).

Outro estudo de repertório de anticorpos em humanos foi realizado após a vacinação contra a hepatite B, onde foi constatado que os grupos específicos de vacinas também apresentavam um perfil diferente de uso do segmento gênico IGHV, com surpreendente aumento no uso de IGHV3-7 (Galson et al., 2016).

Quando analisamos o CDR-H3 entre os 3 grupos estudados, pudemos perceber que apesar de muito semelhantes, nos clones expandidos algumas diferenças foram encontradas em relação ao seu tamanho, com maior frequência de CDR-H3 de 5, 20 e 24 aminoácidos no repertório expandido do que nos demais grupos estudados. Em estudos em humanos, observou-se que após a vacina contra hepatite B, os CDR-H3 foram mais curtos quando comparados ao grupo controle (Galson et al., 2016). Quando analisamos a hidrofobicidade do CDR-H3 pudemos observar que os resíduos de aminoácidos são mais hidrofílicos no grupo expandido em relação ao grupo imune ($p = 0,04$), indicando que essas sequências são mais ricas em aminoácidos carregados negativamente, sendo que alguns desses aminoácidos como Glicina e Tirosina (neutra) estão relacionadas à interação com o antígeno (Ivanov I, Link J., Ippolito G.C., 2002).

Em equinos, pouco se sabe sobre o índice de hidrofobicidade de CDR-H3 e ainda menos após a imunização com veneno de aranhas *Loxosceles*. Estudos em humanos revelam que os aminoácidos CDR-H3 são significativamente mais hidrofílicos em anticorpos de pacientes controle quando comparado a pacientes com doença nefrológica, e sabe-se também que um aumento em sequências altamente hidrofóbicas e hidrofílicas durante a diferenciação de células B imaturas para *naives* e de memória, acompanham uma diminuição progressiva na proporção de especificidades de células B autorreativas durante a ontogenia de células B. (Su et al., 2021; Wu et al., 2016).

Neste trabalho, observamos também que tanto a frequência quanto no perfil de mutação de aminoácidos na região IGHV foram muito semelhantes em anticorpos de cavalo antes e após a imunização, bem como nos clones expandidos. Interessante perceber que quando o perfil de mutação por nucleotídeos na região IGHV foi analisado vimos um pico importante de mutação na posição 9 que altera o nucleotídeo “A” para um “G”, mutação que não causa alteração de aminoácidos, ou mutação sinônima. Como a mutação está na região FR1, uma das hipóteses é que essa mutação possa ter vindo de um alelo de IGHV não relatado ainda. No entanto, isto precisa ser validado experimentalmente.

O estudo de Wang et al, em 2018, ao analisar o repertório de anticorpos em macacos com diferentes vacinas contra o HIV, constatou que não houve diferença de SHM nos anticorpos do grupo expandido e o repertório total para o isotipo IgG, mas encontraram diferenças significativas no isotipo IgM, o que pode indicar que a maturação por afinidade de células naive foi induzida. Juntos, esses resultados sugerem que os clones de IGH expandidos por imunização selecionados podem ser derivados principalmente de linhas celulares derivadas da população de células B de memória que precedeu a imunização (L. Wang et al., 2018).

Por fim, a diversidade juncional foi comparada entre os grupos estudados, onde pode-se observar uma frequência semelhante nos três grupos, tanto na adição de nucleotídeos quanto na atividade exonuclease. Pouco se sabe sobre a diversidade juncional antes e após estímulos antigênicos, mas sabe-se que a adição de nucleotídeos aleatórios é maior na idade adulta quando comparada à neonatal em humanos (Hong et al., 2018).

6. CONCLUSÃO

O desenvolvimento deste trabalho permitiu avaliar, de forma independente, sequências de cadeia pesada de anticorpos IgM e IgG de quatro cavalos antes e depois da imunização com veneno de aranha do gênero *Loxosceles*, a partir da amplificação dessas sequências utilizando *primers* específicos para a cadeia constante das imunoglobulinas. Sendo assim, foi possível avaliar as características de ambos isotipos de imunoglobulina, caracterizar e comparar o repertório de anticorpos dos animais.

Este é o primeiro estudo de sequenciamento de alto rendimento que caracteriza o repertório equino dos isotipos IgM e IgG. Mostramos um uso altamente restrito dos genes IGHV e IGHJ no repertório de anticorpos de cavalos em que cerca de 80% dos anticorpos são compostos por apenas 3 segmentos gênicos IGHV e quase 60% deles com o mesmo segmento gênico IGHJ. Observamos um repertório complexo e diversificado para IGH, dado principalmente pela diversidade juncional, muito maior e frequente que a presente em outras espécies.

Nosso estudo sobre o repertório de anticorpos equinos contribui para a compreensão da geração de sua diversidade e abre novas questões sobre as particularidades de TdT de cavalo para gerar tal diversidade.

Podemos concluir também que após a imunização com veneno de aranha do gênero *Loxosceles*, os equinos mantêm o uso dos mesmos segmentos gênicos IGHD e IGHJ em seus anticorpos, e que há um aumento de segmentos gênicos IGHV que não eram abundantes antes da imunização, em clones expandidos, que possuem CDR-H3 mais hidrofóbicos.

Os resultados deste trabalho são importantes para compreendermos melhor a imunidade adaptativa dos cavalos a nível molecular, além de abrir caminhos para o desenvolvimento de novas alternativas terapêuticas.

7. PERSPECTIVAS

Aprofundar a análise comparativa do repertório de anticorpos IgG de equinos antes e após a imunização com veneno de aranha do gênero *Loxosceles*.

- Comparar o repertório de anticorpos IgM de equinos antes e após a imunização com veneno de aranha do gênero *Loxosceles*.
- Analisar a cadeia leve *kappa* e *lambda* de equinos após imunização com veneno de aranha do gênero *loxosceles*.
- Emparelhar cadeias leves e pesadas e, com o auxílio de análises proteômicas, estabelecer sequências de anticorpos capazes de reconhecer componentes do veneno de aranhas do gênero *Loxosceles*.
- Produzir a TdT de cavalo de forma recombinante e analisar seu mecanismo de ação.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- Acácia Maria Lourenço Francisco Nasr, Andrea Carmen Mattos, Dora Yoko Nozaki Goto, Elenice Vieira Torres Rita, Joseana Cardoso de Souza, Nelson Ricetti de Nazareno, Paulo Collodel Junior, Raul Junior Bely, Viviane Serra Melanda, & Wilse Gorges de Almeida. (2017, May). *APOSTILA TABWIN BÁSICO -SIM/SINASC*. Centro de Epidemiologia/Divisão de Informações Epidemiológicas.
- Alamyar, E., Giudicelli, V., Li, S., Duroux, P., & Lefranc, M. P. (2012). IMGT/Highv-quest: The IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Research*, 8(1), 1–15. <https://doi.org/10.4172/1745-7580.1000056>
- Alberts, J., Lewis Julian, Bruce Alberts, Martin Raff, Keith Roberts, & Peter Walter. (2002). B Cells and Antibodies. In *Molecular Biology of the Cell* (4th edition).
- Almagro, J., Martinez, L., Smith, S., 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(11), 1836–1845. <https://doi.org/10.1016/j.molimm.2005.10.017>
- Alvarenga, L. M., Zahid, M., di Tommaso, A., Juste, M. O., Aubrey, N., Billiald, P., & Muzard, J. (2014). Engineering Venom’s toxin-neutralizing antibody fragments and its therapeutic potential. *Toxins*, 6(8), 2541–2567. <https://doi.org/10.3390/toxins6082541>
- Anderson, C. G. (1955). The distribution of diphtheria antitoxin in pepsin-digested horse antiserum. *The Biochemical Journal*, 59(1), 47–52. <https://doi.org/10.1042/bj0590047>
- Aoufouchi, S., Flatter, E., Dahan, A., Faili, A., Bertocci, B., Storck, S., Delbos, F., Cocea, L., Gupta, N., Weill, J.-C., & Reynaud, C.-A. (2000). Two novel human and mouse DNA polymerases of the polX family. In *Nucleic Acids Research* (Vol. 28, Issue 18).
- Aouinti, S., Giudicelli, V., Duroux, P., Malouche, D., Kossida, S., & Lefranc, M. P. (2016). IMGT/statclonotype for pairwise evaluation and visualization of NGS IG and TR IMGT clonotype (AA) diversity or expression from IMGT/HighV-QUEST. *Frontiers in Immunology*, 7(SEP), 1–14. <https://doi.org/10.3389/fimmu.2016.00339>
- Arnaut, R., Lee, W., Cahill, P., Honan, T., Sparrow, T., Weiland, M., Nusbaum, C., Rajewsky, K., & Koralov, S. B. (2011). High-resolution description of antibody heavy-

- chain repertoires in humans. *PLoS ONE*, 6(8).
<https://doi.org/10.1371/journal.pone.0022365>
- Arun, S. S., Breuer, W., & Hermanns, W. (1996). Immunohistochemical examination of light-chain expression (λ/κ ratio) in canine, feline, equine, bovine and porcine plasma cells. *Journal of Veterinary Medicine Series A: Physiology Pathology Clinical Medicine*, 43(9), 573–576. <https://doi.org/10.1111/j.1439-0442.1996.tb00489.x>
- Attaf, M., Huseby, E., & Sewell, A. K. (2015). $\alpha\beta$ T cell receptors as predictors of health and disease. *Cellular and Molecular Immunology*, 12(4), 391–399.
<https://doi.org/10.1038/cmi.2014.134>
- Barbaro, K. C., Knysak, I., Martins, R., Hogan, C., & Winkel, K. (2005). Enzymatic characterization, antigenic cross-reactivity and neutralization of dermonecrotic activity of five *Loxosceles* spider venoms of medical importance in the Americas. *Toxicon*, 45(4), 489–499. <https://doi.org/10.1016/j.toxicon.2004.12.009>
- Bassing, C. H., Swat, W., & Alt, F. W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell*, 109(2 SUPPL. 1), 45–55.
[https://doi.org/10.1016/S0092-8674\(02\)00675-X](https://doi.org/10.1016/S0092-8674(02)00675-X)
- Beutler, B. (2004). Innate immunity: An overview. *Molecular Immunology*, 40(12), 845–859.
<https://doi.org/10.1016/j.molimm.2003.10.005>
- Bogan, A. A., & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, 280(1), 1–9. <https://doi.org/10.1006/jmbi.1998.1843>
- Bolland, D. J., Koohy, H., Wood, A. L., Matheson, L. S., Krueger, F., Stubbington, M. J. T., Baizan-Edge, A., Chovanec, P., Stubbs, B. A., Tabbada, K., Andrews, S. R., Spivakov, M., & Corcoran, A. E. (2016). Two Mutually Exclusive Local Chromatin States Drive Efficient V(D)J Recombination. *Cell Reports*, 15(11), 2475–2487.
<https://doi.org/10.1016/j.celrep.2016.05.020>
- Bolotin, D., Poslavsky, S., Mitrophanow, I., Shugay, M., Mamedov, I., Putintseva, E. v., & Chudakov, D. (2015). MiXCR: software for comprehensive adaptive immunity profiling. In *Nature Methods* (Vol. 12, Issue 5, pp. 379–380). Nature Publishing Group.
<https://doi.org/10.1038/nmeth.3371>
- 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., ... Chudzinski-

- Tavassi, A. M. (2022). Anti-SARS-CoV-2 equine F (Ab')₂ immunoglobulin as a possible therapy for COVID-19. *Scientific Reports*, *12*(1).
<https://doi.org/10.1038/s41598-022-07793-1>
- Bowers, E., Scamurra, R. W., Asrani, A., Beniguel, L., MaWhinney, S., Keays, K. M., Thurn, J. R., & Janoff, E. N. (2014). Decreased mutation frequencies among immunoglobulin G variable region genes during viremic HIV-1 infection. *PLoS ONE*, *9*(1), 1–13.
<https://doi.org/10.1371/journal.pone.0081913>
- Briney, B., Inderbitzin, A., Joyce, C., & Burton, D. R. (2019). Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature*, *566*(7744), 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 Research*, *36*(Web Server issue), 503–508.
<https://doi.org/10.1093/nar/gkn316>
- Burnet, F. M. (1976). A Modification of Jerne's Theory of Antibody Production using the Concept of Clonal Selection. *CA: A Cancer Journal for Clinicians*, *26*(2), 119–121.
<https://doi.org/10.3322/canjclin.26.2.119>
- Butler J. (1998). Immunoglobulin diversity, B-cell end antibody repertoire development in large farm animals - PDF Free Download. *Rev. Sci. Tech.* , *17*(43), 1–17.
- Butler, J. E., Santiago-Mateo, K., Sun, X.-Z., Wertz, N., Sinkora, M., & Francis, D. H. (2011). Antibody Repertoire Development in Fetal and Neonatal Piglets. XX. B Cell Lymphogenesis Is Absent in the Ileal Peyer's Patches, Their Repertoire Development Is Antigen Dependent, and They Are Not Required for B Cell Maintenance. *The Journal of Immunology*, *187*(10), 5141–5149. <https://doi.org/10.4049/jimmunol.1101871>
- Butler, J. E., Wertz, N., Deschacht, N., & Kacs Kovics, I. (2009). Porcine IgG: Structure, genetics, and evolution. *Immunogenetics*, *61*(3), 209–230.
<https://doi.org/10.1007/s00251-008-0336-9>
- Chailyan, A., Marcatili, P., & Tramontano, A. (2011). The association of heavy and light chain variable domains in antibodies: Implications for antigen specificity. *FEBS Journal*, *278*(16), 2858–2866. <https://doi.org/10.1111/j.1742-4658.2011.08207.x>
- Chaudhary, N., & Wesemann, D. R. (2018). Analyzing immunoglobulin repertoires. *Frontiers in Immunology*, *9*(MAR), 1–18. <https://doi.org/10.3389/fimmu.2018.00462>

- Chippaux, J. P., & Goyffon, M. (2008). Epidemiology of scorpionism: A global appraisal. *Acta Tropica*, *107*(2), 71–79. <https://doi.org/10.1016/j.actatropica.2008.05.021>
- Chothia, C., & Lesk, A. M. (1987). Canonical Structures for the Hypervariable Regions of Immunoglobulins. In *J. Mol. Hid.*
- Cole, R., & Moore, H. F. (1917). The production of antipneumococcic serum. *Journal of Experimental Medicine*, *26*(4), 537–561. <https://doi.org/10.1084/jem.26.4.537>
- Collins, A. M., Wang, Y., Roskin, K. M., Marquis, C. P., & Jackson, K. J. L. (2015). The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1676). <https://doi.org/10.1098/rstb.2014.0236>
- Cooper, M. D., & Alder, M. N. (2006). The evolution of adaptive immune systems. *Cell*, *124*(4), 815–822. <https://doi.org/10.1016/j.cell.2006.02.001>
- Cortina-Ceballos, B., Godoy-Lozano, E. E., Sámano-Sánchez, H., Aguilar-Salgado, A., Velasco-Herrera, M. D. C., Vargas-Chávez, C., Velázquez-Ramírez, D., Romero, G., Moreno, J., Téllez-Sosa, J., & Martínez-Barnetche, J. (2015). Reconstructing and mining the B cell repertoire with ImmuneDiversity. *MAbs*, *7*(3), 516–524. <https://doi.org/10.1080/19420862.2015.1026502>
- Cowell, L. G., Kim, H. J., Humaljoki, T., Berek, C., & Kepler, T. B. (1999). Enhanced evolvability in immunoglobulin V genes under somatic hypermutation. *Journal of Molecular Evolution*, *49*(1), 23–26. <https://doi.org/10.1007/PL00006530>
- Cunha, L. E. R., Stolet, A. A., Strauch, M. A., Pereira, V. A. R., Dumard, C. H., 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., Zingali, R. B., Oliveira, G. A. P., Souza, T. M. L., Tanuri, A., Gomes, A. M. O., ... Silva, J. L. (2020). Equine hyperimmune globulin raised against the SARS-CoV-2 spike glycoprotein has extremely high neutralizing titers. *BioRxiv*. <https://doi.org/10.1101/2020.08.17.254375>
- Cupo, P. (2015). Clinical update on scorpion envenoming. *Revista Da Sociedade Brasileira de Medicina Tropical*, *48*(6), 642–649. <https://doi.org/10.1590/0037-8682-0237-2015>
- Darsley, M. J., & Rees, A. R. (1985). Nucleotide sequences of five anti-lysozyme monoclonal antibodies. *The EMBO Journal*, *4*(2), 393–398. <https://doi.org/10.1002/j.1460-2075.1985.tb03641.x>

- Da Silva, P. H., da Silveira, R. B., Helena Appel, M., Mangili, O. C., Gremski, W., & Veiga, S. S. (2004). Brown spiders and loxoscelism. In *Toxicon* (Vol. 44, Issue 7, pp. 693–709). <https://doi.org/10.1016/j.toxicon.2004.07.012>
- De Almeida, D. M., Fernandes-Pedrosa, M. D. F., Gonçalves De Andrade, R. M., Marcelino, J. R., Gondo-Higashi, H., Junqueira De Azevedo, I. D. L. M., Ho, P. L., van den Berg, C., & Tambourgi, D. v. (2008). A new anti-loxoscelic serum produced against recombinant sphingomyelinase D: Results of preclinical trials. *American Journal of Tropical Medicine and Hygiene*, *79*(3), 463–470. <https://doi.org/10.4269/ajtmh.2008.79.463>
- Deiss, T. C., Vadnais, M., Wang, F., Chen, P. L., Torkamani, A., Mwangi, W., Lefranc, M. P., Criscitiello, M. F., & Smider, V. v. (2019). Immunogenetic factors driving formation of ultralong VH CDR3 in *Bos taurus* antibodies. *Cellular and Molecular Immunology*, *16*(1), 64–75. <https://doi.org/10.1038/cmi.2017.117>
- Dekosky, B. J., Ippolito, G. C., Deschner, R. P., Lavinder, J. J., Wine, Y., Rawlings, B. M., Varadarajan, N., Giesecke, C., Dörner, T., Andrews, S. F., Wilson, P. C., Hunicke-Smith, S. P., Willson, C. G., Ellington, A. D., & Georgiou, G. (2013). High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nature Biotechnology*, *31*(2), 166–169. <https://doi.org/10.1038/nbt.2492>
- Delarue, M., Boulé, J. B., Lescar, J., Expert-Bezançon, N., Jourdan, N., Sukumar, N., Rougeon, F., & Papanicolaou, C. (2002). Crystal structures of a template-independent DNA polymerase: Murine terminal deoxynucleotidyltransferase. *EMBO Journal*, *21*(3), 427–439. <https://doi.org/10.1093/emboj/21.3.427>
- De Wildt, R. M. T., van Venrooij, W. J., Winter, G., Hoet, R. M. A., & Tomlinson, I. M. (1999). Somatic insertions and deletions shape the human antibody repertoire. *Journal of Molecular Biology*, *294*(3), 701–710. <https://doi.org/10.1006/jmbi.1999.3289>
- Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, *14*(927 930), 1–4.
- Doherty, A. J., Serpell, L. C., & Ponting, C. P. (1996). The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. In *Nucleic Acids Research* (Vol. 24, Issue 13).
- Early, P., Huang, H., Davis, M., Calame, K., & Hood, L. (1980). *An Immunoglobulin Heavy Chain Variable Region Gene Is Generated from Three Segments of DNA: VH, D and JH* (Vol. 9).

- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Eronique Giudicelli, V. ´, & Lefranc, M.-P. (1999). Ontology for immunogenetics: the IMGT-ONTOLOGY. In *BIOINFORMATICS* (Vol. 15, Issue 12). <http://imgt>.
- Ewing, B., & Green, P. (n.d.). *Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities*. www.genome.org
- Fiala, K. A., Duym, W. W., Zhang, J., & Suo, Z. (2006). Up-regulation of the fidelity of human DNA polymerase λ by its non-enzymatic proline-rich domain. *Journal of Biological Chemistry*, 281(28), 19038–19044. <https://doi.org/10.1074/jbc.M601178200>
- Fitzsimmons, C. M., Falcone, F. H., & Dunne, D. W. (2014). Helminth allergens, parasite-specific IgE, and its protective role in human immunity. In *Frontiers in Immunology* (Vol. 5, Issue FEB). Frontiers Research Foundation. <https://doi.org/10.3389/fimmu.2014.00061>
- Flajnik, M. F. (2002). Comparative analyses of immunoglobulin genes: Surprises and portents. In *Nature Reviews Immunology* (Vol. 2, Issue 9, pp. 688–698). <https://doi.org/10.1038/nri889>
- Foley, R. C. (1992). Analysis of immunoglobulin light chain loci in sheep. In *Animal Genetics* (Vol. 23).
- Fu, L., Li, X., Zhang, W., Wang, C., Wu, J., Yang, H., Wang, J., & Liu, X. (2017). A comprehensive profiling of T- And Blymphocyte receptor repertoires from a Chinese-origin rhesus macaque by highthroughput sequencing. *PLoS ONE*, 12(8). <https://doi.org/10.1371/journal.pone.0182733>
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Futrell, J. M. (1992). Loxoscelism. *American Journal of the Medical Sciences*, 304(4), 261–267. <https://doi.org/10.1097/00000441-199210000-00008>
- Galson, J. D., Pollard, A. J., Trück, J., & Kelly, D. F. (2014). Studying the antibody repertoire after vaccination: Practical applications. *Trends in Immunology*, 35(7), 319–331. <https://doi.org/10.1016/j.it.2014.04.005>
- Galson, J. D., Trück, J., Clutterbuck, E. A., Fowler, A., Cerundolo, V., Pollard, A. J., Lunter, G., & Kelly, D. F. (2016). B-cell repertoire dynamics after sequential hepatitis B

- vaccination and evidence for cross-reactive B-cell activation. *Genome Medicine*, 8(1).
<https://doi.org/10.1186/s13073-016-0322-z>
- Galson, J. D., Trück, J., Fowler, A., Münz, M., Cerundolo, V., Pollard, A. J., Lunter, G., & Kelly, D. F. (2015). In-depth assessment of within-individual and inter-individual variation in the B cell receptor repertoire. *Frontiers in Immunology*, 6(OCT), 1–13.
<https://doi.org/10.3389/fimmu.2015.00531>
- Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. In *Annual Review of Biochemistry* (Vol. 71, pp. 101–132).
<https://doi.org/10.1146/annurev.biochem.71.090501.150203>
- 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. *Nature Biotechnology*, 32(2), 158–168. <https://doi.org/10.1038/nbt.2782>
- Gervásio, J., Ferreira, A., & Felicori, L. F. (2022). Yclon: Ultrafast clustering of B cell clones from high-throughput immunoglobulin repertoire sequencing data. *BioRxiv*, 1–15.
<https://doi.org/10.1101/2022.02.17.480909>
- Giudicelli, V., Chaume, D., & Lefranc, M. P. (2004). IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. *Nucleic Acids Research*, 32(WEB SERVER ISS.).
<https://doi.org/10.1093/nar/gkh412>
- Glatman-Freedman, A., & Casadevall, A. (1998). Serum therapy for tuberculosis revisited: Reappraisal of the role of antibody-mediated immunity against *Mycobacterium tuberculosis*. *Clinical Microbiology Reviews*, 11(3), 514–532.
<https://doi.org/10.1128/cmr.11.3.514>
- Gonçalves, E. S., Salomão, M. G., & Almeida-santos, S. M. de. (2007). O uso do monitoramento espaço-temporal da expansão urbana no diagnóstico de áreas passíveis de risco epidemiológico peçonhento em Guarulhos-Estado de São Paulo, Brasil. *Anais Do XIII Simpósio Brasileiro de Sensoriamento Remoto*, 3171–3178.
- Gremski, L. H., da Silveira, R. B., Chaim, O. M., Probst, C. M. A., Ferrer, V. P., Nowatzki, J., Weinschutz, H. C., Madeira, H. M. I., Gremski, W., Nader, H. B., Senff-Ribeiro, A., & Veiga, S. S. (2010). A novel expression profile of the *Loxosceles intermedia* spider venomous gland revealed by transcriptome analysis. *Molecular BioSystems*, 6(12), 2403–2416. <https://doi.org/10.1039/c004118a>

- Guo, K., Halemano, K., Schmitt, K., Katuwal, M., Wang, Y., Harper, M. S., Heilman, K. J., Kuwata, T., Stephens, E. B., & Santiago, M. L. (2015). Immunoglobulin VH gene diversity and somatic hypermutation during SIV infection of rhesus macaques. *Immunogenetics*, *67*(7), 355–370. <https://doi.org/10.1007/s00251-015-0844-3>
- 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*(20), 3356–3358. <https://doi.org/10.1093/bioinformatics/btv359>
- Hershberg, U., & Luning Prak, E. T. (2015). The analysis of clonal expansions in normal and autoimmune B cell repertoires. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1676). <https://doi.org/10.1098/rstb.2014.0239>
- Hoehn, K. B., Fowler, A., Lunter, G., & Pybus, O. G. (2016). The Diversity and Molecular Evolution of B-Cell Receptors during Infection. *Molecular Biology and Evolution*, *33*(5), 1147–1157. <https://doi.org/10.1093/molbev/msw015>
- Hoffman, W., Lakkis, F. G., & Chalasani, G. (2016). B cells, antibodies, and more. *Clinical Journal of the American Society of Nephrology*, *11*(1), 137–154. <https://doi.org/10.2215/CJN.09430915>
- Hogan, C. J., Barbaro, K. C., & Winkel, K. (2004). Loxoscelism: Old obstacles, new directions. *Annals of Emergency Medicine*, *44*(6), 608–624. <https://doi.org/10.1016/j.annemergmed.2004.08.028>
- Holm, L., & Sander, C. (1995). DNA polymerase B belongs to an ancient nucleotidyltransferase superfamily. *TIBc*, *20*, 1–3.
- Hong, B., Wu, Y., Li, W., Wang, X., Wen, Y., Jiang, S., Dimitrov, D. S., & Ying, T. (2018). In-depth analysis of human neonatal and adult IgM antibody repertoires. *Frontiers in Immunology*, *9*(FEB), 1–13. <https://doi.org/10.3389/fimmu.2018.00128>
- Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, *7*(12), 1451–1456. <https://doi.org/10.1111/2041-210X.12613>
- H Stern, J. N., Yaari, G., vander Heiden, J. A., Church, G., Donahue, W. F., Hintzen, R. Q., Huttner, A. J., Laman, J. D., Nagra, R. M., Nylander, A., Pitt, D., Ramanan, S., Siddiqui, B. A., Vigneault, F., Kleinstein, S. H., Hafler, D. A., & O, K. C. (n.d.). *B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes*. www.ScienceTranslationalMedicine.org

- Hwang, M. H., Darzentas, N., Bienzle, D., Moore, P. F., Morrison, J., & Keller, S. M. (2018). Characterization of the canine immunoglobulin heavy chain repertoire by next generation sequencing. *Veterinary Immunology and Immunopathology*, *202*, 181–190. <https://doi.org/10.1016/j.vetimm.2018.07.002>
- Ito, J., & Braithwaite, D. K. (1991). Compilation and alignment of DNA polymerase sequences. In *Nucleic Acids Research* (Vol. 19, Issue 15). Oxford University Press.
- Ivanov I, Link J., Ippolito G.C., S. H. W. J. (2002). Constraints on the Hydrophobicity and Sequence Composition of HCDR3 are Conserved Across Evolution. In Zanetti Maurizio & Capra Donald (Eds.), *The Antibodies* (Vol 7, Vol. 7, pp. 1–225).
- Jackson, K. J. L., Liu, Y., Roskin, K. M., Glanville, J., Hoh, R. A., Seo, K., Marshall, E. L., Gurley, T. C., Moody, M. A., Haynes, B. F., Walter, E. B., Liao, H. X., Albrecht, R. A., García-Sastre, A., Chaparro-Riggers, J., Rajpal, A., Pons, J., Simen, B. B., Hanczaruk, B., ... Boyd, S. D. (2014). Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell Host and Microbe*, *16*(1), 105–114. <https://doi.org/10.1016/j.chom.2014.05.013>
- Johansson, S. G. O. (2014). The discovery of immunoglobulin e and its role in allergy. *Chemical Immunology and Allergy*, *100*, 150–154. <https://doi.org/10.1159/000358621>
- Johnston, C. M., Wood, A. L., Bolland, D. J., & Corcoran, A. E. (2006). Complete Sequence Assembly and Characterization of the C57BL/6 Mouse Ig Heavy Chain V Region. *The Journal of Immunology*, *176*(7), 4221–4234. <https://doi.org/10.4049/jimmunol.176.7.4221>
- Jung, S., Spinelli, S., Schimmele, B., Honegger, A., Pugliese, L., Cambillau, C., & Plückthun, A. (2001). The importance of framework residues H6, H7 and H10 in antibody heavy chains: Experimental evidence for a new structural subclassification of antibody VH domains. *Journal of Molecular Biology*, *309*(3), 701–716. <https://doi.org/10.1006/jmbi.2001.4665>
- Kaplinsky, J., & Arnaout, R. (2016). Robust estimates of overall immune-repertoire diversity from high-throughput measurements on samples. *Nature Communications*, *7*(May). <https://doi.org/10.1038/ncomms11881>
- Kitaura, K., Yamashita, H., Ayabe, H., Shini, T., Matsutani, T., & Suzuki, R. (2017). Different somatic hypermutation levels among antibody subclasses disclosed by a new next-generation sequencing-based antibody repertoire analysis. *Frontiers in Immunology*, *8*(MAY), 1–11. <https://doi.org/10.3389/fimmu.2017.00389>

- Könning, D., Zielonka, S., Grzeschik, J., Empting, M., Valldorf, B., Krah, S., Schröter, C., Sellmann, C., Hock, B., & Kolmar, H. (2017). Camelid and shark single domain antibodies: structural features and therapeutic potential. In *Current Opinion in Structural Biology* (Vol. 45, pp. 10–16). Elsevier Ltd. <https://doi.org/10.1016/j.sbi.2016.10.019>
- Koshland, M. E. (1985). THE COMING OF AGE OF THE IMMUNOGLOBULIN J CHAIN. In *Ann. Rev. Immunol* (Vol. 3). www.annualreviews.org
- Koti, M., Kataeva, G., & Kaushik, A. K. (2010). Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies. *Molecular Immunology*, *47*(11–12), 2119–2128. <https://doi.org/10.1016/j.molimm.2010.02.014>
- Kuchenbecker, L., Nienen, M., Hecht, J., Neumann, A. U., Babel, N., Reinert, K., & Robinson, P. N. (2015). IMSEQ-A fast and error aware approach to immunogenetic sequence analysis. *Bioinformatics*, *31*(18), 2963–2971. <https://doi.org/10.1093/bioinformatics/btv309>
- Kuroda, D., Shirai, H., Kobori, M., & Nakamura, H. (2008). Structural classification of CDR-H3 revisited: A lesson in antibody modeling. *Proteins: Structure, Function and Genetics*, *73*(3), 608–620. <https://doi.org/10.1002/prot.22087>
- Kyte, J., & Doolittle, R. F. (1982). A Simple Method for Displaying the Hydrophobic Character of a Protein. In *J. Mol. Biol* (Vol. 157).
- Lafaille, J. J., Decloux, A., Bonneville, M., Takagaki, Y., & Howard, S. T. (1989). Junctional Sequences of T Cell Receptor γ 6 Genes: Implications for γ 6 T Cell Lineages and for a Novel Intermediate of V-(D)-J Joining. In *Cell* (Vol. 59).
- Landau, N. R., St Johns, T. P., WEISSMAN, I. L., Wolf, S. C., Silverstone, A. E., & Baltimore, D. (1984). Cloning of terminal transferase cDNA by antibody screening (A \sim gt11/pre-B-lymphoid cells). In *Immunology* (Vol. 81).
- Lang, J., Kanga-Fotso, L., Peyrieux, J. C., Blondeau, C., Lutsch, C., & Forrat, R. (2000). Safety and immunogenicity of a new equine tetanus immunoglobulin associated with tetanus-diphtheria vaccine. *American Journal of Tropical Medicine and Hygiene*, *63*(5–6), 298–305. <https://doi.org/10.4269/ajtmh.2000.63.298>
- Larsen, P. A., & Smith, T. P. L. (2012). *Application of circular consensus sequencing and network analysis to characterize the bovine IgG repertoire*. <http://www.biomedcentral.com/1471-2172/13/52>

- Lavinder, J. J., Hoi, K. H., Reddy, S. T., Wine, Y., & Georgiou, G. (2014). Systematic characterization and comparative analysis of the rabbit immunoglobulin repertoire. *PLoS ONE*, *9*(6). <https://doi.org/10.1371/journal.pone.0101322>
- Lefranc, M. P. (2003). IMGT, the international ImMunoGeneTics database®. *Nucleic Acids Research*, *31*(1), 307–310. <https://doi.org/10.1093/nar/gkg085>
- Lefranc, M. P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., Hadi-Saljoqi, S., Sasorith, S., Lefranc, G., & Kossida, S. (2015). IMGT R, the international ImMunoGeneTics information system R 25 years on. *Nucleic Acids Research*, *43*(D1), D413–D422. <https://doi.org/10.1093/nar/gku1056>
- Lefranc, M. P., & Lefranc, G. (2020). Immunoglobulins or antibodies: IMGT® bridging genes, structures and functions. In *Biomedicines* (Vol. 8, Issue 9). MDPI AG. <https://doi.org/10.3390/biomedicines8090319>
- Lewis, S. M. (1994). P nucleotide insertions and the resolution of hairpin DNA structures in mammalian cells [V(D)J recombination/end joining/illegitimate recombination/severe combined immunodeficiency/scidJ]. In *Biochemistry* (Vol. 91).
- Liljavirta, J., Niku, M., Pessa-Morikawa, T., Ekman, A., & Iivanainen, A. (2014). Expansion of the preimmune antibody repertoire by junctional diversity in *Bos taurus*. *PLoS ONE*, *9*(6). <https://doi.org/10.1371/journal.pone.0099808>
- Litman, G. W., Anderson, M. K., & Rast, J. P. (1999). Evolution of Antigen Binding Receptors. *Annual Review of Immunology*, *17*(1), 109–147. <https://doi.org/10.1146/annurev.immunol.17.1.109>
- Lo'ch, J., & Delarue, M. (2018). Terminal deoxynucleotidyltransferase: the story of an untemplated DNA polymerase capable of DNA bridging and templated synthesis across strands. In *Current Opinion in Structural Biology* (Vol. 53, pp. 22–31). Elsevier Ltd. <https://doi.org/10.1016/j.sbi.2018.03.019>
- Lo'ch, J., Rosario, S., & Delarue, M. (2016). Structural Basis for a New Templated Activity by Terminal Deoxynucleotidyl Transferase: Implications for V(D)J Recombination. *Structure*, *24*(9), 1452–1463. <https://doi.org/10.1016/j.str.2016.06.014>
- López-Santibáñez-Jácome, L., Avendaño-Vázquez, S. E., & Flores-Jasso, C. F. (2019). The pipeline repertoire for Ig-Seq analysis. *Frontiers in Immunology*, *10*(APR), 1–15. <https://doi.org/10.3389/fimmu.2019.00899>

- MAGURRAN, A. (1983). *Ecological Diversity and Its Measurement*.
- Ma, L., Qin, T., Chu, D., Cheng, X., Wang, J., Wang, X., Wang, P., Han, H., Ren, L., Aitken, R., Hammarström, L., Li, N., & Zhao, Y. (2016). Internal Duplications of DH, JH, and C Region Genes Create an Unusual IgH Gene Locus in Cattle. *The Journal of Immunology*, *196*(10), 4358–4366. <https://doi.org/10.4049/jimmunol.1600158>
- Manso, T. (2018). *Sequenciamento de Nova Geração na Análise do Repertório de Anticorpos de Cavalos* [Dotorado]. UNIVERSIDADE FEDERAL DE MINAS GERAIS.
- 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*(August 2018), 251–259. <https://doi.org/10.1016/j.molimm.2018.11.017>
- Martin, V., Wu, Y. C., Kipling, D., & Dunn-Walters, D. (2015). Ageing of the B-cell repertoire. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1676). <https://doi.org/10.1098/rstb.2014.0237>
- 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. *Nature Protocols*, *11*(3), 429–442. <https://doi.org/10.1038/nprot.2016.024>
- Miranda, A. L. S. de, Guerra-Duarte, C., Lima, S. de A., Chávez-Olórtegui, C., & Soto-Blanco, B. (2021). History, challenges and perspectives on *Loxosceles* (brown spiders) antivenom production in Brazil. In *Toxicon* (Vol. 192, pp. 40–45). Elsevier Ltd. <https://doi.org/10.1016/j.toxicon.2021.01.004>
- Monod, M. Y., Giudicelli, V., Chaume, D., & Lefranc, M. P. (2004). IMGT/JunctionAnalysis: The first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. *Bioinformatics*, *20*(SUPPL. 1). <https://doi.org/10.1093/bioinformatics/bth945>
- Monteiro, R. C. (2014). Immunoglobulin A as an anti-inflammatory agent. *Clinical and Experimental Immunology*, *178*(S1), 108–110. <https://doi.org/10.1111/cei.12531>
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, *102*(5), 553–563. [https://doi.org/10.1016/S0092-8674\(00\)00078-7](https://doi.org/10.1016/S0092-8674(00)00078-7)

- Murphy, K. (2014). *Imunologia - Janeway - 8ed* (8th ed.).
- Paixão-Cavalcante, D., van den Berg, C. W., Gonçalves-De-Andrade, R. M., Fernandes-Pedrosa, M. D. F., Okamoto, C. K., & Tambourgi, D. v. (2007). Tetracycline protects against dermonecrosis induced by *Loxosceles* spider venom. *Journal of Investigative Dermatology*, *127*(6), 1410–1418. <https://doi.org/10.1038/sj.jid.5700688>
- Pasman, Y., Merico, D., & Kaushik, A. K. (2017). Preferential expression of IGHV and IGHD encoding antibodies with exceptionally long CDR3H and a rapid global shift in transcriptome characterizes development of bovine neonatal immunity. *Developmental and Comparative Immunology*, *67*, 495–507. <https://doi.org/10.1016/j.dci.2016.08.020>
- Peled, J. U., Kuang, F. L., Iglesias-Ussel, M. D., Roa, S., Kalis, S. L., Goodman, M. F., & Scharff, M. D. (2008). The Biochemistry of Somatic Hypermutation. *Annual Review of Immunology*, *26*(1), 481–511. <https://doi.org/10.1146/annurev.immunol.26.021607.090236>
- Petersen-Mahrt, S. (2005). DNA deamination in immunity. *Immunological Reviews*, *203*(80–97), 1–18.
- Ralph, D. K., & Matsen, F. A. (2016). Consistency of VDJ Rearrangement and Substitution Parameters Enables Accurate B Cell Receptor Sequence Annotation. *PLoS Computational Biology*, *12*(1), 1–25. <https://doi.org/10.1371/journal.pcbi.1004409>
- Reddy, S. T., Ge, X., Miklos, A. E., Hughes, R. A., Kang, S. H., Hoi, K. H., Chrysostomou, C., Hunicke-Smith, S. P., Iverson, B. L., Tucker, P. W., Ellington, A. D., & Georgiou, G. (2010). Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nature Biotechnology*, *28*(9), 965–969. <https://doi.org/10.1038/nbt.1673>
- Rettig, T. A., Ward, C., Bye, B. A., Pecaute, M. J., & Chapes, S. K. (2018). Characterization of the naive murine antibody repertoire using unamplified high-throughput sequencing. *PLoS ONE*, *13*(1), 1–20. <https://doi.org/10.1371/journal.pone.0190982>
- Richterich, P. (n.d.). *Estimation of Errors in “Raw” DNA Sequences: A Validation Study*. www.genome.org
- Rio, D. C., Ares, M., Hannon, G. J., & Nilsen, T. W. (2010). Purification of RNA using TRIzol (TRI Reagent). *Cold Spring Harbor Protocols*, *5*(6), 1–4. <https://doi.org/10.1101/pdb.prot5439>

- Robins, H. (2013). Immunosequencing: applications of immune repertoire deep sequencing. In *Current Opinion in Immunology* (Vol. 25, Issue 5, pp. 646–652). Elsevier Ltd. <https://doi.org/10.1016/j.coi.2013.09.017>
- Robinson, W. H. (2015). Sequencing the functional antibody repertoire—diagnostic and therapeutic discovery. In *Nature Reviews Rheumatology* (Vol. 11, Issue 3, pp. 171–182). Nature Publishing Group. <https://doi.org/10.1038/nrrheum.2014.220>
- Rosner, K., Winter, D. B., Tarone, R. E., & Skovgaard, G. L. (2001). *Third complementarity-determining region of mutated V H immunoglobulin genes contains shorter V , D , J , P , and N components than non-mutated genes.*
- Schatz, D. G., Oettinger, M. A., & Baltimore, D. (1989). The V(D)J Recombination Activating Gene, RAG-1. In *Cell* (Vol. 59).
- Schenone Fernández, H. (1998). Cutaneous loxoscelism with dema predominance. *Bol. Chil. Parasito*, 53(78–83), 78–83.
- Schlissel, M. S. (1998). Structure of Nonhairpin Coding-End DNA Breaks in Cells Undergoing V(D)J Recombination. In *MOLECULAR AND CELLULAR BIOLOGY* (Vol. 18, Issue 4).
- Schroeder, H. W. (2015). The evolution and development of the antibody repertoire. *Frontiers in Immunology*, 6(FEB), 5–6. <https://doi.org/10.3389/fimmu.2015.00033>
- Schroeder, H. W., Zemlin, M., Khass, M., Nguyen, H. H., & Schelonka, R. L. (2010). Genetic control of DH reading frame and its effect on B-cell development and antigen-specific antibody production. *Critical Reviews in Immunology*, 30(4), 327–344. <https://doi.org/10.1615/critrevimmunol.v30.i4.20>
- Sezerino, U. M., Zannin, M., Coelho, L. K., Gonçalves, J., Grando, M., Mattosinho, S. G., Cardoso, J. L. C., von Eickstedt, V. R., França, F. O. S., Barbaro, K. C., & Fan, H. W. (1998). A clinical and epidemiological study of *Loxosceles* spider envenoming in Santa Catarina, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 92(5), 546–548. [https://doi.org/10.1016/S0035-9203\(98\)90909-9](https://doi.org/10.1016/S0035-9203(98)90909-9)
- Shi, B., Ma, L., He, X., Wang, X., Wang, P., Zhou, L., & Yao, X. (2014). Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST. *Theoretical Biology and Medical Modelling*, 11(1), 1–11. <https://doi.org/10.1186/1742-4682-11-30>

- Shugay, M., Bagaev, D. v., Turchaninova, M. A., Bolotin, D. A., Britanova, O. v., Putintseva, E. v., Pogorelyy, M. v., Nazarov, V. I., Zvyagin, I. v., Kirgizova, V. I., Kirgizov, K. I., Skorobogatova, E. v., & Chudakov, D. M. (2015). VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Computational Biology*, *11*(11).
<https://doi.org/10.1371/journal.pcbi.1004503>
- Silva, S. T. da, Tiburcio, I. C. S., Correia, G. Q. C., & Aquino, R. C. T. de. (2005). Escorpiões: origem e diversidade. *Escorpiões, Aranhas e Serpentes: Aspectos Gerais e Espécies de Interesse Médico No Estado de Alagoas*, 10–15.
http://www.usinaciencia.ufal.br/multimedia/livros-digitais-cadernos-tematicos/Escorpioes_Aranhas_e_Serpentes.pdf
- Šinkora, M., Sun, J., Šinkorová, J., Christenson, R. K., Ford, S. P., & Butler, J. E. (2003). Antibody Repertoire Development in Fetal and Neonatal Piglets. VI. B Cell Lymphogenesis Occurs at Multiple Sites with Differences in the Frequency of In-frame Rearrangements. *The Journal of Immunology*, *170*(4), 1781–1788.
<https://doi.org/10.4049/jimmunol.170.4.1781>
- Sirisinha, S. (2014). Evolutionary insights into the origin of innate and adaptive immune systems: Different shades of grey. *Asian Pacific Journal of Allergy and Immunology*, *32*(1), 3–15.
- Smith, A. D., & de Sena Brandine, G. (2021). Falco: High-speed FastQC emulation for quality control of sequencing data. *F1000Research*, *8*.
<https://doi.org/10.12688/f1000research.21142.2>
- Soto, C., Bombardi, R. G., Branchizio, A., Kose, N., Matta, P., Sevy, A. M., Sinkovits, R. S., Gilchuk, P., Finn, J. A., & Crowe, J. E. (2019). High frequency of shared clonotypes in human B cell receptor repertoires. *Nature*, *566*(7744), 398–402.
<https://doi.org/10.1038/s41586-019-0934-8>
- Spencer, J., & Dunn-Walters, D. K. (2005). Hypermutation at A-T Base Pairs: The A Nucleotide Replacement Spectrum Is Affected by Adjacent Nucleotides and There Is No Reverse Complementarity of Sequences Flanking Mutated A and T Nucleotides. *The Journal of Immunology*, *175*(8), 5170–5177.
<https://doi.org/10.4049/jimmunol.175.8.5170>
- 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(1), 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>
- Steitz, T. A., & Steitz, J. A. (1993). *A general two-metal-ion mechanism for catalytic RNA (phosphoryl transfer mechanism/ribozyme/group I splicing/spliceosome/group H splicing)* (Vol. 90). <https://www.pnas.org>
- Strokappe, N. M., Hock, M., Rutten, L., McCoy, L. E., Back, J. W., Caillat, C., Haffke, M., Weiss, R. A., Weissenhorn, W., & Verrips, T. (2019). Super Potent Bispecific Llama VHH Antibodies Neutralize HIV via a Combination of gp41 and gp120 Epitopes. *Antibodies*, 8(2), 38. <https://doi.org/10.3390/antib8020038>
- Sukumar, N., Boule Â, J. B., Expert-Bezanc Èon, N., Jourdan, N., Lescar, J., Rougeon, F., Papanicolaou, C., & Delarue, M. (2000). Biological Crystallography Crystallization of the catalytic domain of murine terminal deoxynucleotidyl transferase. *Terminal Deoxynucleotidyl Transferase Acta Cryst.*
- Sun, Y., Liu, Z., Ren, L., Wei, Z., Wang, P., Li, N., & Zhao, Y. (2012). Immunoglobulin genes and diversity: What we have learned from domestic animals. *Journal of Animal Science and Biotechnology*, 3(1), 1–5. <https://doi.org/10.1186/2049-1891-3-18>
- 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 and Comparative Immunology*, 34(9), 1009–1020. <https://doi.org/10.1016/j.dci.2010.05.003>
- Su, Z., Jin, Y., Zhang, Y., Guan, Z., Li, H., Chen, X., Xie, C., Zhang, C., Liu, X., Li, P., Ye, P., Zhang, L., Kong, Y., & Luo, W. (2021). The Diagnostic and Prognostic Potential of the B-Cell Repertoire in Membranous Nephropathy. *Frontiers in Immunology*, 12. <https://doi.org/10.3389/fimmu.2021.635326>
- Swanson, D. L., & Vetter, R. S. (2006). Loxoscelism. *Clinics in Dermatology*, 24(3), 213–221. <https://doi.org/10.1016/j.clindermatol.2005.11.006>
- Taggart, D. J., Dayeh, D. M., Fredrickson, S. W., & Suo, Z. (2014). N-terminal domains of human DNA polymerase lambda promote primer realignment during translesion DNA synthesis. *DNA Repair*, 22, 41–52. <https://doi.org/10.1016/j.dnarep.2014.07.008>

- Tallmadge, R. L., Tseng, C. T., & Felipe, M. J. B. (2014). Diversity of immunoglobulin lambda light chain gene usage over developmental stages in the horse. *Developmental and Comparative Immunology*, *46*(2), 171–179.
<https://doi.org/10.1016/j.dci.2014.04.001>
- Tallmadge, R. L., Tseng, C. T., King, R. A., & Felipe, M. J. B. (2013). Developmental progression of equine immunoglobulin heavy chain variable region diversity. *Developmental and Comparative Immunology*, *41*(1), 33–43.
<https://doi.org/10.1016/j.dci.2013.03.020>
- Theakston, R. D. G., Warrell, D. A., & Griffiths, E. (2003). Report of a WHO workshop on the standardization and control of antivenoms. *Toxicon*, *41*(5), 541–557.
[https://doi.org/10.1016/S0041-0101\(02\)00393-8](https://doi.org/10.1016/S0041-0101(02)00393-8)
- 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*(13), 1930–1932. <https://doi.org/10.1093/bioinformatics/btu138>
- Van Egmond, N. D. (n.d.). *Historical Perspective and Future Outlook*.
- Veiga, S. S., da Silveira, R. B., Dreyfuss, J. L., Haoach, J., Pereira, A. M., Mangili, O. C., & Gremski, W. (2000). Identification of high molecular weight serine-proteases in *Loxosceles intermedia* (brown spider) venom. *Toxicon*, *38*(6), 825–839.
[https://doi.org/10.1016/S0041-0101\(99\)00197-X](https://doi.org/10.1016/S0041-0101(99)00197-X)
- Wagner, B. (2006). Immunoglobulins and immunoglobulin genes of the horse. *Developmental and Comparative Immunology*, *30*(1–2), 155–164.
<https://doi.org/10.1016/j.dci.2005.06.008>
- Walther, S., Rusitzka, T. v., Diesterbeck, U. S., & Czerny, C. P. (2015). Equine immunoglobulins and organization of immunoglobulin genes. *Developmental and Comparative Immunology*, *53*(2), 303–319. <https://doi.org/10.1016/j.dci.2015.07.017>
- Wang, B., Kluwe, C. A., Lungu, O. I., DeKosky, B. J., Kerr, S. A., Johnson, E. L., Jung, J., Rezig, A. B., Carroll, S. M., Reyes, A. N., Bentz, J. R., Villanueva, I., Altman, A. L., Davey, R. A., Ellington, A. D., & Georgiou, G. (2015). Facile Discovery of a Diverse Panel of Anti-Ebola Virus Antibodies by Immune Repertoire Mining. *Scientific Reports*, *5*. <https://doi.org/10.1038/srep13926>
- Wang, F., Ekiert, D. C., Ahmad, I., Yu, W., Zhang, Y., Bazirgan, O., Torkamani, A., Raudsepp, T., Mwangi, W., Criscitiello, M. F., Wilson, I. A., Schultz, P. G., & Smider,

- V. v. (2013). XReshaping antibody diversity. *Cell*, 153(6).
<https://doi.org/10.1016/j.cell.2013.04.049>
- Wang, L., Zhang, W., Lin, L., Li, X., Saksena, N. K., Wu, J., Wang, S., Joyce, J. G., Zhang, X., Yang, H., Wang, J., Wang, I. M., & Liu, X. (2018). A Comprehensive Analysis of the T and B Lymphocytes Repertoire Shaped by HIV Vaccines. *Frontiers in Immunology*, 9. <https://doi.org/10.3389/fimmu.2018.02194>
- Watson, C. T., Glanville, J., & Marasco, W. A. (2017). The Individual and Population Genetics of Antibody Immunity. In *Trends in Immunology* (Vol. 38, Issue 7, pp. 459–470). Elsevier Ltd. <https://doi.org/10.1016/j.it.2017.04.003>
- Watson, C. T., Steinberg, K. M., Graves, T. A., Warren, R. L., Malig, M., Schein, J., Wilson, R. K., Holt, R. A., Eichler, E. E., & Breden, F. (2015). Sequencing of the human IG light chain loci from a hydatidiform mole BAC library reveals locus-specific signatures of genetic diversity. *Genes and Immunity*, 16(1), 24–34.
<https://doi.org/10.1038/gene.2014.56>
- Weiss, S., Lehmann, K., Raschke, W. C., & Cohn, M. (1984). Mice completely suppressed for the expression of immunoglobulin K light chain (immunoglobulin A light chain/compensation of Kchain loss/heterogeneity/Immunological repertoire). In *Proc. Nad. Acad. Sci. USA* (Vol. 81).
- Wertz, N., Vazquez, J., Wells, K., Sun, J., & Butler, J. E. (2013). Antibody repertoire development in fetal and neonatal piglets. XII. Three IGLV genes comprise 70% of the pre-immune repertoire and there is little junctional diversity. *Molecular Immunology*, 55(3–4), 319–328. <https://doi.org/10.1016/j.molimm.2013.03.012>
- 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 of the United States of America*, 110(8), 2993–2998. <https://doi.org/10.1073/pnas.1213737110>
- Wolf, K. J., Adeyemo, A. A., & Williamson, K. C. (2021). Plasma cell immunoglobulin heavy chain repertoire dynamics before and after tetanus booster vaccination. *Immunogenetics*, 73(4), 321–332. <https://doi.org/10.1007/s00251-021-01215-8>
- Wu, J., Jia, S., Wang, C., Zhang, W., Liu, S., Zeng, X., Mai, H., Yuan, X., Du, Y., Wang, X., Hong, X., Li, X., Wen, F., Xu, X., Pan, J., Li, C., & Liu, X. (2016). Minimal residual disease detection and evolved IGH clones analysis in acute b lymphoblastic leukemia

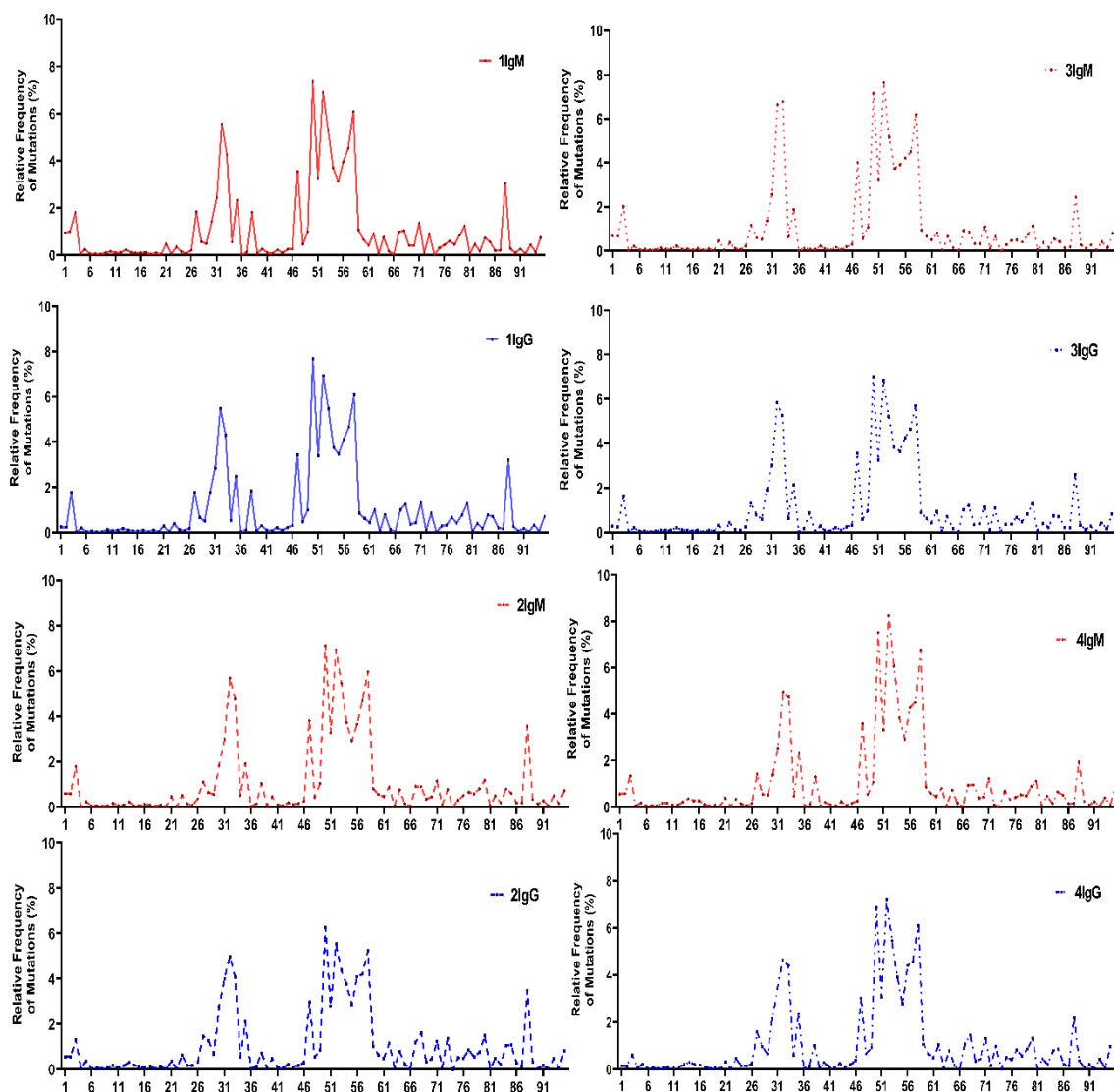
- using IGH deep sequencing. *Frontiers in Immunology*, 7(OCT).
<https://doi.org/10.3389/fimmu.2016.00403>
- Yaari, G., & Kleinstein, S. H. (2015). Practical guidelines for B-cell receptor repertoire sequencing analysis. In *Genome Medicine* (Vol. 7, Issue 1). BioMed Central Ltd.
<https://doi.org/10.1186/s13073-015-0243-2>
- Yangs, B., Gathy, K. N., & Coleman, M. S. (1994). Mutational Analysis of Residues in the Nucleotide Binding Domain of Human Terminal Deoxynucleotidyl Transferase*. In *OF BIOLOGICAL CHEWISTRY* (Vol. 269, Issue 16).
- Ye, J., Ma, N., Madden, T. L., & Ostell, J. M. (2013). IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Research*, 41(Web Server issue).
<https://doi.org/10.1093/nar/gkt382>
- Yu, Y., Ceredig, R., & Seoighe, C. (2015). LymAnalyzer: A tool for comprehensive analysis of next generation sequencing data of T cell receptors and immunoglobulins. *Nucleic Acids Research*, 44(4). <https://doi.org/10.1093/nar/gkv1016>
- Zylberman, V., Sanguineti, S., Pontoriero, A. v., Higa, S. v., Cerutti, M. L., Seijo, S. M. M., Pardo, R., Muñoz, L., Intrieri, M. E. A., 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., ... Goldbaum, F. A. (2020). Development of a hyperimmune equine serum therapy for covid-19 in Argentina. *Medicina*, 80, 1–6.

ANEXO I

Frequência de mutação nos picos da região V dos três principais aminoácidos para os isotipos IgM e IgG

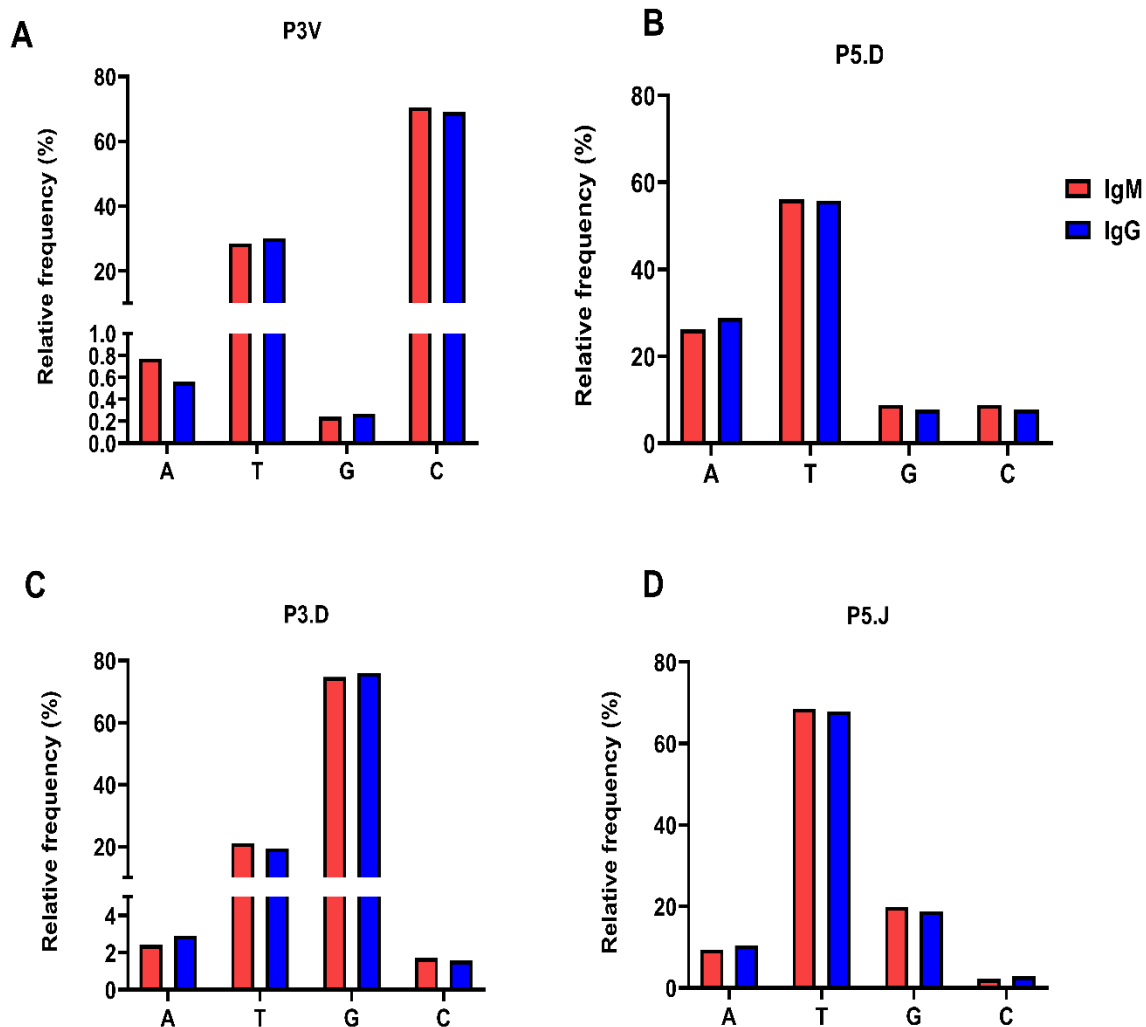
Posição	Mudança de aminoácido no isotipo IgM	%	Mudança de aminoácido no isotipo IgG	%
3	Q por K	1,84	Q por K	0,97
	Q por P	0,014	K por Q	0,010
	K por Q	0,012	Q por H	0,005
27	L por S	0,13	L por F	0,48
	L por Y	0,098	L por S	0,079
	L por T	0,082	L por Y	0,077
32	Y por N	2,48	Y por N	1,41
	Y por S	0,49	Y por S	0,38
	Y por V	0,43	Y por V	0,31
33	A por G	0,59	G por A	0,54
	A por S	0,52	A por T	0,30
	A por R	0,48	A por S	0,25
50	A por G	0,59	A por G	0,49
	A por S	0,52	A por S	0,45
	A por R	0,48	A por T	0,34
52	S por T	0,44	S por T	0,32
	S por K	0,42	S por G	0,32
	S por G	0,42	S por K	0,29
58	N por R	0,72	N por Y	0,43
	N por Y	0,65	N por R	0,40
	N por D	0,48	N por D	0,35
88	G por S	1,73	G por S	1,06
	N por R	0,72	S por Y	0,17
	N por Y	0,65	Y por T	0,10

ANEXO II



Perfil de mutação de aminoácidos na região V dos cavalos não imunes analisados (1, 2, 3, 4). Avaliação da frequência das mutações observadas. IgM em vermelho e IgG em azul.

ANEXO III



Frequência de nucleotídeos nas diferentes regiões da junção. (A) Frequência relativa de nucleotídeos na região 3VP. (B) Frequência relativa de nucleotídeos na região 5DP. (C) Frequência relativa de nucleotídeos na região 3DP. (D) Frequência relativa de nucleotídeos na região 5JP. IgM em vermelho e IgG em azul.

ANEXO IV

Molecular Immunology
THE MAJOR ROLE OF JUNCTIONAL DIVERSITY IN THE HORSE ANTIBODY
REPertoire
 --Manuscript Draft--

Manuscript Number:	
Article Type:	Full Length Article
Keywords:	horse; antibody repertoire; BCR-seq; junctional diversity
Corresponding Author:	Liza Felicori Universidade Federal de Minas Gerais Belo Horizonte, BRAZIL
First Author:	Carlena Navas
Order of Authors:	Carlena Navas Taciana Manso Fabio Martins Lucas Minto Rennan Moreira João Minozzo Bruno Antunes André Vale Jonathan R. McDaniel Gregory Ippolito Liza Felicori
Abstract:	<p>The sequencing of the antibody repertoire (Rep-seq) revolutionized the diversity of antigen B cell receptor studies, allowing deep and quantitative analysis to decipher the role of adaptive immunity in health and disease. Particularly, horse (<i>Equus caballus</i>) polyclonal antibodies have been produced and used since the century XIX to treat and prophylaxis of diphtheria, tuberculosis, tetanus, pneumonia, and, more recently, COVID-19. However, our knowledge about the horse B cell receptors repertoires is minimal. We present a deep horse antibody heavy chain repertoire (IGH) characterization of non-immunized horses using HTS technology. In this study, we obtained a mean of 248,169 unique IgM clones and 66,141 unique IgG clones from four domestic adult horses. Rarefaction analysis showed sequence coverage was between 52 and 82% in IgM and IgG isotypes. We observed that besides horses antibody can use all of the functional IGHV genes, around 80% of their antibodies use only three IGHV gene segments, and around 55% use only one IGHJ gene segment. This limited VJ diversity seems to be compensated by the junctional diversity of these antibodies. We observed that the junctional diversity in horses antibodies is highly frequent, present in more than 90% of horse antibodies. Besides this, the length of this region seems to be higher in horse antibodies than in other species. N1 and N2 nucleotides addition range from 0 to 111 nucleotides. In addition, around 45% of the antibody clones have more than ten nucleotides in both N1 and N2 junction regions. This diversity mechanism may be one of the most important in providing variability to the equine antibody repertoire. This study provides new insights regarding horse antibody composition, diversity generation, and particularities compared to other species, such as the frequency and length of N nucleotide addition. This study also points out the urgent need to better characterize TdT in horses and in other species to better understand antibody repertoire characteristics.</p>
Suggested Reviewers:	Rebecca L. Tallmadge rlt8@cornell.edu Expert in horse antibodies. Published previous works characterizing horse antibodies repertoire using Sanger sequencing

	<p>Antti Iivanainen anti.iivanainen@helsinki.fi Published a work characterizing cattle antibodies repertoire (PMID: 24926997).</p>
	<p>Harry W Schroeder Jr hwsj@uab.edu Expert in IGHD gene characterization</p>



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

Dear Editor,

We are pleased to send to you a manuscript entitled **“THE MAJOR ROLE OF JUNCTIONAL DIVERSITY IN THE HORSE ANTIBODY REPERTOIRE”** which we would like to have published in *Molecular Immunology*. This article reports high-throughput sequencing of the antibody repertoire for the IgM and IgG isotypes of PBMC from non-immunized horses. A previously study of our group published in *Molecular Immunology* characterized horse antibodies, in a high-throuput way, for the first time (PMID: 30562645). Now, with a depper analysis of 4 horses instead of 2 previsously analysed, we observed that horses antibodies has a restrict VJ usage in their antibodies. We observed that around 80% of IgM and IgG repertoires are composed by only 3 different IGHV4 gene segments, and 60% of them use only the IGHJ6 gene segment. This restrict gene usage in horses antibodies seems to be compensated by a high frequent and in many times, long junctional diversity of N1 and N2 regions, that can have from 0 to 111 nucleotides. In this way, this study provides new insights regarding horse antibody composition, diversity generation, and particularities compared to other species, such as the frequency and length of N nucleotide addition.

For that, we believe that our results are very relevant and of general interest. In addition, studies of the antibody repertoire of dogs (PMID: 24509215), salmon (PMID: 30593934), pigs (<https://doi.org/10.1016/j.molimm.2005.10.017>) among other species have also been published in *Molecular Immunology*. Given the interest of the *Molecular Immunology* community in understand the characteristics of antibody rcpertoires from different species, we believe this work present a depper and significant contribution for a better understanding of horse antibody characteristics.

Looking forward to hearing from you,

Yours truly,

Felicori



Liza Felicori

Associate Professor |
Principal investigator
Synbiom Lab
Coordinator of the IdeaReal
lab: www.ideareal.org

SYNBIOM

Laboratório de Biologia
Sintética e Biomiméticos
Departamento de
Bioquímica e Imunologia |
ICB
Universidade Federal de
Minas Gerais

☎ +55 (31) 99330-4420

✉ liza@icb.ufmg.br

📍 Av. Antonio Carlos, 6627 –
Bloco N4 – sala 202 –
Pampulha
Belo Horizonte – MG –
Brazil
CEP 31270-901

Highlights

- Horse antibodies have a restrict VJ gene usage
- Only 3 different IGHV4 and the IGJH6 genes are used for most of the horses antibodies
- The junctional diversity in horse antibodies looks bigger compared to other species
- More than 80% of horse antibodies use IGHD reading frame 1

Abstract

The sequencing of the antibody repertoire (Rep-seq) revolutionized the diversity of antigen B cell receptor studies, allowing deep and quantitative analysis to decipher the role of adaptive immunity in health and disease. Particularly, horse (*Equus caballus*) polyclonal antibodies have been produced and used since the century XIX to treat and prophylaxis of diphtheria, tuberculosis, tetanus, pneumonia, and, more recently, COVID-19. However, our knowledge about the horse B cell receptors repertoires is minimal. We present a deep horse antibody heavy chain repertoire (IGH) characterization of non-immunized horses using HTS technology. In this study, we obtained a mean of 248,169 unique IgM clones and 66,141 unique IgG clones from four domestic adult horses. Rarefaction analysis showed sequence coverage was between 52 and 82% in IgM and IgG isotypes. We observed that besides horses antibody can use all of the functional IGHV genes, around 80% of their antibodies use only three IGHV gene segments, and around 55% use only one IGHV gene segment. This limited VJ diversity seems to be compensated by the junctional diversity of these antibodies. We observed that the junctional diversity in horses antibodies is highly frequent, present in more than 90% of horse antibodies. Besides this, the length of this region seems to be higher in horse antibodies than in other species. N1 and N2 nucleotides addition range from 0 to 111 nucleotides. In addition, around 45% of the antibody clones have more than ten nucleotides in both N1 and N2 junction regions. This diversity mechanism may be one of the most important in providing variability to the equine antibody repertoire. This study provides new insights regarding horse antibody composition, diversity generation, and particularities compared to other species, such as the frequency and length of N nucleotide addition. This study also points out the urgent need to better characterize TdT in horses and in other species to better understand antibody repertoire characteristics.

THE MAJOR ROLE OF JUNCTIONAL DIVERSITY IN THE HORSE ANTIBODY REPERTOIRE

Carlena Navas^{a, b}, Taciana Manso^{a, c}, Fabio Martins^a, Lucas Minto^a, Rennan Moreira^d, João Minozzo^e, Bruno Antunes^e, André Vale^f, Jonathan R. McDaniel^g, Gregory C. Ippolito^g, Liza F. Felicori^{a*}

^aLaboratory of Synthetic Biology and Biomimetics, Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas - ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

^b University of Carabobo, Faculty of Health Sciences, School of Biomedical and Technological Sciences Department of Morphological and Forensic Sciences, Valencia Venezuela.

^c The International Immunogenetics Information System / IMGT Institut de Génétique Humaine / IGH – CNRS Montpellier / France.

^d Multi-users Laboratories Center, Instituto de Ciências Biológicas - ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

^eProduction and Research Centre of Immunobiological Products, Department of Health of the State of Paraná, Piraquara 83302-200, Brazil.

^fProgram in Immunobiology, Carlos Chagas Filho Institute of Biophysics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

^g Department of Molecular Biosciences, The University of Texas at Austin, 100 E. 24th Street, Stop A5000, Austin, TX, 78712, USA

Correspondence:

*Liza F. Felicori

Email: liza@icb.ufmg.br

Keywords: horse, antibody repertoire, BCR-seq, junctional diversity

Abstract

The sequencing of the antibody repertoire (Rep-seq) revolutionized the diversity of antigen B cell receptor studies, allowing deep and quantitative analysis to decipher the role of adaptive immunity in health and disease. Particularly, horse (*Equus caballus*) polyclonal antibodies have been produced and used since the century XIX to treat and prophylaxis of diphtheria, tuberculosis, tetanus, pneumonia, and, more recently, COVID-19. However, our knowledge about the horse B cell receptors repertoires is minimal. We present a deep horse antibody heavy chain repertoire (IGH) characterization of non-immunized horses using HTS technology. In this study, we obtained a mean of 248,169 unique IgM clones and 66,141 unique IgG clones from four domestic adult horses. Rarefaction analysis showed sequence coverage was between 52 and 82% in IgM and IgG isotypes. We observed that besides horses antibody can use all of the functional IGHV genes, around 80% of their antibodies use only three IGHV gene segments, and around 55% use only one IGHJ gene segment. This limited VJ diversity seems to be compensated by the junctional diversity of these antibodies. We observed that the junctional diversity in horses antibodies is highly frequent, present in more than 90% of horse antibodies. Besides this, the length of this region seems to be higher in horse antibodies than in other species. N1 and N2 nucleotides addition range from 0 to 111 nucleotides. In addition, around 45% of the antibody clones have more than ten nucleotides in both N1 and N2 junction regions. This diversity mechanism may be one of the most important in providing variability to the equine antibody repertoire. This study provides new insights regarding horse antibody composition, diversity generation, and particularities compared to other species, such as the frequency and length of N nucleotide addition. This study also points out the urgent need to better characterize TdT in horses and in other species to better understand antibody repertoire characteristics.

Introduction

The effective humoral immune response depends partly on having a variety of B cells with different B cell receptors (BCRs) capable of recognizing and binding to many different antigens. The entire set of B cells with different BCRs is called the antibody repertoire (Glanville et al., 2009). In humans, it has the theoretical potential to reach a size of up to 10^{16} - 10^{18} unique antibody sequences (Briney et al., 2019).

The sequencing of the repertoire (Rep-seq) revolutionized the antigen B-cell receptors studies, allowing deep and quantitative analysis to decipher the role of adaptive immunity in health and disease (Georgiou et al., 2014). However, besides being less common, antibody repertoire analysis in species such as chicken, sheep, pig, cattle, and horses revealed new insights into the many different mechanisms that can create antibodies diversity in vertebrates (Butler et al., 2009; Liljavirta et al., 2014; Reynaud et al., 1989; Sun et al., 2012).

Particularly, horses (*Equus caballus*) polyclonal antibodies have been produced and used since the century XIX for the treatment and prophylaxis of diseases such as diphtheria, tuberculosis, tetanus, and pneumonia (ANDERSON, 1955; Cole & Moore, 1917; Glatman-Freedman & Casadevall, 1998; Gonçalves et al., 2007; Lang et al., 2000) to the present day. It is even being used in the current COVID-19 pandemic as a treatment in some countries (Cunha et al., 2020; Zylberman et al., 2020).

Similar to other vertebrates, horses have three types of immunoglobulin chains: light lambda (IGL), light kappa (IGK), and heavy (IGH). The horse antibody V(D)J gene segments were annotated by Sun et al. (2010) and reviewed by Walter et al. (2015), using an EquCab 2.0 genome composed of several scaffolds. After that, the EquCab3.0 genome was published, and the international ImMunoGeneTics information system® (IMGT®) annotated the IG locus. In this annotation, the horse IGH locus present on chromosome 24 has 104 IGHV (21 functional, 74 pseudogenes, and nine ORFs), 44 IGHD (16 functional, twenty-eight ORFs), and nine IGHJ (six functional, and three ORFs).

So far, analyzes of the repertoire of equine antibodies have been carried out by different methodologies, most of them by Sanger sequencing, with low deepness (Almagro et al., 2006; Tallmadge et al., 2013, 2014). Only in 2019, our group carried out a deeper horse antibody

repertoire analysis using the new generation technology (NGS) was carried out, showing some characteristics of 45,000 IGH clones and 30,000 IGL clones as new gene transcripts (IGHV6S1 and IGLV4S2) and the amino acids composition and features of CDR-H3 (Manso et al., 2019). However, some essential horse antibody repertoire characteristics are still unclear, such as somatic hypermutation frequency and the characteristics of the junction, among others. Furthermore, the fraction of the potential repertoire expressed in an individual is unknown, and how similar repertoires are between individuals who have lived in similar environments.

We present a deeper horse antibody heavy chain repertoire (IGH) characterization of non-immunized horses using HTS technology, where we obtained a mean of 248,169 unique IgM clones and 66,141 unique IgG clones from four domestic adult horses. Sequence coverage was between 52 and 82% in IgM and IgG isotypes. We observed that the IGHV4 subgroup is expressed in around 80% of horse's antibodies, and between 50% and 56% use IGHJ6 indicating limited use of combinations of gene segments. However, most horse antibody IgM and IgG clones (~91%) present N-nucleotide addition, reaching 78 nucleotides in N1 and 62 in N2 regions for IgM and 111 nucleotides in N1 and 104 in N2 for IgG. These results suggest a major role of junctional diversity in generating equine antibody repertoire variability.

MATERIALS AND METHODS

Horse blood samples

The peripheral blood samples from four healthy, mixed male breed adult horses, aged 5 to 9 years old, were obtained in partnership with the Immunobiological Research and Production Center (CPPI) of the State of Paraná.

About 35 ml of peripheral blood was obtained from each animal using Vacutainer tubes with EDTA anticoagulant. The PBMC were isolated by Ficoll-Paque™ gradient centrifugation. The cells (1×10^7 cells) were cryopreserved in FBS 90%/ DMSO 10% at -80 °C until use.

The Ethics Committee approved the experimental design on the Use of Animals of the Federal University of Minas Gerais (CEUA - UFMG) under protocol number 190/2018.

Amplification of the horse antibody BCR repertoire

Mononuclear cells (PBMCs) were isolated for RNA extraction and subsequent cDNA synthesis. Total RNA extraction was performed by the TRIzol method (Rio et al., 2010), and the RNA

concentrations were verified by the Qubit RNA BR Assay kit (Thermo Fisher Scientific). According to the manufacturer's instructions, approximately 500 ng of RNA was used for cDNA synthesis using the SuperScript IV enzyme (Thermo Fisher Scientific). The IGH amplification of the gene segments V and the constant region was carried out by multiplex PCR. A set of forward specific primers (F) for the heavy chain variable region (Manso et al., 2019) was used with new reverse specific primers (R) for the heavy chain constant region designed in this study: IgM isotype 5' ATGACGTTGGGTAGGAAGTCCCG 3' and IgG isotype 5' CCACCGTGGMGTCAGAYGTG 3'. All primers have incorporated the Illumina overhang adaptors sequence to prepare the Illumina library.

Multiplex PCR reactions were conducted to obtain IGH amplicons from each of the four horses. All reactions were prepared to contain 10X High Fidelity buffer, 50 mM MgSO₄, 10 mM dNTPs, 0.5 μM of each F primer, 0.5 μM of each R primer, and 0.5 U Taq DNA polymerase Platinum High Fidelity (Thermo Fisher Scientific). The cycling parameters were 94 °C for 2 min; 4 cycles of 94 °C for 1 min, 50 °C for 1 min and 72 °C for 1 min; 4 cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min; 26 cycles of 94 °C for 1 min, 63 °C for 1 min and 72 °C for 1 min, and 72 °C for 7 min. The amplifications were analyzed on 1% agarose gels and stained with Sybr Safe (Invitrogen). The bands were excised, and purified with PCR clean-up Gel extraction (NucleoSpin).

Library preparation and sequencing

The purified cDNAs were quantified by Qubit DNA High Sensitivity kit (Thermo Fisher Scientific). Then, it was used for sequencing libraries prepared by the Nextera XT DNA Library Prep kit (Illumina) according to the manufacturer's instructions. The P5 and P7 indexes and adapters were incorporated into the 500 bp amplicons by the overhang adapters added to the primers. The library concentration was verified using Qubit DNA High Sensitivity kit (Thermo Fisher Scientific), and the size and quality of amplicons were confirmed with the Bioanalyzer High Sensitivity DNA Analysis (Agilent).

The IGH samples (18 pM) from the four equines were sequenced using Illumina MiSeq platform 2 × 300 bp read length. The sequences have been deposited at the NIH SRA (Sequence Read Archive) under accession number PRJNA851406.

Bioinformatic analysis of the horse immunoglobulin heavy chain (IgH) variable-region repertoire

The reads were preprocessed by the pRESTO pipeline (vander Heiden et al., 2014), and the IG genes were annotated using IMG/HighV-QUEST (Alamyar et al., 2012). The unproductive V(D)J rearrangements were eliminated from the dataset, as well as the productive sequences containing insertions, deletions (indels), or stop codons in V- and J-gene segments. The sequences with the same VJ segment and identical CDR H3 size were grouped using the IMG/StatClonotype (Aouinti et al., 2016) for clonotype analyses.

After processing the sequences, analyses of the antibodies diversity were conducted, evaluating the frequencies of gene segment usage, gene subgroups, and the combination of V(D)J genes in each animal using IMG/StatClonotype. The size, composition, and amino acids groups (defined by Crooks et al., 2004) of CDR-H3 amino acid sequences (numbered according to IMG) (Lefranc et al., 2003) were analyzed using R studio. R studio was also used to get public repertoire antibodies defined as different horses containing antibodies with the same V and J and CDR3. To determine the reading frame (RF) of IGHD genes, we first determined the hydrophobicity index, according to Kyte-Doolittle scale, of each frame using R studio peptide package. The most hydrophilic reading frame was defined as RF1, the most hydrophobic as RF2, and the one hydrophobic with stop codons was defined as RF3 (Ivanov et al., 2002)..

Other parameters such as somatic hypermutation and junction were analyzed from data extracted from IMG/HighV-QUEST.

Rarefaction analysis and constructing species-richness curves clonotypes.

We used the program iNEXT (Hsieh et al., 2016) to subsample populations of clonotypes from immunoglobulin heavy chains that belonged to four horses based on the frequency of their occurrence in productive reads. iNEXT was also used to extrapolate beyond the number of experimentally observed productive reads that we might expect with additional sequencing.

Recon (Kaplinsky & Arnaout, 2016) was used to estimate the number of missing clonotypes in the immunoglobulin heavy-chain datasets

Statistical analysis

To compare antibody isotypes (IgG and IgM) differences, we used the Shapiro-Wilk test followed by the Mann-Whitney test.

For all analyses, the media of each horse-specific parameter followed by the average media of all four horses, and also the standard deviation of the average media of the 4 horses were used.

RESULTS

Restricted VJ gene usage in horse antibodies

We analyzed IgG and IgM variable heavy chains from four individual horses. Overall, the mean of raw reads per IgM samples was 1,082,148 (354,598- 1,558,035) and 347,302 (210,964-481,150) for IgG samples. We obtained 31-64% of productive reads and between 40,018 to 328,300 horse antibody clones (Table 1).

The EquCab3.0 horse's genome includes 21 IGHV, 16 IGHD, and 6 IGHJ functional gene segments, leading to 2,016 possible germline coding antibodies. However, in our study, we observed a strong preference for IGHV4 subgroup gene segments, where the IGHV4-21, IGHV4-29, and IGHV4-22 are used by 80% of the horse antibodies in both IgM and IgG isotypes (Figure 1A). In addition, only 13 (of which three are present in less than 0.1% of the antibodies) from 21 IGHV seem to be used in horse antibodies.

Similarly, IGHJ4 and IGHJ6 are the preferred J gene used by both IgM and IgG isotypes (Figure 1B). Interestingly, IGHJ6 is present in almost 60% of all horse antibody clones, showing a restricted use of IGHV and IGHJ gene segments. All the 16 functional IGHD genes seem to be used by horses' antibodies (Supplementary Figure 1).

The most frequent VDJ combinations used by horse antibodies were IGHV4-21, IGHD2-26, and IGHJ6-1, found in 2.8% (± 0.9) of the IgM and 2.6% (± 0.6) IgG isotypes (Figure 1C).

Our analysis also showed that more rare clones were observed in IgM samples than in IgG since the rarefaction curves do not begin to plateau, indicating that we were unlikely to capture this population's full diversity (Figures 2A and 2B). However, we were able to capture between 52 to 66% and between 62 to 82% of all IgM and IgG, respectively (Figure 2C), with no statistical difference between IgM and IgG horse antibodies diversity compared to Shannon's (Figure 2D) and Simpsons test (Figure 2E).

Public horse antibody repertoire is enriched in shorter CDR-H3

In this study, we observed that the four horses shared (public repertoire) shared only 0.05% (44 clones) of their IgM repertoire and 0.0099% of the IgG repertoire (4 clones) (Figures 3A and 3B). For the IgM public repertoire, most of the clones present the IGHV4-21 gene (77%) of the IGHV genes in the IgM public repertoire, while in the total repertoire, it represents approximately 32% (Figure 1A). In the case of IGHJ genes, we observed an increased gene usage of IGHJ4 (from

4% to 9%) and IGHJ5 (from 29 to 32%) in the public IgM repertoire (Figure 1B).

Interesting to note that more than 90% of the CDR-H3 found in the IgM public antibody repertoire presented only five amino acids length (Figure 3C). In general, the CDR-H3 size distribution of horses follows a bi-modal pattern, with sizes ranging from 4 to 51 amino acids residues with a median length of 16 residues for both IgM and IgG isotypes (Figure 3C).

Interestingly, polar amino acids such as glycine (G) and tyrosine (Y) are increased in IgM public repertoire, differently from the acidic aspartic (D) and glutamic acids (E) and the hydrophobic phenylalanine (F), tryptophan (W) and alanine (A) that decreases (Figure 3D).

Our results suggest that the horse IGHV repertoire appears to be derived from limited germline gene families.

Characterization of somatic hypermutation (SHM) frequency and pattern found in horse antibodies

Based on our previous results, we hypothesized that the biggest horse immunoglobulin diversity comes from somatic hypermutation and junctional diversity.

Here, we observed that the frequency of mutations in IgG isotype (media: 7.22%) was similar to IgM (media: 6.46%) compared to their germline mapped on EquCab3.0 genome (Figure 4A). The majority of mutations were found in CDR regions, especially at positions 32 (CDR1), 50, 52 and 58, from CDR2 and 88 (FR3) for both IgM and IgG isotypes (Figure 4B). We also observed that an average of 16.79% of nucleotides are mutated in CDRs of region IGHV, from which the majority of them (45 to 51%) are present in AID motifs (RGYW and complementary WRCY nucleotide motifs) (Supplementary Table 1).

Characterization of Horse Antibodies Junctional Diversity

An essential source of antibody diversification is the addition and deletion of nucleotides between VDJ junctions. Therefore, we analyzed the occurrence of the P/N nucleotide addition and exonuclease trimming for both IgM and IgG horse antibodies. We observed very similar characteristics in all the junctional regions of IgM and IgG horse antibodies (Figures 5A and 5B). N1 and N2 nucleotides media vary from 8.6 to 9.2 present in around 92 % of the IgM and IgG antibody clone (n = 998,756 clones for IgM, n = 264,566 clones for IgG (Figura 5A and 5B, Table 2). Interestingly, 43-44% of the antibodies have N1 (ranging from 10 to 111 nt) and N2

junctions (ranging from 10 to 104 nt) with 10 or more nucleotides, and 6.9 to 9.8% of these regions with 22 or more nucleotides (Table 2). It was also possible to observe that half of the 10 biggest CDR3 present cysteines (Supplementary table 4). We also noticed that N1 region is highly enriched in G (35.59%), and the N2 region is enriched in G and T (30%) for both isotypes (Supplementary Table 2).

Similarly, exonuclease trimming was observed in around 70% to 97% of the Ig clones, with the biggest number of nucleotides trimmed in the IGHJ gene ends (mean of 10 for IgM and 11 for IgG) (Figure 5A and 5B). When analyzing the components that contribute to the length of the CDR-H3, we observed that an average of 15 nucleotides from IGHD gene segment contribute to the length of the IgM and IgG CDR3s with contain an average of 54 nucleotides (around 26-27% of IGHD contribution to the CDR3). Surprisingly, when the 5 biggest CDR-H3 of the IgG clones where analysed we observed a contribution of 12 to 22 nucleotides of IGHD genes which represents only 9% to 12% of the CDR3. The biggest contribution in these cases came from N1 addition that can contribute with 111 nucleotides of a CDR3 with 150 nucleotides (74%) or the N2 addition that can contribute with 91 nucleotides of a CDR3 containing 129 nucleotides (70%) (Figure 5C).

We also observed a preference for the use of RF1 (more of 80%) in the horse's antibodies (Supplementary Figure 2), enriched in polar amino acids such as tyrosine and glycine (Table 2). Of the 96 possible sets of IGHD amino acid sequences ($16 \times 6 = 96$), 36 (37.5%) include one or more tyrosine, while only 13 (13.5%) have one or two arginines (Supplementary Table 3).

Discussion

In this work, we investigate the antibody-heavy chain repertoire of four different horses, presenting the largest collection of adaptive immune receptor sequences described to date for horses. We analyzed 40,018 to 328,300 horse IgG or IgM clones, with good coverage, from 52 to 82% of the repertoire. Similar to this work, it was observed a difference in depths for IgM (36%) when compared to IgG (64%) for human antibodies (Galson et al., 2015). Although not much difference was observed between IgM and IgG isotypes, this is the first high-throughput sequencing study that characterizes both isotype's horse repertoires.

Interestingly, we found that approximately 80% of the IgM and IgG antibodies present the IGHV4 group as a gene segment used in their antibodies, corroborating works from, Similar results were also observed by Tallmadge and collaborators (2013) using 5' RACE, in which they found a strong preference (80%) for the IGHV4-29 (previously called IGHV2S3) and IGHJ6-1 (55%) (previously called IGHJ1S5) usage in adult horses' antibodies (Chaudhary & Wesemann, 2018; Manso et al., 2019; Sun et al., 2010; Tallmadge et al., 2013). Similar, humans antibodies also have a preference for the IGHV4 family (Arnaout et al., 2011), differing from other organisms like cattle, dogs, and mice that present a predominance for IGHV1 genes in their antibodies and cats with the presence of IGHV3 genes (Pasman et al., 2017; Rettig et al., 2018; Steiniger et al., 2014, 2017). We also observed a predominance of IGHJ6 in horse antibodies, while dogs and cats antibodies se mostly IGHJ4, and mouse, the IGHJ1 group (Arnaout et al., 2011; Steiniger et al., 2014, 2017). It is interesting to note that horse antibody repertoire is highly dominated by only a few IGHV and IGHJ genes, even if they can use all of their theoretical germline combinations. We observed a high frequency of antibodies ($2.8 \pm 0.9\%$) containing IGHV4-21, IGHD2-26, and IGHJ6-1 gene segments combination, also identified in previous work on non-immunized horses (Manso et al., 2019; Tallmadge et al., 2013). This result

is not different from human antibody repertoires (Arnaout et al., 2011), where 0.1% to 2.7% of sequences have the same V(D)J combinations.

In addition, even for a few clones, we observed the presence of a public horse antibody repertoire in the absence of any specific immune stimulation. We found more clones in the public IgM (0.05%) repertoire than in the public IgG repertoire (0.009%). The small number of shared public antibodies clones can be due to the high diversity of horses antibodies but can also be due to an artifact of the clonotyping method used in this work that considers the same clone only antibodies with identical CDH3 region (IMGT/HighVQuest). The observation of more public IgM (1.4%) than IgG (0.3%) and IgA (0.5%) was also observed in the human antibody (Galson et al., 2015). Interestingly, even with a smaller number of public antibodies, we found differences between the CDR-H3 length of the public and the entire repertoire, observing a higher percentage of short CDR-H3 in the public repertoire, also observed in human (Briney et al., 2019; Galson et al., 2014; Soto et al., 2019). It is supposed that B cells expressing receptors with short CDR-H3 are selected because they increase their affinity for the antigen, make clonal expansion, and differentiate in plasma cells or memory B cells (Rosner et al., 2001).

In this work, we also evaluated the SHMs in the IGHV gene segment of the animals (FR1, CDR1, FR2, CDR2, and FR3 regions). The IgG sequences showed a similar mutation frequency than the IgM sequences, probably due to limited pathogen stimulation since they are not immunized. To better understand how mutations are distributed along the IGHV gene segment, we evaluated the number of mutations present in each position for both studied Ig isotypes. We observed a similar mutation profile between the IgG and IgM, with more mutations found in the CDR regions, even in these non-immunized animals. This corroborates previous studies in adult horses (Tallmadge et al., 2013) and healthy and HIV patients (Bowers et al., 2014).

The mechanism of variability that produces SHM is carried out by the enzyme cytidine deaminase (AID) induced by activation, by deamination of the cytosine base, creating a U:G mismatch. The AID targets SHM mutations on "hotspots" (complementary RGYW and WRCY nucleotide motifs) (Spencer & Dunn-Walters, 2005). This work observed that around 19% of IGHV sequences present AID motifs, similar to human IGHV, presenting 17.8% of these motifs (Bowers et al., 2014). In addition, as in the human repertoires, we found a higher percentage of

motifs in the CDR (average of 13.5%) than in the FR (average of 5.2%), as well as a higher number of mutated nucleotides in this region (Bowers et al., 2014).

It is important to highlight that, besides a very similar VJ gene usage in IgM and IgG antibodies, and also a very similar profile of SHM, the IgM and IgG clones look very dissimilar according to Bray-Curtis dissimilarity index (data not shown).

Very little has been described the characteristics of horse junctional diversity in horse antibodies. Here, we observed N1 and N2 nucleotide additions in most IgM and IgG clones, and also observed exceptionally long N nucleotide additions for both N1 and N2 IgG (10–111 bp) in around 44% of the antibodies. Non-template additions to IGH genes have been reported in humans and mice (Shi et al., 2014), pigs (Šinkora et al., 2003), and cattle (Liljavirta et al., 2014). The mean number of nucleotide addition in humans is 6.6 ± 4.3 in N1 and 6.4 ± 4.6 in N2, while in mice, it is 2.4 ± 2.2 and 2.1 ± 1.8 in N1 and N2, respectively (Shi et al., 2014). The average number of nucleotide additions in cattle that present ultralong CDR3s is not particularly high (2.5 in N1 and 2.6 in N2), reflecting the high frequency (35%) of unions with zero additions (Liljavirta et al., 2014). Such frequency and length of N additions have not been reported in other species, suggesting that this diversity mechanism is essential to generating variability in equine immunoglobulins.

Extensive trimming of IGHD genes in horse antibodies (mean value of 6 and 7.9 nucleotides for the 3D and 5D junction, respectively) observed is not very different from cattle antibodies (trimming of 5 to 6 nucleotides) (Liljavirta et al., 2014). It is interesting to note that the larger trimming followed in horse antibodies was observed in the IGHJ gene, ranging between 10 and 11 nucleotides, while in other species, such as cattle, humans, and mice, have respectively 2, 6, and 4 nucleotides trimmed in this region (Liljavirta et al., 2014, Shi et al., 2014). In our data, between 70% to 97% of antibody clones had nucleotides deleted anywhere in the junction, similar to other species (Liljavirta et al., 2014; Shi et al., 2014).

This impressive N nucleotide addition frequency and length can be due to differences in Terminal deoxynucleotidyl transferase (TdT) enzyme activity in horses compared to other species.

These enzymes are composed of mainly two regions, a catalytic core composed of finger, palm, and thumb domains at the C-terminus and a BRCA1 (breast cancer susceptibility protein) C-terminal (BRCT) domain at the N-terminus. When comparing the horse sequence of TdT

isorform 1 and 2 with human, mouse, pig, and cattle TdT we observed the conservation of the catalytic aspartic acids and the substrate-specific loop 1 (Figure 6). Interestingly, the palm domain region, between the first aspartic acids and the loop1, is one of the most different regions between the TdT from other species.

In addition to this region, we can also observe a very dissimilar region between the BRCT domain and TdT catalytic core. For the best of all knowledge, it is unclear how this non-enzymatic domain contributes to the unique biological function of TdT. Interesting to note that this interdomain region is enriched in Proline amino acids. It looks like for DNA polymerase lambda, which presents a bigger Proline-rich domain in this region compared to TdT, this region can impact DNA polymerase fidelity and with BRCT domain can act cooperatively to promote primer/template realignment between DNA strands of limited sequence homology (Fiala et al., 2006; Taggart et al., 2014). Since the TdT template and untemplated activities (Loc'h et al., 2016) are proposed to be essential for antibodies diversity, future studies need to investigate the role of the interdomain region in these activities, as well as the role of the region in palm domain in between aspartic acids and loop1.

This study also observed that more of 80% of the horse antibodies use reading frame 1 (RF1) for both the IgM and IgG isotypes. Several species, such as humans, mice, and sharks, produce antibodies using IGHD reading frame 1 (RF1) (Schroeder et al., 2010). However, similarly to other species RF1 used for horses antibodies is strongly enriched in tyrosine, representing 39.55% of the IGHD amino acids. We know that tyrosine is the amino acid that typically makes the most significant contribution to binding affinity at protein ligand-receptor interfaces (Bogan & Thorn, 1998). This suggests that natural selection was operating on immunoglobulin diversity gene segments to restrict and control their evolution in such a way as to influence the composition and range of diversity of immunoglobulin antigen-binding sites (Burnet, 1976).

Conclusions

This is the first high-throughput sequencing study that characterizes IgM and IgG isotype horse repertoire to the best of our knowledge. We showed a highly restrict use of IGHV and IGHJ genes in horse antibody repertoire in which around 80% of the antibodies are composed by only 3 IGHV gene segments and almost 60% of them with the same IGHJ gene segment. We

observed a complex and diverse repertoire for IGH, given mainly by the junctional diversity, much bigger and frequent than the one present in other species. Our study on the equine antibody repertoire contributes to understanding the generation of their diversity and open up new questions about horse TdT particularities to generate such diversity.

Author Contributions

CN: designed the study, did all the experiments, analyzed the data and wrote the paper; TM: designed and validated primers, designed the study and discussed the results; FM: help in data analysis; LM: helped in the initial analysis of the data; RM: helped Illumina library preparation and sequenced the samples; JM, BA: collected horse samples; AV: discussed the results; JMD, GI: help primer design, study design and wrote the paper; LFF: designed the study, discussed and analyzed the data and wrote the paper

Conflict of Interest

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

Fundings

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) [grant numbers 88887.506611/2020-00, 88887.504420/2020-00 and 935/19 (COFECUB)]; Fundação de Amparo a Pesquisa de Minas Gerais (FAPEMIG) [grant numbers PPM-00615-18, Rede Mineira de Imunobiológicos grant #REDE-00140-16]; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [Pq to LFF]; National Institutes of Health (NIH) [grant number 1R01AI143552-02]; Pro-Reitoria de Pesquisa da Universidade Federal de Minas Gerais.

Acknowledgments

SynBiom group for fruitful discussions, specially Dr. Marcella Nunes de Mello-Braga and Dra. Marcele Rocha Neves Rocha. A special acknowledge to Regina Maria Fernandes for project management.

References

- Alamyar, E., Giudicelli, V., Li, S., Duroux, P., & Lefranc, M. P. (2012). IMGT/Highv-quest: The IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Research*, 8(1), 1–15. <https://doi.org/10.4172/1745-7580.1000056>
- 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(11), 1836–1845. <https://doi.org/10.1016/j.molimm.2005.10.017>
- ANDERSON, C. G. (1955). The distribution of diphtheria antitoxin in pepsin-digested horse antiserum. *The Biochemical Journal*, 59(1), 47–52. <https://doi.org/10.1042/bj0590047>
- Aouinti, S., Giudicelli, V., Duroux, P., Malouche, D., Kossida, S., & Lefranc, M. P. (2016). IMGT/statclonotype for pairwise evaluation and visualization of NGS IG and TR IMGT clonotype (AA) diversity or expression from IMGT/HighV-QUEST. *Frontiers in Immunology*, 7(SEP), 1–14. <https://doi.org/10.3389/fimmu.2016.00339>
- Arnaout, R., Lee, W., Cahill, P., Honan, T., Sparrow, T., Weiland, M., Nusbaum, C., Rajewsky, K., & Korolov, S. B. (2011). High-resolution description of antibody heavy-chain repertoires in humans. *PLoS ONE*, 6(8). <https://doi.org/10.1371/journal.pone.0022365>
- Bogan, A. A., & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, 280(1), 1–9. <https://doi.org/10.1006/jmbi.1998.1843>
- Bowers, E., Scamurra, R. W., Asrani, A., Beniguel, L., MaWhinney, S., Keays, K. M., Thurn, J. R., & Janoff, E. N. (2014). Decreased mutation frequencies among immunoglobulin G variable region genes during viremic HIV-1 infection. *PLoS ONE*, 9(1), 1–13. <https://doi.org/10.1371/journal.pone.0081913>
- Briney, B., Inderbitzin, A., Joyce, C., & Burton, D. R. (2019). Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature*, 566(7744), 393–397. <https://doi.org/10.1038/s41586-019-0879-y>
- Burnet, F. M. (1976). A Modification of Jerne's Theory of Antibody Production using the Concept of Clonal Selection. *CA: A Cancer Journal for Clinicians*, 26(2), 119–121. <https://doi.org/10.3322/canjclin.26.2.119>
- Butler, J. E., Wertz, N., Deschacht, N., & Kacs Kovics, I. (2009). Porcine IgG: Structure, genetics, and evolution. *Immunogenetics*, 61(3), 209–230. <https://doi.org/10.1007/s00251-008-0336-9>
- Chaudhary, N., & Wesemann, D. R. (2018). Analyzing immunoglobulin repertoires. *Frontiers in Immunology*, 9(MAR), 1–18. <https://doi.org/10.3389/fimmu.2018.00462>
- Cole, R., & Moore, H. F. (1917). The production of antipneumococcic serum. *Journal of Experimental Medicine*, 26(4), 537–561. <https://doi.org/10.1084/jem.26.4.537>
- Crooks, G., Hon, G., Chandonia, J., & Brenner, S. (2004). NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Res*, 14, 1188–1190. <https://doi.org/10.1101/gr.849004.1>

- Cunha, L. E. R., Stolet, A. A., Strauch, M. A., Pereira, V. A. R., Dumard, C. H., 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., Zingali, R. B., Oliveira, G. A. P., Souza, T. M. L., Tanuri, A., Gomes, A. M. O., ... Silva, J. L. (2020). Equine hyperimmune globulin raised against the SARS-CoV-2 spike glycoprotein has extremely high neutralizing titers. *BioRxiv*. <https://doi.org/10.1101/2020.08.17.254375>
- Delarue, M., Boule J.B, Lescar J, Expert-Bezancacon N, Jourdan N, Sukumar N, Rougeon F, & Papanicolaou C. (2002). Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. *The EMBO Journal*, 21, 427–439. <https://doi.org/10.1093/emboj/21.3.427>
- Fiala, K. A., Duym, W. W., Zhang, J., & Suo, Z. (2006). Up-regulation of the fidelity of human DNA polymerase λ by its non-enzymatic proline-rich domain. *Journal of Biological Chemistry*, 281(28), 19038–19044. <https://doi.org/10.1074/jbc.M601178200>
- Galson, J. D., Pollard, A. J., Trück, J., & Kelly, D. F. (2014). Studying the antibody repertoire after vaccination: Practical applications. *Trends in Immunology*, 35(7), 319–331. <https://doi.org/10.1016/j.it.2014.04.005>
- Galson, J. D., Trück, J., Fowler, A., Münz, M., Cerundolo, V., Pollard, A. J., Lunter, G., & Kelly, D. F. (2015). In-depth assessment of within-individual and inter-individual variation in the B cell receptor repertoire. *Frontiers in Immunology*, 6(OCT), 1–13. <https://doi.org/10.3389/fimmu.2015.00531>
- 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. *Nature Biotechnology*, 32(2), 158–168. <https://doi.org/10.1038/nbt.2782>
- Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G., Mehta, G. R., Ni, I., Mei, L., Sundar, P. D., Day, G. M. R., Cox, D., Rajpal, A., & Pons, J. (2009). Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20216–20221. <https://doi.org/10.1073/pnas.0909775106>
- Glatman-Freedman, A., & Casadevall, A. (1998). Serum therapy for tuberculosis revisited: Reappraisal of the role of antibody-mediated immunity against *Mycobacterium tuberculosis*. *Clinical Microbiology Reviews*, 11(3), 514–532. <https://doi.org/10.1128/cmr.11.3.514>
- Gonçalves, E. S., Salomão, M. G., & Almeida-santos, S. M. de. (2007). O uso do monitoramento espaço-temporal da expansão urbana no diagnóstico de áreas passíveis de risco epidemiológico peçonhento em Guarulhos-Estado de São Paulo, Brasil. *Anais Do XIII Simpósio Brasileiro de Sensoriamento Remoto*, 3171–3178.
- Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, 7(12), 1451–1456. <https://doi.org/10.1111/2041-210X.12613>
- Ivanov I, Link J., Ippolito G.C., Schoroder, H. W. J. (2002). Constraints on the Hydropathicity of HCDR3 are conserved across evolution. Chapter 3, 43-67. *The Antibodies* (Zaneti Maurizio &

- Capra Donald, Eds.; Vol 7).
- Kaplinsky, J., & Arnaout, R. (2016). Robust estimates of overall immune-repertoire diversity from high-throughput measurements on samples. *Nature Communications*, 7(May). <https://doi.org/10.1038/ncomms11881>
- Lang, J., Kamga-Fotso, L., Peyrieux, J. C., Blondeau, C., Lutsch, C., & Forrat, R. (2000). Safety and immunogenicity of a new equine tetanus immunoglobulin associated with tetanus-diphtheria vaccine. *American Journal of Tropical Medicine and Hygiene*, 63(5–6), 298–305. <https://doi.org/10.4269/ajtmh.2000.63.298>
- Lefranc, M. P., Pommié, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L., Thouvenin-Contet, V., & Lefranc, G. (2003). IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Developmental and Comparative Immunology*, 27(1), 55–77. [https://doi.org/10.1016/S0145-305X\(02\)00039-3](https://doi.org/10.1016/S0145-305X(02)00039-3)
- Liljavirta, J., Niku, M., Pessa-Morikawa, T., Ekman, A., & Iivanainen, A. (2014). Expansion of the preimmune antibody repertoire by junctional diversity in *Bos taurus*. *PLoS ONE*, 9(6). <https://doi.org/10.1371/journal.pone.0099808>
- Loc'h, J., Rosario, S., & Delarue, M. (2016). Structural Basis for a New Templated Activity by Terminal Deoxynucleotidyl Transferase: Implications for V(D)J Recombination. *Structure*, 24(9), 1452–1463. <https://doi.org/10.1016/j.str.2016.06.014>
- Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, 48(D1), D265–D268. <https://doi.org/10.1093/nar/gkz991>
- 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(August 2018), 251–259. <https://doi.org/10.1016/j.molimm.2018.11.017>
- Pasman, Y., Merico, D., & Kaushik, A. K. (2017). Preferential expression of IGHV and IGHD encoding antibodies with exceptionally long CDR3H and a rapid global shift in transcriptome characterizes development of bovine neonatal immunity. *Developmental and Comparative Immunology*, 67, 495–507. <https://doi.org/10.1016/j.dci.2016.08.020>
- Rettig, T. A., Ward, C., Bye, B. A., Pecaut, M. J., & Chapes, S. K. (2018). Characterization of the naive murine antibody repertoire using unamplified high-throughput sequencing. *PLoS ONE*, 13(1), 1–20. <https://doi.org/10.1371/journal.pone.0190982>
- Reynaud, C. A., Dahan, A., Anquez, V., & Weill, J. C. (1989). Somatic hyperconversion diversifies the single VH gene of the chicken with a high incidence in the D region. *Cell*, 59(1), 171–183. [https://doi.org/10.1016/0092-8674\(89\)90879-9](https://doi.org/10.1016/0092-8674(89)90879-9)
- Rio, D. C., Ares, M., Hannon, G. J., & Nilsen, T. W. (2010). Purification of RNA using TRIzol (TRI Reagent). *Cold Spring Harbor Protocols*, 5(6), 1–4. <https://doi.org/10.1101/pdb.prot5439>

- Rosner, K., Winter, D. B., Tarone, R. E., & Skovgaard, G. L. (2001). Third complementarity-determining region of mutated V H immunoglobulin genes contains shorter V , D , J , P , and N components than non-mutated genes.
- Schroeder, H. W., Zemlin, M., Khass, M., Nguyen, H. H., & Schelonka, R. L. (2010). Genetic control of DH reading frame and its effect on B-cell development and antigen-specific antibody production. *Critical Reviews in Immunology*, 30(4), 327–344. <https://doi.org/10.1615/critrevimmunol.v30.i4.20>
- Shi, B., Ma, L., He, X., Wang, X., Wang, P., Zhou, L., & Yao, X. (2014). Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST. *Theoretical Biology and Medical Modelling*, 11(1), 1–11. <https://doi.org/10.1186/1742-4682-11-30>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7. <https://doi.org/10.1038/msb.2011.75>
- Šinkora, M., Sun, J., Šinkorová, J., Christenson, R. K., Ford, S. P., & Butler, J. E. (2003). Antibody Repertoire Development in Fetal and Neonatal Piglets. VI. B Cell Lymphogenesis Occurs at Multiple Sites with Differences in the Frequency of In-frame Rearrangements. *The Journal of Immunology*, 170(4), 1781–1788. <https://doi.org/10.4049/jimmunol.170.4.1781>
- Soto, C., Bombardi, R. G., Branchizio, A., Kose, N., Matta, P., Sevy, A. M., Sinkovits, R. S., Gilchuk, P., Finn, J. A., & Crowe, J. E. (2019). High frequency of shared clonotypes in human B cell receptor repertoires. *Nature*, 566(7744), 398–402. <https://doi.org/10.1038/s41586-019-0934-8>
- Spencer, J., & Dunn-Walters, D. K. (2005). Hypermutation at A-T Base Pairs: The A Nucleotide Replacement Spectrum Is Affected by Adjacent Nucleotides and There Is No Reverse Complementarity of Sequences Flanking Mutated A and T Nucleotides. *The Journal of Immunology*, 175(8), 5170–5177. <https://doi.org/10.4049/jimmunol.175.8.5170>
- 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(1), 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., Liu, Z., Ren, L., Wei, Z., Wang, P., Li, N., & Zhao, Y. (2012). Immunoglobulin genes and diversity: What we have learned from domestic animals. *Journal of Animal Science and Biotechnology*, 3(1), 1–5. <https://doi.org/10.1186/2049-1891-3-18>
- 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 and Comparative Immunology*, 34(9), 1009–1020. <https://doi.org/10.1016/j.dci.2010.05.003>

- Taggart, D. J., Dayeh, D. M., Fredrickson, S. W., & Suo, Z. (2014). N-terminal domains of human DNA polymerase lambda promote primer realignment during translesion DNA synthesis. *DNA Repair*, 22, 41–52. <https://doi.org/10.1016/j.dnarep.2014.07.008>
- Tallmadge, R. L., Tseng, C. T., & Felipe, M. J. B. (2014). Diversity of immunoglobulin lambda light chain gene usage over developmental stages in the horse. *Developmental and Comparative Immunology*, 46(2), 171–179. <https://doi.org/10.1016/j.dci.2014.04.001>
- Tallmadge, R. L., Tseng, C. T., King, R. A., & Felipe, M. J. B. (2013). Developmental progression of equine immunoglobulin heavy chain variable region diversity. *Developmental and Comparative Immunology*, 41(1), 33–43. <https://doi.org/10.1016/j.dci.2013.03.020>
- 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(13), 1930–1932. <https://doi.org/10.1093/bioinformatics/btu138>
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Zylberman, V., Sanguineti, S., Pontoriero, A. v., Higa, S. v., Cerutti, M. L., Seijo, S. M. M., Pardo, R., Muñoz, L., Intriери, M. E. A., 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., ... Goldbaum, F. A. (2020). Development of a hyperimmune equine serum therapy for covid-19 in Argentina. *Medicina*, 80, 1–6.

Tables

Table 1: Overview of the IgM and IgG heavy chain variable sequencing results from four non-immunized horses

Horse Sample	Ig Isotype	Raw Reads	Preprocessed Reads	Annotated Reads	Productive Reads	Clones
1	IgM	1,558,035	779,236	526,221	484,236	294,600
2	IgM	1,499,997	900,306	891,250	790,308	328,300
3	IgM	915,963	690,915	529,584	485,888	283,600
4	IgM	354,598	268,020	203,623	193,644	86,177
1	IgG	210,964	187,229	182,918	175,686	43,861
2	IgG	288,611	233,147	219,402	209,248	40,018
3	IgG	408,483	306,844	304,394	271,204	91,136
4	IgG	481,141	390,341	382,438	361,320	89,551

Table 2: Analysis of nucleotide additions in Horse Antibodies

Sample	Number of clones	Mean number of N1 nucleotides (range)	Mean number of N2 nucleotides (range)	Clones with longer nucleotides/and than 22 nt (%)	Clones with 10 additions longer nucleotides/and than 22 nt (%)	N2 longer than 10 nucleotides/and longer than 22 nt (%)
IgM	998,756	9.54 (0-78)	9.71 (0-62)	43.95/7.4	43.74/8.6	
IgG	264,566	8.72(0-111)	9.24 (0-104)	43.67/7.6	45.17/10.0	

Table 3: Amino acid percentage composition per reading frame of horse D gene

	RF1	RF2	RF3	iRF1	iRF2	iRF3	
D	5.22	0.75	0	0	0	0	Acid
E	0.74	0	2.90	0	0	3.00	
R	0.74	1.50	4.34	5.97	0.74	0	Basic
K	0.74	0	0	0.74	0	0	
H	0	0	0.72	21.64	1.49	0	
F	0	0	0.72	0	0	0.75	Hydrophobic
P	0	2.25	0	2.24	4.48	17.29	
V	0	27.82	0.72	0.74	29.10	3.00	
L	0	5.26	42.75	1.49	1.49	18.04	
I	2.23	6.77	1.44	0.74	18.66	3.00	
A	5.97	3.00	1.45	3.54	5.26	2.93	
W	3.73	0	13.76	0	0	0	
M	0	18.04	0	0	0	0.75	
N	5.97	6.66	0.72	15.67	2.99	0	Neutral
Q	0	0	2.90	1.49	0	4.51	
C	0.74	0	5.80	5.97	0	2.94	Polar
S	14.92	0	0	23.88	4.48	6.77	
G	18.65	4.5	1.45	0	5.97	1.50	
Y	39.55	0	2.90	11.94	0.74	2.25	
T	0.74	29.32	0	0.74	26.12	1.50	
*	0	0	17.39	4.47	0.74	32.33	
AA Media by RF	7.88	7.82	8.11	7.88	7.88	7.82	

Figures

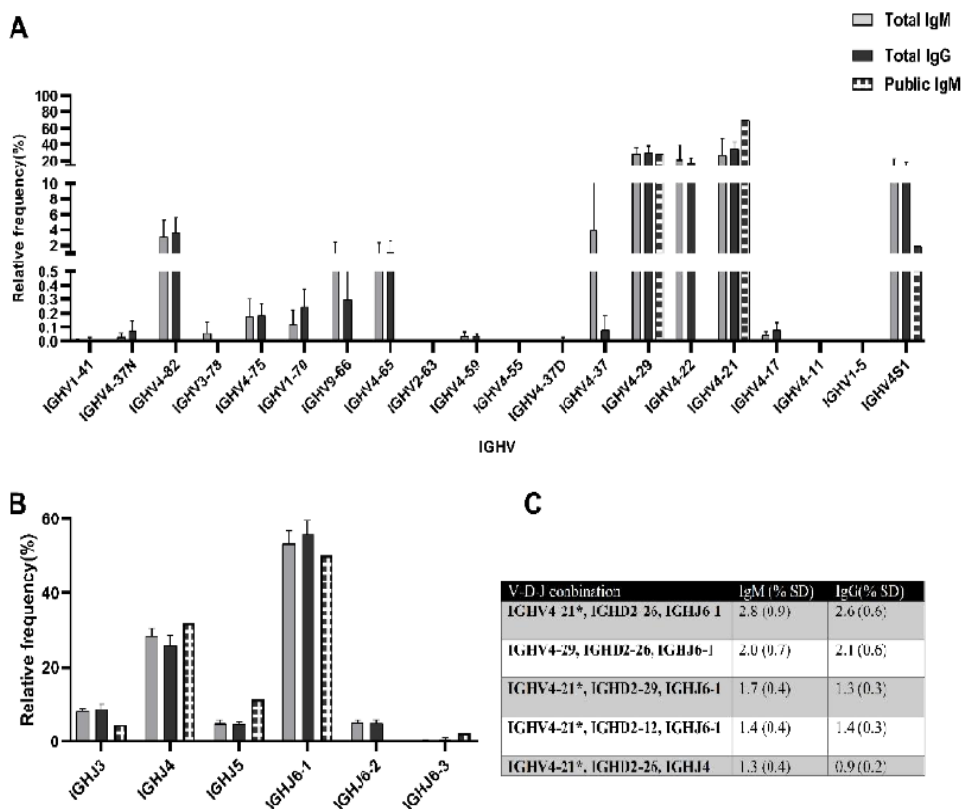


Figure 1: IGHV and IGHJ gene segments frequency present in public IgM and total IgM and IgG horses' antibodies.

Median of relative frequency (%) of IGHV (A), IGHJ (B) and V(D)J more frequent combination (C) in IgM and IgG isotype from four horses. The genes are organized in the order that it appears in the EquCab 3.0 genome (5'-3').

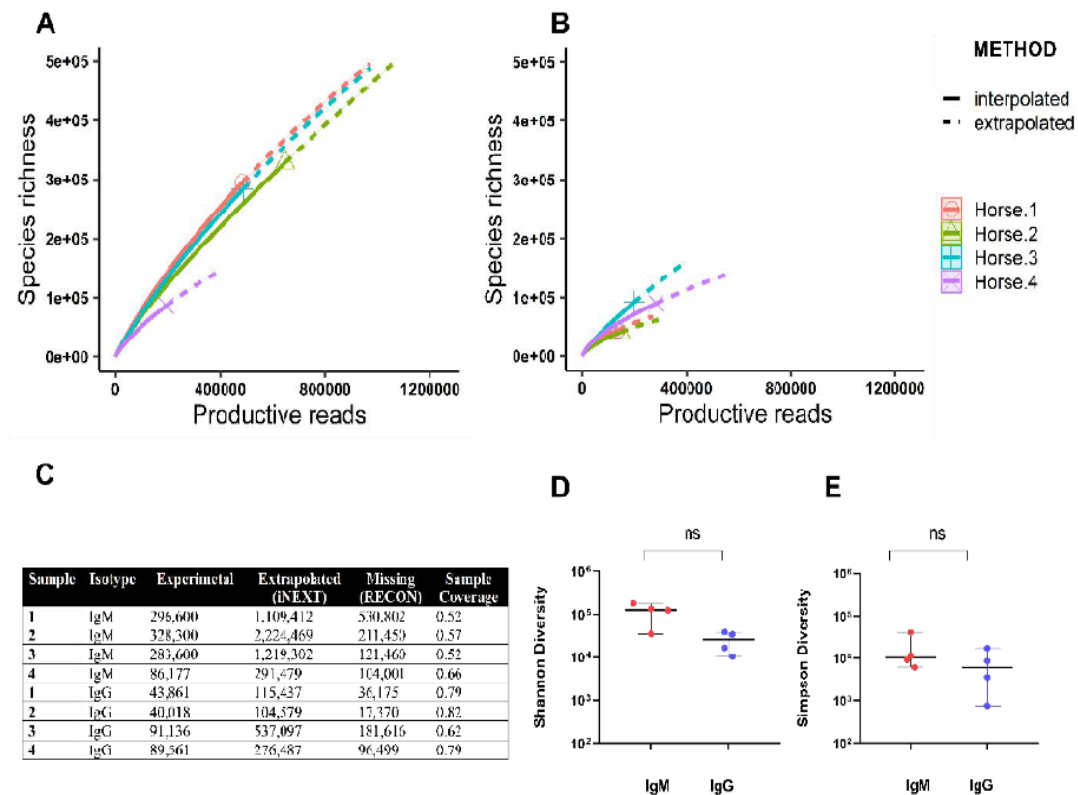


Figure 2: Antibody heavy chain repertoire richness and diversity estimates for IgM and IgG in the four non-immunized horses.

Interpolation and extrapolation of species richness were obtained using iNEXT for IgM (A) and IgG (B). Solid lines correspond to the interpolation (based on experimental data), and the dashed lines belong to the extrapolated data. Summary of estimates for repertoire size, including missing clones (C). The comparison of Shannon's (D) and Simpson's (E) diversity between the IgM and IgG isotypes ($p < 0.05$).

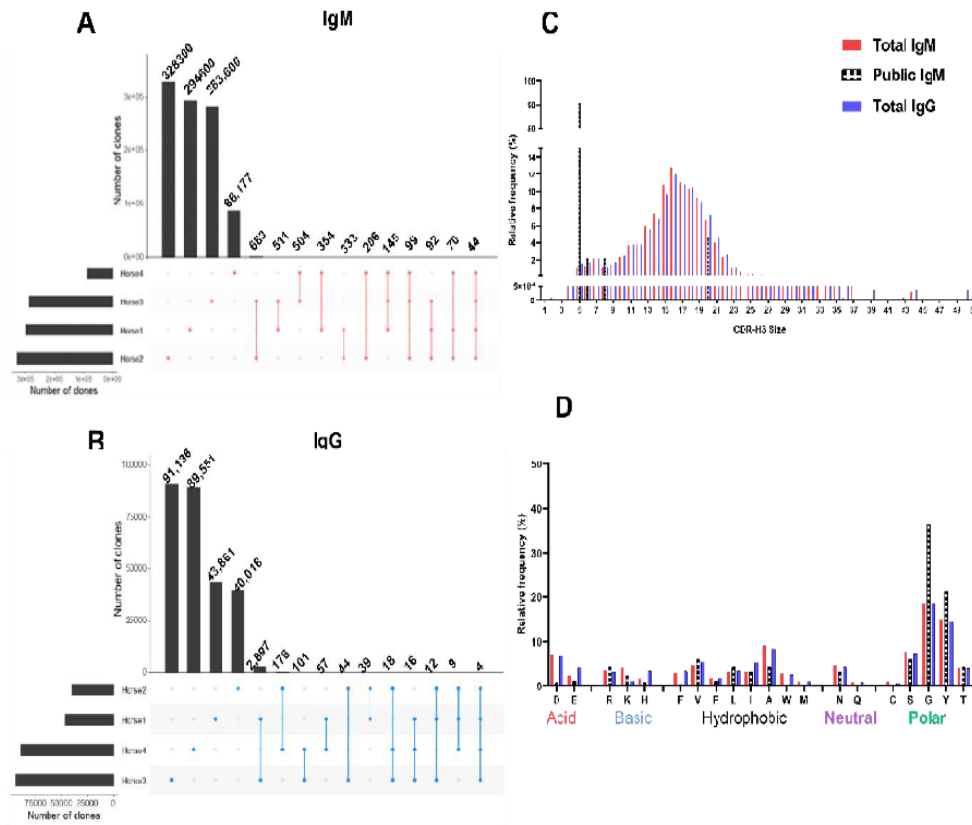


Figure 3: Horse public and private heavy chain variable region repertoire.

The number of antibody clones presented by the different horses and the shared number of clones between the 2, 3, or 4 horses in this study for the IgM (A) and IgG (B). Comparison of the CDR-H3 length (C) and amino acid composition (D) between the Total IgM, Total IgG, and the public IgM repertoire.

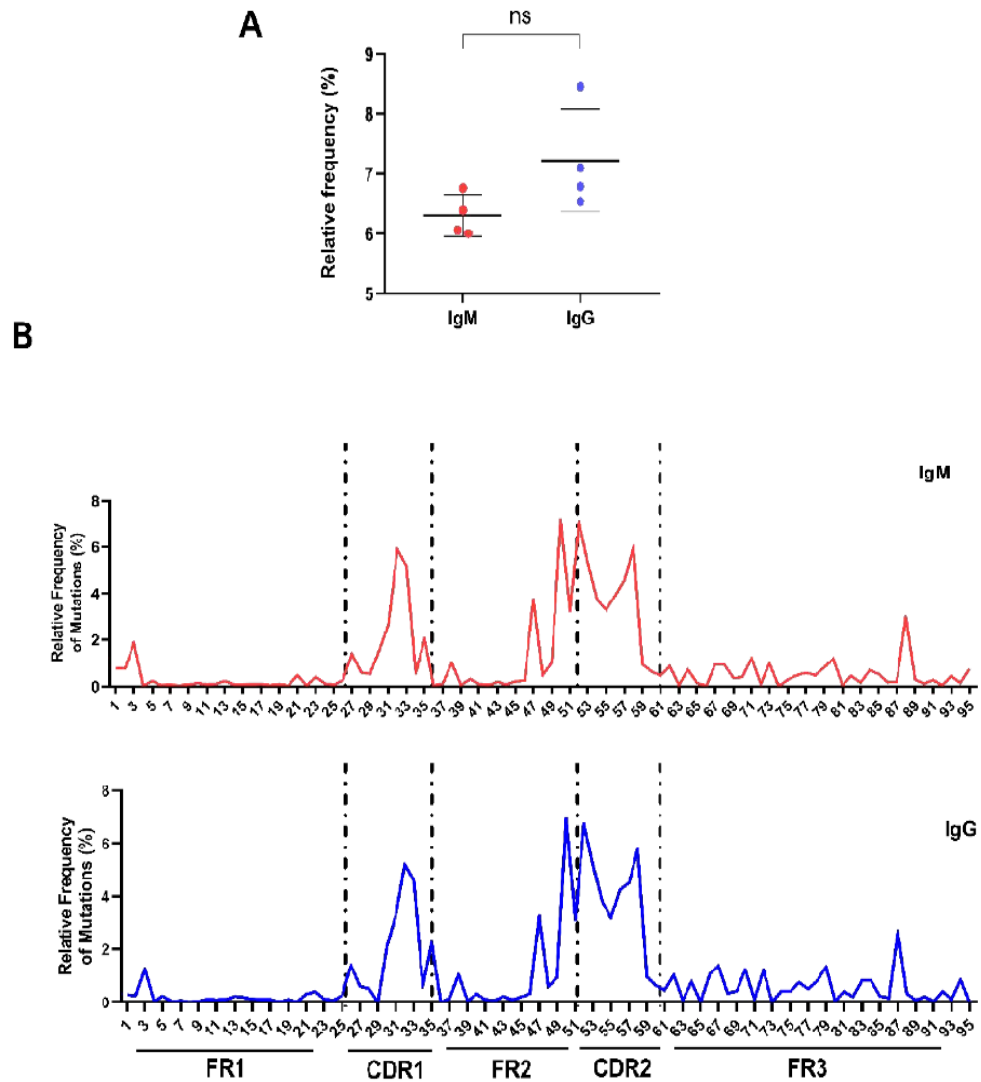


Figure 4: Somatic Hypermutation (SHM) characteristics of Horse IgG and IgM heavy chain variable region.

(A) Media of SHM frequency (%) at the IGHV gene segment from IgM and IgG isotypes ($p < 0.05$). (B) The number of mutations by amino acid position in the IGHV gene segment of the horses' heavy chain (According to the IMGT numbering, without gaps). The dotted lines delimit the FR and CDR regions. IgM is shown in red and IgG in blue.

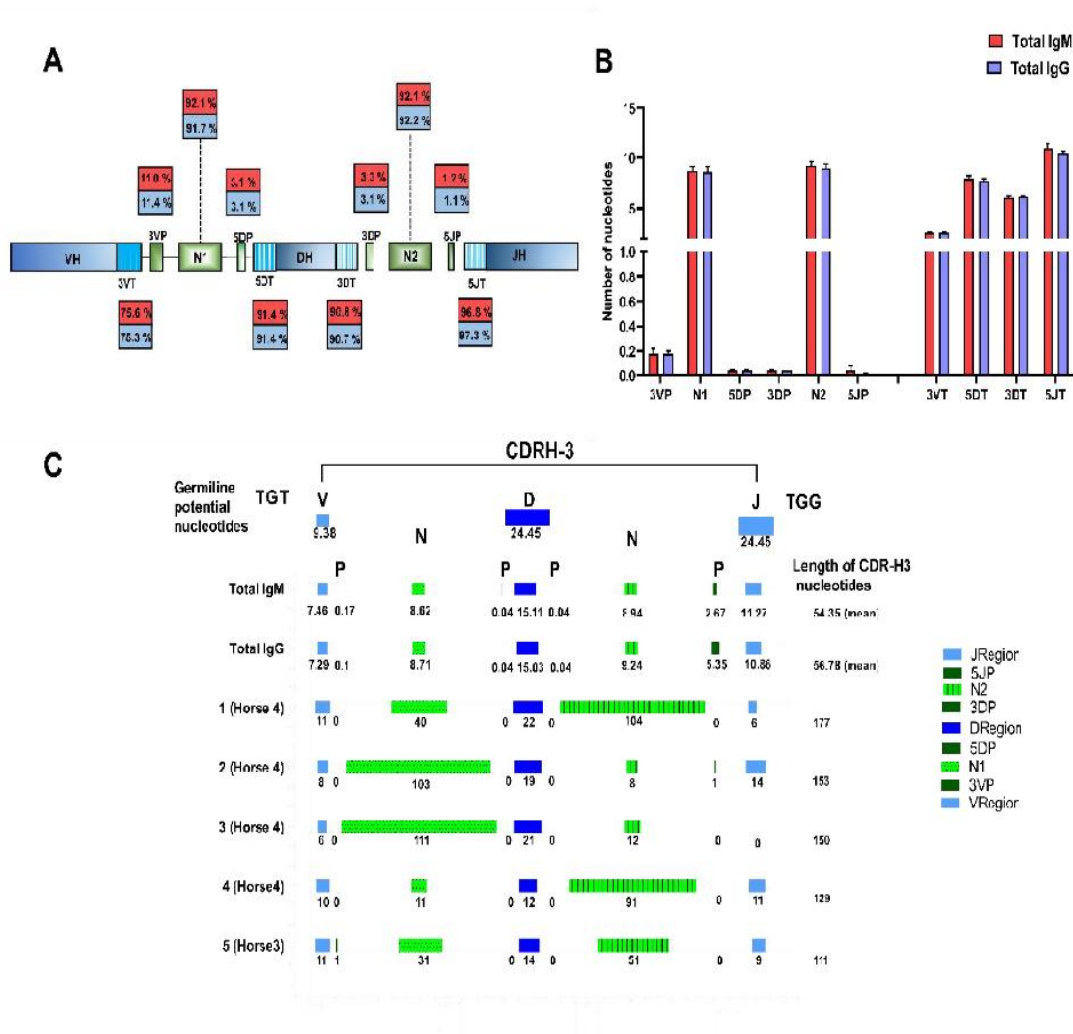


Figure 5: VDJ junction analysis.

(A) Junctional modifications schema during VDJ rearrangement, showing the locations and the occurrence of different types of junctional modifications. Into the box are the mutation frequency (%) of each junctional segment in all the antibodies clones analyzed: 3VP and 3VT for 3'V region, 5DP and 5DT for 5'D genes, 3DP and 3DT for 3'D region, and 5JP and 5JT for 5'J genes, where P means palindromic nucleotides additions, and T means exonuclease trimmings; N1 the non-template randomized nucleotides additions at the 3'V and the 5'D genes; N2 for N additions at the 3'D and the 5'J genes. (B) The median number of nucleotides per junction region added or trimmed. (C) Deconstruction of the components that contribute to the length of the CDR-H3 in

the media of total of the IgM and IgG clones, as well as the 5 biggest CDR-H3 of the IgG clones. The mean of nucleotides of the germline sequence of the VH gene segment, P and N junctions, the DH gene segment, and the JH gene segment to the CDR-H3 length is illustrated.

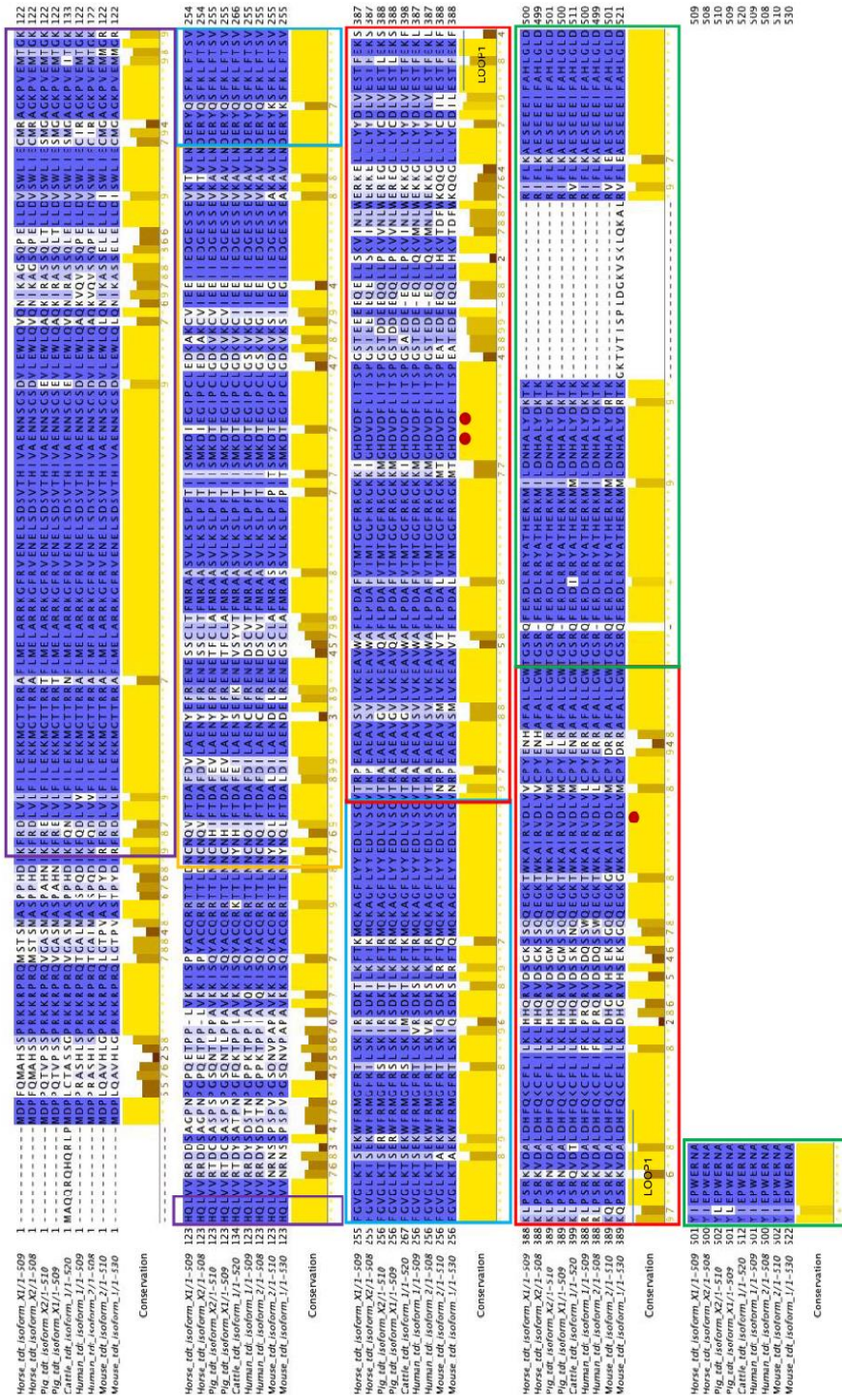


Figure 6: Multiple Sequence alignment of TdI from different species. An amino acid sequence alignment of the two equine TdI isoforms with other vertebrate TdI was done using the Clustal Omega program (Sievers et al., 2011) present in Jalview package (Waterhouse et al., 2009) that also provided Conservation analysis. The different box colors represent TdI domains, BRCT domain annotated by CDD (Lu et al., 2020) and TdI

catalytic core as described (Delarue et al., 2002) Cliquez ou toque aqui para inserir o texto. : Purple: BRCA1 C Terminus (BRCT) domain (CL0459), Yellow: Helix-hairpin-helix domain (HHH_8), Blue: Fingers domain of DNA polymerase lambda, Red: DNA polymerase beta palm, highlighting the 3 catalytic Aspartic Acids (red circle) and loop1, Green: DNA polymerase beta thumb.

NCBI Reference Sequences: Horse_tdt_isoform_X1 (XP_005602408.1); Horse_tdt_isoform_X2 (XP_001501812.3); Pig_tdt_isoform_X2 (XP_003133204.1); Pig_tdt_isoform_X1 (XP_005671421.1); Cattle_tdt_isoform_1 (NP_803461.1); Human_tdt_isoform_1 (NP_004079.3); Human_tdt_isoform_2 (NP_001017520.1); Mouse_tdt_isoform_2 (NP_001036693.1); Mouse_tdt_isoform_1 (NP_033371.2)

Supplementary Table 1: AID (RGYW/WRCY) motifs and targeted mutation frequencies in CDR and FR regions

	IgM	IgG
Number of RGYW/WRCY motifs per IGHV segment	19 (1.18)	20 (1.17)
% of CDR nucleotides mutated	16.79	18.81
% of CDR mutations present in RGYW/WRCY motifs	51.61	45.85
% of FR nucleotides mutated	4.24	4.71
% of FR mutations present in RGYW/WRCY motifs	16.92	18.76
% of all nucleotides mutated	6.37	7.12
% of all mutations present in RGYW/WRCY motifs	24.20	25.59

Supplementary Table 2: Percentage of nucleotides present at the N1 and N2 junctions of horse IgM and IgG antibodies

	%A	%T	%G	%C	
IgM	24.82	20.21	35.59	19.37	N1.REGION
IgG	24.28	21.47	34.90	19.35	
IgM	21.29	28.20	30.32	20.19	N2.REGION
IgG	21.57	29.60	28.87	19.96	

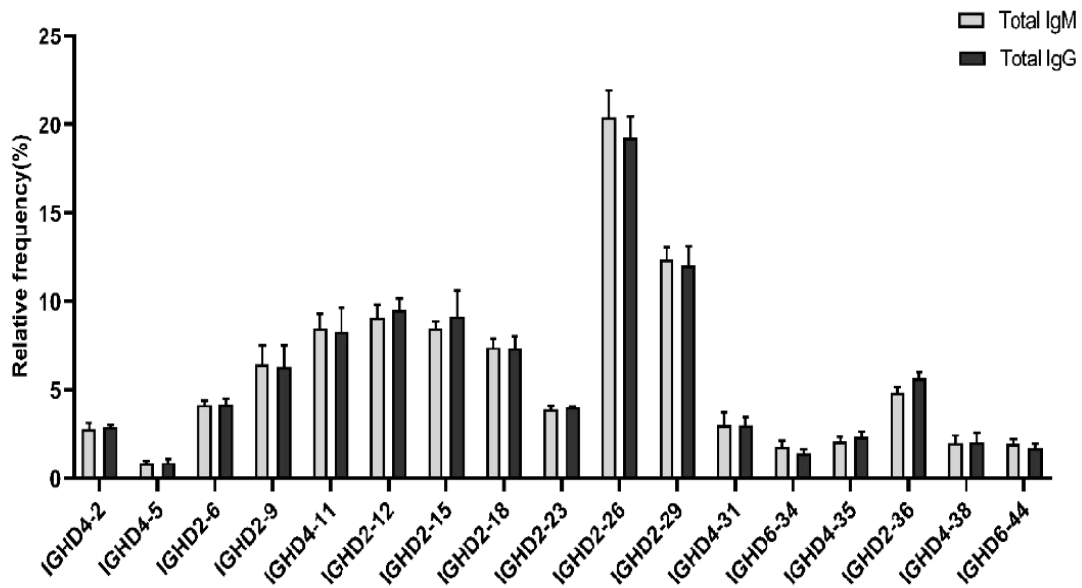
Supplementary Table 3: Amino Acid Composition per reading frame of IGHD functional gene segments in horse antibodies.

IMGT Group	IMGT gene	RF1	RF2	RF3	iRF1	iRF2	iRF3
IGHD2	IGHD2-6	GYYSRSCYT	MVTIVGVAI	WLL**ELLY	YSNSYSNH	GIATPIIVT	V*QLLL**P
	IGHD2-9	GYASGYDY	MVTMLVVVTTI	WLLC*WL*L	CSNH*HSNH	V VITTSIVT	*S*PLA**P
	IGHD2-12	YSYGSYYA	TIVMVVVM	L*LW*LLC	HSNYHNYS	GIVTTITI	A*_LP*L*
	IGHD2-15	CYYGSYYSSY YA	MVTMVVTTIVV TM	WLLW*LLQ*LLC	HSNYCSNYHSNH	GIVTTVTTIVT	A*_L*_LP**P
	IGHD2-18	GYAGSYA	MVTMLVVIM	WLLCW*LLC	HSNYQHSNH	GIVTTSIVT	A*_LP**P
	IGHD2-23	DYYGIDSY	MITMVLVTPI	*LLWY*LL	CRSH*YHSNH	VGVTNIIVI	*ESLIP**S
	IGHD2-26	YGYGGAYY	TMVMVLLTTI	LWLWCLLL	CSSKHHHNS	VVSTIIII	**APP*P*
	IGHD2-29	SYGGSSWYS	TVTMMVVVPEI	QLLW*FLVL	STRNYHNSC	GVPGTTTIVT	EYQELPP**L
	IGHD2-36	DYYGAIDYI	MIIMVLLIT*	*LLWCY*LHN	LCSQ*HNNH	YVNSIIII	VM*SIAP**S
	IGHD4	IGHD4-2	YYGWGN	TMAGV	LLWLG*	YPSHS	VTPAIV
IGHD4-5		YNYNY	TTTIT	LIQL*L	SYSCS	VVIVVV	*L*L*
IGHD4-11		NYGYGA	TTVMVML	*LRLWLCY	VA*P*P*L	*HFNRS	SITITVV
IGHD4-31		YDDGYN	TMTDTT	LR*RILQ	CSIRHR	VVSVIV	_P*YPSS*
IGHD4-35		NYGSYNY	TMAPIIT	*LWLL*LL	SNYRSHS	VVIIGAVI	**_*EP*L
IGHD4-38		EKSWSN	RRVGV	GEELE*	YSNSS	VTPTLL	_LQLFS
IGHD6-34		YGSGW	TVAVG	LR*RLA	ANRYR	PTATV	GQPLP*
IGHD6-44	YGSGW	TVVVG	IR*WLA	ANHYR	PTTTV	GQPLPY	

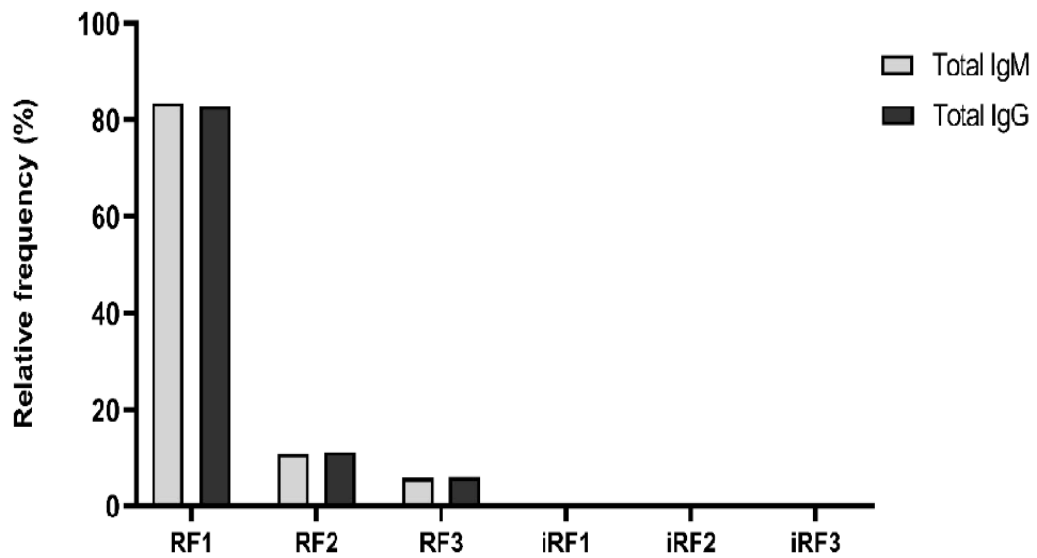
* means Stop codon

Supplementary Table 4: Top 10 bigger CDRH3 found in this study

Horse ID	IGHV gene and allele	IGHJ gene and allele	CDR3 amino acids	CDR3 length (aa)	N1 number (nt)	N2 nt number
4	IGHV4-21*01 ORF	IGHJ4*01 F	IGGKRGGFQLKEGGGGSSYSSGTGSPQRDRYFGYWGDTPV KAVAQSVETEYYTY	59	40	104
4	IGHV4-21*01 ORF	IGHJ4*01 F	AGGEVCEECEKCYDSGYSTITVQEKKCVRRSVNDSEYYSRSCCC RYYFAY	51	103	8
4	IGHV4S1*01	IGHJ6-1*01 F	AGADYGGTMHGKIFWQQGILVTVSSGESHSPLYCCTGADYGGTY HGKIF	50	111	12
4	IGHV4-21*01 ORF	IGHJ4*01 F	AGVWGDWVKGLVYAIDEWGPGLSTVSSGESHDDRGGLLYSIDY	43	11	91
1	IGHV4-21*01 ORF	IGHJ2*01 ORF	AGGNMVGYCAMMRCGHEYCVQGILGTVSSWESRSTEN	37	31	51
2	IGHV4-29*01	IGHJ4*01 F	GASLTVVGEI.PPGPI.I.ETGVADDDYDDTFAFTSEFVY	36	62	12
2	IGHV4-21*01 ORF	IGHJ6-1*01 F	SGGEGRVKDSTIYADEAIMEGRVKDSTVSVDEAILY	36	16	66
4	IGHV4-37D*01	IGHJ6-1*01 F	ATALAQVVLDPWPRWYCLKNVLLGYKLLVYWGINS	35	13	62
4	IGHV4-37D*01	IGHJ6-1*01 F	ATALAQVVLDPWPRWYCLKNVLLGYKFLVYWGINS	35	13	62
4	IGHV4S1*01	IGHJ6-1*01 F	KGLVARDAGGSESLRRRRELRLRIMPVSVYVSVNY	34	12	56



Supplementary Figure 1: Median of Relative frequency of functional IGHD gene segments in horses in IgM and IgG repertoires.



Supplementary Figure 2- Reading Frame (RF) preference for IGHD gene segment in non-immunized horses. Relative frequency of the six possible open reading frames for the IGHD gene segment. RF1, RF2, and RF3 are generated by deletion, while iRF1, iRF2, and iRF3 are generated by inversion.

ANEXO V

Comparação do índice de dissimilaridade de Bray-Curtis entre cavalos IgM e IgG

Amostra	Índice de dissimilaridade de Bray-Curtis (%)
1 IgM / IgG	0,99
2 IgM / IgG	0,99
3 IgM / IgG	0,99
4 IgM / IgG	0,98