Molecular Immunology 105 (2019) 251–259

Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm

Next-generation sequencing reveals new insights about gene usage and CDR-H3 composition in the horse antibody repertoire

Taciana Conceição Manso[a], Michele Groenner-Penna[a], João Carlos Minozzo[b], Bruno Cesar Antunes[b], Gregory C. Ippolito[c], Franck Molina[d], Liza F. Felicori[a,*]

[a] Laboratory of Synthetic Biology and Biomimetics, Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas - ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[b] Production and Research Centre of Immunobiological Products, Department of Health of the State of Paraná, Piraquara 83302-200, Brazil
[c] Department of Molecular Biosciences, The University of Texas at Austin, 100 E. 24th Street, Stop A5000, Austin, TX, 78712, USA
[d] Sys2diag, CNRS/Alcediag, Montpellier, France

ARTICLE INFO

Keywords:
Horse
Antibody repertoire
Immunoglobulin
Diversity
CDR-H3 composition

ABSTRACT

Horse serum antibodies have been used for greater than a century for the treatment and prophylaxis of infectious diseases and envenomations. Little is known, however, about the immunogenetic diversity that produces horse serum antibodies. Here, we employed next-generation sequencing for a first-in-kind comprehensive analysis of the equine B-cell repertoire. Nearly 45,000 and 30,000 clonotypes were obtained for the heavy-chain (IGH) and lambda light-chain (IGL) loci, respectively. We observed skewed use of the common subgroups IGHV2 (92.49%) and IGLV8 (82.50%), consistent with previous reports, but also novel use of the rare genes IGHV6S1 and IGLV4S2. CDR-H3 amino acid composition revealed different amino acid patterns at positions 106 and 116 compared to human, rabbit, and mouse, suggesting that an extended conformation predominates among horse CDR-H3 loops. Our analysis provides new insights regarding the mechanisms employed to generate antibody diversity in the horse, and could be applicable to the optimized design of synthetic antibodies intended for future therapeutic use.

1. Introduction

Whereas the antibody repertoire has been analyzed extensively in humans and mice (de los Rios et al., 2015), such studies are less common in other vertebrate species (Sun et al., 2012). Although uncommon and limited in scope, these comparative analyses in other species such as chicken, sheep, pig, and horse have revealed fresh insights into the multiple distinct mechanisms which can create antibody diversity in vertebrates (Flajnik and Kasahara, 2010; Sun et al., 2012).

Polyclonal horse (*Equus caballus*) serum antibodies (equine immunoglobulins) can be readily elicited by intentional hyperimmunization and have been produced since the 19th century for the effective treatment and prophylaxis of infectious diseases such as diphtheria, tuberculosis, tetanus, and pneumonia (Anderson, 1955; Cole and Moore, 1917; Glatman-Freedman and Casadevall, 1998; Lang et al., 2000). The therapeutic use of equine immunoglobulins continues to the present day. Despite their ease of manufacture and their compelling medical utility, our knowledge of how the humoral immune system generates such immunoglobulins in horses is lacking—especially at the molecular level—even though such knowledge could impinge on the improved development of new vaccines for infectious diseases, the treatment of immunological disorders like allergy, or the production of potent, lifesaving antivenoms to countervail snake and spider envenomings (Alvarenga et al., 2014; Horohov, 2015; Keggan et al., 2013;

Abbreviations: CDR-H3, third complementarity-determining region of the variable heavy chain; FR1, framework 1; FR4, framework 4; HTS, high-throughput sequencing; IG, immunoglobulin; IGH, immunoglobulin heavy chain; IGHV, immunoglobulin heavy variable gene; IGHD, immunoglobulin heavy diversity gene; IGHJ, immunoglobulin heavy joining gene; IGK, immunoglobulin kappa light chain; IGKV, immunoglobulin kappa variable gene; IGKJ, immunoglobulin kappa joining gene; IGKC, immunoglobulin kappa constant gene; IGL, immunoglobulin lambda light chain; IGLV, immunoglobulin lambda variable gene; IGLJ, immunoglobulin lambda joining gene; IGLC, immunoglobulin lambda constant gene; IMGT, the international ImMunoGeneTics information system; NCBI, National Center for Biotechnology Information; ORF, open reading frame; P, pseudogenes; PBMC, peripheral blood mononuclear cell; SRA, sequence read archive; VDJ, immunoglobulin heavy chain variable diversity and joining regions

* Corresponding author at: Bloco N4, sala 202, Instituto de Ciências Biológicas - ICB, Universidade Federal de Minas Gerais, Avenida Antônio Carlos 6627, CEP: 31270-901, Belo Horizonte, Brazil.
E-mail address: liza@icb.ufmg.br (L.F. Felicori).

https://doi.org/10.1016/j.molimm.2018.11.017
Received 10 August 2018; Received in revised form 11 November 2018; Accepted 30 November 2018
Available online 15 December 2018

Wagner, 2006).

Horse antibody constant regions can be formed by 11 different heavy chain (IGH) classes (IGHM, IGHD, IGHA, IGHE, and seven IGHG subclasses) (Wagner et al., 2004) and two light chains, kappa (IGK) and lambda (IGL). In equines there is a predominance of IGL light chain expression (~95% of the circulating IG) similar to sheep, cattle, dog and cat (Arun et al., 1996; Foley and Beh, 1992), in contrast to mice and rabbits that have a predominance of the kappa light chain isotype (Lavinder et al., 2014; Weiss et al., 1984).

According to IMGT nomenclature guidelines (Lefranc, 2007), the horse IGH has 56 IGHV genes (variable genes) (of which 16 are functional, 34 are pseudogenes, and 6 are defined as open reading frames - ORFs) grouped in 7 subgroups plus some pseudogenes, 40 IGHD genes (diversity genes) grouped in 28 subgroups, and eight IGHJ genes (joining genes) grouped in two subgroups (Sun et al., 2010; Walther et al., 2015). The IGK contains 60 IGKV genes, five IGKJ genes, and one IGKC gene, whereas the IGL locus has 144 IGLV genes (of which 27 are functional genes, 112 pseudogenes and 5 ORFs) grouped in 11 subgroups plus some pseudogenes, 7 IGLJ genes classified in 7 different subgroups and 7 IGLC genes (Sun et al., 2010; Walther et al., 2015).

Informative yet low-throughput equine antibody repertoire analyses have been performed by analysing the composition at different ages (Tallmadge et al., 2014, 2013), against infectious diseases (Khatibzadeh et al., 2015; Tallmadge et al., 2015), for antivenom production (Almagro et al., 2006), and in naïve horses (Almagro et al., 2006; Hara et al., 2012; Sun et al., 2010). It's known that antibody repertoire analysis can be very useful for the generation of recombinant antibodies (Lee et al., 2017), to monitor the immune response in health, autoimmunity, microbial infection, biothreat detection, vaccination and cancer (Robinson, 2015), for vaccine and immunomodulatory drug development, to study B-cell development, and for antigen and epitope discovery (Calis and Rosenberg, 2014; Georgiou et al., 2014; Robinson, 2015; Yaari and Kleinstein, 2015). However, due to the high diversity of the IG repertoire (for humans it's estimated to be 10^{11} (Boyd and Joshi, 2014)) only high-throughput sequencing (HTS) technologies can capture this diversity (Calis and Rosenberg, 2014).

Herein we present a first-in-kind characterization of the non-immunized equine antibody repertoire (IGH and IGL) using HTS technology. From this analysis we obtained approximately 45,000 unique IGH sequences and 30,000 unique IGL sequences obtained from two domestic adult horses. We observed that subgroup IGHV2 is highly expressed in horse PBMCs with a mean 92.49%. Regarding lambda light-chain, IGLV8 had a higher relative frequency in both horses, with a mean 82.50%. Additionally, the CDR-H3 amino-acid pattern by position revealed a different enrichment at IMGT positions 106 and 116 (Lefranc et al., 2003) compared to human, rabbit and mouse, suggesting that the CDR-H3 extended-loop conformation predominates in horse CDR-H3 structure in contrast to these other species in which the kinked structure is predominant.

2. Methods

2.1. Horse blood samples and RNA extraction

Peripheral blood samples from healthy, non-immunized equines of undefined breed ($n = 2$) were collected in the Production and Research Centre of Immunobiological Products, Paraná, Brazil. These experiments were performed in accordance with the Board of Ethics, at Federal University of Minas Gerais with the number 190/2018.

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/ml) were cryopreserved in FBS 90%/DMSO 10% at −80 °C until use. The equines were named as Horse 1 and Horse 2.

The RNA from PBMC was extracted with RNeasy Mini kit (Qiagen) following the manufacturer's instruction. RNA quantity and quality were determined by Qubit RNA BR Assay kit (Thermo Fisher Scientific, Inc.) and RNA Assay Reagent kit for LabChip (Perkin Elmer), respectively. The RNA quality was assessed using the RNA Quality Score on the LabChip results above 9.0. About 500 ng of RNA was used for the cDNA synthesis with SuperScript IV RT (Thermo Fisher Scientific, Inc.).

2.2. IGH and IGL horse primers design

Horse antibody heavy chain (IGH) and lambda light chain (IGL) primers were designed using the annotated horse IG loci (Sun et al., 2010; Walther et al., 2015) from EquCab2 genome (Genbank assembly ID GCA_000002305.1).

With the aim to optimize forward primers design, the sequences from 16 IGHV and 27 IGLV functional genes (leader and framework 1 - FR1) were aligned separately by Clustal Omega (Sievers and Higgins, 2014), a distance matrix was generated and used for hierarchical clustering by *pvclust* package in R program (R Core Team, 2013). The clusters were used for degenerate primers design. Similarly, five IGLJ functional genes were aligned and reverse primers were designed on the framework 4 (FR4) for IGL. We used the primer reverse IGHJ described by Tallmadge et al. (2013) for the eight IGHJ functional genes. Each forward primer was analyzed individually by PCR.

The expected amplicons size for the IGH and IGL are 350 bp and 300 bp, respectively. We incorporated the Illumina overhang adaptors sequence in all designed primers what resulted in an increase of 70 bp in the amplicon size. After PCR optimization, the primers were used in a multiplex PCR for either the IGH or IGL libraries.

2.3. IGH and IGL amplification

After cDNA synthesis and primer design, a multiplex PCR reaction was conducted to obtain IGH and IGL amplicons. The reactions were conducted with 80 ng cDNA, 1X high fidelity buffer, 2 mM $MgSO_4$, 0.2 mM dNTPs, 0.4 mg/ml BSA, the mix forward primers and mix reverse primers (0.25 μM each primer), and 1 U Platinum Taq DNA polymerase High Fidelity (Thermo Fisher Scientific, Inc.). Thermal 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; 28 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 concentrated with Zymo Research Genomic DNA Clean & Concentrator™-10 kit and the products were analyzed on 2% agarose gels and stained with GelGreen® Nucleic Acid stain (Biotium). The bands were excised and purified with NucleoSpin® Extract II kit (Macherey Nagel). The purified DNA were quantified by Qubit DNA High Sensitivity kit (Thermo Fisher Scientific, Inc.) and analyzed by LabChip DNA High Sensitivity (Perkin Elmer). The IGH and IGL samples from two equines were sequenced by Macrogen Inc (Seoul, Korea) using Illumina MiSeq platform 2 × 300 bp read length. The sequences have been deposited at the NIH SRA (Sequence Read Archive) under accession number SRP156553.

2.4. Bioinformatics repertoire analysis

The quality of the raw data was analyzed using FASTQC (Andrews, 2010). The adapters were removed and the sequences were processed (assembly, Phred quality score trim, primer identification and reads count) by pRESTO pipeline (Vander Heiden et al., 2014). Using a horse germline database (Sun et al., 2010; Walther et al., 2015) as reference, IgBLAST (Ye et al., 2013) was used for V(D)J annotation using as parameters: mismatch D penalty -4, mismatch J penalty -3 and e-value 25. To identify CDR3 end (coming from J gene), using IgBLAST, an auxiliary file was created containing: the frame start position, the gene type (IGHJ or IGLJ), and the CDR3-end position for each J gene.

The clonotypes were clustering based on the following criteria: same VJ gene segment, an identical CDR3 length and CDR3 90% of similarity

at amino acids sequences, according to Rogosch et al. (2012) and (Wine et al., 2013). CD-HIT (Fu et al., 2012) was used to identify the clusters. The representative sequences for each cluster were used in the analysis. Gene usage, subgroup and amino acids frequency were computed using Perl and Python *in house* scripts. To compare the results from two horses we used unpaired *t*-test by GraphPad Prism 5.00 (GraphPad Software, Inc., San Diego California USA).

2.5. CDR-H3 amino acid length and composition analysis

To compare horse CDR-H3 amino acid sequences with others mammals, the Sequence Read Archive (SRA) from National Center for Biotechnology Information (NCBI, U.S. National Library of Medicine) was used to obtain the IGH non-immunized repertoire sequences of human (SRR4034888), mouse (SRR3575990), and rabbit (SRR1300150) (Lavinder et al., 2014). The sequences were submitted to the same pipeline of the horse sequences: pRESTO toolkit (Vander Heiden et al., 2014) followed by VDJ and CDR-H3 assignment using IgBLAST (Ye et al., 2013) with their gemline databases from IGMT, and clonotypes clustering by CD-HIT. Mann-Whitney Rank Sum test was used to assess the differences of the medians of CDR-H3 lengths.

To assess amino acid composition, IGH amino acid sequences were aligned and gapped according IMGT numbering (Lefranc et al., 2003). The positions 105–117, that corresponding to CDR-H3, with the additional positions between positions 111 and 112, were analyzed by Python *in house* script. To analyze amino acid composition by position, the additional positions were omitted, and the amino acids frequencies were analyzed. The two-way ANOVA with Bonferroni post-test was used to analyze amino acid composition among different species. The statistical methods were performed using GraphPad Prism 5.00 (GraphPad Software, Inc., San Diego California USA).

3. Results

3.1. Primer design to enable horse IGH and IGL deep sequencing

With the aim to analyze the antibody repertoire from two adult non-immunized equines, the PBMC from peripheral blood was isolated and used to generate IGH and IGL libraries. Forward primers targeting the leader peptide and FR1 regions for the IGHV and IGLV functional genes as well as a single reverse primer targeting FR4 of the IGLJ genes were designed (Table 1). A reverse primer targeting FR4 on the IGHJ gene was previously described (Tallmadge et al., 2013).

Each forward primer was individually tested by PCR showing amplifications for most of them, except to hoHV1S6f, hoHV4S17f, and hoHV7S1f primers for the IGH and hoLVxf primer for IGL (Fig. S1). However, when complete mix of the forward primers was used to PCR amplify horse antibodies, the genes covered by these primers were nonetheless identified in our dataset (Table 2), demonstrating a successful primer design which enables the amplification of all IGH and IGL genes.

After obtaining successful amplifications using the designed immunoglobulin primers, the PCR amplicons were purified and sequenced using the Illumina MiSeq 2 × 300 bp platform. Approximately 500,000 sequences were obtained from each library. Raw data processing using the pRESTO toolkit retrieved about 165,000 unique IGH sequences and 125,000 unique IGL sequences from each horse. All the sequences were annotated with IgBLAST tool; approximately 50% of them could be completely annotated for all component V(D)J gene segments (Table 3), indicating that the other 50% of the reads had only a partial annotation in the J region. Around 79,736 IGH and 65,403 IGL yielded a complete VDJ annotation, among these we have about 45,000 IGH and 30,000 IGL unique clones. This represents more than 430 times the number of horse VDJ annotated sequences than previously described in the literature (Almagro et al., 2006; Hara et al., 2012; Sun et al., 2010; Tallmadge et al., 2014, 2013) (Table 3).

Table 1
Sequences of the designed primers used to amplify IGH and IGL genes.

Primers sequences	
IGHV forward primers	
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
IGHJ reverse primer	
IGHJr*	CTC GCC TGA GGA GAC GGT GAC CAG
IGLV forward primers	
hoLV8f	CCC AGT CTS TGA CBC AGC CC
hoLV5f	TCT GTG KCC YCT TCT GAG CTK ACT CA
hoLV2_4f	CYG TGG CYT CTT CTA WGC TGA C
hoLV1_4f	TCG GAG STG ACT CAG CCA
hoLV4f	TCT CTT CTG MRG TGA CTC AGC CA
hoLV6f	CAG TCT GCC CTG RYT CAG CC
hoLV9f	AGC CTG TGC TGA CCC AG
hoLVxf	CRG ATT CTC AGK CTG TGG TG
hoLV3S1f	TCT GTG ACC ACC TAT GAC TTG ACG
hoLV7S1f	ACA GGA TGT GGT GAT TCA GG
IGLJ reverse primers	
hoLJ1r	CTT ACC CAG GAC GCT GAG GTG
hoLJ4-7r	CTC ACC TGC GAT GGT CAG GT

(*)Tallmadge et al. (2013).
Nomenclature: ho: horse; H: heavy chain; L: lambda light chain; V: variable gene; J: joing gene; f: forward; r: reverse; numbers: subgroup classification; x: not classified into subgroups.

3.2. IGH and IGL gene usage

The designed primers could amplify all known IGH and IGL gene subgroups described in horse (Sun et al., 2010; Walther et al., 2015). A skewed preference for the IGHV2 gene subgroup with a mean of 92.49% ± 0.024 was observed for both equines (Fig. S2A). Within the IGHV2 subgroup, the gene segments IGHV2S2, IGHV2S3, and IGHV2S4 were the most highly expressed in both animals (Fig. 1A). For the IGHD genes all 28 subgroups were expressed with a similar relative frequency; the three most frequent genes were the same between the 2 animals (IGHD18S1, IGHD15S2, and IGHD10S1). Comparing the two IGHJ gene subgroups, IGHJ1 (composed of seven genes) is over-represented with a mean 99.71% ± 0.002 for the two equines. The more highly expressed genes within this group were IGHJ1S5, IGHJ1S3 and IGHJ1S2 (Fig. 1A).

With respect to the light chain repertoire, among the 11 IGLV gene subgroups, IGLV8, specifically the genes IGLV8S1 and IGLV8S2 (Fig. 1B) had the highest relative frequency in both horses, with a mean frequency of 82.50% ± 0.073. The IGLV4 subgroup had important relative expression also in both horses (15.48% ± 0.083), whereas the other subgroups had reduced expression (Fig. S2B). From the seven IGLJ genes, IGLJ1S1 and IGLJ5S1 had higher expression frequency in the two horses, with frequency means 35.46% ± 0.045 and 58.60% ± 0.040, respectively. The pseudogene IGLJ4S1(P) was found among the more frequent ones in both horses (frequency mean of 5.94% ± 0.005) (Fig. 1B).

For IGH, 1,130 and 1,559 different VDJ recombined genes were found in Horse 1 and Horse 2, respectively. In IGH recombined genes, a greater diversity was observed. The most frequent recombined fragments was similar in both equines, IGHV2S3, IGHD18S1, IGHJ1S5 with frequency 2.63% for Horse 1 and IGHV2S2, IGHD18S1, IGHJ1S5 with frequency 2.37% for Horse 2 sequences (Table S1).

Table 2
Representation of individual primers tested, the corresponding gene each primer amplifies, their results in PCR and the frequency of the group identified by each primer in 79,732 IGH and 65,403 IGL annotated reads.

IG genes	Primers	PCR test result	Frequency of sequences identified
Heavy chain *locus*			
IGHV2S1	hoHV2_4f	+ +	75.07%
IGHV2S2			
IGHV2S3			
IGHV2S4			
IGHV4S5			
IGHV4S1	hoHV4f	+ +	0.66%
IGHV4S2			
IGHV4S3			
IGHV4S4			
IGHV1S1	hoHV1f	+ +	0.45%
IGHV1S2			
IGHV1S3			
IGHV1S6	hoHV1S6f	–	2.16%
IGHV4S17	hoHV4S17f	+	4.36%
IGHV6S1	hoHV6S1f	+ +	9.94%
IGHV7S1	hoHV7S1f	–	0.02%
Lambda light chain *locus*			
IGLV8S1	hoLV8f	+ +	48.73%
IGLV8S2			
IGLV8S3			
IGLV8S4			
IGLV8S5			
IGLV8S6			
IGLV8S7			
IGLV8S8			
IGLV8S9			
IGLV5S1	hoLV5f	+ +	0.11%
IGLV5S2			
IGLV2S1	hoLV2_4f	+ +	4.16%
IGLV4S2			
IGLV1S1	hoLV1_4f	+ +	3.80%
IGLV4S3			
IGLV4S1	hoLV4f	+ +	10.92%
IGLV4S4			
IGLV6S1	hoLV6f	+ +	3.89%
IGLV6S2			
IGLV6S3			
IGLV6S4			
IGLV9S1	hoLV9f	+ +	5.18%
IGLV9S2			
IGLVxS60	hoLVxf	+	0.18%
IGLVxS64			
IGLV3S1	hoLV3S1f	+ +	0.15%
IGLV7S1	hoLV7S1f	+ +	0.89%

3.3. CDR-H3 length in horse is intermediate between human and mouse repertoires

CDR-H3 amino acid length in horses varies from 4 to 26 amino acids residues, following a bimodal model distribution with a clear peak at five and six residues and a more extended peak ranging from 10 to 19 residues (Fig. S3). The median length was 14 residues for Horse 1 and 15 residues for Horse 2. When CDR-H3 lengths derived from horses were compared with CDR-H3 derived from other species we noticed that human CDR-H3 median lengths are significantly longer (16 residues, $p < 0.0001$), while mouse shows shorter CDR-H3 length median (12 residues, $p < 0.0001$) than horse CDR-H3 (Fig. 2).

CDR-H3 amino acids residues composition in both horses exhibited a similarity frequency, only an Ala frequency has higher in Horse 1 than Horse 2. The five most frequently used amino acids were Gly, Tyr, Ala, Asp and Ser (Fig. S4). It can also be noticed that horse has higher Gly composition (17.47%) in comparison with rabbit (8.89%, $p < 0.05$) and low Leu composition (4.24%) than rabbit (13.89%, $p < 0.05$) (Fig. 3).

3.4. Horse CDR-H3 differs from rabbit, human and mouse CDR-H3 in their extremities

Horse 1 and Horse 2 had their CDR-H3 sequences aligned from positions 105–117, where a limited variability in the amino acids residues at positions 105 (high frequency of Ala - 60.73%), 106 (high frequency of Gly - 43.58%) and 117 (high frequency of Tyr - 53.16%), were observed in Horse 1 (Fig. 4A). For Horse 2 results, see Fig. S5.

The amino acid residues composition by position in human, mouse, and rabbit were also analyzed (Fig. 4B) and compared to horse. Similar to horse, the position 105 has a high prevalence of Ala for human and mouse, while rabbit have a higher prevalence of Arg (41.71%, $p < 0.001$) than horses (1.15%).

Interestingly, the position 106 has significantly higher frequency for Arg residues in human (55.79%, $p < 0.001$) than horse (16.34%), while horse shows higher frequency for Gly in this position (43.58%) significantly different than mouse and rabbit (5.76% and 5.27%, respectively, $p < 0.05$) and human (1.99%, $p < 0.01$). At position 116, Asp residues were more frequent in human (62.49%, $p < 0.001$) than horse (20.06%). At position 117, tyrosine was the most frequent residue in human, mouse, and horse. However, rabbit shows higher frequency for Leu in this position (46.82%, $p < 0.01$) than horse (2.91%) and low frequency for Tyr (0.19%, $p < 0.001$) than horse (53.16%).

These results indicate a remarkable difference in the amino acid composition in the extremities of the horse CDR-H3 that could result in a different conformation in the base of these CDR-H3s and consequently impact affinity or specificity of these antibodies.

4. Discussion

Our analysis of nearly 45,000 IGH and 30,000 IGL unique antibody genes represents the first examination of the horse antibody repertoire using high-throughput sequencing technology. In addition to the insights gained from analyzing the IGH repertoire, our study also adds

Table 3
MiSeq data analysis summary.

Chain	Samples	Raw reads*	pRESTO analysis	IgBLAST annotation		Clonotypes	Unique reads from literature
			Unique reads	Annotated reads	V(D)J annotated**		
Heavy	Horse 1	478,188	150,655	150,655	77,879	49,957	102[a], 30[b], 46[c]
	Horse 2	516,852	183,951	183,950	81,593	43,474	
Lambda light	Horse 1	495,644	114,806	114,806	52,766	25,515	97[a], 30[d]
	Horse 2	557,202	133,614	133,614	78,040	32,743	

(*) Total reads sequenced (reverse and forward).
(**) Reads with the V(D)J segments annotated. The other reads were partial annotated since they lack one of the segments.
[a] Sun et al. (2010).
[b] Tallmadge et al. (2013).
[c] Almagro et al. (2006).
[d] Tallmadge et al. (2014).

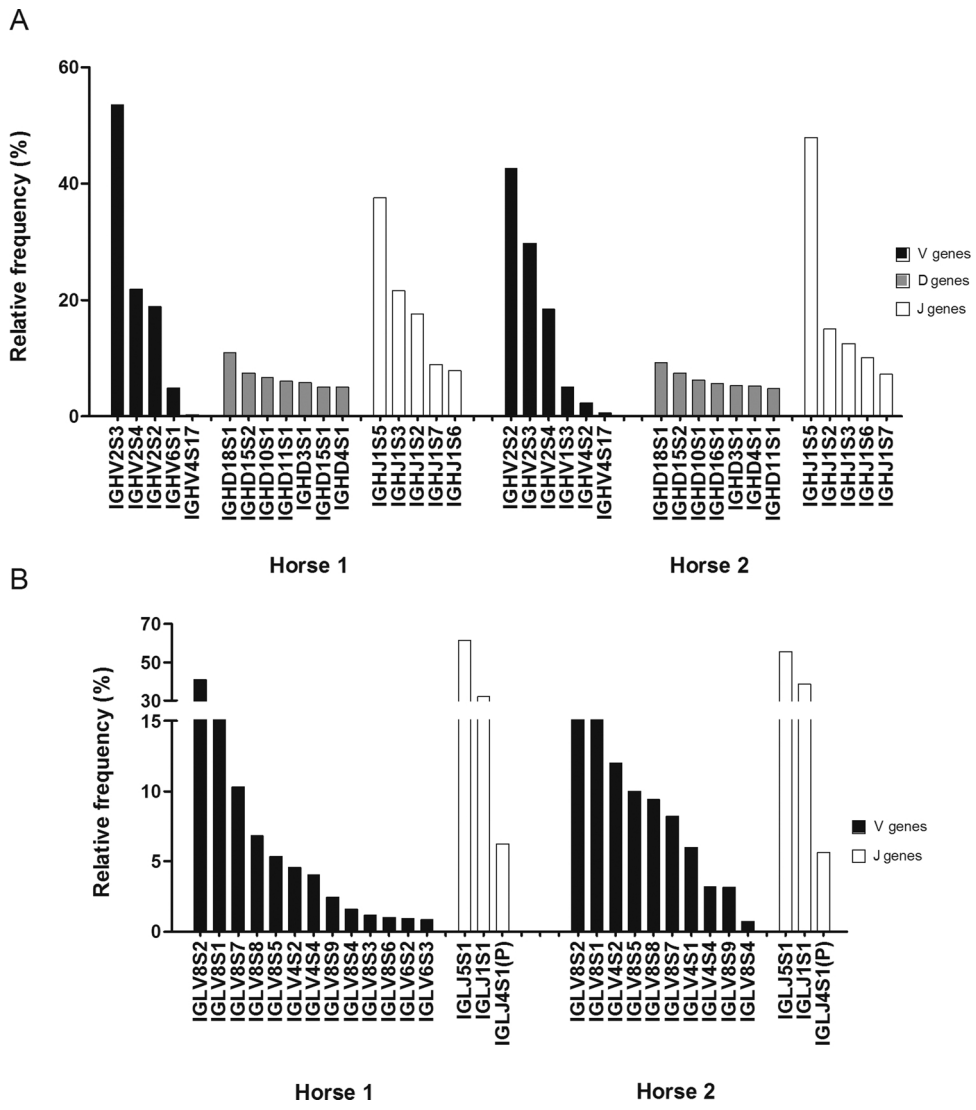

Fig. 1. Relative frequency (%) of the most frequent horse gene segments from (A) heavy chain and (B) lambda light chain in Horse 1 and Horse 2. V genes are show in black, D genes in gray, and J genes in white. (P) denotes gene annotated as pseudogene.

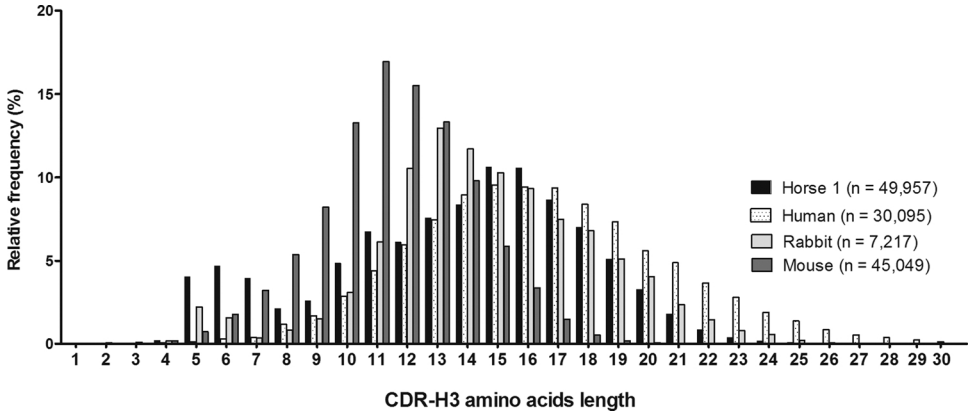

Fig. 2. CDR-H3 amino acids length distribution. The relative frequency in percentage is plotted on the y-axis. CDR-H3 amino acids length distribution of Horse 1 is show in black in comparison with: human (white with black dots), rabbit (light gray), and mouse (dark gray) repertoires. Since Horse 1 and 2 have similar CDR-H3 length distribution only Horse 1 is shown here. The immunoglobulin sequences were obtained from SRA database: human (SRR4034888), mouse (SRR3575990), and rabbit (SRR1300150).

substantially to the global community's database of horse IGL sequences, which is significant since fully 95% of horse antibodies use lambda light-chains (IGL) (Tallmadge et al., 2014; Walther et al., 2015).

We conclude that the primer sets used for V-gene PCR amplification had been adequately designed because all previously described functional V genes were also identified in our datasets. The reverse IGH primers used here target the IGHJ region, similar to prior strategies which had used PCR followed by Sanger sequencing to identify horse DNA antibody sequences (Almagro et al., 2006; Tallmadge et al., 2013). Although adequate in its design, we conclude that our primer was nevertheless suboptimal since, in our analysis we could unambiguously identify only about 50% of IGHJ (or IGLJ) genes, and therefore we were

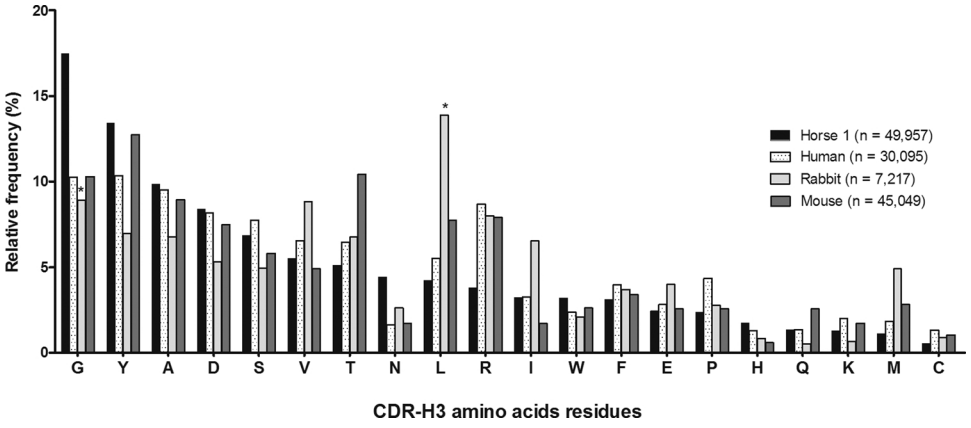

Fig. 3. Composition of total amino acids usage at CDR-H3 positions 105–117. Horse 1 is show in black, human in white with black dots, rabbit in light gray and mouse in dark gray. The relative frequency in percentage of each amino acid is plotted on the y-axis. The two-way ANOVA with Bonferroni post-test was used to compare horse *vs.* human, horse *vs.* mouse, and horse *vs.* rabbit. The asterisk (*) indicates the significance level was better than 0.05% for comparing individual residues frequencies.

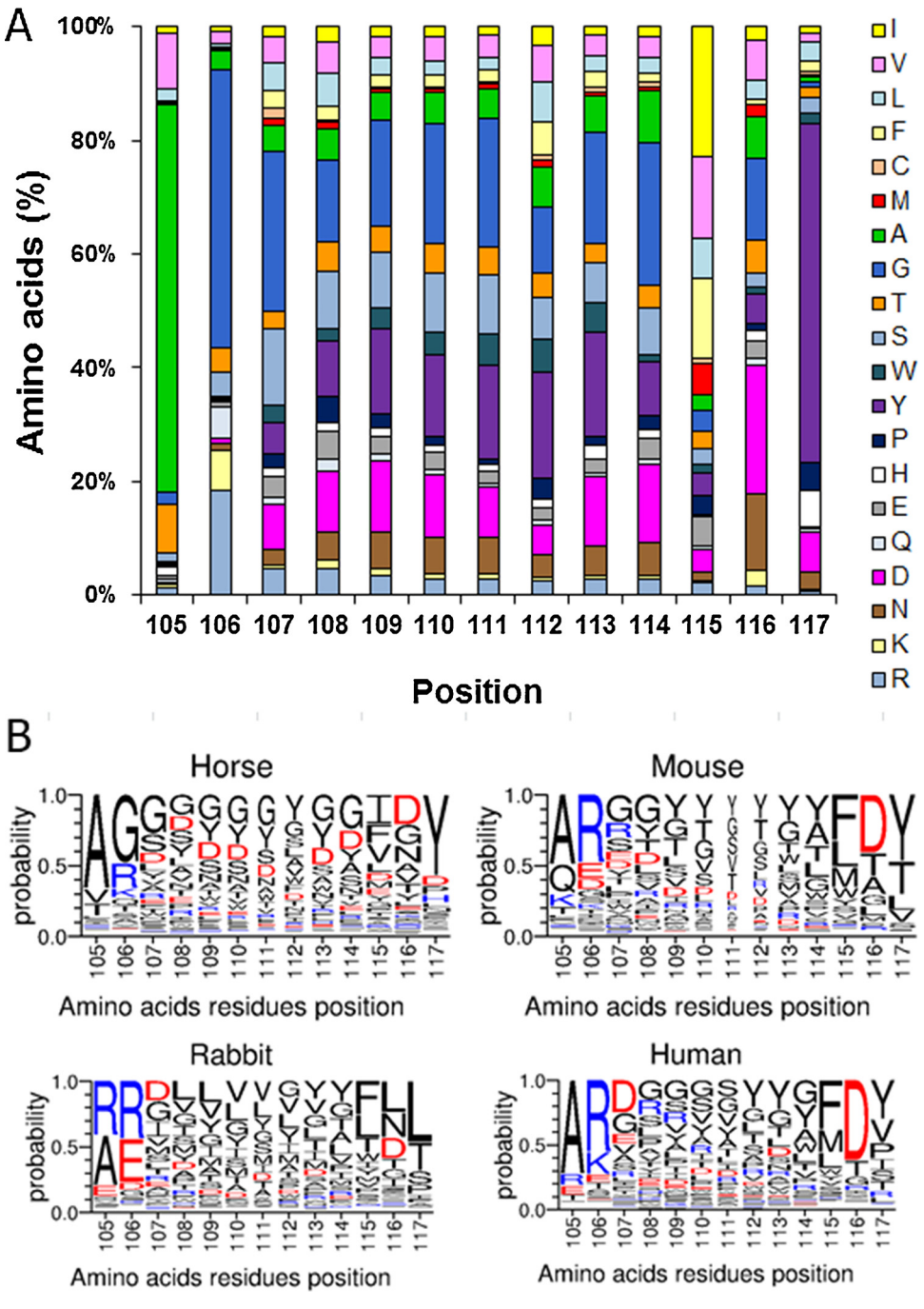

Fig. 4. CDR-H3 amino acid composition by position. CDR-H3 positions 105 to 117 were used according to IMGT. The additional positions between positions 111 and 112 at the top of the CDR3-IMGT loop were omitted. (A) CDR-H3 amino acids relative frequencies by position from CDR-H3 Horse 1 sequences. Horse 2 shows a similar pattern. Each bar represents 100% of the amino acid residue identified by position. Amino acids are stacked in the order of their hydrophobicity. (B) CDR-H3 Amino acids weblogo at the amino acids residues at the positions 105–117 for Horse 1, mouse, rabbit and human CDR-H3 sequences. The amino acids residues are colored by charge, red for negative charged amino acids, blue for positive charged amino acids, and black for not charged amino acids (For the printed version: the amino acids residues are colored by charge, light gray for negative charged amino acids, dark gray for positive charged amino acids, and black for not charged amino acids).

able to conclusively annotate only about 50% of full-length V(D)J IGH (or IGL) sequences. The use of IGHJ/IGLJ reverse primers for PCR amplification made it difficult to annotate these genes, and, in retrospect, it would have been optimal to use primers targeting the constant regions instead.

Despite this technical shortcoming, this work succeeded to annotate nearly 430-fold more IGH and IGL horse antibody sequences than previously existed, when compared to the 102 unique clones for heavy chain (Sun et al., 2010) and 97 unique clones for lambda light chain (Tallmadge et al., 2014) that had been previously reported. IGH and IGL gene usage identified by high-throughput sequencing used here were similar to the ones described by others when using Sanger sequencing (Almagro et al., 2006; Sun et al., 2010; Tallmadge et al., 2014, 2013), but additional information about gene usage could be observed in this work. Non-immunized horses showed a restricted repertoire for IGHV genes, since 92.49% of the V genes expressed belongs to the subgroup IGHV2, similar to previous studies (Sun et al., 2010; Tallmadge et al., 2013). From the 4 genes in the subgroup IGHV2, the more frequently expressed genes were IGHV2S2, IGHV2S3, and IGHV2S4, which appears to be expressed throughout horse life (Tallmadge et al., 2013). Here, however, for the first time, we detected expression of the genes IGHV4S17 (0.59% in Horse 2), IGHV6S1 (4.82% in Horse 1), IGHV1S3 (5.02% in Horse 2) and IGHV4S2 (2.35% in Horse 2) in adult horses; the subgroup IGHV6 had never been described before in the literature.

In agreement with Sun et al. (2010), IGHJ genes displayed a bias for IGHJ1S5 gene usage. With regard to IGHD genes, we observed that members of all 28 IGHD gene subgroups can indeed be expressed in the horse repertoire; to the best of our knowledge, this constitutes a first documentation of pan-IGHD expression in the horse. Previous studies, which did not use NGS for in-depth repertoire profiling, concluded that few IGHD genes are actually expressed, and also concluded that the IGHD18S1 gene is the most frequently used gene segment (Sun et al., 2010; Tallmadge et al., 2014). Herein, we too see a similar bias for IGHD18S1 (10.11% ± 0.12), as well as IGHD15S2 (7.37% ± 0.001), and IGHD10S1 (6.39% ± 0.3, whereas all other IGHD genes were used at frequencies ranging from 6% to as little as 1%. Counterbalancing the restrictive use of IGHV and IGHJ gene segments, we propose that this restriction is compensated by a wide-ranging use of all the IGHD gene segments, thereby augmenting the repertoire diversity of CDR-H3 antigen binding sites.

For IGL lambda light-chains, we observed the IGLV8 gene subgroup (comprising IGLV8S1 and IGLV8S2) to be the most frequently used by horses, consistent with previous reports; however, contrary to other reports, we detected additional high-frequency IGLV genes expressed in adult horses, such as IGLV4S2, IGLV4S4, and IGLV4S1 (Hara et al., 2012; Sun et al., 2010; Tallmadge et al., 2014). Regarding IGLJ genes, the most expressed genes were IGLJ1S1 and IGLJ5S1; both were previously described by Tallmadge et al. (2013, 2014) in horse fetuses but not in adult horses. Sun et al. (2010) did not observe IGLJ5S1 and IGLJ7S1 genes in their examination of two horses, whereas Hara et al. (2012) identified all IGLJ genes when interrogating two different horse breeds. In our study, we cannot differentiate the genes IGLJ5S1, IGLJ6S1 and IGLJ7S1, because they have the same gene sequence (to differentiate them it would be necessary to have the sequence of their constant region).

IGLJ4S1(P) was one of the three most frequent IGLJ genes we observed. Based on genome sequencing, this gene has been annotated as a pseudogene, although expressed mRNA sequences curiously do not contain the putative single C-insertion (Hara et al., 2012; Sun et al., 2010; Walther et al., 2015). These collective results suggest a careful consideration to change its classification to a functional gene.

In naïve antibody repertoires, a Gaussian-like CDR-H3 length distribution is frequently observed (Miqueu et al., 2007).This is different from our observations that showed a bimodal distribution of horse CDR-H3 length as previously observed by Almagro et al. (2006). The CDR-H3 length distribution of 4–26 amino acids found in this study was similar with other studies (Almagro et al., 2006; Tallmadge et al., 2013). About 13.70% ± 0.011 of the CDR-H3 sequences contained < 7 amino acids, whereas 86.30% ± 0.011 of the sequences were > 8 amino acids.

The non-immunized horse repertoire CDR-H3 median length (14 aa) is shorter than naïve human CDR-H3s (16 aa) and longer than mouse ones (12 aa). The CDR-H3 length is known to affect the folding, the local shape of the binding site and the antigen binding specificity (den Hartog et al., 2013; Gibson et al., 2009; Miqueu et al., 2007). Antibodies that have CDR-H3 longer than 12 amino acids are more exposed to solvent (Ramsland et al., 2001), extending out of the combining site to enhance interaction with distal antigenic epitopes.

Horse CDR-H3 amino acid composition was compared with that of human, mouse, and rabbit antibody repertoires. The three most frequently used amino acids in horse Gly, Tyr, and Ala were the same as already described in human and mouse (Chang et al., 2016; Ippolito et al., 2006; Zemlin et al., 2003). Cys residues are the least frequent CDR-H3 amino acids in these species, including horse (0.53%), consistent with the literature (Almagro et al., 2006; Chang et al., 2016; Lavinder et al., 2014). Horse shows only higher frequencies for Gly (17.47%) than rabbit (8.89%, $p < 0.05$).

CDR-H3 amino acids composition by position has never been analyzed before in horses. The Ala residue was highly prevalent at position 105 in human and mouse and the position 117 shows highly frequency of Tyr in horse (53.16%), human (30.48%), and mouse (47.29%), as also described for canine and feline repertoires (Steiniger et al., 2017, 2014); however, rabbit CDR-H3s contain more Leu residues (46.82%) in this position. It already had been shown that the positions 105 and 117 typically map to the base of the CDR-H3 loop and are encoded by the IGHV and IGHJ gene segments (Zemlin et al., 2003). Thus, binding ability differences may be expected.

Interestingly, we identified main differences between horse and others mammals in other CDR-H3 positions. Different from other species, the positions 106 and 116 are not highly conserved for charged amino acids in horse (Steiniger et al., 2014; Zemlin et al., 2003). In general, these positions show conserved charged amino acids as Arg/Lys106 and Asp116. This creates a salt-bridge which stabilizes the CDR-H3 base and contributes to the kinked form (Kuroda et al., 2008; Shirai et al., 1996; Wu et al., 2012) as seen in human, rabbit, canine, feline and chicken (Chang et al., 2016; Lavinder et al., 2014; Steiniger et al., 2017, 2014; Wu et al., 2012). In horses however, only 9.8% ± 0.99 of the sequences have this amino acids composition at positions 106 and 116. In addition, the kinked form is favored when a large hydrophobic residue at position 116 and a small amino acid, such as Gly or Ala at the position 114 are present, composition present in about 4.43% ± 0.26 of horse sequences.

According to Kuroda et al. (2008), CDR-H3 extended form is favored when length is six or seven residues, in our analysis this represents 8.62% ± 0.0 of the sequences. The extended form tends also to exist when the residue 116 is not Asp and there is a hydrophilic residue at position 115 (47.92% ± 0.13 of the horse sequences) or there are Phe/Met115 and Ala/Gly114 (about 3.21% ± 1.69 of the horse sequences) (Kuroda et al., 2008; Shirai et al., 1996). Consequently, based on amino acids composition, the extended form seems to be predominant in horse antibodies, totalizing about 59.75% of the sequences. However, experiments to prove horse antibody structures and the impact of this structure in the affinity and specificity of these antibodies should be performed.

5. Conclusion

This work presents the first analysis of PBMC horse antibody repertoire by high-throughput sequence. We showed a complex and diverse repertoire for IGH, whereas IGL repertoire shows a lower diversity as seen before. We confirm the frequent genes described for horse and show additional genes used in horse repertoire never described before.

Our results suggest that horses use different amino acid architecture than humans and mice to stabilize the CDR-H3 loop, favoring the extended form. The knowledge of the horse repertoire is necessary for understand gene recombination and diversity generation what can help the design of new therapeutics based on horse immunoglobulins.

Funding

This research was supported by Brazilian Funding Agencies: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Acknowledgements

We are grateful to Dr. Rebecca L. Tallmadge and Dr. Tiago A.O. Mendes for expert advice, and to Dr. Larissa M. Alvarenga and Dr. Camila O. Franco for their help with sample collection and processing.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.molimm.2018.11.017.

References

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. Mol. Immunol. 43, 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 (Basel) 6, 2541–2567. https://doi.org/10.3390/toxins6082541.

Anderson, C.G., 1955. The distribution of diphtheria antitoxin in pepsin-digested horse antiserum. Biochem. J. 59, 47–52. https://doi.org/10.1042/bj0590047.

Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data.

Arun, S.S., Breuer, W., Hermanns, W., 1996. Immunohistochemical examination of light-chain expression (lambda/kappa ratio) in canine, feline, equine, bovine and porcine plasma cells. Zentralbl. Veterinarmed. A 43, 573–576.

Boyd, S.D., Joshi, S.A., 2014. High-throughput DNA sequencing analysis of antibody repertoires. Microbiol. Spectr. 2, 1–13. https://doi.org/10.1128/microbiolspec.AID-0017-2014.

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

Chang, Y.H., Kuan, H.C., Hsieh, T.C., Ma, K.H., Yang, C.H., Hsu, W., Tsai, S.F., Chao, A., Liu, H.H., 2016. Network signatures of IgG immune repertoires in hepatitis B associated chronic infection and vaccination responses. Sci. Rep. 6, 26556. https://doi.org/10.1038/srep26556.

Cole, R., Moore, H.F., 1917. The production of antipneumonococcic serum. J. Exp. Med. 26, 537–561. https://doi.org/10.1084/jem.26.4.537.

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

den Hartog, G., Crooijmans, R.P.M.A., Parmentier, H.K., Savelkoul, H.F.J., Bos, N.A., Lammers, A., 2013. Ontogeny of the avian intestinal immunoglobulin repertoire: modification in CDR3 length and conserved VH-pseudogene usage. Mol. Immunol. 56, 811–818. https://doi.org/10.1016/j.molimm.2013.08.005.

Flajnik, M.F., Kasahara, M., 2010. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat. Rev. Genet. 11, 47–59. https://doi.org/10.1038/nrg2703.

Foley, R.C., Beh, K.J., 1992. Analysis of immunoglobulin light chain loci in sheep. Anim. Genet. 23, 31–42. https://doi.org/10.1111/j.1365-2052.1992.tb00015.x.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

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

Gibson, K.L., Wu, Y.C., Barnett, Y., Duggan, O., Vaughan, R., Kondeatis, E., Nilsson, B.O., Wikby, A., Kipling, D., Dunn-Walters, D.K., 2009. B-Cell diversity decreases in old age and is correlated with poor health status. Aging Cell 8, 18–25. https://doi.org/10.1111/j.1474-9726.2008.00443.x.

Glatman-Freedman, A., Casadevall, A., 1998. Serum therapy for tuberculosis revisited: reappraisal of the role of antibody-mediated immunity against Mycobacterium tuberculosis. Clin. Microbiol. Rev. 11, 514–532.

Hara, S., Diesterbeck, U.S., König, S., Czerny, C.P., 2012. Transcriptional analysis of equine λ-light chains in the horse breeds Rhenish-German Coldblood and Hanoverian Warmblood. Vet. Immunol. Immunopathol. 145, 50–65. https://doi.org/10.1016/j.vetimm.2011.10.006.

Horohov, D.W., 2015. The equine immune responses to infectious and allergic disease: a model for humans? Mol. Immunol. 66, 89–96. https://doi.org/10.1016/j.molimm.2014.09.020.

Ippolito, G.C., Schelonka, R.L., Zemlin, M., Ivanov, I.I., Kobayashi, R., Zemlin, C., Gartland, G.L., Nitschke, L., Pelkonen, J., Fujihashi, K., Rajewsky, K., Schroeder, H.W., 2006. Forced usage of positively charged amino acids in immunoglobulin CDR-H3 impairs B cell development and antibody production. J. Exp. Med. 203, 1567–1578. https://doi.org/10.1084/jem.20052217.

Keggan, A., Freer, H., Rollins, A., Wagner, B., 2013. Production of seven monoclonal equine immunoglobulins isotyped by multiplex analysis. Vet. Immunol. Immunopathol. 153, 187–193. https://doi.org/10.1016/j.vetimm.2013.02.010.

Khatibzadeh, S.M., Gold, C.B., Keggan, A.E., Perkins, G.A., Glaser, A.L., Dubovi, E.J., Wagner, B., 2015. West Nile virus–specific immunoglobulin isotype responses in vaccinated and infected horses. Am. J. Vet. Res. 76, 92–100. https://doi.org/10.2460/ajvr.76.1.92.

Kuroda, D., Shirai, H., Kobori, M., Nakamura, H., 2008. Structural classification of CDR-H3 revisited: a lesson in antibody modeling. Proteins Struct. Funct. Genet. 73, 608–620. https://doi.org/10.1002/prot.22087.

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. Am. J. Trop. Med. Hyg. 63, 298–305. https://doi.org/10.1001/jama.2014.3633.

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, 1–11. https://doi.org/10.1371/journal.pone.0101322.

Lee, C.H., Romain, G., Yan, W., Watanabe, M., Charab, W., Todorova, B., Lee, J., Triplett, K., Donkor, M., Lungu, O.I., Lux, A., Marshall, N., Lindorfer, M.A., Goff, O.R., Balbino, B., Kang, T.H., Tanno, H., Delidakis, G., Alford, C., Taylor, R.P., Nimmerjahn, F., Varadarajan, N., Bruhns, P., Zhang, Y.J., Georgiou, G., 2017. IgG Fc domains that bind C1q but not effector Fc3 receptors delineate the importance of complement-mediated effector functions. Nat. Immunol. 18, 889–898. https://doi.org/10.1038/ni.3770.

Lefranc, M.P., 2007. WHO-IUIS nomenclature subcommittee for immunoglobulins and T cell receptors report. Immunogenetics 59, 899–902. https://doi.org/10.1007/s00251-007-0260-4.

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. Dev. Comp. Immunol. 27, 55–77. https://doi.org/10.1016/S0145-305X(02)00039-3.

Miqueu, P., Guillet, M., Degauque, N., Doré, J.C., Soulillou, J.P., Brouard, S., 2007. Statistical analysis of CDR3 length distributions for the assessment of T and B cell repertoire biases. Mol. Immunol. 44, 1057–1064. https://doi.org/10.1016/j.molimm.2006.06.026.

R Core Team, 2013. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Ramsland, P.A., Kaushik, A., Marchalonis, J.J., Edmundson, A.B., 2001. Incorporation of long CDR3s into V domains: implications for the structural evolution of the antibody-combining site. Exp. Clin. Immunogenet. 18, 176–198. https://doi.org/10.1159/000049197.

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

Rogosch, T., Kerzel, S., Hoi, K.H., Zhang, Z., Maier, R.F., Ippolito, G.C., Zemlin, M., 2012. Immunoglobulin analysis tool: a novel tool for the analysis of human and mouse heavy and light chain transcripts. Front. Immunol. 3, 176. https://doi.org/10.3389/fimmu.2012.00176.

Shirai, H., Kidera, A., Nakamura, H., 1996. Structural classification of CDR-H3 in antibodies. FEBS Lett. 399, 1–8.

Sievers, F., Higgins, D.G., 2014. Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. Methods in Molecular Biology, Clifton, N.J. https://doi.org/10.1007/978-1-62703-646-7_6. pp. 105–116.

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. Mol. Immunol. 59, 71–78. https://doi.org/10.1016/j.molimm.2014.01.010.

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

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

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. J. Anim. Sci. Biotechnol. 3, 1. https://doi.org/10.1186/2049-1891-3-18.

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

Tallmadge, R.L., Tseng, C.T., Felippe, M.J.B., 2014. Diversity of immunoglobulin lambda light chain gene usage over developmental stages in the horse. Dev. Comp. Immunol.

46, 171–179. https://doi.org/10.1016/j.dci.2014.04.001.
Tallmadge, R.L., Shen, L., Tseng, C.T., Miller, S.C., Barry, J., Felippe, M.J.B., 2015. Bone marrow transcriptome and epigenome profiles of equine common variable immunodeficiency patients unveil block of B lymphocyte differentiation. Clin. Immunol. 160, 261–276. https://doi.org/10.1016/j.clim.2015.05.005.
Vander Heiden, J.A., Yaari, G., Uduman, M., Stern, J.N.H., O'connor, K.C., Hafler, D.A., Vigneault, F., Kleinstein, S.H., 2014. PRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics 30, 1930–1932. https://doi.org/10.1093/bioinformatics/btu138.
Wagner, B., 2006. Immunoglobulins and immunoglobulin genes of the horse. Dev. Comp. Immunol. 30, 155–164. https://doi.org/10.1016/j.dci.2005.06.008.
Wagner, B., Miller, D.C., Lear, T.L., Antczak, D.F., 2004. The complete map of the ig heavy chain constant gene region reveals evidence for seven IgG isotypes and for IgD in the horse. J. Immunol. 173, 3230–3242. https://doi.org/10.4049/jimmunol.173.5.3230.
Walther, S., Rusitzka, T.V., Diesterbeck, U.S., Czerny, C.P., 2015. Equine immunoglobulins and organization of immunoglobulin genes. Dev. Comp. Immunol. 53, 303–319. https://doi.org/10.1016/j.dci.2015.07.017.
Weiss, S., Lehmann, K., Raschke, W.C., Cohn, M., 1984. Mice completely suppressed for the expression of immunoglobulin kappa light chain. Proc. Natl. Acad. Sci. U. S. A. 81, 211–215. https://doi.org/10.1073/pnas.81.1.211.
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. Proc. Natl. Acad. Sci. 110, 2993–2998.
Wu, L., Oficjalska, K., Lambert, M., Fennell, B.J., Darmanin-Sheehan, A., Ni Shuilleabhain, D., Autin, B., Cummins, E., Tchistiakova, L., Bloom, L., Paulsen, J., Gill, D., Cunningham, O., Finlay, W.J.J., 2012. Fundamental characteristics of the immunoglobulin VH repertoire of chickens in comparison with those of humans, mice, and camelids. J. Immunol. 188, 322–333. https://doi.org/10.4049/jimmunol.1102466.
Yaari, G., Kleinstein, S.H., 2015. Practical guidelines for B-cell receptor repertoire sequencing analysis. Genome Med. 7, 1–14. https://doi.org/10.1186/s13073-015-0243-2.
Ye, J., Ma, N., Madden, T.L., Ostell, J.M., 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. 41, 34–40. https://doi.org/10.1093/nar/gkt382.
Zemlin, M., Klinger, M., Link, J., Zemlin, C., Bauer, K., Engler, J.A., Schroeder, H.W., Kirkham, P.M., 2003. Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. J. Mol. Biol. 334, 733–749. https://doi.org/10.1016/j.jmb.2003.10.007.