Organization and transcription analyses of the immunoglobulin genes in cattle and horses

Dissertation
to obtain the Ph.D. degree
in the Ph.D. Program for Agricultural Sciences in Göttingen (PAG)
at the Faculty of Agricultural Sciences,
Georg-August-University Göttingen, Germany

presented by

Stefanie Walther

born in Sangerhausen, Germany

Göttingen, May 2016

D7

1. Name of supervisor: Prof. Dr. Dr. Claus-Peter Czerny

2. Name of co-supervisor: Prof. Dr. Sven König

Date of dissertation: May 12, 2016

Table of Contents

Summary **1**

General Introduction **4**

References 6

Chapter 1: Immunoglobulins **7**

General Structure of immunoglobulins 8

Immunoglobulin heavy chains 10

Immunoglobulin light chains 13

Development of the immunoglobulin repertoire 16

References 20

Chapter 2: Equine Immunoglobulins **25**

Equine immunoglobulins and organization of immunoglobulin genes **26**

Abstract 27

1. Introduction 28

2. Immunoglobulins in equine offspring 30

3. The equine immunoglobulin heavy chain gene locus 31

4. Equine immunoglobulin light chain gene loci 44

5. Transcription analyses of heavy and light chain genes 55

6. New allotypic variants of IGLC 58

7. Future Directions 60

Acknowledgments 61

References 62

Transcriptional analysis of equine λ-light chains in the horse breeds Rhenish-German Coldblood and Hanoverian Warmblood **69**

Abstract 70

1. Introduction 71

2. Material and methods 73

3. Results 78

4. Discussion 97

Acknowledgment 101

Appendix A. Supplementary data 101

References 103

Chapter 3: Bovine Immunoglobulins 107

Bovine immunoglobulin heavy and light chains 108

Bovine immunoglobulin heavy chain gene locus 108

The bovine immunoglobulin light chain gene loci 115

Development of B cells and the bovine antibody repertoire 119

References 122

Exceptionally Long CDR3H Are Not Isotype Restricted in Bovine Immunoglobulins 128

Abstract 129

Introduction 130

Materials and Methods 132

Results 136

Discussion 144

Supporting Information 147

Acknowledgments 164

References 165

Development of a bioinformatics framework for the detection of gene conversion and the analysis of combinatorial diversity in immunoglobulin heavy chains in four cattle breeds 169

Abstract 170

Author Summary 171

Introduction 172

Results 175

Discussion 197

Material and Methods 202

Acknowledgments 209

References 210

Supporting information 214

Chapter 4: General Discussion 231

General Discussion 232

Conclusions 244

Future Prospects 246

References 248

Chapter 5: Appendix 254

Curriculum Vitae 255

Summary

Initial studies on genetic aspects of immunoglobulins were performed on humans and mice but were successfully applied to various other animals such as chicken, rabbit, swine, cattle, and horses, too. Especially in cattle and horses, fundamental research in immunoglobulin genetics still needs more attention to complete previous information such as the number of available gene segments, gene families, and allotypes of different isotypes of the immunoglobulin heavy and light chains. Results will enable the analysis and generation of synthetic recombinant antibodies, as well as an alternating treatment of infectious diseases to prevent resistance to antibiotics.

As reviewed in the first publication, the understanding of the organization of equine immunoglobulin genes has increased significantly in the recent years. For equine heavy chains, 52 immunoglobulin heavy chain variable gene segments (IGHV), 40 immunoglobulin heavy chain diversity gene segments (IGHD), 8 immunoglobulin heavy chain joining gene segments (IGHJ) and 11 immunoglobulin heavy chain constant region genes (IGHC) are present. Seven of these IGHCs are gamma chain genes. Sequence diversity is increasing between fetal, neonatal, foal and adult age. The kappa light chain contains 60 immunoglobulin kappa light chain variable gene segments (IGKV), 5 immunoglobulin kappa light chain joining gene segments (IGKJ) and 1 immunoglobulin kappa light chain constant region gene (IGKC), whereas there are 144 immunoglobulin lambda light chain variable gene segments (IGLV), 7 immunoglobulin lambda light chain joining gene segments (IGLJ), and 7 immunoglobulin lambda light chain constant region genes (IGLC) for the lambda light chain, which is expressed predominantly in horses. A decrease in IGLVs is noted during age development, although nucleotide diversity and significant differences in gene usage increased. A standardization of the existing nomenclature of immunoglobulin genes is suggested.

The first experimental study focused on the identification of allotypic variants of equine IGLC and differences in the expression of IGLV within and between the two horse breeds Rhenish-German Coldblood (RGC) and Hanoverian Warmblood (HW). The two breeds differ in stud book size and breeding goals. After PCR amplification of cDNA and subcloning, 120 samples per breed were isolated and sequenced. Statistical analysis of transcription frequencies were performed applying non-parametric tests. The significant majority of the sequences represented IGLC6/7 in both breeds,

whereas IGLC1, IGLC4, and IGLC5 occurred in significant different frequencies per breed. Five allotypic IGLC1 variants, four allotypic IGLC5 variants, and three allelic as well as two allotypic IGLC6/7 variants were identified in breed specific proportions. Eleven out of 144 known IGLV segments were transcribed of which IGLV15 and IGLV17 were preferred significantly. IGLV25 displayed significant differences in the rearrangement between both breeds. In addition, the pseudogenes IGLV101ψ and IGLV74ψ were also identified. Rearrangements with IGLC genes showed significant differences for IGLV15 in both breeds, whereas IGLV25 also revealed significant differences between the breeds. The transcriptional orientation of the functional segments had no influence on the occurrence of the IGLV.

The second experimental study carried out in cattle dealt with two main topics. On the one hand it focused on the third complementarity determining region of the bovine heavy chain (CDR3H) whose exceptional length previously was described as a specificity of bovine IgG and IgM. On the other hand, the genomic organization of the immunoglobulin heavy chain locus was analyzed with a special focus on the number of IGHV. After isotype-specific cDNA-PCR, subcloning of 20 DNA plasmids per immunoglobulin isotype and sequence analyzes of the variable regions, we proved the existence of exceptionally long CDR3H in all five bovine isotypes. The sequences of CDR3H belong to three distinct groups and possess ≤10, 12 to 31 or ≥48 amino acid residues. Hydrophilic amino acid residues dominated in long and intermediate long CDR3H, while short CDR3H possessed hydrophobic amino acid residues, too. All sequences with exceptionally long CDR3H were related to the germline IGHV10. Further, the germline IGHD2, with 148 bp in size, contributes to those CDR3H.

The genomic organization of the bovine immunoglobulin heavy-chain locus was analyzed using the current genome assembly, *Bos taurus* NCBI build 6.1. A main locus was identified on BTA21. Additional exons coding for immunoglobulin heavy chain variable (IGHV), diversity (IGHD), and joining (IGHJ) segments, as well as for the constant regions of different isotypes, were localized on BTA7, BTA8, BTA20, and on unplaced contigs, too. Altogether, 36 IGHV were detected of which 13 are putatively functional. For the phylogenetic analysis, the complete nucleotide sequences of the 36 bovine IGHV segments were aligned with one member of the human IGHV families 1 to 7. Results proved the existence of two bovine IGHV families (boVH1, boVH2). The boVH1 comprises all functional segments. This study substantially improved the understanding of the generation of immunoglobulin diversity in cattle.

The third study aimed to gain more insight into the combinatorial diversity, somatic hypermutations and putative gene conversions of IgG in the four cattle breeds Aubrac, German Simmental, German Black Pied, and Holstein Friesian. For the more detailed analysis of rearranged bovine heavy chain immunoglobulin variable regions, a new bioinformatics framework was developed by combining and refining widely used alignment algorithms. Immunoglobulin heavy chains possessing exceptionally long CDR3Hs can now be analyzed specifically, as well as the dominantly transcribed IGHV, IGHD, and IGHJ segments and their recombination.

The use of 15 different IGHV segments, 21 IGHD segments, and 2 IGHJ segments was investigated with significant different transcription levels within the breeds. There are preferred rearrangements within the 3 groups of CDR3H lengths. In sequences of group 1 (≤10 aa) and 3 (≥48 aa) a lower number of recombinations were observed than in sequences of group 2 (11-47 aa). The combinatorial diversity revealed 162 significantly different rearrangements of germline IGHV, IGHD, and IGHJ segments. The few preferably rearranged gene segments within group 3 CDR3H regions are supposed to indicate specialized antibodies because this length is unique in cattle.

The main result of this study enabled by the new bioinformatics framework, is the strong evidence for gene conversion as a rare event using pseudogenes fulfilling all definitions for this particular diversification mechanism.

In conclusion, this thesis contributes to a more detailed understanding of the expressed immunoglobulin repertoire in cattle and horses. Breed and husbandry conditions are supposed to influence the repertoire significantly. This thesis also highlights that the bovine heavy chain diversity is not restricted to the use of a limited number of germline genes although there are preferred rearrangements within the three groups of CDR3H lengths. These results will be of future importance in analyzing seroconversion data after infection or vaccination, as well as determining breed specific differences to select healthy, robust animals.

General Introduction

Immunoglobulin genetics focuses on the special genetic aspects of immunoglobulins. As relatively few research groups work on immunogenetics, fundamental research is still necessary. Consequently no definite numbers of germline gene segments leading to immunoglobulins, gene families and allotypic variants are known in many species. Nevertheless, in the last few decades, this field of research has attracted higher attention because of its increasing importance in the regional or global eradication and monitoring of several infectious diseases. The availability of new high throughput technologies and descending prices facilitate and advance the experimental work flow in analyzing the entire immunoglobulin repertoire. Initial studies were performed on humans and mice but successful techniques were applied to various other animals such as chicken, rabbit, cattle, and horses, too. Especially in cattle and horses, investigation in immunoglobulin genetics still needs more attention to complete previous information about fundamentals such as the number of available gene segments, gene families, and allotypes of different isotypes of the immunoglobulin heavy and light chains. Results will contribute to the analysis and generation of synthetic recombinant species-specific antibodies. The production of antibodies from animals may be replaced and first of all, the results will enable an alternating treatment of infectious diseases to avoid antibiotics and resulting resistance of antigens. Side effects of conventional therapeutics might be excluded and higher antigen specificity will be achieved. Therefore, recombinant antibodies and antibody fragments are important tools for research, diagnostics and therapy (Hust et al. 2002; Hust and Dubel 2004). A widely used method for the selection of recombinant antibody fragments is the phage display (Smith 1985; Taussig et al. 2007). Further, monoclonal antibodies have been used successfully for the therapeutic treatment of many disorders, including inflammatory and putative autoimmune diseases as they bind to cell-specific antigens and mediate immune response (Hohlfeld and Wekerle 2005). By adding an appropriate constant domain, a promising antibody fragment is converted into any antibody isotype, for example IgG from different species (Moutel et al. 2009).

This present thesis was performed to contribute to a more detailed understanding of immunoglobulin diversity in cattle and horses. One of the major scopes was to investigate the bovine germline heavy chain gene segments *in silico*, which represent the theoretical immunoglobulin repertoire. The expressed repertoire was then

investigated in four different cattle breeds by transcriptional analyses and statistical methods. Further, the bovine specific mechanism of exceptionally long CDR3H contributing to diversity was proven in all five immunoglobulin isotypes. Gene conversion using pseudogenes was indicated for the first time in bovine heavy chains.
In addition, gene segments of transcribed equine lambda light chains were evaluated statistically in two different horse breeds. The previous findings of horse immunoglobulins were collated and reviewed.

The objectives of this dissertation were:

1. Equine immunoglobulins and organization of immunoglobulin genes

2. Transcriptional analysis of equine λ light chains in the horse breeds Rhenish-German Coldblood and Hanoverian Warmblood

3. Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins

4. Development of a bioinformatics framework for the detection of gene conversion and the analysis of combinatorial diversity in immunoglobulin heavy chains in four cattle breeds

References

Hohlfeld R, Wekerle H. 2005. Drug insight: using monoclonal antibodies to treat multiple sclerosis. *Nature clinical practice Neurology* **1**: 34-44.

Hust M, Dubel S. 2004. Mating antibody phage display with proteomics. *Trends in biotechnology* **22**: 8-14.

Hust M, Maiss E, Jacobsen HJ, Reinard T. 2002. The production of a genus-specific recombinant antibody (scFv) using a recombinant potyvirus protease. *Journal of virological methods* **106**: 225-233.

Moutel S, El Marjou A, Vielemeyer O, Nizak C, Benaroch P, Dubel S, Perez F. 2009. A multi-Fc-species system for recombinant antibody production. *BMC biotechnology* **9**: 14.

Smith GP. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**: 1315-1317.

Taussig MJ, Stoevesandt O, Borrebaeck CA, Bradbury AR, Cahill D, Cambillau C, de Daruvar A, Dubel S, Eichler J, Frank R et al. 2007. ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nature methods* **4**: 13-17.

Chapter 1:

Immunoglobulins

General Structure of immunoglobulins

The three-dimensional structure of immunoglobulins was determined by crystallization. Immunoglobulins (Ig) are described as a Y-shaped tetramer composed of four polypeptide chains containing two identical heavy chains (IGH) and two identical light chains (IGL) that are covalently connected by disulfide bonds (Figure 1) (Edelman 1973). In addition, non-covalent interactions contribute to the connection of IGH and IGL. While an IGH has a molecular mass of about 50 kDa, the molecular mass of an IGL is of about 25 kDa.

Both the IGH and IGL chains are further divided functionally and genetically into variable and constant domains that show a similar structural folding (Marchalonis et al. 2002; Ramsland and Farrugia 2002). Each chain possesses one variable domain (IGHV, IGLV), while the number of constant domains depends on the chain type and isotype. The variable domains of both chains are located at the arms of the Y-shaped structure (Figure 1). The variability, mediated by the first 110 amino-terminal amino acids, accounts for competent and highly specific antigen binding, whereby both the heavy and light chain variable region work together (Ramsland and Farrugia 2002; Torres et al. 2007). At the carboxyl terminus of the IGH, at least two constant regions (IGHC2 and IGHC3) are connected to the arms of the Y shaped structure. The IGHC3 regions of both IGH interact, but steric hindrance by carbohydrate side chains inhibits the interaction of the IGHC2 (Ramsland and Farrugia 2002).

In addition to flexible regions or hinge regions between the IGHC1 and IGHC2 connection between variable and constant domain is adjustable, too. Both torsion and bending are possible and enable simultaneous binding of antigen structures with different distances, as well as the interaction with signal proteins to enable effector mechanisms (Porter 1973; Ramsland and Farrugia 2002).

The proteases papain and pepsin cleave the polypeptide chain at specific amino acids generating fragments of different sizes (Porter 1973). While cleavage with papain occurs at the carboxyl terminal side of the disulfide bonds within the hinge region of an IgG resulting in two fragments, cleavage at the amino terminal side of the disulfide bonds using papain generates three fragments. The resulting fragments are named according to their characteristic features such as antigen binding ($F(ab')_2$ fragment) and the ability to crystallize (Fc fragment) (Yamaguchi et al. 1995). $F(ab')_2$ fragments contain the complete IGL, as well as the variable domain and the first constant region (IGHC1) of the heavy chain. They may still bind antigens. IGHC2 and IGHC3 belong to

the easy crystallizable Fc fragment. These regions mediate effector mechanisms after antigen binding.

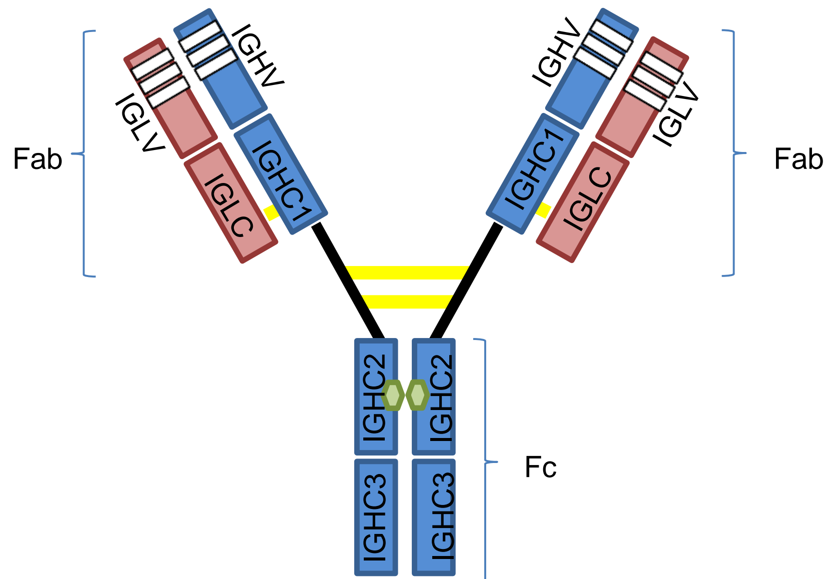


Figure 1: Structure of an immunoglobulin G (IgG)
The heavy chains are shown in blue, the light chains are red. IGLV and IGHV designate the variable regions of light and heavy chain. White boxes within IGLV and IGHV show the complementarity determining regions. The constant regions of light and heavy chain are marked by IGLC and IGHC1 to IGHC3. The hinge region is shown in black, disulfide bonds are colored in yellow. The antigen binding fragments are named Fab and Fc denominates the easy crystallizable fragment. Carbohydrate side chains are colored in green.

The secondary and tertiary structure of all domains is similar but there are differences between variable and constant domains (Ramsland and Farrugia 2002). Both domains consist of seven stacked antiparallel beta-strands that form a beta-barrel (Figure 2). At the end of each beta-strand and change of direction flexible turns are generated. Variable domains possess one more turn than constant domains. The three turns of each the IGH and IGL variable domain generate the area for antigen binding and consist of hypervariable regions contributing to the diversity of the immunoglobulin repertoire as these regions are characterized by extraordinary variability. They are located at restricted areas at the tip of the arms of the Y-shaped molecule (Figure 1, 2). The three-dimensional structure of their amino acid motif is complementary to the three-dimensional structure of the antigen epitope and is called complementarity determining region (CDR1-3, Figure 2) (Wu and Kabat 1970; Decanniere et al. 2000; Ramsland and Farrugia 2002). The length of the CDRs varies. Especially the CDR3 of the heavy chain is highly variable as described in cattle (Walther et al. 2013).

Depending on the amino acid sequence, electrostatic interactions, hydrogen bonds, Van-der-Waals-forces, and hydrophobic interactions also contribute to antigen binding (Braden and Poljak 1995; Braden et al. 1998). The complete specificity for antigens is generated by the combination of IGH and IGL.

The three hypervariable regions are interspersed by four less variable parts called framework regions (FR1-4). Their amino acid sequences within the variable and constant domain are very similar and they are responsible for stability and structure (Ramsland and Farrugia 2002).

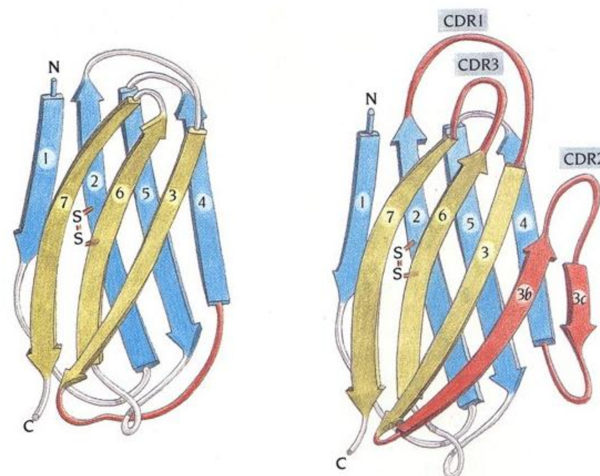


Figure 2: Secondary structure of the constant (left) and variable (right) domains
There are three hypervariable loops (CDR1-3) within the IGHV and the IGLV domains that account for most of the structural variability of the binding site. The CDRs are colored in red. The β-strands build the framework. The insertion of two β-strands (3b and 3c) linked by a loop containing the CDR2 in the variable domain is also marked red. (Branden 1991)

Immunoglobulin heavy chains

One immunoglobulin heavy chain is composed of one variable domain and a varying number of constant domains. The variable region is composed of three different gene segments that were randomly joined together during B-cell development and finally possesses around 110 amino acid residues. These gene segments are the variable (IGHV), diversity (IGHD), and joining (IGHJ) segments existing in multiple copies at the heavy chain locus (Figure 3) (Taussig 1988). Numbers of the gene segments are species specific. For instance in human, 123-129 IGHV, 27 IGHD, and 9 IGHJ segments are known, whereby not all of these segments are functional (Lefranc 2001). Non-functional gene segments are called pseudogenes. Mutations leading to premature stop codons prevent the formation of functional proteins. Further, changes within sequence regions necessary for gene recombination such as promotor or

recombination signal sequences lead to potentially functional gene segments, named open reading frame (ORF) (Lefranc 1998).

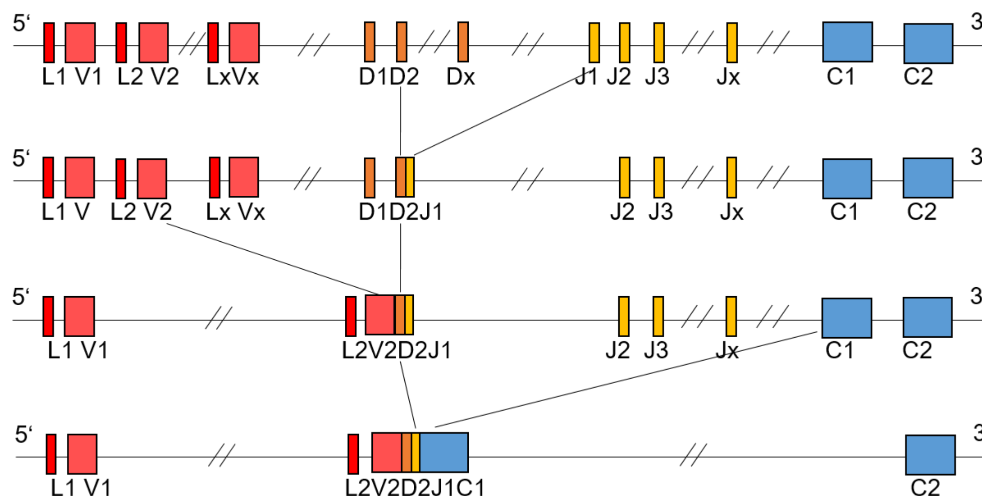


Figure 3: Genomic organization and recombination of the heavy chain gene segments
The immunoglobulin heavy chain variable (V1-x, light red), diversity (D1-x, orange), and joining (J1-x, dark yellow) gene segments, as well as the constant region genes (C1-2, blue) are organized in separated clusters on the genome. The recombination of single gene segments occurs at random in multiple steps and results in the deletion of gene segments between the recombined ones.

In general, the immunoglobulin heavy chain gene segments are located on a single chromosome. However, for instance in cattle there is evidence for gene segments located outside the major locus which are called orphan genes (Walther et al. 2013). Each IGHV is preceded by a leader sequence and they are grouped together upstream of an IGHD cluster. Downstream of IGHD, a separate cluster of IGHJ follows. The constant region genes are located 3' of the IGHJ cluster (Figure 3). Each IGHC corresponds to a different isotype.

The different variable gene segments are divided into several families where the members show sequence identities of at least 80% as recommended for mice (Brodeur and Riblet 1984). Families are further combined to clans. Families within one clan are more similar than families of different clans (Kirkham et al. 1992; Ota and Nei 1994). The variable gene segments of reptiles, amphibians, and mammals belong to the same three clans. During evolution, gene duplication and diversifications led to the variable gene segments known so far.

The number of constant domains in the immunoglobulin molecule depends on the isotype of which five different ones are described in humans, as well as in i.e. mice, cattle, and horses (Berens et al. 1997). These isotpyes are known as IgM, IgD, IgG,

IgE, and IgA (Figure 4), whereby they are encoded by μ, δ, γ, ε, and α genes (Woof and Burton 2004). IgM and IgD are coexpressed in naïve B-cells due to alternative splicing of mRNA or class switch recombination in artiodactyls. Activation by antigen contact leads to a switch of isotypes which is also called class switch recombination.

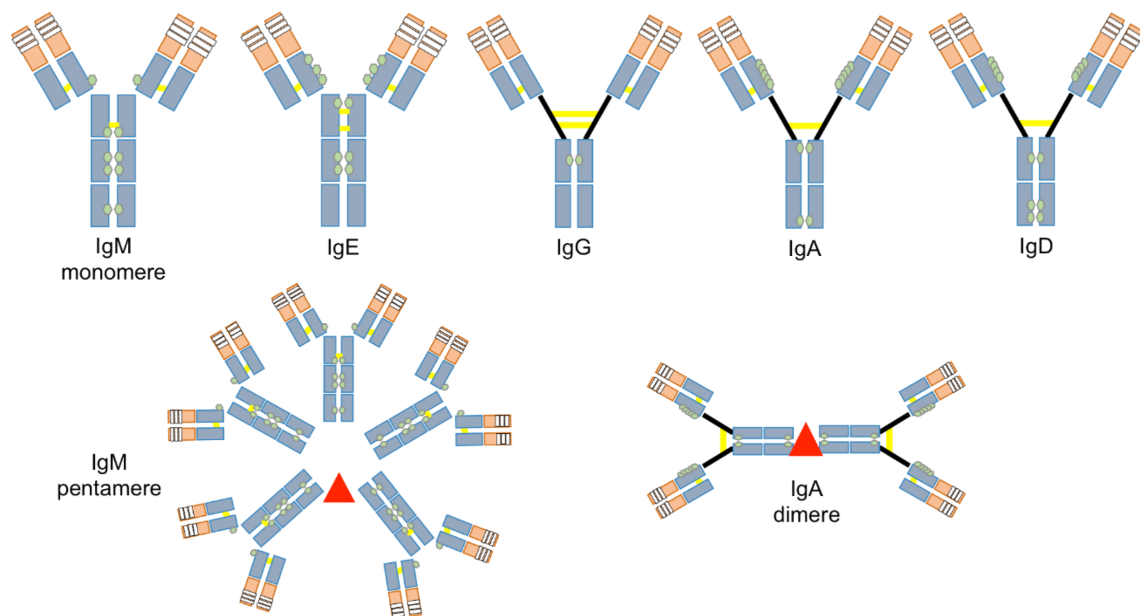


Figure 4: The 5 immunoglobulin isotypes and their multimers
The immunoglobulin variable domains are shown in light red, the constant region genes are shown in blue. The white boxes within the variable domains indicate the complementarity determining regions. Green spots symbolize carbohydrate side chains. Disulfide bonds are shown in yellow and the hinge region is black. The red triangle indicates an additional 15 kDa polypeptide chain contributing to polymerization.

Heavy chains of the α-, δ-, and γ-isotype possess a constant region composed of three domains (IGHC1, IGHC2, IGHC3) whereas the μ- and ε-isotypes have an additional fourth constant region (IGHC4). While IgA, IgD, and IgG possess a flexible hinge region, IgM and IgE achieve flexibility by bending of the antibody binding fragments (Mousavi et al. 1998; Janeway 2001). Furthermore, isotypes vary in the number of disulfide bonds between the chains, connected oligosaccharids, and length of the hinge region. Different sizes and compositions are characteristic for each isotpye. Consequently, α and γ possess 450 amino acids, whereas δ contains approximately 500 amino acids, and μ and ε are composed of 550 amino acids. Further, the antibody classes may occur as monomers (membrane bound in all isotypes), dimers, or multimers (Fudenberg and Warner 1970; Mestecky 1972). Secreted IgM and IgA may

occur as pentamers in plasma or dimers in mucous secretions, respectively (Figure 4). The IGHCs are responsible for complement activation, Fc receptor binding, serum half-life, and flexibility or stabilization of the variable region (Ravetch and Kinet 1991; Woof and Burton 2004).

Immunoglobulin light chains

Light chains contribute to antigen binding and enlarge variability of antibodies. They enable the expression of the heavy chains in pre-B-cells and therefore are responsible for the expression of B-cell receptors, as well as of secreted antibodies (Meffre et al. 2001). Immunoglobulin light chain constant regions support antigen recognition, stabilize the variable region, and are associated to the first constant region of heavy chain isotypes by specific amino acid residues that form an interdomain interface and contribute to non-covalent binding, as well as they contribute to covalent binding due to disulfide bonds (Padlan et al. 1986; Chen et al. 2008).

In mammals, two isotypes of IGL exist, which are called kappa (κ) and lambda (λ) (Korngold and Lipari 1956). As there is no shared origin for the light chain isotypes, they are polyphyletic (Sitnikova and Su 1998). The ratio of the isotypes depends on the species. A κ:λ ratio of 2:1 is found in human and swine, mice possess a ratio of 20:1, whereas in cattle and horse ratios of 1:20 and 1:13 are described. Consequently, in cattle and horses λ-light chains are predominantly expressed (Home et al. 1992; Arun et al. 1996; Butler 1998). Exceptions from these immunoglobulins with either κ- or λ-light chain are found in chicken, camel, and shark where solely λ-light chains are expressed or antibodies without any light chains and heavy chain homodimers were found (Ford et al. 1994; Wernery 2001; Saini et al. 2003). In *Xenopus laevis* an additional IGL of σ-isotype was described (Klein et al. 2002).The light chain isotypes can be distinguished by specific conserved amino acid motifs (Das et al. 2008). Hence, this isotype occurs in more than 90% of horse serum antibodies (Gibson 1974).

The light chain isotypes comprises of 211 to 217 amino acid residues (Janeway 2001). Characteristic amino acid motifs within the framework regions of the variable domain allow differentiation of the three IGL isotypes (Das et al. 2008). Distinctive features are the additional three amino acids in FR3 in σ-isotype compared to κ- and λ-light chains, 22 amino acids within FR1 in λ-isotypes and 23 amino acids building FR1 in the κ-isotypes. Further, amino acids Ser and Thr are distinguished at position 7 using the Ensembl annotation which is based on the IMGT nomenclature (Das et al. 2008). Also

the amino acid residue at position 53 differentiates κ- and λ-light chains (κ: Phe/Tyr vs. λ: Ala/Gly). A conserved amino acid motif in λ-isotype is Asp/Glu/Ala/Asp, which is missing in the σ- and κ-isotypes. Beside the differences within the variable gene segments, differences within the joining gene segments are described (σ: Ser 4, Ser 7; λ and κ: Gly 4, Thr 7; κ: Thr 2, Glu-Ile-Lys/Glu-Leu-Lys 10-12; λ and σ: Thr-Val-Leu/Thr-Val-Thr und Ile-Val-Thr 10-12). Specific amino acids at the positions 14, 32, 34, 79, and 91 using the Ensembl annotation enable to discriminate the constant regions of σ-light chains from κ- and λ-light chains, while particular amino acids at positions 17, 56, 60, 65, and 102 are responsible for the differentiation of κ- and λ-isotypes (Das et al. 2008).

Genes coding for light chain isotypes are located on different chromosomes. While κ- and λ-light chains show similar differences in their sequences compared to the σ-light chains, the genomic organization of joining gene segments and constant region genes of σ- and κ-light chains is analogical and differs from the organization found for λ (Das et al. 2008). Joining gene segments and constant region genes in σ- and κ-loci have an own cluster, whereas in the λ-locus joining gene segments and constant region genes cluster pairwise (Figure 5).

Hitherto no functional differences were described between the light chain isotypes although they appear in connection to specific diseases (Das et al. 2008). For instance, allotypic markers of human light chains were associated with the susceptibility to different infectious diseases (Pandey et al. 1995; Pandey 2000; Giha et al. 2009).

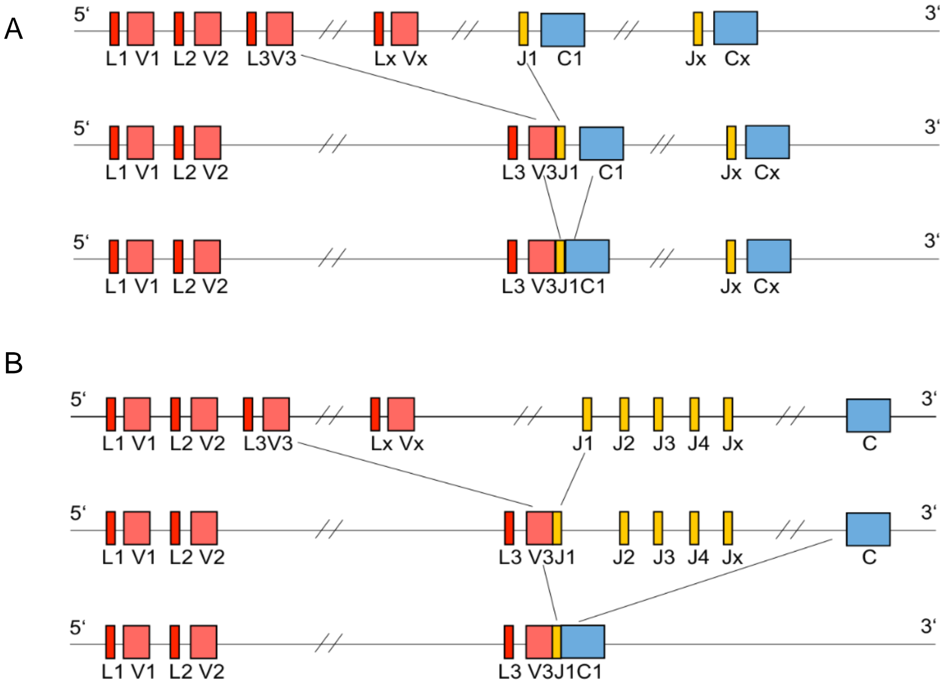


Figure 5: Genomic organization and recombination of the light chain gene segments
The immunoglobulin light chain variable (V1-x, light red), diversity (D1-x, orange), and joining (J1-x, dark yellow) gene segments, as well as the constant region genes (C/ Cx, blue) are organized in separated clusters on the genome. (A) Lambda light chain genes, the joining and constant region genes occur pairwise. (B) Kappa light chain genes, joining gene segments and the constant region gene are separated. The recombination of single gene segments occurs at random in multiple steps and results in the deletion of gene segments between the recombined ones.

Development of the immunoglobulin repertoire

The vertebrate immune system is able to produce a large diversity of antibodies with different specificities from a relatively modest number of gene segments (Parng et al. 1996). Therefore, immunoglobulins are a major component of the humoral immunity (Saini et al. 2007). Immunoglobulins are produced by B-cells, whereas each of these cells is specific for one antigen. Consequently, the immunoglobulin repertoire at a certain point in time is restricted by the number of B-cells and depends on antigen contacts (Janeway CA Jr 2001).

In one individual, the repertoire of immunoglobulins is immense. Immunoglobulins are produced by B-lymphocytes and plasma cells and may be membrane-bound (B-cell receptors) or secreted proteins (antibodies), which become diversified additionally. The development of the whole immunoglobulin repertoire depends on different primary and secondary mechanisms. The original repertoire is generated by the combinatorial diversity due to heavy and light chain pairing, random gene rearrangements, as well as the junctional diversity, which introduces insertions and deletions of nucleotides at the recombination site. Antigen contact and effector functions of the expressed antibodies further increase the primary repertoire. This secondary part of diversification is based on somatic hypermutations, gene conversion, and class switch recombination.

During the early development of lymphatic progenitor B-cells, variable and constant domains are joined together by somatic recombinations of separate heavy and light chain variable (IGHV, IGLV), diversity (IGHD), joining (IGHJ, IGLJ), and constant (IGHC, IGLC) germline components (Tonegawa 1983). In B-lymphocytes, the heavy chain rearrangement precedes the rearrangement of the light chains (Alt et al. 1981). Thus, recombination process starts within the heavy chain locus in the pro-B-cells. The separate gene segments rearrange together to form one complete variable domain exon. This process depends on recombination signal sequences (RSS) consisting of two conserved parts, the heptamer and the nonamer that are separated by a 12 or 23 bp spacer. The heptamer is directly connected to the gene segment. The spacer length is specific for the segment type e.g. the spacer following the IGLV of the λ-light chain (IGVL) has 23 bp, whereas the nonamer and the heptamer of the joining segment (IGJL) are separated by a 12 bp spacer. Spacer lengths between heptamer and nonamer of the IGLV and IGLJ of the κ-light chains (IGVK and IGKJ) possess 12 bp and 23 bp, respectively. Due to the spacer length, the heptamer and nonamer bind to the protein complex catalyzing the somatic recombination (Kim et al. 2015;

Lapkouski et al. 2015). This process follows the 12/23 rule and uses gene segments located on the same chromosome. During the process of recombination, the DNA located between the two joined segments is deleted (Sakano et al. 1979; Akira et al. 1987). After effective rearrangement, a µ-chain is expressed and associates with surrogate light chains. These surrogate light chains simulate the variable and constant region of the light chains that are not expressed yet at this step in B-cell development. Their expression is caused by the transcription factors E2A and EBF. This first checkpoint induces the completion of the heavy chain rearrangement and results in allelic exclusion (Loffert et al. 1996; Melchers et al. 1999; Vettermann and Schlissel 2010). Consequently, only one of the two alleles is expressed in one pro-B-cell. Subsequently, pro-B-cells divide and result in a large number of pro-B-cells that contain the same heavy chain and develop into pre-B-cells. The rearrangement of the light chain genes starts and is repeated until a productive light chain emerges. In case of unsuccessful recombinations of one light chain isotype, the rearrangement may also switch to the second light chain isotype. This process is called light chain rescue to prevent cell death. Also during rearrangements in the light chain genes, allelic exclusion and isotype exclusion occur. Therefore, just one light chain isotype is transcribed in one B-cell (Arakawa et al. 1996; Loffert et al. 1996; Melchers et al. 1999). Finally, the associated µ- and light chains are expressed as B-cell receptors on the surface of immature B-cells. Before these immature cells leave the bone marrow for the periphery, they undergo several types of negative selection, such as clonal deletion (Nemazee and Burki 1989), receptor editing (Gay et al. 1993; Tiegs et al. 1993), clonal anergy (Goodnow et al. 1988), or apoptosis to avoid autoreactivity (Levine et al. 2000).

Beside the random combination of different variable, diversity, and joining gene segments a junctional diversity occurs during rearrangements by the insertion or deletion of nucleotides within the joining area of IGHV-IGHD, IGHD-IGHJ, IGKV-IGKJ, and IGLV-IGLJ, respectively. The inserted nucleotides are called N- and P-nucleotides that are characterized by the addition of non-encoded (N-) or palindromic (P-) nucleotides catalyzed by the enzymes terminal deoxynucleotidyltransferase, as well as RAG-proteins and the artemis enzyme complex. Further, exonucleases may delete nucleotides. These mechanisms result in an increased variability of nucleotides and amino acid residues within the CDR3 of both heavy and light chains, as well as in length differences. Following, the identification of the originating IGHD is sometimes difficult or may even be impossible in some cases.

The secondary diversification mechanisms, somatic hypermutation, gene conversion, and class switch recombination introduce changes into the sequence of functional, secreted antibodies (Figure 6). All these mechanisms are initiated on single stranded DNA by the activation induced cytidine deaminase (AID) (Di Noia and Neuberger 2002; Petersen-Mahrt et al. 2002; Bransteitter et al. 2003; Chaudhuri et al. 2003).

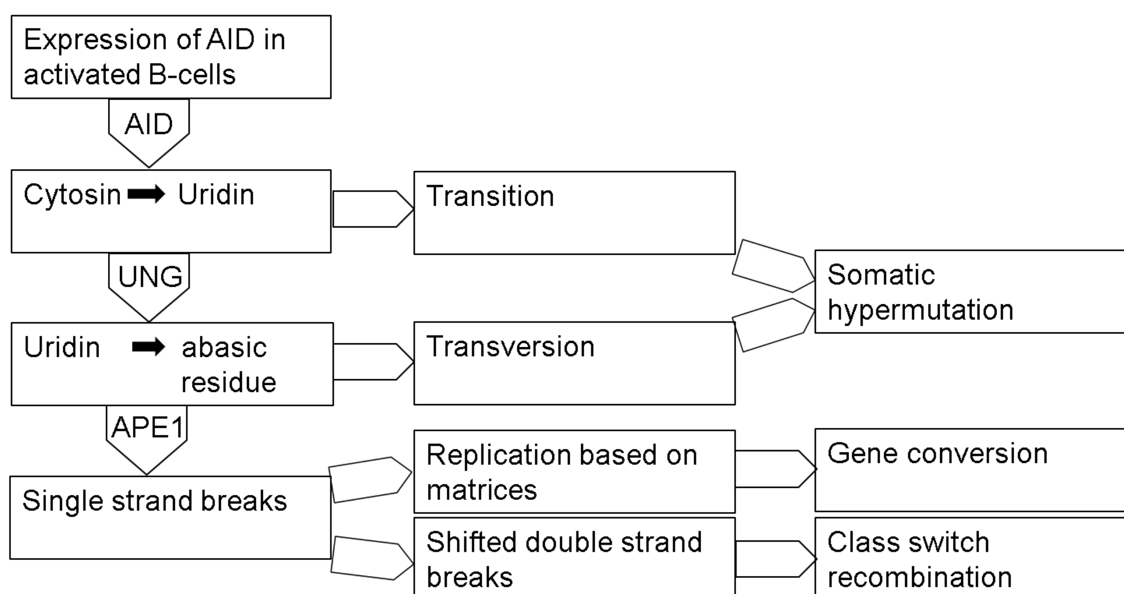


Figure 6: Secondary diversification mechanisms
AID: activation induced cytidine deaminase; UNG: uracil DNA glycosylase; APE1: apurinic/ apyrimidinic endonuclease 1 (modified from "Janeway Immunologie", (Murphy 2009))

Deamination of cytidine to uracil by AID results in transition mutations, one kind of somatic hypermutation. A second kind of somatic hypermutation, transversion mutation, is generated by the base excision enzyme uracil DNA glycosylase (UNG), which deletes the uracil generated by AID. Somatic hypermutation occurs in B-cells located in peripheric lymphoid tissues after stimulation by an antigen and generates point mutations within the complete exon of the variable region of both the heavy and light chains (Muramatsu et al. 2000). While silent mutations accumulate also in FRs, mutations affecting amino acid substitutions and protein structure are mainly found in the CDRs (Maizels 2005; Neuberger 2008). Certain major hotspots targeting somatic hypermutation are known. For instance, a cytosine (C) residue is more likely to be mutated if it is part of a WRCY motif (W= A or T, R= A or G, Y= T or C) and also WA motifs (Li et al. 2004; Wang et al. 2010). If this leads to improved affinity for antigens, the resulting affinity maturation of cells expressing such immunoglobulins leads to

further expansion. After deamination of cytidine by AID and deletion of uracil by UNG an abasic residue exists. This abasic residue is exised by the apurinic/ apyrimidinic endonuclease 1 (APE1) leading to a single strand break which is assumed to result in matrix-based replication and gene conversion or shifted double strand brakes and class switch recombination, respectively. Gene conversion also affects the complete variable regions. Parts of pseudogene sequences replace the original sequence, which is assumed to be a homology based repair mechanism characteristically found in chicken and rabbits. This mechanism increases antibody diversification in species with small number of germline gene segments for the variable region such as chicken, sheep, rabbits, cattle, and is assumed in horses (Reynaud et al. 1985; Reynaud et al. 1987; Reynaud et al. 1989; Becker and Knight 1990; Reynaud et al. 1995; Parng et al. 1996; Sun et al. 2010). Class switch recombination concerns the constant region. Hence, the same heavy chain variable region associates with different IGHC during one immune response. Once a B-cell was stimulated by an antigen, CD40 and toll like receptors provide the activation for B cells to undergo class switch recombination. For this purpose, toll like receptors on the surface of major B cells respond to microbial products such as lipopolysaccharides and CpG-enriched DNA. Both ligands (lipopolysaccharides and CpG DNA) of toll like receptors stimulate cell proliferation, AID expression and class switch recombination, as well as differentiation into antibody secreting cells by signals transduced through the toll like receptors (Edry et al. 2008; Pone et al. 2012). During class switch recombination, the primary IgM is replaced by an alternative IGHC isotype resulting in an increased functional diversity of the immunoglobulin molecule. This process is directed by repetitive nucleotide sequences (switch-regions). For instance, common elements are GGGGT, GGGCT, or GAGCT within the introns upstream of the IGHC and downstream of the IGHJ. Switch regions possess tandem repeats of short consensus elements that function as hotspot target for the AID. Class switches are supposed to occur by a non-homologous end joining mechanism. Cytokines produced by T-helper cells and dendritic cells regulate this intrachromosomal deletional recombination by inducing transcription form promotors located upstream to the acceptor switch region. Consequently, cytokines target the class switch recombination to a specific isotype (Stavnezer et al. 2008). In mice, IL-4 induces the switch to IgG1 and IgE, while TGF-β induces the switch to IgG2b and IgA (Stavnezer and Amemiya 2004). Class switch recombination always results in functional immunoglobulins.

References

Akira S, Okazaki K, Sakano H. 1987. Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. *Science* **238**: 1134-1138.

Alt F, Rosenberg N, Lewis S, Thomas E, Baltimore D. 1981. Organization and reorganization of immunoglobulin genes in A-MULV-transformed cells: rearrangement of heavy but not light chain genes. *Cell* **27**: 381-390.

Arakawa H, Shimizu T, Takeda S. 1996. Re-evaluation of the probabilities for productive arrangements on the kappa and lambda loci. *International immunology* **8**: 91-99.

Arun SS, Breuer W, Hermanns W. 1996. Immunohistochemical examination of light-chain expression (lambda/kappa ratio) in canine, feline, equine, bovine and porcine plasma cells. *Zentralblatt fur Veterinarmedizin Reihe A* **43**: 573-576.

Becker RS, Knight KL. 1990. Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. *Cell* **63**: 987-997.

Berens SJ, Wylie DE, Lopez OJ. 1997. Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. *International immunology* **9**: 189-199.

Braden BC, Goldman ER, Mariuzza RA, Poljak RJ. 1998. Anatomy of an antibody molecule: structure, kinetics, thermodynamics and mutational studies of the antilysozyme antibody D1.3. *Immunological reviews* **163**: 45-57.

Braden BC, Poljak RJ. 1995. Structural features of the reactions between antibodies and protein antigens. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **9**: 9-16.

Branden CT, J. 1991. *Introduction to Protein Structure*. Garland Publishing.

Bransteitter R, Pham P, Scharff MD, Goodman MF. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 4102-4107.

Brodeur PH, Riblet R. 1984. The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *European journal of immunology* **14**: 922-930.

Butler JE. 1998. Immunoglobulin diversity, B-cell and antibody repertoire development in large farm animals. *Rev Sci Tech* **17**: 43-70.

Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **422**: 726-730.

Chen L, Li M, Li Q, Yang X, An X, Chen Y. 2008. Characterization of the bovine immunoglobulin lambda light chain constant IGLC genes. *Veterinary immunology and immunopathology* **124**: 284-294.

Das S, Nikolaidis N, Klein J, Nei M. 2008. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 16647-16652.

Decanniere K, Muyldermans S, Wyns L. 2000. Canonical antigen-binding loop structures in immunoglobulins: more structures, more canonical classes? *Journal of molecular biology* **300**: 83-91.

Di Noia J, Neuberger MS. 2002. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**: 43-48.

Edelman GM. 1973. Antibody structure and molecular immunology. *Science* **180**: 830-840.

Edry E, Azulay-Debby H, Melamed D. 2008. TOLL-like receptor ligands stimulate aberrant class switch recombination in early B cell precursors. *International immunology* **20**: 1575-1585.

Ford JE, Home WA, Gibson DM. 1994. Light chain isotype regulation in the horse. Characterization of Ig kappa genes. *Journal of immunology* **153**: 1099-1111.

Fudenberg HH, Warner NL. 1970. Genetics of immunoglobulins. *Advances in human genetics* **1**: 131-209.

Gay D, Saunders T, Camper S, Weigert M. 1993. Receptor editing: an approach by autoreactive B cells to escape tolerance. *The Journal of experimental medicine* **177**: 999-1008.

Gibson D. 1974. Structural studies on normal horse immunoglobulin light chains. Detection of k-type N-terminal sequences. *Biochemistry* **13**: 2776-2785.

Giha HA, Nasr A, Iriemenam NC, Arnot D, Troye-Blomberg M, Theander TG, Berzins K, ElGhazali G, Pandey JP. 2009. Antigen-specific influence of GM/KM allotypes on IgG isotypes and association of GM allotypes with susceptibility to Plasmodium falciparum malaria. *Malaria journal* **8**: 306.

Goodnow CC, Crosbie J, Adelstein S, Lavoie TB, Smith-Gill SJ, Brink RA, Pritchard-Briscoe H, Wotherspoon JS, Loblay RH, Raphael K et al. 1988. Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature* **334**: 676-682.

Home WA, Ford JE, Gibson DM. 1992. L chain isotype regulation in horse. I. Characterization of Ig lambda genes. *Journal of immunology* **149**: 3927-3936.

Janeway CA Jr TP, Walport M, et al. . 2001. The generation of diversity in immunoglobulins. . in *Immunobiology: The Immune System in Health and Disease* New York: Garland Science.

Janeway CA, Jr. 2001. How the immune system protects the host from infection. *Microbes and infection / Institut Pasteur* **3**: 1167-1171.

Kim MS, Lapkouski M, Yang W, Gellert M. 2015. Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature* **518**: 507-511.

Kirkham PM, Mortari F, Newton JA, Schroeder HW, Jr. 1992. Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. *EMBO J* **11**: 603-609.

Klein SL, Strausberg RL, Wagner L, Pontius J, Clifton SW, Richardson P. 2002. Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. *Dev Dyn* **225**: 384-391.

Korngold L, Lipari R. 1956. Multiple-myeloma proteins. III. The antigenic relationship of Bence Jones proteins to normal gammaglobulin and multiple-myeloma serum proteins. *Cancer* **9**: 262-272.

Lapkouski M, Chuenchor W, Kim MS, Gellert M, Yang W. 2015. Assembly Pathway and Characterization of the RAG1/2-DNA Paired and Signal-end Complexes. *The Journal of biological chemistry* **290**: 14618-14625.

Lefranc M-PaL, G. 2001. *The Immunoglobulin FactsBook*, Academic Press.

Lefranc MP. 1998. IMGT (ImMunoGeneTics) locus on focus. A new section of Experimental and Clinical Immunogenetics. *Experimental and clinical immunogenetics* **15**: 1-7.

Levine MH, Haberman AM, Sant'Angelo DB, Hannum LG, Cancro MP, Janeway CA, Jr., Shlomchik MJ. 2000. A B-cell receptor-specific selection step governs immature to mature B cell differentiation. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 2743-2748.

Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD. 2004. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev* **18**: 1-11.
Loffert D, Ehlich A, Muller W, Rajewsky K. 1996. Surrogate light chain expression is required to establish immunoglobulin heavy chain allelic exclusion during early B cell development. *Immunity* **4**: 133-144.
Maizels N. 2005. Immunoglobulin gene diversification. *Annu Rev Genet* **39**: 23-46.
Marchalonis JJ, Jensen I, Schluter SF. 2002. Structural, antigenic and evolutionary analyses of immunoglobulins and T cell receptors. *Journal of molecular recognition : JMR* **15**: 260-271.
Meffre E, Milili M, Blanco-Betancourt C, Antunes H, Nussenzweig MC, Schiff C. 2001. Immunoglobulin heavy chain expression shapes the B cell receptor repertoire in human B cell development. *The Journal of clinical investigation* **108**: 879-886.
Melchers F, ten Boekel E, Yamagami T, Andersson J, Rolink A. 1999. The roles of preB and B cell receptors in the stepwise allelic exclusion of mouse IgH and L chain gene loci. *Seminars in immunology* **11**: 307-317.
Mestecky J. 1972. Structure of antibodies. *Journal of oral pathology* **1**: 288-300.
Mousavi M, Rabbani H, Pilstrom L, Hammarstrom L. 1998. Characterization of the gene for the membrane and secretory form of the IgM heavy-chain constant region gene (C mu) of the cow (Bos taurus). *Immunology* **93**: 581-588.
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**: 553-563.
Murphy KT, P.; Walport, M. 2009. *Janeway Immunologie*, Spektrum Akademischer Verlag Heidelberg
Nemazee DA, Burki K. 1989. Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes. *Nature* **337**: 562-566.
Neuberger MS. 2008. Antibody diversification by somatic mutation: from Burnet onwards. *Immunol Cell Biol* **86**: 124-132.
Ota T, Nei M. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Molecular biology and evolution* **11**: 469-482.
Padlan EA, Cohen GH, Davies DR. 1986. Antibody Fab assembly: the interface residues between CH1 and CL. *Molecular immunology* **23**: 951-960.
Pandey JP. 2000. Immunoglobulin GM and KM allotypes and vaccine immunity. *Vaccine* **19**: 613-617.
Pandey JP, Elson LH, Sutherland SE, Guderian RH, Araujo E, Nutman TB. 1995. Immunoglobulin kappa chain allotypes (KM) in onchocerciasis. *The Journal of clinical investigation* **96**: 2732-2734.
Parng CL, Hansal S, Goldsby RA, Osborne BA. 1996. Gene conversion contributes to Ig light chain diversity in cattle. *Journal of immunology* **157**: 5478-5486.
Petersen-Mahrt SK, Harris RS, Neuberger MS. 2002. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**: 99-103.
Pone EJ, Xu Z, White CA, Zan H, Casali P. 2012. B cell TLRs and induction of immunoglobulin class-switch DNA recombination. *Front Biosci (Landmark Ed)* **17**: 2594-2615.
Porter RR. 1973. Structural studies of immunoglobulins. *Science* **180**: 713-716.
Ramsland PA, Farrugia W. 2002. Crystal structures of human antibodies: a detailed and unfinished tapestry of immunoglobulin gene products. *Journal of molecular recognition : JMR* **15**: 248-259.
Ravetch JV, Kinet JP. 1991. Fc receptors. *Annual review of immunology* **9**: 457-492.

Reynaud CA, Anquez V, Dahan A, Weill JC. 1985. A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell* **40**: 283-291.

Reynaud CA, Anquez V, Grimal H, Weill JC. 1987. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell* **48**: 379-388.

Reynaud CA, Dahan A, Anquez V, Weill JC. 1989. Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell* **59**: 171-183.

Reynaud CA, Garcia C, Hein WR, Weill JC. 1995. Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell* **80**: 115-125.

Saini SS, Farrugia W, Muthusamy N, Ramsland PA, Kaushik AK. 2007. Structural evidence for a new IgG1 antibody sequence allele of cattle. *Scandinavian journal of immunology* **65**: 32-38.

Saini SS, Farrugia W, Ramsland PA, Kaushik AK. 2003. Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted VH + Vlambda pairings. *International immunology* **15**: 845-853.

Sakano H, Huppi K, Heinrich G, Tonegawa S. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280**: 288-294.

Sitnikova T, Su C. 1998. Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. *Mol Biol Evol* **15**: 617-625.

Stavnezer J, Amemiya CT. 2004. Evolution of isotype switching. *Seminars in immunology* **16**: 257-275.

Stavnezer J, Guikema JE, Schrader CE. 2008. Mechanism and regulation of class switch recombination. *Annual review of immunology* **26**: 261-292.

Sun Y, Wang C, Wang Y, Zhang T, Ren L, Hu X, Zhang R, Meng Q, Guo Y, Fei J et al. 2010. A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. *Developmental and comparative immunology* **34**: 1009-1020.

Taussig MJ. 1988. Molecular genetics of immunoglobulins. *Immunol Suppl* **1**: 7-15.

Tiegs SL, Russell DM, Nemazee D. 1993. Receptor editing in self-reactive bone marrow B cells. *The Journal of experimental medicine* **177**: 1009-1020.

Tonegawa S. 1983. Somatic generation of antibody diversity. *Nature* **302**: 575-581.

Torres M, Fernandez-Fuentes N, Fiser A, Casadevall A. 2007. The immunoglobulin heavy chain constant region affects kinetic and thermodynamic parameters of antibody variable region interactions with antigen. *The Journal of biological chemistry* **282**: 13917-13927.

Vettermann C, Schlissel MS. 2010. Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunological reviews* **237**: 22-42.

Walther S, Czerny CP, Diesterbeck US. 2013. Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins. *PloS one* **8**: e64234.

Wang M, Rada C, Neuberger MS. 2010. Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *The Journal of experimental medicine* **207**: 141-153.

Wernery U. 2001. Camelid immunoglobulins and their importance for the new-born--a review. *Journal of veterinary medicine B, Infectious diseases and veterinary public health* **48**: 561-568.

Woof JM, Burton DR. 2004. Human antibody-Fc receptor interactions illuminated by crystal structures. *Nature reviews Immunology* **4**: 89-99.

Wu TT, Kabat EA. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *The Journal of experimental medicine* **132**: 211-250.

Yamaguchi Y, Kim H, Kato K, Masuda K, Shimada I, Arata Y. 1995. Proteolytic fragmentation with high specificity of mouse immunoglobulin G. Mapping of proteolytic cleavage sites in the hinge region. *Journal of immunological methods* **181**: 259-267.

Chapter 2:

Equine Immunoglobulins

The nature and extent of my contribution to the work was the following:

1. Equine immunoglobulins and organization of immunoglobulin genes

Nature of contribution	Extent of contribution
1. Scientific design	70%
2. Laboratory work	-
3. Evaluation	70%
4. Scientific Writing	80%

2. Transcriptional analysis of equine λ light chains in the horse breeds Rhenish-German Coldblood and Hanoverian Warmblood

Nature of contribution	Extent of contribution
1. Scientific design	50%
2. Laboratory work	-
3. Evaluation	70%
4. Scientific Writing	70%

Equine immunoglobulins and organization of immunoglobulin genes

Stefanie Walther[a], Tamara V. Rusitzka[a], Ulrike S. Diesterbeck[b], Claus-Peter Czerny[a]*

[a]Department of Animal Sciences, Institute of Veterinary Medicine, Division of Microbiology and Animal Hygiene, Faculty of Agricultural Sciences, Georg-August University Göttingen, Burckhardtweg 2, 37077 Göttingen, Germany
[b]Building 33, 33 North Drive, Bethesda, MD 20892

*Corresponding author:
Claus-Peter Czerny
Tel.: +49 551 3933375
Fax: +49 551 3913513
E-mail: cczerny@gwdg.de
Postal address: Prof. Claus-Peter Czerny, DVM, PhD
Department of Animal Sciences,
Georg-August University Göttingen
Burckhardtweg 2
37077 Göttingen, Germany

Published in “Developmental and Comparative Immunology”

DOI: 10.1016/j.dci.2015.07.017
URL: http://www.sciencedirect.com/science/article/pii/S0145305X15300227

Abstract

Our understanding of how equine immunoglobulin genes are organized has increased significantly in recent years. For equine heavy chains, 52 IGHV, 40 IGHD, 8 IGHJ and 11 IGHC are present. Seven of these IGHCs are gamma chain genes. Sequence diversity is increasing between fetal, neonatal, foal and adult age. The kappa light chain contains 60 IGKV, 5 IGKJ and 1 IGKC, whereas there are 144 IGLV, 7 IGLJ, and 7 IGLC for the lambda light chain, which is expressed predominantly in horses. Significant transcriptional differences for IGLV and IGLC are identified in different breeds. Allotypic and allelic variants are observed for IGLC1, IGLC5, and IGLC6/7, and two IGLV pseudogenes are also transcribed. During age development, a decrease in IGLVs is noted, although nucleotide diversity and significant differences in gene usage increased. The following paper suggests a standardization of the existing nomenclature of immunoglobulin genes.

Keywords: horse, immunoglobulin genes, immunoglobulin heavy and light chains, allotype, diversity, equine developmental stages

1. Introduction

Equine immunoglobulins have played a key role throughout the history of human and veterinary immunology. Since the 19th century, horses have been important blood donors for serum therapy in heterospecific hosts. In the early 20th century, common horse serum or serum from immunized horses was produced, especially for the treatment and prophylaxis of diphtheria in humans (Bingel, 1918; Daniels, 1921). Later, horses assisted with the production of sera not only for the treatment of diphtheria but also for the treatment of other human infectious diseases such as tuberculosis, tetanus, and pneumonia (Behring von, 1918; Cole and Moore, 1917; Glatman-Freedman and Casadevall, 1998; Winau and Winau, 2002). Albert Calmette employed this method to produce antivenoms against poisonous bites from snakes and spiders (Calmette, 1896; Hawgood, 1999). Today, equine immunoglobulins are also used to support immunosuppression after organ or stem cell transplantation, or to manage autoimmune diseases in humans (Leleu et al., 2006; Zand, 2006). However, repeated systemic injections of hyperimmune sera or polyclonal and monoclonal antibodies from different species resulted in serum sickness and therefore, were not a feasible option for repeated therapy (Lang et al., 2000; Theakston et al., 2003). Early studies on the structure and function of equine immunoglobulins have already generated a considerable amount of data on the characterization of equine immunoglobulin genes and their genomic organization (Helms and Allen, 1970; Hill and Cebra, 1965; Pahud and Mach, 1972; Rockey, 1967; Sandor et al., 1964a; Vaerman et al., 1971; Wagner, 2006; Weir et al., 1966; Zolla and Goodman, 1968). As is known for humans and mice (Edelman, 1973), and nearly all jawed vertebrates, equine immunoglobulins are heterotetramers with two identical heavy and light chains. Both of them can be divided functionally and genetically into a variable region and a constant region. The variable regions are created by the random fusion of germline variable (V), diversity (D), and joining gene segments (J) that are combined with a constant region gene (Tonegawa, 1983), which are found in species-specific numbers in the genome. The immunoglobulin diversity depends on several processes of combinatorial and junctional diversity due to the imprecise joining of the single gene segments and non-templated or palindromic nucleotide insertions between two adjacent gene segments, as well as somatic hypermutations. In addition, several secondary mechanisms such as gene conversion and isotype switch may increase diversity further. The resulting variable regions of the heavy and light chains together are responsible for antigen binding.

Highly specialized, complementary determining regions form a perfect counterpart of the antigen epitope and are stabilized by conserved framework regions (Kabat and Wu, 1991). The repertoire of immunoglobulins in one individual is immense because they are produced by B-lymphocytes and plasma cells and may be membrane-bound (B-cell receptors) or secreted proteins (antibodies), which additionally become diversified. In B-lymphocytes, the heavy chain rearrangement precedes the rearrangement of the light chains (Alt et al., 1981). In mammals there are two types of light chains – the lambda and kappa light chains – which are expressed in species-specific ratios. In contrast to humans and mice, where the kappa isotype dominates in serum antibodies, the lambda isotype is predominantly found in cattle and horses (Almagro et al., 1998; Arun et al., 1996). In addition to what is already known about the general structural features of equine immunoglobulins and their function, most recent studies on diversity provide the scientific basis for the production of highly specific and effective recombinant antibodies or antibody libraries.

2. Immunoglobulins in equine offspring

Evidently, as early as in the equine fetus, an initial B-cell repertoire is developed despite the lack of exogenous antigenic stimulation (Tallmadge et al., 2009). Corresponding antibodies such as IgM, IgG1, and IgG4/7 are detectable on a limited scale at birth. Nevertheless, newborn foals are immunocompetent but do not possess an effective humoral immunity to infections and, therefore, depend on the absorption of maternal colostral immunoglobulins by specialized cells lining the small intestine (Jenvey et al., 2012). During the first 6 h after birth, absorption is highest and decreases gradually within 24 h (Franz et al., 1998). After this time, absorption of antibodies is no longer possible (McGuire and Crawford, 1973) leading to the rapid decrease of IgA and IgG levels in the mare´s milk for the first days after parturition. Failure of passive transfer (FPT) – meaning the insufficient transfer of immunoglobulins via the mare´s colostrum in the first 12-24 h after birth – results in a considerably higher risk of sepsis, bacteremia and localized infections (Haas et al., 1996; Koterba et al., 1984). Both the foal and the dam may suffer from FPT, which could be attributed to poor colostrum quality, lack of colostrum ingestion, poor intestinal absorption or a combination of these factors (Drogoul et al., 2008). At least 60 g of Ig/l are regarded as a sufficient quality colostrum (Drogoul et al., 2008). Nevertheless 23% to 32% of mares produce colostrum of low qualities (LeBlanc et al., 1992). With up to 70%, IgG is the main isotype present (Turtinen and Allen, 1982). An average of about 100 g IgG is secreted per lactation. Furthermore, colostrum comprises 20% IgA that shows limited absorption but has a local protective local function within the digestive tract of newborn foals (Sedlinska et al., 2006). Similarly, horse type and breed are known to have an influence on the overall Ig concentration. Arabian and Quarter Horse mares show better colostrum qualities than Thoroughbred and Standardbred (Leblanc and Tran, 1987). Six hours after delivery, Haflinger exhibit higher whey protein amounts than Arabian and Trotter, but these quantities decrease more rapidly (Civardi et al., 2002). Age and rank of lactation, nutrition and body condition of mares, as well as vaccination programs, season, and temperature may be further variables concerning colostrum quality (Drogoul et al., 2008).

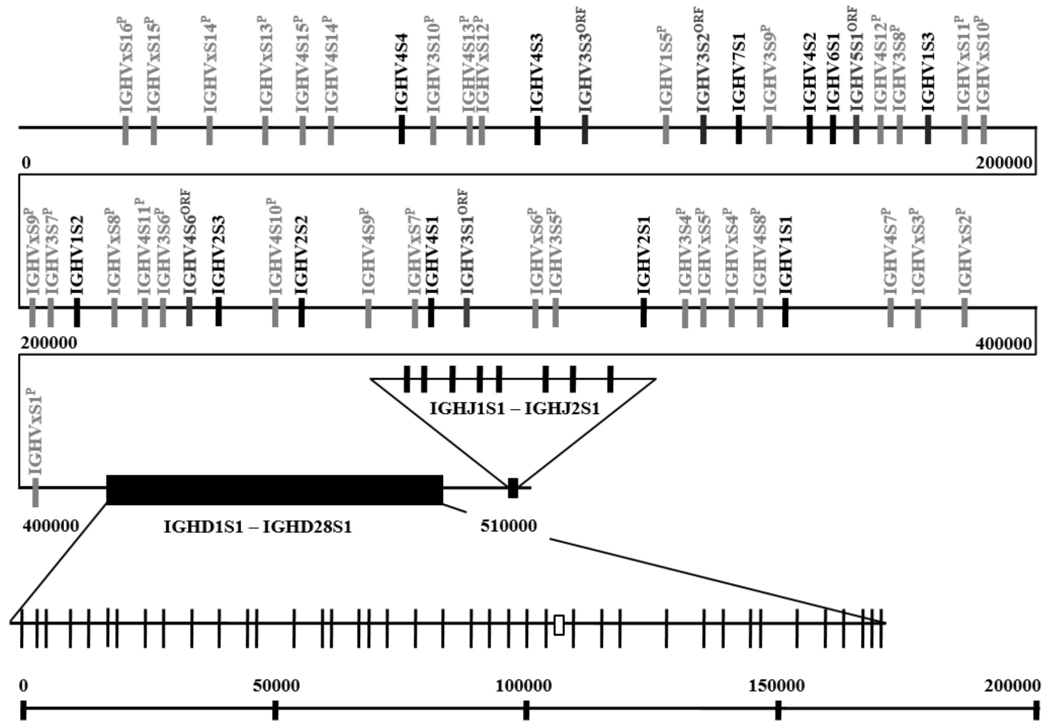
After maternal colostrum antibodies disappear and the appropriate antigenic stimulation is provided, the active antibody synthesis begins in the foal. The immune system starts the production of IgM followed by the other Ig classes (Wagner, 2006).

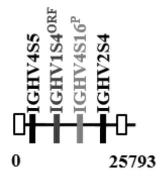
Endogenous IgG and IgA synthesis begins within the first four weeks of life. Stable levels were reached by 8 weeks of age (IgGa), 12 weeks of age (IgG(T) and IgA), and 51 weeks of age (IgGb) (Holznagel et al., 2003).

3. The equine immunoglobulin heavy chain gene locus

In silico analyses identified the heavy chain locus on two unplaced contigs, which are called Un0011 and Un0038 (Fig. 1a and b, Tables 1-3). The nomenclature used in previous studies to denominate Ig heavy and light chain gene segments varied, although they conformed to current definitions by the international ImMunoGeneTics (IMGT) information system, as well as to the WHO–IUIS Nomenclature Subcommittee for immunoglobulins and T-cell receptors (Hara et al., 2012; Lefranc, 2001b, 2007; Sun et al., 2010; Tallmadge et al., 2014; Tallmadge et al., 2013). Nevertheless, in the designation system (Tallmadge et al., 2014) used most recently, pseudogenes and open reading frames are not indicated precisely. This led us to propose a supplemented taxonomical designation for all known Ig heavy and light chain genes investigated in the most recent studies (Sun et al., 2010; Tallmadge et al., 2013; Wagner et al., 2006), which is shown in Tables 1-10 and in Figs 1-3. The gene segments IGHV/IGLV/IGKV, IGHD, IGHJ/IGLJ/IGKJ, IGHC, IGLC, and IGKC (without superscript letters) are potentially functional variable gene segments. Superscript ORF was used to indicate variable gene segments with open reading frames that have either a defect in splicing sites, recombination signal sequence (RSS) and/or regulatory elements, and/or changes to the conserved amino acids, and therefore have been suggested to lead to incorrect folding (Lefranc, 1998). Superscript P indicates pseudo-variable gene segments. The genes were named according to the subgroup they belong to (Sun et al., 2010) and their number within this subgroup. The former 'VH1' was renamed 'IGHV1S1' to designate sequence 1 of subgroup 1. Based on >75% nucleotide identity 28 subgroups were established for the 40 IGHD genes and 2 subgroups were established for the 8 IGHJ genes. The classification of variable genes followed previous research. In Sun et al. (2010) and Tallmadge et al. (2013) sequences with at least 75% identity were grouped to the same family (Giudicelli and Lefranc, 1999). Their genes were named accordingly (Tallmadge et al., 2013). However, for future analyses we suggest to use 80% nucleotide identity as already recommended in 1984 for mouse immunoglobulin genes (Brodeur and Riblet, 1984). The IGHC

nomenclature also conforms to IMGT. Positions identified on several contigs are listed as well. The contigs are Un0011/NW_001876796, *Un0388/NW_001871527, **NW_001869767, and ***NW_001872990. The variable gene segments that were not classified into subgroups because they were too divergent or truncated are marked with n.c. (Sun et al., 2010).





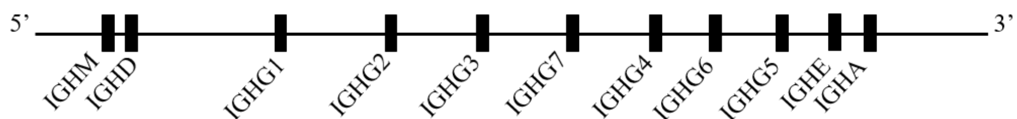


Fig. 1: Map of the equine heavy chain gene segments
(A) Within the scaffold Un0011 (NW_001876796), IGHV, IGHD and IGHJ gene clusters span a 510 kb region. (B) Scaffold Un0388 (NW_001871527) contains two potentially functional IGHVs (IGHV4S5 and IGHV2S4, according to the supposed designation), one IGHV ORF (IGHV1S4^{ORF}) and one IGHVΨ (IGHV4S16^{P}). (C) Map of the IGHC region of the horse, indicating the order of the eleven IGHC genes (adapted from Wagner et al., 2004). Boxes indicate IGHC genes. This map is adapted from Wagner et al. (2004) and Sun et al. (2010) to ensure that the newly proposed nomenclature (Tables 1-4) is designated to the respective positions on the locus. IGHV without superscript letters are potentially functional variable gene segments. Superscript ORF indicates variable gene segments with open reading frames that either have a defect in splicing sites, RSS and/or regulatory elements, and/or changes to the conserved amino acids, and therefore have been suggested to lead to incorrect folding (Lefranc, 1998). Superscript P indicates pseudo-variable gene segments.

The heavy chain contig Un0011 contains 50 variable gene segments (IGHV), 40 diversity gene segments (IGHD), as well as eight joining gene segments (IGHJ). Twelve out of 50 IGHV were described as functional, whereas 33 IGHV were defined as pseudogenes. Five open reading frames (ORF) were also described by Sun and coworkers. Two additional functional IGHV, as well as one ORF and one pseudogene were identified on Un0038 (Sun et al., 2010). All IGHV, IGHJ, and IGHC show the same transcriptional orientation (Fig. 1a, b). Most of the IGHV segments are flanked by 23 bp-spaced RSSs at their 3'ends, except IGHV7S1, IGHVxS1^{P}, IGHVxS2^{P}, IGHVxS7^{P}, IGHVxS8^{P}, IGHV3S7^{P}, and IGHV1S5^{P}, which either lack the nonamer or carry spacers shorter than 23 bp. The segments IGHV1S6 and IGHV4S17 were identified on the unplaced contigs NW_001869767 and NW_001872990, respectively (Table 1) (Tallmadge et al., 2013). IGHJ2S1 and IGHJ1S4 have 22 bp-spaced RSSs at their 5'ends. The remaining 6 IGHJs show 23 bp-spaced RSSs. All the IGHDs have 12 bp-spaced RSSs on both sides (Sun et al., 2010). With 40 IGHD identified, horses belong the mammalian species that possess the most IGHD. For instance, in guinea pig and the African elephant 41 and 87 IGHD gene segments were identified so far (Guo et al., 2011; Guo et al., 2012).
The equine immunoglobulin heavy chain constant region gene locus was localized on chromosome 24 (ECA24qter) and comprises 11 genes. All five isotypes known from humans were also identified in horses (Fig. 1c, Table 4). The entire equine IGHC region is located on a minimum of two overlapping clones from the CHORI-241 Horse Bacterial Artificial Chromosome library, suggesting that the size of the IGHC region is 250–350 kb (Wagner et al., 2004).

Table 1: Nomenclature of the equine IGHV gene segments

IGHV subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
1	VH1	IGHV1S1	IGHV1S1	349128-349423
1	VH6	IGHV1S2	IGHV1S2	211197-211492
1	VH7	IGHV1S3	IGHV1S3	175891-176186
1	VHORF6	IGHV1S4	IGHV1S4[ORF]	7876-8169*
1	pVH24	IGHV1S5	IGHV1S5[P]	126998-127283
1	Novel Vh[a]	IGHV1S6	IGHV1S6	38420-38913**
2	VH2	IGHV2S1	IGHV2S1	322988-323280
2	VH4	IGHV2S2	IGHV2S2	254382-254674
2	VH5	IGHV2S3	IGHV2S3	237777-238067
2	VH14	IGHV2S4	IGHV2S4	19736-20028*
3	VHORF1	IGHV3S1	IGHV3S1[ORF]	283617-283909
3	VHORF4	IGHV3S2	IGHV3S2[ORF]	132488-132804
3	VHORF5	IGHV3S3	IGHV3S3[ORF]	110030-110316
3	pVH8	IGHV3S4	IGHV3S4[P]	330696-330991
3	pVH9	IGHV3S5	IGHV3S5[P]	303932-304217
3	pVH14	IGHV3S6	IGHV3S6[P]	227389-227680
3	pVH17	IGHV3S7	IGHV3S7[P]	203930-204210
3	pVH21	IGHV3S8	IGHV3S8[P]	172321-172616
3	pVH23	IGHV3S9	IGHV3S9[P]	147144-147434
3	pVH27	IGHV3S10	IGHV3S10[P]	79642-79940
4	VH3	IGHV4S1	IGHV4S1	280198-280496
4	VH9	IGHV4S2	IGHV4S2	156258-156556
4	VH11	IGHV4S3	IGHV4S3	100632-100927
4	VH12	IGHV4S4	IGHV4S4	75777-76072
4	VH13	IGHV4S5	IGHV4S5	4967-5265*
4	VHORF2	IGHV4S6	IGHV4S6[ORF]	232063-232361
4	pVH4	IGHV4S7	IGHV4S7[P]	370921-371206
4	pVH5	IGHV4S8	IGHV4S8[P]	344891-345187
4	pVH12	IGHV4S9	IGHV4S9[P]	266901-267198
4	pVH13	IGHV4S10	IGHV4S10[P]	248659-248968
4	pVH15	IGHV4S11	IGHV4S11[P]	223950-224266
4	pVH22	IGHV4S12	IGHV4S12[P]	169462-169754
4	pVH26	IGHV4S13	IGHV4S13[P]	87064-87357
4	pVH28	IGHV4S14	IGHV4S14[P]	59569-59865
4	pVH29	IGHV4S15	IGHV4S15[P]	54540-54838
4	pVH34	IGHV4S16	IGHV4S16[P]	13567-13866*
4	Novel VH[b]	IGHV4S17	IGHV4S17	11760-12258***
5	VHORF3	IGHV5S1	IGHV5S1[ORF]	165239-165534
6	VH8	IGHV6S1	IGHV6S1	159437-159729

IGHV subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
7	VH10	IGHV7S1	IGHV7S1	141209-141509
n.c.	pVH1		IGHVxS1[P]	400324-400623
n.c.	pVH2		IGHVxS2[P]	385845-386119
n.c.	pVH3		IGHVxS3[P]	375741-376041
n.c.	pVH6		IGHVxS4[P]	339915-340211
n.c.	pVH7		IGHVxS5[P]	333740-334025
n.c.	pVH10		IGHVxS6[P]	301302-301579
n.c.	pVH11		IGHVxS7[P]	277258-277526
n.c.	pVH16		IGHVxS8[P]	217225-217491
n.c.	pVH18		IGHVxS9[P]	200550-200817
n.c.	pVH19		IGHVxS10[P]	188124-188419
n.c.	pVH20		IGHVxS11[P]	185173-185466
n.c.	pVH25		IGHVxS12[P]	89740-89991
n.c.	pVH30		IGHVxS13[P]	47308-47605
n.c.	pVH31		IGHVxS14[P]	35699-35979
n.c.	pVH32		IGHVxS15[P]	24565-24859
n.c.	pVH33		IGHVxS16[P]	19976-20184

Un0011 NW_001876796
*Un0388=NW_001871527
**NW_001869767
***NW_001872990

Table 2: Nomenclature of the equine IGHD gene segments

IGHD subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
1	DH1	IGHD1S1	IGHD1S1	415757-415787
2	DH2	IGHD2S1	IGHD2S1	417275-417293
3	DH3	IGDH3S1	IGDH3S1	417450-417497
4	DH4	IGHD4S1	IGHD4S1	419955-419986
5	DH5	IGHD5S1	IGHD5S1	422080-422110
5	DH10	IGHD5S2	IGHD5S2	430784-430814
5	DH15	IGHD5S3	IGHD5S3	441806-441836
5	DH21	IGHD5S4	IGHD5S4	453012-453042
5	DH25	IGHD5S5	IGHD5S5	459989-460019
5	DH28	IGHD5S6	IGHD5S6	466318-466348
6	DH6	IGHD6S1	IGHD6S1	423268-423286
7	DH7	IGHD7S1	IGHD7S1	423733-423761
7	DH17	IGHD7S2	IGHD7S2	445696-445724
8	DH8	IGHD8S1	IGHD8S1	426759-426777
8	DH16	IGHD8S2	IGHD8S2	443012-443030
8	DH19	IGHD8S3	IGHD8S3	447486-447504
9	DH9	IGHD9S1	IGHD9S1	428333-428363
10	DH11	IGHD10S1	IGHD10S1	433564-433589
11	DH12	IGHD11S1	IGHD11S1	435422-435455
12	DH13	IGHD12S1	IGHD12S1	436872-436890
12	DH37	IGHD12S2	IGHD12S2	486726-486744
13	DH14	IGHD13S1	IGHD13S1	439594-439631
14	DH18	IGHD14S1	IGHD14S1	446390-446423
15	DH20	IGHD15S1	IGHD15S1	450401-450433
15	DH27	IGHD15S2	IGHD15S2	463956-463987
16	DH22	IGHD16S1	IGHD16S1	454814-454843
17	DH23	IGHD17S1	IGHD17S1	456420-456444
17	DH26	IGHD17S2	IGHD17S2	461177-461201
18	DH24	IGHD18S1	IGHD18S1	457832-457861
19	DH29	IGHD19S1	IGHD19S1	467519-467540
20	DH30	IGHD20S1	IGHD20S1	471601-471618
20	DH40	IGHD20S2	IGHD20S2	490706-490721
21	DH31	IGHD21S1	IGHD21S1	475562-475586
22	DH32	IGHD22S1	IGHD22S1	477123-477153
23	DH33	IGHD23S1	IGHD23S1	478925-478957
24	DH34	IGHD24S1	IGHD24S1	480174-480192
25	DH35	IGHD25S1	IGHD25S1	483056-483084
26	DH36	IGHD26S1	IGHD26S1	485482-485512

IGHD subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
27	DH38	IGHD27S1	IGHD27S1	489459-489477
28	DH39	IGHD28S1	IGHD28S1	489958-489978

Un0011 NW_001876796

Table 3: Nomenclature of the equine IGHJ gene segments

IGHJ subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
1	JH1	IGHJ1S1	IGHJ1S1	504144-504197
1	JH3	IGHJ1S2	IGHJ1S2	504660-504713
1	JH4	IGHJ1S3	IGHJ1S3	505003-505050
1	JH5	IGHJ1S4	IGHJ1S4	505280-505330
1	JH6	IGHJ1S5	IGHJ1S5	505855-505908
1	JH7	IGHJ1S6	IGHJ1S6	506275-506331
1	JH8	IGHJ1S7	IGHJ1S7	506729-506782
2	JH2	IGHJ2S1	IGHJ2S1	504350-504402

Un0011 NW_001876796

Table 4: Nomenclature of the equine IGHC genes

Original nomen-clature	Current desig-nation	Gene	Allel	Genome Position on NW_001876796
IgM	IgM	IGHM	M_02	519420-524201
				514993-518108
			M_01	516522-517025
				515374-515717
IgD	IgD	IGHD		531198-534510
				526129-527334
IgGa	IgG1	IGHG1		576506-576800
				577273-577607
				577685-578019
			G1_01	576506-578019
	IgG2	IGHG2		608720-610787
				625782-627849
IgG(T)	IgG3	IGHG3	G3_01	651171-651465
				651958-652292
				652361-652695
			G3_02	650866-653171
	IgG7	IGHG7	G7_01	674467-676688
			G7_02	674945-676422
IgGb	IgG4	IGHG4	G4_02	674954-676422
				651171-651701
			G4_01	674827-676691
IgGc	IgG6	IGHG6		698724-700204
				651171-651723
IgG(T)	IgG5	IGHG5		730626-732347
IgGB				
IgE	IgE	IGHE	E_01	754864-756464
			E_02	756129-756464
				755315-755639
				755732-756058
				754864-755155
			E_03	756129-756585
				755732-756058
			E_04	756129-756551
				755732-756058
				755315-755639
				754858-755155
IgA	IgA	IGHA		766888-768383
				768662-769379

3.1. IgM

IgM is the first isotype that appears phylogenetically, as well as ontogenetically. It is the first immunoglobulin produced by fetuses. This isotype is responsible for the primary immune response. IgM may be bound to the membrane as a monomer or secreted as a pentameric molecule with high avidity. IgM, as well as IgG, is present in the plasma (Tallmadge et al., 2009).

The equine IGHM locus spans 1472 bp coding for 451 amino acid residues (Fig. 1c, Table 4). This isotype exhibits six potential N-glycosylation sites. Similar to IgE, no separated hinge region was identified. To enable flexibility of this istopye, the second constant region may act as hinge. The third and fourth constant regions of IgM are conserved even between different species, which indicates a specific role of IgM in complement activation (Schrenzel et al., 1997).

3.2. IgD

Although the importance of this isotype remains unclear and a receptor-like function is supposed, constant region genes coding for IgD are not restricted to primates and rodents. After successful description of the IgD isotype in cattle, sheep and swine, IgD was also identified in horses (Wagner et al., 2004; Zhao et al., 2002). The equine IgD locus was identified 5 kb downstream of the genes coding for IgM (Fig. 1c, Table 4). The entire genomic locus spans about 9.1 kb and contains eight exons that are: CH1, two exons coding for the hinge region, CH2, CH3, one exon for the secreted molecule, as well as two exons for the transmembrane anchor; the first of which contains a specific splicing site. For IgD, the secreted form consists of 391 amino acid residues and the transmembrane form of 418 amino acid residues. The coding nucleotide sequences of CH1 to CH3 exons show sequence identities of 64% to human and porcine IgD genes and 61% to bovine and ovine IgD genes. In analogy to the human IgD, there is no switch-region between the genes coding for IgM and IgD, indicating that there is an alternative splice mechanism (Wagner et al., 2004). In contrast, a c-switch region has only been identified for cattle and pig (Zhao et al., 2002; Zhao et al., 2003).

3.3. IgG

The IgG isotype is considered to be the most versatile immunoglobulin in all mammals and birds. Only in the porcine immune system has a similarly high number of eleven genomic constant region genes representing six putative subclasses been described (Fig. 1c, Table 4) (Butler et al., 2009). IgG performs all functions of a typical immunoglobulin molecule such as antigen binding, complement fixation, and binding to various cells like phagocytic cells, lymphocytes, platelets, mast cells, and basophils. It contributes about 75% of the antibodies. The isotype was defined at a biochemical and serological level and the different genomic elements were characterized. The subclasses IgGa, IgGb, IgGc, IgG(T), and IgG(B) were classified (Klinman et al., 1965; Rockey, 1967; Rockey et al., 1964; Sandor et al., 1964b; Sheoran and Holmes, 1996; Widders et al., 1986) according to their antigenic differences and their serological and electrophoretic properties. Restriction enzyme analyses of the genomic DNA of horse lambda phage clones, deletion analyses of heterohybridomas, and nucleotide sequencing of individual IGHC genes revealed six constant region genes in the equine haploid genome (Overesch et al., 1998; Wagner et al., 2006). The characterization of the complete IGHC region and the analysis of 34 overlapping genomic clones from the CHORI-241 BAC library revealed seven constant region genes of equine IgG (Table 4) and constituted the current repertoire of IgG IGHCs based on their sequences, phylogeny, and homologous structures (Wagner et al., 2002; Wagner et al., 2004). Multiple duplications, gene conversions, and crossovers are supposed to explain these seven IgG isotypes, which are unique in all species investigated. IgG4 and IgG7 seem to be the latest duplication events in evolution as their sequences show 96% identity. As available reagents are insufficient for differentiation, isotypes IgG2 and IgG7 could not yet be diversified in serum and other body fluids (Wagner et al., 2006). The individual subclasses possess specific roles in protective immunity. Thus, the FcγR and complement-binding isotypes IgG1, IgG4, and IgG7 (former IgGa and IgGb) contribute to the protection against several equine pathogens such as equine influenza virus (Breathnach et al., 2006; Nelson et al., 1998), *Streptococcus equi* (Sheoran et al., 1997) and *Rhodococcus equi* (Lopez et al., 2002). In addition, systemic and mucosal IgG responses were described to play an important role in limiting the spread and severity of equine herpes virus 1 (Kydd et al., 2006). Although IgG4 and IgG7 represent the dominating serum antibodies, these isotypes are rarely found in clinically healthy horses (Keggan et al., 2013).

All these seven IgG subclasses were expressed *in vivo* (Wagner et al., 2004). Based on these insights, the first recombinant versions of all seven equine IgG subclasses (mouse λ-light chains, horse IGHG1-7) were expressed in Chinese Hamster Ovary cells to analyze their individual physical and biological properties (Lewis et al., 2008). In contrast to these first recombinant equine immunoglobulins, which were expressed for functional analyses, seven distinct monoclonal equine antibodies (IgM, IgG1, IgG3, IgG4/7, IgG5, IgG6, and IgE) were produced in equine-murine heterohybridomas to be used for quantification of isotypes in diagnostic testing and immunological research (Keggan et al., 2013).

All recombinant IgGs were N-glycosylated and maintained the affinity for their antigen. Heavy and light chains assembled in intact IgG were stabilized by interchain disulfide bridges except in a small proportion of IgG4 and 7. Complete IgG molecules possessed molecular weights of approximately 150 kDa. Nevertheless, differences in complement activation, Fc-receptor binding and the bacterial protein binding capacity between the IgG isotypes were observed. IgG3 showed O-glycosylation in addition to N-glycans. As its hinge region is rich in Proline at position -1 and +3 relative to Threonine residues, which are known sites for O-glycan attachment, O-glycans may be attached to the hinge region of the equine IgG3 (Lewis et al., 2008; Wilson et al., 1991). In IgG4, IgG6, and IgG7 the Cystein equivalent to the human Cys131 was missing within the CH1 region. Disulfide bond formation – with a Cystein at position 220 in the upper hinge region – is assumed to support the formation of intact IgG. Amino acid substitutions at key positions (234-239, 251-154, 270, 311, 314, 318, 320, 322, 329, 331, 380, 382, 428, 432-436, 438) of the IgG molecule, which is responsible for binding to protein A and G, were described (Lewis et al., 2008). IgG1, IgG3, IgG4, IgG5, and IgG7 elicited strong respiratory burst activity in equine peripheral blood leukocytes (PBL), indicating an efficient interaction with FcγR. In contrast, IgG2 and IgG6 induced little or no response. Apparently, they are unable to interact with FcγR. An amino acid residue motif at the N-terminus of the CH2 domain (Leu234-Leu235-Gly236-Gly237-Pro238-Ser239) seemed to be involved in Fc-receptor binding. In complement activation, IgG1, IgG3, IgG4, and IgG7 showed positive reactions, whereas IgG2, IgG5, and IgG6 failed. Responsible C1q binding sites appeared not to be conserved between different mammalian species, but there was a Lysine corresponding to human and mouse Lysine at position 322, which was found in all complement activating equine IgG, as well as in IgG6, but Lys322 was replaced by Serine in IgG2 and IgG5 (Keggan et al., 2013; Lewis et al., 2008).

Evidently, IgG plays an important role both in serum and mucosal compartments in horses. Therefore, the summarized considerations appear to be relevant to systemic and mucosal vaccination strategies. IgG antibodies of the subclasses IgG1, IgG3, IgG4/IgG7 should be induced by vaccines to achieve maximal protection via FcγR and complement-mediated elimination mechanisms. Less effective protection is provided by vaccines that trigger only IgG2, IgG5 or IgG6 antibodies (Lewis et al., 2008).

3.4. IgE

IgE is the least common serum immunoglobulin isotype (Oethinger et al., 1997). Besides the immune response against parasites, IgE is also known to be responsible for type 1 allergies and resulting diseases like summer eczema, urticaria, COPD (recurrent airway obstruction) or head shaking (Wagner et al., 2006). For primary investigations, equine IgE was prepared from the serum of foals suffering from endoparasitic infections (Suter and Fey, 1981). Compared to humans, horses revealed a high concentration of IgE in serum (Wagner et al., 2006). The constant region gene is located between the constant region genes coding for IgG5 and IgA, and it contains four exons. Similar to IgM, there is no separated hinge region for IgE and the second constant region enables flexibility. The switch region is located 2-4 kb upstream of the constant region gene. Three haplotypes are described differing in two SNPs, which result in two amino acid residue substitutions (Fig. 1c, Table 4) (Navarro et al., 1995; Wagner et al., 2006). The polymorphic character of the IgE genes may influence effector functions, binding mast cells and basophils, and finally the level of allergic reactions.

3.5. IgA

IgA dominates in the mucosal immune response of respiratory, gastrointestinal and genitourinary tracts (Souza et al., 2010). In body fluids such as milk, nasal secretions and saliva, it appears as a dimeric secreted molecule and is called secretory IgA. In serum it shows itself in monomeric form (Sheoran et al., 1997; Wagner et al., 2003). IgA produced by B1 lymphocytes represents a T-lymphocyte independent source of IgA, which recognizes commensal bacteria. IgA produced by B2 lymphocytes in the

mucosal-associated lymphoid tissues represents T-dependent secreted IgA, e.g. directed against exotoxins (Macpherson et al., 2000).

The IgA locus possesses three exons, two introns, and one hinge region. This hinge region is an extension of the CH2 sequence. Within the equine IGHC region, IGHA is the most 3'located gene (Fig. 1c, Table 4). The switch region is located 1.4 kb upstream of the constant region genes coding for IgA. Phylogenetic comparisons show a common cluster with the IgA genes of humans, dogs, swine, cattle, and sheep. In contrast to rabbits possessing 13 subclasses of IgA and humans with two subclasses of IgA, horses have only one IgA class (Burnett et al., 1989; Mestecky et al., 2004; Wagner et al., 2003).

Investigations concerning the concentration of IgA in serum showed significantly lower IgA quantities in sport horses when compared to extensively reared horses (Souza et al., 2010). Comparable amounts of IgA were found in the colostrum of Warmblood and Standardbred mares after parturition (Kohn et al., 1989). In milk, the reduced IgA concentration is likely due to passively transferred IgA, which was observed in a similar way for IgG (Jenvey et al., 2012). Once the window for absorbing maternal IgA has passed, IgA levels in the mare's milk persisted but the significance of this observation is yet to be fully established. However, investigations in human neonates demonstrated that IgA and other factors within breast milk may provide an ongoing contribution to local passive protection and immunological development (Corthesy, 2007). Equine serum and secretory IgA was quantified in Shetland ponies 182 days after parturition. While IgG levels decreased with time after parturition, IgA levels increased until IgA was the predominant immunoglobulin (McGuire and Crawford, 1972). In contrast, the quantity of IgA in colostrum and milk from Thoroughbred mares rapidly decreased during the first 19 days after parturition, and fluctuated at this level until day 60 p.p. (Jenvey et al., 2012).

4. Equine immunoglobulin light chain gene loci

As known from other mammals, horses also express two isotypes of light chains (λ and κ) differentiated by isotype-specific amino acid residue motifs, antigenic properties, and chemical structure. In contrast to other mammals, 95% of the equine serum antibodies combine λ-light chains with the heavy chains, although the reason for this phenomenon is not yet clear (Gibson, 1974; Home et al., 1992). The assumption that amino acid residue substitutions in IGKC result in structures with reduced binding efficiency to the heavy chains was not verified (Ford et al., 1994; Sun et al., 2010). Therefore, a correlation between the number of IGLV gene segments and the preferred usage of a particular isotype is supposed. The λ-light chain locus was mapped on equine autosome 8 (ECA8) and the gene segments of the κ-light chain were located on ECA15 (Das et al., 2008).

First studies on the organization of the equine light chain loci revealed a restricted number of variable gene segments for both isotypes. The assembly of variable, joining, and constant region genes was also found to be similar to other species (Ford et al., 1994; Home et al., 1992). Today 144 IGLV, 7 IGLJ and 7 IGLC are identified within 1310 kb on ECA8 (Fig. 2, Tables 5-7). Each of the 7 IGLJ is preceded by one IGLC and there are two clusters of IGLV possessing different transcriptional orientations. Within each IGLV cluster there are functional genes and pseudogenes (Sun et al., 2010). Similarly, pseudogenes are described in different species such as chicken, where it is already known that pseudogenes are used for gene conversion (Reynaud et al., 1985).

As there were differences in the nomenclature, we also suggest a supplemented designation for the light chain gene segments (Tables 5-7). The name of Ig lambda light chain variable (IGLV) and joining (IGLJ) gene segments were assigned according to IMGT nomenclature guidelines; the international ImMunoGeneTics information system (www.imgt.org). The genes were named according to the subgroup they belong to (Sun et al., 2010) and their number within this subgroup. Furthermore, to denote open reading frames or pseudogenes, sequence names were complemented by superscript ORF and P, respectively. The former 'Vλ1' was renamed 'IGLV1S1' to designate sequence 1 of subgroup 1. In analogy, IGLJ and IGLC genes were designated IGLJ1S1 through IGLJ7S1 and IGLC1S1 through IGLC7S1, instead of "Jλ1" and "Cλ1", which is consistent with the human nomenclature of Ig lambda genes

(Hara et al., 2012; Lefranc, 2001a; Sun et al., 2010; Tallmadge et al., 2014). Positions identified on several contigs are also listed. The contigs are NW_001867428.1 and chromosome 8 (Sun et al., 2010). Superscript P indicates pseudo-variable gene segments. The variable gene segments that were not classified to subgroups because they were too divergent or truncated were marked with n.c. (Sun et al., 2010).

According to the supposed nomenclature, which is in analogy to the heavy and λ-light chain, IGLJ5S1, IGLJ6S1, and IGLJ7S1 show identical sequences. IGLC2S1^{P} is defined as pseudogene due to premature stop codons, as well as IGLJ2S1^{P}, IGLJ3S1^{P}, and IGLJ4S1$^{(P)}$ (Sun et al., 2010). Among the 144 IGLV segments identified by Sun et al. (2010), 108 have RSS heptamers and nonamers, 9 lack conserved nonamers and the remaining segments lack both heptamers and nonamers. IGLV3S1, IGLV9S1, IGLV9S2, and IGLV9S3^{ORF} have 22-spaced RSS and the remaining potentially functional and ORF-IGLVs carry 23-spaced RSS. The spacer length is more varied among the 76 IGLV^{P} segments, which have the intact RSSs. As many as 55 IGLVP segments maintain 23 bp-spaced RSSs, whereas 16 IGLVP segments carry 22 bp-spaced RSSs. The IGLVxS36^{P}, IGLV9S4^{P} and IGLVxS59^{P} have spacer lengths of 21, 21 and 19 bp, respectively. All 7 IGLJ segments have 12 bp-spaced RSSs at their 5’end.

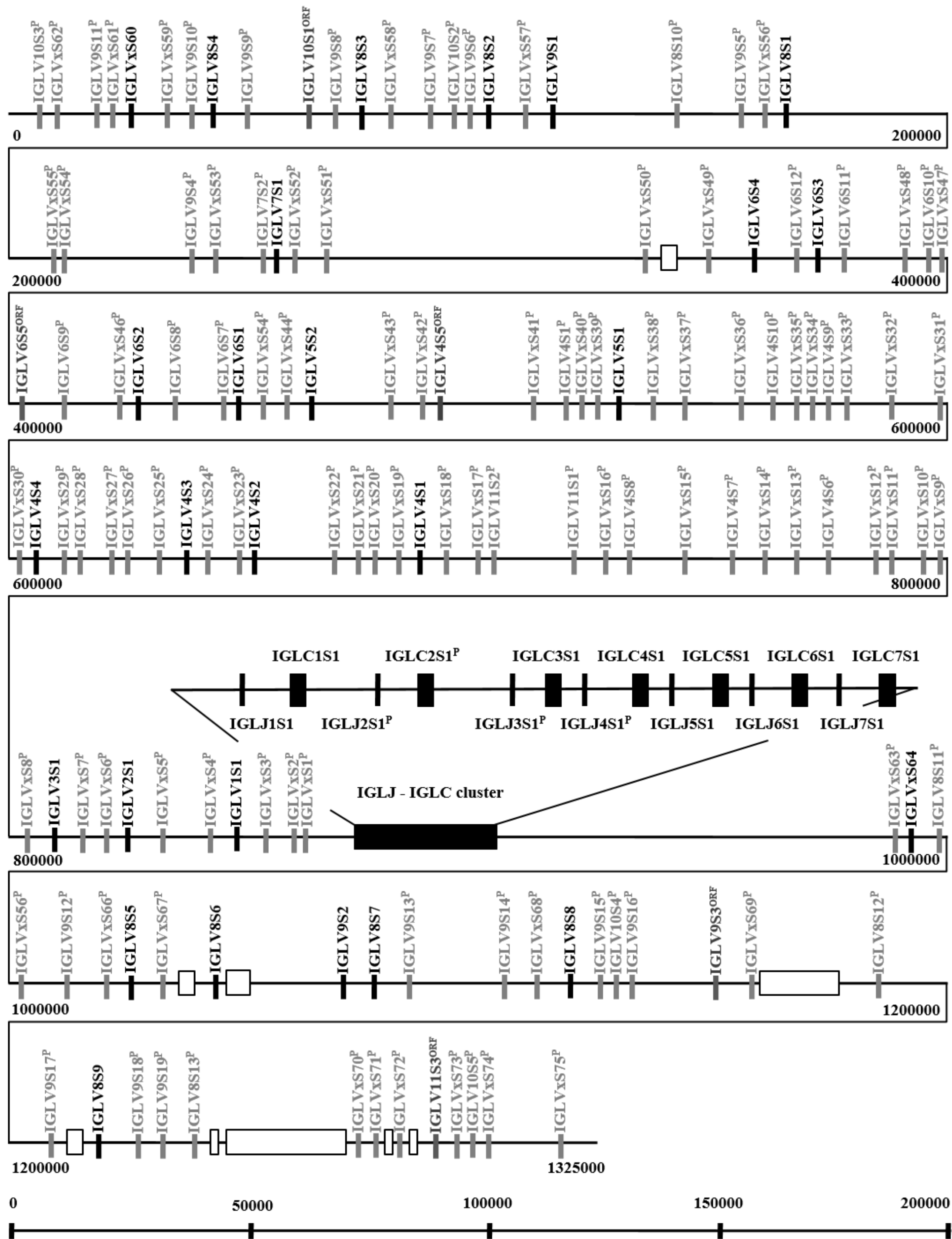


Fig. 2: Map of the equine lambda light chain gene locus
The lambda light chain genes are located on chromosome 8 spanning about 1310 kb. The IGLJ-IGLC pairs and the IGLVs upstream of the (IGLJ-IGLC)7 show opposite transcriptional orientation to the IGLVs downstream of the (IGLJ-IGLC)7 cluster. This physical map is adapted from Sun et al. (2010) to ensure that the newly proposed nomenclature (table 4-7) is designated to the respective positions on the locus. IGLV without superscript letters are potentially functional variable gene segments. Superscript ORF indicates variable gene segments with open reading frames that either have a defect in splicing sites, RSS and/or regulatory elements, and/or changes to the conserved amino acids, and therefore have been suggested to lead to incorrect folding (Lefranc, 1998). Superscript P indicates pseudo-variable gene segments.

Table 5: Nomenclature of the equine IGLV gene segments

IGLV subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
1	Vλ1	IGLV1-38	IGLV1S1	4171705-4171992
2	Vλ2	IGLV2-41	IGLV2S1	4195092-4195381
3	Vλ3	IGLV3-44	IGLV3S1	4210612-4210896
4	Vλ4	IGLV4-61	IGLV4S1	4330800-4331090
4	Vλ5	IGLV4-66	IGLV4S2	4366543-4366833
4	Vλ6	IGLV4-69	IGLV4S3	4381050-4381340
4	Vλ7	IGLV4-75	IGLV4S4	4411343-4411633
4	VλORF1	IGLV4-92	IGLV4S5^{ORF}	4529112-4529396
4	pVλ13	IGLV4-50	IGLV4S6^{P}	4245996-4246283
4	pVλ16	IGLV4-53	IGLV4S7^{P}	4267110-4267396
4	pVλ18	IGLV4-55	IGLV4S8^{P}	4287903-4288193
4	pVλ39	IGLV4-80	IGLV4S9^{P}	4447087-4447374
4	pVλ42	IGLV4-83	IGLV4S10^{P}	4458683-4458970
4	pVλ48	IGLV4-90	IGLV4S11^{P}	4502561-4502854
5	Vλ8	IGLV5-87	IGLV5S1	4489461-4489759
5	Vλ9	IGLV5-95	IGLV5S2	4554617-4554901
6	Vλ10	IHLV6-98	IHLV6S1	4571250-4571546
6	Vλ11	IGLV6-101	IGLV6S2	4590316-4590612
6	Vλ12	IGLV6-109	IGLV6S3	4648068-4648364
6	Vλ13	IGLV6-111	IGLV6S4	4660775-4661062
6	VλORF2	IGLV6-104	IGLV6S5^{ORF}	4614091-4614386
6	pVλ54	IGLV6-99	IGLV6S7^{P}	4574617-4574913
6	pVλ55	IGLV6-100	IGLV6S8^{P}	4584807-4585103
6	pVλ57	IGLV6-103	IGLV6S9^{P}	4607689-4607985
6	pVλ59	IGLV6-106	IGLV6S10^{P}	4624835-4625128
6	pVλ61	IGLV6-108	IGLV6S11^{P}	4643773-4644068
6	pVλ62	IGLV6-110	IGLV6S12^{P}	4654126-4654437
7	Vλ14	IGLV7-116	IGLV7S1	4763155-4763451
7	pVλ67	IGLV7-117	IGLV7S2^{P}	4766554-4766835
8	Vλ15	IGLV8-122	IGLV8S1	4855083-4855381
8	Vλ17	IGLV8-128	IGLV8S2	4916905-4917203
8	Vλ18	IGLV8-133	IGLV8S3	4944794-4945086
8	Vλ19	IGLV8-137	IGLV8S4	4975350-4975639
8	Vλ22	IGLV8-28	IGLV8S5	3994214-3994503
8	Vλ23	IGLV8-26	IGLV8S6	3976833-3977125
8	Vλ25	IGLV8-24	IGLV8S7	3942211-3942500
8	Vλ26	IGLV8-20	IGLV8S8	3902029-3902327
8	Vλ27	IGLV8-12	IGLV8S9	3799404-3799702
8	pVλ74	IGLV8-125	IGLV8S10^{P}	4878404-4878692

IGLV subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
8	pVλ89	IGLV8-32	IGLV8S11^{P}	4023156-4023449
8	pVλ101	IGLV8-14	IGLV8S12^{P}	3838136-3838430
8	pVλ105	IGLV8-9	IGLV8S13^{P}	3781419-3781717
9	Vλ16	IGLV9-126	IGLV9S1	4904149-4904473
9	Vλ24	IGLV9-25	IGLV9S2	3948638-3948962
9	VλORF4	IGLV9-16	IGLV9S3^{ORF}	3870582-3870906
9	pVλ69	IGLV9.119	IGLV9S4^{P}	4780685-4780996
9	pVλ73	IGLV9-124	IGLV9S5^{P}	4866218-4866534
9	pVλ76	IGLV9-129	IGLV9S6^{P}	4922628-4922938
9	pVλ78	IGLV9-131	IGLV9S7^{P}	4929757-4930046
9	pVλ80	IGLV9-134	IGLV9S8^{P}	4950548-4950857
9	pVλ81	IGLV9-136	IGLV9S9^{P}	4968911-4969235
9	pVλ82	IGLV9-138	IGLV9S10^{P}	4981125-4981435
9	pVλ85	IGLV9-142	IGLV9S11^{P}	4999541-4999848
9	pVλ91	IGLV9-30	IGLV9S12^{P}	4006944-4007268
9	pVλ94	IGLV9-23	IGLV9S13^{P}	3936435-3936744
9	pVλ95	IGLV9-22	IGLV9S14^{P}	3914834-3915157
9	pVλ97	IGLV9-19	IGLV9S15^{P}	4780685-4780996
9	pVλ99	IGLV9-17	IGLV9S16^{P}	4866218-4866534
9	pVλ102	IGLV9-13	IGLV9S17^{P}	3812261-3812584
9	pVλ103	IGLV9-11	IGLV9S18^{P}	3793651-3793937
9	pVλ104	IGLV9-10	IGLV9S19^{P}	3788296-3788612
10	VλORF3	IGLV10-135	IGLV10S1^{ORF}	4954678-4954976
10	pVλ77	IGLV10-130	IGLV10S2^{P}	4926370-4926670
10	pVλ87	IGLV10-144	IGLV10S3^{P}	5012699-5012989
10	pVλ98	IGLV10-18	IGLV10S4^{P}	3892471-3892771
10	pVλ110	IGLV10-3	IGLV10S5^{P}	3723141-3723424
11	pVλ20	IGLV11-140	IGLV11S1^{P}	4300066-4300348
11	pVλ21	IGLV11-33	IGLV11S2^{P}	4316629-4316892
11	VλORF5	IGLV11-5	IGLV11S3^{ORF}	3730193-3730489
n.c.	pVλ1		IGLVxS1^{P}	4157587-4157847
n.c.	pVλ2		IGLVxS2^{P}	4160537-4160835
n.c.	pVλ3		IGLVxS3^{P}	4166903-4167176
n.c.	pVλ4		IGLVxS4^{P}	4178429-4178741
n.c.	pVλ5		IGLVxS5^{P}	4188046-4188311
n.c.	pVλ6		IGLVxS6^{P}	4199645-4199909
n.c.	pVλ7		IGLVxS7^{P}	4205575-4205866
n.c.	pVλ8		IGLVxS8^{P}	4216947-4217233

IGLV subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
n.c.	pVλ9		IGLVxS9^{P}	4221861-4222159
n.c.	pVλ10		IGLVxS10^{P}	4226486-4226798
n.c.	pVλ11		IGLVxS11^{P}	4232825-4233111
n.c.	pVλ12		IGLVxS12^{P}	4236991-4237275
n.c.	pVλ14		IGLVxS13^{P}	4252560-4252850
n.c.	pVλ15		IGLVxS14^{P}	4259340-4259638
n.c.	pVλ17		IGLVxS15^{P}	4276218-4276516
n.c.	pVλ19		IGLVxS16^{P}	4293600-4293868
n.c.	pVλ22		IGLVxS17^{P}	4320273-4320548
n.c.	pVλ23		IGLVxS18^{P}	4327257-4327515
n.c.	pVλ24		IGLVxS19^{P}	4336476-4336776
n.c.	pVλ25		IGLVxS20^{P}	4341721-4341986
n.c.	pVλ26		IGLVxS21^{P}	4345368-4345651
n.c.	pVλ27		IGLVxS22^{P}	4350798-4351060
n.c.	pVλ28		IGLVxS23^{P}	4370580-4370835
n.c.	pVλ29		IGLVxS24^{P}	4377206-4377464
n.c.	pVλ30		IGLVxS25^{P}	4387212-4387524
n.c.	pVλ31		IGLVxS26^{P}	4393759-4394024
n.c.	pVλ32		IGLVxS27^{P}	4397408-4397690
n.c.	pVλ33		IGLVxS28^{P}	4403625-4403886
n.c.	pVλ34		IGLVxS29^{P}	4407546-4407804
n.c.	pVλ35		IGLVxS30^{P}	4417549-4417861
n.c.	pVλ36		IGLVxS31^{P}	4423822-4424107
n.c.	pVλ37		IGLVxS32^{P}	4433296-4433588
n.c.	pVλ38		IGLVxS33^{P}	4444562-4444847
n.c.	pVλ40		IGLVxS34^{P}	4450690-4451002
n.c.	pVλ41		IGLVxS35^{P}	4454114-4454375
n.c.	pVλ43		IGLVxS36^{P}	4464714-4465026
n.c.	pVλ44		IGLVxS37^{P}	4476525-4476807
n.c.	pVλ45		IGLVxS38^{P}	4483271-4483580
n.c.	pVλ46		IGLVxS39^{P}	4495317-4495629
n.c.	pVλ47		IGLVxS40^{P}	4498517-4498783
n.c.	pVλ49		IGLVxS41^{P}	4508307-4508617
n.c.	pVλ50		IGLVxS42^{P}	4532982-4533235
n.c.	pVλ51		IGLVxS43^{P}	4539033-4539326
n.c.	pVλ52		IGLVxS44^{P}	4560252-4560551
n.c.	pVλ53		IGLVxS45^{P}	
n.c.	pVλ56		IGLVxS46^{P}	4595924-4596217

IGLV subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
n.c.	pVλ58		$IGLVxS47^P$	4621790-4622080
n.c.	pVλ60		$IGLVxS48^P$	4629113-4629400
n.c.	pVλ63		$IGLVxS49^P$	4670998-4671281
n.c.	pVλ64		$IGLVxS50^P$	4684017-4684281
n.c.	pVλ65		$IGLVxS51^P$	4753071-4753353
n.c.	pVλ66		$IGLVxS52^P$	4760021-4760299
n.c.	pVλ68		$IGLVxS53^P$	4775528-4775824
n.c.	pVλ70		$IGLVxS54^P$	4807475-4807769
n.c.	pVλ71		$IGLVxS55^P$	4808989-4809266
n.c.	pVλ72		$IGLVxS56^P$	4860864-4861172
n.c.	pVλ75		$IGLVxS57^P$	4910505-4910786
n.c.	pVλ79		$IGLVxS58^P$	4938347-4938661
n.c.	pVλ83		$IGLVxS59^P$	4985429-4985745
n.c.	Vλ20		$IGLVxS60^P$	4994608-4994903
n.c.	pVλ84		$IGLVxS61^P$	4998412-4998621
n.c.	pVλ86		$IGLVxS62^P$	5010496-5010800
n.c.	pVλ88		$IGLVxS63^P$	4033346-4033647
n.c.	Vλ21		$IGLVxS64^P$	4031616-4031911
n.c.	pVλ90		$IGLVxS65^P$	4017794-4018083
n.c.	pVλ92		$IGLVxS66^P$	4000617-4000857
n.c.	pVλ93		$IGLVxS67^P$	3988326-3988631
n.c.	pVλ96		$IGLVxS68^P$	3908433-3908722
n.c.	pVλ100		$IGLVxS69^P$	3864249-3864531
n.c.	pVλ106		$IGLVxS70^P$	3745800-3746092
n.c.	pVλ107		$IGLVxS71^P$	3742171-3742365
n.c.	pVλ108		$IGLVxS72^P$	3738392-3738672
n.c.	pVλ109		$IGLVxS73^P$	3726480-3726695
n.c.	pVλ111		$IGLVxS74^P$	3719722-3720011
n.c.	pVλ112		$IGLVxS75^P$	3703267-3703501

IGLV: Sun et al., 2010, Chromosom 8, NW_001867428.1

Table 6: Nomenclature of the equine IGLJ gene segments

IGLJ subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
1	Jλ1	IGLJ1	IGLJ1S1	4146337-4146302 [a]
2	Jλ2	IGLJ2	IGLJ2S1[P]	4141033-4140996 [a]
3	Jλ3	IGLJ3	IGLJ3S1[P]	4135576-4135541 [a]
4	Jλ4	IGLJ4	IGLJ4S1[(P)]	4132448-4132410 [a]
5	Jλ5	IGLJ5	IGLJ5S1	4129086-4129050 [a]
6	Jλ6	IGLJ6	IGLJ6S1	4125743-4125707 [a]
7	Jλ7	IGLJ7	IGLJ7S1	4122272-4122236 [a]

[a] Contig_ NW_001867428.1

Table 7: Nomenclature of the equine IGLC genes

IGLC subgroup	Designation Sun et al. 2010	Designation Tallmadge et al. 2013	Proposed designation	Genome Position
1	Cλ1	IGLC1	IGLC1S1	4144802-4144483 [a]
2	pCλ2	IGLC2	IGLC2S1[P]	4139692-4139371 [a]
3	Cλ3	IGLC3	IGLC3S1	4134292-4133973 [a]
4	Cλ4	IGLC4	IGLC4S1	4131021-4130702 [a]
5	Cλ5	IGLC5	IGLC5S1	4127666-4127347 [a]
6	Cλ6	IGLC6	IGLC6S1	4124320-4124002 [a]
7	Cλ7	IGLC7	IGLC7S1	4120851-4120531 [a]

[a] Contig_ NW_001867428.1

The λ-light chain locus shows a unique feature of two IGLV clusters compared to this locus in all other species analyzed, whereas the κ-light chain genes are arranged as usual in one cluster. In the 820 kb locus, there is one cluster of 60 IGKV, one cluster of five IGKJ, and a single IGKC (Fig. 3, Tables 8-10) (Sun et al., 2010). Moreover, the functional and nonfunctional IGKV are located within one locus. All IGKV segments have 12 bp-spaced RSSs at their 3'ends except IGKV1S21[P], which possesses an 11 bp spacer and additional 8 IGKV[P] segments where RSSs are lacking completely. IGKJ1S1, IGKJ2S1, and IGKJ4S1 have 23 bp-spaced RSSs at their 5'ends in contrast to IGKJ3S1, which has a 20 bp-spaced RSS. IGKJ5S1[P] is nonfunctional as there is no conserved heptamer within the RSS (Sun et al., 2010). Transcriptional orientations in both directions are found without grouping. The single IGKC is separated from the IGKJ by a 2926 bp intron. IGKJ5S1[P] has no heptamer in the recombination signal sequence (RSS) and is therefore described as pseudogene. In IGKJ3S1, the spacer of the RSS shows 20 bp instead of 23 bp, which does not result in reduced recombination frequency (Ford et al., 1994; Sun et al., 2010).

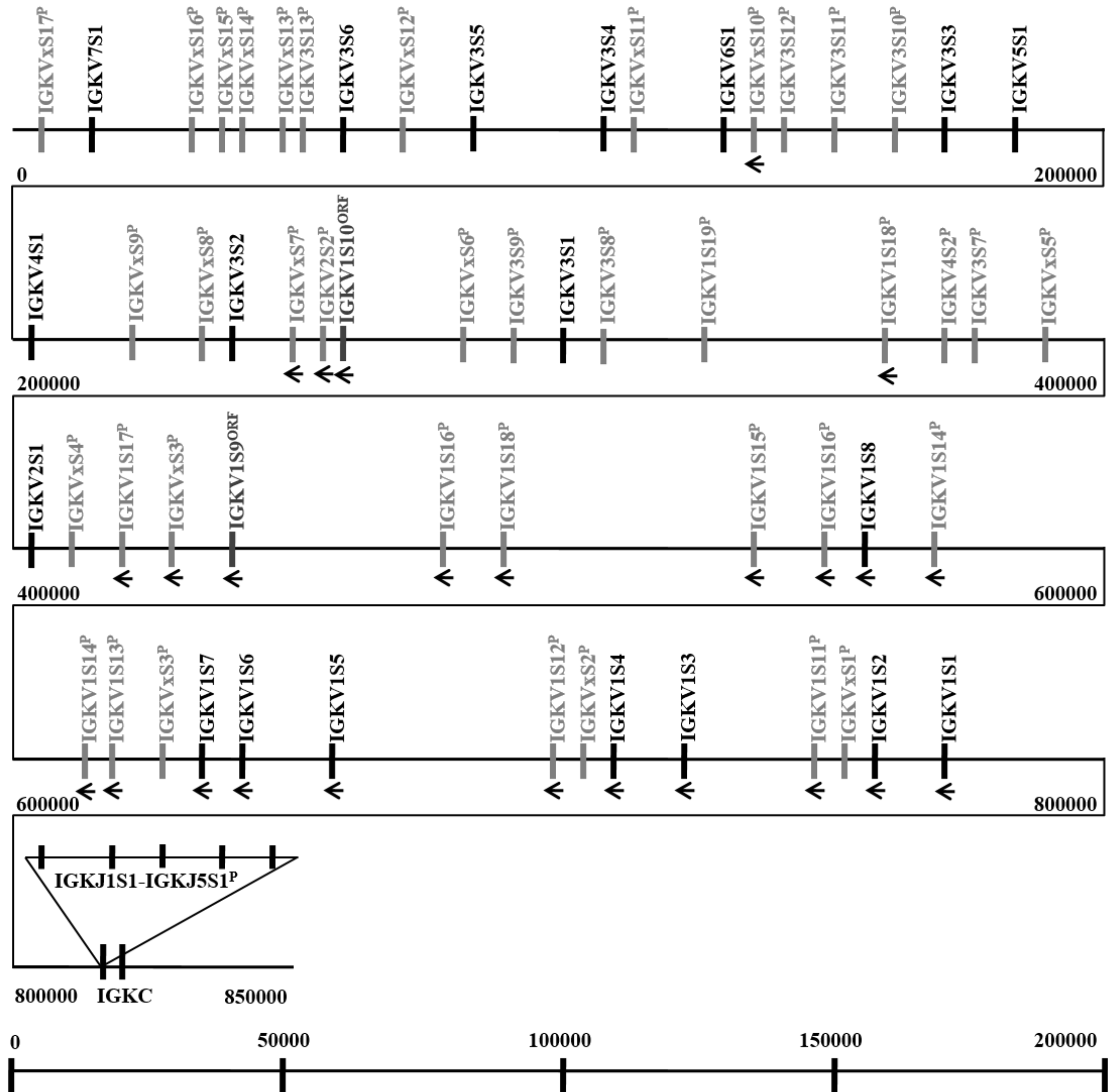


Fig. 3: Physical map of the equine kappa light chain gene locus
820 kb of the kappa light chain genes are located on chromosome 15. The potentially functional IGKVs, 2 ORFs and 15 pseudogenes show opposite transcriptional orientation to the (IGKJ)5-IGKC cluster. Arrows below the vertical lines directed 5' indicate transcriptional orientation. This physical map is adapted from Sun et al. (2010) to ensure that the newly proposed nomenclature (table 3) is designated to the respective positions on the locus. IGKV without superscript letters are potentially functional variable gene segments. Superscript ORF indicates variable gene segments with open reading frames that either have a defect in splicing sites, RSS and/or regulatory elements, and/or changes to the conserved amino acids, and therefore have been suggested to lead to incorrect folding (Lefranc, 1998). Superscript P indicates pseudo-variable gene segments.

Table 8: Nomenclature of the equine IGKV gene segments

IGKV subgroup	Designation Sun et al. 2010	Proposed designation	Genome Position
1	Vκ1	IGKV1S1	17071497-17071783
1	Vκ2	IgKV1S2	17057542-17057828
1	Vκ3	IGKV1S3	17023418-17023704
1	Vκ4	IGKV1S4	17010212-17010498
1	Vκ5	IGKV1S5	16958378-16958664
1	Vκ6	IGKV1S6	16941855-16942141
1	Vκ7	IGKV1S7	16933657-16933943
1	Vκ8	IGKV1S8	16855771-16856057
1	VκORF1	IGKV1S9^{ORF}	16739610-16739895
1	VκORF2	IGKV1S10^{ORF}	16560248-16560552
1	pVκ2	IGKV1S11^{P}	17046559-17046846
1	pVκ4	IGKV1S12^{P}	16998516-16998803
1	pVκ6	IGKV1S13^{P}	16917360-16917646
1	pVκ7	IGKV1S14^{P}	16912894-16913180
1	pVκ8	IGKV1S15^{P}	16868993-16869281
1	pVκ9	IGKV1S16^{P}	16849772-16850060
1	pVκ10	IGKV1S17^{P}	16835919-16836203
1	pVκ11	IGKV1S18^{P}	16789755-16790042
1	pVκ12	IGKV1S19^{P}	16778531-16778817
1	pVκ13	IGKV1S20^{P}	16728653-16728934
1	pVκ14	IGKV1S21^{P}	16720475-16720760
1	pVκ19	IGKV1S22^{P}	16659096-16659390
1	pVκ20	IGKV1S23^{P}	16626800-16627104
2	Vκ9	IGKV2S1	16700635-16700951
2	pVκ24	IGKV2S2^{P}	16557308-16557612
3	Vκ10	IGKV3S1	16599670-16599971
3	Vκ11	IGKV3S2	16540025-16540326
3	Vκ14	IGKV3S3	16470843-16471144
3	Vκ16	IGKV3S4	16407261-16407562
3	Vκ17	IGKV3S5	16383431-16383732
3	Vκ18	IGKV3S6	16360142-16360443
3	pVκ17	IGKV3S7^{P}	16676005-16676305
3	pVκ21	IGKV3S8^{P}	16608233-16608528
3	pVκ22	IGKV3S9^{P}	16591859-16592160
3	pVκ28	IGKV3S10^{P}	16461689-16461983
3	pVκ29	IGKV3S11^{P}	16450171-16450478
3	pVκ30	IGKV3S12^{P}	16441110-16441414
3	pVκ34	IGKV3S13^{P}	16351930-16352220
4	Vκ12	IGKV4S1	16502696-16502982
4	pVκ18	IGKV4S2^{P}	16672345-16672638
5	Vκ13	IGKV5S1	16484542-16484828
6	Vκ15	IGKV6S1	16431518-16431804
7	Vκ19	IGKV7S1	16313053-16313339

IGKV subgroup	Designation Sun et al. 2010	Proposed designation	Genome Position
n.c.	pVκ1	IGKVxS1^{P}	17053257-17053545
n.c.	pVκ3	IGKVxS2^{P}	17004266-17004536
n.c.	pVκ5	IGKVxS3^{P}	16927468-16927746
n.c.	pVκ15	IGKVxS4^{P}	16709612-16709911
n.c.	pVκ16	IGKVxS5^{P}	16689071-16689357
n.c.	pVκ23	IGKVxS6^{P}	16581856-16582113
n.c.	pVκ25	IGKVxS7^{P}	16550552-15550840
n.c.	pVκ26	IGKVxS8^{P}	16534333-16534598
n.c.	pVκ27	IGKVxS9^{P}	16520693-16520979
n.c.	pVκ31	IGKVxS10^{P}	16434887-16435177
n.c.	pVκ32	IGKVxS11^{P}	16413133-16413356
n.c.	pVκ33	IGKVxS12^{P}	16370408-16370629
n.c.	pVκ35	IGKVxS13^{P}	16348990-16349239
n.c.	pVκ36	IGKVxS14^{P}	16342798-16343043
n.c.	pVκ37	IGKVxS15^{P}	16338745-16339014
n.c.	pVκ38	IGKVxS16^{P}	16332882-16333152
n.c.	pVκ39	IGKVxS17^{P}	16305235-16305453

IGKV: Sun et al., 2010, Chromosom 15, NW_001867379.1

Table 9: Nomenclature of the equine IGKJ gene segments

IGKJ subgroup	Designation Sun et al. 2010	Proposed designation	Genome Position
1	Jκ1	IGKJ1S1	13727254-13727217*
2	Jκ2	IGKJ2S1	13727624-13727586*
3	Jκ3	IGKJ3S1	13727931-13727894*
4	Jκ4	IGKJ4S1	13728269-13728232*
5	pJκ5	IGKJ5S1^{P}	13728570-13728534*

*Contig: NW_001867379.1

Table 10: Nomenclature of the equine IGKC genes

IGKC subgroup	Designation Sun et al. 2010	Proposed designation	Genome Position
1	Cκ1	IGKC1S1	13731499-13731821*

*Contig: NW_001867379.1

5. Transcription analyses of heavy and light chain genes

5.1. Heavy chain genes

Newborn foals are unable to mount an effective humoral immune response despite many features of the equine immune system having been developed during fetal life. Immunohistochemical experiments showed the expression of essential B-cell genes during equine gestation. Active gene recombination and isotype switching were suggested and underpinned by limited but detectable production of IgM and IgG at birth (Tallmadge et al., 2009). After demonstrating that equine neonatal B-cells produce all known five Ig isotypes, usage frequencies of the equine variable, diversity, and joining gene segments of the heavy and light chains were first investigated in two unrelated horses (Sun et al., 2010). Both animals showed a similar usage pattern of IGHV, IGHD, and IGHJ. Out of 50 IGHV, four (IGHV4S1, IGHV2S2, IGHV2S3, and IGHV4S5) were expressed but IGHV2S2 and IGHV2S3 were used preferably. Six out of eight IGHJ were recombined. IGHJ1S1 and IGHJ2S1 were not detected, whereas IGHJ1S5 was mainly rearranged. The analysis of the usage frequencies of the 40 germline IGHD revealed the dominant frequent use of IGHD18S1. The transcription of five additional IGHDs (IGDH3S1, IGHD12S1, IGHD8S2, IGHD8S3, IGHD28S1) was not observed. Eight clones indicated the incorporation of more than one IGHD into the IGHV-IGHD-IGHJ rearrangement resulting in IGHV-IGHD-IGHD-IGHJ. The second incorporated IGHD was supposed to be located 5' of the first IGHD in the germline locus (Sun et al., 2010). In addition to these studies of immunoglobulin diversity in adult horses, the level of Ig heavy chain variable region diversity was analyzed from fetal spleen, neonatal, foal, and adult horse mesenteric lymph node samples (Tallmadge et al., 2009). Interestingly, combinatorial and junctional levels of VDJ sequences in equine fetus were comparable to those of adult horses and a similar set of variable gene segments was used during fetal and post-natal life stages (Tallmadge et al., 2009; Tallmadge et al., 2013). During fetal life stage, 7 IGHV genes were identified with nucleotide identities of at least 97% compared to the donor germline sequences. Besides the gene segments IGHV2S2, IGHV2S3, IGHV1S3, IGHV4S2, IGHV4S5 and IGHV2S4, a newly identified IGHV4S17 was recognized. A reduced repertoire was observed in neonates where only 4 segments were used: IGHV2S2, IGHV2S3, IGHV2S4, and a novel IGHV1S6. In foals and adult horses, 5 and 4 IGHV were transcribed, respectively (foal: IGHV2S2, IGHV2S3, IGHV2S4, IGHV1S3, IGHV4S2; adult: IGHV2S2, IGHV2S3,

IGHV2S4, IGHV4S2). The segments IGHV2S2, IGHV2S3, and IGHV2S4 were used predominantly at all ages. Between the expressed gene segments, there are 250,000 bases suggesting that the entire IGHV locus is available throughout equine life (Tallmadge et al., 2013). Similarly, in fetal piglets the IGHV usage was described to be independent of the genome position although there is a limited number of IGHV representing the Ig repertoire, too (Butler et al., 2011; Eguchi-Ogawa et al., 2010). In contrast, the biased usage of germline IGHV in the mouse fetus was explained with different accessible positions within the IGHV locus (Jeong et al., 1988).

Analysis of the IGHD gene segments revealed that 13, 17, 15, and 20 different IGHDs were used in fetuses, neonates, foals, and adult horses. The predominant IGHD was IGHD18S1 at all ages. Four out of 8 IGHJ had an influence on diversity in all life stages, whereas IGHJ1S5 was the most common (Tallmadge et al., 2013).

VDJ diversity increased mainly during the last two thirds of gestation. Overall, there was no biased usage of a single IGHV, IGHD, or IGHJ gene segment at any age. In foals, first variations in the length of CDR2H were found. Sequence diversity and length variation further increased in CDR1H and CDR2H, and framework regions of adult horses in accordance to somatic hypermutation (Tallmadge et al., 2013). Non-templated nucleotide insertions of about 8 nucleotides were observed at the IGHV-IGHD junction in fetuses and neonates, although there was no significant increase in CDR3H length, varying from 7 to 21 amino acids in contrast to cattle (Walther et al., 2013). The length of the IGHV-IGHD and IGHD-IGHJ junctions increased significantly until birth and then remained similar in foals and adult horses. At least 5 amino acid residues were observed within the CDR3H of an adult horse, whereas 25 amino acid residues were counted in the longest CDR3H, which was identified in an equine neonate (Tallmadge et al., 2013). In comparison, recent studies showed the existence of very short CDR3H (5-10 amino acids), midlength CDR3H (11-31 amino acids) and exceptionally long CDR3H (more than 47 amino acids) in both bovine fetuses and adult cattle. This length heterogeneity is not isotype restricted and very long CDR3H contribute to diversity by uniquely folded small domains (Koti et al., 2010; Saini et al., 1999; Saini and Kaushik, 2002; Shojaei et al., 2003; Walther et al., 2013; Wang et al., 2013).

5.2. Light chain genes

For the IGKV, a preferential usage pattern was also observed. IGKV1S1, IGKV1S4, and IGKV2S1 were used most often. These gene segments showed an opposite transcriptional orientation to the IGKJ and IGKC cluster. All four functional IGKJ were expressed (Sun et al., 2010).
Within the transcribed λ-light chains, IGLVs belonging to subgroup 8 were preferred. IGLV from subgroup 6 and 4 were also transcribed but less frequently. The recombined IGLVs – especially IGLV8S2 and IGLV8S7 – were located within the IGLV cluster downstream of the IGLJ-IGLC cluster and were oriented in opposite transcriptional polarity. In both animals, examined by Sun and coworkers, genomic IGLJ-IGLC were amplified, whereas IGLJ-IGLC5 could not be amplified and IGLJ-IGLC7 was amplified successfully in one of these animals. IGLJ4S1$^{(P)}$ was found to be functional but without the single C-insertion, which is present in the database sequence of the equine genome project (Sun et al., 2010). Subsequent investigations of the transcription frequencies of λ-light chain genes in ten animals of Rhenish-German Coldblood (RGC) and Hanoverian Warmblood (HW), respectively, showed similar results (Hara et al., 2012). Eleven out of 144 IGLV were identified; all members of subgroup 8 as caused by the applied primers. In both breeds analyzed, the preferred use of two IGLV (IGLV8S1 and IGLV8S2) was observed. In addition, significant differences in usage frequencies of the IGLV (both within and between the breeds) were noted. Two pseudogenes were also recombined (IGLV8S10^{P} and IGLV8S12^{P}). The transcriptional orientation of the IGLV had no influence on the transcription frequencies. All four functional IGLC were also transcribed. IGLC6S1/7S1 was predominantly used in both breeds. These two constant region genes could not be distinguished due to high sequence similarity. Isotypes IGLC1S1 and IGLC5S1 revealed significantly higher transcription frequencies than IGLC4S1 in RGC and showed significant usage differences in HW. New allotypic variants were observed for IGLC1S1, IGLC5S1 and IGLC6S1/7S1 (Hara et al., 2012; Tallmadge et al., 2014).
The constant gene usage was also investigated in fetal spleen, neonatal, foal, and adult horse mesenteric lymph node tissues (Tallmadge et al., 2014). In equine fetuses, IGLC1S1, IGLC4S1 and IGLC5S1 were expressed. This feature changed in neonates, where IGLC1S1 and IGLC7S1 were transcribed. Both of these IGLC were still expressed in foal and adult horses but they also showed IGLC4S1. IGLC1S1, solely

found in all four developmental stages, was not the predominantly used constant region gene.

Furthermore, the usage and diversity of the equine IGLV was analyzed in fetal spleen and mesenteric lymph nodes of neonates, foals, and adult horses (Tallmadge et al., 2014). Analysis of the transcribed IGLV revealed usage of 13 IGLV in equine fetuses; 14 IGLV were used in neonates; 11 IGLV were identified in foals; and 9 IGLV were expressed in adult horses. Out of 27 functional genomic IGLVs, 18 different were identified in this study. In addition, the pseudogene IGLV8S10^{P} (Table 4) was also transcribed, as it had already been observed in Hanoverian Warmblood and Rhenish-German Coldblood horses (Hara et al., 2012). Interestingly, five sequences of equine fetus and neonate were not assigned to the currently defined germline IGLVs. These new IGLVs belonged to subgroups 4, 6, and 8, whereas the already known expressed IGLVs belonged to subgroups 2, 3, and 11. Within each age group, subgroup 8 was the most prevalent. Furthermore, the first allelic variants in horse IGLV were identified. Nine IGLV genes were found in only one developmental stage and 5 IGLV genes were identified in all life stages investigated. More and varying IGLVs were used during equine fetal life than in later stages. Nucleotide diversity increased significantly with time (Tallmadge et al., 2014).

6. New allotypic variants of IGLC

It is well known today that immunoglobulin light chains increase antibody variability and contribute to antigen binding. In addition to the use of non-functional variable gene segments due to gene conversion, genes coding for the constant regions also influence antibody diversity (Emorine et al., 1983; Moxley and Gibbs, 1992). Allotypic variants from alleles of IGHCs and IGLCs originate from single nucleotide polymorphisms (SNPs) and amino acid residue substitutions. Resulting epitope variability can improve immune responses as already shown for human and cattle. Allotypic markers of human light chains were associated with the susceptibility of infectious diseases caused by *Plasmodium falciparum* malaria, *Haemophilus influenza*, and *Meningococcus polysaccharides* for example (Giha et al., 2009; Granoff et al., 1984; Pandey et al., 1979). For allotypes of the bovine IgG2, correlations regarding complement activation, age-dependent expression, and influences on the effector function were reported in

binding of *Haemophilus somnus* immunoglobulin binding protein (IgBP) and cleavage by *Tritrichomonas foetus* extracellular cysteine proteinase (Bastida-Corcuera et al., 2000; Bastida-Corcuera et al., 1999a; Bastida-Corcuera et al., 1999c; Corbeil et al., 1997). Allelic variations of the bovine light chain have been described for different cattle breeds (Diesterbeck et al., 2012; Stein et al., 2012). Similarly, the constant region genes of the λ-light chains of the horse breeds RGC and HW showed five allotypic IGLC1S1 variants (IGLC1S1$^{a\text{-}e}$), four allotypic IGLC5S1 variants (IGLC5S1$^{a\text{-}d}$), and three allelic, as well as two allotypic IGLC6S1/7S1 variants (IGLC6S1/7S1$^{a1\text{-}a3,b}$). In RGC, IGLC1S1^{b,d}, IGLC5S1^{c,d} and IGLC6S1/7S1$^{a3,\ b}$ were observed, whereas IGLC1S1^{c} and IGLC5S1^{b} were detected in HW. Amino acid residue substitutions were identified inside the molecule, as well as on the solvent accessible surface. They were shown by homology-based 3D modeling. Conservative amino acid residue substitutions as well as changes of charge or hydrophobicity were identified. Replacements adjacent to the interface and within the interface to IGHC1 were supposed to potentially influence the stable connection between both the heavy and light chains (Hara et al., 2012). These investigations of the Ig lambda constant region genes in adult horses of two different breeds were complemented by the analyses of the Ig lambda joining gene segments in combination with the constant genes at different equine developmental stages (Tallmadge et al., 2014). In fetuses, the usage of germline IGLJ1S1-IGLC1S1, IGLJ4S1$^{(P)}$-IGLC4S1 and IGLJ5S1-IGLC5S1 was found. New allelic variants were identified for IGLJ1S1, IGLC1S1, IGLJ4S1$^{(P)}$ and IGLC4S1. The previously described isotype IGLC5S1^{b} was also expressed. Neonatal sequences showed the transcription of IGLJ1S1-IGLC1S1, as well as IGLJ7S1-IGLC7S1, whereas there were two allelic variants for IGLJ7S1-IGLC7S1. The same allelic IGLJ7S1-IGLC7S1 variants were also expressed in addition to one version of IGLJ1S1-IGLC1S1 in neonates. Both in foals and adult horses, IGLJ1S1-IGLC1S1, IGLJ4S1$^{(P)}$-IGLC4S1, and IGLJ7S1-IGLC7S1 were transcribed. Two alleles were identified for IGLC7S1 in both developmental stages. IGLC4S1 revealed two allelic versions in foal. Only IGLC1S1 was detected at all four life stages, but it was not the most frequently used gene at any stage (Tallmadge et al., 2014).

7. Future Directions

Since the late 19th century, horse antibodies in terms of anti-sera were closely connected with the medication of human diseases. Today equine antibody applications come more and more to the fore in order to prevent or treat equine infectious diseases. Amongst others they can be regarded as potential alternatives to antibiotic therapy in the near future.

Starting with the germline repertoire, it is, therefore, necessary to investigate the fundamental immunoglobulin genetics underlying the equine immunoglobulin immune response.

Recent analyzes already revealed individual and breed specific differences in the genomic fundament, as well as the gene usage at discrete life stages and an increased variability throughout life (Hara et al., 2012; Sun et al., 2010; Tallmadge et al., 2014; Tallmadge et al., 2013). However, more detailed studies are required to substantiate allelic or haplotypic differences for both variable and constant region genes and link them to individuals or breeds. Such detailed maps of the germline repertoire are currently available for human and mice. *In vitro* analyses of bovine IgG already determined the influences of genetic variations on the immune response (Bastida-Corcuera et al., 2000; Bastida-Corcuera et al., 1999a; Bastida-Corcuera et al., 1999c; Corbeil et al., 1997). Similar studies are still missing in horses as well as analyzes of their *in vivo* impact.

Subsequent investigations in genetic, structural, and configurational properties of equine immunoglobulins might offer opportunities for the development of new antibody-based immunotherapeutics. For instance, recombinant antibodies or antibody-fragment related products are among the fastest growing new therapeutics worldwide (Dübel, 2010). However, only few engineered species-specific antibodies have been developed against veterinary pathogens (Koti et al., 2014). The most promising outcome from latest equine immunoglobulin research might be the construction of tailor-designed antibodies based on the format of single chain fragments variable (scFv). In these antibodies VH and VL domains are linked together by, e.g., a $(Gly_4Ser)_3$ linker. When naïve or immunized immunoglobulin libraries are constructed for this reason by PCR amplification with well established primer sets for the equine VH and VL domains, highly affine and neutralizing recombinant scFv molecules can be assembled by phage display against any desired viral, bacterial, and parasitic antigenic site, as well as toxins. To elongate the short half-life time and clearance of scFvs *in vivo*, and to

achieve effector functions, such as opsonization, complement-dependent cytotoxicity (CDC), and antibody-dependent cellular cytotoxicity (ADCC), equine scFv-molecules can be enhanced in the same way as established for human antibodies to scFv-FCGRT (FcRn) fusion proteins (Kontermann, 2011; Koti et al., 2014) or to complete IgG antibodies through genetic manipulation.

Meanwhile, X-ray crystal structure analyses of antigen/antibody binding are increasing (Saini et al., 1999; Saini and Kaushik, 2002; Wang et al., 2013) and together with immunoglobulin sequencing data, even from high-throughput next generation sequencing, they will give valuable information on antigen-/antibody interactions. Subsequent amino acid replacement, especially in the CDR-regions, may contribute to enhanced binding efficiency of equine recombinant antibody molecules.

Consequently, future investigations of the basic biology once might lead into translation of new prophylaxis and treatment options of diseases.

Acknowledgments

We are grateful to Lynne Riddles for completing the critical review of this manuscript.

References

Almagro, J.C., Hernandez, I., Ramirez, M.C., Vargas-Madrazo, E., 1998. Structural differences between the repertoires of mouse and human germline genes and their evolutionary implications. Immunogenetics 47, 355e363.

Alt, F., Rosenberg, N., Lewis, S., Thomas, E., Baltimore, D., 1981. Organization and reorganization of immunoglobulin genes in A-MULV-transformed cells: rearrangement of heavy but not light chain genes. Cell 27, 381e390.

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. Zent. Vet. A 43, 573e576.

Bastida-Corcuera, F., Butler, J.E., Heyermann, H., Thomford, J.W., Corbeil, L.B., 2000. Tritrichomonas foetus extracellular cysteine proteinase cleavage of bovine IgG2 allotypes. J. Parasitol. 86, 328e332.

Bastida-Corcuera, F.D., Butler, J.E., Yahiro, S., Corbeil, L.B., 1999a. Differential complement activation by bovine IgG2 allotypes. Vet. Immunol. Immunopathol. 71, 115e123.

Bastida-Corcuera, F.D., Nielsen, K.H., Corbeil, L.B., 1999c. Binding of bovine IgG2a and IgG2b allotypes to protein A, protein G, and Haemophilus somnus IgBPs. Vet. Immunol. Immunopathol. 71, 143e149.

Behring von, E., 1918. Epidemiologie, Aetiologie und Bek€ampfung der Diphtherie. In: Beiträge zur experimentellen Therapie. Springer.

Bingel, A., 1918. Über Behandlung der Diphtherie mit gew€ohnlichem Pferdeserum. F. C. W. Vogel, Leipzig.

Breathnach, C.C., Clark, H.J., Clark, R.C., Olsen, C.W., Townsend, H.G., Lunn, D.P., 2006. Immunization with recombinant modified vaccinia Ankara (rMVA) constructs encoding the HA or NP gene protects ponies from equine influenza virus challenge. Vaccine 24, 1180e1190.

Brodeur, P.H., Riblet, R., 1984. The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. Eur. J. Immunol. 14, 922e930.

Burnett, R.C., Hanly, W.C., Zhai, S.K., Knight, K.L., 1989. The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 C alpha genes. EMBO J. 8, 4041e4047.

Butler, J.E., Sun, X., Wertz, N., Lager, K.M., Chaloner, K., Urban Jr., J., Francis, D.L., Nara, P.L., Tobin, G.J., 2011. Antibody repertoire development in fetal and neonatal piglets XXI. Usage of most VH genes remains constant during fetal and postnatal development. Mol. Immunol. 49, 483e494.

Butler, J.E., Wertz, N., Deschacht, N., Kacskovics, I., 2009. Porcine IgG: structure, genetics, and evolution. Immunogenetics 61, 209e230.

Calmette, A., 1896. Venin des Serpents. Physiologie de l`Envenimation Traitment des Morsures Venimeuses par le Seárum des Animaux Vaccineás. Socieáte ád`Eáditions Scienti ques, Paris.

Civardi, G., Curadi, M.C., Orlandi, M., Cattaneo, M., Giangiacomo, R., 2002. Capillar electrophoresis (CE) applied to analysis of mare's milk. Milchwissenschaft 57, 515e517.

Cole, R., Moore, H.F., 1917. The production of Antipneumococcic serum. J. Exp. Med. 26, 537e561.

Corbeil, L.B., Gogolewski, R.P., Kacskovics, I., Nielsen, K.H., Corbeil, R.R., Morrill, J.L., Greenwood, R., Butler, J.E., 1997. Bovine IgG2a antibodies to Haemophilus somnus and allotype expression. Can. J. Vet. Res. 61, 207e213.

Corthesy, B., 2007. Roundtrip ticket for secretory IgA: role in mucosal homeostasis? J. Immunol. 178, 27e32.

Daniels, J., 1921. Über die prophylaktische Wirkung des Diphtherieheilserums und des normalen antitoxinfreien Pferdeserums auf die Diphtherieinfektion des Meerschweinchens. Feller, Neubrandenburg.

Das, S., Nikolaidis, N., Klein, J., Nei, M., 2008. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. Proc. Natl. Acad. Sci. U. S. A. 105, 16647e16652.

Diesterbeck, U.S., Aboelhassan, D.M., Stein, S.K., Czerny, C.P., 2012. Detection of new allotypic variants of bovine lambda-light chain constant regions in different cattle breeds. Dev. Comp. Immunol. 36, 130e139.

Drogoul, C., Cl_ement, F., Ventorp, M., Curadi, M.C., Orlandi, M., 2008. Euine passive immune transfer trough colostrum. Horse Health Nutrition. In: 4th European Nutrition & Health Congress, vol. 18e19, pp. 23e27.

Dübel, S., 2010. Handbook of Therapeutic Antibodies. Weiley-Blackwell, Weinheim, Germany.

Edelman, G.M., 1973. Antibody structure and molecular immunology. Science 180, 830e840.

Eguchi-Ogawa, T., Wertz, N., Sun, X.Z., Piumi, F., Uenishi, H., Wells, K., Chardon, P., Tobin, G.J., Butler, J.E., 2010. Antibody repertoire development in fetal and neonatal piglets. XI. The relationship of variable heavy chain gene usage and the genomic organization of the variable heavy chain locus. J. Immunol. 184, 3734e3742.

Emorine, L., Dreher, K., Kindt, T.J., Max, E.E., 1983. Rabbit immunoglobulin kappa genes: structure of a germline b4 allotype J-C locus and evidence for several b4- related sequences in the rabbit genome. Proc. Natl. Acad. Sci. U. S. A. 80, 5709e5713.

Ford, J.E., Home, W.A., Gibson, D.M., 1994. Light chain isotype regulation in the horse. Characterization of Ig kappa genes. J. Immunol. 153, 1099e1111.

Franz, L.C., Landon, J.C., Lopes, L.A., Marinho, L.A., Sarma, C., Bruemmer, J., Squires, E.L., 1998. Oral and intravenous immunglobulin therapy in neonatal foals. Equine Vet. Sci. 18, 742e748.

Gibson, D., 1974. Structural studies on normal horse immunoglobulin light chains. Detection of k-type N-terminal sequences. Biochemistry 13, 2776e2785.

Giha, H.A., Nasr, A., Iriemenam, N.C., Arnot, D., Troye-Blomberg, M., Theander, T.G., Berzins, K., ElGhazali, G., Pandey, J.P., 2009. Antigen-specific influence of GM/KM allotypes on IgG isotypes and association of GM allotypes with susceptibility to Plasmodium falciparum malaria. Malar. J. 8, 306.

Giudicelli, V., Lefranc, M.P., 1999. Ontology for immunogenetics: the IMGTONTOLOGY. Bioinformatics 15, 1047e1054.

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, 514e532.

Granoff, D.M., Pandey, J.P., Boies, E., Squires, J., Munson Jr., R.S., Suarez, B., 1984. Response to immunization with Haemophilus influenzae type b polysaccharidepertussis vaccine and risk of Haemophilus meningitis in children with the Km(1) immunoglobulin allotype. J. Clin. Investig. 74, 1708e1714.

Guo, Y., Bao, Y.,Wang, H., Hu, X., Zhao, Z., Li, N., Zhao, Y., 2011. A preliminary analysis of the immunoglobulin genes in the african elephant (Loxodonta africana). PLoS One 6, e16889.

Guo, Y., Bao, Y., Meng, Q., Hu, X., Meng, Q., Ren, L., Li, N., Zhao, Y., 2012. Immunoflobulin genomics in the guinea pig (Cavia porcellus). PLoS One 7, e39298.
Haas, S.D., Bristol, F., Card, C.E., 1996. Risk factors associated with the incidence of foal mortality in an extensively managed mare herd. Can. Vet. J. 37, 91e95.
Hara, S., Diesterbeck, U.S., Konig, S., Czerny, C.P., 2012. Transcriptional analysis of equine lambda-light chains in the horse breeds Rhenish-German Coldblood and Hanoverian Warmblood. Vet. Immunol. Immunopathol. 145, 50e65.
Hawgood, B.J., 1999. Doctor Albert Calmette 1863e1933: founder of antivenomous serotherapy and of antituberculous BCG vaccination. Toxicon 37, 1241e1258.
Helms, C.M., Allen, P.Z., 1970. Studies on equine immunoglobulins, II. Antigenic interrelationships among horse IgG(T) and antipneumococcal gamma-1-component. J. Immunol. 105, 1253e1263.
Hill, W.C., Cebra, J.J., 1965. Horse anti-SI immunoglobulins. I. Properties of gamma-M-antibody. Biochemistry 4, 2575e2584.
Holznagel, D.L., Hussey, S., Mihalyi, J.E., Wilson, W.D., Lunn, D.P., 2003. Onset of immunoglobulin production in foals. Equine Vet. J. 35, 620e622.
Home, W.A., Ford, J.E., Gibson, D.M., 1992. L chain isotype regulation in horse. I. Characterization of Ig lambda genes. J. Immunol. 149, 3927e3936.
Jenvey, C., Caraguel, C., Howarth, G.B., Riley, C.B., 2012. Identification of periparturient mare and foal associated predictors of post parturient immunoglobulin A concentrations in thoroughbred foals. Equine Vet. J. Suppl. 73e77.
Jeong, H.D., Komisar, J.L., Kraig, E., Teale, J.M., 1988. Strain-dependent expression of VH gene families. J. Immunol. 140, 2436e2441.
Kabat, E.A.,Wu, T.T., 1991. Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. J. Immunol. 147, 1709e1719.
Keggan, A., Freer, H., Rollins, A., Wagner, B., 2013. Production of seven monoclonal equine immunoglobulins isotyped by multiplex analysis. Vet. Immunol. Immunopathol. 153, 187e193.
Klinman, N.R., Rockey, J.H., Karush, F., 1965. Equine antihapten antibody Ii. The gamma-G(7s-gamma) components and their specific interaction. Int. J. Cancer Suppl. 47, 51e60.
Kohn, C.W., Knight, D., Hueston,W., Jacobs, R., Reed, S.M., 1989. Colostral and serum IgG, IgA, and IgM concentrations in Standardbred mares and their foals at parturition. J. Am. Vet. Med. Assoc. 195, 64e68.
Kontermann, R.E., 2011. Strategies for extended serum half-life of protein therapeutics. Curr. Opin. Biotechnol. 22, 868e876.
Koterba, A.M., Brewer, B.D., Tarplee, F.A., 1984. Clinical and clinicopathological characteristics of the septicaemic neonatal foal: review of 38 cases. Equine Vet. J. 16, 376e382.
Koti, M., Kataeva, G., Kaushik, A.K., 2010. Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies. Mol. Immunol. 47, 2119e2128.
Koti, M., Saini, S.S., Sachan, A., Kaushik, A., 2014. Engineered bovine antibodies in the development of novel therapeutics, immunomodulators and vaccines. Antibodies 3, 205e214.
Kydd, J.H., Townsend, H.G., Hannant, D., 2006. The equine immune response to equine herpesvirus-1: the virus and its vaccines. Vet. Immunol. Immunopathol. 111, 15e30.

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, 298e305.

Leblanc, M.M., Tran, T., 1987. Relationship among colostral electrolytes, colostral IgG concentration and absortion of colostral IgG by foals. J. Reprod. Fertil. 35, 735e736.

LeBlanc, M.M., Tran, T., Baldwin, J.L., Pritchard, E.L., 1992. Factors that influence passive transfer of immunoglobulins in foals. J. Am. Vet. Med. Assoc. 200, 179e183.

Lefranc, M.P., 1998. IMGT (ImMunoGeneTics) locus on focus. A new section of experimental and clinical immunogenetics. Exp. Clin. Immunogenet. 15.

Lefranc, M.P., 2001a. IMGT, the international ImMunoGeneTics database. Nucleic Acids Res. 29, 207e209.

Lefranc, M.P., 2001b. Nomenclature of the human immunoglobulin lambda (IGL) genes. Exp. Clin. Immunogenet. 18, 242e254.

Lefranc, M.P., 2007. WHO-IUIS nomenclature subcommittee for immunoglobulins and T cell receptors report. Immunogenetics 59, 899e902.

Leleu, X., Terriou, L., Duhamel, A., Moreau, A.S., Andrieux, J., Dupire, S., Coiteux, V., Berthon, C., Micol, J.B., Guieze, R., Facon, T., Bauters, F., 2006. Long-term outcome in acquired aplastic anemia treated with an intensified dose schedule of horse antilymphocyte globulin in combination with androgens. Ann. Hematol. 85, 711e716.

Lewis, M.J., Wagner, B., Woof, J.M., 2008. The different effector function capabilities of the seven equine IgG subclasses have implications for vaccine strategies. Mol. Immunol. 45, 818e827.

Lopez, A.M., Hines, M.T., Palmer, G.H., Alperin, D.C., Hines, S.A., 2002. Identification of pulmonary T-lymphocyte and serum antibody isotype responses associated with protection against Rhodococcus equi. Clin. Diagn. Lab. Immunol. 9, 1270e1276.

Macpherson, A.J., Gatto, D., Sainsbury, E., Harriman, G.R., Hengartner, H., Zinkernagel, R.M., 2000. A primitive T cell-independent mechanism of intestinal mucosal IgA responses to commensal bacteria. Science 288, 2222e2226.

McGuire, T.C., Crawford, T.B., 1972. Identification and quantitation of equine serum and secretory immunoglobulin A. Infect. Immun. 6, 610e615.

McGuire, T.C., Crawford, T.B., 1973. Passive immunity in the foal: measurement of immunoglobulin classes and specific antibody. Am. J. Vet. Res. 34, 1299e1303.

Mestecky, J., Jackson, S., Moldoveanu, Z., Nesbit, L.R., Kulhavy, R., Prince, S.J., Sabbaj, S., Mulligan, M.J., Goepfert, P.A., 2004. Paucity of antigen-specific IgA responses in sera and external secretions of HIV-type 1-infected individuals. AIDS Res. Hum. Retrovir. 20, 972e988.

Moxley, G., Gibbs, R.S., 1992. Polymerase chain reaction-based genotyping for allotypic markers of immunoglobulin kappa shows allelic association of Km with kappa variable segment. Genomics 13, 104e108.

Navarro, P., Barbis, D.P., Antczak, D., Butler, J.E., 1995. The complete cDNA and deduced amino acid sequence of equine IgE. Mol. Immunol. 32, 1e8.

Nelson, K.M., Schram, B.R., McGregor, M.W., Sheoran, A.S., Olsen, C.W., Lunn, D.P., 1998. Local and systemic isotype-specific antibody responses to equine influenza virus infection versus conventional vaccination. Vaccine 16, 1306e1313.

Oethinger, M., Wand-Wurttenberger, A., Hengel, H., 1997. Mirkobiologie und Immunologie. Gustav Fischer Verlag, Ulm.

Overesch, G.,Wagner, B., Radbruch, A., Leibold,W., 1998. Organisation of the equine immunoglobulin constant heavy chain genes. II. Equine cgamma genes. Vet. Immunol. Immunopathol. 66, 273e287.
Pahud, J.J., Mach, J.P., 1972. Equine secretory IgA and secretory component. Int. Arch. Allergy Appl. Immunol. 42, 175e186.
Pandey, J.P., Fudenberg, H.H., Virella, G., Kyong, C.U., Loadholt, C.B., Galbraith, R.M., Gotschlich, E.C., Parke Jr., J.C., 1979. Association between immunoglobulin allotypes and immune responses to Haemophilus influenzae and Meningococcus polysaccharides. Lancet 1, 190e192.
Reynaud, C.A., Anquez, V., Dahan, A., Weill, J.C., 1985. A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. Cell 40, 283e291.
Rockey, J.H., 1967. Equine antihapten antibody. The subunits and fragments of antibeta-lactoside antibody. J. Exp. Med. 125, 249e275.
Rockey, J.H., Klinman, N.R., Karush, F., 1964. Equine antihapten antibody. I. 7s Beta-2a- and 1os Gamma-1- globulin components of purified anti-beta-lactoside antibody. J. Exp. Med. 120, 589e609.
Saini, S.S., Allore, B., Jacobs, R.M., Kaushik, A., 1999. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. Eur. J. Immunol. 29, 2420e2426.
Saini, S.S., Kaushik, A., 2002. Extensive CDR3H length heterogeneity exists in bovine foetal VDJ rearrangements. Scand. J. Immunol. 55, 140e148.
Sandor, G., Korach, S., Mattern, P., 1964a. 7 S globulin, immunologically identical to 19 S gamma-1-(beta-2-)-M-globulin, a new protein of horse serum. Nature 204, 795e796.
Schrenzel, M.D., King, D.P., McKnight, M.L., Ferrick, D.A., 1997. Characterization of horse (Equus caballus) immunoglobulin mu chain-encoding genes. Immunogenetics 45, 386e393.
Sedlinska, M., Krejci, J., Vyskocil, M., Kudiackova, H., 2006. Postnatal development of blood serum concentrations of immunoglobulin IgG, IgA and IgM isotypes of suckling foals. Acta Vet. Brno 75, 175e182.
Sheoran, A.S., Holmes, M.A., 1996. Separation of equine IgG subclasses (IgGa, IgGb and IgG(T)) using their differential binding characteristics for staphylococcal protein A and streptococcal protein G. Vet. Immunol. Immunopathol. 55, 33e43.
Sheoran, A.S., Sponseller, B.T., Holmes, M.A., Timoney, J.F., 1997. Serum and mucosal antibody isotype responses to M-like protein (SeM) of Streptococcus equi in convalescent and vaccinated horses. Vet. Immunol. Immunopathol. 59, 239e251.
Shojaei, F., Saini, S.S., Kaushik, A.K., 2003. Unusually long germline DH genes contribute to large sized CDR3H in bovine antibodies. Mol. Immunol. 40, 61e67.
Souza, C.M., Miotto, B.A., Bonin, C.P., Camargo, M.M., 2010. Lower serum IgA levels in horses kept under intensive sanitary management and physical training. Animal 4, 2080e2083.
Stein, S.K., Diesterbeck, U.S., Aboelhassan, D.M., Czerny, C.P., 2012. Comparison of joining and constant kappa-light chain regions in different cattle breeds. Anim. Genet. 43, 776e780.
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, 1009e1020.
Suter, M., Fey, H., 1981. Isolation and characterization of equine IgE. Zent. Vet. B 28, 414e420.

Tallmadge, R.L., McLaughlin, K., Secor, E., Ruano, D., Matychak, M.B., Flaminio, M.J., 2009. Expression of essential B cell genes and immunoglobulin isotypes suggests active development and gene recombination during equine gestation. Dev. Comp. Immunol. 33, 1027e1038.
Tallmadge, R.L., Tseng, C.T., Felippe, M.J., 2014. Diversity of immunoglobulin lambda light chain gene usage over developmental stages in the horse. Dev. Comp. Immunol. 46, 171e179.
Tallmadge, R.L., Tseng, C.T., King, R.A., Felippe, M.J., 2013. Developmental progression of equine immunoglobulin heavy chain variable region diversity. Dev. Comp. Immunol. 41, 33e43.
Theakston, R.D., Warrell, D.A., Griffiths, E., 2003. Report of a WHO workshop on the standardization and control of antivenoms. Toxicon 41, 541e557.
Tonegawa, S., 1983. Somatic generation of antibody diversity. Nature 302, 575e581.
Turtinen, L.W., Allen, G.P., 1982. Identification of the envelope surface glycoproteins of equine herpesvirus type 1. J. Gen. Virol. 63, 481e485.
Vaerman, J.P., Querinjean, P., Heremans, J.F., 1971. Studies on the IgA system of the horse. Immunology 21, 443e454.
Wagner, B., 2006. Immunoglobulins and immunoglobulin genes of the horse. Dev. Comp. Immunol. 30, 155e164.
Wagner, B., Greiser-Wilke, I., Antczak, D.F., 2003. Characterization of the horse (Equus caballus) IGHA gene. Immunogenetics 55, 552e560.
Wagner, B., Greiser-Wilke, I., Wege, A.K., Radbruch, A., Leibold, W., 2002. Evolution of the six horse IGHG genes and corresponding immunoglobulin gamma heavy chains. Immunogenetics 54, 353e364.
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, 3230e3242.
Wagner, B., Miller, W.H., Morgan, E.E., Hillegas, J.M., Erb, H.N., Leibold, W., Antczak, D.F., 2006. IgE and IgG antibodies in skin allergy of the horse. Vet. Res. 37, 813e825.
Walther, S., Czerny, C.P., Diesterbeck, U.S., 2013. Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins. PLoS One 8, e64234.
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. Reshaping antibody diversity. Cell 153, 1379e1393.
Weir, R.C., Porter, R.R., Givol, D., 1966. Comparison of the C-terminal amino-acid sequence of two horse immunoglobulins IgG and IgG(T). Nature 212, 205e206.
Widders, P.R., Stokes, C.R., Bourne, F.J., 1986. Investigation of the antigenic relationship between equine IgG and IgGT. Vet. Immunol. Immunopathol. 13, 255e259.
Wilson, I.B., Gavel, Y., von Heijne, G., 1991. Amino acid distributions around O-linked glycosylation sites. Biochem. J. 275 (Pt 2), 529e534.
Winau, F.,Winau, R., 2002. Emil von Behring and serum therapy. Microbes Infect. 4, 185e188.
Zand, M.S., 2006. B-cell activity of polyclonal antithymocyte globulins. Transplantation 82, 1387e1395.
Zhao, Y., Kacskovics, I., Pan, Q., Liberles, D.A., Geli, J., Davis, S.K., Rabbani, H., Hammarstrom, L., 2002. Artiodactyl IgD: the missing link. J. Immunol. 169, 4408e4416.
Zhao, Y., Kacskovics, I., Rabbani, H., Hammarstrom, L., 2003. Physical mapping of the bovine immunoglobulin heavy chain constant region gene locus. J. Biol. Chem. 278, 35024e35032.

Zolla, S., Goodman, J.W., 1968. An aggregating immunoglobulin in hyperimmune equine anti-pneumococcal sera. J. Immunol. 100, 880e897.

Transcriptional analysis of equine λ-light chains in the horse breeds Rhenish-German Coldblood and Hanoverian Warmblood

Stefanie Hara[1], Ulrike S. Diesterbeck[1]*, Sven König[2], Claus-Peter Czerny[1]

[1]Department of Animal Sciences, Institute of Veterinary Medicine, Division of Microbiology and Animal Hygiene, Faculty of Agricultural Sciences, Georg-August University Göttingen, Burckhardtweg 2, 37077 Göttingen, Germany
[2]Department of Animal Breeding, University of Kassel, Nordbahnhofstr. 1a, 37213 Witzenhausen, Germany

*Corresponding author:
U. S. Diesterbeck
Tel.: +49 551 393375
Fax: +49 551 3913513
E-mail: udieste@gwdg.de
Postal address: Dr. Ulrike Diesterbeck
Department of Animal Sciences,
Georg-August University Göttingen
Burckhardtweg 2
37077 Göttingen, Germany

Published in "Veterinary Immunology and Immunopathology"

DOI: 10.1016/j.vetimm.2011.10.006
URL: http://www.sciencedirect.com/science/article/pii/S0165242711004132

Abstract

The present study analyzed equine λ-light chain genes (IGLV and IGLC) transcribed in the horse breeds Rhenish-German Coldblood (RGC) and Hanoverian Warmblood (HW). Primers were generated for the major expressed IGLV subgroup 8. The significant majority of the sequences represented IGLC6/7. In RGC, IGLC1 and IGLC5 were observed in significant higher frequencies than IGLC4. In HW, significant differences were obtained for the transcription of IGLC1 and IGLC5. IGLC4 was not determined in this breed. Five allotypic *IGLC1* variants, four allotypic IGLC5 variants, and three allelic as well as two allotypic IGLC6/7 variants were identified. IGLC1$^{b, d}$, IGLC5$^{c, d}$, and IGLC6/7$^{a3, b}$ were detected in RGC while IGLC1^{c} and IGLC5^{b} were solely found in HW. Furthermore, 11 out of 144 known IGLV-segments were transcribed of which IGLV15 and IGLV17 were preferred significantly. IGLV25 displayed significant differences in the rearrangement between both breeds. The classified pseudogenes IGLV101ψ and IGLV74ψ were also identified. Rearrangements with IGLC-genes showed significant differences for IGLV15 in both breeds, whereas IGLV25 also revealed significant differences between the breeds. The transcriptional orientation of the functional segments has no influence on the occurrence of the IGLV.

Keywords

Equus caballus; horse; immunoglobulin; immunoglobulin light chains; allotype

1. Introduction

In jawed vertebrates, immunoglobulins (Ig) are usually described as a Y-shaped tetramer composed of two identical heavy chains (IGH) and two identical light chains (IGL) (Edelman, 1973). Both chains are comprised of variable and constant domains. During differentiation of B-lymphocytes they are joined together by combinatorial assembly of separate heavy and light chain variable (V_{H+L}), diversity (D_H), joining (J_{H+L}), and constant (C_{H+L}) germline components (Tonegawa, 1983). While IGH are encoded by a single locus, mammals express 2 isotypes of IGL, which are called kappa (κ) and lambda (λ) (Korngold and Lipari, 1956). In comparison to IGH, evolution of IGL seems to be more complex. Sequences similar to κ light chains described in different species, such as mammals and fish, gave rise to this assumption (Sitnikova and Su, 1998). In *Xenopus laevis* an additional IGL of σ-isotype is present (Klein et al., 2002). The light chain isotypes can be distinguished by specific conserved amino acid motifs as well as by distinct species specific usage frequencies (Das et al., 2008). For instance, in cattle, sheep, and horses λ-light chains are predominantly expressed (Arun et al., 1996; Home et al., 1992). This isotype occurs in more than 90% of horse serum antibodies (Gibson, 1974).

In several genetic investigations equine IGH constant regions (IGHC) as well as the germline V_H, D_H, and J_H segments were mapped to *Equus caballus* autosome (ECA) 24 and the genomic scaffold UN0011. Eleven IGHC, 50 V_H, 40 D_H, and eight J_H segments were described (Sun et al., 2010; Wagner, 2006). The genomic organization of equine IGL is characterized by 144 variable λ-light chain segments (IGLV) and 60 variable κ-light chain segments (IGKV). While IGKV and IGKJ-segments are organized, as it is the case, for example, in humans and mice, IGLV-segments are distributed into two clusters located up- and downstream of the IGLJ-IGLC-cluster (Kawasaki et al., 2001; Roschenthaler et al., 2000; Sun et al., 2010). The genetic information for equine λ-light chains was found to be located on ECA8. Genes for equine κ-light chains are located on ECA15. The number of genes is supposed to be responsible for the dominant λ-light chain expression (Almagro et al., 2006; Sun et al., 2010). 112 IGLV-pseudogenes (IGLVψ) were described for the λ-locus (Sun et al., 2010). Beside IGLV, seven IGLC were described for the equine genomic λ-locus. Each was preceded by an IGLJ-segment. Genomic analyses revealed that three IGLC were functional, while in transcriptional analyses four IGLC were determined to be functional (Das et al., 2008; Sun et al., 2010).

In general, light chains enlarge antibody variability and contribute to antigen binding. Both can be enhanced by mechanisms of somatic hypermutation and gene conversion (Arakawa and Buerstedde, 2009; Parng et al., 1996; Winstead et al., 1999; Zhou et al., 2004). Beside gene conversion utilizing nonfunctional V-gene segments, genes for the constant regions can also contribute to antibody diversity (Emorine et al., 1983; Moxley and Gibbs, 1992). Allotypic variants emerging from alleles of the constant regions of IGH and IGL as well as the resulting epitope variability can improve immune responses (Bastida-Corcuera et al., 2000; Bastida-Corcuera et al., 1999a; Bastida-Corcuera et al., 1999b; Oli et al., 2004). Similar allelic variations of the light chains have already been described for different cattle breeds (Diesterbeck et al., 2012) but not yet for horses. In addition, no structural analyses of the inter-domain interface between the first constant IGH and IGL regions are available. Comparable data were obtained by analyses of the crystal structures of two murine and two human Fabs (Padlan et al., 1986).

We investigated two horse breeds (Rhenish-German Coldblood, RGC, and Hanoverian Warmblood, HW) with different stud book sizes and breeding goals. Our interest was a comparative analysis of IGLC sequences and the identification of different alleles and putative allotypes. The locations of allotype-specific amino acid residue substitutions within the molecule were examined using comparative 3D modeling with known crystal structures. Furthermore, the usage frequency and rearrangement of IGLV-segments were examined for comparison of the λ-light chain repertoire in HW and RGC.

2. Material and methods

2.1. Breed selection, isolation of lymphocytes and cDNA-synthesis

For the analysis of breed specific expression of the immunoglobulin λ-light chain the equine breeds Rhenish-German Coldblood (RGC) and Hanoverian Warmblood (HW) were chosen as these breeds differ in stud book size and breeding goals. In RGC the stud book is closed resulting in a true-breeding population. For HW the stud book is open and the breeding concept is more flexible (Hartmann et al., 2006).

Peripheral blood samples were collected from 10 randomly chosen animals per breed. All animals of one breed were located in one stable. Peripheral blood mononuclear cells (PBMC) were isolated from 20 ml EDTA-blood using Ficoll gradients (GE Healthcare Europe GmbH, Germany) according to the manufacturer's protocol. After staining with trypan blue viable PBMC were counted and total RNA was extracted from $1x10^7$ cells by using the RNeasy® Mini Kit (QIAGEN, Germany). For synthesis of cDNA from 3 µg of total RNA in a total volume of 20 µl $pd(N)_6$-primers were used (First-Strand cDNA Synthesis Kit, GE Healthcare Europe GmbH, Germany).

2.2. PCR amplification of the expressed immunoglobulin λ-light chain repertoire

Complete λ-light chains were amplified using primers binding in the 5'UTR (eIgL_5'UTR_for: 5'-GCCACAGAAGGCAGGACTCGG-3') and the 3'UTR (eIgL_3'UTR_rev1: 5'-AGGGGTCCCGTGACAGCAGG-3'). Primer eIgL_5'UTR_for hybridizes at positions 2-22 in the reference sequence L07565 (Home et al., 1992). Primer eIgL_3'UTR_rev is complementary to positions 764-783 in the same reference sequence. The expected product size of the λ-light chain including leader, variable, and constant region was 782 bp. By using the eIgL_5'UTR_for primer, subgroup 8-IGLV were amplified.

A total reaction volume of 50 µl contained 2 µl of cDNA, 200 µM dNTPs (Bioline, Germany), 5 µl of 10x PCR buffer (75 mM Tris HCl pH 9.0; 2 mM $MgCl_2$; 50 mM KCl; 20 mM $(NH_4)_2SO_4$), 0.4 µM of each primer, and 2 units DNA polymerase (Biotools, Spain). Thermal cycling was carried out at 95°C for 5 min, 95°C for 1 min, 65°C for 1 min, 72°C for 2 min for 35 cycles with initial denaturation at 95°C for 5 min and terminal elongation at 72°C for 10 min. The length and purity of the PCR products was evaluated by agarose gel electrophoresis.

2.3. PCR amplification of the germline IGLJ-IGLC pairs

To examine the presence of all predicted germline IGLJ-IGLC pairs, they were amplified from genomic DNA, which was isolated from EDTA-blood (QIAamp® DNA Blood Mini Kit, QIAGEN, Germany). Primers for the functional and transcribed IGLC were based on the breed specific sequence information derived from the analyses of the expressed λ-light chains. To generate primers for nonfunctional IGLC, sequences from Acc. No. NW_001867428.1 were used as reference. The IGLJ-IGLC pair specific primers are shown in Table 1 including sequences, start and stop positions within the reference as well as product sizes. For IGLJ5-IGLC5 and IGLJ7-IGLC7 identical reverse primers were used. IGLJ6-IGLC6 was not amplified particularly because only one difference to IGLJ7-IGLC7 concerning one nucleotide could be identified. Product size varies from 1614 bp to 1870 bp.

Table 1
Primers for amplification of germline IGLJ-IGLC pairs. Binding positions for primers amplifiying IGLJ1-IGLC1 to IGLJ7-IGLC7 and referring to NW_001867428.1 in the equine genome assembly EquCab2.0.

Primers	Sequences 5‘ > 3‘	Product Size	Binding positions
eIgL_J1-C1L_for eIgL_J1-C1L_rev	AGAACACTCTGAGGGGGACAGT TGAGACTCAGTCACCGTGGT	1870 bp	4144485…4144506 4146335…4146355
eIgL_J2-C2L_for eIgL_J2-C2L_rev	CCTCTGAGGGGGACACTTTCT TGTGTCACAGTGTTCATAGTTCAC	1667 bp	4139378…4139399 4141021…4141045
eIgL_J3-C3L_for eIgL_J3-C3L_rev	AGCACACTCTGAGGGGGACACTTT CTGTGTCACTGTGTCCTGTGC	1614 bp	4133976…4134000 4135569…4135590
eIgL_J4-C4L_for eIgL_J4-C4L_rev	AAGGACACTCTGAGGGGGACAGT TGCTGTATTCGGCGGAGGCAC	1743 bp	4130704…4130727 4132426…4132447
eIgL_C5L_for eIgL_J7-C5L_rev	TTCCACTCTGTGGAGGTCCGTG ACCTGACCATCGCAGGTGAGTC	1635 bp	4127429…4127450 4129063…4129042
eIgL_C7L_for eIgL_J7-C5L_rev	TTCCACTGTGCGGAAGTCCGC ACCTGACCATCGCAGGTGAGTC	1637 bp	4120613…4120633 4122249…4122228

A total reaction volume of 50 µl contained 2 µl of cDNA, 200 µM dNTPs (Bioline, Germany), 5 µl of 10x PCR buffer (75 mM Tris-HCl pH 9.0; 2 mM $MgCl_2$; 50 mM KCl; 20 mM $(NH_4)_2SO_4$), 3% of DMSO; 0.4 µM of each primer, and 2 units of DNA polymerase (Biotools, Spain). Thermal cycling was performed at 95°C for 5 min, 95°C for 1 min, 60°C for 1 min, 72°C for 2 min for 35 cycles with initial denaturation at 95°C for 5 min and terminal elongation at 72°C for 10 min. Length and purity of the PCR products were evaluated by agarose gel electrophoresis.

2.4. Purification and cloning of PCR products

For reasons of purity and amount, PCR-products were either purified by gel extraction or they were concentrated. For gel extraction the MiniElute Gel Extraction Kit was used (QIAGEN, Germany) according to the manufacturer's protocol. The QG buffer was replaced by a QX1 buffer. Samples were eluted with 13 µl of 10 mmol Tris/HCl; pH 8.5. To increase the concentration of the PCR-products the DNA Clean & Concentrator™ Kit was used (Zymo Research Corporation, CA92867 U.S.A) in accordance to the manufacturer's protocol. Purified products were sub-cloned into the pCR®2.1-TOPO® 3.9 kb TA vector and transformed into chemically competent One Shot TOP 10 *E. coli* cells (Invitrogen™, Germany). After blue-white selection, 12 randomly chosen white transformants, which had been generated individually from each horse, were grown in 5 ml LB-ampicillin broth. Plasmid isolation occurred using a MiniPrep Kit (QIAGEN, Germany). In order to verify the correct insert size, plasmids were cleaved with *Eco*RI (New England Biolabs, Germany).

2.5. Sequencing and sequence analysis of PCR products

A definite number of 12 clones per animal were sequenced according to the chain-termination method (Sanger et al., 1977). The M13 (-20) Forward and M13 Reverse (Invitrogen, Germany) vector specific primers as well as the corresponding gene specific primers eIgL_5'UTR_for/ eIgL_3'UTR_rev were used for sequencing. The sequences were analyzed with the DNAstar program (GATC Biotech AG, Germany) and aligned by ClustalW (Thompson et al., 1994) to the nucleotide sequences from the genomic IGLV and IGLC published by Sun et al. (2010). The nomenclature for the transcribed IGLV and IGLC corresponds to this reference. The IMGT numbering system was used for numbering of amino acid residues (Lefranc et al., 2005). In order to determine the different isotypes, specific amino acid residue motifs were identified. Unique features in IGLC1 are SWK at positions 40-42, KSSSSV at positions 97-102, and VTH at positions 106-108. IGLC6/7 can only be distinguished from IGLC5 due to the combination of special motifs. In IGLC6/7 STPS, DAVTT, and TRTSAQW are characteristic at positions 1.1-3, 45-45.4, and 90-96. In contrast, IGLC5 has SAPS, DAVTN, and TRTSTEW at these positions. Identities between IGLC5 and IGLC4 are KSYSSV and VKH (positions 97-102 and 106-108). IGLC4 and IGLC1 share SWK, GAATT, and PLTPTQW at positions 40-42, 45-45.4, and 90-96. In addition, phylogenetic trees as well as sequence identity of at least 80% were used for grouping the transcribed with the genomic IGLC and IGLV. New alleles of IGLC were defined if

substitutions were detected in at least two clones sequenced from one animal. Furthermore, these substitutions had to occur in at least two animals in order to exclude sequencing errors derived from reverse transcriptase and DNA polymerase amplification. Nucleotide substitutions found in more than four sequences of one animal were also defined as new allele. All alleles were compared with the equine ESTs (expressed sequence tags) database (gp/9796.11760/9796_est; 37199 sequences) using BLASTN. Breed information derived from sequences with 100% coverage and identity were also considered. In sequences received from the two horse breeds RGC and HW combinations of IGLV with IGLC-isotypes were analyzed with regard to occurrence and frequency.

2.6. Homology-based modeling of the λ-constant region

The amino acid sequences of the transcribed IGLC were deduced from the corresponding nucleotide sequences. By using the PHYRE server (Kelley and Sternberg, 2009) the protein sequences were aligned with known crystal structures of λ-light chains. Due to this alignment the generation of a three-dimensional (3D) model was possible. The analysis included the identification of amino acid residue positions which putatively form the connection to the constant region of the heavy chain according to Padlan et al. (1986) and the position of detected amino acid residue substitutions within the 3D-molecule. Graphical analyses and figure preparation were compiled with *vmd* (Humphrey et al., 1996). To demonstrate the accessible surface area, additional 1.4Å were added to the radii of van der Waals forces (Lee and Richards, 1971).

2.7. Statistical analyses

Frequencies of individual IGLC and IGLV genes as well as combinations of both within one breed and among the breeds have been compared by applying non-parametric tests. Such test procedures, i.e. the Chi^2-test for analyzing IGLC and IGLV independently and Fisher's exact test for the gene segment x breed contingency table, are implemented in the software package SAS, Version 9.2.

The association between the count variable usage frequency of IGLV and the effects of transcriptional orientation including either all classified IGLV or only the putative functional IGLV was analyzed separately by breed as well as by combining both breeds via mixed model theory. The association study was carried out by defining the usage frequency of IGLV as a dependent variable as well as by modelling transcriptional

orientation and functionality as fixed effects. Application of linear mixed models considering both fixed and random effects assumes a Gaussian distribution for the variable to be analyzed, whereas the count variable usage frequency of IGLV follows a Poisson-like distribution. Hence, as an extension of linear mixed models a generalized linear mixed model (GLMM), that can be used to analyze data with a Poisson-like distribution, was used for this specific part of the statistical analysis. The main feature of a GLMM is a link function that allows the mean of a population to depend on a linear predictor. The link function f_i between the linear predictor h_i and the observations y_i used for these type of count data was a log link defined as $f_i = \log_e(h_i)$. The statistical model was defined as follows:

$$\log\left[y_{rst}\right] = \eta_r = \varphi + \gamma_s + \delta_t$$

where y_{rst} = value for usage frequency of IGLV r; φ = overall mean effect; γ_s = fixed effect of transcriptional orientation s; δ_t = fixed effect of functionality t.

3. Results

3.1. Amplification of germline IGLJ-IGLC pairs

To examine the presence of all predicted germline IGLJ-IGLC pairs, specific primers were generated. For functional IGLJ-IGLC pairs, primers based on the sequences of the transcribed λ-light chain repertoire (IGLJ1-IGLC1, IGLJ4-IGLC4, IGLJ5-IGLC5, IGLJ6/7-IGLC6/7). The genomic contig NW_001867428.1 was used for primer generation of the nonfunctional IGLJ2-IGLC2 and IGLJ3-IGLC3 pairs. Nomenclature of the IGLC-genes refers to Sun et al. (2010). DNA of two RGC animals were not available. All six IGLJ-IGLC pairs were amplified in two RGC animals and one HW animal. IGLJ1-IGLC1 was demonstrated in all RGC animals as well as in eight animals of the breed HW. IGLJ2-IGLC2 was proven in all animals analyzed in RGC, whereas in HW eight animals showed the presence of this IGLJ-IGLC pair. IGLJ3-IGLC3 was amplified in five animals of the breed RGC and in six HW animals, respectively. IGLJ4-IGLC4, which seemed to be nonfunctional due to a single C-insertion within the IGLJ4 exon (Sun et al., 2010), was demonstrated in eight and seven animals of RGC and HW. In four RGC as well as in five HW animals IGLJ5-IGLC5 was amplified. IGLJ6/7-IGLC6/7 was amplified in all RGC and HW animals analyzed. IGLJ6 and IGLJ7 as well as IGLC6 and IGLC7 were not distinguished, because primer hybridization sites were identical due to the high sequence similarities.

3.2. Transcriptional use of IGLC-genes

A total of 12 sequences per animal were analyzed to establish potential differences in usage frequencies of IGLC-genes in the breeds RGC and HW. IGLC were differentiated by specific amino acid residue motifs. The analysis of 120 sequences per breed revealed that IGLC-genes were transcribed in different frequencies (Fig. 1a). The isotype IGLC6/7 was predominantly used in both horse breeds examined and was identified in 55% (n=66) of the 120 sequences analyzed in each breed. IGLC5 had a lower frequency in RGC (n=25, 20.8%) than in HW (n=42, 35.0%). Twenty-seven sequences (22.5%) of RGC were classified to IGLC1, whereas in HW 12 sequences (10%) were identified as IGLC1. IGLC4 was transcribed in two RGC animals (1.7%), but not in the breed HW. In the breed RGC highly significant differences (P<0.001) in the transcription were confirmed by the Chi² test between IGLC1 and IGLC4, IGLC1 and IGLC6/7, IGLC4 and IGLC5, IGLC4 and IGLC6/7, as well as IGLC5 and IGLC6/7. The same P-values were obtained for the statistical comparison of IGLC1 and IGLC5,

as well as IGLC1 and IGLC6/7 in HW. With *P*≤0.05 IGLC5 and IGLC6/7 were significantly different in this breed (Fig. 1a). Statistical analysis did not show significant differences of IGLC usage between the breeds.

In most of the animals more than one IGLC-gene was transcribed (Fig. 1b) but without a significant preference of a specific combination. The isotype combinations IGLC1/IGLC5/IGLC6/7 (four animals) as well as IGLC1/IGLC6/7 (three animals) were preferred in RGC. The combinations of IGLC1/IGLC4/IGLC5 and IGLC1/IGLC4/IGLC5/IGLC6/7 were exhibited by one RGC animal. Exclusive transcription of IGLC6/7 was detected in four animals of the breed HW but in only one RGC animal. Other IGLC-genes transcribed simultaneously were detected in two HW animals (IGLC5/IGLC6/7, IGLC1/IGLC5, IGLC1/IGLC5/IGLC6/7; Fig 1b).

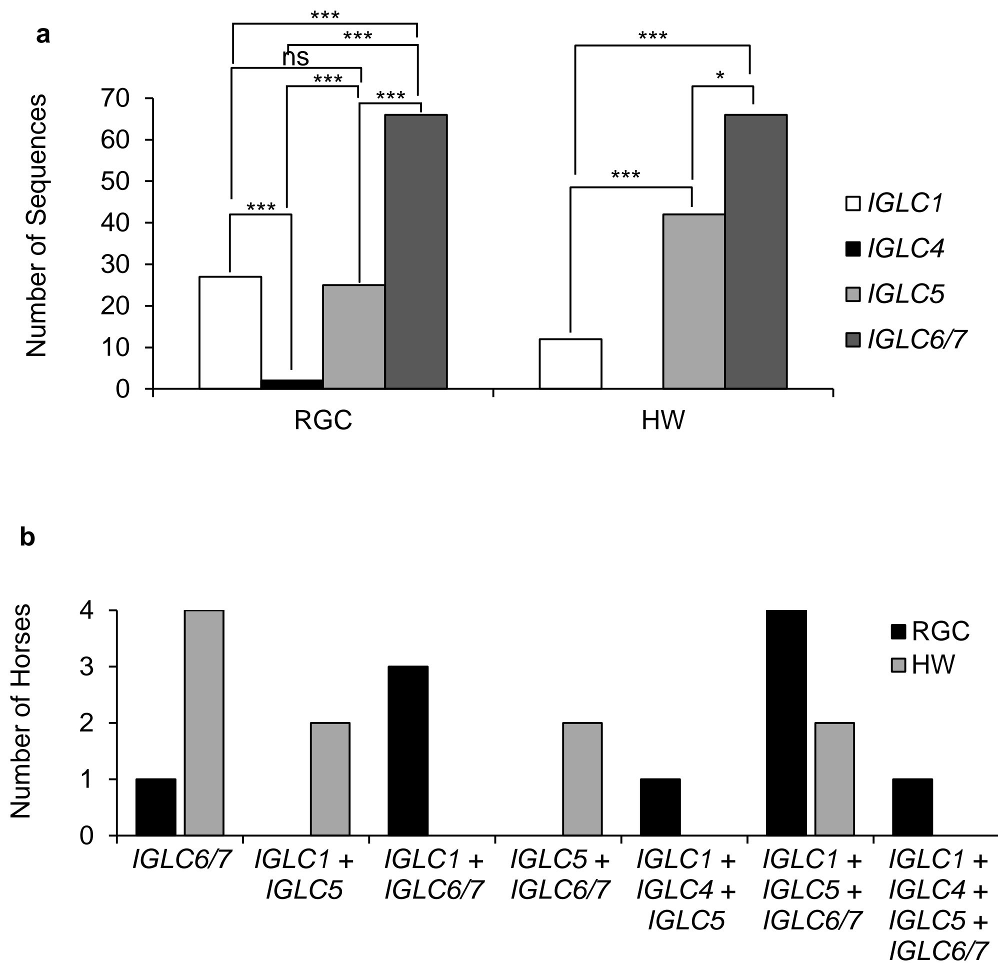


Fig. 1 Transcription of IGLC-genes in the horse breeds Rhenish-German Coldblood (RGC) and Hanoverian Warmblood (HW). **a** Usage frequency of IGLC-genes in RGC and HW. Significant differences between transcriptional frequency of the four IGLC are indicated by asterisks (*P≤0.05, **P≤0.01, ***P≤0.001). **b** Concurrently transcribed IGLC-genes in RGC- and HW animals. IGLC1, IGLC4, IGLC5, and IGLC6/7 describe immunoglobulin λ-light chain constant regions that were transcribed in this study. Nomenclature of the IGLC-genes corresponds to a previous study (Sun et al., 2010).

3.3. Transcriptional use of IGLV-segments

In analogy to the IGLC-genes, the 120 sequences available per breed were analyzed for transcriptional use of IGLV-segments. With a threshold of at least 80% sequence identity, all sequences were classified to 11 of 144 genomic IGLV-segments. Those 11 germline IGLV-segments were assigned to subgroup 8 of 11 known IGLV subgroups as previously described (Sun et al., 2010). All eleven IGLV-segments found were transcribed in RGC, whereas nine IGLV-segments were transcribed in HW (Fig. 2).

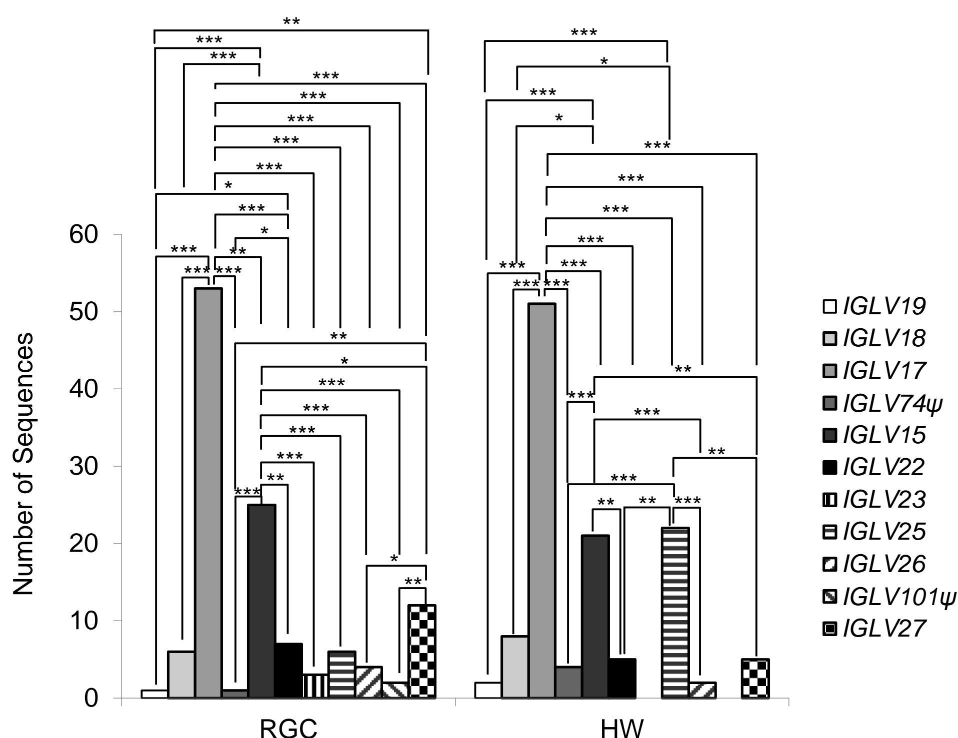


Fig. 2 Usage frequency of IGLV in Rhenish-German Coldblood (RGC) and Hanoverian Warmblood (HW) resulting from transcription analysis. IGLV describe the numbers of immunoglobulin λ-light chain variable gene segments as designated in a previous study (Sun et al., 2010). Significant differences between transcriptional frequency of the IGLV are indicated by asterisks (*P≤0.05, **P≤0.01, ***P≤0.001). For IGLV25, a significant difference was calculated between the two breeds (P=0.0233).

In both breeds examined, IGLV17 was rearranged preferably (RGC: n=53, 44.2%; HW: n=51, 42.5%). Similarly, IGLV15 was used with high frequencies in RGC (n=25, 20.8%) as well as in HW (n=21, 17.5%). In both breeds the Chi^2 statistics revealed significant differences of the transcriptional use of IGLV15 and IGLV17 in comparison to the other transcribed IGLVs. P-values ≤0.001 were obtained for the transcription of IGLV17 compared to IGLV19, IGLV18, IGLV74ψ, IGLV22, IGLV25, IGLV26, and IGLV27 within the both breeds examined. Comparing IGLV17 and IGLV15, we received a P-value ≤0.01 for RGC and ≤0.001 for HW. Very high significances (P≤0.001) were observed for the use IGLV15 in comparison to IGLV19, IGLV74ψ, and IGLV26 in RGC and HW animals. The same P-values were calculated for IGLV15 and IGLV23 as well as for IGLV15 and IGLV101ψ in RGC. Differences in the significance level were obtained between IGLV15 and IGLV18 (P_{HW}=0.0004, P_{RGC}=0.0015) and IGLV15 and IGLV27 (P_{HW}=0.0017, P_{RGC}=0.0326). Beside these similarities in transcriptional use of IGLV15 and IGLV17, both breeds differed in usage of further

IGLV-segments. While HW preferred IGLV25 as a third major IGLV-segment (*n*=22, 18.3%), RGC favored IGLV27 (*n*=12, 10%) as verified by Chi². In HW significant differences ($P≤0.001$) in transcription frequency were calculated for IGLV25 compared to IGLV19, IGLV74ψ, and IGLV26. Comparing IGLV25-IGLV18 ($P≤0.05$), IGLV22, and IGLV27 ($P≤0.01$) additional significant differences were calculated. In RGC IGLV27 revealed significant differences in transcriptional use to IGLV19, IGLV74ψ, IGLV101ψ ($P≤0.01$), IGLV23, and IGLV26 ($P≤0.05$). Statistical analysis further revealed significant differences in the expression of IGLV25 between the breeds (P=0.0233).

Moreover, it is noteworthy that two pseudogenes (IGLV101ψ and IGLV74ψ) were completely transcribed. For IGLV74ψ we identified nucleotide variations in CDR1 sequences that started at nucleotide position 96 of IGLV and neutralized the stop codons. These nucleotide variations within CDR1 resulted in a frame shift by what the deduced amino acid sequence changed and replaced the first stop codon by A, while S, K, I, or R substituted the second stop codon.

In order to describe the impact of the transcriptional orientation on the usage of IGLV-segments two assumptions were examined. The first calculation was based on all classified IGLV independently of their functionality. The second analysis included only the putative functional segments. Both statistical analysis showed that the *F*-ratios provided in the analysis of variance are identical to the Wald/rank(K) F-statistics as defined by Littell et al. (2004). Those F-statistics have been used to identify the impact of fixed effects on transcriptional orientation and functionality. For both breeds, RGC and HW, the transcriptional orientation has no influence on the transcription frequency of the IGLV-segments.

3.4. Rearrangement of IGLV- and IGLC-genes

In order to investigate the λ-light chain repertoire in the breeds RGC and HW, we analyzed rearrangement frequencies of the transcribed IGLV- and IGLC-genes. In addition, we used non-parametric tests for the analysis of significant differences within the rearrangements in the breeds as well as between the two breeds. The most frequently transcribed IGLC6/7 was rearranged with all of the transcribed IGLV-segments. In RGC, IGLV19, IGLV18, IGLV74ψ, and IGLV101ψ were only combined with IGLC6/7 (Fig. 3a), while IGLV26 was combined solely with IGLC6/7 in HW (Fig. 3b).

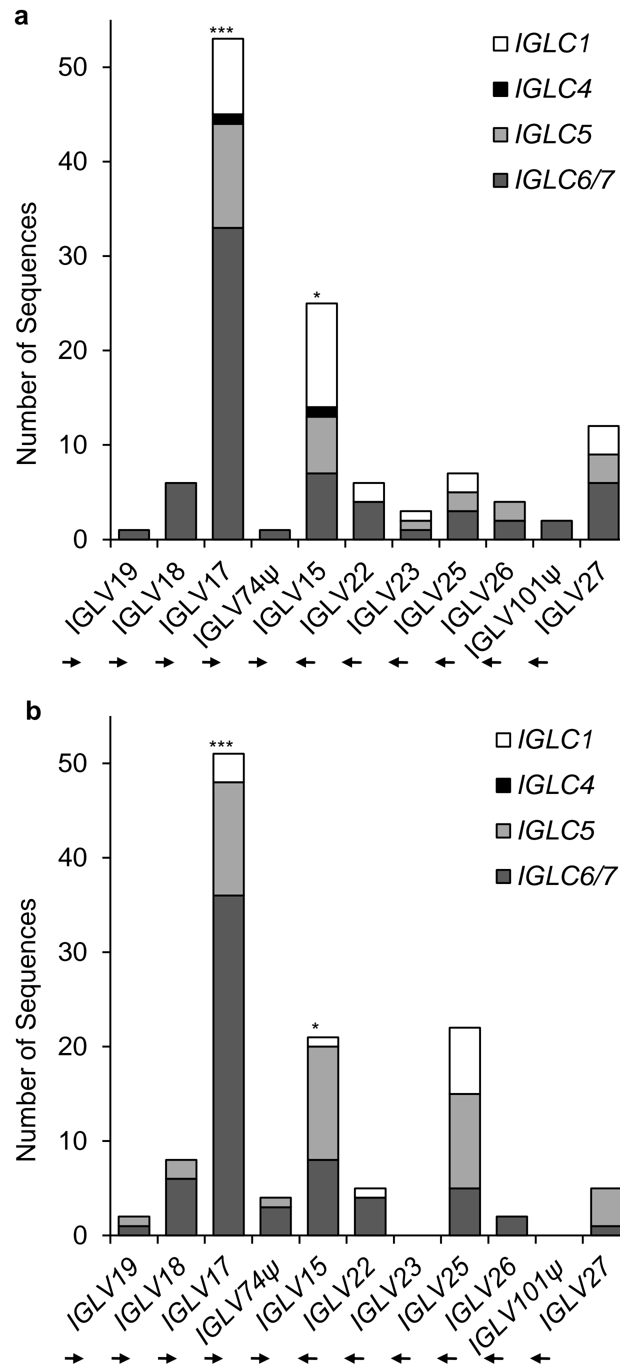


Fig. 3 Frequency of rearrangement of IGLV and IGLC genes in Rhenish-German Coldblood (RGC; **a**) and Hanoverian Warmblood (HW; **b**). IGLC1, IGLC4, IGLC5, and IGLC6/7 describe immunoglobulin λ-light chain constant regions that were transcribed in this study. IGLV describe the numbers of immunoglobulin λ-light chain variable gene segments as designated in a previous study (Sun et al., 2010). Ψ indicates completely transcribed pseudogenes. Arrows below annotation of the horizontal axis designate transcriptional orientation of IGLV. Arrows pointing to the right hand side indicate the same transcriptional orientation as germline IGLJ-IGLC cluster. Asterisks show significant differences in IGLC usage within this IGLV (*$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$). For IGLV15 there is a significant difference between the breeds (P=0.011).

Furthermore, in RGC IGLV22 and IGLV26 were rearranged either with IGLC1/IGLC6/7 or IGLC5/IGLC6/7, respectively. The IGLV-segments IGLV23, IGLV25, and IGLV27 were rearranged with IGLC1, IGLC5, and IGLC6/7, whereas in RGC IGLV15 and IGLV17 were rearranged with all four transcribed IGLC-genes, wherein in both IGLV IGLC4 was rearranged once (Fig. 3a). 44.0% of IGLV15 sequences in RGC were rearranged with IGLC1, 24.0% were joined to IGLC5, and 28.0% were joined to IGLC6/7. In HW IGLC5 was used preferentially (57.1%) for recombination with IGLV15. IGLC6/7 (38.1%) and IGLC1 (4.8%) were recombined less frequently with IGLV15. Analyses of the sequence numbers showed significant differences (P=0.011) of rearrangements of IGLV15 with the IGLC-genes between both breeds examined. These analyses also revealed significant differences of IGLC rearrangements within the breeds (P_{RGC}=0.0436; P_{HW}=0.0119). IGLV17 preferred recombination with IGLC6/7 (RGC: 62.3%; HW: 70.6%) followed by IGLC5 (RGC: 20.8%; HW: 23.5%) and IGLC1 (RGC: 15.1%; HW: 5.9%). For IGLV17 statistical analyses showed highly significant differences of IGLC rearrangements within the two breeds examined ($P_{RGC}<0.0001$; $P_{HW}<0.0001$).

In comparison to RGC, in HW IGLV19, IGLV18, IGLV74ψ, IGLV22, and IGLV27 were rearranged with either IGLC1/IGLC6/7 or IGLC5/IGLC6/7, whereas IGLV17, IGLV15, and IGLV25 were combined with IGLC1, IGLC5, and IGLC6/7 (Fig. 3b).

3.5. Allelic variants of IGLC1

In the breeds analyzed, single nucleotide polymorphisms (SNPs) of IGLC1 were detected (Table 2, Fig. 4a). The analyses of the deduced amino acid sequences revealed five allotypic variants denominated as IGLC1^{a}, IGLC1^{b}, IGLC1^{c}, IGLC1^{d}, and IGLC1^{e}. Germline IGLC1 was set as allotype a. The corresponding sequence can be found in the contig NW_001867428.1 (nucleotides 4144483-4144802) of the equine genome assembly EquCab2.0. SNPs were compared to the positions within this exon. Deduced amino acid residue substitutions were referred to the IMGT nomenclature (Lefranc et al., 2005; Table 2, Fig. 4a). The homology-based modeling showed the closest model for all IGLC1 allotypic variants to the crystal structures of either Fab KOL or Fab NEW derived from human mAbs (IGLC1$^{a,\ e}$ PDB code 2FB4, IGLC1$^{b\text{-}d}$ PDB code 1AQK; Faber et al., 1998; Kratzin et al., 1989). Identities of at least 63% were calculated.

Table 2
Allotypic variants of IGLC1. Nucleotide substitutions of allotypes are shown as SNP at corresponding nucleotide positions of IGLC. Resulting amino acid residue substitutions are shown as follows: initial amino acid residue, position, and replacing amino acid residue. Numbering of amino acid residue positions occurred in accordance with the IMGT numbering system for constant domains (Lefranc et al., 2005).

Isotype	Allotype/ allele	Position: SNP[a]	Position amino acid substitution
IGLC1	b	7: C > A	P 1.3 T
		10: A > C	T1.2 P
		146: T > C	I 45.3 T
		149: G > C	S 45.4 T
		156: A > C	-
		189: T > C	-
		201: G > A	-
		204: T > C	-
		217: A > C	T 90 S
		229: G > A	A 94 S
		230: C > G	
		316: T > C	S 127 P
	c	10: A > C	T 1.2 L
		11: C > T	
		16: G > C	A 1 P
		62: C > T	A 16 V
		146: T > C	I 45.3 T
		149: G > C	S 45.4 T
		156: A > C	-
		170: A > G	K 82 R
		189: T > C	-
		201: G > A	-
		204: T > C	-
		217: A > C	T 90 P
		229: G > A	A 94 T
		316: T > C	S 127 P
	d	16: G > A	A 1 T
		146: T > C	I 45.3 T
		149: G > C	S 45.4 T
		156: A > C	-
		189: T > C	-
		201: G > A	-
		204: T > C	-
		217: A > C	T 90 S
		229: G > A	A 94 S
		230: C > G	
		316: T > C	S 127 P
	e	149: G > C	S 45.4 T
		170: A > G	K 82 R
		217: A > C	T 90 P
		229: G > A	A 94 T
		316: T > C	S 127 P

[a] Single nucleotide polymorphism.

The allotypic variant IGLC1[b] differed from IGLC1[a] in two conservative as well as four non-conservative amino acid residue substitutions represented by P1.3T, T1.2P, S45.4T, T90S, A94S, and S127P which are located at the solvent accessible surface (Table 2, Fig. 4b). A fifth non-conservative amino acid residue substitution namely I45.3T was located inside the molecule. The adjoining residues 1.3 and 1.2 are located at the linking area to IGLV. At position 1.3 the hydrophobic P was replaced by the hydrophilic T. Due to the reverse substitution at position 1.2 the polarity in this molecule part was converted. At position 94 the replacement of A by S also led to a polar side chain. Residue 90 was adjacent to the putative interface to CH1, whereas positions 45.4 and 94 were located at the outer side of the molecule as determined by homology-based 3D modeling (Fig. 4b). Another non-conservative amino acid residue substitution represented by S127P was located at the C-terminal end of IGLC1, leading to the replacement of the neutral and hydrophilic amino acid residue by the stabilizing heterocyclic and hydrophobic P. Altogether 12 sequences of five RGC horses were assigned to allotypic variant IGLC1[b] (JN228100). Three of these five animals transcribed this allotype solely, two were heterozygous with IGLC1[d]. In HW amino acid residue substitutions regarding IGLC1[b] were not observed (Table 3).

a

```
                                   ____________________________ A ______________________________   ___ AB ___
                                   G   G   P   T   S   A   P   S   V   S   L   F   P   P   S   S   E   E   L   S
                                   .5  .4  .3  .2  .1  1   2   3   4   5   6   7   8   9   10  11  12  13  14  15  .1  .2  .3
NW_001867428_IGLC1[a]  4144802 (G)GT GGT CCC ACG TCT GCA CCC TCG GTC TCT CTC TTC CCG CCC TCC TCT GAG GAG CTC AGC --- --- ---
IGLC1[b]                     1 . .. ... A.. C.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ---
IGLC1[c]                     1 . .. ... ... CT. ... C.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ---
IGLC1[d]                     1 . .. ... ... ... ... A.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ---
IGLC1[e]                     1 . .. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ---

                                ______________________ B _______________________  _______________ BC ___________________
                                A   N   K   A   T   V   V   C   L   I   S   D   F   S   P           S   D   L   T
                                16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  34  35  36  37  38
NW_001867428_IGLC1[a]  4144743  GCC AAC AAG GCC ACA GTG GTG TGT CTC ATC AGT GAC TTC TCC CCC --- --- AGC GAC TTG ACG
IGLC1[b]                    61  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ... ... ... ...
IGLC1[c]                    61  .T. ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ... ... ... ...
IGLC1[d]                    61  ..G ... ... ... ... ... ... C.. ... ... ... ... ... ... ... --- --- ... ... ... ...
IGLC1[e]                    61  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ... ... ... ...

                                ___________ C ____________  __________ CD __________
                                V   S   W   K   V   N   G   A   A   I   S   Q
                                39  40  41  42  43  44  45  .1  .2  .3  .4  .5  .6  .7
NW_001867428_IGLC1[a]  4144686  GTG AGC TGG AAG GTA AAT GGC GCC GCC ATC AGC CAG --- ---
IGLC1[b]                   118  ... ... ... ... ... ... ... ... ... .C. .C. ... --- ---
IGLC1[c]                   118  ... ... ... ... ... ... ... ... ... .C. .C. ... --- ---
IGLC1[d]                   118  ... ... ... ... ... ... ... ... ... .C. .C. ... --- ---
IGLC1[e]                   118  ... ... ... ... ... ... ... ... ... ... .C. ... --- ---

                                _______________ D ________  ________________________ DE _____________________________
                                G   V   Q   T   T   K   P   S   K   Q   S   N                           G   K   Y   A
                                77  78  79  80  81  82  83  84  .1  .2  .3  .4  .5  .6  .7  .7  .6  .5  .4  .3  .2  .1
NW_001867428_IGLC1a    4144650  GGA GTC CAG ACC ACC AAG CCC TCG AAA CAG AGC AAT --- --- --- --- --- --- GGC AAG TAC GCG
IGLC1b                     154  ..C ... ... ... ... ... ... ... ... ... ... ..C --- --- --- --- --- --- ... ... ... ..A
IGLC1c                     154  ..C ... ... ... ... .G. ... ... ... ... ... ..C --- --- --- --- --- --- ... ... ... ...
IGLC1d                     154  ..C ... ... ... ... ... ... ... ... ... ... ..C --- --- --- --- --- --- ... ... ... ..A
IGLC1e                     154  ..C ... ... ... ... ... ... ... ... ... ... ..C --- --- --- --- --- --- ... ... ... ...

                                ______________________ E ____________________  _ EF __
                                A   S   S   Y   L   T   L   T   P   A   Q   W
                                85  86  87  88  89  90  91  92  93  94  95  96  .1  .2
NW_001867428_IGLC1[a]  4144602  GCT AGC AGC TAC CTG ACG CTG ACC CCC GCC CAG TGG --- ---
IGLC1[b]                   202  ..C ... ... ... ... T.. ... ... ... AG. ... ... --- ---
IGLC1[c]                   202  ..C ... ... ... ... C.. ... ... ... A.. ... ... --- ---
IGLC1[d]                   202  ..C ... ... ... ... T.. ... ... ... AG. ... ... --- ---
IGLC1[e]                   202  ..C ... ... ... ... C.. ... ... ... A.. ... ... --- ---

                                _____________ F ______________  _________________________ FG ________________________
                                K   S   S   S   S   V   S   C   Q   V   T   H   Q                   G   K   T   V
                                97  98  99  100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117
NW_001867428_IGLC1[a]  4144566  AAA TCG TCC AGC AGC GTC AGC TGC CAG GTC ACG CAC CAA --- --- --- --- GGG AAA ACC GTG
IGLC1[b]                   238  ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- ... ... ... ...
IGLC1[c]                   238  ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- ... ... ... ...
IGLC1[d]                   238  ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- ... ... ... ...
IGLC1[e]                   238  ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- ... ... ... ...

                                _______________ G _______________
                                E   K   K   L   S   P   S   E   C   S   *
                                118 119 120 121 122 123 124 125 126 127
NW_001867428_IGLC1[a]  4144515  GAG AAG AAA CTG TCC CCC TCA GAG TGT TCT TAG
IGLC1[b]                   289  ... ... ... ... ... ... ... ... ... C.. ...
IGLC1[c]                   289  ... ... ... ... ... ... ... ... ... C.. ...
IGLC1[d]                   289  ... ... ... ... ... ... ... ... ... C.. ...
IGLC1[e]                   289  ... ... ... ... ... ... ... ... ... C.. ...
```

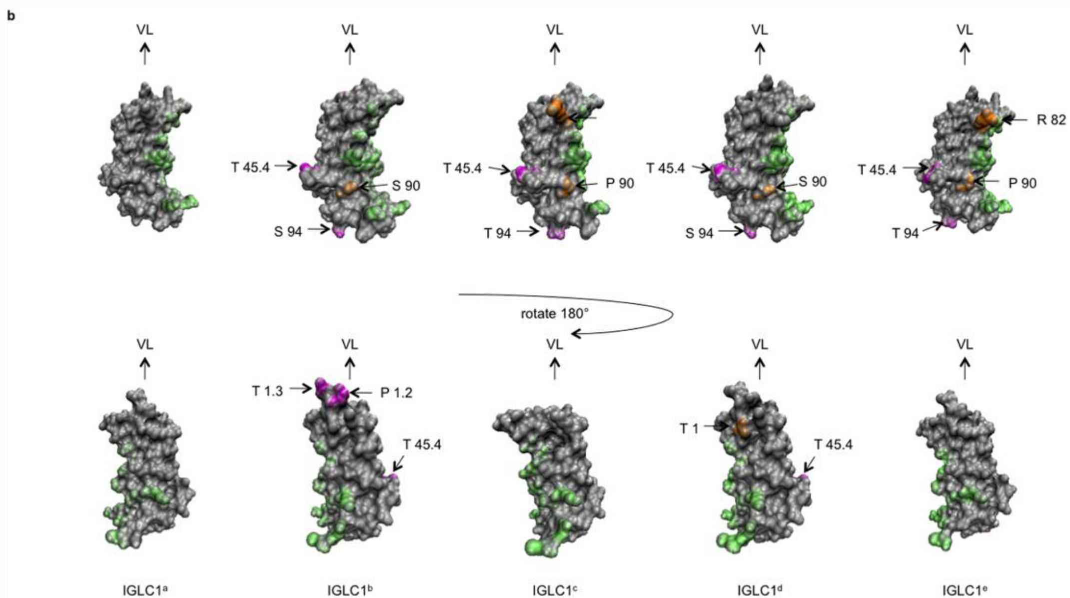


Fig. 4 Comparison of IGLC1 and its allotypic variants. **a** Alignment of isotype IGLC1 by ClustalW (Thompson et al., 1994) using detected allotypes. Sequences were numbered according to the IMGT numbering system for constant domains (Lefranc et al., 2005). A, B, C, D, E, F, and G correspond to the sandwich fold beta strands, whereas AB, BC, CD, DE, EF, and FG classify the turns and loops of constant domain immunoglobulin structure. Dots indicate identical nucleotides and dashes indicate gaps. The asterisk marks the stop codon. Amino acid residues providing the contact to CH1 on the interface between IGHC1 and IGLC in the crystal structure of human mAb NEW are indicated by bold letters and in the crystal structure of human mAb KOL they are indicated by italic letters (Padlan et al., 1986). Bold and italic letters demonstrate interface residues in both of the mAbs. **b** Homology-based modeling of IGLC isotype IGLC1 and its allotypes. For the predicted 3D structures solvent accessible surfaces are shown from two opposite sides. The images at the bottom are turned by 180° over the vertical axis compared to those at the top. The putative interface to CH1 is green. Locations of amino acid residue differences to the basic allotype IGLC1^{a} within the interface are marked in orange, whereas the remaining substitutions are marked in magenta. VL indicates the location of the corresponding variable region. Description of the amino acid residues followed the IMGT nomenclature (Lefranc et al., 2005).

IGLC1^{c} showed nine amino acid residue substitutions in comparison to IGLC1^{a} (Table 2, Fig. 4b). Seven of them occurred in residues exposed to the surface whereas only four were shown in figure 4a due to sequence predictions of the homology-based PDB-file. Amino acid residue substitutions at positions 16 and 45.3 were placed inside the molecule. While amino acid residues T45.4 and T94 were adjacent to the putative interface to CH1, R82 and P90 were located at the outer side of the molecule. The amino acid residue substitutions at positions 45.4 and 82 were conservative, while at position 90 the hydrophilic T was replaced by the hydrophobic P and the hydrophilic T was incorporated instead of hydrophobic A at position 94. There was no change of charge beside the change of hydrophobicity. Allotype IGLC1^{c} (JN228101) was detected in one HW animal as well as in one RGC animal that were heterozygous and also exhibited IGLC1^{e} (Table 3).

The fourth allotypic variant IGLC1^{d} also revealed two conservative amino acid residue substitutions resulting in T45.4 and S90, whereas the hydrophilic S at position 94 replaced the hydrophobic A. These positions were located at the surface of the molecule. IGLC1^{d} also differed from IGLC1^{a} by T45.3. An additional amino acid residue substitution enabling the differentiation of IGLC1^{d} from the other four allotypic variants of IGLC1 was located at amino acid residue position 1. Here, A was replaced by T exhibiting a polar side chain (Table 2, Fig. 4a, b). IGLC1^{d} (JN228102) was found in three animals of the breed RGC but was not detected in HW. In two animals this allotype was transcribed heterozygously together with IGLC1^{b}, while it was transcribed homozygously in one animal (Table 3).

Table 3
Transcriptional occurrence and frequencies of the allotypic variants in the horse breeds Rhenish-German Coldblood (RGC) and Hanoverian Warmblood (HW). Homozygous and heterozygous transcription is indicated by the combination of allotypic variants. Only transcribed allotypic variants are shown in the table.

	Allotypes/ Alleles		Horse breeds	
			RGC	HW
IGLC1	b/b		3	0
	b/d		2	0
	c/e		1	1
	d/d		1	0
	e/e		2	3
		∑	9	4
Frequencies	b		0.44	0
	c		0.06	0.13
	d		0.22	0
	e		0.28	0.87
		∑	1	1
IGLC5	a/a		2	5
	a/b		0	1
	a/c		1	0
	a/d		1	0
	d/d		2	0
		∑	6	6
Frequencies	a		0.5	0.92
	b		0	0.08
	c		0.08	0
	d		0.42	0
		∑	1	1
IGLC6/7	a1/a1		1	0
	a1/a2		3	7
	a1/a3		1	0
	a2/a2		0	1
	a2/a3		1	0
	a3/a3		1	0
	a2/b		1	0
	a3/b		1	0
		∑	9	8
Frequencies	a1		0.33	0.44
	a2		0.28	0.56
	a3		0.28	0
	b		0.11	0
		∑	1	1

IGLC1^{e} revealed five amino acid residue substitutions compared to IGLC1^{a} (S45.5T, K82R, T90P, A94T, S127P; Table 2, Fig. 4a). In accordance to IGLC1^{b-d}, IGLC1^{e} differs in S127P IGLC1^{a}. This stabilizing amino acid substitution is located at the C-terminal end of IGLC1 and also had an influence on hydrophobicity. In HW, three horses were homozygous for IGLC1^{e} as well as two animals of RGC (RGC: JN228103, HW: JN228104). In each breed, one animal showed heterozygous transcription together with IGLC1^{c} (Table 3).

In silico analyses of the ESTs database gave evidence to one nucleotide sequence coding for IGLC1^{d} in a Dartmore Pony (CD470899). For further allotypes no entries were observed matching 100% sequence identity and coverage. Alleles IGLC1^{a}, IGLC1^{b}, IGLC1^{c}, IGLC1^{d}, and IGLC1^{e} were at least 96.0% identical to each other (Online Resource 1).

3.6. Allelic variants of IGLC5

For *IGLC5*, four allotypic variants were identified and were denominated as IGLC5^{a}, IGLC5^{b}, IGLC5^{c}, and IGLC5^{d}. The germline sequence of the equine genome assembly EquCab2.0 (NW_001867428, nucleotide no. 4127347-4127666) was set IGLC5^{a}. The allotypic variants were distinguished by the deduced amino acid residue substitutions resulting from 10 different SNPs (Table 4, Fig. 5a). For IGLC5, the deduced amino acid residues of the allotypic variants were aligned and 3D structures of the accessible surfaces were calculated (Fig. 5b). IGLC5^{a} shared the best homology with the PDB code 2FB4 (Kratzin et al., 1989). The closest model for IGLC5^{b-d} was PDB code 1AQK (Faber et al., 1998). Homology to selected PDB codes was at least 63%.

For IGLC5^{a} five of ten HW animals were homozygous (JN228106) and one animal also transcribed IGLC5^{b}. In RGC four of 10 animals transcribed this allotype (JN228105). In one of these animals IGLC5^{a} was transcribed heterozygously together with IGLC5^{c}. A second RGC animal transcribed IGLC5^{d} (Table 3).

IGLC5^{b} differed from IGLC5^{a} in three amino acid residue substitutions (Table 4, Fig. 5b). A1P and K107T represented conservative amino acid residue substitutions concerning charge and hydrophobicity. In spite of that, the long amino acid side chain of K107 was replaced by the shorter side chain of T and a steric influence on the molecule has to be considered. The replacement T1.2L resulted in a neutral amino acid with hydrophobic side chain which was located at the opposite site of the interface to CH1. Allotype IGLC5^{b} (JN228107) was only identified in five sequences of one HW animal. In this animal IGLC5^{a} was also determined (Table 3).

a

```
                                       ______________________________ A ______________________________ ___ AB ___
                                        G   G   P   T   S   A   P   S   V   S   L   F   P   P   S   S   E   E   L   S
                                       .5  .4  .3  .2  .1   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  .1  .2  .3
NW_001867428_IGLC5a   4127666 (G)GT GGT CCC ACG TCT GCA CCC TCG GTC TCT CTC TTC CCG CCC TCC TCT GAG GAG CTC AGC --- --- ---
IGLC5b                      1 . .. ... ... CT. ... C.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ---
IGLC5c                      1 . .. ... ... ... ... C.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ---
IGLC5d                      1 A .. ... ... ... ... A.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ---

                                _____________________ B ___________________ ____________________ BC _________________
                                 A   N   K   A   T   V   V   C   L   I   S   D   F   S   P           S   G   L   E
                                16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  34  35  36  37  38
NW_001867428_IGLC5a   4127607   GCC AAC AAG GCC ACA GTG GTG TGT CTC ATC AGT GAC TTC TCC CCC --- --- AGC GGC TTG GAG
IGLC5b                     61   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ... ... ... ...
IGLC5c                     61   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ... ... ... ...
IGLC5d                     61   ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ... --- --- ... ... C.. ...

                                ___________ C _____________ ___________ CD ___________
                                 V   I   W   K   V   N   D   A   V   T   N   D
                                39  40  41  42  43  44  45  .1  .2  .3  .4  .5  .6  .7
NW_001867428_IGLC5a   4127550   GTG ATC TGG AAA GTA AAT GAC GCT GTC ACC AAC GAC --- ---
IGLC5b                    118   ... ... ... ..G ... ... ... ... ... ... ... ... --- ---
IGLC5c                    118   ... ... ... ... ... ... ... ... ... ... ... ... --- ---
IGLC5d                    118   ... ... ... ..G ... ... ... ... ... ... .C. ... --- ---

                                ______________ D ______________ ___________________________ DE ___________________________
                                 R   V   Q   T   T   R   P   S   K   Q   S   N                           G   K   Y   A
                                77  78  79  80  81  82  83  84  .1  .2  .3  .4  .5  .6  .7  .7  .6  .5  .4  .3  .2  .1
NW_001867428_IGLC5a   4127514   CGC GTC CAG ACC ACC AGG CCC TCG AAA CAG AGC AAC --- --- --- --- --- --- GGC AAG TAC GCG
IGLC5b                    154   ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- ... ... ... ...
IGLC5c                    154   ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- ... ... ... ...
IGLC5d                    154   G.. ... ... ... ... ... T.. ... ... ... ... ... --- --- --- --- --- --- ... ... ... ...

                                ______________________ E ______________________ _ EF __
                                 A   S   S   Y   L   T   R   T   S   T   E   W
                                85  86  87  88  89  90  91  92  93  94  95  96  .1  .2
NW_001867428_IGLC5a   4127466   GCC AGC AGC TAC CTG ACA CGG ACC TCC ACA GAG TGG --- ---
IGLC5b                    202   ... ... ... ... ... ... ... ... ... ... ... ... --- ---
IGLC5c                    202   ... ... ... ... ... ... ... ... ... ... ... ... --- ---
IGLC5d                    202   ... ... ... ... ... ... ... ... ... ... ... ... --- ---

                                ______________ F ______________ _____________________ FG ________________________
                                 K   S   Y   S   S   V   S   C   Q   V   K   H   Q                   G   K   T   V
                                97  98  99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117
NW_001867428_IGLC5a   4127430   AAA TCG TAC AGC AGC GTC AGC TGC CAG GTC AAG CAC CAA --- --- --- --- GGG AAA ACC GTG
IGLC5b                    238   ... ... ... ... ... ... ... ... ... ... .C. ... ... --- --- --- --- ... ... ... ...
IGLC5c                    238   ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- ... ... ... ...
IGLC5d                    238   ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- ... ... ... ...

                                __________________ G __________________
                                 E   K   K   L   S   P   S   E   C   P   *
                               118 119 120 121 122 123 124 125 126 127
NW_001867428_IGLC5a   4127379   GAG AAG AAA CTG TCC CCC TCA GAG TGT CCT TAG
IGLC5b                    289   ... ... ... ... ... ... ... ... ... ... ...
IGLC5c                    289   ... ... ... ... ... ... ... ... ... ... ...
IGLC5d                    289   ... ... ... ... ... ... ... ... ... ... ...
```

b

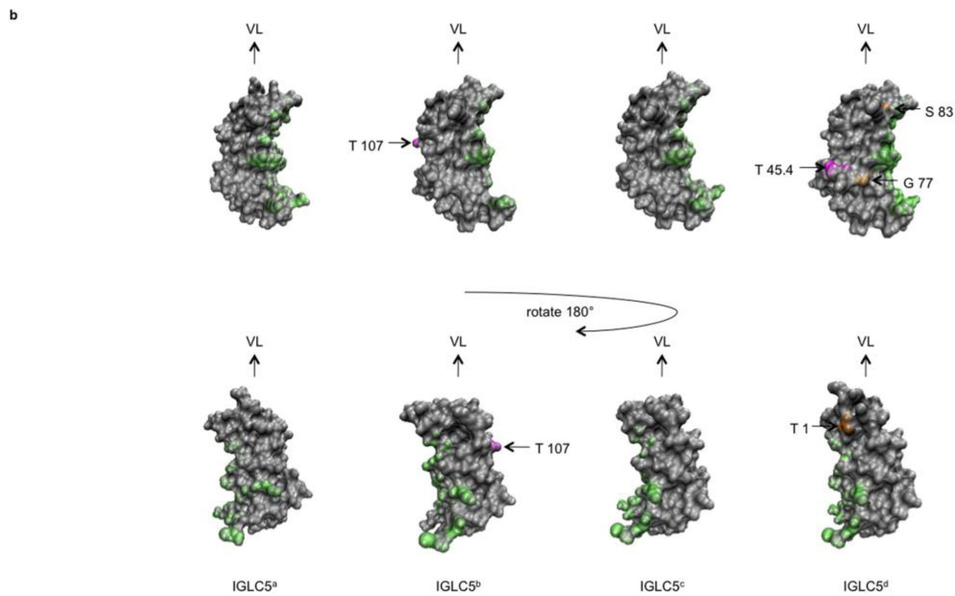


Fig. 5 Comparison of IGLC5 and its allotypic variants **a** Alignment of isotype IGLC5 by ClustalW (Thompson et al., 1994) using detected allotypes. Sequences were numbered according to the IMGT numbering system for constant domains (Lefranc et al., 2005). A, B, C, D, E, F, and G correspond to the sandwich fold beta strands, whereas AB, BC, CD, DE, EF, and FG classify the turns and loops of constant domain immunoglobulin structure. Dots indicate identical nucleotides and dashes indicate gaps. The asterisk marks the stop codon. Amino acid residues providing the contact to CH1 on the interface between IGHC1 and IGLC in the crystal structure of human mAb NEW are indicated by bold letters and in the crystal structure of human mAb KOL they are indicated by italic letters (Padlan et al., 1986). Bold and italic letters demonstrate interface residues in both of the mAbs. **b** Homology-based modeling of IGLC isotype IGLC5 and its allotypes. For the predicted 3D structures solvent accessible surfaces are shown from two opposite sides. The images at the bottom are turned by 180° over the vertical axis compared to those at the top. The putative interface to CH1 is green. Locations of amino acid residue differences to the basic allotype $IGLC5^{a}$ within the interface are marked in orange, whereas the remaining substitutions are marked in magenta. VL indicates the location of the corresponding variable region. Description of the amino acid residues followed the IMGT nomenclature (Lefranc et al., 2005).

$IGLC5^{c}$ (JN228108) revealed a single SNP located near the putative interface to CH1, which resulted in the conservative amino acid residue substitution A1P (Table 4, Fig. 5b). This substitution was solely observed in one RGC animal which also expressed $IGLC5^{a}$ (Table 3).

Three non-conservative amino acid residue substitutions at positions 1, 77, and 83 as well as one conservative amino acid residue substitution at position 45.4 characterized $IGLC5^{d}$ (Table 4, Fig. 5a). Due to the amino acid residue substitution P83S a new hydrophilic amino acid residue was exposed to the accessible surface. In addition, R77G converts the charge at this position into neutral as the hydrophobic G was incorporated instead of the hydrophilic R. Position 77 was adjacent to the putative interface to CH1. Amino acid residue substitutions at positions 45.4 and 107 were located at the opposite site to the interface to CH1. As $IGLC5^{c}$, allotypic variant $IGLC5^{d}$ (JN228109) was only found in one RGC animal that also transcribed $IGLC5^{a}$. $IGLC5^{d}$ was represented by one sequence in two further animals of RGC (Table 3).

For allotypic variants of IGLC5 no EST was observed with 100% identity and coverage. Alleles $IGLC5^{a}$, $IGLC5^{b}$, $IGLC5^{c}$, and $IGLC5^{d}$ showed at least 96.9% identity to each other (Online Resource 2).

Table 4
Allotypic variants of IGLC5. Nucleotide substitutions of further allotypes are shown as SNP at corresponding nucleotide positions of IGLC. Resulting amino acid residue substitutions are shown as follows: initial amino acid residue, position, and replacing amino acid residue. Numbering of amino acid residue positions occurred in accordance with the IMGT numbering system for constant domains (Lefranc et al., 2005).

Isotype	Allotype/ allele	Position: SNP[a]	Position amino acid substitution
IGLC5	b	10: A > C	T 1.2 L
		11: C > T	
		16: G > C	A 1 P
		129: A > G	-
		269: A > C	K 107 T
	c	16: G > C	A 1 P
	d	16: G > A	A 1 T
		112: T > C	-
		129: A > G	-
		149: A > C	N 45.4 T
		154: C > G	R 77 G
		172: C > T	P 83 S

[a] Single nucleotide polymorphism.

3.7. Allelic variants of IGLC6/7

The analyses of the predominantly expressed isotypes revealed two allotypic variants (IGLC6/7^{a}, IGLC6/7^{b}). For IGLC6/7^{a} three allelic variants were differentiated (IGLC6/7^{a1}, IGLC6/7^{a2}, IGLC6/7^{a3}; Table 5, Fig. 6a). The germline sequence of the equine genome assembly EquCab2.0 was defined as IGLC6/7^{a1}. The corresponding nucleotides are presented by positions 4120531-4120851 coding for IGLC7 in the contig NW_001867428. IGLC6 and IGLC7 were not distinguished from each other as their sequences differed only in nucleotide position 16 within their genes corresponding to positions 4124306 (G) and 4120837 (A) (NW_001867428). This SNP resulted in different motifs at amino acid positions 1.1 to 3 (SAPS for IGLC6, STPS for IGLC7). In all of our sequences analyzed nucleotide A and therefore amino acid residue motif STPS was determined similarly to germline IGLC7.

Table 5
Allelic and allotypic variants of IGLC6/7. Nucleotide substitutions of the variants are shown as SNP at corresponding nucleotide positions of IGLC. Resulting amino acid residue substitutions are shown as follows: initial amino acid residue, position, and replacing amino acid residue. Numbering of amino acid residue positions occurred in accordance with the IMGT numbering system for constant domains (Lefranc et al., 2005).

Isotype	Allotype/ allele	Position: SNP[a]	Position amino acid substitution
IGLC6/ IGLC7	a2	93: T > C	-
	a3	112: C > T	-
	b	32: T > C	L 6 P
		93: T > C	-
		167: C > T	T 81 I

[a] Single nucleotide polymorphism.

Alignments of the deduced amino acid sequences to the crystal structures of human mAb KOL and NEW derived Fabs enabled calculation of the accessible surfaces (Fig. 6b). For IGLC6/7$^{a, b}$ the closest model was PDB code 1AQK (Faber et al., 1998). Homology to this PDB code was at least 63%.
The three alleles were distinguished by SNPs at nucleotide positions 4120760 with T>C (IGLC6/7^{a2}) and 4120741 with C>T (IGLC6/7^{a3}) within the gene (NW_001867428, Table 5, Fig. 6a). Only one animal of the breeds analyzed showed homozygous transcription of IGLC6/7^{a1} (RGC: JN228110, HW: JN228111). IGLC6/7^{a2} was found in five RGC animals (JN228112) and in eight horses of HW (JN228113). One HW animal was homozygous for this allele. All remaining animals were heterozygous and transcribed additionally either IGLC6/7^{a1}, IGLC6/7^{a3}, or IGLC6/7^{b}. IGLC6/7^{a3} (JN228114) was transcribed in four animals of the breed RGC (Table 3).
In comparison to IGLC6/7^{a}, IGLC6/7^{b} is characterized by the two amino acid residue exchanges L6P and T81I. At position six a neutral and hydrophobic amino acid side chain was replaced by an aliphatic and hydrophobic side chain. The amino acid residue substitution resulting in a hydrophobic side chain at position 81 affected one of the conserved residues that were responsible for the interaction between CL and CH1 (Fig. 6b). In comparison to IGLC6/7^{a2} the same SNP like at nucleotide position 4,120,760 caused a silent mutation in IGLC6/7^{b}.
IGLC6/7^{b} was only detected in two animals of the breed RGC (JN228115). In these two animals either IGLC6/7^{a2} or IGLC6/7^{a3} were observed additionally. For all allotypic variants of IGLC6/7 no EST showed 100% identity and coverage. The four alleles of IGLC6/7 were at least 98.4% identical to each other (Online Resource 3).

a

```
                                             ___________________________________ A _________________________________ ___ AB ____
                                               G   G   P   T   S   T   P   S   V   S   L   F   P   P   S   S   E   E   L   S
                                              .5  .4  .3  .2  .1   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  .1  .2  .3
NW_001867428_IGLC7a1          4120850  (G)GT GGT CCC ACG TCT ACA CCC TCG GTC TCT CTC TTC CCG CCC TCC TCT GAG GAG CTC AGC --- --- ---
NW_001867428_IGLC6a1          4124320  (.).. ... ... ... ... G.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ---
IGLC6/7a2                           1     . .. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ---
IGLC6/7a3                           1     . .. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ---
IGLC6/7b                            1     . .. ... ... ... ... ... ... ... ... ... ... .C. ... ... ... ... ... ... ... ... --- --- ---

                                         ____________________ B _______________ _______________ BC ______________
                                          A   N   K   A   T   V   V   C   L   I   S   D   F   S   P           S   G   L   E
                                         16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  34  35  36  37  38
NW_001867428_IGLC7a1          4120791    GCC AAC AAG GCC ACA GTG GTG TGT CTC ATC AGT GAC TTC TCC CCC --- --- AGC GGC CTG GAG
NW_001867428_IGLC6a1          4124261    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ... ... ... ...
IGLC6/7a2                          61    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ... ... ... ...
IGLC6/7a3                          61    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- ... ... T.. ...
IGLC6/7b                           61    ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ... --- --- ... ... ... ...

                                         ___________ C __________ ___________ CD __________
                                          V   I   W   K   V   N   D   A   V   T   T   D
                                         39  40  41  42  43  44  45  .1  .2  .3  .4  .5  .6  .7
NW_001867428_IGLC7a1          4120734    GTG ATC TGG AAG GTA AAT GAC GCT GTC ACC ACC GAC --- ---
NW_001867428_IGLC6a1          4124204    ... ... ... ... ... ... ... ... ... ... ... ... --- ---
IGLC6/7a2                         118    ... ... ... ... ... ... ... ... ... ... ... ... --- ---
IGLC6/7a3                         118    ... ... ... ... ... ... ... ... ... ... ... ... --- ---
IGLC6/7b                          118    ... ... ... ... ... ... ... ... ... ... ... ... --- ---

                                         ______________ D ___________ ________________________ DE ______________________
                                          G   V   Q   T   T   R   S   S   K   Q   S   N                           G   K   Y   A
                                         77  78  79  80  81  82  83  84  .1  .2  .3  .4  .5  .6  .7  .7  .6  .5  .4  .3  .2  .1
NW_001867428_IGLC7a1          4120698    GGC GTC CAG ACC ACC AGG TCC TCG AAA CAG AGC AAC --- --- --- --- --- --- GGC AAG TAC GCG
NW_001867428_IGLC6a1          4124168    ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- ... ... ... ...
IGLC6/7a2                         154    ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- ... ... ... ...
IGLC6/7a3                         154    ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- --- --- ... ... ... ...
IGLC6/7b                          154    ... ... ... ... .T. ... ... ... ... ... ... ... --- --- --- --- --- --- ... ... ... ...

                                         _____________________ E ___________________ _ EF __
                                          A   S   S   Y   L   T   R   T   S   A   Q   W
                                         85  86  87  88  89  90  91  92  93  94  95  96  .1  .2
NW_001867428_IGLC7a1          4120650    GCC AGC AGC TAC CTG ACG CGG ACT TCC GCA CAG TGG --- ---
NW_001867428_IGLC6a1          4124120    ... ... ... ... ... ... ... ... ... ... ... ... --- ---
IGLC6/7a2                         202    ... ... ... ... ... ... ... ... ... ... ... ... --- ---
IGLC6/7a3                         202    ... ... ... ... ... ... ... ... ... ... ... ... --- ---
IGLC6/7b                          202    ... ... ... ... ... ... ... ... ... ... ... ... --- ---

                                         _____________ F _____________ _____________________ FG ________________________
                                          K   S   Y   S   S   V   S   C   Q   V   K   H   Q                   G   K   T   V
                                         97  98  99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117
NW_001867428_IGLC7a1          4120614    AAA TCG TAC AGC AGC GTC AGC TGC CAG GTC AAG CAC CAA --- --- --- --- GGG AAA ACC GTG
NW_001867428_IGLC6a1          4124084    ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- ... ... ... ...
IGLC6/7a2                         238    ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- ... ... ... ...
IGLC6/7a3                         238    ... C.. ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- ... ... ... ...
IGLC6/7b                          238    ... ... ... ... ... ... ... ... ... ... ... ... ... --- --- --- --- ... ... ... ...

                                         _________________ G _________________
                                          E   K   K   L   S   P   S   E   C   P   *
                                        118 119 120 121 122 123 124 125 126 127
NW_001867428_IGLC7a1          4120563    GAG AAG AAA CTG TCC CCC TCA GAG TGT CCT TAG
NW_001867428_IGLC6a1          4124033    ... ... ... ... ..- --- --- --- --- --- ---
IGLC6/7a2                         289    ... ... ... ... ... ... ... ... ... ... ...
IGLC6/7a3                         289    ... ... ... ... ... ... ... ... ... ... ...
IGLC6/7b                          289    ... ... ... ... ... ... ... ... ... ... ...
```

b

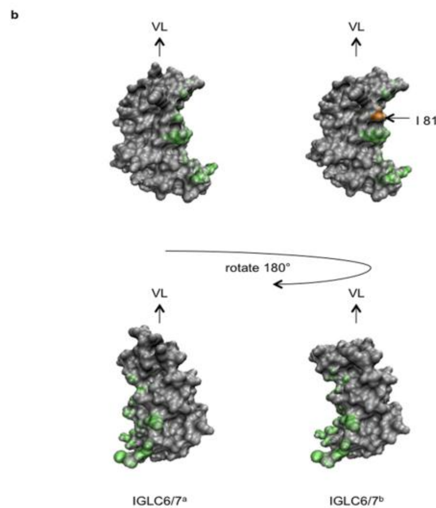


Fig. 6 Comparison of IGLC6/7 and its allotypic variants **a** Alignment of isotype IGLC6/7 by ClustalW (Thompson et al., 1994) using detected allotypes. Sequences were numbered according to the IMGT numbering system for constant domains (Lefranc et al., 2005). A, B, C, D, E, F, and G correspond to the sandwich fold beta strands, whereas AB, BC, CD, DE, EF, and FG classify the turns and loops of constant domain immunoglobulin structure. Dots indicate identical nucleotides and dashes indicate gaps. The asterisk marks the stop codon. Amino acid residues providing the contact to CH1 on the interface between IGHC1 and IGLC in the crystal structure of human mAb NEW are indicated by bold letters and in the crystal structure of human mAb KOL they are indicated by italic letters (Padlan et al. 1986). Bold and italic letters demonstrate interface residues in both of the mAbs. **b** Homology-based modeling of IGLC isotype IGLC6/7 and its allotypes. For the predicted 3D structures solvent accessible surfaces are shown from two opposite sides. The images at the bottom are turned by 180° over the vertical axis compared to those at the top. The putative interface to CH1 is green. Locations of amino acid residue differences to the basic allotype IGLC7^{a} within the interface are marked in orange, whereas the remaining substitutions are marked in magenta. VL indicates the location of the corresponding variable region. Description of the amino acid residues followed the IMGT nomenclature (Lefranc et al., 2005).

4. Discussion

Our study is the first detailed molecular genetic description of allotypic variants within the equine λ-light chain locus. The present analyses revealed the existence of five IGLC1, four IGLC5, and two IGLC6/7 allotypic variants. For IGLC6/7[a] three alleles were observed. The amino acid residue variations at positions 1.3, 1.2, 1, 45.3, 45.4, 81, 83, 90, 94, 95, and 107 were located at the accessible surface as confirmed by homology-based predicted 3D structural analysis. These modifications are assumed to possess distinct serological properties resulting in allotypic variants.
The transcriptional use of IGLC6/7 was significantly preferred in both breeds analyzed as confirmed by Chi^2 statistics. There were highly significant differences in the transcription frequency of the four functional IGLC isotypes within the two breeds. IGLC6 and IGLC7 were distinguished by only one amino acid residue substitution which was caused by one SNP. Due to the similarity of our corresponding sequences it was not obvious whether they were different alleles of IGLC6 or IGLC7 or allotypic variants of one gene. It is noteworthy that the sequences analyzed resembled IGLC7. All allotypic variants identified showed distinct amino acid residue substitutions in all seven β-sheets according to the IMGT nomenclature (Lefranc et al., 2005). Most of the substitutions were located within the accessible surface area (Padlan et al., 1986). Seven of these substitutions (A1T, R77G, T81I, 82R, P83S, T90S/P) were positioned within the interface between the constant region of the light chain and the first constant region of the heavy chain. These changes might influence the stable connection between both the heavy and light chains. Due to crystal structures of Fabs derived from human mAbs KOL and NEW, six amino acids (F7, E12, E13, T20, V22, and T81) were defined as conserved residues. They are expected for the main interaction of IGLC and IGHC1. The remaining interface residues may be variable (Padlan et al., 1986). Regarding these facts, T81I affected one of the conserved residues for the contact to IGHC1 leading to the conclusion, that the assembly of heavy and light chains could be changed. Otherwise we do not have any information about allelic changes in CH1 of the heavy chains and this conclusion remains to be verified. In previous studies IGHC1 was related to antibody affinity (Pritsch et al., 1996). IGLC and IGHC1 were found to have an influence on the whole immunoglobulin molecule due to time differences in the assembly of full size antibodies and in the kinetics of antigen binding (Montano and Morrison, 2002).

Amino acid residue substitutions located at the solvent accessible surface area might be distinguished by specific sera and therefore might be used as markers. Specific markers differing in frequencies within human populations were observed in λ- and κ-light chains as well as in heavy chains of IgG and IgA (Calderon et al., 2007; Matsumoto et al., 1984; Schanfield et al., 2008). Due to either direct association or linkage disequilibrium with the causative gene, allotypic markers of human immunoglobulin chains could be related to susceptibility of different infectious diseases (Giha et al., 2009; Granoff et al., 1984; Pandey, 2000; Pandey et al., 1995; Pandey et al., 1979). Influences of bovine allotypes on effector functions, complement activation, and age dependent expression have also been pointed out in previous studies (Bastida-Corcuera et al., 2000; Bastida-Corcuera et al., 1999a; Bastida-Corcuera et al., 1999b; Corbeil et al., 1997). As investigations on these issues are missing in horses, we suppose heterozygotic horses to be advantageous to the humoral immune response.

In addition to IGLC1, IGLC5, and IGLC6/7, a fourth IGLC-gene (IGLC4) was found in one sequence of two RGC horses. Both of them varied from each other as well as from the genomic sequence in the genome assembly EquCab2.0. Since our set criteria for the definition of a new allotype were not fulfilled, this isotype was not analyzed further. This constant region was described as potentially nonfunctional due to a single C-insertion within the corresponding joining gene segment IGLJ4. We observed transcription without the C-insertion within the IGLJ-segment as mentioned in previous studies (Sun et al., 2010).

In our investigation we also analyzed the presence of all predicted germline IGLJ-IGLC pairs. Each of the six distinguishable IGLJ-IGLC pairs was amplified from two RGC and one HW animal. IGLJ5-IGLC5 was detected in four animals of RGC and five animals of HW. In contrast, in previous investigations germline IGLJ5-IGLC5 as well as IGLJ7-IGLC7 were not amplified. In the same study germline IGLC6 and IGLC7 were distinguished by PCR-products amplified with primers that anneal within the genomic introns (Sun et al., 2010). As we used exon specific primer pairs to minimize genomic sequence variability and to avoid false-negative results we could not differentiate between IGLC6 and IGLC7.

Beside the presence of predicted germline IGLJ-IGLC pairs and the transcription frequency of IGLC-genes, we had a closer look at the usage of IGLV-segments of subgroup 8 because this subgroup contains the highest number of functional IGLV. In addition, Sun et al. (2010) showed this subgroup is transcribed preferentially. As there

are nine functional IGLV in subgroup 8, we sequenced 12 clones per animal to find subgroup 8-IGLV transcribed at low frequency in horses, too. Finally, there were two variable segments, IGLV15 and IGLV17, which were preferred in both breeds. For IGLV15, statistical analyses showed significant differences of the rearrangements with the four IGLC-isotypes transcribed. While rearrangement of IGLV17 was highly significant different from the other IGLV-segments in both breeds examined, rearrangement of IGLV15 showed *P*-values of 0.0119 in HW and 0.0436 in RGC. Between both breeds analyzed, only IGLV15 differed significantly in rearrangements with the IGLC-genes (*P*=0.011). In HW a third variable segment, IGLV25, was transcribed at high frequency. For this segment Chi^2 revealed significant differences of transcription frequency within both breeds examined (*P*=0.0233). These results indicate that the V-domain repertoire is dominated by two breed independent IGLV genes and a third breed specific IGLV gene that are effectively used for combinatorial joining of IGLJ-IGLC-genes and thus for antibody production. Furthermore, two putative pseudogenes (IGLV74ψ, IGLV101ψ) were transcribed. For IGLV74ψ a frameshift in CDR1 resulting in two in frame stop codons in FR3 and CDR3 was described (Sun et al., 2010). In addition, no TATA-box was identified. As known for humans and *Saccharomyces cerevisiae*, the TATA-box is not essential for accurate transcription initiation (Bjornsdottir and Myers, 2008; Yang et al., 2007). Consequently the premature stop codons are seen as reason for nonfunctionality. The isolated sequences did not show premature termination. We identified variations in CDR1 sequences that were introduced by the deletion of C, which was present at nucleotide position 96 of germline IGLV74ψ. This deletion resulted in a frame shift replacing the first stop codon by A, and the second one either by S, K, I, or R. As cDNA sequences were identified as IGLV74ψ due to sequence identities between 82.5% and 90.1%, we suppose individual mutations within the breeds to enable the transcriptional use of pseudogenes. Further we have to consider that germline sequences are prepared from a Thoroughbred, which is a very conserved breed founded by few stallions. So mutations resulting in pseudogenes may be spread within this breed but can be eliminated in other breeds. In RGC IGLV101ψ, also caused by a frame shift within the CDR1 of the germline sequence, was identified in two sequences. Transcription of pseudo V-gene segments has already been described for cattle, chicken, and rabbits (Arakawa and Buerstedde, 2009; Parng et al., 1996; Winstead et al., 1999). In these species usage of pseudogenes occurs by gene conversion which is a process contributing to antibody diversity. Our results indicated presence of similar mechanisms

in horses. In addition, we have to take into account that the spacer sequences of the recombination signal sequences shown in Sun et al. (2010) of the transcribed IGLVs seem to be conserved. The spacer sequence in IGLV101ψ is identical to that in IGLV17 and IGLV15, which are the mainly used IGLV. As shown in Nadel et al. (1998), spacer sequences contribute to rearrangement of human IGKV segments. In analogy these identical sequences of the spacers are an additional option for the transcription of potential pseudogenes. Moreover, transcriptional use of IGLV19 is limited in comparison to the other IGLV. The spacer of this segment shows three nucleotide substitutions relative to IGLV17 at positions 3, 10, and 12 of the spacer which also indicate the spacer to be a determinant factor for rearrangements.

As analyses took place in two breeds with different stud book sizes and breeding goals, some of the results could be linked to breeding. RGC is a quiet conserved breed with approximately 1348 mares and 153 stallions presently in Germany. Animals of this breed are mainly used for agricultural purposes and breed representations (Bremond and Balzer, 2011). In contrast, animals of the breed HW participate in all disciplines of equestrian sports and are internationally well placed. This breed influenced other horse breeds and it presently is one of the biggest horse breeds with 19492 registered mares and more than 522 stallions (Bremond and Balzer, 2011). As there was a large number of founder animals of genetically different breeds, the permission to choose stallions from different sire lines, and modest migration of sires from other breeds is also granted. The inbreeding coefficient in the Hanoverian population is at low level (1.33%) than those of smaller populations like the Rhenish German Coldblood (1.73%; Biedermann et al., 2002; Hamann and Distl, 2008). Nevertheless, effective management of breeding and therefore restricted use of single stallions is necessary to prevent the random loss of alleles caused by large genetic contributions of few individuals over a long time period (bottleneck effect). Genetic drift and homogenization of genetic information may also occur and have to be controlled as the rate of inbreeding has increased in Hanoverians in the last ten years (Constans et al., 1985; Hamann and Distl, 2008; Simianer and Kohn, 2010).

Our study aimed to characterize genetic and transcriptional differences of λ-light chain in two horse breeds. Distinct alleles and putative allotypic variants were described for the first time in horses. Future investigations should evaluate the distribution of allotypic markers of immunoglobulin chains in further horse breeds. Moreover, the linkage to infectious diseases such as in humans and cattle should be examined.

Acknowledgment

We are grateful to Jessica Olbricht for critical review of the manuscript and thank Tina Henkel as well as Caroline Bierschenk for excellent technical assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.vetimm.2011.10.006.

Legends to the Supplemental tables

Supplemental table 1 Percent identity (upper triangle) and divergence (lower triangle) of the detected allotypes of IGLC1 were divided by black boxes. Database entries were indicated by their accession number, calculations were based on the pairwise alignment using ClustalW (Thompson et al., 1994). Percent identity was compared to sequence pairs without regard to their phylogenetic relationships. Divergence was calculated in relation to the phylogeny.

Percent Identity (upper triangle); Divergence (lower triangle)

	1	**2**	**3**	**4**	**5**	**6**		
1		96.3	96.0	96.3	97.5	96.3	**1**	IGLC1[a]
2	3.8		97.5	98.8	97.8	98.8	**2**	IGLC1[b]
3	4.2	2.5		97.2	98.4	97.2	**3**	IGLC1[c]
4	3.8	1.3	2.9		97.8	100.0	**4**	IGLC1[d]
5	2.5	2.2	1.6	2.2		97.8	**5**	IGLC1[e]
6	3.8	1.3	2.9	0.0	2.2		**6**	CD470899
	1	**2**	**3**	**4**	**5**	**6**		

Supplemental table 2 Percent identity (upper triangle) and divergence (lower triangle) of the detected allotypes of IGLC5 were divided by black boxes. Calculations were based on the pairwise alignment using ClustalW (Thompson et al., 1994). Percent identity was compared to sequence pairs without accounting for their phylogenetic relationships. Divergence was calculated in relation to the phylogeny.

Percent Identity

Divergence

	1	2	3	4		
1		98.4	99.7	97.5	**1**	IGLC5[a]
2	1.6		98.8	96.9	**2**	IGLC5[b]
3	0.3	1.3		97.5	**3**	IGLC5[c]
4	2.5	3.2	2.5		**4**	IGLC5[d]
	1	**2**	**3**	**4**		

Supplemental table 3 Percent identity (upper triangle) and divergence (lower triangle) of the detected alleles and allotypes of IGLC6/7 were divided by black boxes. Calculations were based on the pairwise alignment using ClustalW (Thompson et al., 1994). Percent identity was compared to sequence pairs without regard to their phylogenetic relationships. Divergence was calculated in relation to the phylogeny.

Percent Identity

Divergence

	1	2	3	4	5		
1		99.7	100.0	99.4	99.1	**1**	IGLC7[a1]
2	0.3		99.7	99.0	98.7	**2**	IGLC6[a1]
3	0.0	0.3		99.4	99.1	**3**	IGLC6/7[a2]
4	0.6	1.0	0.6		98.4	**4**	IGLC6/7[a3]
5	0.9	1.4	0.9	1.6		**5**	IGLC6/7[b]
	1	**2**	**3**	**4**	**5**		

References

Almagro, J.C., Martinez, L., Smith, S.L., Alagon, A., Estevez, J., Paniagua, J., 2006. Analysis of the horse V(H) repertoire and comparison with the human IGHV germline genes, and sheep, cattle and pig V(H) sequences. Mol. Immunol. 43, 1836-1845.

Arakawa, H., Buerstedde, J.M., 2009. Activation-induced cytidine deaminase-mediated hypermutation in the DT40 cell line. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364, 639-644.

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.

Bastida-Corcuera, F., Butler, J.E., Heyermann, H., Thomford, J.W., Corbeil, L.B., 2000. Tritrichomonas foetus extracellular cysteine proteinase cleavage of bovine IgG2 allotypes. J. Parasitol. 86, 328-332.

Bastida-Corcuera, F.D., Butler, J.E., Yahiro, S., Corbeil, L.B., 1999a. Differential complement activation by bovine IgG2 allotypes. Vet. Immunol. Immunopathol. 71, 115-123.

Bastida-Corcuera, F.D., Nielsen, K.H., Corbeil, L.B., 1999b. Binding of bovine IgG2a and IgG2b allotypes to protein A, protein G, and Haemophilus somnus IgBPs. Vet. Immunol. Immunopathol. 71, 143-149.

Biedermann, G., Clar, U., Finke, A., Bickel, M., 2002. Analyse der Population des Rheinisch-Deutschen Kaltbluts. Züchtungskunde 74, 237-249.

Bjornsdottir, G., Myers, L.C., 2008. Minimal components of the RNA polymerase II transcription apparatus determine the consensus TATA box. Nucleic Acids Res. 36, 2906-2916.

Bremond, J., Balzer, B., 2011. Zentrale Dokumentation Tiergenetischer Ressourcen in Deutschland (TGRDEU).

Calderon, R., Lodeiro, R., Varela, T.A., Farina, J., Ambrosio, B., Guitard, E., Gonzalez-Martin, A., Dugoujon, J.M., 2007. GM and KM immunoglobulin allotypes in the Galician population: new insights into the peopling of the Iberian Peninsula. BMC Genet. 8, 37.

Constans, J., Hazout, S., Garruto, R.M., Gajdusek, D.C., Spees, E.K., 1985. Population distribution of the human vitamin D binding protein: anthropological considerations. Am. J. Phys. Anthropol. 68, 107-122.

Corbeil, L.B., Gogolewski, R.P., Kacskovics, I., Nielsen, K.H., Corbeil, R.R., Morrill, J.L., Greenwood, R., Butler, J.E., 1997. Bovine IgG2a antibodies to Haemophilus somnus and allotype expression. Can. J. Vet. Res. 61, 207-213.

Das, S., Nikolaidis, N., Klein, J., Nei, M., 2008. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. Proc. Natl. Acad. Sci. U. S. A. 105, 16647-16652.

Diesterbeck, U.S., Aboelhassan, D.M., Stein, S.K., Czerny, C.-P., 2011 in press. Detection of new allotypic variants of bovine λ-light chain constant regions in different cattle breeds. Dev. Comp. Immunol.

Edelman, G.M., 1973. Antibody structure and molecular immunology. Science 180, 830-840.

Emorine, L., Dreher, K., Kindt, T.J., Max, E.E., 1983. Rabbit immunoglobulin kappa genes: structure of a germline b4 allotype J-C locus and evidence for several b4-related sequences in the rabbit genome. Proc. Natl. Acad. Sci. U. S. A. 80, 5709-5713.

Faber, C., Shan, L., Fan, Z., Guddat, L.W., Furebring, C., Ohlin, M., Borrebaeck, C.A., Edmundson, A.B., 1998. Three-dimensional structure of a human Fab with high affinity for tetanus toxoid. Immunotechnology 3, 253-270.

Gibson, D., 1974. Structural studies on normal horse immunoglobulin light chains. Detection of k-type N-terminal sequences. Biochemistry 13, 2776-2785.

Giha, H.A., Nasr, A., Iriemenam, N.C., Arnot, D., Troye-Blomberg, M., Theander, T.G., Berzins, K., ElGhazali, G., Pandey, J.P., 2009. Antigen-specific influence of GM/KM allotypes on IgG isotypes and association of GM allotypes with susceptibility to Plasmodium falciparum malaria. Malar. J. 8, 306.

Granoff, D.M., Pandey, J.P., Boies, E., Squires, J., Munson, R.S., Jr., Suarez, B., 1984. Response to immunization with Haemophilus influenzae type b polysaccharide-pertussis vaccine and risk of Haemophilus meningitis in children with the Km(1) immunoglobulin allotype. J. Clin. Invest. 74, 1708-1714.

Hamann, H., Distl, O., 2008. Genetic variability in Hanoverian warmblood horses using pedigree analysis. J. Anim. Sci. 86, 1503-1513.

Hartmann, O., Bruns, E., Hertsch, B., Schade, W., 2006. Pferdezucht. Eugen Ulmer KG, Stuttgart (Hohenheim).

Home, W.A., Ford, J.E., Gibson, D.M., 1992. L chain isotype regulation in horse. I. Characterization of Ig lambda genes. J. Immunol. 149, 3927-3936.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33-38, 27-38.

Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Asakawa, S., Sasaki, T., Klobeck, H.G., Combriato, G., Zachau, H.G., Shimizu, N., 2001. Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the Vkappa genes. Eur. J. Immunol. 31, 1017-1028.

Kelley, L.A., Sternberg, M.J., 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4, 363-371.

Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W., Richardson, P., 2002. Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. Dev. Dyn. 225, 384-391.

Korngold, L., Lipari, R., 1956. Multiple-myeloma proteins. III. The antigenic relationship of Bence Jones proteins to normal gammaglobulin and multiple-myeloma serum proteins. Cancer 9, 262-272.

Kratzin, H.D., Palm, W., Stangel, M., Schmidt, W.E., Friedrich, J., Hilschmann, N., 1989. [The primary structure of crystallizable monoclonal immunoglobulin IgG1 Kol. II. Amino acid sequence of the L-chain, gamma-type, subgroup I]. Biol. Chem. Hoppe Seyler 370, 263-272.

Lee, B., Richards, F.M., 1971. The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. 55, 379-400.

Lefranc, M.P., Pommie, C., Kaas, Q., Duprat, E., Bosc, N., Guiraudou, D., Jean, C., Ruiz, M., Da Piedade, I., Rouard, M., Foulquier, E., Thouvenin, V., Lefranc, G., 2005. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Dev. Comp. Immunol. 29, 185-203.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R., 2004. SAS® system for Mixed Models, 6th Edition. SAS Institute Inc, Cary, NC, USA.

Matsumoto, H., Miyazaki, T., Rychkov, Y.G., Zhukova, O., Lebedeva, I.A., Kondik, V.M., Udina, I.G., Spitzin, V.A., Batsuur, J., Shnader, Y.V., et al., 1984. Studies on the human immunoglobulin allotypes in five populations in the USSR. Jinrui Idengaku Zasshi 29, 105-111.

Montano, R.F., Morrison, S.L., 2002. Influence of the isotype of the light chain on the properties of IgG. J. Immunol. 168, 224-231.

Moxley, G., Gibbs, R.S., 1992. Polymerase chain reaction-based genotyping for allotypic markers of immunoglobulin kappa shows allelic association of Km with kappa variable segment. Genomics 13, 104-108.

Nadel, B., Tang, A., Escuro, G., Lugo, G., Feeney, A.J., 1998. Sequence of the spacer in the recombination signal sequence affects V(D)J rearrangement frequency and correlates with nonrandom Vkappa usage in vivo. J. Exp. Med. 187, 1495-1503.

Oli, M.W., Rhodin, N., McArthur, W.P., Brady, L.J., 2004. Redirecting the humoral immune response against Streptococcus mutans antigen P1 with monoclonal antibodies. Infect. Immun. 72, 6951-6960.

Padlan, E.A., Cohen, G.H., Davies, D.R., 1986. Antibody Fab assembly: the interface residues between CH1 and CL. Mol. Immunol. 23, 951-960.

Pandey, J.P., 2000. Immunoglobulin GM and KM allotypes and vaccine immunity. Vaccine 19, 613-617.

Pandey, J.P., Elson, L.H., Sutherland, S.E., Guderian, R.H., Araujo, E., Nutman, T.B., 1995. Immunoglobulin kappa chain allotypes (KM) in onchocerciasis. J. Clin. Invest. 96, 2732-2734.

Pandey, J.P., Fudenberg, H.H., Virella, G., Kyong, C.U., Loadholt, C.B., Galbraith, R.M., Gotschlich, E.C., Parke, J.C., Jr., 1979. Association between immunoglobulin allotypes and immune responses to Haemophilus influenzae and Meningococcus polysaccharides. Lancet 1, 190-192.

Parng, C.L., Hansal, S., Goldsby, R.A., Osborne, B.A., 1996. Gene conversion contributes to Ig light chain diversity in cattle. J. Immunol. 157, 5478-5486.

Pritsch, O., Hudry-Clergeon, G., Buckle, M., Petillot, Y., Bouvet, J.P., Gagnon, J., Dighiero, G., 1996. Can immunoglobulin C(H)1 constant region domain modulate antigen binding affinity of antibodies? J. Clin. Invest. 98, 2235-2243.

Roschenthaler, F., Hameister, H., Zachau, H.G., 2000. The 5' part of the mouse immunoglobulin kappa locus as a continuously cloned structure. Eur. J. Immunol. 30, 3349-3354.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467.

Schanfield, M.S., Ferrell, R.E., Hossaini, A.A., Gerald Sandler, S., Stevenson, J.C., 2008. Immunoglobulin allotypes in Southwest Asia: populations at the crossroads. Am. J. Hum. Biol. 20, 671-682.

Simianer, H., Kohn, F., 2010. Genetic management of the Gottingen Minipig population. J. Pharmacol. Toxicol. Methods 62, 221-226.

Sitnikova, T., Su, C., 1998. Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. Mol. Biol. Evol. 15, 617-625.

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.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

Tonegawa, S., 1983. Somatic generation of antibody diversity. Nature 302, 575-581.

Wagner, B., 2006. Immunoglobulins and immunoglobulin genes of the horse. Dev. Comp. Immunol. 30, 155-164.

Winstead, C.R., Zhai, S.K., Sethupathi, P., Knight, K.L., 1999. Antigen-induced somatic diversification of rabbit IgH genes: gene conversion and point mutation. J. Immunol. 162, 6602-6612.

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., Martinez, E., 2007. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene 389, 52-65.

Zhou, J., Lottenbach, K.R., Barenkamp, S.J., Reason, D.C., 2004. Somatic hypermutation and diverse immunoglobulin gene usage in the human antibody response to the capsular polysaccharide of Streptococcus pneumoniae Type 6B. Infect. Immun. 72, 3505-3514.

Chapter 3:

Bovine Immunoglobulins

The nature and extent of my contribution to the work was the following:

3. Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins

Nature of contribution	Extent of contribution
1. Scientific design	50%
2. Laboratory work	80%
3. Evaluation	70%
4. Scientific Writing	65%

4. Development of a bioinformatics framework for the detection of gene conversion and the analysis of combinatorial diversity in immunoglobulin heavy chains in four cattle breeds

Nature of contribution	Extent of contribution
1. Scientific design	50%
2. Laboratory work	50%
3. Evaluation	70%
4. Scientific Writing	65%

Bovine immunoglobulin heavy and light chains

Bovine immunoglobulin heavy chain gene locus

The bovine immunoglobulin heavy chain gene (IGH) locus was assigned to the *Bos taurus* autosome (BTA) 21 and spans approximately 250 kb (Miller et al. 1992; Niku et al. 2012). Southern blot analyses indicated the existence of two IGHV families (boVH1 and boVH2) in the bovine germline repertoire. Before the study of Walther et al. (2013), the expression of only boVH1 genes has been observed and boVH2 was found comprising only pseudogenes (Tutter and Riblet 1989; Berens et al. 1997; Saini et al. 1997; Sinclair et al. 1997; Lopez et al. 1998). The family boVH1 offers a restricted set of functional genes, which shares homologies to the murine Q52 family and human VHII family. The definite number and organization of IGHV has not been determined yet. The latest *in silico* analysis using blastn on all bovine genome assemblies revealed IGHV on the contigs NW_003104530.1 and NW_003104538.1 on *Bos taurus* chromosome 21 (BTA21; AC_000178.1), NW_003064289.1, NW_003064290.1, NW_003064296.1, NW_003064297.1, NW_003064298.1, and NW_003064299.1 on BTA7 (AC_000164.1), as well as on several unplaced contigs (Walther et al. 2013). In total, 36 IGHV are identified in the bovine genome assembly. On BTA 21, IGHV1Ψ and IGHV2 were found at the centromeric region in NW_003104530.1. Eight IGHV located at the telomeric region were described within about 146 kb on NW_003104538.1 (IGHV3, IGHV4Ψ, IGHV5Ψ, IGHV6, IGHV7Ψ, IGHV8Ψ, IGHV9Ψ, and IGHV10). The contigs NW_003064289.1, NW_003064290.1, NW_003064296.1, NW_003064297.1, NW_003064298.1, and NW_003064299.1 were annotated to 144 kb of the centromeric region on BTA7 possessing nine IGHV (IGHV11Ψ, IGHV12Ψ, IGHV13Ψ, IGHV14Ψ, IGHV15Ψ, IGHV16(ORF), IGHV17(ORF), IGHV18Ψ, and IGHV19Ψ). Thirteen out of the 36 IGHV segments are putatively functional. Eleven IGHV segment pairs shared sequence identity of 100% (IGHV3/33, IGHV10/34, IGHV9Ψ/35Ψ, IGHV4Ψ/32Ψ, IGHV7Ψ/22Ψ, IGHV2/26, IGHV1Ψ/27Ψ, IGHV18Ψ/30Ψ, IGHV16(ORF)/25, IGHV14Ψ/23Ψ, and IGHV36/29(F)). Phylogenetic analyses revealed a relation of all functional IGHV to huIGHV2-05, whereas this group corresponds to the bovine IGHV family 1 (boVH1) (Berens et al. 1997; Saini et al. 1997; Sinclair et al. 1997).

In cattle, ten IGHD genes are classified into four families that are organized in sub-clusters (Koti et al. 2008; Koti et al. 2010). Further, the IGHD exons reveal huge size

differences (Koti et al. 2008). On the telomeric end of BTA21, the bovine IGHD8 and IGHD4 were localized between IGHV6 and IGHV7Ψ (Walther et al. 2013). Within the contig NW_003064411.1 on BTA7 one IGHD cluster of 1131 bp, encoding IGHD1(ORF) to IGHD3(ORF), was identified. On BTA8 a third location enclosing about 43 kb possesses five IGHD segments, a δ chain pseudogene, a μ chain gene, and six IGHJ segments. In general, loci for the IGHD were observed on three chromosomes (Walther et al. 2013).

By hybridization experiments, six joining segments (IGHJ) were detected on BTA11q23 (Hosseini et al. 2004). The number of IGHJ segments preceding the IGHC genes was described by screening a bovine BAC and Cosmid library. Only two out of six IGHJ are functional (IGHJ1, IGHJ2). IGHJ1 is recombined predominantly while IGHJ2 is rearranged at low frequency (Zhao et al. 2003; Hosseini et al. 2004).

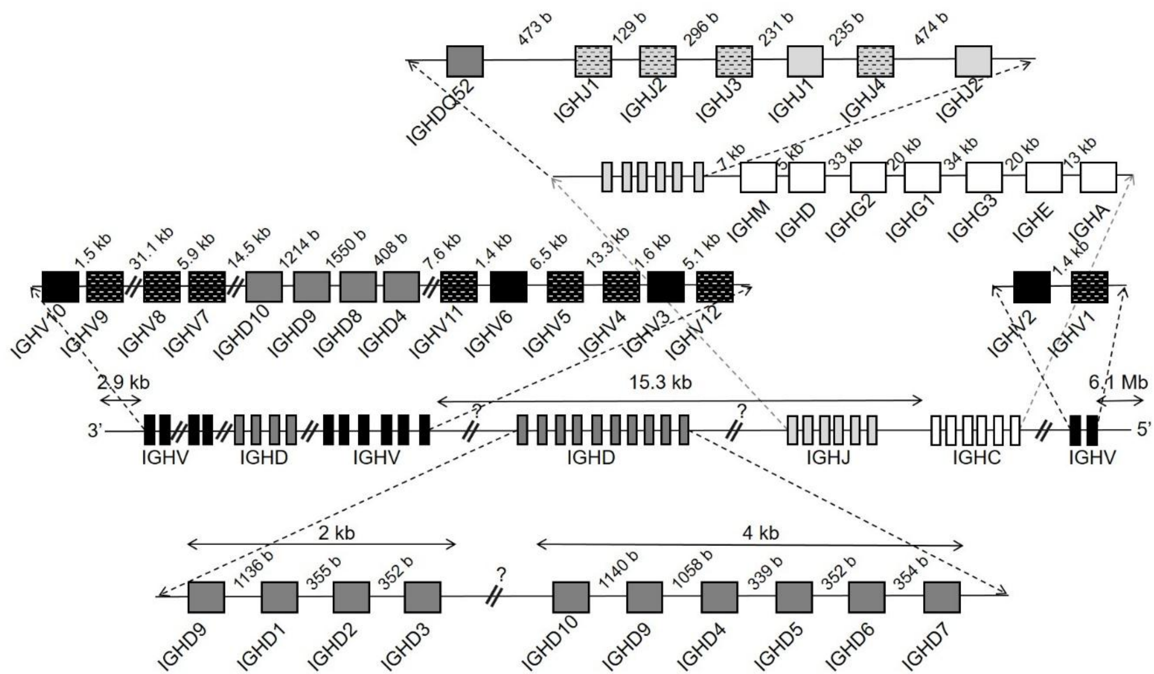


Figure 1: Physical map of the bovine heavy chain locus
The locus is shown in 3'-5' direction on chromosome 21. This figure is adapted from current literature (Zhao et al. 2003; Zimin et al. 2009; Koti et al. 2010; Niku et al. 2012; Walther et al. 2013; Pasman 2014).

The CDR3H encoded by IGHV-IGHD-IGHJ recombinations varies from 3 to 67 amino acid residues (Berens et al. 1997; Saini et al. 1999; Saini and Kaushik 2002; Saini et al. 2003; Kaushik et al. 2009; Larsen and Smith 2012). These exceptionally long CDR3H are specific for cattle and possess characteristic hydrophilic Glycine and Tyrosine residues. The 148 bp long IGHD2 contributes to those CDR3H (Berens et al. 1997; Shojaei et al. 2003; Koti et al. 2008). Further, a high number of Cysteine residues was described and is demonstrated to promote intra-CDR3H disulfide bonds that contribute to the “knob” domain in exceptionally long CDR3H (Sinclair et al. 1997). In contrast, inter CDR disulfide bonds may exist in mid-length CDR3H due to one to three Cys residues in CDR3H and one Cys within CDR2H (Saini et al. 1999; Ramsland et al. 2001). Additional mechanisms contributing to the unusual length of bovine CDR3H were studied because the imprecise junction of germline encoded IGHV, IGHD, and IGHJ cannot fully explain the existence of exceptionally long CDR3H in cattle. The insertion of conserved short nucleotide sequences of 13 to 18 nucleotides specifically into the IGHV and IGHD junction was found to contribute to a further elongation of the CDR3H. So far, this mechanism is unique for cattle (Koti et al. 2010).
The genomic organization of the bovine IGHC locus was described by screening a bovine BAC and Cosmid library. All heavy chain isotype classes that have been detected in other mammals were also described for cattle (Knight and Becker 1987; Zhao et al. 2002). Three sub-classes comprise the γ-isotype, namely γ1, γ2, and γ3 (Knight and Becker 1987; Symons et al. 1989). The IGHC locus spans approximately 150 kb enclosing the genes in the following order: 5’-JH–7 kb–μ–5 kb–δ–33 kb–γ3–20 kb–γ1–34 kb–γ2–20 kb–ε–13 kb–α-3’ (Zhao et al. 2003).

Bovine immunoglobulin M (IgM)

IgM is the major serum antibody in the primary immune response as it provides the first antibody-mediated host defense (Saini and Kaushik 2001; Woof and Burton 2004). IgM combats septicaemia when administered passively to calves and it is an important bactericidal antibody directed against mastitis in cattle and other ruminants (Mousavi et al. 1998). High amounts of IgM are described in colostrum and milk (3 mg/ml) (Butler 1995).
Two loci comprising functional IGHM genes were identified on BTA21 and BTA11q23 by *in situ* hybridization, whereas the latter is currently assigned to BTA8 (Tobin-Janzen and Womack 1992; Hayes and Petit 1993). The IGHM gene comprises four exons

(IGHM1-4) encoding the constant domains, as well as two additional exons encoding the transmembrane domain (TM1, TM2) (Mousavi et al. 1998). Three bovine IgM allotypes are described, designated as IgMa, IgMb, and IgMc. While these allotypes are classified by amino acid substitutions within the exons, further IgM variants may result via alternative splicing e.g. inserting in frame codons at the IGHM1 and IGHM2 junction. IGHM2 possesses few proline residues and functions as a hinge restricting flexibility and antigen binding ability of IgM (Saini and Kaushik 2001; Pasman 2014).

Bovine immunoglobulin D (IgD)

On BTA7 (contig NW_003064411.1), a cluster of 1131 bp was identified encoding exons for IGHD1(ORF) to IGHD3(ORF) and a pseudogene for a δ chain. The exons 1 and 2 of the IGHD pseudogene are fragmented and exon 6 is missing (codes for the secretory region). Nevertheless, IgD is also found in a secreted form although showing lower concentrations than IgG, IgA, and IgM (Preud'homme et al. 2000; Zhao et al. 2002). Further, frame shifts were identified. Genomic information for IGHD4 to 7 and a δ chain pseudogene were found on the contig NW_001503306.1, while IGHD4 and 8 were detected on NW_001504477.2. A locus involving a putative functional δ chain gene was detected on NT_186572. In addition to NW_003100387.1, including a δ chain pseudogene locus, a putative functional δ chain gene was noted on NW_003100112.1 (Walther et al. 2013).

The bovine, ovine, and porcine germline IgD genes are transcriptionally active at low levels and show structural similarities to the human IgD due to the three constant domains and the hinge region. As phylogenetic analyses revealed the Cδ gene duplicated from the Cμ gene more than 300 million years ago, the exon of the first constant domain of the IgD (IGHD1) is very similar to the IGHM1 exon. Further, unlike other species, the bovine IGHD has a switch region that may permit class switch recombination (Zhao et al. 2002).

Bovine immunoglobulin G (IgG)

Immunoglobulin G is the most abundant isotype in bovine serum as it constitutes more than 75% of serum immunoglobulins (Saini et al. 2007). It is also found in blood, lymph, peritoneal fluid, and cerebrospinal fluid. Furthermore, IgG contributes to different effector functions. For instance, it is involved in the enhancement of phagocytosis,

antibody-dependent cellular cytotoxicity, the regulation of cytokine and antibody production, antigen processing and presentation, as well as in the selective transfer via the FcRn-like receptor on alveolar epithelial cells of mammary glands (Burg and Pillinger 2001; Kacskovics 2004; Mestecky 2005; Pentsuk and van der Laan 2009).

Three IgG subclasses (IgG1, IgG2, and IgG3) were identified in cattle, whereas two subclasses (IgG1 and IgG2) were found in sheep, seven IgG subclasses (IgG1-7) were detected in horses, four subclasses were described in humans, five in swine, and one in rabbits (Knight et al. 1988; Symons et al. 1989; Clarkson et al. 1993; Kacskovics et al. 1995; Kacskovics and Butler 1996; Rabbani et al. 1997; Wagner et al. 2004; Pasman 2014). Depending on characterization methods, two different types of designation exist. IgG1, IgG2a, and IgG2b were used in serological analyses, whereas IgG1, IgG2, and IgG3 were used in molecular analyses, respectively (Knight et al. 1988). Serum concentrations of IgG1 and IgG2 reveal 10 mg/ml and the concentration of IgG1 can exceed 60 mg/ml in hyperimmunized animals. In colostrum, IgG1 concentration was also found to exceed 100 mg/ml (Butler 1995). The bovine IgG3 shows low serum concentrations, which may be caused by the longer hinge region that is a preferred proteolytic site (Rabbani et al. 1997).

These three bovine Cγ genes (γ1, γ2, and γ3) have been mapped to chromosome 21q24 and are located between the δ and ε genes of the heavy chain locus (Gu et al. 1992; Tobin-Janzen and Womack 1992; Chowdhary et al. 1996; Saini et al. 2007). Interestingly, in an *in silico* study, on the bovine genome assembly BTA20 revealed the genomic information for a γ2 chain. A genomic order of γ3, γ2, and γ1 was identified on NT_185580.1. Genes coding for γ1 and γ3 were found on NW_003100065.1 and NW_003099305.1 (Walther et al. 2013). Gene duplication of bovine IGHG1 led to IGHG2 and IGHG3. Sequence similarities of 85.1% and 83.4% exist between the homologues IGHG1 and IGHG3, as well as IGHG1 and IGHG2, respectively (Zhao et al. 2003). For IgG1, four allotypic variants were described (IgG1a-d) (Symons et al. 1989; Jackson et al. 1992; Kacskovics and Butler 1996; Saini et al. 2007). The bovine IgG2 and IgG3 reveal the two allotypic variants IgG2a and IgG2b (Kacskovics and Butler 1996) and IgG3a and IgG3b (Rabbani et al. 1997), respectively. Significant structural differences within the hinge regions and the CH3 domain result in different biological effector functions of IGHG1, IGHG2, and IGHG3. Amino acid residue substitutions were also described within the CH1 and CH2 domains of these three genes (Saini et al. 2007).

The four known IgG1 allotypic variants are characterized by individual amino acid

differences at the positions 190 and 192 (CH1), 218, 224, 225, and 226 (hinge region), as well as 281 (CH2), and 402 (CH3), which were designated using the EU numbering system. IgG1a showed the amino acids Gly190, Thr192, Arg218, Thr224, and Thr226, Asp281, Gly402, as well as a deletion of the amino acid at position 225. The most similar allelic variants IgGb and IgGd possess Gly190, Thr192, Thr218, Pro224, Ser225, Pro226, and Asp281. They solely differ by Gly and Ser at position 402. The allotypic variant IgGc shows the amino acid residues Ala190, Ser192, Arg218, Arg224, Pro226, Asn281, Ser402 and a deletion at position 225. Three-dimensional modeling revealed that all substitutions are located on the surface of the IgG molecule and therefore may be recognized by antisera and effector molecules such as cellular adhesion molecules (Symons et al. 1989; Saini et al. 2007). Additionally, the repeating motif Pro-Ala-Ser-Ser at the positions 189–192 and 205–208 within the CH1 domain of IgG1c may stabilize the core immunoglobulin fold or promote interactions with fibronectin or similar adhesion molecules. The replacement of Thr with Arg residues within the hinge region reduces the number of O-linked glycosylation sites and increases the susceptibility to degradation by protease secreting bacteria (Saini et al. 2007).

The allotypic variants IgG2a and IgG2b differed in amino acid residues at 19 positions. These amino acid residue substitutions were found in the three constant region domains (CH1, CH2, and CH3) but also in the hinge region. In CH1, IgG2a and IgG2b differed by Ser and Ala at position 129 (Ser/Ala129), Cys/Ser131, Lys/Thr135, Gly/Ala190, and Thr/Ser192. Within the hinge region Ser/Ile219, Pro/Cys224, Asn/His225, and His/Asn228 were observed. In CH2, Thr/Ser290, Ile/Asn326, and Ser/Pro332 differentiate IgG2a and IgG2b. Val/Leu364, Ile/Thr370, Asp/Ala378, Asp/Asn384, Ala/Thr402, Arg/Glu419, and Thr/Ala422 are the characteristic amino acids described for IgG2a and IgG2b, respectively. These amino acid differences result in different immune responses of the two IgG2 allotypes (Kacskovics and Butler 1996; Corbeil et al. 1997). In comparison to IgG2a, IgG2b activates the complement more than twice (Bastida-Corcuera et al. 1999a).

Previous studies indicated that IgG3a is expressed at a lower level than IgG3b. These two allotypic variants differ in six positions. The amino acids Ser/Arg131, Arg/Leu174, and Ser/Thr192 are located in the CH1 domain, whereas Pro/Gln237 and Lys/Glu238 were found in the hinge region, and Lys/Arg4431 characterizes individual amino acid residues of IgG3a and IgG3b in the CH3 domain (Rabbani et al. 1997).

The allotypic variants of IgG provide polymorphic immunoglobulin genetic markers with specific abilities. In particular, modified qualities in complement activation, age-dependent expression, and influences on the effector function against *Haemophilus somnus* and *Tritrichomonas foetus* were made for the allotypic variants of bovine IgG2 (Corbeil et al. 1997; Bastida-Corcuera et al. 1999b; Bastida-Corcuera et al. 2000; Saini et al. 2007).

Bovine immunoglobulin E (IgE)

Whereas IgE is found in low amounts in the serum, IgE is mainly found during immune responses against parasite infections mediated by basophiles and mast cells. Furthermore, the presence of IgE gives rise to type I hypersensitivity reactions such as asthma, conjunctivitis, and rhinitis (Mousavi et al. 1997). The bovine IgE also shows heat labile skin sensitizing ability as the human IgE does (Hammer et al. 1971).

The latest *in silico* analysis revealed genomic information for an ε chain on NT_185723.1 (Walther et al. 2013). This finding is consistent with previous studies, where only a single copy of the bovine IgE was identified (Knight et al. 1988). IgE has four exons (IGHE1-4) showing homologies of 87% to sheep IGHE, 58% to horse IGHE, 55% to human IGHE, and 52% to mouse IGHE. Due to the additional fourth constant domain, IGHE possesses a higher molecular weight than other Ig isotypes. The additional domain is supposed to be important for high affinity binding to Fc receptors on mast cells (Mousavi et al. 1997).

Bovine immunoglobulin A (IgA)

Immunoglobulin A is the dominating isotype on mucosal surfaces such as mammary glands with an important part in the immune response against microbial infections.

A single IGHA gene was identified in the bovine genome and was assigned to the contig NT_185723.1 (Knight et al. 1988; Walther et al. 2013). IgA possesses three IGHA exons (Brown et al. 1997). Two allelic variants were identified by restrictions fragment length polymorphism and serological analyses but were not confirmed by genomic DNA analyses of 50 Swedish cattle (De Benedictis G 1984; Brown et al. 1997; Pasman 2014).

The bovine immunoglobulin light chain gene loci

In cattle, two distinct immunoglobulin light chain isotypes (lambda λ and kappa κ) have been described (Pilstrom 2002; Das et al. 2008). In serum antibodies, bovine λ-light chains are predominantly expressed in a proportion of 95% (Arun et al. 1996). About 5% of the heavy chains in serum antibodies are connected with functional κ-light chains (Arun et al. 1996; Aitken et al. 1999). Similarly, the light chain repertoire is dominated by λ-light chains in horses (Ford et al. 1994) and sheep (Foley and Beh 1992; Griebel and Ferrari 1994; Broad et al. 1995), although a functional kappa system was described, too (Home et al. 1992). In contrast, human and mice serum antibodies dominantly associate κ-light chains (60% and 95%) (Chen et al. 2008). Genomic conditions such as genomic complexity of the loci, recombination signal sequences, the ordered rearrangement of κ- and λ-light chains, as well as antigen selection are supposed to influence differences in the κ:λ ratio across species (Pasman et al. 2010).

Bovine lambda light chain genes

The bovine variable, joining, and constant region genes of the λ-light chain (IGL) are located on *Bos taurus* autosome 17 (BTA17) (Tobin-Janzen and Womack 1992). The locus spans 412 kb and contains 25 to 31 variable gene segments (IGLV) whereof 14 to 17 are potentially functional (Ekman et al. 2009). A total of 63 IGLV were identified on ten scaffolds in the current genome assembly (Btau_3.1). Out of these 63 IGLV, 32 were identified on unplaced contigs and 11 of them are also potentially functional (Ekman et al. 2009). The IGLV locus possessed both transcriptional orientations. All IGLV are organized in three clusters separated by 126.8 kb and 138.3 kb introns. The three IGLV clusters also define separate IGLV families designated as IGLV1, IGLV2, and IGLV3. The predominantly expressed IGLV1 genes are subdivided into five subfamilies (IGLV1a, IGLV1b, IGLV1c, IGLV1d, IGLV1x) (Sinclair et al. 1995; Saini et al. 2003). These IGLV1 are located within the two 5' subclusters, whereas the minor expressed IGLV2 and IGLV3 genes were identified within the IGLJ-proximal cluster. The subfamilies IGLV1a and IGLV1b are mainly expressed, whereas IGLV1c, IGLV1e, and IGLV1x specifically pair with heavy chains possessing CDR3H with at least 50 amino acid residues (Saini et al. 1999; Saini et al. 2003). Analyses of genomic sequences of the cattle breed Herford revealed 17 IGLV1 genes of which ten are described functional. Further, three out of four IGLV2 genes as well all four IGLV3 genes were designated functional (Pasman et al. 2010). Phylogenetic comparisons revealed that four of six ovine subgroups cluster with the bovine ones, which are

supposedly ruminant specific genes. Consequently, due to sequence identities of 92.6% and 88.9%, the bovine IGLV1 and IGLV2 families are closely related with the ovine families IGLVI and IGLVII. The bovine IGLV3 family is related to the unclassified ovine IGLV family (Pasman et al. 2010; Pasman 2014). In addition, the CDR1L length varies from 6, 8, or 9 codons. The length of CDR2L is restricted to 3 or 7 codons. The number of unique CDR1L–CDR2L combinations is lower than in mice and humans (Ekman et al. 2009).

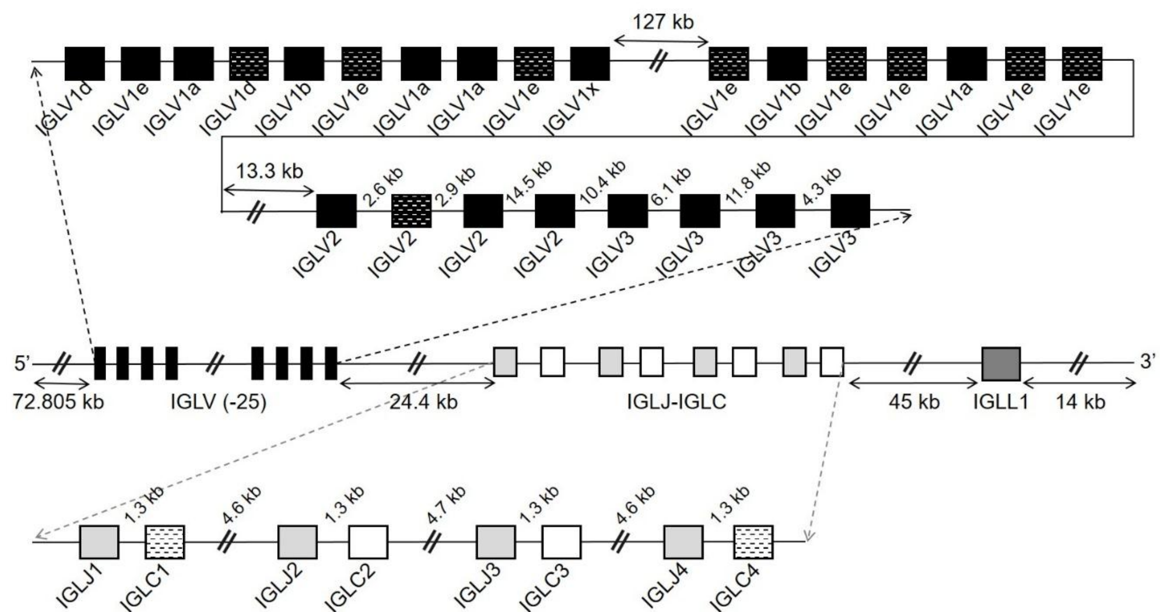


Figure 2: Physical map of the bovine lambda light chain locus
The locus is shown in 5'-3' direction on chromosome 17. This figure is adapted from (Pasman et al. 2010; Pasman 2014).

In cattle, four λ-light chain constant region (IGLC) genes (IGLC1, IGLC2, IGLC3, and IGLC4) were identified yet, each preceded by a joining gene segment (IGLJ1-4) (Parng et al. 1995; Parng et al. 1996; Chen et al. 2008; Ekman et al. 2009; Pasman et al. 2010). Only IGLC2 and IGLC3 are functional whereas IGLC3 is preferentially expressed during rearrangements. IGLC1 and IGLC4 are pseudogenes (Chen et al. 2008; Ekman et al. 2009; Pasman et al. 2010). A fifth bovine IGLC gene was detected but could not be located on a chromosome until now (Ekman et al. 2009). In comparison, in sheep only two IGLC genes were described (IGLC1 and IGLC2). The

ovine IGLC1 is functional and in the IGLC2 gene possess a premature stop codon (Jenne et al. 2003). Furthermore, in previous studies seven equine IGLC genes were identified. Three of these genes are functional, whereas the others were described to be pseudogenes (Home et al. 1992; Das et al. 2008; Sun et al. 2010).

Transcription analyses in several cattle breeds revealed single nucleotide polymorphisms (SNPs) for both the bovine IGLC2 and IGLC3 genes resulting in silent mutations or amino acid residue substitutions (Diesterbeck et al. 2012). Within the breeds German Simmental (GS) and Aubrac (A), single nucleotide polymorphisms of IGLC2 were identified. Analyses in the breeds German Black Pied (GBP) and Holstein Friesian (HF) revealed the already known allele and allotype IGLC2a. Two additional alleles, IGLC2b and IGLC2c, were detected and represent putative new allotypes. Amino acid residue substitutions were either conservative or led to charge changes. They were located within the molecule by homology based three-dimensional modeling. Most amino acid residue substitutions were found on the outer side of the molecule in the solvent accessible surface area, within or near the putative interface to IGHC1. Additionally, by in silico analyses IGLC2a was also identified in other cattle breeds such as Angus (DY170709) and Hereford (EH173018). IGLC2a and IGLC2b showed 98.4% sequence identity while IGLC2c was 97.8% identical to IGLC2a and 96.9% identical to IGLC2b (Diesterbeck et al. 2012). Similar to IGLC2, five allelic variants were identified for the predominantly expressed IGLC3 (IGLC3a, IGLC3b, IGLC3c, IGLC3d, IGLC3e) within the same cattle breeds. Amino acid residue substitutions were also determined on the accessible surface of the molecule, as well as in and next to the interface to IGHC1. Reversed charges were observed due to several amino acid residue replacements. Further, nucleotide sequences coding for IGLC3a were found in Hereford (EV679232) and Angus (DY173535), while IGLC3b was only found in Angus (DY149783). IGLC3c was identified both in Holstein Friesian crossbreed (DY145594) and purebred (CK950153). Angus and Holstein Friesian crossbreeds (EV608839, DY216512), as well as Holstein Friesian purebred (CK979405) and Hereford (DT858283) exhibited IGLC3d (Diesterbeck et al. 2012).

Bovine kappa light chain genes

The bovine κ-light chain genes are located on *Bos taurus* autosome 11 (BTA11) (Ekman et al. 2009) and span approximately 280 kb, although the complete Ig κ-light chain (IGK) locus encompasses a size of 412 kb (Pasman et al. 2010). The analysis of the bovine κ-light chain locus on Herford genome sequence revealed 22 IGKV, 3 IGKJ,

and 1 IGKC (Ekman et al. 2009). The IGKV genes are organized in a cluster that covers 150 kb and which is followed by the three IGKJ genes and the IGKC gene. Out of 22 IGKV, only eight are functional (Ekman et al. 2009). They are present in both transcriptional orientations (Sitnikova and Nei 1998). The 22 IGKV genes are classified into four subfamilies. Twenty-one of them are closely related to the ovine gene families 1, 2, and 4. Seven of the eight functional segments belong to bovine subgroup 2 (IGKV2), which therefore is supposed to be mainly expressed. This subgroup is phylogenetically related to sheep (91%), killer whale (84%), and dolphin (84%). Bovine IGKV vary from six, ten, or 11 codons within CDR1L and differs from human, mice, and sheep CDR1L possessing 11 codons, whereby CDR2L constantly possesses three codons (Ekman et al. 2009; Ehrenmann et al. 2010; Pasman 2014).

The three IGKJ genes (IGKJ1, IGKJ2, IGKJ3) are clustered within 800 bp on the genome and are followed by one IGKC gene. Although all IGKJ genes are functional, they were all transcribed in different frequencies. IGKJ1 showing the conserved 23 bp RSS, was transcribed predominantly in animals of the breeds GS, HF, A, and GBP (Stein et al. 2012). IGKJ2 was detected in two animals of GS despite of the shortened spacer with a size of 22 bp instead of 23 bp (Ekman et al. 2009). IGKJ3 was also described in each of the four cattle breeds. The distinct cladistic protein motif ‘EIK’ was found changed in IGKJ3 to ‘EIK’ (Das et al. 2008).

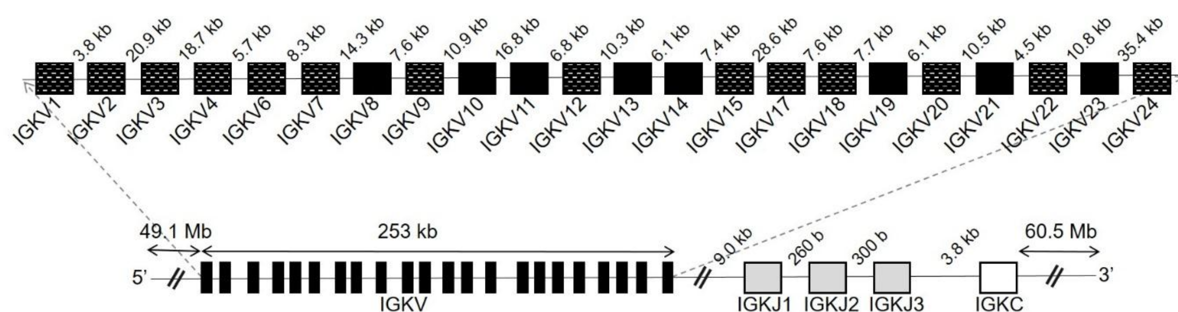


Figure 3: Physical map of the bovine kappa light chain locus
The locus is shown in 5’-3’ direction on chromosome 11. The locus is identical to both BTAU 4.6 and UMD 3.1 assembly of the Hereford cow genome. This figure is adapted from Pasman and Kaushik (2014) and Ekman et al. (2009).

Transcriptional analysis of IGKC within four cattle breeds revealed three alleles distributed in different frequencies (Stein et al. 2012). In animals of HF, GS, and A, all three alleles were identified, whereas in GBP solely IGKCa was found. Additional *in silico* analyses of the ESTs database revealed allele 1 and 2 of IGKCa in Hereford Shorthorns (CF763228, CF768037). The three alleles differed at four nucleotide positions with two of them leading to amino acid residue substitutions at positions 100 and 116 within the constant region resulting in the putative allotypic variants of IGKCa and IGKCb. The two amino acid residue replacements caused either a change in polarity or a change in electric charge of the amino acid side chain. The amino acid residue substitutions were located at the solvent accessible surface. Also for humans and rabbits three and five allotypes were described, respectively (Emorine et al. 1983; Moxley and Gibbs 1992). In humans, an improvement of immune responses by IGKC allotypes was observed (Pandey et al. 1979; Granoff et al. 1984). Due to the similar substitutions in bovine allotypes, different serological properties are assumed in cattle allotypes, too (Stein et al. 2012).

Development of B cells and the bovine antibody repertoire

The development of immunoglobulin producing cells already starts in the bovine fetus. While IgM-bearing B-cells have been detected at day 59 of gestation (Schultz et al. 1973), recombinations of heavy and light chain gene segments were observed in splenic B-cells at day 125 of gestation and serum immunoglobulins were detectable at day 145 of gestation (Saini and Kaushik 2002). The perinatal diversification occurs in the ileal Peyer's patches (Yasuda et al. 2004; Yasuda et al. 2006) but λ-light chain diversification was observed in bovine fetal spleen prior to the diversification in the ileum (Lucier et al. 1998; Aida 2014). Pre-B like cells were also noted in the fetal bone marrow and lymph node indicating B lymphopoieses in parallel to ileal Peyer's patches (Ekman et al. 2012; Aida 2014).

The generation of antibody diversity in vertebrates depends on several mechanisms that are very similar in all vertebrate species. Such processes are the recombination of separated germline gene segments (IGHV, IGHD, and IGHJ; IGLV/IGKV and IGLJ/IGKJ), the imprecise junction of these gene segments due to nucleotide deletions or additions (N, P), the combination of two identical heavy and light chains (Alt and Baltimore 1982; Kurosawa and Tonegawa 1982; Tonegawa 1983; Desiderio et al. 1984), as well as somatic hypermutations (Reynaud et al. 1995; Wagner and

Neuberger 1996; Berens et al. 1997). The available pool of the germline IGHV, IGHD, and IGHJ segments differs between species. Whereas humans and mice possess a high number of IGHV-IGHD-IGHJ genes, in livestock such as chicken (Reynaud et al. 1989), pigs (Sun et al. 1994), sheep (Dufour et al. 1996), and cattle (Berens et al. 1997; Saini et al. 1997; Sinclair et al. 1997) relatively few genes were detected resulting in a restricted combinatorial diversity. Therefore, several diversifying mechanisms dominate or additional options are employed in such species. For instance, in chicken and bovine λ-light chains gene conversion is a post-recombinatorial strategy for the generation of the preimmune antibody repertoire (Reynaud et al. 1995; Arakawa et al. 2004). Further, this mechanism is discussed in horses (Sun et al. 2010).

Today, differences in gene usage of bovine fetus and adult animals are known. In fetal IGHV-IGHD-IGHJ recombinations two IGHV are preferentially used (Saini and Kaushik 2002). Furthermore, IGHD5 and IGHD7 dominate both in bovine fetus and adult animals in contrast to IGHDQ52 (Koti et al. 2010). Within the CDRs transition nucleotide substitutions predominate over transversion mutations and in framework region three somatic hypermutations lead to isotype dependent degree of diversification (Kaushik et al. 2009).

In both bovine fetal and adult heavy chains, exceptionally long CDR3H regions contribute to higher variability, which is restricted by the use of only one VH-family (Armour et al. 1994; Saini et al. 1999; Saini and Kaushik 2002; Wang et al. 2013). Those long CDR3H regions were first described in IgM but were also detected in IgG and in all other isotypes by now and occur in 8-10% of circulating B cells (Saini et al. 1997; Sinclair et al. 1997; Saini et al. 1999; Saini and Kaushik 2002; Saini et al. 2003; Kaushik et al. 2009; Larsen and Smith 2012; Walther et al. 2013). The length of CDR3H varies by class. Consequently, the bovine CDR3H has an average length of 22.7 ± 3.2 amino acids, whereas in IgM there are 21.7 ± 1.8 and in IgG there are 18.2 ± 1.3 amino acids (Almagro et al. 2006; Kaushik et al. 2009; Aida 2014). While the CDR3H region in human and mice encodes 12 amino acid residues, in cattle it extends up to 67 amino acid residues (Sun et al. 1994; Walther et al. 2013) and up to 30 in chicken (McCormack and Thompson 1990) where this phenomenon is caused by the usage of two IGHD segments within on CDR3H region. Therefore the length of the CDR3H region is 50% greater than in human, mice, swine, and rabbit sequences (Butler 1997). The extensive size heterogeneity in CDR3H together with intrachain disulfide bridges results in a configurational diversity of this region (Saini and Kaushik

2002; Wang et al. 2013). Crystallization of such bovine antibodies revealed a unique 'stalk and knob' structure (Wang et al. 2013). Although both fetal and adult exceptionally long CDR3H originate from identical IGHV-IGHD-IGHJ rearrangements, there are structural differences. In the bovine fetus, there is no addition of CSNS at the IGHV-IGHD junction resulting in a shorter or non-existent stalk. Due to a structural support, bovine heavy chains with exceptionally long CDR3H always pair with λ-light chains that possess a conserved Ser at position 90 (Wang et al. 2013).

In contrast to human and mice bovine antibody diversity increases by somatic hypermutation, which may also occur without previous antigen contact (Tomlinson et al. 1996; Berens et al. 1997; Lopez et al. 1998). Furthermore, diversity already exists at the foetal stage as a result of somatic hypermutation.

In bovine light chains, gene conversion and somatic hypermutation are the major mechanisms leading to diversification (Parng et al. 1995; Parng et al. 1996). Initial diversification of λ-light chains was identified in the early fetal spleen before the establishment of a diverse repertoire in the ileum in cattle and sheep (Lee and Richards 1971). Gene conversion is supposed to be less significant in sheep than in cattle as there is a higher number of IGLV pseudogenes in the bovine genome (Parng et al. 1996; Ekman et al. 2009). This mechanism is also highly relevant regarding light chain diversification in chicken (Arakawa and Buerstedde 2009), as well as in rabbits using gene conversion to diversify Ig heavy chains (Weinstein et al. 1994). In addition, allotypes diversify the immunoglobulin repertoire, which has already been described in species like humans and rabbits. For instance, in humans three allotypic variants of the so called Km 1, Km 1.2, and Km 3 (Moxley and Gibbs 1992) are described. Similarly, rabbits possess five allotypic variants of the κ-light chain constant region gene called b4 (a+b), b5, b6, and b9 (Emorine et al. 1983). In cattle, allotypes have been identified for both the heavy and light chains (Bastida-Corcuera et al. 1999a; Bastida-Corcuera et al. 1999b; Diesterbeck et al. 2012; Stein et al. 2012).

References

Aida Y, Takeshima, S.-N., Baldwin, C. L. and Kaushik, A. K. 2014. Bovine Immunogenetics. in *The Genetics of Cattle* (ed. DJaR Garrick, A.). CAB International, CAB International.

Aitken R, Hosseini A, MacDuff R. 1999. Structure and diversification of the bovine immunoglobulin repertoire. *Veterinary immunology and immunopathology* **72**: 21-29.

Almagro JC, Martinez L, Smith SL, Alagon A, Estevez J, Paniagua J. 2006. Analysis of the horse V(H) repertoire and comparison with the human IGHV germline genes, and sheep, cattle and pig V(H) sequences. *Molecular immunology* **43**: 1836-1845.

Alt FW, Baltimore D. 1982. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proc Natl Acad Sci U S A* **79**: 4118-4122.

Arakawa H, Buerstedde JM. 2009. Activation-induced cytidine deaminase-mediated hypermutation in the DT40 cell line. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **364**: 639-644.

Arakawa H, Saribasak H, Buerstedde JM. 2004. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PLoS Biol* **2**: E179.

Armour KL, Tempest PR, Fawcett PH, Fernie ML, King SI, White P, Taylor G, Harris WJ. 1994. Sequences of heavy and light chain variable regions from four bovine immunoglobulins. *Molecular immunology* **31**: 1369-1372.

Arun SS, Breuer W, Hermanns W. 1996. Immunohistochemical examination of light-chain expression (lambda/kappa ratio) in canine, feline, equine, bovine and porcine plasma cells. *Zentralblatt fur Veterinarmedizin Reihe A* **43**: 573-576.

Bastida-Corcuera F, Butler JE, Heyermann H, Thomford JW, Corbeil LB. 2000. Tritrichomonas foetus extracellular cysteine proteinase cleavage of bovine IgG2 allotypes. *The Journal of parasitology* **86**: 328-332.

Bastida-Corcuera FD, Butler JE, Yahiro S, Corbeil LB. 1999a. Differential complement activation by bovine IgG2 allotypes. *Veterinary immunology and immunopathology* **71**: 115-123.

Bastida-Corcuera FD, Nielsen KH, Corbeil LB. 1999b. Binding of bovine IgG2a and IgG2b allotypes to protein A, protein G, and Haemophilus somnus IgBPs. *Veterinary immunology and immunopathology* **71**: 143-149.

Berens SJ, Wylie DE, Lopez OJ. 1997. Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. *International immunology* **9**: 189-199.

Broad TE, Burkin DJ, Cambridge LM, Jones C, Lewis PE, Morse HG, Geyer D, Pearce PD, Ansari HA, Maher DW. 1995. Six loci mapped on to human chromosome 2p are assigned to sheep chromosome 3p. *Animal genetics* **26**: 85-90.

Brown WR, Rabbani H, Butler JE, Hammarstrom L. 1997. Characterization of the bovine C alpha gene. *Immunology* **91**: 1-6.

Burg ND, Pillinger MH. 2001. The neutrophil: function and regulation in innate and humoral immunity. *Clinical immunology* **99**: 7-17.

Butler JE. 1995. Antigen receptors, their immunomodulation and the immunoglobulin genes of cattle and swine. *Livest Prod Sci* **42**: 105-121.

Butler JE. 1997. Immunoglobulin gene organization and the mechanism of repertoire development. *Scandinavian journal of immunology* **45**: 455-462.

Chen L, Li M, Li Q, Yang X, An X, Chen Y. 2008. Characterization of the bovine immunoglobulin lambda light chain constant IGLC genes. *Veterinary immunology and immunopathology* **124**: 284-294.

Chowdhary BP, Fronicke L, Gustavsson I, Scherthan H. 1996. Comparative analysis of the cattle and human genomes: detection of ZOO-FISH and gene mapping-based chromosomal homologies. *Mammalian genome : official journal of the International Mammalian Genome Society* **7**: 297-302.

Clarkson CA, Beale D, Coadwell JW, Symons DB. 1993. Sequence of ovine Ig gamma 2 constant region heavy chain cDNA and molecular modelling of ruminant IgG isotypes. *Molecular immunology* **30**: 1195-1204.

Corbeil LB, Gogolewski RP, Kacskovics I, Nielsen KH, Corbeil RR, Morrill JL, Greenwood R, Butler JE. 1997. Bovine IgG2a antibodies to Haemophilus somnus and allotype expression. *Canadian journal of veterinary research = Revue canadienne de recherche veterinaire* **61**: 207-213.

Das S, Nikolaidis N, Klein J, Nei M. 2008. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 16647-16652.

De Benedictis G CP, Dragone A. 1984. Identification of an allotypic IgA in cattle serum. *Comp Immunol Microbiol Infect Dis* **7**: 35-42.

Desiderio SV, Yancopoulos GD, Paskind M, Thomas E, Boss MA, Landau N, Alt FW, Baltimore D. 1984. Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* **311**: 752-755.

Diesterbeck US, Aboelhassan DM, Stein SK, Czerny CP. 2012. Detection of new allotypic variants of bovine lambda-light chain constant regions in different cattle breeds. *Developmental and comparative immunology* **36**: 130-139.

Dufour V, Malinge S, Nau F. 1996. The sheep Ig variable region repertoire consists of a single VH family. *J Immunol* **156**: 2163-2170.

Ehrenmann F, Kaas Q, Lefranc MP. 2010. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic acids research* **38**: D301-307.

Ekman A, Ilves M, Iivanainen A. 2012. B lymphopoiesis is characterized by pre-B cell marker gene expression in fetal cattle and declines in adults. *Developmental and comparative immunology* **37**: 39-49.

Ekman A, Niku M, Liljavirta J, Iivanainen A. 2009. Bos taurus genome sequence reveals the assortment of immunoglobulin and surrogate light chain genes in domestic cattle. *BMC immunology* **10**: 22.

Emorine L, Dreher K, Kindt TJ, Max EE. 1983. Rabbit immunoglobulin kappa genes: structure of a germline b4 allotype J-C locus and evidence for several b4-related sequences in the rabbit genome. *Proceedings of the National Academy of Sciences of the United States of America* **80**: 5709-5713.

Foley RC, Beh KJ. 1992. Analysis of immunoglobulin light chain loci in sheep. *Animal genetics* **23**: 31-42.

Ford JE, Home WA, Gibson DM. 1994. Light chain isotype regulation in the horse. Characterization of Ig kappa genes. *Journal of immunology* **153**: 1099-1111.

Granoff DM, Pandey JP, Boies E, Squires J, Munson RS, Jr., Suarez B. 1984. Response to immunization with Haemophilus influenzae type b polysaccharide-pertussis vaccine and risk of Haemophilus meningitis in children with the Km(1) immunoglobulin allotype. *The Journal of clinical investigation* **74**: 1708-1714.

Griebel PJ, Ferrari G. 1994. Evidence for a stromal cell-dependent, self-renewing B cell population in lymphoid follicles of the ileal Peyer's patch of sheep. *European journal of immunology* **24**: 401-409.

Gu F, Chowdhary BP, Andersson L, Harbitz I, Gustavsson I. 1992. Assignment of the bovine immunoglobulin gamma heavy chain (IGHG) gene to chromosome 21q24 by in situ hybridization. *Hereditas* **117**: 237-240.

Hammer DK, Kickhofen B, Schmid T. 1971. Detection of homocytotropic antibody associated with a unique immunoglobulin class in the bovine species. *European journal of immunology* **1**: 249-257.

Hayes HC, Petit EJ. 1993. Mapping of the beta-lactoglobulin gene and of an immunoglobulin M heavy chain-like sequence to homoeologous cattle, sheep, and goat chromosomes. *Mamm Genome* **4**: 207-210.

Home WA, Ford JE, Gibson DM. 1992. L chain isotype regulation in horse. I. Characterization of Ig lambda genes. *Journal of immunology* **149**: 3927-3936.

Hosseini A, Campbell G, Prorocic M, Aitken R. 2004. Duplicated copies of the bovine JH locus contribute to the Ig repertoire. *Int Immunol* **16**: 843-852.

Jackson T, Morris BA, Sanders PG. 1992. Nucleotide sequences and expression of cDNAs for a bovine anti-testosterone monoclonal IgG1 antibody. *Molecular immunology* **29**: 667-676.

Jenne CN, Kennedy LJ, McCullagh P, Reynolds JD. 2003. A new model of sheep Ig diversification: shifting the emphasis toward combinatorial mechanisms and away from hypermutation. *Journal of immunology* **170**: 3739-3750.

Kacskovics I. 2004. Fc receptors in livestock species. *Veterinary immunology and immunopathology* **102**: 351-362.

Kacskovics I, Butler JE. 1996. The heterogeneity of bovine IgG2--VIII. The complete cDNA sequence of bovine IgG2a (A2) and an IgG1. *Molecular immunology* **33**: 189-195.

Kacskovics I, Wittum TE, Butler JE, Littledike ET. 1995. The heterogeneity of bovine IgG2. VII. The phenotypic distribution of the A1 and A2 allotypes of IgG2a among beef cows with known clinical history. *Veterinary immunology and immunopathology* **48**: 89-96.

Kaushik AK, Kehrli ME, Jr., Kurtz A, Ng S, Koti M, Shojaei F, Saini SS. 2009. Somatic hypermutations and isotype restricted exceptionally long CDR3H contribute to antibody diversification in cattle. *Veterinary immunology and immunopathology* **127**: 106-113.

Knight KL, Becker RS. 1987. Isolation of genes encoding bovine IgM, IgG, IgA and IgE chains. *Vet Immunol Immunopathol* **17**: 17-24.

Knight KL, Suter M, Becker RS. 1988. Genetic engineering of bovine Ig. Construction and characterization of hapten-binding bovine/murine chimeric IgE, IgA, IgG1, IgG2, and IgG3 molecules. *Journal of immunology* **140**: 3654-3659.

Koti M, Kataeva G, Kaushik AK. 2008. Organization of D(H)-gene locus is distinct in cattle. *Developments in biologicals* **132**: 307-313.

Koti M, Kataeva G, Kaushik AK. 2010. Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies. *Molecular immunology* **47**: 2119-2128.

Kurosawa Y, Tonegawa S. 1982. Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA segments. *J Exp Med* **155**: 201-218.

Larsen PA, Smith TP. 2012. Application of circular consensus sequencing and network analysis to characterize the bovine IgG repertoire. *BMC immunology* **13**: 52.

Lee B, Richards FM. 1971. The interpretation of protein structures: estimation of static accessibility. *Journal of molecular biology* **55**: 379-400.

Lopez O, Perez C, Wylie D. 1998. A single VH family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains. *Immunological reviews* **162**: 55-66.

Lucier MR, Thompson RE, Waire J, Lin AW, Osborne BA, Goldsby RA. 1998. Multiple sites of V lambda diversification in cattle. *Journal of immunology* **161**: 5438-5444.

McCormack WT, Thompson CB. 1990. Somatic diversification of the chicken immunoglobulin light-chain gene. *Advances in immunology* **48**: 41-67.

Mestecky J. 2005. *Mucosal immunology*, Amsterdam; Boston.

Miller JR, Thomsen PD, Dixon SC, Tucker EM, Konfortov BA, Harbitz I. 1992. Synteny mapping of the bovine IGHG2, CRC and IGF1 genes. *Animal genetics* **23**: 51-58.

Mousavi M, Rabbani H, Hammarstrom L. 1997. Characterization of the bovine epsilon gene. *Immunology* **92**: 369-373.

Mousavi M, Rabbani H, Pilstrom L, Hammarstrom L. 1998. Characterization of the gene for the membrane and secretory form of the IgM heavy-chain constant region gene (C mu) of the cow (Bos taurus). *Immunology* **93**: 581-588.

Moxley G, Gibbs RS. 1992. Polymerase chain reaction-based genotyping for allotypic markers of immunoglobulin kappa shows allelic association of Km with kappa variable segment. *Genomics* **13**: 104-108.

Niku M, Liljavirta J, Durkin K, Schroderus E, Iivanainen A. 2012. The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed. *Developmental and comparative immunology* **37**: 457-461.

Pandey JP, Fudenberg HH, Virella G, Kyong CU, Loadholt CB, Galbraith RM, Gotschlich EC, Parke JC, Jr. 1979. Association between immunoglobulin allotypes and immune responses to Haemophilus influenzae and Meningococcus polysaccharides. *Lancet* **1**: 190-192.

Parng CL, Hansal S, Goldsby RA, Osborne BA. 1995. Diversification of bovine lambda-light chain genes. *Annals of the New York Academy of Sciences* **764**: 155-157.

Parng CL, Hansal S, Goldsby RA, Osborne BA. 1996. Gene conversion contributes to Ig light chain diversity in cattle. *Journal of immunology* **157**: 5478-5486.

Pasman Y, Saini SS, Smith E, Kaushik AK. 2010. Organization and genomic complexity of bovine lambda-light chain gene locus. *Veterinary immunology and immunopathology* **135**: 306-313.

Pasman YK, A. 2014. Bovine Immunoglobulin Genetics: Novel Phylogenetic Perspective. in *Comparative Immunoglobulin Genetics* (ed. AaP Kaushik, Y.). Apple Academic Press, Toronto, Canada.

Pentsuk N, van der Laan JW. 2009. An interspecies comparison of placental antibody transfer: new insights into developmental toxicity testing of monoclonal antibodies. *Birth defects research Part B, Developmental and reproductive toxicology* **86**: 328-344.

Pilstrom L. 2002. The mysterious immunoglobulin light chain. *Developmental and comparative immunology* **26**: 207-215.

Preud'homme JL, Petit I, Barra A, Morel F, Lecron JC, Lelievre E. 2000. Structural and functional properties of membrane and secreted IgD. *Molecular immunology* **37**: 871-887.

Rabbani H, Brown WR, Butler JE, Hammarstrom L. 1997. Polymorphism of the IGHG3 gene in cattle. *Immunogenetics* **46**: 326-331.

Ramsland PA, Kaushik A, Marchalonis JJ, Edmundson AB. 2001. Incorporation of long CDR3s into V domains: implications for the structural evolution of the antibody-combining site. *Exp Clin Immunogenet* **18**: 176-198.

Reynaud CA, Dahan A, Anquez V, Weill JC. 1989. Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell* **59**: 171-183.

Reynaud CA, Garcia C, Hein WR, Weill JC. 1995. Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell* **80**: 115-125.

Saini SS, Allore B, Jacobs RM, Kaushik A. 1999. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *European journal of immunology* **29**: 2420-2426.

Saini SS, Farrugia W, Muthusamy N, Ramsland PA, Kaushik AK. 2007. Structural evidence for a new IgG1 antibody sequence allele of cattle. *Scandinavian journal of immunology* **65**: 32-38.

Saini SS, Farrugia W, Ramsland PA, Kaushik AK. 2003. Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted VH + Vlambda pairings. *International immunology* **15**: 845-853.

Saini SS, Hein WR, Kaushik A. 1997. A single predominantly expressed polymorphic immunoglobulin VH gene family, related to mammalian group, I, clan, II, is identified in cattle. *Molecular immunology* **34**: 641-651.

Saini SS, Kaushik A. 2001. Origin of bovine IgM structural variants. *Molecular immunology* **38**: 389-396.

Saini SS, Kaushik A. 2002. Extensive CDR3H length heterogeneity exists in bovine foetal VDJ rearrangements. *Scandinavian journal of immunology* **55**: 140-148.

Schultz RD, Dunne HW, Heist CE. 1973. Ontogeny of the bovine immune response. *Infection and immunity* **7**: 981-991.

Shojaei F, Saini SS, Kaushik AK. 2003. Unusually long germline DH genes contribute to large sized CDR3H in bovine antibodies. *Mol Immunol* **40**: 61-67.

Sinclair MC, Gilchrist J, Aitken R. 1995. Molecular characterization of bovine V lambda regions. *Journal of immunology* **155**: 3068-3078.

Sinclair MC, Gilchrist J, Aitken R. 1997. Bovine IgG repertoire is dominated by a single diversified VH gene family. *Journal of immunology* **159**: 3883-3889.

Sitnikova T, Nei M. 1998. Evolution of immunoglobulin kappa chain variable region genes in vertebrates. *Molecular biology and evolution* **15**: 50-60.

Stein SK, Diesterbeck US, Aboelhassan DM, Czerny CP. 2012. Comparison of joining and constant kappa-light chain regions in different cattle breeds. *Animal genetics* **43**: 776-780.

Sun J, Kacskovics I, Brown WR, Butler JE. 1994. Expressed swine VH genes belong to a small VH gene family homologous to human VHIII. *Journal of immunology* **153**: 5618-5627.

Sun Y, Wang C, Wang Y, Zhang T, Ren L, Hu X, Zhang R, Meng Q, Guo Y, Fei J et al. 2010. A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. *Developmental and comparative immunology* **34**: 1009-1020.

Symons DB, Clarkson CA, Beale D. 1989. Structure of bovine immunoglobulin constant region heavy chain gamma 1 and gamma 2 genes. *Molecular immunology* **26**: 841-850.

Tobin-Janzen TC, Womack JE. 1992. Comparative mapping of IGHG1, IGHM, FES, and FOS in domestic cattle. *Immunogenetics* **36**: 157-165.

Tomlinson IM, Walter G, Jones PT, Dear PH, Sonnhammer EL, Winter G. 1996. The imprint of somatic hypermutation on the repertoire of human germline V genes. *Journal of molecular biology* **256**: 813-817.

Tonegawa S. 1983. Somatic generation of antibody diversity. *Nature* **302**: 575-581.

Tutter A, Riblet R. 1989. Conservation of an immunoglobulin variable-region gene family indicates a specific, noncoding function. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 7460-7464.

Wagner B, Miller DC, Lear TL, Antczak DF. 2004. The complete map of the Ig heavy chain constant gene region reveals evidence for seven IgG isotypes and for IgD in the horse. *Journal of immunology* **173**: 3230-3242.

Wagner SD, Neuberger MS. 1996. Somatic hypermutation of immunoglobulin genes. *Annu Rev Immunol* **14**: 441-457.

Walther S, Czerny CP, Diesterbeck US. 2013. Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins. *PloS one* **8**: e64234.

Wang F, Ekiert DC, Ahmad I, Yu W, Zhang Y, Bazirgan O, Torkamani A, Raudsepp T, Mwangi W, Criscitiello MF et al. 2013. Reshaping antibody diversity. *Cell* **153**: 1379-1393.

Weinstein PD, Anderson AO, Mage RG. 1994. Rabbit IgH sequences in appendix germinal centers: VH diversification by gene conversion-like and hypermutation mechanisms. *Immunity* **1**: 647-659.

Woof JM, Burton DR. 2004. Human antibody-Fc receptor interactions illuminated by crystal structures. *Nature reviews Immunology* **4**: 89-99.

Yasuda M, Fujino M, Nasu T, Murakami T. 2004. Histological studies on the ontogeny of bovine gut-associated lymphoid tissue: appearance of T cells and development of IgG+ and IgA+ cells in lymphoid follicles. *Developmental and comparative immunology* **28**: 357-369.

Yasuda M, Jenne CN, Kennedy LJ, Reynolds JD. 2006. The sheep and cattle Peyer's patch as a site of B-cell development. *Veterinary research* **37**: 401-415.

Zhao Y, Kacskovics I, Pan Q, Liberles DA, Geli J, Davis SK, Rabbani H, Hammarstrom L. 2002. Artiodactyl IgD: the missing link. *Journal of immunology* **169**: 4408-4416.

Zhao Y, Kacskovics I, Rabbani H, Hammarstrom L. 2003. Physical mapping of the bovine immunoglobulin heavy chain constant region gene locus. *The Journal of biological chemistry* **278**: 35024-35032.

Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS et al. 2009. A whole-genome assembly of the domestic cow, Bos taurus. *Genome biology* **10**: R42.

Exceptionally Long CDR3H Are Not Isotype Restricted in Bovine Immunoglobulins

Stefanie Walther[1], Claus-Peter Czerny[1], Ulrike S. Diesterbeck[1]*

[1]Department of Animal Sciences, Institute of Veterinary Medicine, Division of Microbiology and Animal Hygiene, Faculty of Agricultural Sciences, Georg-August University Göttingen, Burckhardtweg 2, 37077 Göttingen, Germany

*Corresponding author:
U. S. Diesterbeck
Tel.: +49 551 393375
Fax: +49 551 3913513
E-mail: udieste@gwdg.de
Postal address: Dr. Ulrike Diesterbeck
Department of Animal Sciences,
Georg-August University Göttingen
Burckhardtweg 2
37077 Göttingen, Germany

Published in "PlosOne"

DOI: 10.1371/journal.pone.0064234
URL: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0064234

Abstract

Exceptionally long third complementarity determining regions of the heavy chain (CDR3H) were previously described as a specificity of bovine IgG and IgM immunoglobulins. In addition, the genomic organization of the immunoglobulin heavy chain locus remains to be elucidated with a special focus on the number of variable segments (IGHV).

By analyzing the variable regions according to the isotype-specific PCR using cDNA-PCR, we were able to prove the existence of exceptional long CDR3H in all bovine isotypes. The corresponding sequences of three distinct amplicons were grouped according to the length of the CDR3H. Sequences of CDR3H possessed 5 to 10, 12 to 31 or at least 48 amino acid residues. Long and mid-length CDR3H were composed of mainly hydrophilic amino acid residues, while short CDR3H also contained hydrophobic amino acid residues. All sequences with long CDR3H were related to the germline variable segment 10.

Using the current genome assembly, *Bos taurus* NCBI build 6.1, the genomic organization of the bovine immunoglobulin heavy-chain locus was analyzed. A main locus was investigated on BTA21. Exons coding for variable, diversity, and joining segments, as well as for the constant regions of different isotypes, were also localized on BTA7, BTA8, and BTA20. Together with the information from unplaced contigs, 36 IGHV were detected of which 13 are putatively functional. Phylogenetic analysis revealed two bovine IGHV families (boVH1, boVH2). Thus, the existence of the two bovine families suggested was demonstrated, where boVH1 comprises all functional segments.

This study substantially improves the understanding of the generation of immunoglobulin diversity in cattle.

Introduction

The generation of antibody diversity in vertebrates is subjected to a sequence of steps such as the recombination of separated germline gene segments for both heavy (V, D, and J) and light (V and J) chains. Furthermore, the imprecise junction of the germline gene segments occurs as a result of nucleotide deletions or additions (N, P), introduced by the terminal deoxynucleotidyl transferase during the recombination process. The assembly of two identical heavy and light chains completes the tetrameric molecule [1,2,3,4]. In addition, somatic hypermutations contribute to antibody diversity – dependent or independent of antigen contact [5,6,7]. While these general processes of diversification are very similar in all vertebrate species, considerable differences were found in the available pool of the germline V, D, and J segments. Although humans and mice possess a large pool of VDJ genes [8], livestock such as chicken [9], pigs [10], sheep [11], and cattle [6,12,13] are relatively restricted in the generation of combinatorial diversity. Therefore, species-dependent mechanisms dominate the different diversification steps or additional options are employed. For instance, in chicken gene conversion, the use of pseudogene sequences is a frequent post-recombinatorial strategy for the generation of the preimmune antibody repertoire [5,14]. This mechanism was confirmed for λ-light chains in cattle [15] and is discussed in horses [16].

All heavy-chain isotype classes detected in other mammals were also described for cattle [17,18], whereas the γ-isotype encompasses three sub-classes, namely γ1, γ2, and γ3 [18,19]. The bovine IGH locus was assigned to the *Bos taurus* autosome (BTA) 21 [20] and localized on the q23-q24 bands [21] or on the q24 band respectively [22,23]. An IgM-like chain was assigned to BTA11q23 by hybridization [24,25], which was supported by the detection of six IGHJ segments on the same chromosome [26]. By screening a bovine BAC and Cosmid library, the genomic organization of the IGHC locus was described, as well as the number of the preceding joining segments (IGHJ). Only two out of six IGHJ were classified as functional – of which only one seems to be involved predominantly in the recombination process [21,26]. The IGHV itself codes for the complementarity determining regions 1 and 2 (CDR1H, CDR2H) and for the N-terminal part of the complementarity determining region 3 (CDR3H). Bovine IGHV offer a restricted set of genes related to one family (boVH1), which shares homologies to the murine Q52 family and human VHII family. Southern blot analyses indicated one additional IGHV family in the germline repertoire but only expression of boVH1 has

been observed yet [6,12,13,27,28]. The definite number and organization of IGHV remains under further investigation.

Another peculiarity is the organization of the bovine IGHD locus. Ten IGHD genes classified into four families are organized in sub-clusters [29,30]. A comparison of the IGHD exons revealed huge size differences [29]. Cattle antibodies provide exceptionally long CDR3H consisting of up to 62 amino acid residues (aa) [6,31,32,33,34,35]. IGHD2, with 148 bp in size, contributes to those CDR3H and encodes the characteristic hydrophilic Glycine and Tyrosine residues [6,29,36]. The high number of Cysteine residues detected is supposed to promote intra-CDR3H disulfide bonds [13]. Mid-length CDR3H – containing one to three Cys residues – were almost always accompanied by one Cys residue found in the CDR2H, which may result in intra CDR disulfide bond formation [31,37]. The germline encoded IGHV, IGHD, and IGHJ and their imprecise junction during rearrangement cannot fully explain the remarkable length of the CDR3H. Conserved short nucleotide sequences of 13 to 18 nucleotides are specifically inserted into the IGHV and IGHD junction, leading to a further extension of the CDR3H. This mechanism is unique for cattle [30].

To date, these exceptionally long CDR3H have been attributed exclusively to the γ1-3- and μ-isotype [33,35]. In our study, we demonstrate the expression of exceptionally long CDR3H in all bovine immunoglobulin isotypes. We were able to observe three distinct groups of CDR3H sizes, which were related to their genomic origin. Loci of IGHV were determined on BTA7, BTA21 and seven unplaced contigs.

Materials and Methods

In silico Analysis of the Bovine IGHV Segments

A sequence search was performed with blastn on *Bos taurus* in the Reference genomic sequences (refseq_genomic) database using the leader and variable region of one mRNA sequence (accession number AY145128). On the identified contigs, the IGHV and their respective leader were annotated together with the recombination signal sequences (RSS). The octamers, TATA boxes, and splicing sites were also noted. The nucleotide sequences of the bovine IGHD1 to 8 and Q52 [30], as well as all IGHJ coding sequences (AY158087, AY149283), were used for a similar alignment approach. The detected IGHV, IGHD, and IGHJ were used in the further analysis of the amplified immunoglobulin sequences. To annotate the constant region locus, IgM (U63637), IgD (AF411240, AF515672), IgG1 (X16701), IgG2 (S82407, X16702), IgG3 (U63638), IgE (AY221098), and IgA (AF109167) bovine coding sequences were applied. Missing transmembrane regions were determined in bovine ESTs (expressed sequence tags). Based on the available sequence data, functionality was defined according to Lefranc [38]. In brief, functional sequences exhibited an open reading frame (ORF) without stop codon, and no defects in the splicing sites, RSS, or in the regulatory elements. If sequence information was missing due to end of contigs or N's introduced in the sequence but the available sequence offered putative functionality, genes were marked with (F). Classification to ORF included either alterations in the splicing site, RSS, regulatory elements, substitutions of conserved amino acid residues (Cys23, Trp41, Leu89, Cys104 within IGHV or IGHC and a Phe/Trp118-Gly119-X120-Gly121 motif within IGHJ [39]) or orphons ((ORF)). In this case, orphons are located outside of BTA21 [20,21,22,23]. Pseudogenes (Ψ) were characterized by the presence of stop codons or frameshifts. Fragmented loci were also defined as pseudogenes. Functional recombination assays revealed the spacer lengths, the first three nucleotides of the heptamer as well as three consecutive adenosine residues within the nonamer to be crucial for efficient recombination [40,41].

For the purpose of phylogenetic analysis, the complete nucleotide sequences of bovine IGHV segments were aligned with one member of the human IGHV families 1 to 7, respectively, using the ClustalW algorithm with the ClustalX 2.1 interface [42]. The phylogenetic tree was calculated using the neighbor-joining method, with the exclusion of gaps. The confidence values were compiled with 1000 bootstrap replicates [43]. To

root the tree, the sequence of one IGHV segment of the horned shark (accession number X13449) and little skate (X15124) were defined as an outgroup, similar to the method performed by Sitnikova and Su [44] and Almagro et al. [45]. Visualization of the phylogenetic tree was performed using the program NJplot [46].

Ethical Statement

To collect B-lymphocytes, 20 ml of EDTA blood were taken from the tail vein of a German Simmental bull kept by the Division of Microbiology and Animal Hygiene for demonstrations in claw-treatment within student courses and to study the clinical development of *Mycobacteria avium spp. paratuberculosis* infection. The bull was owned, because he had acquired a natural infection of MAP and showed positive antibody-titers already with an age of 18 months. Similar to other cattle herds, he has to be tested for cattle diseases periodically. The blood sample was taken from the tail vein during regular investigation of infectious diseases in the bull. The plasma was applied e.g. in an indirect ELISA testing for antibodies against *Mycobacteria avium spp. paratuberculosis* or BHV-1. Therefore, no specific approval is required.

Isolation of PBMCs and cDNA Synthesis

Peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll gradients (GE Healthcare Europe GmbH, Germany) according to the manufacturer's protocol. Viable B cells were stained with trypan blue and counted. Total RNA was isolated from $1x10^7$ cells using the RNeasy® Mini Kit (Qiagen, Germany). The first-strand cDNA was synthesized using $pd(N)_6$ primers from 3 µg of total RNA in a total volume of 20 µl (SuperScript™III First-Strand Synthesis SuperMix, Life Technologies GmbH, Germany).

Amplification of Immunoglobulin Heavy-chain Isotype Restricted Variable Regions

To amplify the variable region restricted to each isotype, a primer set was generated with one primer hybridizing in the leader region, and individual primers with binding sites in the constant region (CH) of the immunoglobulin heavy chains. The primers were based on database entries and their own sequence information (data not shown). For α, γ1-3, and ε isotypes primers anneal to the CH1. For δ and µ isotypes, primers bind within the CH2 (Table 1). To monitor the integrity and purity of the cDNA, 527 bp

of the bovine GAPDH (Glycerinaldehyde 3-phosphate dehydrogenase) were amplified as a positive control. A no template control served as a negative control for the PCR.

Table 1. Primer for the isotype-specific amplification of the complete variable regions.

Forward primer	Reverse primer	Primersequence 5'-3'	Approximated product size in bp
BoLH_BACK		ACCCACTGTGGACCCTCCTC	
	BoIgMCH2_FOR	TGCCGTCACCAGAGAGGCTGT	795
	BoIgDCH2_FOR	TGCGTGCTGACCGCCTTGTT	805
	BoIgG1-3CH1_FOR	GGCACCCGAGTTCCAGGTCA	536
	BoIgECH1_FOR	GCCCAGCCTTACACGGGCTT	467
	BoIgACH1_FOR	GCCAGCACGGCAGGGAAGTT	574
GAPDH_for		TGGTCACCAGGGCTGCT	
	GAPDH_rev	GGAGGGGCCATCCACAGTCT	527

One universal forward primer was used for annealing within the leader region. For each isotype, a reverse primer was generated for specific amplification. The annealing sites were selected in the first constant region (IGCH1), with the exception of IgM and IgD. Both isotypes share high homologies in the IGCH1 and therefore, specific reverse primers were generated for binding in the second constant region. The IgG subtypes were not distinguished further. Primers for bovine GAPDH served as cDNA quality control.

The total reaction volume of 50 µl included 1 µl of cDNA, 200 µM dNTPs (Bioline, Germany), 5 µl of 10x PCR buffer (75 mM Tris HCl pH 9.0; 2 mM $MgCl_2$; 50 mM KCl; 20 mM $(NH_4)_2SO_4$), 5% DMSO (Dimethyl sulfoxide), 0.4 µM of each primer pair, and 2 units of DNA polymerase (Biotools, Spain). PCR was performed under cycling conditions of 95°C for 5 min, followed by 35 cycles of 95°C for 1 min, 58°C for 1 min, 72°C for 2 min, and terminated with elongation at 72°C for 10 min. The length and purity of the PCR products were evaluated by means of electrophoresis on 1% agarose gels.

Cloning and Sequencing of the PCR Products

The PCR products were purified and concentrated using the DNA Clean & Concentrator Kit (Zymo Research, USA). Purified products were cloned into the pCR® 2.1-TOPO® 3.9 Kb TA vector (Invitrogen™, Karlsruhe, Germany) and transformed into chemically competent One Shot TOP10 *E. coli* cells (Invitrogen™, Karlsruhe, Germany). Transformants were plated on LB agar containing 0.3 mM ampicillin, 40 µl 2.44 µM X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside), and 40 µl 1 M IPTG (Isopropyl β-D-1-thiogalactopyranoside) for blue/white selection. After incubation at 37°C, overnight cultures of randomly selected white transformants were grown in a 5 ml LB-ampicillin broth. Plasmids were isolated using the MiniPrep Kit (Qiagen,

Germany). In order to assess the insert size, plasmid DNA was cleaved with *Eco*RI (New England Biolabs, Germany) and DNA sizes were confirmed by agarose gel electrophoresis.
Twenty plasmids of each PCR product were sequenced according to the chain-termination method [47]. The M13 (-20) Forward and M13 Reverse (Invitrogen, Germany) vector-specific primers, as well as the corresponding gene specific primers, were used for sequencing.

Nucleotide and Amino Acid Sequence Analyses

The genetic information of the VDJ recombinations was used for further analysis. The amplified part of the constant regions served as verification of the respective isotype. The sequences were analyzed using the DNAStar program (GATC Biotech AG, Germany) and aligned by ClustalW [48].
The deduced amino acid residues of the variable parts were aligned to the IMGT nomenclature [39] using the IMGT/DomainGapAlign [49,50]. Framework regions, as well as CDRs, were identified and analyzed with regard to their biochemical properties such as the hydrophobicity, polarity, and charge of the amino acid residues incorporated. The CDR3H regions were classified according to their number of amino acid residues. The amino acid compositions of CDR2H and CDR3H were examined for their numbers of Tyr, Gly, aromatic amino acid residues and Cys, since some bovine CDR3Hs are characterized by exceptional length and preferred amino acid residues.

Results

Annotation of the Bovine Germline Immunoglobulin Heavy-chain Locus

For the identification of germline IGHV, a search using blastn on all bovine genome assemblies was performed. The contigs NW_003104530.1 and NW_003104538.1 were identified on *Bos taurus* chromosome 21 (BTA21; AC_000178.1). NW_003104530.1 was located at the centromeric region with two IGHV (IGHV1Ψ and IGHV2). A region of about 146 kb on NW_003104538.1, located at the telomeric region, comprised eight IGHV: IGHV3, IGHV4Ψ, IGHV5Ψ, IGHV6, IGHV7Ψ, IGHV8Ψ, IGHV9Ψ, and IGHV10. Upstream of them, the exons coding for the μ, ε and α chains, were identified. Two IGHD (8 and 4) were localized between IGHV6 and IGHV7Ψ (Figure 1A).

The contigs NW_003064289.1, NW_003064290.1, NW_003064296.1, NW_003064297.1, NW_003064298.1, and NW_003064299.1 were localized to the centromeric region on BTA7 (AC_000164.1) involving nine IGHV (IGHV11Ψ, IGHV12Ψ, IGHV13Ψ, IGHV14Ψ, IGHV15Ψ, IGHV16(ORF), IGHV17(ORF), IGHV18Ψ, and IGHV19Ψ). This cluster spans approximately 144 kb. Downstream on BTA7, a cluster of 1131 bp, with exons encoding IGHD1(ORF) to IGHD3(ORF), was identified on contig NW_003064411.1, followed by a pseudogene for a δ chain. The latter one is characterized by fragmented exons 1 and 2, a deleted exon 6 (codes for the secretory region), and frame shifts (Figure 1B).

A third location comprising five IGHD segments, a δ chain pseudogene, a μ chain gene, and six IGHJ segments in about 43 kb was detected on BTA8. Two loci for the IGHD were observed. IGHDQ52(ORF) was the most downstream segment on contig NW_003066919.1, whereas IGHD4(ORF) to 7(ORF) were found to be the most upstream on contig NW_003066918.1 (Figure 1C). Interestingly, BTA20 revealed the genomic information for a γ2 chain (Figure 1D).

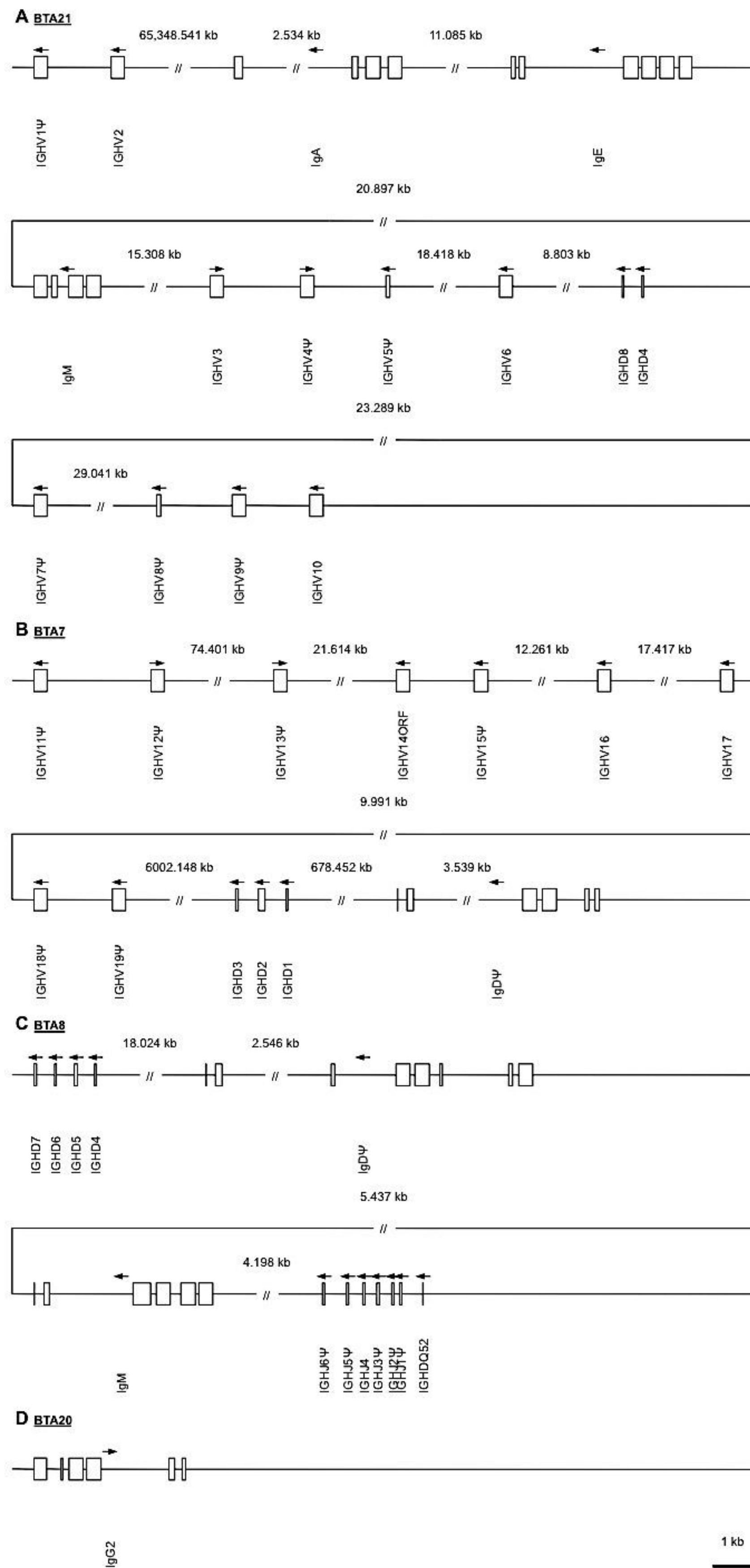

A BTA21
65,348.541 kb
2.534 kb
11.085 kb
IGHV1Ψ
IGHV2
IgA
IgE
20.897 kb
15.308 kb
18.418 kb
8.803 kb
IgM
IGHV3
IGHV4Ψ
IGHV5Ψ
IGHV6
IGHD8
IGHD4
23.289 kb
29.041 kb
IGHV7Ψ
IGHV8Ψ
IGHV9Ψ
IGHV10
B BTA7
74.401 kb
21.614 kb
12.261 kb
17.417 kb
IGHV11Ψ
IGHV12Ψ
IGHV13Ψ
IGHV14ORF
IGHV15Ψ
IGHV16
IGHV17
9.991 kb
6002.148 kb
678.452 kb
3.539 kb
IGHV18Ψ
IGHV19Ψ
IGHD3
IGHD2
IGHD1
IgDΨ
C BTA8
18.024 kb
2.546 kb
IGHD7
IGHD6
IGHD5
IGHD4
IgDΨ
5.437 kb
4.198 kb
IgM
IGHJ6Ψ
IGHJ5Ψ
IGHJ4
IGHJ3Ψ
IGHJ2Ψ
IGHJ1Ψ
IGHDQ52
D BTA20
IgG2
1 kb

Figure 1. Chromosomal organization of variable (IGHV), diversity (IGHD), joining (IGHJ) segments, and the constant regions of the heavy chains. The physical map displays the order of functional segments (F), pseudogenes (Ψ), and open reading frames (ORF). Classification to "functional" includes an ORF without stop and exhibition of conserved amino acid residues as well as no defects in splicing signals, recombination signal sequences (RSS) or regulatory elements. ORF are defined by alterations in the splicing signals, recombination signal sequences, and/or regulatory elements. In addition, changes to conserved amino acid residues, which may lead to misfolding were included in the ORF classification. Functional elements on orphon localizations are highlighted with ORF in parenthesis (ORF) [38,39]. Pseudogenes possessed stop codons, frameshifts or mutations of the spacer lengths within the first three nucleotides of the heptamer as well as in three consecutive adenosines residues within the nonamer abolish the recombination [40,41]. In addition, fragmented loci were also defined as pseudogenes. Arrows indicate the transcription direction.

In addition, unplaced contigs (NT_182448.1, NT_182449.1, NT_183109.1, NT_185036.1, NT_185907.1, NT_186922.1, and NW_003100762.1) were discovered to contain IGHV segments. Likewise, IGHD1(ORF) and 2 genes were localized on NT_186153.1. NW_001494075.1 includes a gene coding for IGHDQ52 downstream of an IGHJ1 to 6 locus. Genes coding for a μ chain and a δ chain pseudogene were found most upstream in this contig. The contig NW_001503306.1 comprises the genomic information for IGHD4 to 7 and a δ chain pseudogene while IGHD4 and 8 were detected on NW_001504477.2. A locus involving IGHJ4ORF to 6ORF and a putative functional δ chain gene was detected on NT_186572. Additional genomic information for α and ε chains were discovered on NT_185723.1. A genomic order of γ3, γ2, and γ1 was identified on NT_185580.1. Furthermore, a gene coding for γ1 and γ3 was found on NW_003100065.1 and NW_003099305.1, respectively. Along with NW_003100387.1, including a δ chain pseudogene locus, a putative functional δ chain gene was noted on NW_003100112.1.

Thirteen out of the 36 IGHV segments identified are putatively functional (Supplemental Table S1). Eleven IGHV segment pairs shared sequence identity of 100%, namely IGHV3/33, IGHV10/34, IGHV9Ψ/35Ψ, IGHV4Ψ/32Ψ, IGHV7Ψ/22Ψ, IGHV2/26, IGHV1Ψ/27Ψ, IGHV18Ψ/30Ψ, IGHV16(ORF)/25, IGHV14Ψ/23Ψ, and IGHV36/29(F). Since the human IGHV locus has been fully investigated [51], one member of each family was chosen for phylogenetic analysis. The sequences clustered into two distinct branches either with huIGHV2-05 or with huIGHV4-04 and huIGHV6-1. All functional IGHV were phylogenetically related to huIGHV2-05 (Figure 2). This group corresponds to the bovine IGHV family 1 (boVH1) described previously [6,12,13]. Multiple alignments revealed two distinct families with identities of at least 80.5% for boVH1 and 79.5% for boVH2. IGHV5Ψ and IGHV8Ψ represent fragmented loci consisting of 77 bp. They share the highest identity of 70.1% with IGHV1Ψ,

IGHV18Ψ, IGHV27Ψ, and IGHV30Ψ which are members of boVH2. The highest identity between members of both families was calculated to be 69.7%.

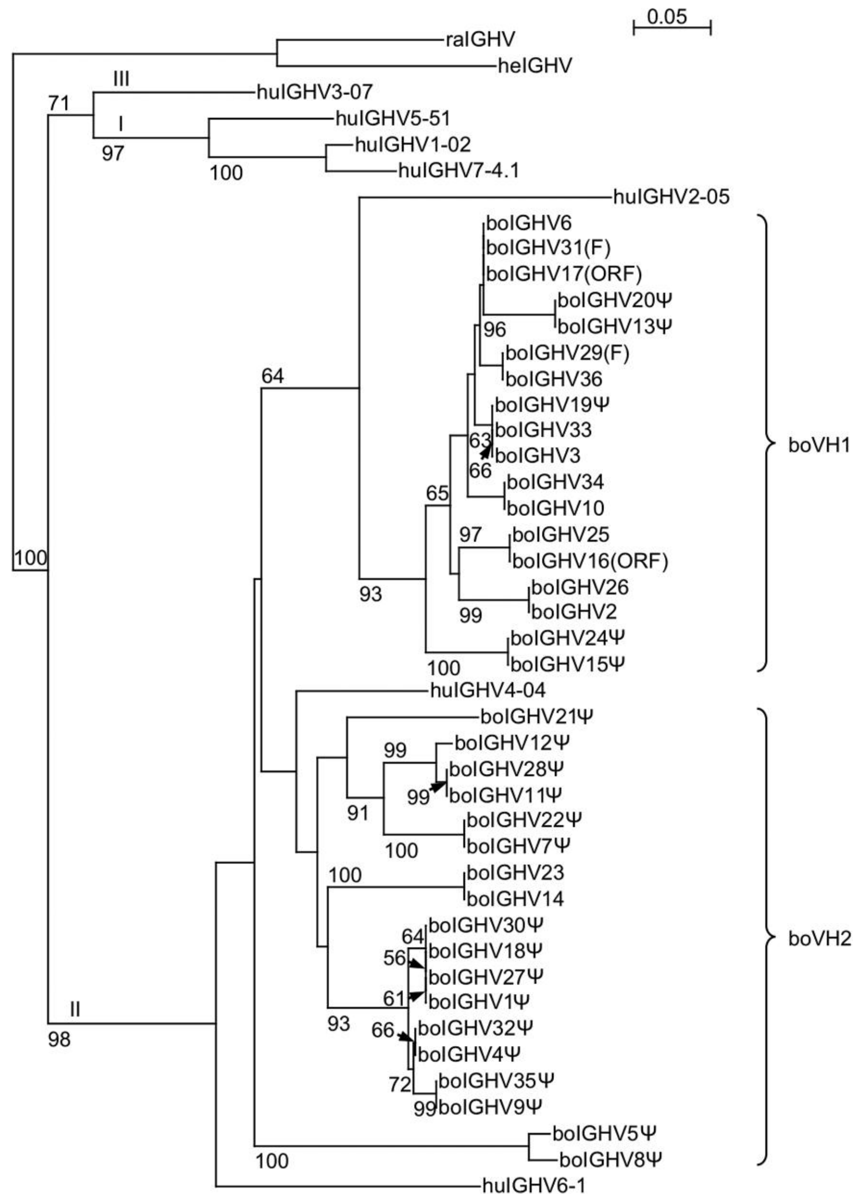


Figure 2. Neighbor-joining phylogenetic tree of the genomic bovine IGHV segments. The complete sequences of the bovine IGHV segment (boIGHV), and one representative sequence of each human family (huIGHV1 to huIGHV7), were used for the comparison. The reliability of the tree was estimated using 1000 bootstrap replicates [43]. Numbers at each node are the percentage bootstrap value and are indicated only when greater than 50%. Arrows mark the respective node. The Roman numerals I, II, and III describe the clans [64]. Two clusters of bovine IGHV were visible and corresponded to two families. The bovine IGHV family 1 (boVH1) comprises all functional segments, whereas boVH2 consists only of pseudogenes. IGHV5Ψ and IGHV8Ψ present fragmented loci, which consist of only 77 bp. They share 70.1% sequence identity with IGHV1Ψ, IGHV18Ψ, IGHV27Ψ, and IGHV30Ψ. We would therefore propose to assign IGHV5Ψ and IGHV8Ψ to boVH2. Horned shark (heIGHV from accession number X13449) and little skate (raIGHV; X15124) represent the outgroup in this analysis, similar to that performed by Sitnikova and Su [44] and Almagro et al. [45]. The scale bar indicates the number of nucleotide substitutions per site.

Amplification of Isotype-specific Variable Regions and Sequence Analyses

The immunoglobulin heavy chains were amplified by PCR for each bovine isotype. Three distinct bands became visible following agarose gel electrophoresis of the amplicons of the μ, δ, γ1 to 3, ε, and α isotypes. The expected product sizes of 467 bp (IgE) to 805 bp (IgD) corresponded with the lowest band. The isotypes γ1 to 3, ε, and α revealed the lowest band, always approximately 100 bp below the middle band, which was again 100 bp smaller than the largest band (Figure 3). The dominant product was observed in the middle band. In contrast, a faint lower band was noticed in the products of IgM and IgD – also with differences in size of 100 bp. The middle and upper bands showed strong amplification. The three bands observed per amplified isotype should allow for grouping of the respective CDR3H lengths. After sub-cloning of the purified products, 20 sequences per isotype were evaluated.

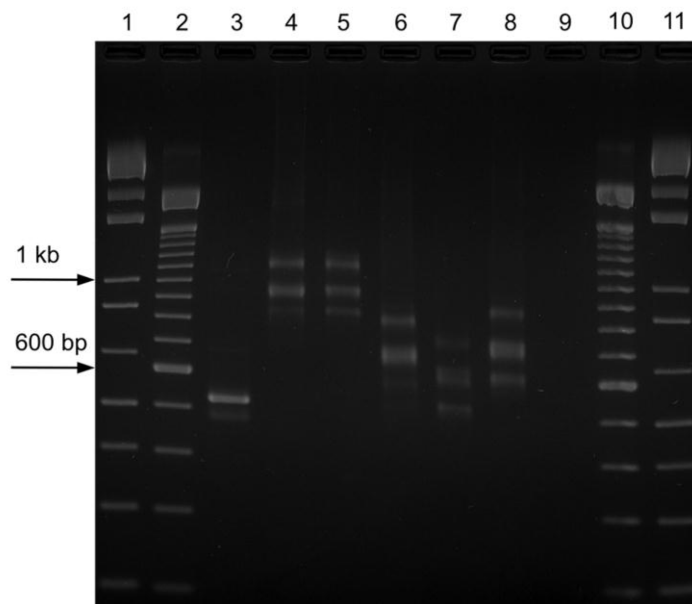


Figure 3. PCR products of the bovine μ, δ, γ1 to 3, ε, and α isotypes.
The amplicons of the heavy chain variable domain of each bovine isotype were resolved by 2.5% agarose gel electrophoresis and revealed three distinct products. Lanes 1 and 11: 1kb ladder, lanes 2 and 10: 100 bp ladder. Lane 1: product of the GAPDH positive control; Lanes 4 and 5: in the products of μ and δ, a faint lower band was noticed also with differences in size of 100 bp. The middle and upper bands showed strong amplification. Lanes 6 to 8: the dominant product of isotypes γ1 to 3, ε, and α was observed in the middle with a size difference of about 100 bp compared to both the upper and the lower band. Lane 9 is the no template negative control of the PCR.

The deduced amino acid sequences were aligned to the IMGT nomenclature for variable domains [39]. The size of the variable regions varied between 111 and 173 amino acid residues (Table 2). The Cysteine residues, forming the intra-chain disulfide bond, were always observed at position 23 and 104. Trp41 and Leu89 were also conserved. These amino acid residue positions are coded by the V segment [39]. A Trp at position 118 and subsequently a Gly119, coded by the J segment, were also found in most sequences analyzed. In four sequences, amino acid residue substitutions were observed at these positions. Two SNPs resulted in the replacement of Gly119 by Asp (SNP: GGC>GAC, in one sequence) or Ser (SNP: GGC>AGC, in two sequences). In one sequence, we found Trp118 substituted by Ser, which was caused by the SNP TTG>TCG.

The CDR1H always comprised eight amino acid residues, whereas CDR2H possessed seven amino acid residues. They consisted of both hydrophobic and hydrophilic amino acid residues.

Table 2. Analysis of the complement determining regions (CDR) of the different isotypes.

Isotype	Length of the variable region	Length of the CDR1H	Length of the CDR2H	Length of the CDR3H
IgM	131.65 ±14.62	8 ±0	7 ±0	25.65 ±14.62
IgD	138.70 ±17.54	8 ±0	7 ±0	32.70 ±17.54
IgG1-3	138.70 ±19.39	8 ±0	7 ±0	32.70 ±19.39
IgE	136.55 ±17.97	8 ±0	7 ±0	30.55 ±17.97
IgA	127.05 ±12.17	8 ±0	7 ±0	21.05 ±12.17

The average amino acid residue lengths of the complete variable region and the CDR regions, according to the IMGT nomenclature [39], are provided with their standard deviations. The size of the variable regions ranged from 111 to 173 amino acid residues. In total, 20 sequences per isotype were analyzed.

Composition of the CDR3H with Different Lengths

The three distinct sizes of the CDR3H were noted in all bovine immunoglobulin isotypes. A specific composition of amino acid residues was found within the CDR3H, which seems to correlate with the different lengths of CDR3H. The shortest group (group 1) – comprising five to ten amino acid residues according to Lopez et al. [27] – was characterized by hydrophobic, as well as hydrophilic amino acid residues, without Cysteines (Figure 4). Twelve to 31 amino acid residues formed the middle size (group 2). Within this group, hydrophilic amino acid residues were incorporated predominantly into the CDR3H. The Gly, Tyr, and Cys-rich, long CDR3H included more than 47 amino acid residues (group 3). Gly and Tyr were frequently found within the CDR3H, whereas Gly dominated in most sequences analyzed. Within the exceptionally long CDR3H, four, six, seven or eight Cys were detected, which accumulated in the middle of the CDR3H. Group 1 and 2 only possessed none, one, or two Cys. In two sequences of IgE, with one Cys in the mid-length CDR3H, one additional Cys was found in CDR2H. Gly, Pro and multiple Cys, as well as aromatic residues, were observed in long and intermediate CDR3H.

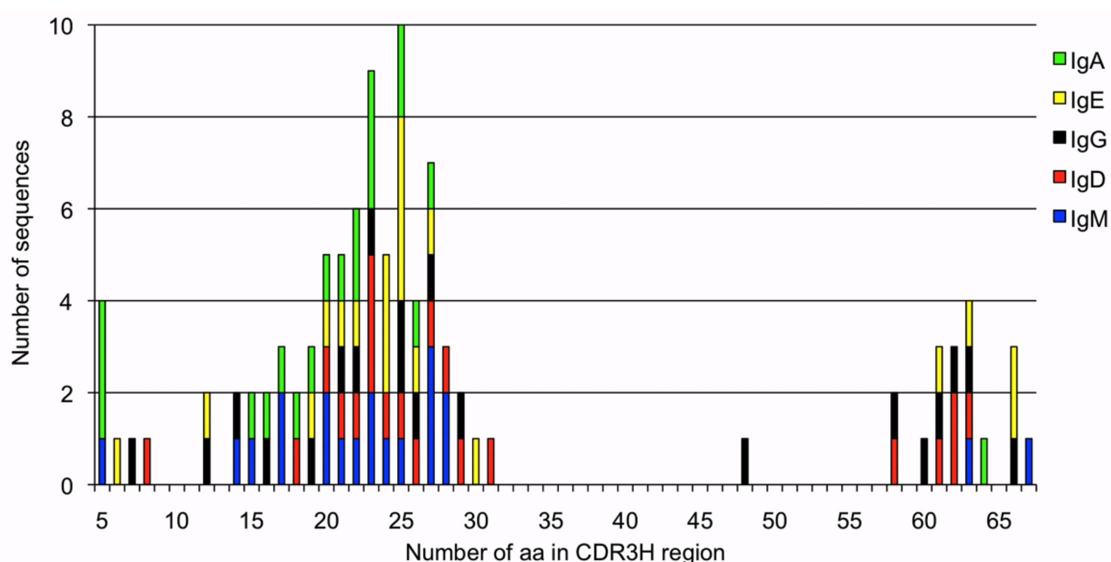


Figure 4. Three distinct sizes of CDR3H in all isotypes. The lengths of CDR3H of all sequences analyzed was divided into three groups. These groups were identified in each isotype, marked in different colors. The numbers of amino acid residues (aa) in CDR3H are indicated on the horizontal axis, whereas the number of sequences possessing each number of aa is provided on the vertical axis.

The IGHV, IGHD, and IGHJ segments were related to their genomic origin by phylogenetic alignment. In general, IGHJ1 was related to all sequences and heavy-chain isotypes were analyzed. Although IGHJ2 is also functional, we did not identify this segment. In addition to the exclusive use of one IGHJ segment, six IGHV segments were recombined. For group 1, segment IGHV3 was detected in six out of seven sequences. One sequence showed IGHV6. The segments IGHV2 (1 sequence), IGHV3 (5 sequences), IGHV6 (18 sequences), IGHV10 (6 sequences), IGHV17(ORF) (35 sequences), and IGHV36 (9 sequences) were identified within group 2. These IGHV segments were distributed between the isotypes in similar proportions. Only the segment IGHV10 was identified in all group 3 sequences. The phylogenetic alignment of IGHD segments revealed dissimilarities between recombined and genomic sequences, which resulted in low sequence identities. In addition, there were homologies between the germline IGHD segments ranging from 81.4% (IGHD2 and IGHD8) up to 97.6% (IGHD1 and IGHD6). Therefore, it was not possible to unambiguously annotate all CDR3H regions of the sequences analyzed. Nevertheless, IGHD2 was preferentially recombined in sequences possessing exceptionally long CDR3H. At the IGHV-IGHD junctions, conserved short nucleotide sequences (CSNS) were identified, which are rich in A nucleotides. The inclusion of N and P nucleotides was noted within the IGHD-IGHJ junction. In addition to the presence of CSNS in sequences possessing very long CDR3H, these CSNS were also discovered in sequences with CDR3H of mid-length. IGHD4 was used most often in these group 2 sequences. All other known IGHD segments were recombined less frequently in this group. The short IGHD segments IGHDQ52 and IGHD4(1 and 2) – possessing 15 and 36 nucleotides – were identified in very short CDR3H regions.

Discussion

The *in silico* analyses of the genomic organization of the heavy-chain locus revealed differences from previous mapping and annotation results. The functional locus was mapped to BTA21q23-q24 [21,22,23], where we also detected ten variable and two diversity segments together with the exons coding for IgM(F), IgE, and IgA(F). The order of the respective segments highlighted deviations from other fully described mammalian loci [16,52]. Zhao et al. [21] have already described the constant heavy-chain locus by means of BAC clone analysis, as well as the joining segments organized upstream of IgM. These BAC clones were not introduced into the genomic assembly. Since the UMD2 assembly used mapping data, synteny with the human genome, and paired-end sequence information [53], we did not expect heavy-chain loci on BTA7, BTA8, and BTA20. Upstream of IGHV11Ψ, the bovine contig-NW_003064289 on BTA7-harbors the genes for FBXL12 (F-box and leucine-rich repeat protein 12), UBL5 (ubiquitin-like 5), PIN1 (peptidylprolyl cis/trans isomerase, NIMA-interacting 1), and OLFM2 (olfactomedin 2) which share homology with HSA19. The human gene following downstream is COL5A3 (collagen, type 5, alpha 3). No variable segment was found between OLFM2 and COL5A3 on HSA19p13.2. The bovine pseudo δ-chain locus on BTA7 revealed no human equivalent. The human 5pter part is inversely syntenic to BTA20qter [54]. On contig NW_003104522.1, we found the gene ADCY2 (adenylat cyclase 2) as an anchor. The bovine IgG2(ORF) (IGCGAMMA) gene, annotated between AHRR (aryl-hydrocarbon receptor repressor) and PDCD6 (programmed cell death 6), has no homologue on the HSA5pter, although it has been noted that AHRR and PDCD6 overlap on HSA5pter. For IGHD4 to 7, and pseudo δ-chain on BTA8, no specific genes were identified. Human syntenic groups on BTA8 were described for HSA8p and HSA9 [54]. No additional genes were determined on contig NW_003066919, with a complete IgDΨ, IgM(ORF) gene and IGHJ1(ORF) to 6(ORF), as well as IGHDQ52(ORF). Hybridization investigations assigned an IgM-like chain (probe IGHML1) to the syntenic group U16; corresponding to HSA9q [24,55]. Later, IGHML1 was assigned to BTA11q23 by hybridization [24,25], which was supported by the detection of six IGHJ segments on the same chromosome using BAC clone and locus-specific PCR analysis [26]. We were not able to identify an IgM-like locus on BTA11. Based on these results and the fact that we observed a transcribed IGHV from a putative orphon, we concluded an incorrect and incomplete annotation of the bovine immunoglobulin heavy-chain locus,

which may be solved by the re-sequencing of the described localizations and underpinned by different authors and methods.

Previous studies have already classified bovine IGHV segments into clan II, with the closest homology to the human VH2 family [6,12,13,44]. Quite in contrast to the statement made by Tutter and Riblet [28] and Berens et al. [6], none of the genomic sequences clustered with the human clan III family VH3. All of the functional bovine IGHV segments are most closely related to the human VH2 family represented by IGHV2-05, which explains the exclusive transcription of only one bovine VH family [6,13,56]. The comparison with human sequences was performed by Saini et al. [12] using one VH4 family sequence only. The second bovine VH family described here consisted only of pseudogenes and clustered with the human VH4 and VH6 family. We were now able to describe the second bovine VH family (boVH2) previously proposed [6], but we had no indications of the possible gene conversions using boVH2 segments in the sequences investigated, as shown for the bovine λ-light chains [15]. We would propose that IGHV5Ψ and IGHV8Ψ should be assigned to the boVH2 family as they are fragmented loci and share 70.1% sequence identity with IGHV1Ψ, IGHV18Ψ, IGHV27Ψ, and IGHV30Ψ.

With regard to the description of unusually long CDR3H in bovine IgM, long and short IGHD segments have already been described. Independent of nucleotide addition during rearrangement they contribute directly to CDR3H length heterogeneity. Nevertheless, the genomic IGHD segments showed high homologies among themselves, which resulted in the complicated annotation of the transcribed IGHD. In particular, many hypermutations within the recombined IGHD segments led to low sequence identities. Intrinsic hot spots as targets for somatic hypermutations within CDR1H, CDR2H, and CDR3H were already found in a bovine fetus. Furthermore, CDR3H length heterogeneity, junctional flexibility, and somatic hypermutation are thought to contribute solely to IgM antibody diversification in both bovine fetus and adult cattle [32]. As we found CDR3H length heterogeneity in all isotypes, exceptionally long CDR3H are apparently not primarily generated to compensate the restricted flexibility of IgM caused by reduced Pro numbers within Cμ2. The most recent study on the IgG repertoire in calves also showed very long CDR3H [35]. Therefore, antigen selection of variable domains and class switch recombination seem to be of higher impact. We did not observe any evidence to suggest that the combination of two different IGHD segments enhances diversity, which was the case for horses [16]. The exceptionally long CDR3H were generated by the direct fusion of a single IGHV

segment (IGHV10), the longest IGHD segment (IGHD2), and one functional IGHJ segment (IGHJ1), as described previously [30]. In very short CDR3H, we noted a preferred use of the short IGHD segments, IGHDQ52 and IGHD4. In sequences of CDR3H group 1 and 2, no predominant use of one special IGHV segment was determined. According to structural analyses, spatial distances are not thought to contribute to preferred IGHV-IGHD-IGHJ rearrangements, as there are conformational changes of chromatin resulting in the repositioning of the IGHD cluster and the merging of proximal and distal IGHV regions during early B cell development [57]. All IGHV segments identified were found to be functional. Thus, there is no evidence for gene conversion in bovine immunoglobulin heavy chains, which is already known to contribute to the diversity of chicken immunoglobulin heavy chains and bovine λ-light chains [15,58]. Conserved short nucleotide sequences (CSNS), which were inserted into the IGHV-IGHD junction, were found both in intermediate length and very long CDR3H. This novel mechanism, which contributes to antibody diversification, is neither restricted to immunoglobulin heavy chains with exceptionally long CDR3H, nor is it isotype restricted. As the insertion of CSNS is supposed to directly follow antigen exposure during the development of an immune response [30], we conclude class switch recombination to be responsible for isotype-independent, long CDR3H in cattle. In addition, exceptionally long CDR3H protrude from the variable domain with support from the λ-light chains. Thus, there is no conventional combining site and the other two CDRH do not contribute to antigen binding. Instead, this function is undertaken by side chains that are exclusively contained within long CDR3H regions, as investigated by structural comparisons with protein toxins [37].

Transcribed CDR1H, CDR2H, and short CDR3H sequences showed both hydrophilic and hydrophobic amino acid residues. In long CDR3H, hydrophilic amino acid residues were represented mainly by repetitive Gly, Tyr, and Ser, whereas Gly dominated in most of the sequences analyzed – which is consistent with previous findings [27]. The major usage of the hydrophilic reading frame was already described in humans [59], mice [60], chicken [58], and rabbits [61]. Their occurrence in antigen-binding loops is thought to enhance flexibility and recruit somatic hypermutations for advantageous antigen binding [59].

Moreover, in accordance with descriptions provided previously, we identified multiple and mainly even numbered Cys within exceptionally long CDR3H, which were accumulated in the middle of the CDR3H [31,33]. These Cys are predicted to form intra and inter CDRH disulfide bonds, rigidifying the combining site or helping to stabilize

long CDR3H, as demonstrated in the crystallized human Fab Kol [62] and the camel cAb-Lys3 single domain antibody [63]. In this context, additional Cys in CDR2H were noticed before, when there were only one or three Cys in CDR3H [31]. Concerning the IMGT numbering system, we solely identified one additional Cys in CDR2H, when one Cys in CDR3H was found in two sequences of IgE with middle-length CDR3H. When numbering the amino acids according to Kabat et al. [8], ten sequences classified in group 2 possessed one Cys in CDR3H and showed an additional Cys in CDR2H. We did not note analogues for sequences from group 1 or group 3. In contrast to the findings that there is at least one Cys in CDR3H regions containing more than 12 amino acid residues, and that there is no Cys if the CDR3H possessed less than 10 amino acid residues [27], we also observed CDR3H sequences of intermediate length without Cys residue in CDR3H.

We annotated the bovine immunoglobulin heavy-chain locus, demonstrated the expression of unusually long CDR3H in the five bovine immunoglobulin isotypes, and specified their genomic origin. Thus, this study reviewed the opinion that exceptionally long CDR3H are a unique feature of bovine IgG1-3 and IgM.

Note

Sequences can be found under GenBank Acc.No. KC471523 to KC471622.

Supporting Information

Table S1 Genomic annotation of the bovine immunoglobulin heavy-chain locus. For a functional gene (F), the complete coding sequence, octamer motif, TATA box, splicing signals or recombination signal sequences (RSS), and poly A motif were identified. Putative functional genes ((F)) lacked some of the described parts due to end of the contig or N's introduced into gaps for example. ORFs are classified by alterations in the splicing signals, RSS or regulatory elements. The extension (ORF) describes fully functional genes in putative orphons. Pseudogenes (Ψ) revealed stop codons, frameshifts or mutations within the RSS, which lead to abolition of recombination. Fragmented loci were also defined as Ψ [38,39,40,41].

(XLS)

Contig	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Octamer-motif	Octamer Start	Octamer Stop	Octamer Start-ATG bp	Tata-Box-motif	Tata-Box Start	Tata-Box Stop	Exon 1 Start	Exon 1 Stop	bp	Intronsize	Splice signals	Exon 2a Start	Exon 2a Stop	bp	Exon 2b Start	Exon 2b Stop	bp
NW_003104530.1	21	AC_000178.1	IGHV1	-	Ψ	ATGAAAAT	6100830	6100823	114	TAGAAT	6100803	6100798	6100716	6100665	52	78	GT/GG	6100586	6100576	11	6100575	6100282	294
	21		IGHV2	-	F	ATGCAAAT	6102591	6102584	115	TTAAATT	6102542	6102536	6102476	6102431	46	82	GT/AG	6102348	6102338	11	6102337	6102045	293
NW_003104538.1	21		IGHV3	+	F	TATTCACT	71507390	71507397	156	TAT	71507430	71507432	71507546	71507591	46	82	GT/AG	71507674	71507684	11	71507685	71507977	293
	21		IGHV4	+	Ψ	ATGCAAAT	71509490	71509497	115	TAT	71509535	71509537	71509605	71509656	52	78	GT/GG	71509735	71509745	11	71509746	71510041	296
	21		IGHV5	+	Ψ	AAACAAAT	71511442	71511449		TAA	71511506	71511508					/AG	71511682	71511692	11	71511693	71511769	77
	21		IGHV6	+	F	TATTCACT	71529904	71529911	156	TAT	71529944	71529946	71530060	71530105	46	82	GT/AG	71530188	71530192	11	71530199	71530491	293
	21		IGHV7	-	Ψ	ATGGAAAC	71563607	71563600	112	TACAAT	71563580	71563575	71563495	71563444	52	76	GT/AG	71563367	71563357	11	71563356	71563070	287
	21		IGHV8	-	Ψ	AGTAAAAT	71592715	71592708		TTATATT	71592668	71592662					/AG	71592496	71592486	11	71592485	71592409	77
	21		IGHV9	-	Ψ	ATGCAAAT	71594678	71594671	115	TAT	71594633	71594631	71594563	71594513	51	78	GT/GG	71594434	71594424	11	71594423	71594130	294
	21		IGHV10	-	F	ATGCAAAT	71596456	71596449	115	TTAAATT	71596407	71596401	71596341	71596296	46	82	GT/AG	71596213	71596203	11	71596202	71595902	301
NW_003064289.1	7	AC_000164.1	IGHV11	-	Ψ							END OF CONT	15578469	15578424	46	82	GT/AG	15578341	15578332	10	15578331	15578043	289
NW_003064290.1	7		IGHV12	+	Ψ	ATGGAAAC	15580470	15580477	113				15580583	15580628	46	82	GT/AG	15580711	15580720	10	15580721	15581009	289
NW_003064296.1	7		IGHV13	+	Ψ	ATGCAGGT	15655161	15655168	122	TTAAATT	15655217	15655223	15655283	15655328	46	82	GG/AG	15655411	15655421	11	15655422	15655708	287
NW_003064297.1	7		IGHV14	-	Ψ	ATGCAAAT	15677874	15677867	115	TAAAAT	15677847	15677842	15677759	15677708	52	89	GT/CC	MISSING			15677618	15677323	296
	7		IGHV15	-	Ψ	ATGCAAAT	15679658	15679651	115	TTAAATT	15679609	15679603	15679543	15679494	50	78	GT/AG	15679413	15679404	10	15679403	15679093	311
NW_003064298.1	7		IGHV16	-	(ORF)	GTGCAAAT	15692221	15692214	115	TTAAATT	15692172	15692166	15692106	15692061	46	82	GT/AG	15691978	15691968	11	15691967	15691675	293
NW_003064299.1	7		IGHV17	-	(ORF)	ATGCAAAT	15709942	15709935	115	TTAAATT	15709893	15709887	15709827	15709782	46	82	GT/AG	15709699	15709689	11	15709688	15709396	293
	7		IGHV18	-	Ψ	ATGCAAAT	15720238	15720231	115	TAAAAT	15720211	15720206	15720123	15720072	52	78	GT/GG	15719993	15719983	11	15719982	15719691	292
	7		IGHV19	-	Ψ	ATGCAAAT	15722029	15722022	115	TTAAATT	15721980	15721974	15721869	15721914	46	82	GT/AG	15721786	15721776	11	15721775	15721483	293
NT_182448.1	unplaced		IGHV20	+	Ψ	ACAGAAAT	8940	8947	128	TTAAATT	9002	9008	9068	9117	50	78	GT/AG	9196	9206	11	9207	9494	288
	unplaced		IGHV21	+	Ψ	ATGGAAAC	12693	12700	112	TAA	12737	12739	12805	12850	46	83	GT/AG	12934	12941	8	12942	13218	277
	unplaced		IGHV22	+	Ψ	ATGGAAAC	16420	16427	112	TACAAT	16447	16452	16532	16583	52	76	GT/AG	16660	16670	11	16671	16957	287
NT_182449.1	unplaced		IGHV23	-	Ψ	ATGCAAAT	18836	18829	115	TAAAAT	18809	18804	18721	18676	52	78	GT/CC	18591	18581	11	18580	18285	296
	unplaced		IGHV24	-	Ψ	ATGCAAAT	20620	20613	115	TTAAATT	20571	20565	20505	20456	50	78	GT/AG	20377	20367	11	20366	20055	312

Contig	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	RSS Start	RSS Stop	Heptamer-motif	Spacer bp	Nonamer-motif	Feature
NW_003104530.1	21	AC_000178.1	IGHV1	-	Ψ	6100281	6100243	CACACTG	23	CCACAAGCC	AGG start codon, splicing signal, in frame stop codon, nonamer
	21		IGHV2	-	F	6102044	6102006	CACAGTG	23	ACAAAAACC	
NW_003104538.1	21		IGHV3	+	F	71507978	71508016	CACAGTG	23	ACAAAAACC	
	21		IGHV4	+	Ψ	71510042	71510080	CACAGTG	23	ACACAGGCC	ACG start codon, splicing signal, in frame stop codon, nonamer
	21		IGHV5	+	Ψ			CACAGAG			IGHV fragmented
	21		IGHV6	+	F	71530492	71530530	CACAGTG	23	ACAAAAACC	
	21		IGHV7	-	Ψ	71563069	71563031	CACAGTG	23	ACACCAGCC	ACA start codon, stop, nonamer
	21		IGHV8	-	Ψ			CACAGAG			IGHV fragmented
	21		IGHV9	-	Ψ	71594129	71594091	CACACTG	23	CCACAAGCC	14bp are missing at the start of V-segment, nonamer
	21		IGHV10	-	F	71595901	71595863	CACAGTG	23	ACAAAAACC	
NW_003064289.1	7	AC_000164.1	IGHV11	-	Ψ	15578042	15578004	CACAGTG	23	ACACCAGCC	
NW_003064290.1	7		IGHV12	+	Ψ	15581010	15581048	CACAGTG	23	ACACCAGCC	in frame stop codon, nonamer
NW_003064296.1	7		IGHV13	+	Ψ	15655709	15655743	CACAGTG	19	CCAAAAACC	in frame stop codon
NW_003064297.1	7		IGHV14	-	Ψ	15677322	15677284	CACAGTG	23	ACACCAGCC	splicing signal: ag in the 1st codon (CAG) of the V-segment, IMGT C104L, nonamer
	7		IGHV15	-	Ψ	15679092	15679054	CACAGTG	23	ACAAAAACC	in frame stop codon
NW_003064298.1	7		IGHV16	-	(ORF)	15691674	15691636	CACAGTG	23	ACAAAAACC	orphon
NW_003064299.1	7		IGHV17	-	(ORF)	15709395	15709357	CACAGTG	23	ACAAAAACC	orphon
	7		IGHV18	-	Ψ	15719690	15719652	CACAGTG	23	ACACCACAC	splicing signal, in frame stop codon, nonamer
	7		IGHV19	-	Ψ	15721482	15721444	CACAGTG	23	ACAAAAACC	in frame stop codon
NT_182448.1	unplaced		IGHV20	+	Ψ	9495	9529	CACAGTG	19	CCAAAAACC	stop
	unplaced		IGHV21	+	Ψ	13219	13257	CACAGTG	23	ACACCAGCC	stop, nonamer
	unplaced		IGHV22	+	Ψ	16958	16996	CACAGTG	23	ACACCAGCC	in frame stop, nonamer
NT_182449.1	unplaced		IGHV23	-	Ψ	18284	18246	CACAGTG	23	ACACCAGCC	splicing signal: ag in the 1st codon (CAG) of the V-segment, IMGT C104L, nonamer
	unplaced		IGHV24	-	Ψ	20054	20016	CACAGTG	23	ACAAAAACC	stop

Contig	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Octamer-motif	Octamer Start	Octamer Stop	Octamer Start-ATG bp	Tata-Box-motif	Tata-Box Start	Tata-Box Stop	Exon 1 Start	Exon 1 Stop	bp	Intronsize	Splice signals	Exon 2a Start	Exon 2a Stop	bp	Exon 2b Start	Exon 2b Stop	bp
NT_183109.1	unplaced		IGHV25	+	F	GTGCAAAT	3023	3030	115	TTAAATT	3072	3078	3138	3183	46	82	GT/AG	3266	3276	11	3277	3569	293
	unplaced		IGHV26	+	F	ATGCAAAT	13931	14045	115	TTAAATT	13980	13986	14046	14091	46	82	GT/AG	14174	14184	11	14185	14477	293
	unplaced		IGHV27	+	Ψ	ATGAAAAT	15693	15700	115	TAGAAT	15720	15725	15808	15859	52	78	GT/GG	15938	15948	11	15949	16242	294
	unplaced		IGHV28	+	Ψ	ATGGAAAC	22332	22339	113				22445	22490	46	82	GT/AG	22573	22583	11	22584	22871	288
	unplaced		IGHV29	+	(F)	N's			>80	TTAAATT	34027	34033	34093	34138	46	82	GT/AG	34221	34231	11	34232	34524	293
	unplaced		IGHV30	+	Ψ	ATGCAAAT	45790	45797	115	TAAAAT	45817	45822	45905	45956	52	78	GT/GG	46035	46045	11	46046	46337	292
	unplaced		IGHV31	+	(F)	N's			>102	TTAAATT	55960	55966	56026	56071	46	82	GT/AG	56154	56164	11	56165	56458	294
NT_185036.1	unplaced		IGHV32	-	Ψ	ATGCAAAT	10919	10912	115	AAAAAT	10892	10887	10804	10753	52	78	GT/GG	10674	10664	11	10663	10368	296
	unplaced		IGHV33	-	F	ATGCAAAT	12978	12971	115	TTAAATT	12929	12923	12863	12818	46	82	GT/AG	12735	12725	11	12724	12432	293
NT_185907.1	unplaced		IGHV34	-	F	ATGCAAAT	1113	1106	115	TTAAATT	1064	1058	998	953	46	82	GT/AG	870	860	11	859	559	301
NT_186922.1	unplaced		IGHV35	+	Ψ								<1	10	10	78	GT/GG	89	99	11	100	393	294
NW_003100762.1	unplaced		IGHV36	+	F	ATGCAAAT	1471	1478	115	TTAAATT	1520	1526	1586	1631	46	82	GT/AG	1714	1724	11	1725	2017	293

Contig	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	RSS Start	RSS Stop	Heptamer-motif	Spacer bp	Nonamer-motif	Feature
NT_183109.1	unplaced		IGHV25	+	F	3570	3608	CACAGTG	23	ACAAAAACC	
	unplaced		IGHV26	+	F	14478	14516	CACAGTG	23	ACAAAAACC	
	unplaced		IGHV27	+	Ψ	16243	16281	CACAGTG	23	ACACAGGCC	in frame stop codon, nonamer
	unplaced		IGHV28	+	Ψ	22872	22910	CACAGTG	23	ACACCAGCC	in frame stop codon, nonamer
	unplaced		IGHV29	+	(F)	34525	34563	CACAGTG	23	ACAAAAACC	sequence information incomplete
	unplaced		IGHV30	+	Ψ	46338	46376	CACAGTG	23	ACACCACAC	stop, nonamer
	unplaced		IGHV31	+	(F)	56459	56497	CACAGTG	23	ACAAAAACC	sequence information incomplete
NT_185036.1	unplaced		IGHV32	-	Ψ	10367	10329	CACAGTG	23	ACACAGGCC	ACG start codon, splicing signal, in frame stop codon, nonamer
	unplaced		IGHV33	-	F	12431	12393	CACAGTG	23	ACAAAAACC	
NT_185907.1	unplaced		IGHV34	-	F	558	520	CACAGTG	23	ACAAAAACC	
NT_186922.1	unplaced		IGHV35	+	Ψ	394	432	CACACTG	23	CCACAAGCC	splicing signal, stop, nonamer
NW_003100762.1	unplaced		IGHV36	+	F	2018	2055	CACAGTG	23	ACAAAACCT	

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	RSS 1 Stop	Nonamer-1-motif	Spacer-1 bp	Heptamer-1-motif	Exon 3 Start	Exon 3 Stop	bp	RSS 2 Start	RSS 2 Stop	Heptamer-2-motif	Spacer-2 bp	Nonamer-2-motif	Feature
NW_003104538.1		21	AC_000178.1	IGHD8	71539338	GGTTTCTGA	12	CACGGTG	71539337	71539295	43	71539294	71539267	CACAGTG	12	ACAAAAACC	
				IGHD4	71539781	GGTTTCTGA	12	CACGGTG	71539780	71539745	36	71539744	71539717	CACAGTG	12	ACAGAAACC	
NW_003064411.1		7	AC_000164.1	IGHD3	21723971	GGTTTCTGA	12	CACGGTG	21723970	21723935	58	21723912	21723885	CACAGTG	12	ACAAAAACC	
				IGHD2	21724535	GGTTTCTGA	12	CACGGTG	21724471	21724434	154	21724380	21724353	CACAGTG	12	ACAAAAACC	
				IGHD1	21724988	GGTTTCTGA	12	TGTGGTG	21724987	21724946	42	21724945	21724918	CACAGTG	12	TCAGAAACC	
NW_003066918.1		8	AC_000165.1	IGHD7	73053087	GGTTTCTGA	12	CACGGTG	73053086	73053029	58	73053028	73053001	CACAGTG	12	ACAAAAACC	
		8		IGHD6	73053539	GGTTTCTGA	12	TGTGGTG	73053538	73053497	42	73053496	73053469	CACAGTG	12	TCAGAAACC	
		8		IGHD5	73054014	GGTTTTTGA	12	CACGGTG	73054013	73053947	67	73053946	73053919	CACAGTG	12	ACAAAAACC	
		8		IGHD4	73054445	GGTTTCTGA	12	CACGGTG	73054444	73054409	36	73054408	73054381	CACAGTG	12	ACAGAAACC	
NW_003066919.1		8		IGHDQ52	73095869	AGTTTTGGC	12	CACAGTG	73095868	73095855	14	73095854	73095827	CACAGAG	12	ACAAAAACC	
NT_186153.1		unplaced		IGHD1	734	GGTTTCTGA	13	TGTGGTG	733	692	42	691	664	CACAGTG	12	TCAGAAACC	spacer-1
		unplaced		IGHD2	281	GGTTTCTGA	12	CACGGTG	280	127	154	126	99	CACAGTG	12	ACAAAAACC	
NW_001494075.1		unplaced		IGHDQ52	40361	AGTTTTGGC	12	CACAGTG	40360	40347	14	40346	40319	CACAGAG	12	ACAAAAACC	
NW_001503306.1		unplaced		IGHD4	37221	GGTTTCTGA	12	CACGGTG	37222	37257	36	37258	37285	CACAGTG	12	ACAGAAACC	
		unplaced		IGHD5	37652	GGTTTTTGA	12	CACGGTG	37653	37719	67	37720	37747	CACAGTG	12	ACAAAAACC	
		unplaced		IGHD6	38127	GGTTTCTGA	12	TGTGGTG	38128	38169	42	38170	38197	CACAGTG	12	TCAGAAACC	
		unplaced		IGHD7	38579	GGTTTCTGA	12	CACGGTG	38580	38637	58	38638	38665	CACAGTG	12	ACAAAAACC	
NW_001504477.2		unplaced		IGHD4	6702	GGTTTCTGA	12	CACGGTG	6701	6666	36	6665	6638	CACAGTG	12	ACAGAAACC	
				IGHD8	6259	GGTTTCTGA	12	CACGGTG	6258	6216	43	6215	6188	CACAGTG	12	ACAAAAACC	

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	RSS 1 Stop	Nonamer-1-motif	Spacer-1 bp	Heptamer-1-motif	Exon 3 Start	Exon 3 Stop	bp	RSS 2 Start	RSS 2 Stop	Heptamer-2-motif	Spacer-2 bp	Nonamer-2-motif	Feature
AY149283.1	Hosseini et al. 2004	11		IGHDQ52	112	GGTTTTGGC	12	CACAGTG	113	126	14	127	154	CACAAAA	12	ACAAAAACC	
AY559838.1	Shojaei et al. 2003	21		IGHD1	29	GGTTTCTGA	12	TGTGGTG	30	71	42	72	99	CACAGTG	12	TCAGAAACC	
AY559839.1	Shojaei et al. 2003	21		IGHD2	29	GGTTTCTGA	12	CACGGTG	30	177	148	178	205	CACAGTG	12	ACAGAAACC	
AY559840.1	Shojaei et al. 2003	21		IGHD3	29	GGTTTCTGA	12	CACGGTG	30	87	58	88	115	CACAGTG	12	ACAGAAACC	
HM543732.1	Koti et al. 2010	21		IGHD4	29	GGTTTCTGA	12	CACGGTG	30	65	36	66	93	CACAGTG	12	ACAGAAACC	
HM543733.1	Koti et al. 2010	21		IGHD5	29	GGTTTTTGA	12	CACGGTG	30	96	67	97	124	CACAGTG	12	ACAAAAACC	
HM543734.1	Koti et al. 2010	21		IGHD6	29	GGTTTCTGA	12	TGTGGTG	30	71	42	72	99	CACAGTG	12	TCAGAAACC	
HM543735.1	Koti et al. 2010	21		IGHD7	29	GGTTTCTGA	12	TGTGGTG	30	87	58	88	115	CACAGTG	12	ACAAAAAAC	
HM543736.1	Koti et al. 2010	21		IGHD8	29	GGTTTCTGA	12	CACGGTG	30	72	43	73	100	CACAGTG	12	ACAAAAACC	
HM543737.1	Koti et al. 2010	21		IGHD4(2)	29	GGTTTCTGA	12	CACGGTG	30	65	36	66	93	CACAGTG	12	ACAGAAACC	

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	RSS Start	RSS Stop	Nonamer-motif	Spacer bp	Heptamer-motif	Exon 4 Start	Exon 4 Stop	bp	Splicing signal	Feature
NW_003066919.1		8	AC_000165.1	JH6	-	ORF	73093665	73093628	GGTTTTTGT	22	CACTGTG	73093627	73093574	54	GT/	spacer
		8		JH5	-	Ψ	73094212	73094175	GAGCTCTTG	22	CACTGTG	73094174	73094123	52	CT/	nonamer, spacer, splicing signal
		8		JH4	-	(ORF)	73094578	73094540	GGTTTTTGC	23	CACTGTG	73094539	73094493	47	GT/	possible orphon
		8		JH3	-	Ψ	73094891	73094861	GGTTCATGT	15	GACTGCG	73094860	73094803	58	GT/	nonamer, spacer, heptamer
		8		JH2	-	Ψ	73095239	73095202	TGTTTTTGT	22	AGCCATG	73095201	73095149	53	GG/	spacer, heptamer, splicing signal
		8		JH1	-	ORF	73095419	73095383	GGTTTCCGT	21	CACTGTG	73095382	73095329	54	CT/	spacer, splicing signal
NT_186572.1		unplaced		JH4	+	ORF	2998	3035	GGTTTTGGC	22	CTCGGTG	3036	3085	50	GT/	spacer
				JH5	+	Ψ	3388	3425	GAGCTCTTG	22	CACCGTG	3426	3477	52	CT/	nonamer, spacer, splicing signal
				JH6	+	ORF	3915	3952	GGTTTTTGT	22	CACTGTG	3953	4006	54	GT/	spacer
NW_001494075.1		unplaced		JH6	-	ORF	38158	38120	GGTTTTTGT	22	CACTGTG	38119	38066	54	GT/	spacer
				JH5	-	Ψ	38704	38667	GAGCTCTTG	22	CACTGTG	38666	38615	52	CT/	nonamer, spacer, splicing signal
				JH4	-	F	39070	39032	GGTTTTTGC	23	CACTGTG	39031	38985	47	GT/	
				JH3	-	Ψ	39383	39353	GGTTCATGT	15	GACTGCG	39352	39295	58	GT/	nonamer, spacer, heptamer
				JH2	-	Ψ	39731	39694	TGTTTTTGT	22	AGCCATG	39693	39641	53	GG/	spacer, heptamer, splicing signal
				JH1	-	ORF	39911	39875	GGTTTCCGT	21	CACTGTG	39874	39821	54	CT/	spacer, splicing signal
AY149283.1	Hosseini et al. 20	11		JH1	+	Ψ	564	601	GGTTTCCGT	22	CACTGTG	602	655	54	GT/	spacer
		11		JH2	+	Ψ	745	782	TGTTTTTGT	22	AGCCATG	783	835	53	GG/	spacer, splicing signal
		11		JH3	+	Ψ	1093	1123	GGTTCATGT	15	GACTGCG	1124	1181	58	GT/	nonamer, spacer, heptamer
		11		JH4	+	(ORF)	1406	1444	GGGTTTTGC	23	CACTGTG	1445	1491	47	GT/	possible orphon
		11		JH5	+	Ψ	1772	1809	GAGCTCTTG	22	CACTGTG	1810	1861	52	CT/	nonamer, spacer, splicing signal
		11		JH6	+	Ψ	2319	2356	GGTTTTTGT	22	CACTGTG	2357	2410	54	GT/	spacer
AY158087.1	Zhao et al. 2003	21		PSJH1	+	ORF	171	207	GGTTTCCGT	21	CACTGTG	208	261	54	CT/	spacer, splicing signal
		21		PSJH2	+	ORF	351	388	TGTTTTTGT	22	AGCCATG	389	441	53	GG/	spacer, splicing signal
		21		PSJH3	+	Ψ	696	727	GGTTCATGT	16	GACTGTG	728	785	58	GT/	nonamer, spacer
		21		JH1	+	F	977	1015	GGTTTTTGC	23	CACTGTG	1016	1063	48	GT/	
		21		PSJH4	+	ORF	1360	1397	GAGCTCTTG	22	CACCGTG	1398	1449	52	CT/	nonamer, spacer, splicing signal
		21		JH2	+	F	1885	1922	GGTTTTTGT	22	CACTGTG	1923	1976	54	GT/	spacer

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Splice signal	Exon 1	Exon1-Start	Exon 1-Stop	bp	Intronsize	Splice signals	Exon 2	Exon 2-Start	Exon 2-Stop	bp	Intronsize	Splice signals	Exon 3	Exon 3-Start	Exon 3-Stop	bp
NW_003104538.1		21	C_000178.1	IgA	-	(F)	/AG	IgA-CH1	7,1E+07	7,1E+07	306	180	GT/AG	IgA-CH2	7,1E+07	7,1E+07	330	188	GT/AG	IgA-CH3	7,1E+07	<7145361	>125
		21		IgE	-	F	/AG	IgE-CH1	7,1E+07	7,1E+07	291	118	GT/AG	IgE-CH2	7,1E+07	7,1E+07	324	87	GT/AG	IgE-CH3	7,1E+07	71468806	324
						F																	
		21		IgM	-	(F)	/AG	IgM-CH1	7,1E+07	7,1E+07	324	75	GT/AG	IgM-CH2	7,1E+07	7,1E+07	333	252	GT/AG	IgM-CH3	7,1E+07	<7149124	>134
						(F)																	
NW_003064414.1		7	AC_000164.	IgD	-	Ψ	/TT	IgD-CH1	2,2E+07	2,2E+07	98	140	GT/AG	gD-Hinge	2,2E+07	2,2E+07	90	666	GT/AG				
NW_003066919.1		8	C_000165.1	IgD	-	Ψ	/AG	IgD-CH1	7,3E+07	7,3E+07	324	140	GT/AG	IgD-Hinge 1	7,3E+07	7,3E+07	90	1508	GT/AG	IgD-Hinge 2	7,3E+07	73077856	62
						Ψ																	
		8		IgM	-	(ORF)	/AG	IgM-CH1	7,3E+07	7,3E+07	324	74	GT/AG	IgM-CH2	7,3E+07	7,3E+07	333	251	GT/AG	IgM-CH3	7,3E+07	73088079	315
						(ORF)																	
NW_003104522.1		20	AC_000177.	IgG2	+	(ORF)	/AG	IgG2-CH1	7,2E+07	7,2E+07	291	331	GT/AG	IgG2-Hinge	7,2E+07	7,2E+07	39	137	GT/AG	IgG2-CH2	7,2E+07	71903590	321
						(ORF)																	
NT_185580.1		unplaced		IgG3	+	(F)	/AG	IgG3-CH1	14760	15119	360	290	GT/AG	IgG3-Hinge	15410	15454	45	129	GT/AG	IgG3-CH2	15584	15907	324
				IgG2	+	F	/AG	IgG2-CH1	16548	16838	291	331	GT/AG	IgG2-Hinge	17170	17208	39	137	GT/AG	IgG2-CH2	17346	17666	321
						F																	
				IgG1	+	(F)	/AG	IgG1-CH1	44699	44989	291	331	GT/AG	IgG1-Hinge	45321	45359	39	126	GT/AG	IgG1-CH2	45486	45815	330
						F																	
NT_185723.1		unplaced		IgE	+	F	/AG	IgE-CH1	16225	16515	291	118	GT/AG	IgE-CH2	16634	16957	324	87	GT/AG	IgE-CH3	17045	17368	324
						F																	
				IgA	+	(F)	/AG	IgA-CH1	31434	31739	306	180	GT/AG	IgA-CH2	31920	32249	330	187	GT/AG	IgA-CH3.1	32437	32770	333
						F														IgA-CH3.2 (secretory)	32437	32831	395
NT_186572.1		unplaced		IgD	+	(F)	/AG	IgD-CH1	9421	9744	324	>37	GT/										
NW_001494075.1		unplaced		IgD	-	Ψ	/AG	IgD-CH1	24470	24147	324	140	GT/AG	IgD-Hinge 1	24006	23917	90	1508	GT/AG	IgD-Hinge 2	22408	22347	62
						Ψ																	
				IgM	-	F	/AG	IgM-CH1	33867	33544	324	74	GT/AG	IgM-CH2	33469	33137	333	251	GT/AG	IgM-CH3	32885	32571	315
						F																	

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Intronsize	Splice signals	Exon 4	Exon 4-Start	Exon 4-Stop	bp	Intronsize	Splice signals	Exon 5	Exon 5-Start	Exon 5-Stop	bp	Intronsize	Splice signals
NW_003104538.1		21	C_000178.1	IgA	-	(F)	<2534	/AG	TM	71451077	7,1E+07	188								
		21		IgE	-	F	72	GT/AG	IgE-CH4.1	71468733	7,1E+07	333	2253	GT/AG	TM1	7,1E+07	7,1E+07	134	98	GT/AG
						F			IgE-CH4.2 (secretory)	71468733	7,1E+07	341								
		21		IgM	-	(F)	>101		IgM-CH4.1	>7149114	7,1E+07	>239								
						(F)			IgM-CH4.2 (secretory)	>7149114	7,1E+07	>301								
NW_003064414.1		7	AC_000164.	IgD	-	Ψ			IgD-CH2	22408093	2,2E+07	324	104	GT/AG	IgD-CH3	2,2E+07	2,2E+07	321	3539	GT/AG
NW_003066919.1		8	C_000165.1	IgD	-	Ψ	247	GT/AG	IgD-CH2	73077608	7,3E+07	325	104	GT/AG	IgD-CH3	7,3E+07	7,3E+07	321	1401	GT/AG
						Ψ														
		8		IgM	-	(ORF)	131	GT/AG	IgM-CH4.1	73087947	7,3E+07	333	1972	GT/AG	TM1	7,3E+07	7,3E+07	116	101	GT/AG
						(ORF)		GT/AG	IgM-CH4.2 (secretory)	73087947	7,3E+07	395								
NW_003104522.1		20	AC_000177.	IgG2	+	(ORF)	82	GT/AG	IgG2-CH3	71903673	7,2E+07	321	1553	GT/AG	TM1	7,2E+07	7,2E+07	134	166	GT/AG
						(ORF)			IgG2-CH3.2 (secretory)	71903673	7,2E+07	329								
NT_185580.1		unplaced		IgG3	+	(F)	>95	GT/												
				IgG2	+	F	82	GT/AG	IgG2-CH3.1	17749	18069	321	1553	GT/AG	TM1	19623	19756	134	167	GT/AG
						F			IgG2-CH3.2 (secretory)	17749	18077	329								
				IgG1	+	(F)	82	GT/AG	IgG1-CH3.1	45898	46218	321	1556	GT/AG	TM1	47775	47908	134	2845	GT/AG
						F			IgG1-CH3.2 (secretory)	45898	46226	329								
NT_185723.1		unplaced		IgE	+	F	72	GT/AG	IgE-CH4.1	17441	17773	333	2253	GT/AG	TM1	20027	20160	134	98	GT/AG
						F			IgE-CH4.2 (secretory)	17441	17781	341								
				IgA	+	(F)	2274	GT/AG	IgA-TM	35135	>35275	>141								
						F														
NT_186572.1		unplaced		IgD	+	(F)														
NW_001494075.1		unplaced		IgD	-	Ψ	247	GT/AG	IgD-CH2	22099	21775	325	104	GT/AG	IgD-CH3	21670	21350	321	1401	GT/AG
						Ψ														
				IgM	-	F	131	GT/AG	IgM-CH4.1	32439	32107	333	1972	GT/AG	TM1	30134	30019	116	101	GT/AG
						F			IgM-CH4.2 (secretory)	32439	32045	395								

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Exon 6	Exon 6-Start	Exon 6-Stop	bp	Intronsize	Splice signals	Exon 7	Exon 7-Start	Exon 7-Stop	bp	Intronsize	Splice signals	Exon 8	Exon 8-Start	Exon 8-Stop	bp
NW_003104538.1		21	C_000178.1	IgA	-	(F)																
		21		IgE	-	F	TM2	7,1E+07	7,1E+07	90												
						F																
		21		IgM	-	(F)																
						(F)																
NW_003064414.1		7	AC_000164.	IgD	-	Ψ							TM1	2,2E+07	2,2E+07	147	211	GT/AG	TM2	2,2E+07	2,2E+07	9
NW_003066919.1		8	C_000165.1	IgD	-	Ψ	IgD-CH3 (secretory)	7,3E+07	7,3E+07	77	2546	GT/AG	TM1	7,3E+07	7,3E+07	147	210	GT/AG	TM2	7,3E+07	7,3E+07	9
						Ψ																
		8		IgM	-	(ORF)	TM2	7,3E+07	7,3E+07	9												
						(ORF)																
NW_003104522.1		20	AC_000177.	IgG2	+	(ORF)	TM2	7,2E+07	7,2E+07	84												
						(ORF)																
NT_185580.1		unplaced		IgG3	+	(F)																
				IgG2	+	F	TM2	19924	20007	84												
						F																
				IgG1	+	(F)	TM2	50754	50837	84												
						F																
NT_185723.1		unplaced		IgE	+	F	TM2	20259	20348	90												
						F																
				IgA	+	(F)																
						F																
NT_186572.1		unplaced		IgD	+	(F)																
NW_001494075.1		unplaced		IgD	-	Ψ	IgD-CH3 (secretory)	19948	19872	77	2615	GT/AG	TM1	17256	17110	147	210	GT/AG	TM2	16899	16891	9
						Ψ																
				IgM	-	F	TM2	29917	29909	9												
						F																

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Poly A-motif	Start	Stop	Feature
NW_003104538.1		21	C_000178.1	IgA	-	(F)	AATAAA	7,1E+07	7,1E+07	sequence information incomplete: 1390 N's in Exon 3
		21		IgE	-	F	AATAA	7,1E+07	7,1E+07	
						F	AATAAA	7,1E+07	7,1E+07	
		21		IgM	-	(F)				sequence information incomplete: 100 N's between Exon 3 and 4
						(F)				
NW_003064414.1		7	AC_000164.	IgD	-	Ψ	ATAAA	2,2E+07	2,2E+07	Exon 1 fragmented, Frame shift, Exon 2 fragmented, Exon 6 (secretory region) missing, insertion of GCAG in Exon 7
NW_003066919.1		8	C_000165.1	IgD	-	Ψ	ATAAA	7,3E+07	7,3E+07	Frame shift, deletion of CAGA in Exon 3, insertion of G in Exon 4, insertion of GCAG in Exon 7
						Ψ	AAATAAA	7,3E+07	7,3E+07	
		8		IgM	-	(ORF)	AAATAAA	7,3E+07	7,3E+07	possible orphon
						(ORF)	ATAAAAATTAGAAATAAAAA	7,3E+07	7,3E+07	possible orphon
NW_003104522.1		20	AC_000177.	IgG2	+	(ORF)	AATA	7,2E+07	7,2E+07	possible orphon
						(ORF)	AATAAA	7,2E+07	7,2E+07	possible orphon
NT_185580.1		unplaced		IgG3	+	(F)				sequence information incomplete: N's in Intron 3
				IgG2	+	F	AATA	20628	20631	
						F	AATAAA	18189	18194	
				IgG1	+	(F)	AATA	51367	51370	sequence information incomplete: 373 N's in Intron 5
						F	AATAAA	46338	46343	
NT_185723.1		unplaced		IgE	+	F	AATAA	20878	20882	
						F	AATAAA	17850	17855	
				IgA	+	(F)				sequence information incomplete: poly A not detectable due to N's
						F	AATAAA	32853	32858	
NT_186572.1		unplaced		IgD	+	(F)				sequence information incomplete: N's in Intron 1, best homology to IgD-CH1
NW_001494075.1		unplaced		IgD	-	Ψ	ATAAA	16626	16622	Frame shift, deletion of CAGA in Exon 3, insertion of G in Exon 4, insertion of GCAG in Exon 7
						Ψ	AAATAAA	19370	19364	
				IgM	-	F	AAATAAA	29674	29668	
						F	AAAAATTAGAAATAAAAA	31956	31939	

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Splice signal	Exon 1	Exon1-Start	Exon 1-Stop	bp	Intronsize	Splice signals	Exon 2	Exon 2-Start	Exon 2-Stop	bp	Intronsize	Splice signals	Exon 3	Exon 3-Start	Exon 3-Stop	bp
NW_001503306.1		unplaced		IgD	+	Ψ	/TT	IgD-CH1	11925	12022	98	140	GT/AG	IgD-Hinge 1	12163	12252	90	666	GT/AG				
NW_003099305.1		unplaced		IgG3	-	(F)	/AG	IgG3-CH1	1351	992	360	290	GT/AG	IgG3-Hinge	701	657	45	129	GT/AG	IgG3-CH2	527	204	324
						(F)																	
NW_003100065.1		unplaced		IgG1	+	(F)	/AG	IgG1-CH1	4083	4373	291	331	GT/AG	IgG1-Hinge	4705	4743	39	129	GT/AG	IgG1-CH2	4870	5199	330
						F																	
NW_003100112.1		unplaced		IgD	+	(F)	/AG	IgD-CH1	5077	5400	324	>36	GT/										
NW_003100387.1		unplaced		IgD	+	Ψ												>134(	/AG	IgD-Hinge 2	1341	1406	66
AF109167.1	Brown et al. 1997	21		IgA	+	(F)	/AG	IgA-CH1	68	373	306	180	GT/AG	IgA-CH2	554	883	330	188	GT/AG	IgA-CH3.1	1072	1404	333
						F														IgA-CH3.2 (secretory)	1072	1466	395
AY158087.1	Zhao et al. 2003	21		IgM	+	(F)	/AG	IgM-CH1	7819	8142	324	75	GT/AG	IgM-CH2	8218	>8259							
AY230207.1	Zhao et al. 2003	21		IgM	+	(F)		IgM-CH1	<1	308	308	75	GT/AG	IgM-CH2	384	716	333	251	GT/AG	IgM-CH3	968	1282	315
						(F)																	
U63637.2	Mousavi et al. 1998			IgM	+	F	/AG	IgM-CH1	293	616	324	74	GT/AG	IgM-CH2	691	1023	333	252	GT/AG	IgM-CH3	1276	1590	315
						F																	
U63638.1	Rabbani et al. 1997			IgG3	+	(F)	/AG	IgG3-CH1	122	481	360	290	GT/AG	IgG3-Hinge	772	816	45	129	GT/AG	IgG3-CH2	946	1269	324
						F																	
U63639.1	Rabbani et al. 1997			IgG3	+	(F)	/AG	IgG3-CH1	166	525	360	290	GT/AG	IgG3-Hinge	816	860	45	129	GT/AG	IgG3-CH2	990	1313	324
						F																	
U63640.2				IgE	+	(F)	/AG	IgE-CH1	264	554	291	118	GT/AG	IgE-CH2	673	996	324	87	GT/AG	IgE-CH3	1084	1407	324
						F																	

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Intronsize	Splice signals	Exon 4	Exon 4-Start	Exon 4-Stop	bp	Intronsize	Splice signals	Exon 5	Exon 5-Start	Exon 5-Stop	bp	Intronsize	Splice signals
NW_001503306.1		unplaced		IgD	+	Ψ			IgD-CH2	12919	13242	324	104	GT/AG	IgD-CH3	13347	13667	321	3541	GT/AG
NW_003099305.1		unplaced		IgG3	-	(F)	82	GT/AG	IgG3-CH3.1	121	<1									
						(F)			IgG3-CH3.2 (secretory)	121	<1									
NW_003100065.1		unplaced		IgG1	+	(F)	82	GT/AG	IgG1-CH3.1	5282	5602	321	1553	GT/AG	TM1	7156	7289	134	>1041	
						F			IgG1-CH3.2 (secretory)	5282	5610	329								
NW_003100112.1		unplaced		IgD	+	(F)														
NW_003100387.1		unplaced		IgD	+	Ψ	247	GT/AG	IgD-CH2	1654	1977	324	104	GT/AG	IgD-CH3	2082	2402	321	4196	GT/AG
AF109167.1	Brown et al. 1997	21		IgA	+	(F)	>126	GT/												
						F														
AY158087.1	Zhao et al. 2003	21		IgM	+	(F)														
AY230207.1	Zhao et al. 2003	21		IgM	+	(F)	131	GT/AG	IgM-CH4.1	1414	1746	333	1975	GT/AG	TM1	3722	>3799			
						(F)			IgM-CH4.2 (secretory)	1414	1808	395								
U63637.2	Mousavi et al. 1998			IgM	+	F	131		IgM-CH4.1	1722	2054	333	1972	GT/AG	TM1	4027	4142	116	93	GT/AG
						F			IgM-CH4.2 (secretory)	1722	2116	395								
U63638.1	Rabbani et al. 1997			IgG3	+	(F)	166	GT/AG	IgG3-CH3.1	1436	1756	321	>8	GT/						
						F			IgG3-CH3.2 (secretory)	1436	1764	329								
U63639.1	Rabbani et al. 1997			IgG3	+	(F)	82	GT/AG	IgG3-CH3.1	1396	1716	321	>136	GT/						
						F			IgG3-CH3.2 (secretory)	1396	1724	329								
U63640.2				IgE	+	(F)	72	GT/AG	IgE-CH4.1	1480	1812	333	>39	GT/						
						F			IgE-CH4.2 (secretory)	1480	1820	341								

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Exon 6	Exon 6-Start	Exon 6-Stop	bp	Intronsize	Splice signals	Exon 7	Exon 7-Start	Exon 7-Stop	bp	Intronsize	Splice signals	Exon 8	Exon 8-Start	Exon 8-Stop	bp
NW_001503306.1		unplaced		IgD	+	Ψ							TM1	17209	17355	147	211	GT/AG	TM2	17567	17575	9
NW_003099305.1		unplaced		IgG3	-	(F)																
						(F)																
NW_003100065.1		unplaced		IgG1	+	(F)																
						F																
NW_003100112.1		unplaced		IgD	+	(F)																
NW_003100387.1		unplaced		IgD	+	Ψ							TM1	6599	6745	147	211	GT/AG	TM2	6957	6965	9
AF109167.1	Brown et al. 1997	21		IgA	+	(F)																
						F																
AY158087.1	Zhao et al. 2003	21		IgM	+	(F)																
AY230207.1	Zhao et al. 2003	21		IgM	+	(F)																
						(F)																
U63637.2	Mousavi et al. 1998			IgM	+	F	TM2	4236	4244	9												
						F																
U63638.1	Rabbani et al. 1997			IgG3	+	(F)																
						F																
U63639.1	Rabbani et al. 1997			IgG3	+	(F)																
						F																
U63640.2				IgE	+	(F)																
						F																

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Poly A-motif	Start	Stop	Feature
NW_001503306.1		unplaced		IgD	+	Ψ	ATAAA	17856	17860	Exon 1 fragmented, Frame shift, Exon 2 fragmented, Exon 6 (secretory region) missing, insertion of GCAG in Exon 7
							AAATAAA	15679	15685	Possible poly A of secretory IgD
NW_003099305.1		unplaced		IgG3	-	(F)				sequence information incomplete: end of contig
						(F)				
NW_003100065.1		unplaced		IgG1	+	(F)				sequence information incomplete: end of contig
						F	AATAAA	5722	5727	
NW_003100112.1		unplaced		IgD	+	(F)				sequence information incomplete: end of contig, best homology to IgD-CH1
NW_003100387.1		unplaced		IgD	+	Ψ	ATAAA	7236	7240	start of contig, frameshift, exon 6 (secretory region) missing, insertion of GCAG in exon 7
AF109167.1	Brown et al. 1997	21		IgA	+	(F)				sequence information incomplete: end of sequence
						F	AATAAA	1488	1493	
AY158087.1	Zhao et al. 2003	21		IgM	+	(F)				sequence information incomplete: end of sequence
AY230207.1	Zhao et al. 2003	21		IgM	+	(F)				sequence information incomplete: start and end of sequence
						(F)	ATAAAAATTAGAAATAAAAA	1893	1912	
U63637.2	Mousavi et al. 1998			IgM	+	F	AAATAAA	4479	4485	
						F	AAAAATTAGAAATAAAAA	2205	2222	
U63638.1	Rabbani et al. 1997			IgG3	+	(F)				sequence information incomplete
						F				
U63639.1	Rabbani et al. 1997			IgG3	+	(F)				sequence information incomplete
						F	AATAAA	1826	1831	

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Splice signal	Exon 1	Exon1-Start	Exon 1-Stop	bp	Intronsize	Splice signals	Exon 2	Exon 2-Start	Exon 2-Stop	bp	Intronsize	Splice signals	Exon 3	Exon 3-Start	Exon 3-Stop	bp
X16701.1	Symons et al. 1989			IgG1	+	(F)	/AG	IgG1-CH1	170	460	291	328	GT/AG	IgG1-Hinge	789	827	39	126	GT/AG	IgG1-CH2	954	1283	330
						F																	
X16702.1	Symons et al. 1989			IgG2	+	(F)	/AG	IgG2-CH1	122	415	294	328	GT/AG	IgG2-Hinge	744	782	39	137	GT/AG	IgG2-CH2	920	1240	321
						F																	

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Intronsize	Splice signals	Exon 4	Exon 4-Start	Exon 4-Stop	bp	Intronsize	Splice signals	Exon 5	Exon 5-Start	Exon 5-Stop	bp	Intronsize	Splice signals
X16701.1	Symons et al. 1989			IgG1	+	(F)	80	GT/AG	IgG1-CH3.1	1364	1684	321	>146	GT/						
						F			IgG1-CH3.2 (secretory)	1364	1692	329								
X16702.1	Symons et al. 1989			IgG2	+	(F)	79	GT/AG	IgG2-CH3.1	1320	1640	321	>339	GT/						
						F			IgG2-CH3.2 (secretory)	1320	1648	329								

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Exon 6	Exon 6-Start	Exon 6-Stop	bp	Intronsize	Splice signals	Exon 7	Exon 7-Start	Exon 7-Stop	bp	Intronsize	Splice signals	Exon 8	Exon 8-Start	Exon 8-Stop	bp
X16701.1	Symons et al. 1989			IgG1	+	(F)																
						F																
X16702.1	Symons et al. 1989			IgG2	+	(F)																
						F																

Contig	Reference	BTA	Acc.No. BTA	Walther et al. 2013	Strand	Classification	Poly A-motif	Start	Stop	Feature
U63640.2				IgE	+	(F)				sequence information incomplete
						F				
X16701.1	Symons et al. 1989			IgG1	+	(F)				sequence information incomplete
						F				
X16702.1	Symons et al. 1989			IgG2	+	(F)				sequence information incomplete
						F				

Acknowledgments

We are grateful to Lynne Armstrong for critical review of the manuscript. We also thank the Göttingen Genomics Laboratory for sample sequencing and Caroline Bierschenk for her excellent technical assistance.

References

1. Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302: 575-581.
2. Desiderio SV, Yancopoulos GD, Paskind M, Thomas E, Boss MA, et al. (1984) Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. Nature 311: 752-755.
3. Alt FW, Baltimore D (1982) Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. Proc Natl Acad Sci U S A 79: 4118-4122.
4. Kurosawa Y, Tonegawa S (1982) Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA segments. J Exp Med 155: 201-218.
5. Reynaud CA, Garcia C, Hein WR, Weill JC (1995) Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. Cell 80: 115-125.
6. Berens SJ, Wylie DE, Lopez OJ (1997) Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. Int Immunol 9: 189-199.
7. Wagner SD, Neuberger MS (1996) Somatic hypermutation of immunoglobulin genes. Annu Rev Immunol 14: 441-457.
8. Kabat EA, Wu TT, Perry HM, Gottesmann KS, Foeller C (2006) Sequences of proteins of immunological interest. Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health.
9. Reynaud CA, Dahan A, Anquez V, Weill JC (1989) Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell 59: 171-183.
10. Sun J, Kacskovics I, Brown WR, Butler JE (1994) Expressed swine VH genes belong to a small VH gene family homologous to human VHIII. J Immunol 153: 5618-5627.
11. Dufour V, Malinge S, Nau F (1996) The sheep Ig variable region repertoire consists of a single VH family. J Immunol 156: 2163-2170.
12. Saini SS, Hein WR, Kaushik A (1997) A single predominantly expressed polymorphic immunoglobulin VH gene family, related to mammalian group, I, clan, II, is identified in cattle. Mol Immunol 34: 641-651.
13. Sinclair MC, Gilchrist J, Aitken R (1997) Bovine IgG repertoire is dominated by a single diversified VH gene family. J Immunol 159: 3883-3889.
14. Arakawa H, Saribasak H, Buerstedde JM (2004) Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. PLoS Biol 2: E179.
15. Parng CL, Hansal S, Goldsby RA, Osborne BA (1996) Gene conversion contributes to Ig light chain diversity in cattle. J Immunol 157: 5478-5486.
16. Sun Y, Wang C, Wang Y, Zhang T, Ren L, et al. (2010) A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. Dev Comp Immunol 34: 1009-1020.
17. Zhao Y, Kacskovics I, Pan Q, Liberles DA, Geli J, et al. (2002) Artiodactyl IgD: the missing link. J Immunol 169: 4408-4416.
18. Knight KL, Becker RS (1987) Isolation of genes encoding bovine IgM, IgG, IgA and IgE chains. Vet Immunol Immunopathol 17: 17-24.
19. Symons DB, Clarkson CA, Beale D (1989) Structure of bovine immunoglobulin constant region heavy chain gamma 1 and gamma 2 genes. Mol Immunol 26: 841-850.
20. Miller JR, Thomsen PD, Dixon SC, Tucker EM, Konfortov BA, et al. (1992) Synteny mapping of the bovine IGHG2, CRC and IGF1 genes. Anim Genet 23: 51-58.

21. Zhao Y, Kacskovics I, Rabbani H, Hammarstrom L (2003) Physical mapping of the bovine immunoglobulin heavy chain constant region gene locus. J Biol Chem 278: 35024-35032.
22. Hayes H, Di Meo GP, Gautier M, Laurent P, Eggen A, et al. (2000) Localization by FISH of the 31 Texas nomenclature type I markers to both Q- and R-banded bovine chromosomes. Cytogenet Cell Genet 90: 315-320.
23. Gu F, Chowdhary BP, Andersson L, Harbitz I, Gustavsson I (1992) Assignment of the bovine immunoglobulin gamma heavy chain (IGHG) gene to chromosome 21q24 by in situ hybridization. Hereditas 117: 237-240.
24. Tobin-Janzen TC, Womack JE (1992) Comparative mapping of IGHG1, IGHM, FES, and FOS in domestic cattle. Immunogenetics 36: 157-165.
25. Hayes HC, Petit EJ (1993) Mapping of the beta-lactoglobulin gene and of an immunoglobulin M heavy chain-like sequence to homoeologous cattle, sheep, and goat chromosomes. Mamm Genome 4: 207-210.
26. Hosseini A, Campbell G, Prorocic M, Aitken R (2004) Duplicated copies of the bovine JH locus contribute to the Ig repertoire. Int Immunol 16: 843-852.
27. Lopez O, Perez C, Wylie D (1998) A single VH family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains. Immunol Rev 162: 55-66.
28. Tutter A, Riblet R (1989) Conservation of an immunoglobulin variable-region gene family indicates a specific, noncoding function. Proc Natl Acad Sci U S A 86: 7460-7464.
29. Koti M, Kataeva G, Kaushik AK (2008) Organization of D(H)-gene locus is distinct in cattle. Dev Biol (Basel) 132: 307-313.
30. Koti M, Kataeva G, Kaushik AK (2010) Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies. Mol Immunol 47: 2119-2128.
31. Saini SS, Allore B, Jacobs RM, Kaushik A (1999) Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. Eur J Immunol 29: 2420-2426.
32. Saini SS, Kaushik A (2002) Extensive CDR3H length heterogeneity exists in bovine foetal VDJ rearrangements. Scand J Immunol 55: 140-148.
33. Kaushik AK, Kehrli ME, Jr., Kurtz A, Ng S, Koti M, et al. (2009) Somatic hypermutations and isotype restricted exceptionally long CDR3H contribute to antibody diversification in cattle. Vet Immunol Immunopathol 127: 106-113.
34. Saini SS, Farrugia W, Ramsland PA, Kaushik AK (2003) Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted VH + Vlambda pairings. Int Immunol 15: 845-853.
35. Larsen PA, Smith TP (2012) Application of circular consensus sequencing and network analysis to characterize the bovine IgG repertoire. BMC Immunol 13: 52.
36. Shojaei F, Saini SS, Kaushik AK (2003) Unusually long germline DH genes contribute to large sized CDR3H in bovine antibodies. Mol Immunol 40: 61-67.
37. Ramsland PA, Kaushik A, Marchalonis JJ, Edmundson AB (2001) Incorporation of long CDR3s into V domains: implications for the structural evolution of the antibody-combining site. Exp Clin Immunogenet 18: 176-198.
38. Lefranc MP (1998) IMGT (ImMunoGeneTics) locus on focus. A new section of Experimental and Clinical Immunogenetics. Exp Clin Immunogenet 15: 1-7.
39. Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, et al. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol 27: 55-77.

40. Akamatsu Y, Tsurushita N, Nagawa F, Matsuoka M, Okazaki K, et al. (1994) Essential residues in V(D)J recombination signals. J Immunol 153: 4520-4529.
41. Hesse JE, Lieber MR, Mizuuchi K, Gellert M (1989) V(D)J recombination: a functional definition of the joining signals. Genes Dev 3: 1053-1061.
42. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.
43. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783-791.
44. Sitnikova T, Su C (1998) Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. Mol Biol Evol 15: 617-625.
45. Almagro JC, Martinez L, Smith SL, Alagon A, Estevez J, et al. (2006) Analysis of the horse V(H) repertoire and comparison with the human IGHV germline genes, and sheep, cattle and pig V(H) sequences. Mol Immunol 43: 1836-1845.
46. Perriere G, Gouy M (1996) WWW-query: an on-line retrieval system for biological sequence banks. Biochimie 78: 364-369.
47. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74: 5463-5467.
48. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.
49. Ehrenmann F, Kaas Q, Lefranc MP (2010) IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. Nucleic Acids Res 38: D301-307.
50. Ehrenmann F, Lefranc MP (2011) IMGT/DomainGapAlign: IMGT standardized analysis of amino acid sequences of variable, constant, and groove domains (IG, TR, MH, IgSF, MhSF). Cold Spring Harb Protoc 2011: 737-749.
51. Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, et al. (1998) The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J Exp Med 188: 2151-2162.
52. Matsuda F, Honjo T (1996) Organization of the human immunoglobulin heavy-chain locus. Adv Immunol 62: 1-29.
53. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, et al. (2009) A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol 10: R42.
54. Everts-van der Wind A, Kata SR, Band MR, Rebeiz M, Larkin DM, et al. (2004) A 1463 gene cattle-human comparative map with anchor points defined by human genome sequence coordinates. Genome Res 14: 1424-1437.
55. Heuertz S, Hors-Cayla MC (1978) [Bovine chromosome mapping with the cell hybridization technic. Localization on the x chromosome of glucose-6-phosphate dehydrogenase, phosphoglycerate kinase, alpha-galactosidase A and hypoxanthine phosphoribosyltransferase]. Ann Genet 21: 197-202.
56. Sinclair MC, Aitken R (1995) PCR strategies for isolation of the 5' end of an immunoglobulin-encoding bovine cDNA. Gene 167: 285-289.
57. Jhunjhunwala S, van Zelm MC, Peak MM, Cutchin S, Riblet R, et al. (2008) The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. Cell 133: 265-279.
58. Reynaud CA, Bertocci B, Dahan A, Weill JC (1994) Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. Adv Immunol 57: 353-378.
59. Corbett SJ, Tomlinson IM, Sonnhammer EL, Buck D, Winter G (1997) Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis

provides no evidence for the use of DIR segments, inverted D segments, "minor" D segments or D-D recombination. J Mol Biol 270: 587-597.
60. Kaartinen M, Mäkelä O (1985) Reading of D genes in variable frames as a source of antibody diversity. Immunol Today 6: 324-327.
61. Friedman ML, Tunyaplin C, Zhai SK, Knight KL (1994) Neonatal VH, D, and JH gene usage in rabbit B lineage cells. J Immunol 152: 632-641.
62. Schmidt WE, Jung HD, Palm W, Hilschmann N (1983) [Three-dimensional structure determination of antibodies. Primary structure of crystallized monoclonal immunoglobulin IgG1 KOL, I]. Hoppe Seylers Z Physiol Chem 364: 713-747.
63. Desmyter A, Transue TR, Ghahroudi MA, Thi MH, Poortmans F, et al. (1996) Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. Nat Struct Biol 3: 803-811.
64. Ota T, Nei M (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. Mol Biol Evol 11: 469-482.

Development of a bioinformatics framework for the detection of gene conversion and the analysis of combinatorial diversity in immunoglobulin heavy chains in four cattle breeds

Stefanie Walther[1¶], Manfred Tietze[2¶], Claus-Peter Czerny[1], Sven König[2], Ulrike S. Diesterbeck[1*]

[1]Department of Animal Sciences, Institute of Veterinary Medicine, Division of Microbiology and Animal Hygiene, Faculty of Agricultural Sciences, Georg-August University Goettingen, Goettingen, Germany
[2]Department of Animal Breeding, University of Kassel, Witzenhausen, Germany

*Corresponding author:
Email: ulrike.diesterbeck@nih.gov (USD)

¶These authors contributed equally to this work.

Submitted to “PloSOne”, manuscript under review

Abstract

We have developed a new bioinformatics framework for the analysis of rearranged bovine heavy chain immunoglobulin (Ig) variable regions by combining and refining widely used alignment algorithms. This bioinformatics framework allowed us to investigate alignments of heavy chain framework regions (FRHs) and the separate alignments of FRHs and heavy chain complementarity determining regions (CDRHs) to determine their germline origin in the 4 cattle breeds Aubrac, German Black Pied, German Simmental, and Holstein Friesian. Now it is also possible to specifically analyze Ig heavy chains possessing exceptionally long CDR3Hs.

In order to gain more insight into breed specific differences in Ig combinatorial diversity, somatic hypermutations and putative gene conversions of IgG, we compared the dominantly transcribed variable (IGHV), diversity (IGHD), and joining (IGHJ) segments and their recombination in the 4 cattle breeds.

The analysis revealed the use of 15 different *IGHV* segments, 21 *IGHD* segments, and 2 *IGHJ* segments with significant different transcription levels within the breeds. Furthermore, there are preferred rearrangements within the 3 groups of CDR3H lengths. In the sequences of group 2 (CDR3H lengths (L) of 11-47 amino acid residues (aa)) a higher number of recombination was observed than in sequences of group 1 (L≤10 aa) and 3 (L≥48 aa). The combinatorial diversity of germline *IGHV*, *IGHD*, and *IGHJ*-segments revealed 162 rearrangements that were significantly different. The few preferably rearranged gene segments within group 3 CDR3H regions may indicate specialized antibodies because this length is unique in cattle.

The most important finding of this study, which was enabled by using the bioinformatics framework, is the discovery of strong evidence for gene conversion as a rare event using pseudogenes fulfilling all definitions for this particular diversification mechanism.

Author Summary

The humoral immune response is an integral part in the specific pathogen defense. Beside universal processes for the generation of the immunoglobulin diversity applying for all higher vertebrates, species-specific mechanisms exist. Cattle possess less functional gene segments available for combinatorial diversity if compared with human and mice.

Thus, we hypothesized that cattle use pseudogenes for a genetically exchange, the so-called gene conversion. Beside the difficulty of the incomplete annotated bovine heavy chain immunoglobulin locus, we were able to verify gene conversion events fulfilling all conditions like clusters of nucleotide exchanges, at least triplets, origin of the pseudogene upstream of the rearranged segment and a high degree of flanking homology.

In addition, we analyzed the combinatorial diversity and hot spots for hypermutations in four different cattle breeds: Aubrac, German Black Pied, German Simmental, and Holstein Friesian. Here we observed a significant non-random distribution in the usage of the segments as well as in their combinations in Aubrac and German Simmental. The highest variability within the variable region was always observed in the transition of framework regions to complementarity determining regions (CDRH) if compared to the middle of the CDRHs.

Introduction

The basic genetic mechanism in developing immunoglobulin diversity is similar in all jawed vertebrates. Immunoglobulins (Ig) are Y-shaped hetero-tetramers consisting of two identical heavy chains (IGH) and two identical light chains, either κ or λ in mammals (IGK, IGL) [1]. Both chains are functionally divided into variable and constant domains that are combined during B-cell development. The variable domain is rearranged by separate heavy and light chain variable (*IGHV*, *IGKV*, *IGLV*), diversity (*IGHD*), and joining (*IGHJ*, *IGKJ*, *IGLJ*) germline gene segments [2]. In addition, the imprecise junction of the germline gene segments and somatic hypermutations contribute to antibody diversity [3-5].

Species differences were primarily found in the number of germline *IGHV*/*IGKV*/*IGLV*, *IGHD*, and *IGHJ*/*IGKJ*/*IGLJ* segments. In livestock species with restricted combinatorial germline diversity such as chicken [4], pigs [6], sheep [7], and cattle [5, 8, 9], species-dependent mechanisms dominate the different diversification steps. For instance, the use of pseudogene sequence parts is a frequent post-recombinatorial strategy for the generation of the preimmune antibody repertoire in chicken, sheep, and rabbit [4, 10-13]. This phenomenon, called gene conversion, was also confirmed for IGLs in cattle [14] and is assumed to be operative in horses [15].

Gene conversions are difficult to detect especially within a large number of sequences e.g. like those obtained from high throughput sequencing. Gene conversion in immunoglobulins is characterized by clusters of nucleotide changes [14], sometimes only triplets [11], originating from upstream genes of the rearranged segment [4, 13]. High degree of flanking homology of the conversion region ensures the genetic exchange [13], whereby 3 to 5 nucleotides seem to be the minimal overlapping requirement [11].

Detection of gene conversion in bovine IGHV is complicated due to the incomplete IGH locus annotation. The main bovine IGH locus was assigned to the *Bos taurus* autosome (BTA) 21 but exons coding for variable, diversity, and joining segments were also found on BTA7, BTA8, and BTA20 [16-18]. Thirty-six *IGHV*s were identified of which 13 are functional and belong phylogenetically to the bovine *IGHV* family 1 (boVH1). The second bovine *IGHV* family consists solely of non-functional *IGHV*s that have not been identified in expression analyses yet. Eleven *IGHV* segment pairs shared 100% sequence identity, whereas two of these pairs contain a functional segment and either an ORF or a putative functional segment, respectively [17]. The

high proportion of pseudogene segments leads to the assumption of their use in gene conversion events. Two *IGHJ* loci possessing six *IGHJ* segments were detected on BTA11 by BAC clone and locus-specific PCR analysis and were found to rearrange at low frequency while those located on BTA21 rearrange at high frequency. Only two out of these six *IGHJ* were classified as functional whereas one is involved predominantly in the recombination process [19, 20]. Fifteen *IGHD* genes were detected and revealed a sub-cluster organization. *IGHD* are classified into four families and the *IGHD* exons revealed huge size differences [21, 22]. The organization of the actual bovine germline repertoire and its possible allelic variants is incomplete and needs to be investigated in more detail [17]. Since, even the organization of the extensively studied human immunoglobulin germline repertoire is questioned and requires ongoing analyses [23].
In all rearranged bovine immunoglobulin isotypes, exceptionally long complementarity determining region 3 of the heavy chain (CDR3H) possessing up to 67 aa were described [17]. Together with *IGHD2* and *IGHJ1*, the germline *IGHV*10 segment was found to be the only variable segment rearranged in these exceptionally long CDR3Hs [17, 24]. This mechanism is also not isotype restricted [17, 25].
An additional bovine specific mechanism for antibody diversification is the insertion of conserved short nucleotide sequences into the *IGHV-IGHD* junction, which was found in intermediate and exceptionally long CDR3Hs [24].
Currently available programs like IMGT/Junction Analysis [26], IMGT/V-QUEST [27, 28] and IMTG/HIGHV-QUEST [29], VBASE2 [30], JoinSolver [31], iHMMun-align [32], and IgBLAST [33] allow the annotation of only the entire IGHV sequence to germline *IGHV* segments. Differentiated analysis of single parts is not directly possible. Most of the databases are focusing on mouse and human immunoglobulin genes (VBASE2, human, [30]). For cattle and other livestock or companion animals, separate databases have to be created (IgBLAST, [33]). Furthermore, the IMGT numbering system does not provide numbering for CDR3H larger than 31 aa. Placing of bovine intermediate as well as of exceptional long CDR3H in this numbering system is therefore not possible and consequently does not allow correct analysis of the rearrangement in those immunoglobulins. In addition, only IgBLAST allows the adjustment of parameters for *IGHD* identification.
Detailed genetic analysis of the pre- and post-immunization humoral immune response is important to describe the developing diversity and the effectiveness of vaccines and to detect possible individual and breed related differences including non-responders.

As a conclusion those analyses help to develop fast recombinant antibodies for passive vaccination, therapy or diagnostic by genetically pre-selection of newly developed or abundant sequences.

As a first attempt to gain more insight into bovine breed specific differences, we compared the dominantly transcribed and the combinatorial diversity of germline *IGHV*, *IGHD*, and *IGHJ* segments as well as somatic hypermutations and putative gene conversions of IgG in the four cattle breeds Aubrac, German Simmental, German Black Pied and Holstein Friesian, by using a newly developed Bioinformatics framework. This new bioinformatics framework combines and extends several analysis tools and takes into account the unique specificities of bovine immunoglobulin sequences of exceptionally long CDR3Hs. In addition, it allows for the adaptation of alignment parameters for the single segments and enables the selective analysis of the different functional regions of the variable domain (namely framework regions and CDRs) to determine putative gene conversions. This new tool should facilitate a fast and detailed analysis of data sets generated by high throughput sequencing.

Results

For sequence analysis, we developed a new bioinformatics framework using MUSCLE [34, 35] for the initial fast and accurate multiple nucleotide sequence alignment. Subsequently, the sequence distances were calculated with ClustalW [36]. For nucleotide alignments of *IGHV* and *IGHJ*, default values of MUSCLE were used. To improve the biological significance of the assignment of germline and sample *IGHD*s, we tested three different procedures using different parameters. To determine the germline origin using the new bioinformatics framework, only the FRHs were aligned to avoid interference with the highly diversified CDRH [18]. To analyze possible gene conversion events, FR1-3Hs and CDR1-3Hs were extracted and aligned separately to the corresponding regions of the *IGHV* reference sequences to find the most similar germline segment.

Using our bioinformatics tool, we established a sample sequence set for use in detailed analysis of the transcribed bovine immunoglobulin repertoire. Blood samples were taken from 10 animals per cattle breed: Aubrac (A), German Simmental (GS), German Black Pied (GBP), and Holstein Friesian (HF).

In total, 160 IgG heavy chain sequences per breed (n=640 sequences) were investigated as described above. The variable regions were identified and extracted at the 5’ end (N-terminal end) using the motif GCCTCCACC/AST (nt/AA) marking the start of the first constant region of all bovine IgGs. Due to premature Stop-codons or incompletely amplified variable regions, 131 sequences were excluded from further analyses. Consequently, 509 sequences were left: 137 in A, 116 in GS, 111 in GBP, and 145 in HF. Sequences analyzed are published under accession numbers KT761498-KT762006.

Transcriptional analyses, assignment of germline gene segments

Assigning the FR1H to FR3H of transcribed IGHV segments to their germline origin

Ig heavy chain gene usage and identity to germline gene segments was determined by comparing the transcribed sequences with the germline Ig heavy chain genes described by Walther et al. [17] and Liljavirta et al. [22]. We found six of the germline *IGHV*s possess 100% sequence identity (presented as *IGHVx/y*) up to the 3' end of FR3H. Analysis of transcribed IGHV segments (here: comprising FR1-3H) revealed germline gene usage of *IGHV3/33*, *IGHV6*, *IGHV10/34*, *IGHV36/29(F)*, *IGHV17(ORF)/31(F)*, *IGHV1S26*, *IGHV1S28*, *IGHV1S32*, *IGHV1S33*, *IGHV1S34*, *IGHV1S35*, *IGHV1S37*, *IGHV1S38*, *IGHV1S39* as well as *IGHV1S40* (Fig 1).

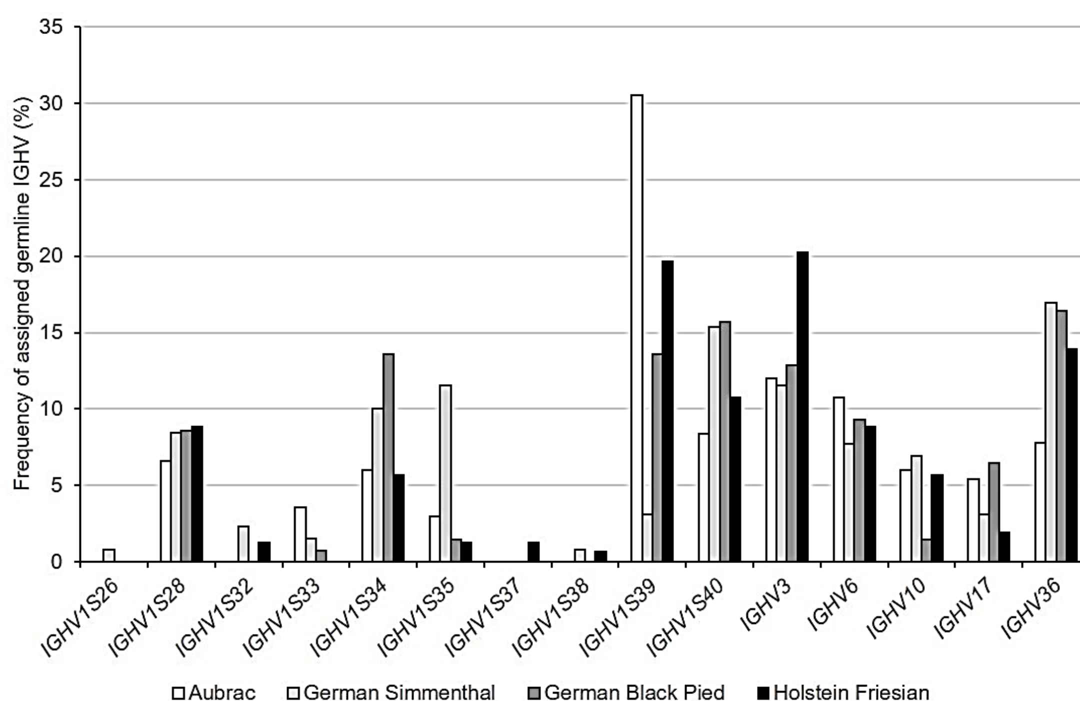


Fig 1. Transcription frequencies of *IGHV* in four cattle breeds. Transcribed IGHV are shown on the horizontal axis, their relative usage frequencies are indicated on the vertical axis. Each breed is marked by the following color code: Aubrac: white, German Simmental: light grey, Holstein Friesian: black, German Black Pied: dark grey

We were not able to unambiguously assign ten of 116 analyzed GS sequences to one germline *IGHV*; this was also true for 14 out of 111 GBP sequences, 12 out of 145 HF sequences, and 19 out of 137 analyzed sequences of A.

These samples showed equal divergence to at least two germline *IGHV* which included combinations of *IGHV3/33*, *IGHV6*, *IGHV10/34*, *IGHV17(ORF)/31(F)*, *IGHV1S28*, and *IGHV1S32-40*. The most frequent multiple assignment affected *IGHV6* and *IGHV1S34*, followed by *IGHV6* and *IGHV1S35* in all A, GS, and HF. *IGHV6*, *IGHV1S34*, and *IGHV17(ORF)/31(F)* could not be differentiated in two sequences of GS, one sequence of GBP and two sequences of HF. Ambiguously assigned germline *IGHV*s and multiple assignments are listed in Table 1.

Table 1. *IGHV* assigned ambiguously and their frequency. Ambiguous sequences can be assigned to more than one germline *IGHV* with the same distance.

Breed	**A**	**GBP**	**GS**	**HF**
Total No. of animals	10	10	10	10
No. of animals with ambiguous sequences	9	5	4	6
Total No. of sequences analyzed	137	111	116	145
No. of ambiguous sequences	19	14	10	12
Total No. of putative *IGHV* recombinations	167	140	130	158
Ambiguous *IGHV*				
IGHV3/IGHV1S33	0	0	1	0
IGHV3/IGHV1S39	0	0	0	2
IGHV6/IGHV17	1	0	0	0
IGHV6/IGHV1S34	3	3	3	5
IGHV6/IGHV1S35	3	0	1	1
IGHV6/IGHV36	0	1	1	1
IGHV6/IGHV1S39	1	0	1	0
IGHV1S28/IGHV1S32	0	0	0	1
IGHV1S33/IGHV1S39	1	1	0	0
IGHV1S33/IGHV1S40	1	0	0	0
IGHV1S37/IGHV1S38	0	0	0	1
IGHV1S39/IGHV17	0	1	0	0
IGHV1S39/IGHV1S40	1	0	0	0
IGHV3/IGHV10/IGHV1S39	2	0	0	0
IGHV6/IGHV17/IGHV1S34	0	2	2	1
IGHV6/IGHV17/IGHV1S40	1	0	0	0
IGHV6/IGHV1S32/IGHV1S34	1	0	1	0
IGHV6/IGHV1S33/IGHV1S35	1	0	0	0
IGHV3/IGHV6/IGHV17/IGHV1S34	0	2	0	0
IGHV6/IGHV10/IGHV1S34/IGHV1S39	1	0	0	0
IGHV6/IGHV17/IGHV1S28/IGHV1S39	0	1	0	0
IGHV6/IGHV17/IGHV1S34/IGHV1S39	1	1	0	0
IGHV6/IGHV36/IGHV1S34/IGHV1S39	0	1	0	0
IGHV3/IGHV6/IGHV17/IGHV1S34/IGHV1S39	0	1	0	0
IGHV6/IGHV17/IGHV1S33/IGHV1S34/IGHV1S35	1	0	0	0

Including the multiple assignments as described above, a total number of 595 possible transcribed germline *IGHV* were observed. Overall, the most frequent variable gene segment was *IGHV1S39*; this was identified in 17.65% of all sequences. This *IGHV* was used in 51 A sequences (n=167, 30.54%), in 19 sequences of GBP (n=140, 13.57%), in four sequences of GS (n=130, 3.08%), and 31 sequences of HF (n=158, 19.62%) (Fig 1, Table 2). *IGHV3/33* was represented by 14.29% of all sequences. The number of transcribed *IGHV3/33* varied from 15 in GS (11.54%) to 32 in HF (20.25%) (Fig 1, Table 2). In similar proportions of 13.45% and 12.27% germline *IGHV36/29(F)* and *IGHV1S40* were used. *IGHV1S28*, *IGHV1S34*, and *IGHV6* were transcribed in proportions of 8.07%, 8.57%, and 9.24%. The other transcribed IGHVs were identified in minor proportions of 0.17% to 5.04%, respectively. These rarely used *IGHV*s were identified once or twice in GS and HF but up to ten times in A (Fig 1). Very high significant differences were calculated for *IGHV* usage within the breeds and between the breeds (P<0.0001).

Table 2. Percentage of *IGHV* assigned to sample sequences of four cattle breeds.

IGHV	**A (%; n=167)**	**GS (%; n=130)**	**GBP (%; n=158)**	**HF (%; n=140)**
IGHV1S26	0.00	0.77	0.00	0.00
IGHV1S28	6.59	8.46	8.57	8.86
IGHV1S32	0.00	2.31	0.00	1.27
IGHV1S33	3.59	1.54	0.71	0.00
IGHV1S34	5.99	10.00	13.57	5.70
IGHV1S35	2.99	11.54	1.43	1.27
IGHV1S37	0.00	0.00	0.00	1.27
IGHV1S38	0.00	0.77	0.00	0.63
IGHV1S39	30.54	3.08	13.57	19.62
IGHV1S40	8.38	15.38	15.71	10.76
IGHV3	11.98	11.54	12.86	20.25
IGHV6	10.78	7.69	9.29	8.86
IGHV10	5.99	6.92	1.43	5.70
IGHV17	5.39	3.08	6.43	1.90
IGHV36	7.78	16.92	16.43	13.92

Separated analyses of FRH 1-3 and CDRH 1-2 to detect gen conversion events

In livestock such as chicken, rabbit, and cattle the use of pseudogene segments is known to contribute to immunoglobulin diversity [4, 10, 11, 13, 14, 37, 38]. Gene conversion in immunoglobulins is characterized by clusters of nucleotide changes [14], sometimes only triplets [11], originating from upstream genes of the rearranged segment [4, 13]. A high degree of flanking homology of the conversion region ensures the genetic exchange [13], whereby three to five nucleotides have been shown to be the minimal overlapping requirement [11].

Consequently, FR1-3H and CDR1-2H were analyzed separately to identify mutations within the FR1-3H and CDR1-2H that would indicate possible gene conversion events. The current genomic annotation of germline immunoglobulin segments in cattle makes a correct prediction of 5' donor segments difficult. Nonetheless, larger contigs have been identified [17] and enabled us to show gene conversion events in bovine immunoglobulin heavy chains variable regions.

For instance, the calculated putative originating germline genes for nucleotide changes identified solely in the CDR2H region of KT761864 were *IGHV4Ψ*, *IGHV9Ψ*, and *IGHV18Ψ* (divergence 0.190). In the calculation covering the whole V-region and using only FR1-3H, *IGHV6* was identified as the originating gene for KT761864 (divergence 0.075). All genes but *IGHV18Ψ* are located on BTA21, whereas *IGHV4Ψ* is located upstream of *IGHV6* and is therefore most likely used for the gene conversion. There are two triplets in KT761864 and *IGHV4Ψ* that are different from *IGHV6* due to a transversion mutation in the first changed triplet (from AAT in *IGHV6* to TAT in KT761864 and *IGHV4Ψ*) and a transition mutation in the second changed triplet (from GAT in *IGHV6* to AAT in KT761864 and *IGHV4Ψ*; Table 3).

Table 3. Possible gene conversion in KT761864.

Sample/Gene	**Sequence in CDR2H**	**Location**	**Divergence**
KT761864	ATT AAT TAT AAT … GGA GAC ACC	BTA21	
IGHV4Ψ	ATA TAT TAT AAT … GGT GAC ACT	BTA21	0.190
IGHV9Ψ	ATA TAT TAT AAT … GGT GAC ACT	BTA21	0.190
IGHV18Ψ	ATA TAT TAT AAT … GGT GAC ACT	BTA7	0.190
IGHV6[1]	ATA GAT AAT GAT … GGA GAC ACA	BTA21	0.075

[1] Divergence in calculations covering the whole V-region and using only FR1-3H. This gene was not observed in the analysis of the separated CDR2H.

Due to high sequence similarities and because we assumed preferable transcription of functional germline gene segments, we first concentrated on pseudogenes that were calculated as unique origin of the sample sequences considering the particular regions.

In all breeds, *IGHV13Ψ* and *IGHV17(ORF)* were used most often in CDR2H (3.34% and 1.77%) as well as in FR2H and FR3H (*IGHV17(ORF)* both 1.38%) and CDR2H (*IGHV13Ψ* 1.77%), respectively. *IGHV15Ψ* was found in 0.59% of CDR2H in sequences of A and HF. *IGHV11Ψ*, *IGHV12Ψ*, and *IGHV21Ψ* always showed the same divergence from GBP and HF sequences but were not solely identified as the potential origin of the sample sequence. Similarly, *IGHV4Ψ*, *IGHV9Ψ*, and *IGHV18Ψ* were identified in triplet as possible originating gene segments but only together with functional germline genes. In one sequence of GBP and 2 sequences of GS both in CDR1H and CDR2H pseudogenes were calculated as possible parental *IGHV*s. In one A sequence only a pseudogene showed lowest divergence from the sample in FR2H and CDR2H (Fig 2).

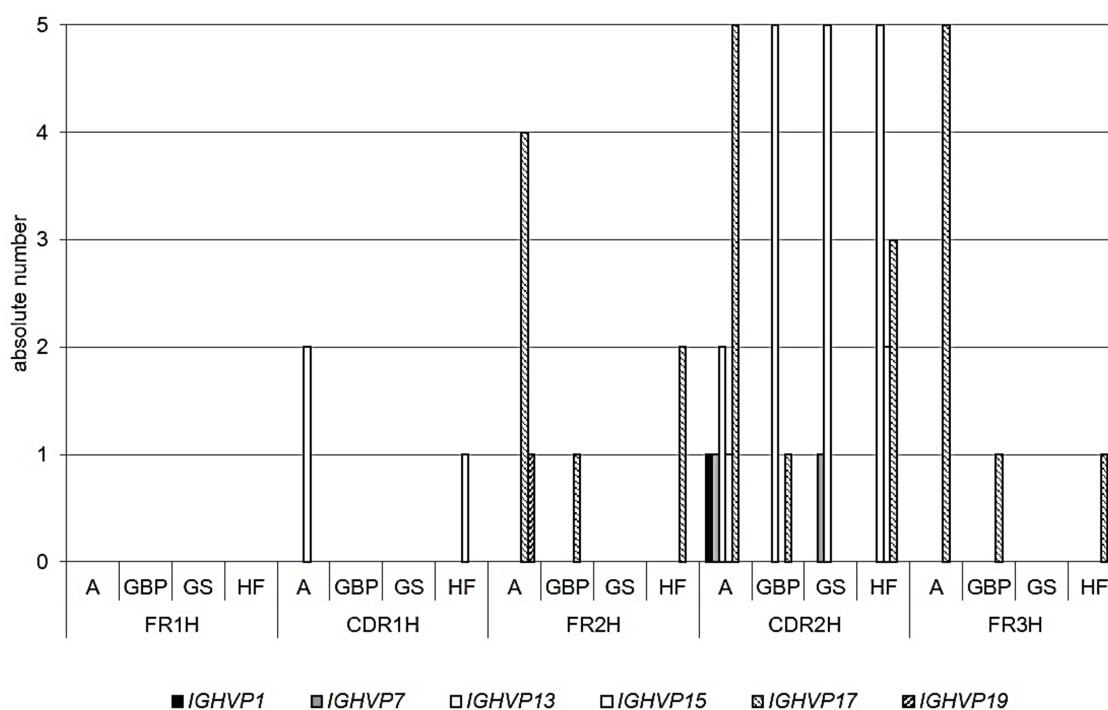


Fig 2. Possible gene conversion in FRH and CDRH. The absolute number of possible gene conversion events is shown for pseudogenes assigned unambiguously to CDR1H, FR2H, and CDR2H in the four cattle breeds Aubrac (A), German Black Pied (GBP), German Simmental (GS), and Holstein Friesian (HF).

Our analyses revealed a less distinct assignment to the germline gene segments. Nevertheless, the majority of identified and putative originating germline sequences were *IGHV3/33*, *IGHV6*, *IGHV10/34*, *IGHV2/26*, *IGHV29(F)/36*, *IGHV16(ORF)/25*, *IGHV17(ORF)/31(F)*, and *IGHV19Ψ*, as well as *IGHV1S19*, *IGHV1S23-28*, *IGHV1S30*, *IGHV1S32-40* [22] within FRHs and CDRHs. Although these are the same gene segments as determined for the complete IGHV segment, calculations indicate exchanges between the IGHV gene segments within FRHs and CDRHs. Additional pseudogenes (*IGHV1Ψ*, *IGHV4Ψ*, *IGHV7Ψ*, *IGHV9Ψ*, *IGHV11Ψ*, *IGHV12Ψ*, *IGHV13Ψ*, *IGHV15Ψ*, *IGHV18Ψ*, *IGHV21Ψ*) were identified by the program and concentrated on CDR1H and CDR2H. *IGHV13Ψ*, *IGHV15Ψ*, and *IGHV19Ψ* showed little divergence in FR1H and FR2H from sample sequences, too. Within these FRHs, different functional *IGHV* were possible originating germline sequences (Table 4).

Table 4. Possible gene conversions in FRH and CDRH.

		A		GBP		GS		HF		Sum		Percent
		a[1]	u[2]	a[1]	u[2]	a[1]	u[2]	a[1]	u[2]	a[1]	u[2]	u[2]
IGHV1Ψ	CDR1H	1	0	0	0	0	0	1	0	2	0	0.00
	CDR2H	1	1	1	0	1	0	2	0	5	1	0.20
IGHV4Ψ	CDR1H	1	0	0	0	0	0	1	0	2	0	0.00
	CDR2H	4	0	1	0	2	0	2	0	9	0	0.00
IGHV7Ψ	CDR2H	2	1	1	0	1	1	4	0	8	2	0.39
IGHV9Ψ	CDR1H	1	0	0	0	0	0	1	0	2	0	0.00
	CDR2H	4	0	1	0	2	0	2	0	9	0	0.00
IGHV11Ψ	CDR2H	0	0	0	0	0	0	2	0	2	0	0.00
IGHV12Ψ	CDR2H	0	0	0	0	0	0	2	0	2	0	0.00
IGHV13Ψ	CDR1H	18	0	11	0	9	0	19	0	57	0	0.00
	FR2H	1	0	0	0	0	0	0	0	1	0	0.00
	CDR2H	16	2	20	5	17	5	22	5	75	17	3.34
IGHV17Ψ	FR1H	80	0	58	0	63	0	75	0	276	0	0.00
	CDR1H	20	0	10	0	2	0	10	0	42	0	0.00
	FR2H	8	4	4	1	4	0	6	2	22	7	1.38
	CDR2H	12	5	7	1	3	0	14	3	36	9	1.77
	FR3H	9	5	10	1	9	0	8	1	36	7	1.38
IGHV15Ψ	CDR1H	7	2	3	0	3	0	4	1	17	3	0.59
	FR2H	0	0	0	0	0	0	0	0	0	0	0.00
	CDR2H	3	1	9	0	1	0	9	2	22	3	0.59
IGHV18Ψ	CDR1H	1	0	0	0	0	0	1	0	2	0	0.00
	CDR2H	4	0	1	0	2	0	2	0	9	0	0.00
IGHV19Ψ	FR1H	80	0	58	0	63	0	75	0	276	0	0.00
	CDR1H	25	0	13	0	27	0	20	0	85	0	0.00
	FR2H	6	1	1	0	0	0	4	0	11	1	0.20
	CDR2H	9	0	15	0	9	0	25	0	58	0	0.00
IGHV21Ψ	CDR2H	0	0	0	0	0	0	2	0	2	0	0.00

[1] ambiguously assigned
[2] unambiguously assigned

CDR3H length distribution

In all four cattle breeds the program identified very short CDR3Hs (less or equal 10 aa, group 1), CDR3Hs of intermediate length (11-47 aa, group 2) as well as exceptionally long CDR3Hs (at least 48 aa, group 3) (Fig 3, Table 5). Very high significant differences were calculated for the number of sequences within the 3 groups of lengths (P<0.0001) when they were compared between all breeds but also within the breeds (GS, GBP, A, HF: P<0.0001). Fourty-four (7.37%) sequences possessed a CDR3H length with 10 or less amino acid (aa) residues. The highest amount of sequences within this group were identified in breed A (12.57%), followed by HF (7.55%), GBP (5.00%) and GS (3.05%). In the breeds A, GS, and HF one and four sequences possessed only four amino acid residues within the CDR3H. Five amino acid residues were the shortest CDR3H in sequences of GBP.

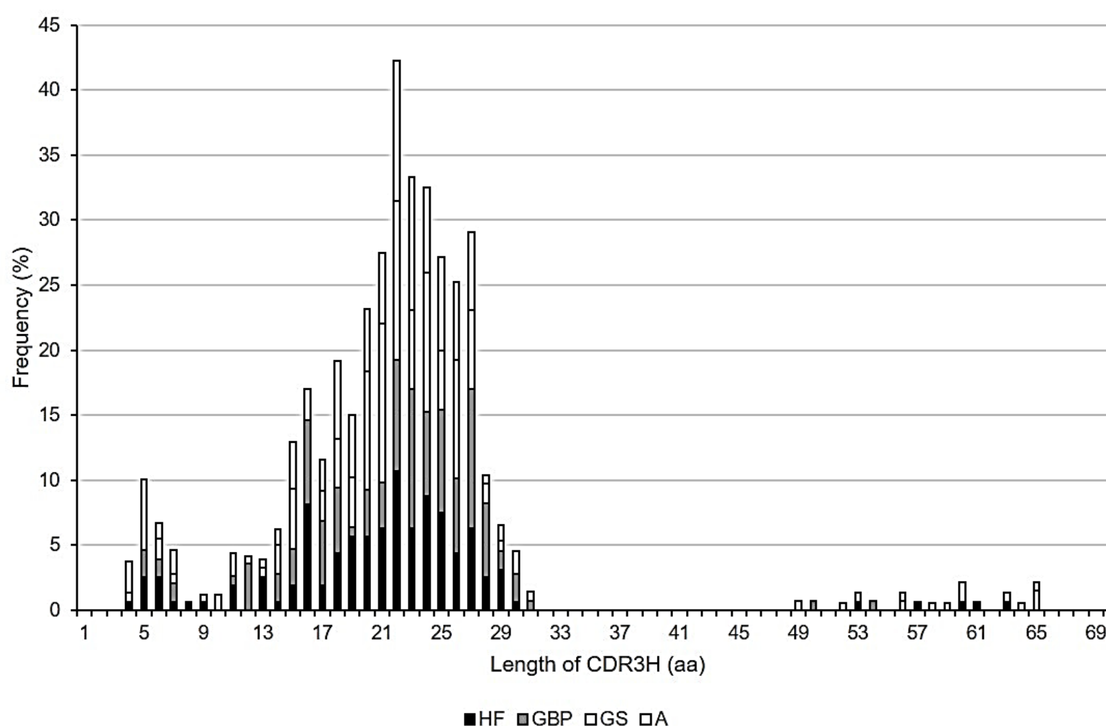


Fig 3. Length distribution of CDR3H in the four cattle breeds. The number of amino acid residues making up the CDR3H lengths identified are shown on the horizontal axis, their relative usage frequencies are indicated on the vertical axis. Each breed is marked by the following color code: Aubrac: white, German Simmental: light grey, Holstein Friesian: black, German Black Pied: dark grey

Table 5. Length distribution of the CDR3Hs.

Number of amino acids within CDR3H	All breeds (%; n=597)	GS (%; n=131)	GBP (%; n=140)	HF (%; n=159)	A (%; n=167)
4	1.01	0.76	0.00	0.63	2.40
5	2.68	0.00	2.14	2.52	5.39
6	1.68	1.53	1.43	2.52	1.20
7	1.17	0.76	1.43	0.63	1.80
8	0.17	0.00	0.00	0.63	0.00
9	0.34	0.00	0.00	0.63	0.60
10	0.34	0.00	0.00	0.00	1.20
11	1.17	0.00	0.71	1.89	1.80
12	1.01	0.00	3.57	0.00	0.60
13	1.01	0.76	0.00	2.52	0.60
14	1.51	2.29	2.14	0.63	1.20
15	3.18	4.58	2.86	1.89	3.59
16	4.36	0.00	6.43	8.18	2.40
17	2.85	2.29	5.00	1.89	2.40
18	4.86	3.82	5.00	4.40	5.99
19	3.85	3.82	0.71	5.66	4.79
20	5.70	9.16	3.57	5.66	4.79
21	6.70	12.21	3.57	6.29	5.39
22	10.55	12.21	8.57	10.69	10.78
23	8.38	6.11	10.71	6.29	10.18
24	8.04	10.69	6.43	8.81	6.59
25	6.87	4.58	7.86	7.55	7.19
26	6.20	9.16	5.71	4.40	5.99
27	7.20	6.11	10.71	6.29	5.99
28	2.51	1.53	5.71	2.52	0.60
29	1.68	0.76	1.43	3.14	1.20
30	1.17	0.00	2.14	0.63	1.80
31	0.34	0.76	0.71	0.00	0.00
49	0.17	0.76	0.00	0.00	0.00
50	0.17	0.00	0.71	0.00	0.00
52	0.17	0.00	0.00	0.00	0.60
53	0.34	0.76	0.00	0.63	0.00
54	0.17	0.00	0.71	0.00	0.00
56	0.34	0.76	0.00	0.00	0.60
57	0.17	0.00	0.00	0.63	0.00
58	0.17	0.00	0.00	0.00	0.60
59	0.17	0.00	0.00	0.00	0.60
60	0.50	1.53	0.00	0.63	0.00
61	0.17	0.00	0.00	0.63	0.00
63	0.34	0.76	0.00	0.63	0.00
64	0.17	0.00	0.00	0.00	0.60
65	0.50	1.53	0.00	0.00	0.60

Complementarity determining regions of group 2 were identified in 532 of all sequences analyzed (89.11%). 93.57% of GBP sequences were found to use 11 up to 47 aa. In HF, 89.31% of the sequences were attributed to this group as well as 83.83% of A and 90.84% of GS sequences. The most frequent CDR3H length was 22 aa, which was found in 10.78% of A sequences, in 8.57% of GBP, in 12.21% of GS, and in 10.69% HF sequences. Simultaneously, this length was identified preferably in CDR3H of A and HF. Nevertheless, in GS CDR3H with a length of 21 aa was identified as often as a length of 22 aa (12.21%). In GBP, CDR3Hs with a length of 23 and 27 aa dominated (10.71%).

CDR3Hs of group 3 were identified in 21 sequences of all four breeds (9.22%). The breed GS showed the highest number of these sequences (6.11%) followed by A (3.59%). German Black Pied and HF sequences possessed smaller proportions of the exceptionally long CDR3H with 3.14% and 1.43%, respectively. Whereas in GS and HF sequences with 65 aa were the longest CDR3Hs (1.53%, 0.6%), 63 and 54 aa were counted in the longest CDR3Hs of A (0.63%) and GBP (0.71%), respectively.

Assigning IGHD to their germline origin using 3 different procedures

We tested three different procedures to assign germline and sample *IGHD*s in order to improve the biological significance. At first, we applied the default values of MUSCLE, in procedure 2 we changed the penalties for gap opening (=-4) and gap extention (=-0.3), and in procedure 3 we additionally incorporated a modified scoring matrix (match +2) to evaluate transversion and transition mutations.

In all three procedures, the assignment of germline and transcribed *IGHD* revealed clear results for the sequences analyzed. Nevertheless, using procedure 3 we obtained results matching short and long sample IGHD sequences best to germline short CDR3H and exceptionally long CDR3H, respectively, whereas procedure 1 and 2 assigned a major number of group 3 CDR3H to germline *IGHD* of moderate length. Twenty-one different germline *IGHD* were transcribed, whereby *IGHD8* located on BTA21, *IGHD3* located on BTA7, *IGHD5* as well as the very short *IGHDQ52* located on BTA8; these were preferred in all breeds (Fig 4, Table 6). While *IGHDQ52*, *IGHDS10* and *14* [22] were solely transcribed in sense orientation, germline *IGHD1* to *IGHD8* gene segments were transcribed in antisense direction in 38 sequences distributed over all four breeds investigated. *IGHD4* (antisense (as), [21]) was identified the most often, followed by *IGHD1* (as, [39]), *IGHD3* (as, [39]), *IGHD2* (as, [39]), *IGHD5* (as,

[21]), *IGHD6* (as, [21]), *IGHD8* (as, [21]) and *IGHD7* (as, [21]). Very high significant differences were calculated for the *IGHD* usage within the cattle breeds (P<0.0001) but not between the breeds (P=0.06).

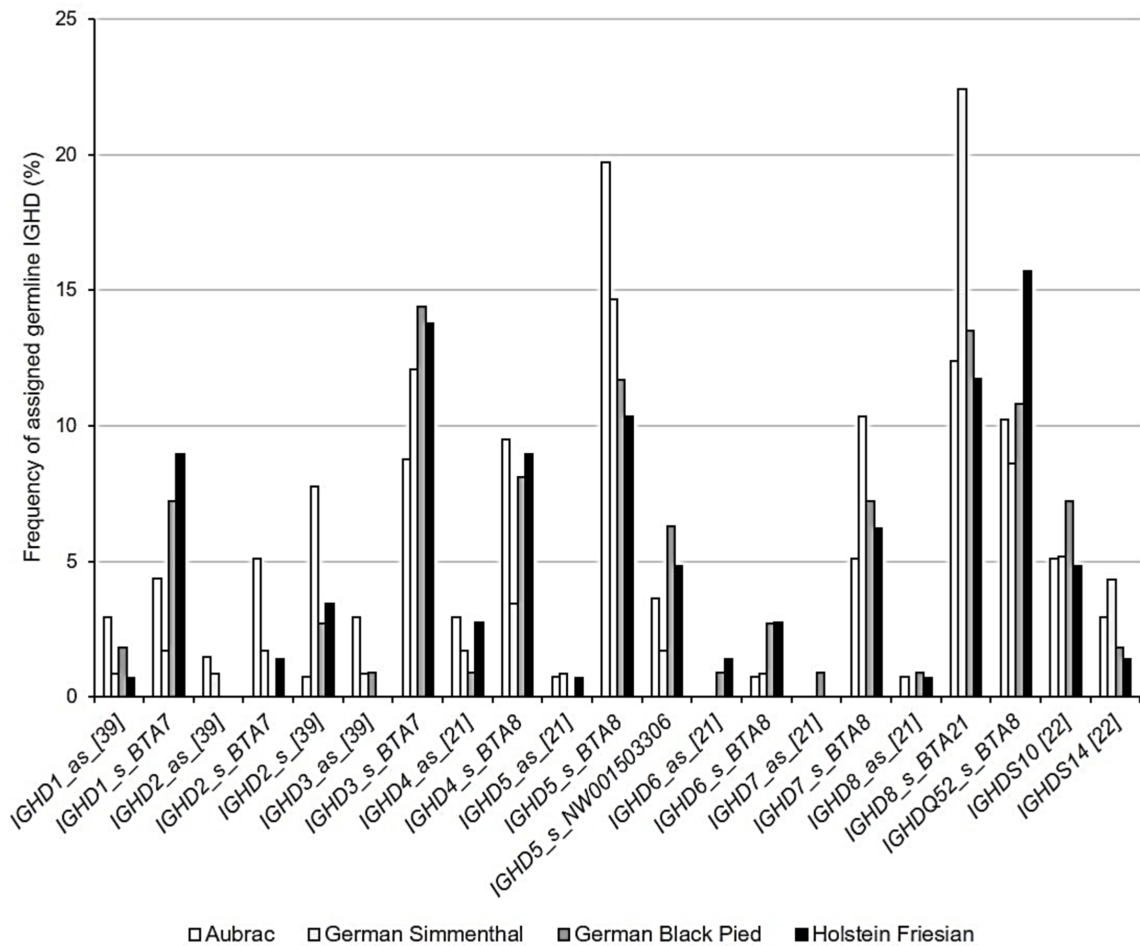


Fig 4. Transcription frequencies of *IGHD* in four cattle breeds using procedure 3. Transcribed IGHD are shown on the horizontal axis, their relative usage frequencies are indicated on the vertical axis. Calculation occurred after changing the default values for gap opening (-4) and gap extention (-0.3) and a modified scoring matrix (match +2) of MUSCLE. Each breed is marked by the following color code: Aubrac: white, German Simmental: light grey, Holstein Friesian: black, German Black Pied: dark grey

Table 6. Percentage of *IGHD* assigned to sample sequences of four cattle breeds.

IGHD	A (%; n=137)	GS (%; n=116)	GBP (%; n=111)	HF (%; n=145)
*IGHD1*_as[1]_[39]	2.92	0.86	1.80	0.69
*IGHD1*_s[2]_BTA7	4.38	1.72	7.21	8.97
*IGHD2*_as_[39]	1.46	0.86	0.00	0.00
*IGHD2*_s_BTA7	5.11	1.72	0.00	1.38
*IGHD2*_s_[39]	0.73	7.76	2.70	3.45
*IGHD3*_as_[39]	2.92	0.86	0.90	0.00
*IGHD3*_s_BTA7	8.76	12.07	14.41	13.79
*IGHD4*_as_[21]	2.92	1.72	0.90	2.76
*IGHD4*_s_BTA8	9.49	3.45	8.11	8.97
*IGHD5*_as_[21]	0.73	0.86	0.00	0.69
*IGHD5*_s_BTA8	19.71	14.66	11.71	10.34
*IGHD5*_s_NW001503306	3.65	1.72	6.31	4.83
*IGHD6*_as_[21]	0.00	0.00	0.90	1.38
*IGHD6*_s_BTA8	0.73	0.86	2.70	2.76
*IGHD7*_as_[21]	0.00	0.00	0.90	0.00
*IGHD7*_s_BTA8	5.11	10.34	7.21	6.21
*IGHD8*_as_[21]	0.73	0.00	0.90	0.69
*IGHD8*_s_BTA21	12.41	22.41	13.51	11.72
*IGHDQ52*_s_BTA8	10.22	8.62	10.81	15.71
*IGHDS10*_[22]	5.11	5.17	7.21	4.83
*IGHDS14*_[22]	2.92	4.31	1.80	1.38

[1] antisense
[2] sense

In comparison, 20 different germline *IGHD* were transcribed using default conditions of MUSCLE for alignments (procedure 1), whereby *IGHD8* being located on BTA21 was the preferred *IGHD* in all breeds. The very short *IGHDQ52* located on BTA8 was transcribed in high frequencies in the breeds A, GBP, and HF (S1 Fig, S1 Table). While *IGHDQ52*, *IGHDS10* and *14*, and *IGHD6* were solely transcribed in sense orientation, germline *IGHD1* to *IGHD8* gene segments were transcribed in antisense direction in 23 sequences that were distributed over all four breeds. *IGHD4* (as, [21]) was identified the most often. Very high significant differences were calculated for the *IGHD* usage within the cattle breeds ($P<0.0001$) but not between the breeds ($P=0.1630$).

After changing penalties for gap opening and gap extension (procedure 2), 17 different germline *IGHD* were transcribed, whereby *IGHD8* located on BTA21, was the most observed *IGHD* in all breeds (S2 Fig, S2 Table). Germline *IGHD1*, *IGHD4*, *IGHD6*, and *IGHD8* gene segments were transcribed in as direction in 24 sequences that were distributed over all four breeds. Again, *IGHD4* (as, [21]) was identified the most often. Very high significant differences were calculated for the *IGHD* usage within the cattle breeds (P<0.0001) but not between the breeds (P=0.6654).

Assigning the FR4H and IGHJ to their germline origin

Located on BTA21, *IGHJ1* and *IGHJ6* [22] were identified as origin to the transcribed gene segments, which defines the FR4H within the samples analyzed. *IGHJ1* was transcribed preferably in the sequences investigated in all breeds (98.83%). *IGHJ6* [22] (1.17%) was detected in only one sequence of each A, GS, and GBP (0.73%, 0.85%, 0.90%), as well as in three sequences of HF (2.05%) animals. One sequence of GS and HF, respectively, showed *IGHJ1* and *IGHJ6* as a possible originating germline segment (Fig 5, Table 7). Statistical analysis revealed very high significant differences (P<0.0001) for the usage of *IGHJ1* and *IGHJ6* [22] within all cattle breeds. No significant differences were calculated between the four breeds.

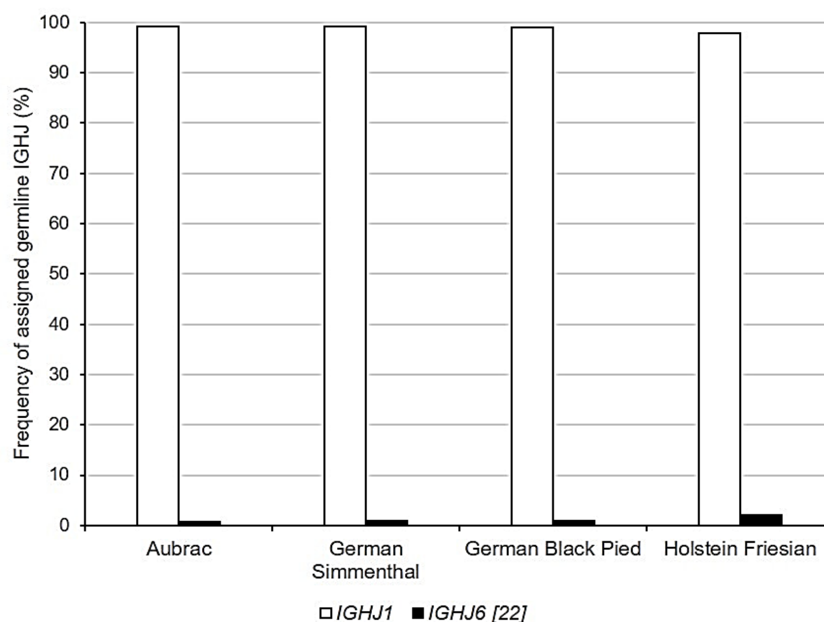


Fig 5. Transcription frequencies of *IGHJ* in four cattle breeds. Transcribed IGHJ are shown on the horizontal axis, their relative usage frequencies are indicated on the vertical axis. Each breed is marked by the following color code: Aubrac: white, German Simmental: light grey, Holstein Friesian: black, German Black Pied: dark grey

Table 7. Percentage of *IGHJ* assigned to sample sequences of four cattle breeds.

IGHJ	A (%; n=137)	GS (%; n=117)	GBP (%; n=111)	HF (%; n=146)
IGHJ1	99.27	99.15	99.10	97.95
IGHJ6 [22]	0.73	0.85	0.90	2.05

Recombination of IGHV, IGHD, and IGHJ in different cattle breeds using procedure 3

Recombined *IGHV*, *IGHD*, and *IGHJ* were identified for each sequence and all possible frequencies were analyzed statistically within and between the breeds examined. In total, 597 recombinations were analyzed including double assigned germline origins for *IGHV* and *IGHJ*.

Applying procedure 3 for the *IGHD* assignment, 162 different combinations of *IGHV*, *IGHD*, and *IGHJ* were revealed (Fig 6, S3 Table). Most combinations occurred in less than ten sequences. Calculations revealed very high significant differences in usage frequencies of the rearranged gene segments between the breeds investigated (P<0.0001). Eleven rearrangements were observed in ten to 27 sequences. They were observed in sequences of all four breeds. These major rearrangements were: IGHV36-IGHD8 (sense (s), BTA21)-IGHJ1 (AY158087) (4.52%), IGHV1S40-IGHD5 (s, BTA8)-IGHJ1 (AY158087) (3.02%), IGHV1S39-IGHD5 (s, BTA8)-IGHJ1 (AY158087) (2.85%), IGHV1S40-IGHD3 (s, BTA7)-IGHJ1 (AY158087) (2.35%), IGHV1S39-IGHD8 (s, BTA21)-IGHJ1 (AY158087) (2.18%), IGHV3-IGHD5 (s, BTA8)-IGHJ1 (AY158087) (2.01%), IGHV3-IGHDQ52 (s, BTA8)-IGHJ1 (AY158087) (1.84%), IGHV36-IGHDQ52 (s, BTA8)-IGHJ1 (AY158087) (1.84%), IGHV1S39-IGHD4 (s, BTA8)-IGHJ1 (AY158087) (1.68%), IGHV1S39-IGHD7 (s, BTA8)-IGHJ1 (AY158087) (1.68%), IGHV3-IGHD8 (s, BTA21)-IGHJ1 (AY158087) (1.68%). Beside these preferred combinations, 52 minor recombinations were identified solely in one sequence distributed with 19 sequences in A, 14 in GS, six in GBP, and 13 in HF. Six variations of rearranged *IGHJ6* were also identified one and two-times. These appear distributed over all four breeds: IGHV3-IGHD3 (s, BTA7)-IGHJ6 [22], IGHV1S28-IGHD4 (s, BTA8)-IGHJ6 [22], IGHV1S39-IGHD3 (s, BTA7)-IGHJ6 [22], IGHV1S40-IGHDS10-IGHJ6 [22], IGHV3-IGHD8 (as, [21])-IGHJ6 [22], as well as IGHV10-IGHD5 (as, [21])-IGHJ6 [22].

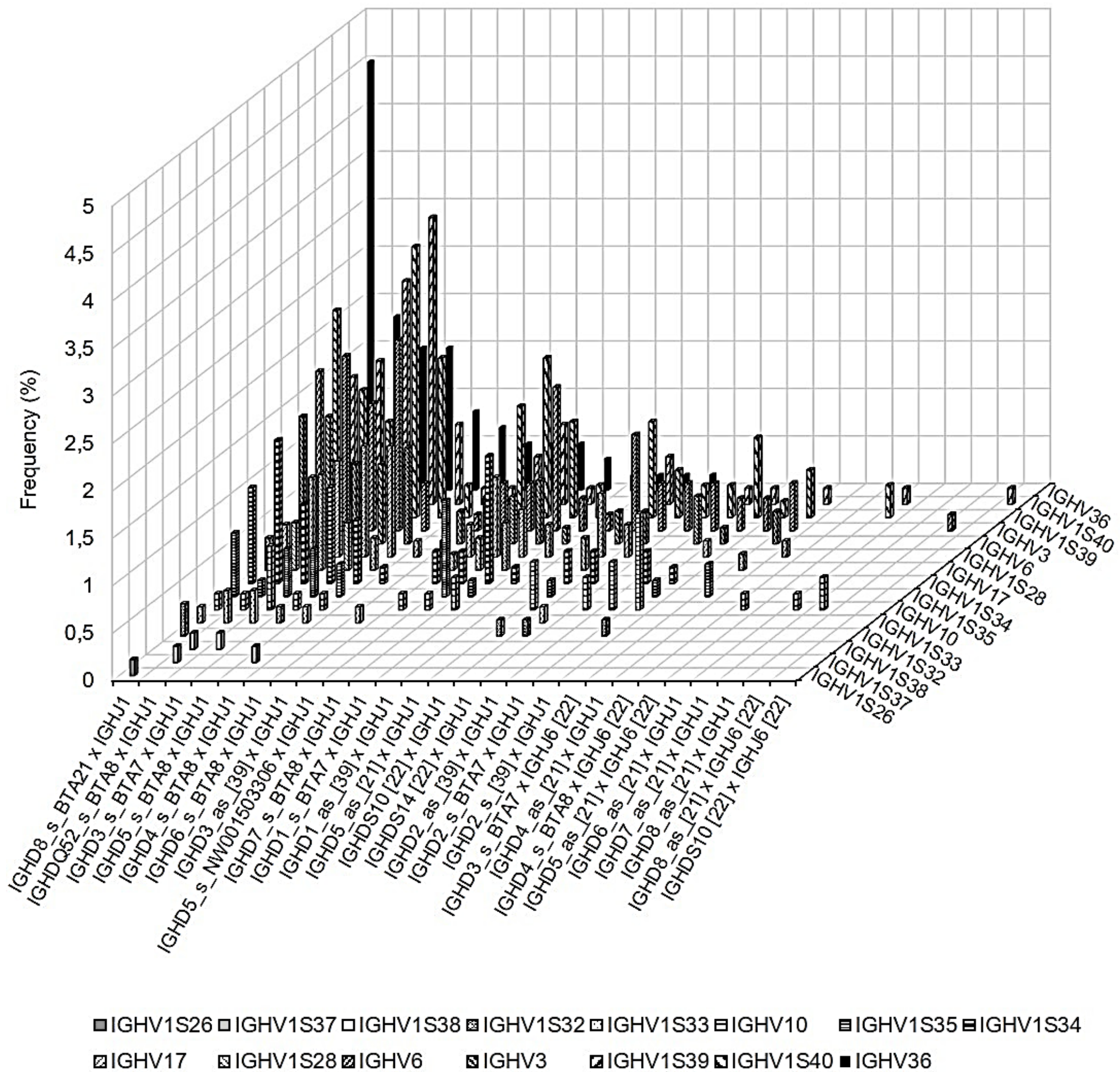


Fig 6. Recombinations of *IGHV*, *IGHD*, and *IGHJ* over all four cattle breeds. In the sequences of all four cattle breeds analyzed 162 different combinations of *IGHV*, *IGHD*, and *IGHJ* were identified. Relative frequencies (%) of the combinations of the 21 transcribed *IGHD* and the 2 transcribed *IGHJ* are shown depending on the rearranged *IGHV* (n=15).

Fifteen out of the 162 IGHV-IGHD-IGHJ combinations were identified in all four cattle breeds investigated. In animals of A, 91 different rearrangements were found, whereas in GBP 74 different recombinations were observed. German Simmental revealed 72 combinations and HF showed 85 variations. Within the cattle breed A (167 rearrangements), the combinations IGHV1S39-IGHD5 (s, BTA8)-IGHJ1 (AY158087), IGHV1S39-IGHD8 (s, BTA21)-IGHJ1 (AY158087), and IGHV1S39-IGHD4 (s, BTA8)-IGHJ1 (AY158087) were most frequently used in 7.78%, 4.79%, and 3.59%,

respectively. Chi square calculations revealed high significant differences in usage frequencies of the recombinations in the breed A (P=0.0004). In animals of HF (159 rearranged sample sequences), the dominant rearrangements were IGHV3-IGHD8 (s, BTA21)-IGHJ1 (AY158087) (4.40%), IGHV3-IGHDQ52 (s, BTA8)-IGHJ1 (AY158087) (3.77%), IGHV36-IGHDQ3 (s, BTA7)-IGHJ1 (AY158087) (3.77%), IGHV3-IGHD5 (s, BTA8)-IGHJ1 (AY158087) (3.14%), and IGHV3-IGHD1 (s, BTA7)-IGHJ1 (AY158087) (3.14%). The Chi square test did not show significant usage differences (P=0.6689) of the different rearranged gene segments in this breed. Very high significant differences were calculated for the recombined IGHV-IGHD-IGHJ in the breeds GS (P<0.0001), while in GBP no significant differences were found (P=0.8870). In GS animals rearranged IGHV36-IGHD8 (s, BTA21)-IGHJ1 (AY158087) (10.69%), IGHV1S40-IGHD3 (s, BTA7)-IGHJ1 (AY158087) (4.58%), IGHV1S40-IGHD5 (s, BTA8)-IGHJ1 (AY158087) (4.58%), IGHV1S34-IGHD8 (s, BTA21)-IGHJ1 (AY158087) (3.82%), and IGHV1S35-IGHD7 (s, BTA8)-IGHJ1 (AY158087) (3.82%) were observed in at least five sequences (131 rearranged samples). Four rearrangements were preferred in GBP (140 rearranged sequences): IGHV36-IGHD8 (s, BTA21)-IGHJ1 (AY158087) (5.00%), IGHV1S40-IGHD5 (s, BTA8)-IGHJ1 (AY158087) (3.57%), and IGHV36-IGHDQ52 (s, BTA8)-IGHJ1 (AY158087) (3.57%), and IGHV1S34-IGHD1 (s, BTA7)-IGHJ1 (AY158087) (3.57%).

Recombination of IGHV, IGHD, and IGHJ in different cattle breeds using procedure 1

With the alignment conditions applied in procedure 1 and 2, different predominantly used recombinations were observed as expected due to the different identified *IGHD*s. Using the default values of MUSCLE (procedure 1), 147 different combinations of *IGHV*, *IGHD*, and *IGHJ* were found (S3 Fig, S4 Table). Most combinations occurred in less than ten sequences. We found very high significant differences in the usage frequencies of the rearranged gene segments between the breeds investigated (P<0.0001). We also observed rearrangements observed in a quantity of ten to 21 sequences that also occurred in sequences spanning of all four breeds. The major rearrangement was IGHV36-IGHD8 (s, BTA21)-IGHJ1 (AY158087) (3.52%). Beside this preferred combination, 49 were identified solely in one sequence, whereby 12 were identified in A, 15 in GS, nine in GBP, and 13 in HF. We also identified seven variations rearranging *IGHJ6* over all animals.

Twenty-one out of the 147 IGHV-IGHD-IGHJ combinations were identified in all four cattle breeds investigated. In animals of A, 81 different rearrangements were found, whereas in GBP 74 different recombinations were observed. German Simmental had 74 combinations and HF had 80 variations. Within the cattle breed A (167 rearranged sample sequences), the combination IGHV1S39-IGHD3 (s. BTA7)-IGHJ1 (AY158087) was the most frequently used in 5.99%. Chi square calculations revealed significant differences in usage frequencies of the recombinations in the breed A (P=0.0108). In animals of HF (159 rearranged sample sequences), the dominant rearrangement was IGHV1S40-IGHD5 (s, BTA8)-IGHJ1 (AY158087) (4.40%). The Chi square test did not show significant usage differences (P=0.0546) for the different rearranged gene segments in this breed. High significant differences were seen for the recombined IGHV-IGHD-IGHJ in the breeds GS (P=0.0002), while in GBP no significant difference was found (P=0.9585). In GS animals rearranged IGHV36-IGHD8 (s, BTA21)-IGHJ1 (AY158087) (9.16%) was observed in at least six sequences (131 rearranged samples). Four rearrangements were preferred in GBP (140 rearranged sequences): IGHV36-IGHD8 (s, BTA21)-IGHJ1 (AY158087) (3.57%), IGHV1S40-IGHD8 (s, BTA8)-IGHJ1 (AY158087) (3.57%), IGHV1S34-IGHD5 (s, BTA8)-IGHJ1 (AY158087) (3.57%), and IGHV36-IGHD5 (s, NW_001503306)-IGHJ1 (AY158087) (3.57%).

Recombination of IGHV, IGHD, and IGHJ in different cattle breeds using procedure 2

When applying procedure 2 with changed values for gap opening and gap extension, we identified 119 different combinations of *IGHV*, *IGHD*, and *IGHJ* (S4 Fig, S5 Table) were identified. As seen for the other procedures, most combinations occurred in less than ten sequences. Calculations revealed very high significant differences in usage frequencies of the rearranged gene segments between the breeds investigated (P<0.0001). Fifteen different rearrangements were observed in a quantity of ten to 42 sequences and were observed across all four breeds. The major rearrangement was IGHV1S39-IGHDQ52 (s, BTA8)-IGHJ1 (AY158087) (7.04%). Beside the dominant combinations, 43 were identified solely in one sequence, whereby six were identified in A, 13 in GS, 12 in GBP, and 12 in HF. Again, seven variations rearranging *IGHJ6* were identified.

Fifteen out of the 119 IGHV-IGHD-IGHJ combinations were identified in all four cattle breeds investigated. For A we found 58 different rearrangements, whereas in GBP 65

different recombinations were observed. German Simmental had 57 combinations and HF had 62 variations. Within A (167 rearrangements), the combination IGHV1S39-IGHDQ52 (s, BTA8)-IGHJ1 (AY158087) was the most frequently used and was seen in 11.38% of the time. Chi square calculations showed significant differences in the usage frequencies of the recombinations for breed A (P=0.0108). In animals of HF (159 rearranged sample sequences), the dominant rearrangement was IGHV1S39-IGHDQ52 (s, BTA8)-IGHJ1 (AY158087) (8.81%). The Chi square test did not show significant usage differences (P=0.0546) for the different rearranged gene segments in this breed. High significant differences were seen for the recombined IGHV-IGHD-IGHJ in the breed GS (P=0.0002), while in GBP no significant differences were observed (P=0.9585). In GS animals rearranged IGHV36-IGHDQ52 (s, BTA8)-IGHJ1 (AY158087) (11.45%) was observed predominantly (131 rearranged sequences). The rearrangement IGHV1S40-IGHDQ52 (s, BTA8)-IGHJ1 (AY158087) (9.29%) was preferred in GBP (140 rearranged sequences).

Recombination of IGHV, IGHD, and IGHJ with different length of CDR3H

Within the three groups of length of CDR3H we identified different preferably expressed recombinations of *IGHV*, *IGHD*, and *IGHJ*. Very high significant differences regarding identified rearrangements within these groups were calculated (procedure 1: among the breeds, GS, A, HF: P<0.0001; GBP: P=0.0016; procedure 2: among the breeds, GS, A, HF, GBP: P<0.0001; procedure 3: among the breeds, A, GBP: P<0.0001; GS: P=0.0003; HF: P=0.0085).

Using changed penalties for gap opening and gap extension and a new scoring matrix (procedure 3) rearrangements of *IGHV3* and *IGHJ1* (AY158087) together with *IGHD1* (s, BTA7), *IGHD2* (as, [39]), *IGHDQ52* (s, BTA8), and *IGHDS10* [22] dominated (0.5-0.84%) in sequences with group 1 CDR3H if breed was not taken into account. The single breeds showed different major recombinations. In A, IGHV6-IGHD2 (s, BTA7)-IGHJ1 was calculated for 1.8% of the sequences, whereas in GS IGHV3-IGHDQ52 (s, BTA8)-IGHJ1 was identified in the same frequencies as IGHV3-IGHD2 (as, [39])-IGHJ1, IGHV3-IGHD1 (as, [39])-IGHJ1, and IGHV1S35-IGHD8 (s, BTA21)-IGHJ1 (0.76%), in GBP IGHV3-IGHDS10 [22]-IGHJ1, IGHV3-IGHDS14 [22]-IGHJ1, IGHV3-IGHD3 (s, BTA7)-IGHJ1, IGHV3-IGHD1 (s, BTA7)-IGHJ1, IGHV1S34-IGHDS10 [22]-IGHJ1, IGHV1S34-IGHDS14 [22]-IGHJ1 and IGHV6-IGHDS10 [22]-IGHJ1 made up

0.71%, and in HF IGHV3-IGHDQ52 (s, BTA8)-IGHJ1 was found the most often (2.52%).

In samples of all breeds possessing CDR3Hs of group 2 the rearrangement of IGHV36-IGHD8 (s, BTA21)-IGHJ1 was found most often (4.52%). This is congruent with our findings from the cattle breeds GS and GBP (10.69%, 5.00%). In A sequences with rearrangement of IGHV1S39-IGHD5 (s, BTA8)-IGHJ1 were used with the highest frequency (7.19%) whereas in HF IGHV3-IGHD8 (s, BTA21)-IGHJ1 and IGHV36-IGHD3 (s, BTA7)-IGHJ1 dominated (3.77%).

The recombination of IGHV10-IGHD2 (s, [39])-IGHJ1 was identified in 0.84% of all sequences with group 3 CDR3Hs. This combination was also dominant in sequences possessing an exceptionally long CDR3H in HF (1.26%). In A, the combination IGHV10-IGHD2 (s, BTA7)-IGHJ1 was preferred (1.2%), whereas in GS, IGHV10-IGHDS10 [22]-IGHJ1 was found the most often (2.29%). In GBP IGHV10-IGHD8 (as, [21])-IGHJ1 and IGHV10-IGHD7 (as, [21])-IGHJ1 were identified (0.71%).

If only new values for gap opening and gap extension were applied (procedure 2), rearrangements of *IGHV3* and *IGHJ1* (AY158087) together with *IGHD1* (as, [39]) dominated (1.01%) in sequences with a very short CDR3H if the breed was not taken into account. The single breeds showed different major recombinations. In A, IGHV6-IGHD1 (s, BTA7)-IGHJ1 was calculated for 2.4% of the sequences, whereas in GS IGHV3-IGHDQ52 (s, BTA8)-IGHJ1 was identified in the same frequencies as IGHV3-IGHD4 (as, [21])-IGHJ1, IGHV3-IGHD1 (s, BTA7)-IGHJ1, and IGHV1S35-IGHD1 (s, BTA7)-IGHJ1 (0.76%), in GBP IGHV3-IGHD1 (as, [39])-IGHJ1 made up 1.43%, whereas in HF IGHV3-IGHD1 (as, [39])-IGHJ1 and IGHV3-IGHD4 (s, BTA8)-IGHJ1were found the most often (1.26%).

In samples of all breeds possessing group 2 CDR3Hs, the rearrangement of IGHV1S39-IGHDQ52 (s, BTA8)-IGHJ1 was found the most often (7.04%). This is congruent with the findings in the cattle breeds A and HF (11.38%, 8.81%). In GS sequences showing the rearrangement IGHV36-IGHDQ52 (s, BTA8)-IGHJ1 were used with the highest frequency (11.45%) whereas in GBP IGHV1S40-IGHDQ52 (s, BTA8)-IGHJ1 dominated (9.29%).

The recombination of IGHV10-IGHD8 (s, BTA21)-IGHJ1 was identified in 1.17% of all sequences with group 3 CDR3Hs. This combination was also dominant in sequences possessing exceptionally long CDR3H in A (2.4%) and GBP (0.71%). In GBP IGHV10-

IGHDS10-IGHJ1 was also found at this frequency. In GS, the combinations IGHV10-IGHD1 (s, BTA7)-IGHJ1 and IGHV10-IGHD4 (s, BTA8)-IGHJ1 were preferred (2.29%), whereas in HF, IGHV10-IGHD3 (s, BTA7)-IGHJ1 was found most often (1.26%).

Using default values (procedure 1) rearrangements of *IGHV3* and *IGHJ1* (AY158087) together with *IGHDQ52* (s, BTA8) as well as IGHV6-IGHD1 (s, BTA7) dominated (1.51 and 0.5%) in sequences with a short CDR3H if the breed was not taken into account. But again the single breeds showed different major recombinations. In A, IGHV3-IGHDQ52 (s, BTA8)-IGHJ1 and IGHV6-IGHD1 (s, BTA7)-IGHJ1 were found for 1.8% of the sequences, whereas in GS IGHV1S35-IGHD4 (s, BTA8)-IGHJ1 was identified at the same frequency as IGHV3-IGHD2 (s, BTA8)-IGHJ1, IGHV3-IGHD1 (as, [39])-IGHJ1, and IGHV3-IGHD8 (as, [21])-IGHJ1 (0.76%), in GBP IGHV3-IGHD1 (s, BTA7)-IGHJ1, IGHV3-IGHD2 (as, [39])-IGHJ1, IGHV3-IGHD5 (s, BTA8)-IGHJ1, IGHV3-IGHDQ52 (s, BTA8)-IGHJ1, IGHV1S34-IGHD5 (s, BTA8)-IGHJ1, IGHV1S34-IGHD5 (s, NW_001503306.)-IGHJ1 and IGHV6-IGHD5 (s, BTA8)-IGHJ1 made up 0.71%, and in HF IGHV3-IGHDQ52 (s, BTA8)-IGHJ1 was found most often (3.14%).
In samples of all breeds possessing CDR3Hs of group 2 the rearrangement of IGHV36-IGHD8 (s, BTA21)-IGHJ1 was found most often (3.52%). This is congruent with the findings in the cattle breed GS (9.16%). In HF sequences showing the rearrangements IGHV1S39-IGHD8 (s, BTA21)-IGHJ1 and IGHV1S40-IGHD5 (s, BTA8)-IGHJ1 were used with the highest frequency (4.4%) whereas in A IGHV1S39-IGHD3 (s, BTA7)-IGHJ1 dominated (5.99%) as well as IGHV1S40-IGHD8 (s, BTA21)-IGHJ1, IGHV36-IGHD5 (s, NW_001503306)-IGHJ1, and IGHV36-IGHD8 (s, BTA21)-IGHJ1 in the cattle breed GBP (3.57%).
The recombination of IGHV10-IGHD8 (s, BTA21)-IGHJ1 was identified in 1.01% of all sequences with exceptionally long CDR3Hs. In sequences possessing those group 3 CDR3H in HF the combinations IGHV10-IGHD4 (s, BTA8)-IGHJ1 and IGHV10-IGHD8 (s, BTA21)-IGHJ1 (1.26%) dominated. In A, the combination IGHV10-IGHD4 (s, BTA8)-IGHJ1 was preferred (1.2%), whereas in GS, IGHV10-IGHD2 (s, BTA7)-IGHJ1 and IGHV10-IGHD8 (s, BTA21)-IGHJ1 were found most often (1.53%), and in GBP only the two recombinations IGHV10-IGHD8 (s, BTA21)-IGHJ1 and IGHV10-IGHD4 (as, [21])-IGHJ1 were identified (0.71%).

Variability based on amino acid substitutions

We then counted the amino acid substitutions at each position to calculate variability as described by Wu and Kabat [40]. The results are shown in the variability plots for each breed separately in Figs 8 a-d. The amino acid positions were numbered in accordance to the IMGT numbering systems [41]. Therefore, FRHs and CDRHs are defined by the following amino acid positions: FR1H: 1…26, CDR1H: 27…38, FR2H: 39…55, CDR2H: 56…65, FR3H: 66…104, CDR3H: 105…117, FR4H: 118…128. In cattle there are no amino acids assigned to positions 10, 31-34, 60-62, and 73.

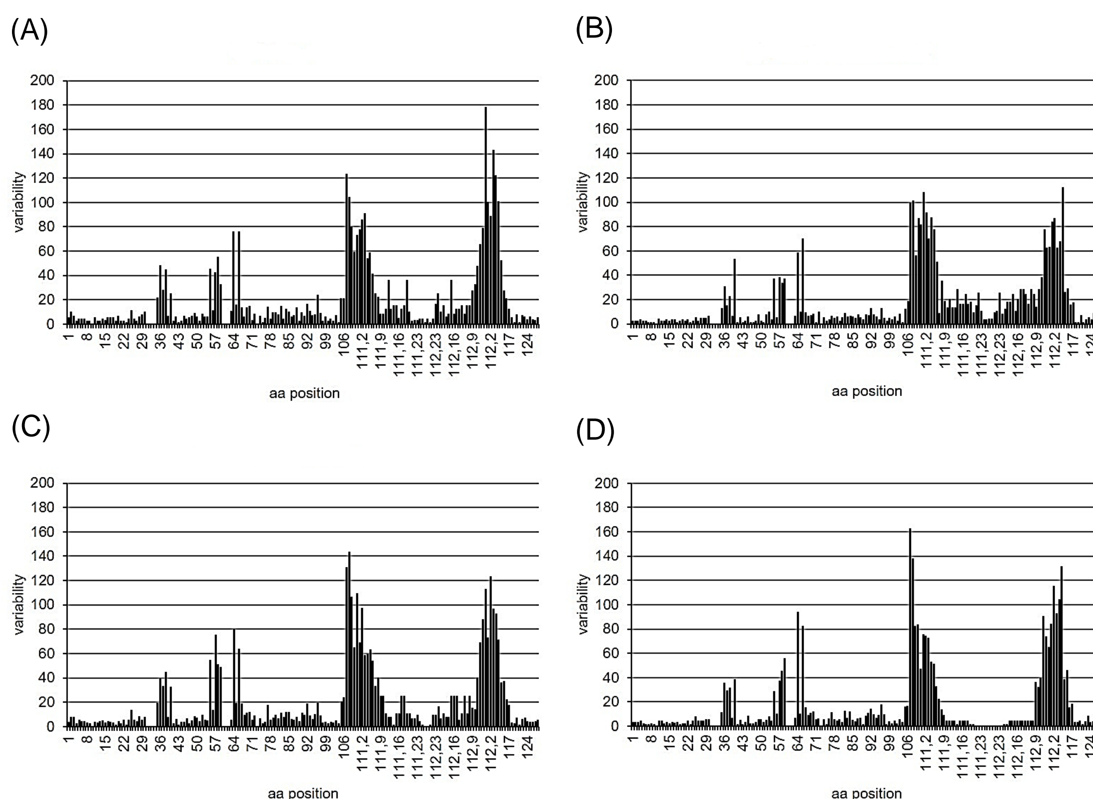


Fig 7. Variability plots of the heavy chain variable regions in the cattle breeds Aubrac, German Simmental, Holstein Friesian, and German Black Pied. The complete variable region is displayed on the horizontal axis. Positions are numbered in accordance to the IMGT numbering system. Within the FR1-4H, little variability is discernable whereas in CDR1-3H increase in variability is observed. (A) Aubrac, (B) German Simmental, (C) Holstein Friesian, (D) German Black Pied

Within the FR1-4H little variability is discernable whereas in CDR1-3H an increase in variability is observed as expected. A maximum of variability in FRHs is calculated for position 96 in FR3H. Variability at this position is 23.7 in A, in HF 19.0, whereas in GS and GBP variability is 12.7 and 17.5, respectively. In CDRHs, variability increases from CDR1H to CDR3H in all breeds. In CDR1H breed A showed the highest variability. In CDR2H variability varies between 58 in GS and 93 in GBP. GBP also showed highest variability in CDR3H. The lowest variability within CDR3H was found in GS. Both HF and A showed moderate variability of CDR2H and CDR3H in comparison to GBP and GS. Overall, variability was much higher at the transition areas between FRH and CDR than in the middle of the CDRHs.

Discussion

This study makes a substantial contribution to the analysis and understanding of the development of the transcribed bovine immunoglobulin repertoire. We examined possible gene conversions within the variable region of bovine heavy chains. We investigated the dominantly transcribed IGHV, IGHD, and IGHJ gene segments and their combinatorial diversity using a newly developed bioinformatics framework, which considers the unique specificity of exceptionally long CDR3 group of bovine immunoglobulin heavy chains. During the development of the program, we applied different conditions (procedure 1 to 3) to improve the alignments of the single gene segments. The progress of assigning germline IGHV, IGHD, and IGHJ to sample sequences is also shown in this study. Unlike previous studies, we complemented our investigation with the analysis of breed specific differences in the four different cattle breeds Aubrac, German Simmental, German Black Pied, and Holstein Friesian.

Due to the limited germline sequence divergence recent studies on bovine immunoglobulin genetics focused on antibody diversification strategies and the junctional diversity of the antibody repertoire. Specific diversification strategies were identified such as the generation of exceptionally long CDR3H [17, 42, 43], the insertion of conserved short nucleotide sequences (CSNS) at the IGHV-IGHD junction [24], the use of pseudogene fragments in lambda light chains as well as gene conversions [14], and somatic hypermutations independent of exposure to external antigens during B-cell development [44]. In addition, more germline gene segments were determined over the last few years [17, 18, 22]. Since the current bovine genome assembly is still incomplete, the full germline repertoire remains under active investigation.

Previous analyses of the bovine immunoglobulin repertoire aimed at identifying rearranged germline gene segments applied various software tools for sequence alignments. As these tools are of limited use for detailed analysis of rearranged immunoglobulin genes due to the difficult and error prone manual assembly of different genes, specialized software tools have been developed. The most familiar ones are IMGT/Junction Analysis [26], IMGT/V-QUEST [27, 28], IMTG/HIGHV-QUEST [29], VBASE2 [30], JoinSolver [31], iHMMun-align [32], and IgBLAST [33]. Only IgBLAST enables the analysis of both nucleotide and protein sequences for FR/CDR and allows the user to either apply the numbering system of Kabat or the IMGT system [41, 45]. Matching germline *IGHV*, *IGHD*, and *IGHJ* genes as well as details at rearrangement

junctions may be analyzed. Searches against germline gene databases and other databases are possible [33]. Tools other than IgBLAST do not provide simultaneous database searches or the analysis of protein sequences. All of these immunoglobulin sequence analysis tools support organisms such as human, mouse, rat, rabbit and rhesus monkey, but representation of livestock is missing or incomplete. We did not apply these analysis tools, as they do not consider the bovine specific occurrence of exceptionally long CDR3H. Therefore, we developed a new bioinformatics framework to address this specific case.

In contrast to the analysis tools mentioned above, our program not only searches our updated bovine specific immunoglobulin germline gene database but also is able to load other pre-designed databases. Matches are identified on the basis of nucleotides as it is the case for all other tools. For the delineation of FR and CDR, we apply the IMGT nomenclature that is currently recommended and most widely used. In addition, we focused on the adjustment of search parameters for *IGHV* and *IGHJ* and especially for the identification of *IGHD*.

The analysis of Ig heavy chain variable regions in four cattle breeds revealed the usage of 15 different *IGHV* segments, 21 *IGHD* segments, as well as two *IGHJ* segments. *IGHV1S39* was used most frequently followed by *IGHV3/33*. Rarely used IGHV segments were *IGHV1S26*, *IGHV1S32*, *IGHV1S33*, *IGHV1S37*, and *IGHV1S38*. In bovine fetal bone marrow, ileum, and spleen high frequencies of *IGHV3/33* (=*IGHV1S3*) and *IGHV1S39* was observed as well as low frequencies of *IGHV1S38* and *IGHV1S26* [22]. The transcription of *IGHV1S32* and *S37* has not yet been described. Among the 20, 17, and 21 transcribed *IGHD* (regarding procedures 1-3), *IGHDS8*, *IGHDS5*, *IGHDS10*, and *IGHDQ52* (=*IGHDS9*) were preferred in all breeds. *IGHDS1* to *IGHDS8* were also found to be transcribed in antisense direction in the third calculation procedure but in low numbers. Using the first procedure, *IGHDS6* was not identified in antisense direction, and applying the second procedure *IGHDS1*, *IGHDS4*, *IGHDS6*, and *IGHDS8* were shown to be transcribed in antisense orientation. Previous studies also elucidated the transcription of 14 *IGHD*, where the occurrence of *IGHDS5* was the most frequent one and was present in 42% of the sequences analyzed in bovine fetus [22]. The assignment of the FR4H to germline *IGHJ* revealed the transcription of *IGHJ1*, and *IGHJ6* [22] with *IGHJ1* clearly preferred. In the cattle breed A, procedure 3 confirmed IGHV1S39-IGHD5-IGHJ1 as the most common recombination of gene segments which is identical to the most frequent finding in

bovine fetus [22]. This recombination belongs to immunoglobulins possessing a CDR3H region of intermediate length. Statistical analyses showed significant different transcription levels of *IGHV*, *IGHD*, and *IGHJ* segments within the breeds.

The usage of pseudogene segments has already been described for animals such as chicken [4, 37]. In bovine lambda light chains, fragments of pseudogenes were also shown to contribute to immunoglobulin diversity in a gene conversion process [14]. In the current analysis, possible gene conversion events were identified by the assignments of parental germline *IGHV* to separate FR1-3H and CDR1-3H. In addition to the *IGHV* identified for the complete variable region based exclusively on FR1-3H, several pseudogenes were assigned as possible originating germline *IGHV* in the separation analysis. For instance, the pseudogenes *IGHV4Ψ*, *IGHV9Ψ*, and *IGHV18Ψ* belong to the boVH2 family [17], but seem to contribute to gene conversion events by nucleotide substitutions. In particular, *IGHV4Ψ*, which was mentioned in the example above meets the criteria for gene conversion such as the location upstream of the rearranged segment [4, 13] and clusters of nucleotide changes [14]. Further, the flanking homology of the conversion region supports the genetic exchange [13] and the separation from *IGHV6* by more than 18 kb on the genome allows looping during rearrangement [17]. In comparison, in chicken the nearest pseudogene is separated by 7 kb [4]. It should be noted that it is difficult to consider the order of gene segments to evaluate the plausibility of other gene conversions due to the incomplete annotation of the bovine genome [17, 18]. Finally, our data indicate an exchange between the two bovine VH families which obviously is rare and which might be an influence of breed or method of analysis when compared to previous results [18].

The length distribution of CDR3H consists of short CDR3H (group 1), intermediate length CDR3H (group 2), and exceptionally long CDR3H (group 3) in all four cattle breeds. In the breed GS the highest percentage of group 3 CDR3H was calculated. The longest CDR3H with 65 aa were found in GS and HF sequences. The longest ever detected CDR3H in cattle was 67 aa long using IMGT numbering [17]. In contrast, four amino acids made up the shortest CDR3H in A animals, GS animals, and HF animals. The maximum length of group 2 CDR3H was 22 aa.

The combinatorial diversity of germline *IGHV*, *IGHD*, and *IGHJ*-segments is represented by 162 different rearrangements that were expressed with significant differences (procedure 3). In comparison, 147 and 119 different recombinations of IGHV-IGHD-IGHJ were identified using calculation procedures 1 and 2, respectively. In

the four breeds, different rearrangements were found. In detail, 91 different combinations occurred in A (procedure 1: 81, and 2: 58), 74 in animals of GBP (procedure 1: 74, and 2: 65), 72 in GS animals (procedure 1: 74, and 2: 57), and 85 in the breed HF (procedure 1: 80, and 2: 62). Most of these combinations were observed in less than ten sequences but seven occurred in up to 21 sequences in all four cattle breeds examined (procedure 1: up to 21, and 2: up to 42).

In sequences belonging to group 1 CDR3H, combinations of *IGHV3* and *IGHJ1* (AY158087) together with *IGHDQ52* (s, BTA8) dominated over all breeds using procedure 3. As *IGHDQ52* is the shortest *IGHD* segment possessing only four amino acids, these results explain best the origin of short CDR3H. Group 3 CDR3H mostly exhibited IGHV10-IGHD2 (s, [39])-IGHJ1 (procedure 3), or IGHV10-IGHD8 (s, BTA21)-IGHJ1 (procedure 1 and 2) in all breeds. Only results from procedure 3 identified biological meaningful combinations of germline IGHV, IGHD, and IGHJ as it gave the best explanation for the origin of group 3 CDR3H. IGHD2 is the longest IGHD segment identified so far. Further, *IGHV10*, which is identical to *IGHV1S1* and *IGHV1S15*, was found to contribute solely to those exceptional lengths [17, 18, 22]. It is assumed, that the “ThrThrValHisGln” terminal motif of IGHV10, that initiates an ascending β strand in the folded antibody enables the formation of the “stalk and knob” structure in addition to inserted conserved short nucleotide sequences (CSNS) [22, 43]. Furthermore, in sequences of group 2 a higher number of recombinations were observed than in sequences of group 1 and 3. As group 3 CDR3H regions are unique in cattle, the few preferably rearranged gene segments within this group may indicate specialized antibodies.

Variability plots indicated quite similar features within the variable region in all breeds. Nevertheless, amino acid residues at each position vary between the samples and the breeds and within the regions. In FR1-4H little variability was calculated, whereas the variability increased from CDR1H to CDR3H, which was described already as concentrated areas of diversity in equine heavy and light chain CDRs [46, 47]. The breed A possesses the highest amino acid variations in CDR1H and HF in CDR2H while GS exhibited lowest variability in these two CDRHs. Further, transition areas between FRHs and CDRHs had a higher variability than the middle of CDRHs. Position 96 in FR3H shows the highest variability within the FRHs. This residue is located on the outer surface of the variable region of the immunoglobulin molecule [43] within the area where the constant region is connected to the variable region. The high variability

at this position may indicate an influence on the position of variable and constant region and their sterical orientation, which may affect light chain pairing as heavy chains possessing group 3 CDR3H are connected to a special type of lambda light chains [43, 48].

Further analysis revealed that no amino acids were assigned to the IMGT amino acid positions 10, 31-34, 60-62, and 73 in cattle. This means, that one amino acid position within FR1H, four positions in CDR1H, three positions in CDR2H, as well as one position in FR3H were not filled. Consequently, in cattle 8 out of 12 amino acid positions within CDR1H are covered. In CDR2H, 10 positions are available and 7 are covered. Compared to FR-IMGT and CDR-IMGT lengths of functional and ORF *IGHV*-genes of human *IGHV*, mouse *IGHV*, rat *IGHV*, arabian camel *IGHV*, sheep *IGHV*, and pig *IGHV* the missing amino acids within FR1H and FR3H are conserved in all animals mentioned [49, 50]. Averaged eight to ten amino acids were positioned in CDR1H whereas in CDR2H six to ten amino acids were placed by the IMGT numbering system in human, mouse, rat, camel, sheep, and pig *IGHV*. Therefore, the positions of missing amino acids are congruent with other species.

In the breed A, the highest number of recombinations and variability were observed when compared to the other breeds investigated. GS possessed the lowest number of recombinations and showed less variability except in the middle of the CDR3H region. This finding indicates the contribution of insertions and deletions to diversity in case of few rearrangements [25]. It should be noted that A and GS were kept under the same management in a mixed herd. GBP and HF were kept at different farms. The breeds kept in different areas were consequently exposed to different antigens. Thus, the individual number of rearrangements per breed and differences in variability additionally indicate a specialized immune response as animals on one farm are challenged with the same environment.

The application of the newly developed bioinformatics framework led to important new results. Our analyses demonstrated that the bovine heavy chain diversity is not restricted to the use of a limited number of germline genes although there are preferred rearrangements within the three groups of CDR3H lengths. We also found strong evidence for gene conversion using pseudogenes. Despite current advances in the understanding of bovine immunoglobulin diversification, future investigations of the germline repertoire are necessary.

Material and Methods

Detailed analyses of immunoglobulin sequences using a newly developed bioinformatics framework

For sequence analysis, we developed a new bioinformatics framework using MUSCLE [34, 35] for the initial fast but accurate multiple nucleotide sequence alignment and following ClustalW [36] for calculating the sequence distances after deduction of the amino acid residues. Both programs are available as stand-alone algorithms and were implemented into our program. The immunogenetics nomenclature (IMGT) was used to assign framework regions 1-4 (FR1-4H) and the complementarity determining regions 1-3 [41].

Therefore, germline nucleotide sequences were imported in FASTA-format (*IGHV*, *IGHD*, *IGHJ* [17, 18, 22]). Using the functional *IGHV*, the nucleotide sequences were translated into amino acids to number the codons of the functional germline *IGHV* gene segments according to the IMGT system from FR1H to FR3H. This required first the identification of the conserved and preassigned positions of Cys23, Trp41, Leu89, and Cys104 defined by Lefranc et al. [41]. Following, the nucleotide sequences of germline pseudo *IGHV* gene segments were aligned separately using MUSCLE [34, 35] to obtain the putative open reading frame. The previously defined positions of the codons using the functional genes were transferred onto the pseudogenes. Insertions as well as deletions of nucleotides within the pseudo gene segment sequence were discarded. The last 33 nucleotides of germline *IGHJ*s were then used to define the FR4H. The region between FR3H and FR4H is defined as CDR3H. This region was later used to align the sample CDR3H to germline *IGHD* segments. Defining FRHs and CDRHs allowed, beside the analyses of the complete transcribed genes, the alignment of *IGHV*, *IGHD* and *IGHJ* using different parameters to improve the biological significance as well as the analyses of the single functionally divergent regions to determine putative gene conversion events in those regions. All functional and pseudo gene germline segments are referred as reference sequences.

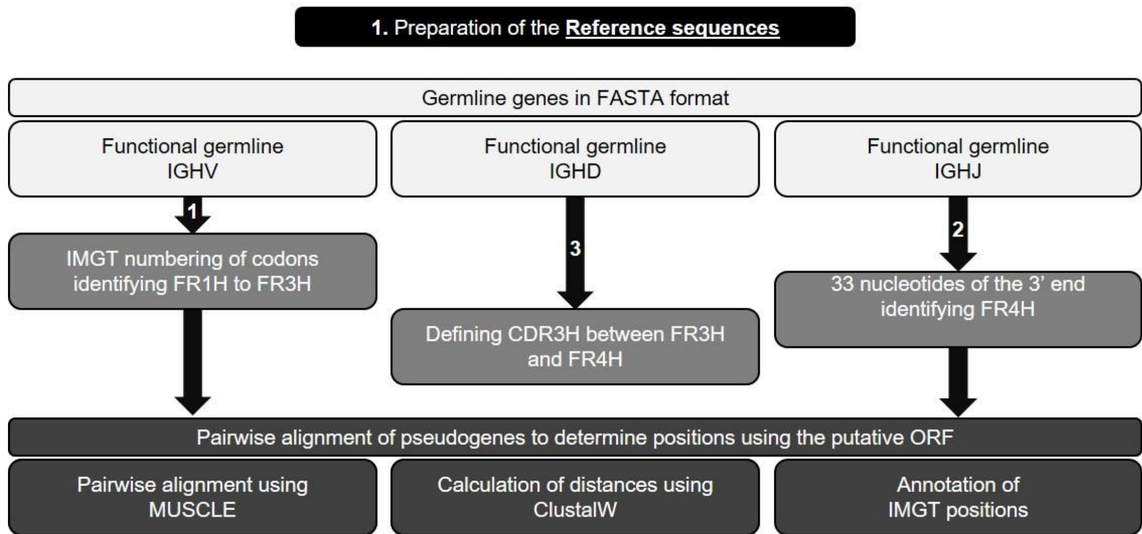
Framework regions and CDRs of transcribed sequences (designated as sample sequences) were aligned pairwise to the isolated reference sequences after isolation from the first IgG constant region. Following, minimal divergence was used to identify the most similar sample sequence and reference sequence pairing.

For nucleotide alignments of *IGHV* and *IGHJ*, default values of MUSCLE were used. We tested three different procedures to assign germline and sample *IGHD*s to improve

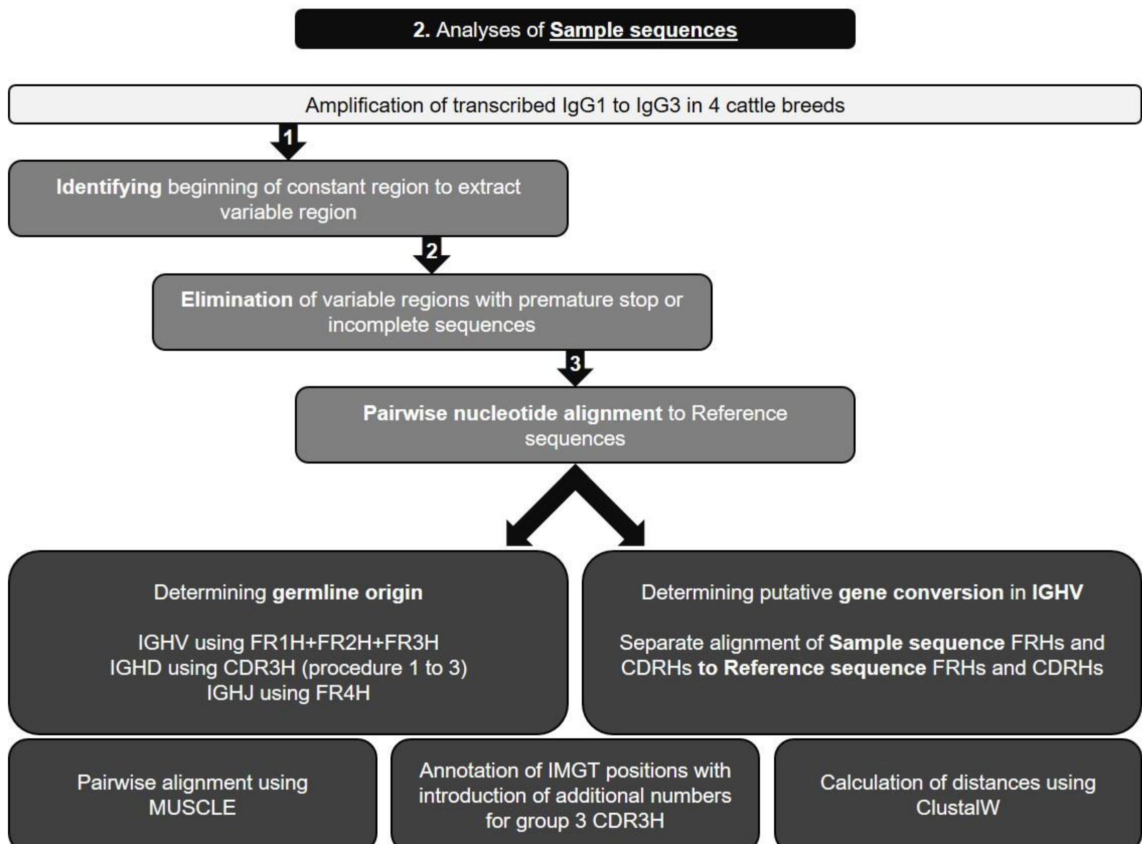
biological significance. In procedure 1, we applied default values of MUSCLE [34, 35], in procedure 2 we changed the penalties for gap opening to -4 and for gap extention to -0.3 [22], and in procedure 3 we additionally incorporated a new scoring matrix with match = 2, transversion = -1, and transition = 1 to evaluate transversion and transition mutations, whereby the IUB (international union of biochemistry) code for single and wobble bases was used.

The three procedures were applied to a set of sample sequences of IgG-derived variable regions from four different cattle breeds. Each nucleotide sequence of our sample sequences was aligned separately to the reference sequences to determine the most similar reference sequence as germline origin. Following, the codons were translated into amino acids. Sample sequences possessing premature Stop codons or not covering the full length of the variable region due to incomplete sequencing were eliminated and were not incorporated in further analyses. The remaining sample sequences were annotated in accordance to the IMGT nomenclature. For exceptional long CDR3H no positions are defined in the IMGT system, therefore positions had to be added as required and designated as 111.1-111.x and 112.y-112.1 in accordance to the IMGT numbering system [41]. To determine the germline origin, only the FRHs were aligned to avoid interference with the highly diversified CDRH [18]. To analyze possible gene conversion events, FR1-3Hs and CDR1-3Hs were extracted and aligned separately to the corresponding regions of the IGHV reference sequences to find the most similar one (Fig 8). The results were presented as an html table.

(A)



(B)



(C)

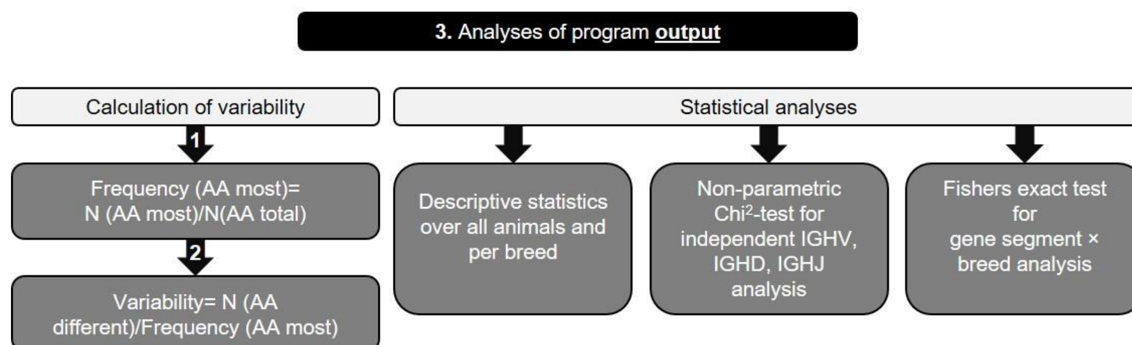


Fig. 8. Graphical presentation of workflow of the developed bioinformatics framework used for analyzing bovine heavy chain IgG. (A) At first the reference sequences were prepared. Using the functional segments the pseudogenes were aligned. (B) The bioinformatics framework analyzed the germline origin of bovine immunoglobulin heavy chain variable segments (IGHV), diversity segments (IGHD), and joining segments (IGHJ). One approach used the framework regions (FRH) 1 to 3 to determine the closest germline *IGHV*. The second approach analyzed single functional regions FRH1 to 3 and complementarity determining regions (CDRH) 1 and 2 independently to reveal putative gene conversion events. (C) The last step included the calculation of the variability and the statistical analyses.

To display accumulation of amino acid substitutions in distinct segments of the variable region, variability was calculated as described by Wu and Kabat [40]. Thus, the frequency of the most common amino acid at a distinct position was calculated first. The number of the most abundant amino acid at a given position was divided by the number of all amino acids observed at this position. This means, only samples possessing an amino acid residue at this position in accordance to the IMGT nomenclature were considered. Subsequently, the number of different amino acids at the given position was divided by the frequency of the most common amino acid residue to determine variability. The variability results were written into a txt-file, which allows further analyses in statistical software.

For the statistical analyses of the distribution of *IGHV*, *IGHD*, and *IGHJ* segments including putative gene conversions within one breed and among breeds have been compared by applying non-parametric tests. Such test procedures, i.e. the Chi^2- test for analyzing *IGHV*, *IGHD* and *IGHJ* independently and Fisher's exact test for the gene segment × breed contingency table, are implemented in the software package SAS, Version 9.2.

Generation of the sample sequence set

Breed selection, isolation of PBMCs and cDNA synthesis

For the analysis of the transcription of IgG heavy chain genes, the four cattle breeds German Black Pied (GBP), German Simmental (GS), Holstein-Friesian (HF), and Aubrac (A) were chosen. The animals selected from the herd of breed A were composed of seven French and three German animals, whereas the sample of breed GS included one Austrian bull. German Black Pied and A represent small populations and have local importance, whereas HF and GS are commonly used in global commercial milk and meat production. German Simmental and A are kept on the same farm under same management conditions and in a mixed herd.

Blood samples were collected from ten randomly chosen animals per breed during routine blood sampling for mandatory examinations in disease control. Peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll gradients (GE Healthcare Europe GmbH, Germany) according to the manufacturer's protocol. Cells were stained with trypan blue and viable cells were counted. Total RNA was isolated from $1x10^7$ cells using the RNeasy® Mini Kit (Qiagen, Germany). The first-strand cDNA was synthesized using $pd(N)_6$ primers from 3 µg of total RNA in a total volume of 20 µl (SuperScript™III First-Strand Synthesis SuperMix, Life Technologies GmbH, Germany).

Ethical Statement

To collect B-lymphocytes, 20 ml of EDTA blood were taken from the tail vein during regular investigation of infectious diseases in the cattle herds. The plasma was applied e.g. in an indirect ELISA testing for antibodies against Mycobacteria avium spp. paratuberculosis or BHV-1. Therefore, no specific approval is required.

Amplification of immunoglobulin heavy-chain isotype restricted variable regions

To amplify the variable regions restricted to γ1-3 isotype heavy chains, a PCR was performed with primers binding within the leader region and the 3'UTR (bIgG_leader: ATG AAC CCA CTG TGG ACC; bIgG_3'UTR: CAG GAG GAA TGC ACA CAG). The primers were based on database sequence information and assigned to accession number X62916. The primer boIgG_leader anneals to position 22-39, and the primer boIgG_3'UTR to position 1518-1535. To monitor the integrity and purity of the cDNA, 527 bp of the bovine GAPDH (Glycerinaldehyde 3-phosphate dehydrogenase) were

amplified as a positive control. A no- template control served as a negative control for the PCR. The total reaction volume of 50 µl included 0.67 µl of cDNA, 200 µM dNTPs (Bioline, Germany), 5 µl of 10x PCR buffer (75 mM Tris-HCl, pH 9.0; 2 mM $MgCl_2$; 50 mM KCl; 20 mM $(NH_4)_2SO_4$), 0.4 µM of each primer, and 2 units of DNA polymerase (Biotools, Spain). PCR was performed under cycling conditions of 95°C for 5 min, followed by 35 cycles of 95°C for 1 min, 59.4°C for 1 min, 72°C for 2 min, and terminated with elongation at 72°C for 10 min. Length and purity of the PCR products were evaluated by means of electrophoresis on 1% agarose gels.

Cloning and sequencing of the PCR products

The PCR products were purified and concentrated using the MiniElute Gel Extraction Kit (Qiagen, Germany) in accordance to the manufacturer's protocol except QX1 buffer replaced QG buffer. Samples were eluted with 13 µl EB buffer (10 mM Tris-HCl, pH 8.5) and were stored at 4°C. Purified products were cloned into the pCR® 2.1-TOPO® 3.9 kb TA vector (Invitrogen™, Karlsruhe, Germany) and transformed into chemically competent One Shot TOP10 *E. coli* cells (Invitrogen™, Karlsruhe, Germany). Transformants were plated on LB agar containing 0.3 mM ampicillin, 40 µl 2.44 µM X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside), and 40 µl 1 M IPTG (Isopropyl β-D-1-thiogalactopyranoside) for blue-white selection. After incubation at 37°C, overnight cultures of randomly selected white transformants were grown in a 5 ml LB-ampicillin broth. Plasmids were isolated using the MiniPrep Kit (Qiagen, Germany). In order to assess the insert size, plasmid DNA was cleaved with *Eco*RI (New England Biolabs, Germany) or a colony PCR was performed. Therefore a 25 µl mixture containing 2 µl cell culture, 0.4 µM of vector specific primers M13 (-20) Forward and M13 Reverse (Invitrogen, Germany), respectively, and one PCR-bead (GE Healthcare Europe GmbH, Germany) were used in a hot start PCR at 95°C for 5 min, denaturation at 95°C for 1 min, annealing at 60°C for 1 min, and extension at 72°C for 2 min up to a total of 30 cycles. A final extension at 72°C for 10 min was included after the final cycle before PCR mixtures were cooled down to 4°C. The size of the resulting fragments and of the PCR products was confirmed by agarose gel electrophoresis.

Sixteen clones per animal were sequenced according to the chain-termination method [51]. The M13 (-20) Forward (5'-GTA AAA CGA CGG CCA G-3') and M13 Reverse (5'-CAG GAA ACA GCT CTG AC-3', Invitrogen, Germany) vector-specific primers, as well as the gene specific primers boIgG_leader, boIgG_3'UTR, boIgG_CH1_for (5'-GCC

TCC ACC ACA GCC CCG AAA G-3’), boIgG_CH3_rev (5’-GAC CTT GCA CTT GAA CTC C-3’) and boIgG_CH1_rev (5’-ACG GTC ACC ATG CTG CTG AG-3’) were used for sequencing.

Acknowledgments

We gratefully thank Dr. Dalia Aboelhassan, Sina Stein and Caroline Bierschenk for their technical support in generating the sequences. We also thank Dr. Apostolos Gittis and Dr. Stephen Dollery for critical review of the manuscript.

References

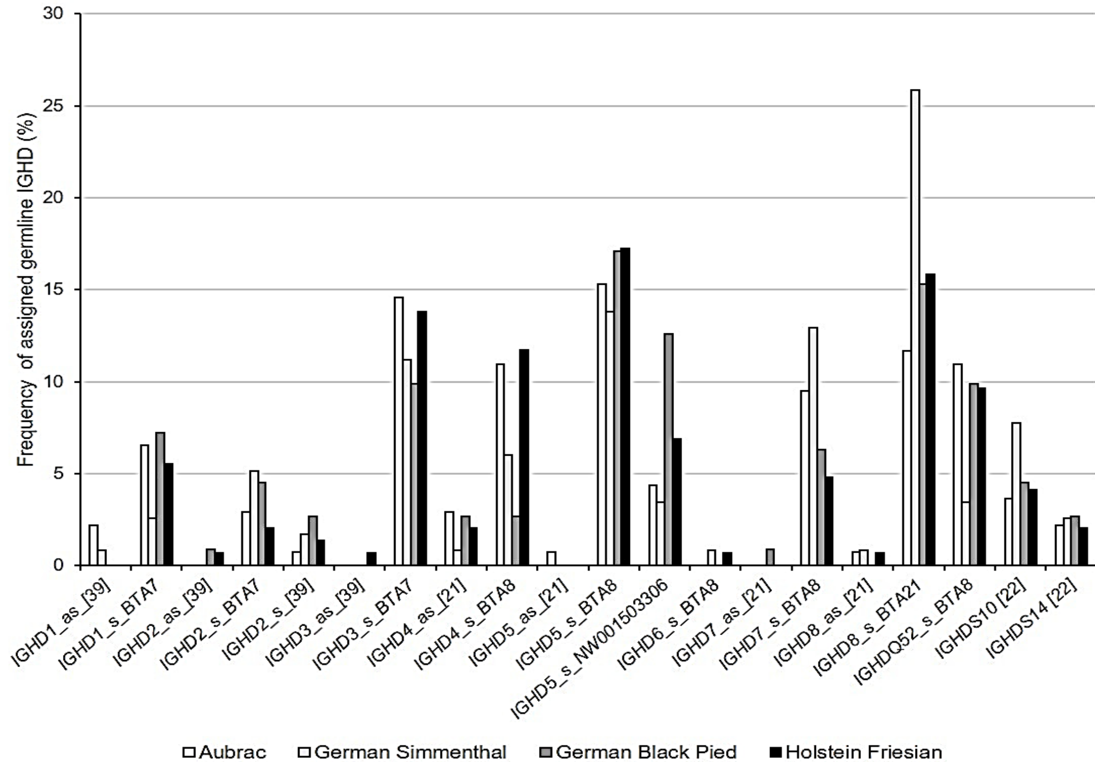
1. Edelman GM. Antibody structure and molecular immunology. Science. 1973;180(88):830-40. Epub 1973/05/25. PubMed PMID: 4540988.
2. Tonegawa S. Somatic generation of antibody diversity. Nature. 1983;302(5909):575-81. Epub 1983/04/14. PubMed PMID: 6300689.
3. Wagner SD, Neuberger MS. Somatic hypermutation of immunoglobulin genes. Annu Rev Immunol. 1996;14:441-57. Epub 1996/01/01. doi: 10.1146/annurev.immunol.14.1.441. PubMed PMID: 8962691.
4. Reynaud CA, Dahan A, Anquez V, Weill JC. Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell. 1989;59(1):171-83. Epub 1989/10/06. doi: 0092-8674(89)90879-9 [pii]. PubMed PMID: 2507167.
5. Berens SJ, Wylie DE, Lopez OJ. Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. Int Immunol. 1997;9(1):189-99. Epub 1997/01/01. PubMed PMID: 9043960.
6. Sun J, Kacskovics I, Brown WR, Butler JE. Expressed swine VH genes belong to a small VH gene family homologous to human VHIII. J Immunol. 1994;153(12):5618-27. Epub 1994/12/15. PubMed PMID: 7989761.
7. Dufour V, Malinge S, Nau F. The sheep Ig variable region repertoire consists of a single VH family. J Immunol. 1996;156(6):2163-70. Epub 1996/03/15. PubMed PMID: 8690905.
8. Saini SS, Hein WR, Kaushik A. A single predominantly expressed polymorphic immunoglobulin VH gene family, related to mammalian group, I, clan, II, is identified in cattle. Mol Immunol. 1997;34(8-9):641-51. Epub 1997/06/01. PubMed PMID: 9393967.
9. Sinclair MC, Gilchrist J, Aitken R. Bovine IgG repertoire is dominated by a single diversified VH gene family. J Immunol. 1997;159(8):3883-9. Epub 1997/10/23. PubMed PMID: 9378976.
10. Reynaud CA, Anquez V, Dahan A, Weill JC. A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. Cell. 1985;40(2):283-91. Epub 1985/02/01. doi: 0092-8674(85)90142-4 [pii]. PubMed PMID: 3917859.
11. Reynaud CA, Anquez V, Grimal H, Weill JC. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell. 1987;48(3):379-88. Epub 1987/02/13. doi: 0092-8674(87)90189-9 [pii]. PubMed PMID: 3100050.
12. Reynaud CA, Garcia C, Hein WR, Weill JC. Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. Cell. 1995;80(1):115-25. Epub 1995/01/13. doi: 0092-8674(95)90456-5 [pii]. PubMed PMID: 7813007.
13. Becker RS, Knight KL. Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. Cell. 1990;63(5):987-97. Epub 1990/11/30. doi: 0092-8674(90)90502-6 [pii]. PubMed PMID: 2124176.
14. Parng CL, Hansal S, Goldsby RA, Osborne BA. Gene conversion contributes to Ig light chain diversity in cattle. J Immunol. 1996;157(12):5478-86. Epub 1996/12/15. PubMed PMID: 8955197.
15. Sun Y, Wang C, Wang Y, Zhang T, Ren L, Hu X, et al. A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. Dev Comp Immunol. 2010;34(9):1009-20. Epub 2010/05/15. doi: S0145-305X(10)00103-5 [pii]
10.1016/j.dci.2010.05.003. PubMed PMID: 20466019.

16. Miller JR, Thomsen PD, Dixon SC, Tucker EM, Konfortov BA, Harbitz I. Synteny mapping of the bovine IGHG2, CRC and IGF1 genes. Anim Genet. 1992;23(1):51-8. Epub 1992/01/01. PubMed PMID: 1349204.
17. Walther S, Czerny CP, Diesterbeck US. Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins. PLoS One. 2013;8(5):e64234. Epub 2013/05/30. doi: 10.1371/journal.pone.0064234
PONE-D-13-04746 [pii]. PubMed PMID: 23717573; PubMed Central PMCID: PMC3661452.
18. Niku M, Liljavirta J, Durkin K, Schroderus E, Iivanainen A. The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed. Dev Comp Immunol. 2012;37(3-4):457-61. doi: 10.1016/j.dci.2012.02.006. PubMed PMID: 22369780.
19. Zhao Y, Kacskovics I, Rabbani H, Hammarstrom L. Physical mapping of the bovine immunoglobulin heavy chain constant region gene locus. J Biol Chem. 2003;278(37):35024-32. Epub 2003/06/28. doi: 10.1074/jbc.M301337200
M301337200 [pii]. PubMed PMID: 12829708.
20. Hosseini A, Campbell G, Prorocic M, Aitken R. Duplicated copies of the bovine JH locus contribute to the Ig repertoire. Int Immunol. 2004;16(6):843-52. Epub 2004/05/06. doi: 10.1093/intimm/dxh085
dxh085 [pii]. PubMed PMID: 15126417.
21. Koti M, Kataeva G, Kaushik AK. Organization of D(H)-gene locus is distinct in cattle. Dev Biol (Basel). 2008;132:307-13. Epub 2008/09/27. PubMed PMID: 18817318.
22. Liljavirta J, Niku M, Pessa-Morikawa T, Ekman A, Iivanainen A. Expansion of the preimmune antibody repertoire by junctional diversity in Bos taurus. PLoS One. 2014;9(6):e99808. doi: 10.1371/journal.pone.0099808. PubMed PMID: 24926997; PubMed Central PMCID: PMC4057420.
23. Wang Y, Jackson KJ, Sewell WA, Collins AM. Many human immunoglobulin heavy-chain IGHV gene polymorphisms have been reported in error. Immunol Cell Biol. 2008;86(2):111-5. doi: 10.1038/sj.icb.7100144. PubMed PMID: 18040280.
24. Koti M, Kataeva G, Kaushik AK. Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies. Molecular immunology. 2010;47(11-12):2119-28. doi: 10.1016/j.molimm.2010.02.014. PubMed PMID: 20435350.
25. Larsen PA, Smith TP. Application of circular consensus sequencing and network analysis to characterize the bovine IgG repertoire. BMC Immunol. 2012;13:52. Epub 2012/09/18. doi: 1471-2172-13-52 [pii]
10.1186/1471-2172-13-52. PubMed PMID: 22978666.
26. Yousfi Monod M, Giudicelli V, Chaume D, Lefranc MP. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. Bioinformatics. 2004;20 Suppl 1:i379-85. doi: 10.1093/bioinformatics/bth945. PubMed PMID: 15262823.
27. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res. 2008;36(Web Server issue):W503-8. Epub 2008/05/27. doi: gkn316 [pii]
10.1093/nar/gkn316. PubMed PMID: 18503082.
28. Giudicelli V, Brochet X, Lefranc MP. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. Cold Spring Harb Protoc. 2011;2011(6):695-715. doi: 10.1101/pdb.prot5633. PubMed PMID: 21632778.

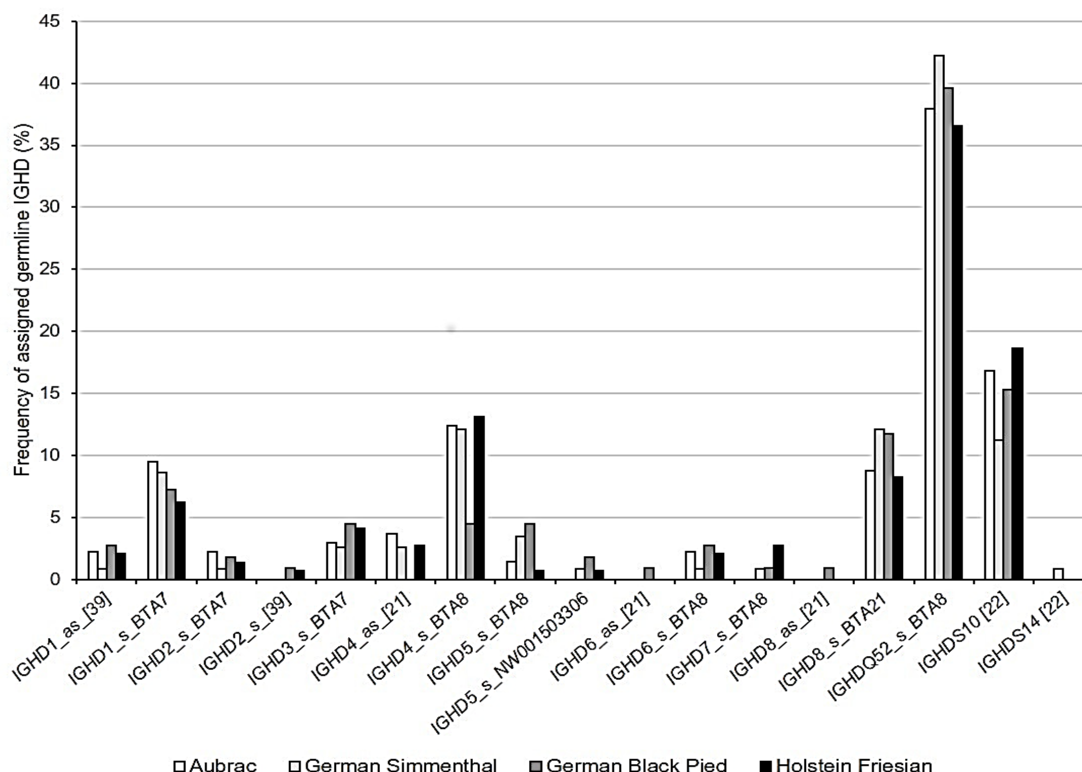
29. Alamyar E, Duroux P, Lefranc MP, Giudicelli V. IMGT((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. Methods Mol Biol. 2012;882:569-604. doi: 10.1007/978-1-61779-842-9_32. PubMed PMID: 22665256.
30. Retter I, Althaus HH, Munch R, Muller W. VBASE2, an integrative V gene database. Nucleic Acids Res. 2005;33(Database issue):D671-4. doi: 10.1093/nar/gki088. PubMed PMID: 15608286; PubMed Central PMCID: PMC540042.
31. Souto-Carneiro MM, Longo NS, Russ DE, Sun HW, Lipsky PE. Characterization of the human Ig heavy chain antigen binding complementarity determining region 3 using a newly developed software algorithm, JOINSOLVER. J Immunol. 2004;172(11):6790-802. PubMed PMID: 15153497.
32. Gaeta BA, Malming HR, Jackson KJ, Bain ME, Wilson P, Collins AM. iHMMune-align: hidden Markov model-based alignment and identification of germline genes in rearranged immunoglobulin gene sequences. Bioinformatics. 2007;23(13):1580-7. doi: 10.1093/bioinformatics/btm147. PubMed PMID: 17463026.
33. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. 2013;41(Web Server issue):W34-40. doi: 10.1093/nar/gkt382. PubMed PMID: 23671333; PubMed Central PMCID: PMC3692102.
34. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113. doi: 10.1186/1471-2105-5-113. PubMed PMID: 15318951; PubMed Central PMCID: PMC517706.
35. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-7. doi: 10.1093/nar/gkh340. PubMed PMID: 15034147; PubMed Central PMCID: PMC390337.
36. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22(22):4673-80. Epub 1994/11/11. PubMed PMID: 7984417.
37. Reynaud CA, Bertocci B, Dahan A, Weill JC. Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. Adv Immunol. 1994;57:353-78. Epub 1994/01/01. PubMed PMID: 7872160.
38. Thompson CB, Neiman PE. Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. Cell. 1987;48(3):369-78. Epub 1987/02/13. doi: 0092-8674(87)90188-7 [pii]. PubMed PMID: 3100049.
39. Shojaei F, Saini SS, Kaushik AK. Unusually long germline DH genes contribute to large sized CDR3H in bovine antibodies. Molecular immunology. 2003;40(1):61-7. PubMed PMID: 12909131.
40. Wu TT, Kabat EA. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J Exp Med. 1970;132(2):211-50. Epub 1970/08/01. PubMed PMID: 5508247.
41. Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol. 2003;27(1):55-77. Epub 2002/12/13. doi: S0145305X02000393 [pii]. PubMed PMID: 12477501.

42. Saini SS, MacGlashan DW, Jr., Sterbinsky SA, Togias A, Adelman DC, Lichtenstein LM, et al. Down-regulation of human basophil IgE and FC epsilon RI alpha surface densities and mediator release by anti-IgE-infusions is reversible in vitro and in vivo. J Immunol. 1999;162(9):5624-30. Epub 1999/05/05. PubMed PMID: 10228046.
43. Wang F, Ekiert DC, Ahmad I, Yu W, Zhang Y, Bazirgan O, et al. Reshaping antibody diversity. Cell. 2013;153(6):1379-93. Epub 2013/06/12. doi: S0092-8674(13)00525-4 [pii]
10.1016/j.cell.2013.04.049. PubMed PMID: 23746848; PubMed Central PMCID: PMC4007204.
44. Lucier MR, Thompson RE, Waire J, Lin AW, Osborne BA, Goldsby RA. Multiple sites of V lambda diversification in cattle. J Immunol. 1998;161(10):5438-44. Epub 1998/11/20. PubMed PMID: 9820519.
45. Kabat EA, Wu TT, Perry HM, Gottesmann KS, Foeller C. Sequences of proteins of immunological interest. 5th ed. Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; 2006.
46. Tallmadge RL, Tseng CT, Felippe MJ. Diversity of immunoglobulin lambda light chain gene usage over developmental stages in the horse. Dev Comp Immunol. 2014;46(2):171-9. Epub 2014/04/15. doi: S0145-305X(14)00097-4 [pii]
10.1016/j.dci.2014.04.001. PubMed PMID: 24726757; PubMed Central PMCID: PMC4107094.
47. Tallmadge RL, Tseng CT, King RA, Felippe MJ. Developmental progression of equine immunoglobulin heavy chain variable region diversity. Dev Comp Immunol. 2013;41(1):33-43. Epub 2013/04/10. doi: S0145-305X(13)00087-6 [pii]
10.1016/j.dci.2013.03.020. PubMed PMID: 23567345; PubMed Central PMCID: PMC3672396.
48. Saini SS, Farrugia W, Ramsland PA, Kaushik AK. Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted VH + Vlambda pairings. Int Immunol. 2003;15(7):845-53. Epub 2003/06/17. PubMed PMID: 12807823.
49. Lefranc MP, Giudicelli V, Ginestoux C, Bodmer J, Muller W, Bontrop R, et al. IMGT, the international ImMunoGeneTics database. Nucleic Acids Res. 1999;27(1):209-12. PubMed PMID: 9847182; PubMed Central PMCID: PMC148137.
50. Scaviner D, Barbie V, Ruiz M, Lefranc MP. Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions. Exp Clin Immunogenet. 1999;16(4):234-40. doi: 19115. PubMed PMID: 10575277.
51. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-7. Epub 1977/12/01. PubMed PMID: 271968.

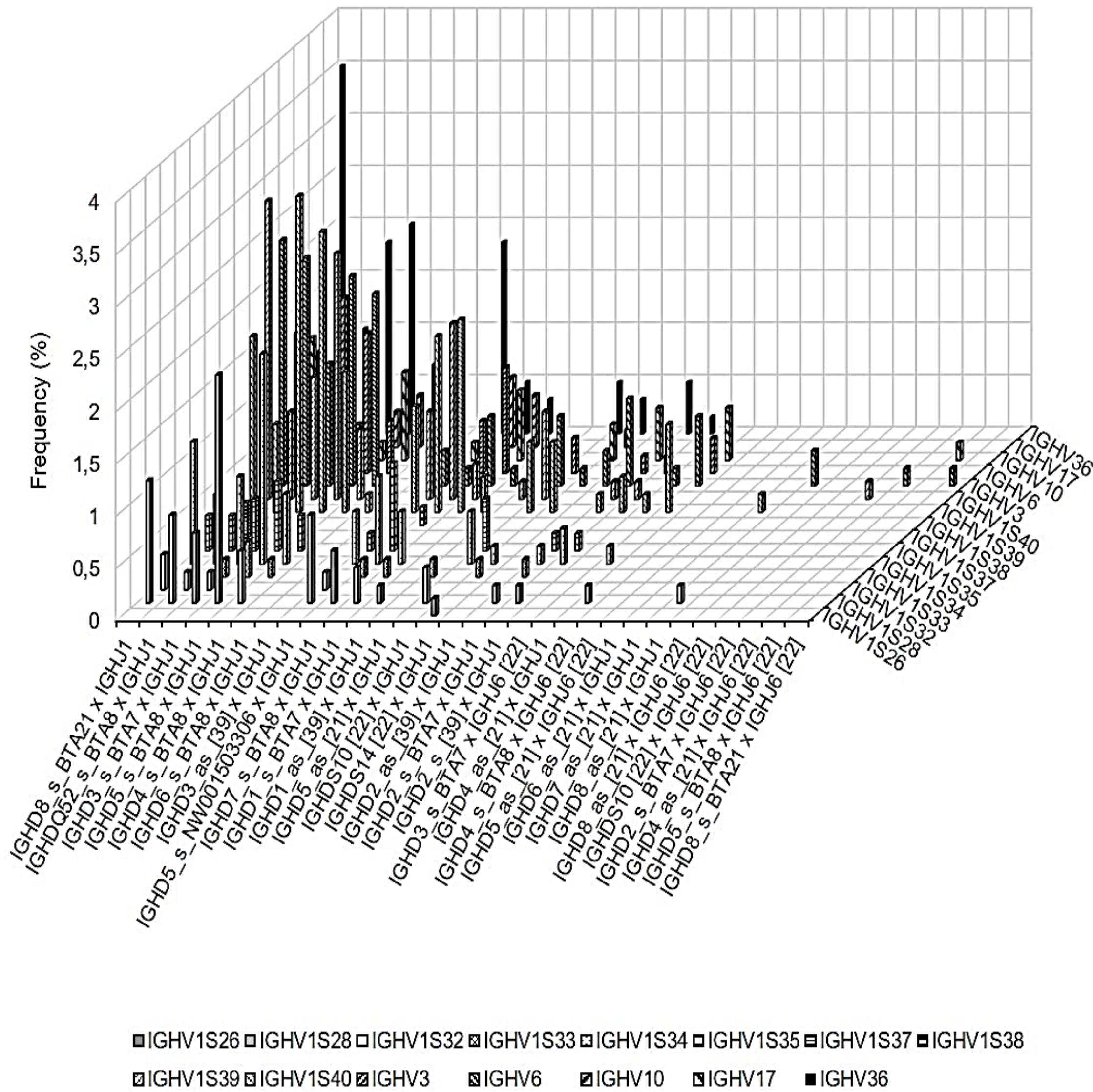
Supporting information



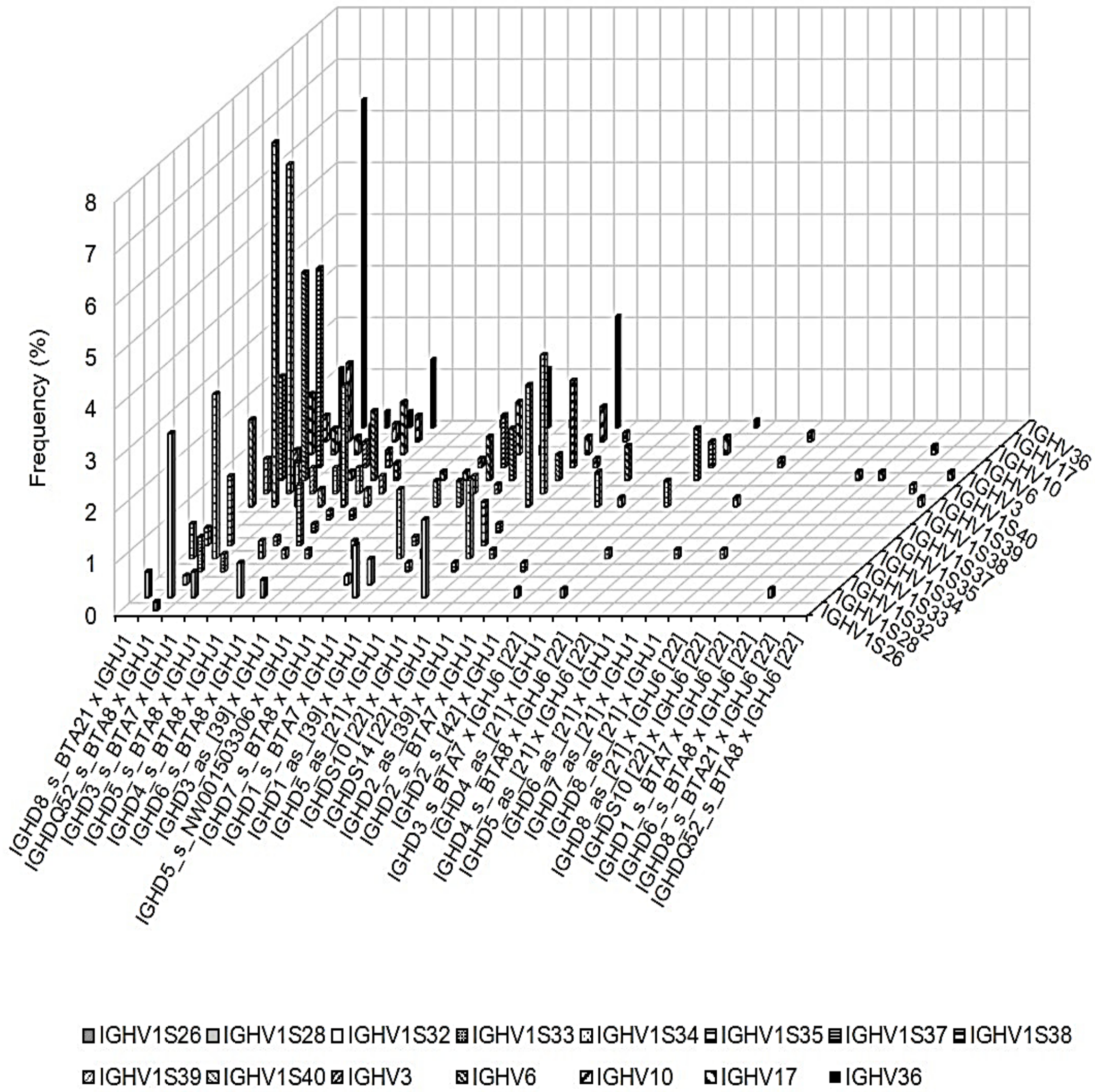
S1 Fig. Transcription frequencies of *IGHD* in four cattle breeds using procedure 1. Transcribed IGHD are shown on the horizontal axis, their relative usage frequencies are indicated on the vertical axis. Calculation occurred using the default values for gap opening and gap extention of MUSCLE. Each breed is marked by the following color code: Aubrac: white, German Simmental: light grey, Holstein Friesian: black, German Black Pied: dark grey



S2 Fig. Transcription frequencies of *IGHD* in four cattle breeds using procedure 2. Transcribed IGHD are shown on the horizontal axis, their relative usage frequencies are indicated on the vertical axis. Calculation occurred after changing the default values for gap opening (-4) and gap extention (-0.3) of MUSCLE. Each breed is marked by the following color code: Aubrac: white, German Simmental: light grey, Holstein Friesian: black, German Black Pied: dark grey



S3 Fig. Recombinations of *IGHV*, *IGHD*, and *IGHJ* over all four cattle breeds using procedure 1. In the sequences of all four cattle breeds analyzed 147 different combinations of *IGHV*, *IGHD*, and *IGHJ* were identified. Relative frequencies (%) of the combinations of the 21 transcribed IGHD and the 2 transcribed IGHJ are shown depending on the rearranged IGHV (n=15).



S4 Fig. Recombinations of *IGHV*, *IGHD*, and *IGHJ* over all four cattle breeds using procedure 2. In the sequences of all four cattle breeds analyzed 119 different combinations of *IGHV*, *IGHD*, and *IGHJ* were identified. Relative frequencies (%) of the combinations of the 21 transcribed IGHD and the 2 transcribed IGHJ are shown depending on the rearranged IGHV (n=15).

S1 Table. Transcription frequencies of *IGHD* in four cattle breeds using procedure 1.

IGHD	A (%; n=137)	GS (%; n=116)	GBP (%; n=111)	HF (%; n=145)
*IGHD1*_as[1]_[39]	2.19	0.86	0.00	0.00
*IGHD1*_s[2]_BTA7	6.57	2.59	7.21	5.52
*IGHD2*_as_[39]	0.00	0.00	0.90	0.69
*IGHD2*_s_BTA7	2.92	5.17	4.50	2.07
*IGHD2*_s_[39]	0.73	1.72	2.70	1.38
*IGHD3*_as_[39]	0.00	0.00	0.00	0.69
*IGHD3*_s_BTA7	14.60	11.21	9.91	13.79
*IGHD4*_as_[21]	2.92	0.86	2.70	2.07
*IGHD4*_s_BTA8	10.95	6.03	2.70	11.72
*IGHD5*_as_[21]	0.73	0.00	0.00	0.00
*IGHD5*_s_BTA8	15.33	13.79	17.12	17.24
*IGHD5*_s_NW001503306	4.38	3.45	12.61	6.90
*IGHD6*_s_BTA8	0.00	0.86	0.00	0.69
*IGHD7*_as_[21]	0.00	0.00	0.90	0.00
*IGHD7*_s_BTA8	9.49	12.93	6.31	4.83
*IGHD8*_as_[21]	0.73	0.86	0.00	0.69
*IGHD8*_s_BTA21	11.68	25.86	15.32	15.86
*IGHDQ52*_s_BTA8	10.95	3.45	9.91	9.66
IGHDS10 [22]	3.65	7.76	4.50	4.14
IGHDS14 [22]	2.19	2.59	2.70	2.07

[1] antisense
[2] sense

S2 Table. Transcription frequencies of *IGHD* in four cattle breeds using procedure 2.

IGHD	A (%; n=137)	GS (%; n=116)	GBP (%; n=111)	HF (%; n=145)
*IGHD1*_as[1]_[39]	2.19	0.86	2.70	2.07
*IGHD1*_s[2]_BTA7	9.49	8.62	7.21	6.21
*IGHD2*_s_BTA7	2.19	0.86	1.80	1.38
*IGHD2*_s_[39]	0.00	0.00	0.90	0.69
*IGHD3*_s_BTA7	2.92	2.59	4.50	4.14
*IGHD4*_as_[21]	3.65	2.59	0.00	2.76
*IGHD4*_s_BTA8	12.41	12.07	4.50	13.10
*IGHD5*_s_BTA8	1.46	3.45	4.50	0.69
*IGHD5*_s_NW001503306	0.00	0.86	1.80	0.69
*IGHD6*_as_[21]	0.00	0.00	0.90	0.00
*IGHD6*_s_BTA8	2.19	0.86	2.70	2.07
*IGHD7*_s_BTA8	0.00	0.86	0.90	2.76
*IGHD8*_as_[21]	0.00	0.00	0.90	0.00
*IGHD8*_s_BTA21	8.76	12.07	11.71	8.28
*IGHDQ52*_s_BTA8	37.96	42.24	39.64	36.55
IGHDS10 [22]	16.79	11.21	15.32	18.62
IGHDS14 [22]	0.00	0.86	0.00	0.00

[1] antisense
[2] sense

S3 Table. Recombinations of *IGHV*, *IGHD*, and *IGHJ* over all four cattle breeds using procedure 3.

IGHV - IGHD - IGHJ	all (n=597)	A (n=167)	GBP (n=140)	GS (n=131)	HF (n=159)
IGHV1S26 - IGHD8_s[1]_BTA21 - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV1S28 - IGHD1_as[2]_[39] - IGHJ1	0.34	0.00	1.43	0.00	0.00
IGHV1S28 - IGHD1_s_BTA7 - IGHJ1	0.50	1.20	0.00	0.00	0.63
IGHV1S28 - IGHD2_s_[39] - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV1S28 - IGHD3_as_[39] - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV1S28 - IGHD3_s_BTA7 - IGHJ1	1.01	1.20	2.14	0.00	0.63
IGHV1S28 - IGHD4_s_BTA8 - IGHJ6_[22]	0.17	0.00	0.00	0.00	0.63
IGHV1S28 - IGHD4_s_BTA8 - IGHJ1	1.17	1.20	1.43	0.76	1.26
IGHV1S28 - IGHD5_s_BTA8 - IGHJ1	1.17	1.20	0.71	2.29	0.63
IGHV1S28 - IGHD5_s_NW_001503306 - IGHJ1	0.34	0.60	0.00	0.76	0.00
IGHV1S28 - IGHD6_s_BTA8 - IGHJ1	0.17	0.00	0.71	0.00	0.00
IGHV1S28 - IGHD7_s_BTA8 - IGHJ1	0.84	0.00	1.43	1.53	0.63
IGHV1S28 - IGHD8_s_BTA21 - IGHJ1	0.34	0.00	0.00	0.76	0.63
IGHV1S28 - IGHDQ52_s_BTA8 - IGHJ1	0.84	0.00	0.71	0.76	1.89
IGHV1S28 - IGHDS10 [22] - IGHJ1	0.67	0.00	0.00	0.00	2.52
IGHV1S28 - IGHDS14 [22] - IGHJ1	0.34	0.60	0.00	0.76	0.00
IGHV1S32 - IGHD2_s_[39] - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV1S32 - IGHD8_s_BTA21 - IGHJ1	0.34	0.00	0.00	1.53	0.00
IGHV1S32 - IGHDS10 [22] - IGHJ1	0.17	0.00	0.00	0.00	0.63
IGHV1S32 - IGHDS14 [22] - IGHJ1	0.17	0.00	0.00	0.00	0.63
IGHV1S33 - IGHD3_as_[39] - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV1S33 - IGHD3_s_BTA7 - IGHJ1	0.34	1.20	0.00	0.00	0.00
IGHV1S33 - IGHD4_s_BTA8 - IGHJ1	0.17	0.00	0.71	0.00	0.00
IGHV1S33 - IGHD5_s_BTA8 - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV1S33 - IGHD8_s_BTA21 - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV1S33 - IGHDQ52_s_BTA8 - IGHJ1	0.34	1.20	0.00	0.00	0.00
IGHV1S33 - IGHDS14 [22] - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV1S34 - IGHD1_as_[39] - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV1S34 - IGHD1_s_BTA7 - IGHJ1	1.34	0.60	3.57	0.76	0.63
IGHV1S34 - IGHD2_s_BTA7 - IGHJ1	0.34	0.60	0.00	0.00	0.63
IGHV1S34 - IGHD2_s_[39] - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV1S34 - IGHD3_s_BTA7 - IGHJ1	0.84	0.00	2.14	0.76	0.63
IGHV1S34 - IGHD4_s_BTA8 - IGHJ1	0.67	0.60	1.43	0.00	0.63
IGHV1S34 - IGHD5_s_BTA8 - IGHJ1	1.01	0.60	1.43	0.76	1.26
IGHV1S34 - IGHD5_s_NW_001503306 - IGHJ1	0.34	0.00	0.71	0.00	0.63
IGHV1S34 - IGHD6_s_BTA8 - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV1S34 - IGHD7_s_BTA8 - IGHJ1	0.34	0.60	0.71	0.00	0.00
IGHV1S34 - IGHD8_s_BTA21 - IGHJ1	1.01	0.00	0.71	3.82	0.00
IGHV1S34 - IGHDQ52_s_BTA8 - IGHJ1	1.51	2.40	2.14	0.76	0.63

IGHV1S34 - IGHDS10 [22] - IGHJ1	0.34	0.00	0.71	0.00	0.63
IGHV1S34 - IGHDS14 [22] - IGHJ1	0.34	0.00	0.00	1.53	0.00
IGHV1S35 - IGHD1_s_BTA7 - IGHJ1	0.17	0.00	0.00	0.00	0.63
IGHV1S35 - IGHD2_s_[39] - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV1S35 - IGHD3_s_BTA7 - IGHJ1	0.50	0.00	0.71	1.53	0.00
IGHV1S35 - IGHD4_as_[21] - IGHJ1	0.34	0.60	0.00	0.76	0.00
IGHV1S35 - IGHD4_s_BTA8 - IGHJ1	0.34	0.60	0.00	0.76	0.00
IGHV1S35 - IGHD5_s_BTA8 - IGHJ1	0.50	0.00	0.00	2.29	0.00
IGHV1S35 - IGHD7_s_BTA8 - IGHJ1	1.01	0.00	0.71	3.82	0.00
IGHV1S35 - IGHD8_s_BTA21 - IGHJ1	0.67	1.20	0.00	1.53	0.00
IGHV1S35 - IGHDQ52_s_BTA8 - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV1S35 - IGHDS10 [22] - IGHJ1	0.17	0.00	0.00	0.00	0.63
IGHV1S37 - IGHD4_s_BTA8 - IGHJ1	0.17	0.00	0.00	0.00	0.63
IGHV1S37 - IGHDQ52_s_BTA8 - IGHJ1	0.17	0.00	0.00	0.00	0.63
IGHV1S38 - IGHD3_s_BTA7 - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV1S38 - IGHDQ52_s_BTA8 - IGHJ1	0.17	0.00	0.00	0.00	0.63
IGHV1S39 - IGHD1_as_[39] - IGHJ1	0.34	0.60	0.00	0.00	0.63
IGHV1S39 - IGHD1_s_BTA7 - IGHJ1	1.01	1.20	1.43	0.00	1.26
IGHV1S39 - IGHD2_as_[39] - IGHJ1	0.34	1.20	0.00	0.00	0.00
IGHV1S39 - IGHD2_s_BTA7 - IGHJ1	0.34	0.60	0.00	0.00	0.63
IGHV1S39 - IGHD2_s_[39] - IGHJ1	0.84	0.00	1.43	0.76	1.26
IGHV1S39 - IGHD3_s_BTA7 - IGHJ6_[22]	0.17	0.00	0.00	0.00	0.63
IGHV1S39 - IGHD3_s_BTA7 - IGHJ1	1.01	1.20	1.43	0.00	1.26
IGHV1S39 - IGHD4_as_[21] - IGHJ1	0.50	0.60	0.00	0.76	0.63
IGHV1S39 - IGHD4_s_BTA8 - IGHJ1	1.68	3.59	0.71	0.00	1.89
IGHV1S39 - IGHD5_s_BTA8 - IGHJ1	2.85	7.78	0.71	0.00	1.89
IGHV1S39 - IGHD5_s_NW_001503306 - IGHJ1	1.17	1.20	2.86	0.76	0.00
IGHV1S39 - IGHD6_as_[21] - IGHJ1	0.34	0.00	0.00	0.00	1.26
IGHV1S39 - IGHD6_s_BTA8 - IGHJ1	0.34	0.00	0.00	0.00	1.26
IGHV1S39 - IGHD7_s_BTA8 - IGHJ1	1.68	2.40	0.71	0.76	2.52
IGHV1S39 - IGHD8_s_BTA21 - IGHJ1	2.18	4.79	1.43	0.00	1.89
IGHV1S39 - IGHDQ52_s_BTA8 - IGHJ1	1.34	2.99	0.00	0.00	1.89
IGHV1S39 - IGHDS10 [22] - IGHJ1	1.01	1.20	2.86	0.00	0.00
IGHV1S39 - IGHDS14 [22] - IGHJ1	0.50	1.20	0.00	0.00	0.63
IGHV1S40 - IGHD1_s_BTA7 - IGHJ1	0.17	0.00	0.00	0.00	0.63
IGHV1S40 - IGHD2_s_BTA7 - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV1S40 - IGHD2_s_[39] - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV1S40 - IGHD3_as_[39] - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV1S40 - IGHD3_s_BTA7 - IGHJ1	2.35	0.00	2.86	4.58	2.52
IGHV1S40 - IGHD4_as_[21] - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV1S40 - IGHD4_s_BTA8 - IGHJ1	0.84	0.60	1.43	0.00	1.26
IGHV1S40 - IGHD5_s_BTA8 - IGHJ1	3.02	2.99	3.57	4.58	1.26
IGHV1S40 - IGHD5_s_NW_001503306 - IGHJ1	0.50	0.60	0.71	0.00	0.63
IGHV1S40 - IGHD6_as_[21] - IGHJ1	0.17	0.00	0.71	0.00	0.00

IGHV1S40 - IGHD6_s_BTA8 - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV1S40 - IGHD7_s_BTA8 - IGHJ1	0.84	0.00	1.43	1.53	0.63
IGHV1S40 - IGHD8_s_BTA21 - IGHJ1	1.34	1.20	2.14	0.76	1.26
IGHV1S40 - IGHDQ52_s_BTA8 - IGHJ1	1.51	0.60	1.43	2.29	1.89
IGHV1S40 - IGHDS10 [22] - IGHJ6_[22]	0.17	0.00	0.71	0.00	0.00
IGHV1S40 - IGHDS10 [22] - IGHJ1	0.50	0.00	0.71	0.76	0.63
IGHV3 - IGHD1_as_[39] - IGHJ1	0.34	0.60	0.00	0.76	0.00
IGHV3 - IGHD1_s_BTA7 - IGHJ1	1.51	0.00	2.14	0.76	3.14
IGHV3 - IGHD2_as_[39] - IGHJ1	0.50	1.20	0.00	0.76	0.00
IGHV3 - IGHD2_s_BTA7 - IGHJ1	0.50	0.60	0.00	0.76	0.63
IGHV3 - IGHD2_s_[39] - IGHJ1	0.34	0.00	0.71	0.00	0.63
IGHV3 - IGHD3_as_[39] - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV3 - IGHD3_s_BTA7 - IGHJ1	1.34	1.20	2.14	0.76	1.26
IGHV3 - IGHD3_s_BTA7 - IGHJ6_[22]	0.34	0.00	0.00	0.76	0.63
IGHV3 - IGHD4_as_[21] - IGHJ1	0.50	0.00	0.71	0.00	1.26
IGHV3 - IGHD4_s_BTA8 - IGHJ1	0.50	1.20	0.71	0.00	0.00
IGHV3 - IGHD5_as_[21] - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV3 - IGHD5_s_BTA8 - IGHJ1	2.01	1.20	1.43	2.29	3.14
IGHV3 - IGHD5_s_NW_001503306 - IGHJ1	0.50	0.60	0.00	0.00	1.26
IGHV3 - IGHD7_s_BTA8 - IGHJ1	0.50	0.60	1.43	0.00	0.00
IGHV3 - IGHD8_as_[21] IGHJ6_[22]	0.17	0.60	0.00	0.00	0.00
IGHV3 - IGHD8_s_BTA21 - IGHJ1	1.68	0.60	0.71	0.76	4.40
IGHV3 - IGHDQ52_s_BTA8 - IGHJ1	1.84	0.00	0.71	3.05	3.77
IGHV3 - IGHDS10 [22] - IGHJ1	1.01	2.40	0.71	0.76	0.00
IGHV3 - IGHDS14 [22] - IGHJ1	0.50	0.00	1.43	0.76	0.00
IGHV6 - IGHD1_as_[39] - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV6 - IGHD1_s_BTA7 - IGHJ1	0.67	0.60	1.43	0.00	0.63
IGHV6 - IGHD2_s_BTA7 - IGHJ1	0.50	1.80	0.00	0.00	0.00
IGHV6 - IGHD2_s_[39] - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV6 - IGHD3_as_[39] - IGHJ1	0.34	0.00	0.71	0.76	0.00
IGHV6 - IGHD3_s_BTA7 - IGHJ1	0.84	0.60	1.43	0.00	1.26
IGHV6 - IGHD4_as_[21] - IGHJ1	0.34	0.60	0.00	0.00	0.63
IGHV6 - IGHD4_s_BTA8 - IGHJ1	1.01	1.20	0.00	0.76	1.89
IGHV6 - IGHD5_s_BTA8 - IGHJ1	0.84	0.60	1.43	0.00	1.26
IGHV6 - IGHD5_s_NW_001503306 - IGHJ1	0.50	0.60	0.71	0.76	0.00
IGHV6 - IGHD7_s_BTA8 - IGHJ1	0.50	0.60	0.71	0.00	0.63
IGHV6 - IGHD8_s_BTA21 IGHJ1	1.34	1.80	0.71	3.05	0.00
IGHV6 - IGHDQ52_s_BTA8 - IGHJ1	1.34	1.80	1.43	0.00	1.89
IGHV6 - IGHDS10 [22] - IGHJ1	0.34	0.00	0.71	0.00	0.63
IGHV6 - IGHDS14 [22] - IGHJ1	0.34	0.00	0.00	1.53	0.00
IGHV10 - IGHD2_as_[39] - IGHJ1	0.34	1.20	0.00	0.00	0.00
IGHV10 - IGHD1_s_BTA7 - IGHJ1	0.34	0.60	0.00	0.00	0.63
IGHV10 - IGHD2_s_BTA7 - IGHJ1	0.50	1.20	0.00	0.76	0.00
IGHV10 - IGHD2_s_[39] - IGHJ1	1.01	0.60	0.00	2.29	1.26

IGHV10 - IGHD3_s_BTA7 - IGHJ1	0.67	0.60	0.00	0.76	1.26
IGHV10 - IGHD4_s_BTA8 - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV10 - IGHD5_as_[21] - IGHJ6_[22]	0.17	0.00	0.00	0.00	0.63
IGHV10 - IGHD5_s_BTA8 - IGHJ1	0.17	0.00	0.00	0.00	0.63
IGHV10 - IGHD5_s_NW_001503306 - IGHJ1	0.17	0.00	0.00	0.00	0.63
IGHV10 - IGHD7_as_[21] - IGHJ1	0.17	0.00	0.71	0.00	0.00
IGHV10 - IGHD7_s_BTA8 - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV10 - IGHD8_as_[21] - IGHJ1	0.34	0.00	0.71	0.00	0.63
IGHV10 - IGHD8_s_BTA21 - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV10 - IGHDQ52_s_BTA8 - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV10 - IGHDS10 [22] - IGHJ1	0.50	0.00	0.00	2.29	0.00
IGHV17 - IGHD1_s_BTA7 - IGHJ1	0.50	0.60	1.43	0.00	0.00
IGHV17 - IGHD3_s_BTA7 - IGHJ1	1.01	1.80	0.71	1.53	0.00
IGHV17 - IGHD4_as_[21] - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV17 - IGHD4_s_BTA8 - IGHJ1	0.34	0.60	0.71	0.00	0.00
IGHV17 - IGHD5_s_BTA8 - IGHJ1	0.50	0.00	1.43	0.00	0.63
IGHV17 - IGHD5_s_NW_001503306 - IGHJ1	0.17	0.00	0.71	0.00	0.00
IGHV17 - IGHD7_s_BTA8 - IGHJ1	0.34	0.00	0.71	0.00	0.63
IGHV17 - IGHD8_s_BTA21 - IGHJ1	0.34	0.00	0.00	1.53	0.00
IGHV17 - IGHDQ52_s_BTA8 - IGHJ1	0.50	1.20	0.00	0.00	0.63
IGHV17 - IGHDS10 [22] - IGHJ1	0.34	0.60	0.71	0.00	0.00
IGHV36 - IGHD1_as_[39] - IGHJ1	0.17	0.60	0.00	0.00	0.00
IGHV36 - IGHD1_s_BTA7 - IGHJ1	0.34	0.60	0.00	0.00	0.63
IGHV36 - IGHD3_as_[39] - IGHJ1	0.50	0.60	0.71	0.76	0.00
IGHV36 - IGHD3_s_BTA7 - IGHJ1	1.51	0.00	2.14	0.00	3.77
IGHV36 - IGHD4_s_BTA8 - IGHJ1	0.84	0.00	0.00	1.53	1.89
IGHV36 - IGHD5_as_[21] - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV36 - IGHD5_s_BTA8 - IGHJ1	1.51	1.80	2.86	0.76	0.63
IGHV36 - IGHD5_s_NW_001503306 - IGHJ1	0.50	0.00	0.71	0.00	1.26
IGHV36 - IGHD6_s_BTA8 - IGHJ1	0.67	0.00	1.43	0.00	1.26
IGHV36 - IGHD7_s_BTA8 - IGHJ1	0.50	0.60	0.00	0.76	0.63
IGHV36 - IGHD8_s_BTA21 - IGHJ1	4.52	1.20	5.00	10.69	2.52
IGHV36 - IGHDQ52_s_BTA8 - IGHJ1	1.84	1.80	3.57	0.76	1.26
IGHV36 - IGHDS10 [22] - IGHJ1	0.17	0.00	0.00	0.76	0.00
IGHV36 - IGHDS14 [22] - IGHJ1	0.17	0.60	0.00	0.00	0.00

[1] sense
[2] antisense

S4 Table. Recombinations of *IGHV*, *IGHD*, and *IGHJ* over all four cattle breeds using procedure 1.

IGHV - IGHD - IGHJ	all (n=597)	GS (n=131)	GBP (140)	A (n=167)	HF (n=159)
IGHV1S26 - IGHDS14 [22] - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S28 - IGHD1_as[1]_[39] - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S28 - IGHD1_s[2]_BTA7 - IGHJ1	0,34	0,00	0,71	0,60	0,00
IGHV1S28 - IGHD2_s_BTA7 - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S28 - IGHD2_s_[39] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S28 - IGHD3_s_BTA7 - IGHJ1	0,67	0,76	0,00	0,60	1,26
IGHV1S28 - IGHD4_s_BTA8 - IGHJ6_[22]	0,17	0,00	0,00	0,00	0,63
IGHV1S28 - IGHD4_s_BTA8 - IGHJ1	0,50	0,00	0,00	0,60	1,26
IGHV1S28 - IGHD5_s_BTA8 - IGHJ1	2,18	3,82	2,86	0,00	2,52
IGHV1S28 - IGHD5_s_NW_001503306 - IGHJ1	0,84	0,76	0,71	1,20	0,63
IGHV1S28 - IGHD7_s_BTA8 - IGHJ1	0,50	0,76	1,43	0,00	0,00
IGHV1S28 - IGHD8_as_[21] - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV1S28 - IGHD8_s_BTA21 - IGHJ1	1,17	1,53	0,71	1,20	1,26
IGHV1S28 - IGHDQ52_s_BTA8 - IGHJ1	0,84	0,76	0,71	1,20	0,63
IGHV1S28 - IGHDS10 [22] - IGHJ1	0,34	0,00	0,00	0,60	0,63
IGHV1S32 - IGHD3_s_BTA7 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S32 - IGHD5_s_NW_001503306 - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV1S32 - IGHD8_s_BTA21 - IGHJ1	0,34	1,53	0,00	0,00	0,00
IGHV1S32 - IGHDQ52_s_BTA8 - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV1S33 - IGHD1_s_BTA7 - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S33 - IGHD2_s_BTA7 - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S33 - IGHD3_s_BTA7 - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S33 - IGHD4_s_BTA8 IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S33 - IGHD5_as_[21] - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S33 - IGHD5_s_BTA8 - IGHJ1	0,34	0,00	0,71	0,60	0,00
IGHV1S33 - IGHD7_s_BTA8 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S33 - IGHDS14 [22] - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S34 - IGHD1_s_BTA7 - IGHJ1	0,50	0,76	0,71	0,60	0,00
IGHV1S34 - IGHD2_s_BTA7 - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S34 - IGHD2_s_[39] - IGHJ1	0,34	0,00	0,71	0,60	0,00
IGHV1S34 - IGHD3_s_BTA7 - IGHJ1	0,84	0,76	1,43	0,60	0,63
IGHV1S34 - IGHD4_as_[21] - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV1S34 - IGHD4_s_BTA8 - IGHJ1	0,67	0,00	2,14	0,00	0,63
IGHV1S34 - IGHD5_s_BTA8 - IGHJ1	2,01	3,05	3,57	1,20	0,63
IGHV1S34 - IGHD5_s_NW_001503306 - IGHJ1	0,50	0,00	1,43	0,00	0,63
IGHV1S34 - IGHD7_s_BTA8 - IGHJ1	0,84	0,76	0,00	1,20	1,26
IGHV1S34 - IGHD8_s_BTA21 - IGHJ1	1,17	3,82	0,71	0,00	0,63
IGHV1S34 - IGHDQ52_s_BTA8 - IGHJ1	0,67	0,00	2,14	0,60	0,00
IGHV1S34 - IGHDS10 [22] - IGHJ1	0,50	0,76	0,00	0,60	0,63
IGHV1S34 - IGHDS14 [22] - IGHJ1	0,17	0,00	0,71	0,00	0,00

IGHV1S35 - IGHD2_s_BTA7 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S35 - IGHD2_s_[39] - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S35 - IGHD3_s_BTA7 - IGHJ1	0,50	0,76	0,00	0,60	0,63
IGHV1S35 - IGHD4_s_BTA8 - IGHJ1	0,34	0,76	0,00	0,00	0,63
IGHV1S35 - IGHD5_s_BTA8 - IGHJ1	0,67	1,53	0,00	1,20	0,00
IGHV1S35 - IGHD5_s_NW_001503306 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S35 - IGHD7_s_BTA8 - IGHJ1	0,84	3,05	0,71	0,00	0,00
IGHV1S35 - IGHD8_s_BTA21 - IGHJ1	0,34	0,76	0,71	0,00	0,00
IGHV1S35 - IGHDQ52_s_BTA8 - IGHJ1	0,34	0,00	0,00	1,20	0,00
IGHV1S35 - IGHDS10 [22] - IGHJ1	0,50	2,29	0,00	0,00	0,00
IGHV1S37 - IGHDQ52_s_BTA8 - IGHJ1	0,34	0,00	0,00	0,00	1,26
IGHV1S38 - IGHD7_s_BTA8 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S38 - IGHDQ52_s_BTA8 - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV1S39 - IGHD1_as_[39] - IGHJ1	0,34	0,00	0,00	1,20	0,00
IGHV1S39 - IGHD1_s_BTA7 - IGHJ1	1,84	0,76	2,86	2,40	1,26
IGHV1S39 - IGHD2_s_BTA7 - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S39 - IGHD2_s_[39] - IGHJ1	0,34	0,00	0,00	0,00	1,26
IGHV1S39 - IGHD3_s_BTA7 - IGHJ6_[22]	0,17	0,00	0,00	0,00	0,63
IGHV1S39 - IGHD3_s_BTA7 - IGHJ1	3,02	0,76	2,14	5,99	2,52
IGHV1S39 - IGHD4_as_[21] - IGHJ1	0,84	0,76	0,00	1,80	0,63
IGHV1S39 - IGHD4_s_BTA8 - IGHJ1	1,34	0,00	0,00	3,59	1,26
IGHV1S39 - IGHD5_s_BTA8 - IGHJ1	2,68	0,00	2,86	5,39	1,89
IGHV1S39 - IGHD5_s_NW_001503306 - IGHJ1	1,01	0,76	1,43	1,20	0,63
IGHV1S39 - IGHD6_s_BTA8 - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV1S39 - IGHD7_as_[21] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S39 - IGHD7_s_BTA8 - IGHJ1	1,68	0,00	0,00	3,59	2,52
IGHV1S39 - IGHD8_s_BTA21 - IGHJ1	1,68	0,00	0,71	1,20	4,40
IGHV1S39 - IGHDQ52_s_BTA8 - IGHJ1	0,84	0,00	1,43	1,20	0,63
IGHV1S39 - IGHDS10 [22] - IGHJ1	0,67	0,00	0,00	1,20	1,26
IGHV1S39 - IGHDS14 [22] - IGHJ1	0,67	0,00	0,71	1,80	0,00
IGHV1S40 - IGHD1_s_BTA7 - IGHJ1	0,34	0,00	0,00	0,00	1,26
IGHV1S40 - IGHD2_s_BTA7 - IGHJ6_[22]	0,17	0,00	0,71	0,00	0,00
IGHV1S40 - IGHD2_s_BTA7 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S40 - IGHD2_s_[39] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S40 - IGHD3_s_BTA7 - IGHJ1	1,17	2,29	1,43	0,60	0,63
IGHV1S40 - IGHD4_s_BTA8 - IGHJ1	0,67	0,76	0,00	1,20	0,63
IGHV1S40 - IGHD5_as_[21] - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S40 - IGHD5_s_BTA8 - IGHJ1	2,35	0,76	1,43	2,40	4,40
IGHV1S40 - IGHD5_s_NW_001503306 - IGHJ1	0,84	0,76	2,14	0,60	0,00
IGHV1S40 - IGHD7_s_BTA8 - IGHJ1	1,68	4,58	1,43	1,20	0,00
IGHV1S40 - IGHD8_s_BTA21 - IGHJ1	2,85	4,58	3,57	1,20	2,52
IGHV1S40 - IGHDQ52_s_BTA8 - IGHJ1	0,84	0,00	2,14	0,60	0,63
IGHV1S40 - IGHDS10 [22] - IGHJ1	0,84	0,76	2,14	0,00	0,63
IGHV3 - IGHD1_as_[39] - IGHJ1	0,17	0,76	0,00	0,00	0,00

IGHV3 - IGHD1_s_BTA7 - IGHJ1	0,67	0,00	1,43	0,00	1,26
IGHV3 - IGHD2_as_[39] - IGHJ1	0,34	0,00	0,71	0,00	0,63
IGHV3 - IGHD2_s_BTA7 - IGHJ1	0,84	1,53	0,71	0,00	1,26
IGHV3 - IGHD3_s_BTA7 - IGHJ6_[22]	0,17	0,00	0,00	0,00	0,63
IGHV3 - IGHD3_s_BTA7 - IGHJ1	1,17	1,53	2,14	0,60	0,63
IGHV3 - IGHD4_as_[21] - IGHJ6_[22]	0,17	0,00	0,00	0,60	0,00
IGHV3 - IGHD4_as_[21] - IGHJ1	0,67	0,00	1,43	1,20	0,00
IGHV3 - IGHD4_s_BTA8 - IGHJ1	1,84	0,76	0,71	2,40	3,14
IGHV3 - IGHD5_s_BTA8 - IGHJ1	2,01	2,29	2,14	0,00	3,77
IGHV3 - IGHD5_s_NW_001503306 - IGHJ1	0,34	0,00	0,71	0,00	0,63
IGHV3 - IGHD7_s_BTA8 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV3 - IGHD8_as_[21] - IGHJ1	0,34	0,76	0,00	0,60	0,00
IGHV3 - IGHD8_s_BTA21 - IGHJ6_[22]	0,17	0,76	0,00	0,00	0,00
IGHV3 - IGHD8_s_BTA21 - IGHJ1	2,35	0,76	1,43	2,99	3,77
IGHV3 - IGHDQ52_s_BTA8 - IGHJ1	2,18	0,00	1,43	2,99	3,77
IGHV3 - IGHDS10 [22] - IGHJ1	0,67	2,29	0,00	0,60	0,00
IGHV3 - IGHDS14 [22] - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV6 - IGHD1_s_BTA7 - IGHJ1	1,01	1,53	0,71	1,80	0,00
IGHV6 - IGHD2_s_BTA7 - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV6 - IGHD2_s_[39] - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV6 - IGHD3_as_[39] - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV6 - IGHD3_s_BTA7 - IGHJ1	1,68	0,00	0,71	2,40	3,14
IGHV6 - IGHD4_as_[21] - IGHJ1	0,34	0,00	0,00	0,00	1,26
IGHV6 - IGHD4_s_BTA8 - IGHJ1	0,50	0,00	1,43	0,00	0,63
IGHV6 - IGHD5_s_BTA8 - IGHJ1	1,34	0,76	2,86	1,20	0,63
IGHV6 - IGHD5_s_NW_001503306 - IGHJ1	0,50	0,76	0,71	0,60	0,00
IGHV6 - IGHD7_s_BTA8 - IGHJ1	0,50	0,00	0,00	0,60	1,26
IGHV6 - IGHD8_s_BTA21 - IGHJ1	1,34	3,82	0,71	0,60	0,63
IGHV6 - IGHDQ52_s_BTA8 - IGHJ1	1,17	0,76	1,43	1,80	0,63
IGHV6 - IGHDS10 [22] - IGHJ1	0,34	0,00	0,71	0,60	0,00
IGHV10 - IGHD1_s_BTA7 - IGHJ1	0,67	0,76	0,00	1,20	0,63
IGHV10 - IGHD2_s_BTA7 - IGHJ1	0,50	1,53	0,00	0,60	0,00
IGHV10 - IGHD3_s_BTA7 - IGHJ1	0,34	0,76	0,00	0,60	0,00
IGHV10 - IGHD4_as_[21] - IGHJ1	0,50	0,00	0,71	1,20	0,00
IGHV10 - IGHD4_s_BTA8 - IGHJ1	0,84	0,76	0,00	1,20	1,26
IGHV10 - IGHD5_s_BTA8 - IGHJ6_[22]	0,17	0,00	0,00	0,00	0,63
IGHV10 - IGHD5_s_BTA8 - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV10 - IGHD5_s_NW_001503306 - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV10 - IGHD8_s_BTA21 - IGHJ1	1,17	1,53	0,71	0,60	1,89
IGHV10 - IGHDQ52_s_BTA8 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV10 - IGHDS14 [22] - IGHJ1	0,34	0,76	0,00	0,00	0,63
IGHV17 - IGHD1_s_BTA7 - IGHJ1	0,50	0,76	1,43	0,00	0,00
IGHV17 - IGHD3_s_BTA7 - IGHJ1	0,84	0,76	0,71	1,80	0,00
IGHV17 - IGHD4_s_BTA8 - IGHJ1	0,50	0,00	1,43	0,00	0,63

IGHV17 - IGHD5_s_BTA8 - IGHJ1	0,34	0,00	0,71	0,60	0,00
IGHV17 - IGHD5_s_NW_001503306 - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV17 - IGHD7_s_BTA8 - IGHJ1	0,67	0,76	0,00	0,60	1,26
IGHV17 - IGHD8_s_BTA21 - IGHJ1	0,50	0,76	0,00	1,20	0,00
IGHV17 - IGHDQ52_s_BTA8 - IGHJ1	0,50	0,00	0,71	1,20	0,00
IGHV17 - IGHDS14 [22] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV36 - IGHD1_s_BTA7 - IGHJ1	0,34	0,00	0,71	0,00	0,63
IGHV36 - IGHD2_s_BTA7 - IGHJ1	0,50	0,00	0,71	0,60	0,63
IGHV36 - IGHD2_s_[39] - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV36 - IGHD3_s_BTA7 - IGHJ1	1,84	0,76	1,43	1,80	3,14
IGHV36 - IGHD4_s_BTA8 - IGHJ1	0,67	1,53	0,00	0,00	1,26
IGHV36 - IGHD5_s_BTA8 - IGHJ1	2,01	0,76	2,86	2,40	1,89
IGHV36 - IGHD5_s_NW_001503306 - IGHJ1	1,84	0,00	3,57	0,60	3,14
IGHV36 - IGHD6_s_BTA8 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV36 - IGHD7_s_BTA8 - IGHJ1	0,50	0,00	1,43	0,60	0,00
IGHV36 - IGHD8_s_BTA21 - IGHJ1	3,52	9,16	3,57	1,80	0,63
IGHV36 - IGHDQ52_s_BTA8 - IGHJ1	1,01	1,53	1,43	0,00	1,26
IGHV36 - IGHDS10 [22] - IGHJ1	0,50	0,76	0,71	0,00	0,63
IGHV36 - IGHDS14 [22] - IGHJ1	0,34	0,76	0,00	0,00	0,63

[1] antisense

[2] sense

S5 Table. Recombinations of *IGHV*, *IGHD*, and *IGHJ* over all four cattle breeds using procedure 2.

IGHV - IGHD - IGHJ	all (n=597)	GS (n=131)	GBP (n=140)	A (n=167)	HF (n=159)
IGHV1S26 - IGHDQ52_s[1]_BTA8 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S28 - IGHD1_s_BTA7 - IGHJ1	1,01	0,00	2,14	1,20	0,63
IGHV1S28 - IGHD2_s_[39] - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV1S28 - IGHD3_s_BTA7 - IGHJ1	0,50	0,76	0,71	0,00	0,63
IGHV1S28 - IGHD4_as[2]_[21] - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S28 - IGHD4_s_BTA8 - IGHJ1	0,67	1,53	0,71	0,60	0,00
IGHV1S28 - IGHD6_s_BTA8 - IGHJ6_[22]	0,17	0,00	0,00	0,00	0,63
IGHV1S28 - IGHD6_s_BTA8 - IGHJ1	0,34	0,00	0,71	0,00	0,63
IGHV1S28 - IGHD8_s_BTA21 - IGHJ1	0,50	2,29	0,00	0,00	0,00
IGHV1S28 - IGHDQ52_s_BTA8 - IGHJ1	3,18	3,05	2,86	2,40	4,40
IGHV1S28 - IGHDS10 [22] - IGHJ1	1,51	0,76	1,43	1,80	1,89
IGHV1S32 - IGHD1_s_BTA7 - IGHJ1	0,50	1,53	0,00	0,00	0,63
IGHV1S32 - IGHD7_s_BTA8 - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV1S32 - IGHDQ52_s_BTA8 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S33 - IGHD1_as_[39] - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S33 - IGHD2_s_BTA7 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S33 - IGHD3_s_BTA7 - IGHJ1	0,34	0,00	0,00	1,20	0,00
IGHV1S33 - IGHDQ52_s_BTA8 - IGHJ1	0,67	0,76	0,71	1,20	0,00
IGHV1S33 - IGHDS10 [22] - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S34 - IGHD1_as_[39] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S34 - IGHD1_s_BTA7 - IGHJ1	1,34	0,76	1,43	2,99	0,00
IGHV1S34 - IGHD4_as_[21] - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S34 - IGHD4_s_BTA8 - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S34 - IGHD5_s_BTA8 - IGHJ1	0,34	0,76	0,71	0,00	0,00
IGHV1S34 - IGHD5_s_NW_001503306 - IGHJ1	0,34	0,76	0,71	0,00	0,00
IGHV1S34 - IGHD6_as_[21] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S34 - IGHD6_s_BTA8 - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S34 - IGHD8_as_[21] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S34 - IGHD8_s_BTA21 - IGHJ1	0,67	1,53	0,71	0,00	0,63
IGHV1S34 - IGHDQ52_s_BTA8 - IGHJ1	3,18	5,34	2,86	1,80	3,14
IGHV1S34 - IGHDS10 [22] - IGHJ1	1,51	0,00	3,57	0,60	1,89
IGHV1S34 - IGHDS14 [22] - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S35 - IGHD1_s_BTA7 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S35 - IGHD4_s_BTA8 - IGHJ1	1,17	3,82	0,71	0,00	0,63
IGHV1S35 - IGHD5_s_BTA8 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S35 - IGHD8_s_BTA21 - IGHJ1	0,34	1,53	0,00	0,00	0,00
IGHV1S35 - IGHDQ52_s_BTA8 - IGHJ1	1,34	3,82	0,00	1,20	0,63
IGHV1S35 - IGHDS10 [22] - IGHJ1	0,84	0,76	0,71	1,80	0,00
IGHV1S37 - IGHD4_s_BTA8 - IGHJ1	0,17	0,00	0,00	0,00	0,63

IGHV1S37 - IGHDS10 [22] - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV1S38 - IGHD4_s_BTA8 - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV1S38 - IGHD6_s_BTA8 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV1S39 - IGHD1_s_BTA7 - IGHJ1	0,50	0,76	0,00	1,20	0,00
IGHV1S39 - IGHD2_s_BTA7 - IGHJ1	0,67	0,00	0,71	1,20	0,63
IGHV1S39 - IGHD2_s_[39] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S39 - IGHD3_s_BTA7 - IGHJ1	0,84	0,00	1,43	1,20	0,63
IGHV1S39 - IGHD4_as_[21] - IGHJ1	0,50	0,00	0,00	1,80	0,00
IGHV1S39 - IGHD4_s_BTA8 - IGHJ1	2,35	0,00	0,71	5,39	2,52
IGHV1S39 - IGHD5_s_BTA8 - IGHJ1	0,34	0,76	0,00	0,60	0,00
IGHV1S39 - IGHD6_as_[21] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV1S39 - IGHD6_s_BTA8 - IGHJ1	0,34	0,00	0,00	1,20	0,00
IGHV1S39 - IGHD7_s_BTA8 - IGHJ1	0,50	0,00	0,71	0,00	1,26
IGHV1S39 - IGHD8_s_BTA21 - IGHJ1	1,68	0,00	2,14	2,40	1,89
IGHV1S39 - IGHDQ52_s_BTA8 - IGHJ6_[22]	0,17	0,00	0,00	0,00	0,63
IGHV1S39 - IGHDQ52_s_BTA8 - IGHJ1	7,04	1,53	5,00	11,38	8,81
IGHV1S39 - IGHDS10 [22] - IGHJ1	2,35	0,00	1,43	4,19	3,14
IGHV1S40 - IGHD1_as_[39] - IGHJ1	0,17	0,00	0,00	0,60	0,00
IGHV1S40 - IGHD1_s_BTA7 - IGHJ1	0,34	0,00	1,43	0,00	0,00
IGHV1S40 - IGHD3_s_BTA7 - IGHJ1	0,50	1,53	0,71	0,00	0,00
IGHV1S40 - IGHD4_s_BTA8 - IGHJ1	0,50	0,00	0,00	0,00	1,89
IGHV1S40 - IGHD5_s_BTA8 - IGHJ1	0,50	0,76	1,43	0,00	0,00
IGHV1S40 - IGHD6_s_BTA8 - IGHJ1	0,34	0,00	0,71	0,60	0,00
IGHV1S40 - IGHD8_s_BTA21 - IGHJ6_[22]	0,17	0,00	0,71	0,00	0,00
IGHV1S40 - IGHD8_s_BTA21 - IGHJ1	0,67	0,76	0,00	1,80	0,00
IGHV1S40 - IGHDQ52_s_BTA8 - IGHJ1	6,37	6,87	9,29	3,59	6,29
IGHV1S40 - IGHDS10 [22] - IGHJ1	2,68	5,34	1,43	1,80	2,52
IGHV3 - IGHD1_as_[39] - IGHJ1	1,01	0,00	1,43	1,20	1,26
IGHV3 - IGHD1_s_BTA7 - IGHJ6_[22]	0,17	0,76	0,00	0,00	0,00
IGHV3 - IGHD1_s_BTA7 - IGHJ1	0,84	1,53	0,00	0,60	1,26
IGHV3 - IGHD2_s_BTA7 - IGHJ1	0,67	0,76	0,71	0,60	0,63
IGHV3 - IGHD4_as_[21] - IGHJ1	1,01	1,53	0,00	1,20	1,26
IGHV3 - IGHD4_s_BTA8 - IGHJ1	1,34	1,53	0,00	0,60	3,14
IGHV3 - IGHD5_s_BTA8 - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV3 - IGHD5_s_NW_001503306 - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV3 - IGHD6_s_BTA8 - IGHJ1	0,34	0,00	0,00	0,60	0,63
IGHV3 - IGHD7_s_BTA8 - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV3 - IGHD8_s_BTA21 - IGHJ1	2,01	1,53	2,86	0,60	3,14
IGHV3 - IGHDQ52_s_BTA8 - IGHJ6_[22]	0,17	0,00	0,00	0,00	0,63
IGHV3 - IGHDQ52_s_BTA8 - IGHJ1	4,02	2,29	5,71	4,19	3,77
IGHV3 - IGHDS10 [22] - IGHJ6_[22]	0,17	0,00	0,00	0,60	0,00
IGHV3 - IGHDS10 [22] - IGHJ1	2,18	1,53	1,43	1,80	3,77
IGHV6 - IGHD1_s_BTA7 - IGHJ1	1,01	0,00	0,71	2,99	0,00
IGHV6 - IGHD3_s_BTA7 - IGHJ1	0,50	0,00	1,43	0,00	0,63

IGHV6 - IGHD4_as_[21] - IGHJ1	0,50	0,00	0,00	0,60	1,26
IGHV6 - IGHD4_s_BTA8 - IGHJ1	0,34	0,00	0,00	0,60	0,63
IGHV6 - IGHD5_s_BTA8 - IGHJ1	0,50	0,76	0,71	0,00	0,63
IGHV6 - IGHD6_as_[21] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV6 - IGHD7_s_BTA8 - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV6 - IGHD8_s_BTA21 - IGHJ1	0,34	0,00	0,71	0,00	0,63
IGHV6 - IGHDQ52_s_BTA8 - IGHJ1	3,85	6,11	3,57	4,19	1,89
IGHV6 - IGHDS10 [22] - IGHJ1	1,68	0,00	1,43	2,40	2,52
IGHV6 - IGHDS14 [22] - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV10 - IGHD1_as_[39] - IGHJ1	0,17	0,00	0,00	0,00	0,63
IGHV10 - IGHD1_s_BTA7 - IGHJ1	1,01	2,29	0,00	1,20	0,63
IGHV10 - IGHD3_s_BTA7 - IGHJ1	0,34	0,00	0,00	0,00	1,26
IGHV10 - IGHD4_as_[21] - IGHJ1	0,34	0,00	0,00	1,20	0,00
IGHV10 - IGHD4_s_BTA8 - IGHJ1	1,01	2,29	0,00	1,20	0,63
IGHV10 - IGHD6_s_BTA8 - IGHJ6_[22]	0,17	0,00	0,00	0,00	0,63
IGHV10 - IGHD8_s_BTA21 - IGHJ1	1,17	1,53	0,71	2,40	0,00
IGHV10 - IGHDQ52_s_BTA8 - IGHJ1	0,50	0,76	0,00	0,00	1,26
IGHV10 - IGHDS10 [22] - IGHJ1	0,34	0,00	0,71	0,00	0,63
IGHV17 - IGHD3_s_BTA7 - IGHJ1	0,34	0,00	0,71	0,60	0,00
IGHV17 - IGHD4_s_BTA8 - IGHJ1	0,50	0,00	0,00	1,80	0,00
IGHV17 - IGHD5_s_BTA8 - IGHJ1	0,34	0,76	0,71	0,00	0,00
IGHV17 - IGHD6_as_[21] - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV17 - IGHD8_s_BTA21 - IGHJ1	0,50	0,00	0,71	0,00	1,26
IGHV17 - IGHDQ52_s_BTA8 - IGHJ1	1,51	0,76	2,86	1,80	0,63
IGHV17 - IGHDS10 [22] - IGHJ1	0,67	0,76	0,71	1,20	0,00
IGHV17 - IGHDS14 [22] - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV36 - IGHD1_as_[39] - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV36 - IGHD1_s_BTA7 - IGHJ1	1,17	0,00	0,71	0,60	3,14
IGHV36 - IGHD3_s_BTA7 - IGHJ1	0,34	0,00	0,71	0,00	0,63
IGHV36 - IGHD4_as_[21] - IGHJ1	0,17	0,76	0,00	0,00	0,00
IGHV36 - IGHD4_s_BTA8 - IGHJ1	1,34	1,53	0,71	0,60	2,52
IGHV36 - IGHD5_s_BTA8 - IGHJ1	0,34	0,00	0,71	0,60	0,00
IGHV36 - IGHD5_s_NW_001503306 - IGHJ1	0,17	0,00	0,71	0,00	0,00
IGHV36 - IGHD8_s_BTA21 - IGHJ1	1,17	1,53	1,43	0,00	1,89
IGHV36 - IGHDQ52_s_BTA8 - IGHJ1	6,37	11,45	7,14	4,19	3,77
IGHV36 - IGHDS10 [22] - IGHJ1	2,18	0,76	4,29	1,80	1,89

[1] sense
[2] antisense

Chapter 4:

General Discussion

General Discussion

Since the resistance of antigens to antibiotics remains a significant problem in livestock production due to considerable economic impacts, research in the alternate treatment of infectious diseases increased. The investigation of special genetic aspects of immunoglobulins has attracted higher attention during the last few decades. New high throughput technologies are available that facilitate and advance the experimental workflow. The price for analyses of the immunoglobulin repertoires is descending. Initial studies on the immunoglobulin repertoire were performed on humans and mice but successful techniques were applied to various other animals such as chicken, rabbit, cattle, horses, and zebrafish, too. Nevertheless, investigation in immunoglobulin genetics still needs more attention to complete previous information about fundamentals, such as the number of genomic available gene segments, gene families, and allotypes of different isotypes of the immunoglobulin heavy and light chains in cattle and horses.

The primary objective of this thesis was to analyze the genomic available and transcribed immunoglobulin heavy and light chain gene segments to contribute to a more detailed understanding of immunoglobulin diversity in cattle and horses.

Similar to most vertebrates, bovine and equine immunoglobulins possess two identical heavy and light chains that consist of a variable region and a constant region. The variable regions are created by the random fusion of germline variable (IGHV/IGLV/IGKV), diversity (IGHD), and joining gene segments (IGHJ/IGLJ/IGKJ) that are combined with a constant region gene (IGHC/IGLC/IGKC) (Tonegawa 1983). The gene segments are found in species-specific numbers in the genome. The immunoglobulin repertoire is generated by several processes of combinatorial and junctional diversity comprising the imprecise joining of the single gene segments and non-templated or palindromic nucleotide insertions between two adjacent gene segments, as well as somatic hypermutations. Secondary mechanisms such as gene conversion and isotype switch may increase the diversity further. Highly specialized, complementarity determining regions form a perfect counterpart to the antigen epitope and are stabilized by conserved framework regions (Kabat and Wu 1991).

In mammals, two types of light chains are described – the lambda and kappa light chains (Korngold and Lipari 1956). They are expressed in species-specific ratios. In

contrast to humans and mice, where the kappa isotype dominates in serum antibodies, the lambda isotype is predominantly found in cattle and horses (Arun et al. 1996; Almagro et al. 1998). In addition to what is already known about the general structural features of both bovine and equine immunoglobulins and their function, this thesis on diversity gave some important baseline information and contributes to the scientific background for the production of highly specific and effective recombinant antibodies or antibody libraries. The findings will also be of future importance in analyzing seroconversion data after infection or vaccination, as well as determining breed specific differences to select healthy, robust animals.

Since the late 19th century, horse antibodies in terms of anti-sera were closely connected with the treatment of human diseases. Today, equine immunoglobulins are also used to support immunosuppression after organ or stem cell transplantation, or to manage autoimmune diseases in humans (Leleu et al. 2006; Zand 2006). In addition, equine antibody applications gain in importance in order to prevent or treat equine infectious diseases. Amongst others they can be regarded as potential alternatives to antibiotic therapy in the near future. Therefore it is necessary to investigate the fundamental immunoglobulin genetics underlying the equine immunoglobulin immune response.

The understanding of the organization of equine immunoglobulin genes has increased significantly in recent years (reviewed in **publication 1**). For equine heavy chains, 52 IGHV, 40 IGHD, 8 IGHJ and 11 IGHC were determined. With 40 IGHD identified, horses belong to the mammalian species that possess the most IGHD. Likewise, in guinea pig and the African elephant 41 and 87 IGHD gene segments are known so far (Guo et al. 2011; Guo et al. 2012). Seven of the equine IGHCs are gamma chain genes (IgG). Only in the porcine immune system a similarly high number of eleven genomic constant region genes representing six putative subclasses has been described (Butler et al. 2009).

Using biochemical and serological methods, the subclasses IgGa, IgGb, IgGc, IgG(T), and IgG(B) were identified and characterized (Rockey et al. 1964; Sandor et al. 1964; Klinman et al. 1965; Rockey 1967; Widders et al. 1986; Sheoran and Holmes 1996). Later, seven IgG isotypes were identified by the analysis of a BAC library. The previous subclasses were renamed. Multiple duplications, gene conversions, and crossovers are supposed to explain these seven IgG isotypes (Wagner et al. 2006). The individual

IgG subclasses possess specific roles in protective immunity. Thus, the FcγR and complement-binding isotypes IgG1, IgG4, and IgG7 (former IgGa and IgGb) contribute to the protection against several equine pathogens such as equine influenza virus (Nelson et al. 1998; Breathnach et al. 2006), *Streptococcus equi* (Sheoran et al. 1997) and *Rhodococcus equi* (Lopez et al. 2002). In addition, systemic and mucosal IgG responses were described to play an important role in limiting the spread and severity of equine herpes virus 1 (Kydd et al. 2006). The first recombinant versions of all seven equine IgG subclasses (mouse λ-light chains, horse IGHG1-7) were expressed in Chinese Hamster Ovary cells to analyze their individual physical and biological properties (Lewis et al. 2008). Further, seven distinct monoclonal equine antibodies (IgM, IgG1, IgG3, IgG4/7, IgG5, IgG6, and IgE) were produced in equine-murine heterohybridomas to be used for quantification of isotypes in diagnostic testing and immunological research (Keggan et al. 2013).

The combinatorial and junctional levels of IGHV-IGHD-IGHJ sequences in equine fetus were comparable to those of adult horses and a similar set of variable gene segments was used during fetal and post-natal life stages (Tallmadge et al. 2009; Tallmadge et al. 2013). Few IGHV segments were used predominantly at all ages. The 250,000 bases between the expressed gene segments suggest that the entire IGHV locus is available throughout equine life (Tallmadge et al. 2013). Similarly, in fetal piglets the IGHV usage was described to be independent of the genome position although there is a limited number of IGHV representing the Ig repertoire, too (Eguchi-Ogawa et al. 2010; Butler et al. 2011). In contrast, the biased usage of germline IGHV in the mouse fetus was explained with different accessible positions within the IGHV locus (Jeong et al. 1988). The incorporation of more than one IGHD into the IGHV-IGHD-IGHJ rearrangement resulting in IGHV-IGHD-IGHD-IGHJ was observed (Sun et al. 2010). In horses, sequence diversity and length variation further increased in complementarity determining region 1 (CDR1H) and CDR2H, and framework regions in accordance to somatic hypermutation (Tallmadge et al. 2013). Within the CDR3H of an adult horse, at least 5 amino acid residues were observed, whereas 25 amino acid residues were counted in the longest CDR3H, which was identified in an equine neonate (Tallmadge et al. 2013). In comparison, recent studies showed the length heterogeneity of very short CDR3H (5-10 amino acids), midlength CDR3H (11-31 amino acids) and exceptionally long CDR3H (more than 47 amino acids) in all isotypes in both bovine fetuses and adult cattle. Very long CDR3H contribute to diversity by uniquely folded

small domains (Saini et al. 1999a; Saini and Kaushik 2002; Shojaei et al. 2003; Koti et al. 2010; Walther et al. 2013; Wang et al. 2013).
The equine kappa light chain contains 60 IGKV, 5 IGKJ and 1 IGKC, whereas there are 144 IGLV, 7 IGLJ, and 7 IGLC for the lambda light chain. Each of the 7 IGLJ is preceded by one IGLC and there are two clusters of IGLV possessing different transcriptional orientations. Within each IGLV cluster there are functional genes and pseudogenes (Sun et al. 2010). Similarly, pseudogenes are described in different species such as chicken, where it is already known that pseudogenes are used for gene conversion (Reynaud et al. 1985). During age development, a decrease in IGLVs is noted, although nucleotide diversity and significant differences in gene usage increased. As in all previous studies different methods were used for designation, the standardization of the existing nomenclature of immunoglobulin genes is suggested.

The study of allotypic variants within the equine λ-light chain locus (**publication 2**) was the first detailed molecular genetic description. Five IGLC1, four IGLC5, and two IGLC6/7 allotypic variants, as well as three alleles for IGLC6/7[a] were observed. Especially for the transcriptional use of IGLC6/7, significant differences were calculated by Chi² tests.
Several amino acid residue variations were located at the accessible surface as confirmed by homology-based predicted 3D structural analysis. Distinct serological properties resulting in allotypic variants due to these modifications are assumed. All allotypic variants showed distinct amino acid residue substitutions in all seven β-sheets according to the IMGT nomenclature (Lefranc et al. 2005). Most of them were located within the solvent accessible surface area (Padlan et al. 1986). Seven of these substitutions were positioned at the interface to the first constant region of the heavy chain (IGHC1). As one of these amino acid substitutions affected one of the conserved residues for the main interaction of constant region of the light chain (IGLC) and IGHC1, the stable assembly of heavy and light chains could be changed. We do not have any information about allelic changes in IGHC1, thus this assumption remains to be verified. Amino acid residue substitutions located at the solvent accessible surface area might be distinguished by specific sera and therefore might be used as markers. Previously, allotypic markers of human immunoglobulin chains were related to the susceptibility of different infectious diseases due to either direct association or linkage disequilibrium with the causative gene (Pandey et al. 1979; Granoff et al. 1984; Pandey

et al. 1995; Pandey 2000; Giha et al. 2009). Furthermore, influences of bovine allotypes on effector functions like complement activation, and age dependent expression have also been shown (Corbeil et al. 1997; Bastida-Corcuera et al. 1999a; Bastida-Corcuera et al. 1999b; Bastida-Corcuera et al. 2000). Specific investigations on functional differences of allotypic variants are missing in horses, but similar to cattle heterozygotic horses are thought to possess an advantage in their humoral immune response.

Beside the presence of predicted germline IGLJ-IGLC pairs and the transcription frequency of IGLC-genes, the usage of IGLV-segments of subgroup 8 was further analyzed. This subgroup contains the highest number of functional IGLV and was shown to be transcribed preferentially (Sun et al. 2010). The variable segments IGLV15, IGLV17, and IGLV25 were preferably transcribed. Significant differences were calculated for the rearrangements with the four IGLC-isotypes transcribed within and among the breeds. Consequently, the V-domain repertoire is dominated by two breed independent IGLV genes and a third breed specific IGLV gene that are productively used for combinatorial joining of IGLJ-IGLC-genes and thus for antibody production. Furthermore, two putative pseudogenes (IGLV74ψ, IGLV101ψ) were transcribed. The isolated sequences showed variations in CDR1 sequences resulting in a frame shift replacing the premature stop codons of germline IGLV74ψ. Individual mutations within the breeds are supposed to enable the transcriptional use of pseudogenes. In addition, *in silico* germline sequences are prepared from a Thoroughbred, which is a conserved breed founded by few stallions. Hence, mutations resulting in pseudogenes may be segregate within this breed but can be functional in other breeds. The transcription of pseudo V-gene segments has already been described in species such as cattle, chicken, and rabbits and occurs by gene conversion (Parng et al. 1996; Winstead et al. 1999; Arakawa and Buerstedde 2009). The results of this study indicated the presence of similar mechanisms in horses. Moreover, the spacer sequences of the recombination signal sequences of the transcribed IGLVs seem to be conserved (Sun et al. 2010). In IGLV101ψ, the spacer sequence is identical to that in IGLV17 and IGLV15, which are the dominantly used IGLV. These identical sequences of the spacers are an additional option for the transcription of potential pseudogenes as was already shown in human IGKV segments (Nadel et al. 1998).

Some of the results could be linked to breeding because two breeds with different stud book sizes and breeding goals were examined. While Rhenish German Coldblood is a

quiet conserved and small breed mainly used for agricultural purposes and breed representations (Bremond and Balzer 2011), animals of Hanoverian Warmblood belong to one of the biggest horse breeds and participate in all disciplines of equestrian sports (Bremond and Balzer 2011). Although the inbreeding coefficient in the Hanoverian population (1.33%) is at lower level than those of smaller populations like the Rhenish German Coldblood (1.73%; (Biedermann et al. 2002; Hamann and Distl 2008), effective management of breeding is necessary to prevent the random loss of alleles caused by large genetic contributions of few individuals over a long time period (bottleneck effect). Genetic drift and homogenization of genetic information may also occur and have to be controlled as the rate of inbreeding has increased in Hanoverians in the last ten years (Constans et al. 1985; Hamann and Distl 2008; Simianer and Kohn 2010).
This study characterized genetic and transcriptional differences of λ-light chain in two horse breeds. For the first time distinct alleles and putative allotypic variants were described in horses. Future investigations should evaluate the distribution of allotypic markers of immunoglobulin chains in further horse breeds. Moreover, the linkage to infectious diseases such as in humans and cattle should be examined.

The *in silico* analyses of the genomic organization of the bovine heavy-chain locus revealed differences from previous mapping and annotation results (**publication 3**). The functional locus was mapped to BTA21 (Gu et al. 1992; Hayes et al. 2000; Zhao et al. 2003). Additional bovine heavy-chain loci were detected on BTA7, BTA8, and BTA20, unexpectedly. The order of the respective segments deviated from other fully described mammalian loci (Matsuda and Honjo 1996; Sun et al. 2010). Further, the constant heavy-chain locus described by means of BAC clone analysis (Zhao et al. 2003), were not introduced into the genomic assembly. Several genes upstream of the heavy-chain locus on the bovine contig-NW_003064289 on BTA7 share homology with HSA19 but some did not show any human equivalent. Similarly, on HSA5pter no equivalent was found to the bovine IgG2(ORF) gene, although the adjacent genes were identified. Hybridization experiments assigned IGHML1 to BTA11 (Tobin-Janzen and Womack 1992; Hayes and Petit 1993), which was supported by the detection of six IGHJ segments on the same chromosome using BAC clone and locus-specific PCR analysis (Hosseini et al. 2004). Nevertheless, *in silico* analyses did not identify an IgM-like locus on BTA11. Consequently, the bovine immunoglobulin heavy-chain locus

seems to be annotated incorrectly and incompletely. This problem may be solved by the re-sequencing of the described localizations and underpinned by different authors and methods.

All of the functional bovine IGHV segments are most closely related to the human clan II family VH2, which explains the exclusive transcription of only one bovine VH family (Sinclair and Aitken 1995; Berens et al. 1997; Sinclair et al. 1997). The second bovine VH family consists only of pseudogenes and clustered with the human VH4 and VH6 family. The description of the second bovine VH family (boVH2) previously proposed was now possible (Berens et al. 1997), but there are still no indications of possible gene conversions using boVH2 segments (Liljavirta et al. 2014) in the sequences investigated, as shown for the bovine λ-light chains (Parng et al. 1996). For further studies the assignment of fragmented loci of bovine IGHV to boVH1 or boVH2 based on sequence identity is recommended.

In bovine IgM, three groups of CDR3H defined by their number of amino acids were described. Independent of nucleotide addition during rearrangement long and short IGHD contribute directly to CDR3H length heterogeneity. Nevertheless, annotation of the transcribed IGHD is complicated due to high homologies among the genomic IGHD segments. In particular, CDR3H length heterogeneity, junctional flexibility, and somatic hypermutation within the recombined IGHD segments led to low sequence identities. Furthermore, intrinsic hot spots targeting somatic hypermutations to CDR1H, CDR2H, and CDR3H were thought to contribute solely to IgM antibody diversification in both bovine fetus and adult cattle (Saini and Kaushik 2002).

CDR3H length heterogeneity was fond in all bovine immunoglobulin isotypes. Hence, exceptionally long CDR3H are apparently not primarily generated to compensate the restricted flexibility of IgM. As the most recent study on the IgG repertoire in calves also showed exceptionally long CDR3H (Larsen and Smith 2012), antigen selection of variable domains and class switch recombination seem to be of higher impact. The exceptionally long CDR3H were generated by the direct fusion of a single IGHV segment (IGHV10), the longest IGHD segment (IGHD2), and one functional IGHJ segment (IGHJ1), as described previously (Koti et al. 2010). In short CDR3H, the preferred use of the short IGHD segments, IGHDQ52 and IGHD4 was observed. All IGHV segments identified were found to be functional. Thus, analyses do not reveal evidence for gene conversion in bovine immunoglobulin heavy chains, which contributes to the diversity of chicken immunoglobulin heavy chains and bovine λ-light

chains (Reynaud et al. 1994; Parng et al. 1996). Both in intermediate and exceptionally long CDR3H conserved short nucleotide sequences (CSNS) were inserted into the IGHV-IGHD junction. This mechanism contributes to antibody diversification in all immunoglobulin heavy chain isotypes and CDR3H lengths. As the insertion of CSNS is supposed to directly follow antigen exposure (Koti et al. 2010), class switch recombination is supposed to be responsible for isotype-independent exceptionally long CDR3H in cattle. In addition, exceptionally long CDR3H do not have a conventional antigen binding site because these CDR3H protrude from the variable domain with support from the λ-light chains. Antigen binding is undertaken by side chains that are exclusively contained within long CDR3H regions, as investigated by structural comparisons with protein toxins (Ramsland et al. 2001).

As described in humans (Corbett et al. 1997), mice (Kaartinen and Mäkelä 1985), chicken (Reynaud et al. 1994), and rabbits (Friedman et al. 1994), mainly hydrophilic amino acid residues such as Tyr, Ser, and Gly were identified in the bovine exceptionally long CDR3H. The occurrence of the hydrophilic reading frame in antigen-binding loops is thought to enhance flexibility and recruit somatic hypermutations for advantageous antigen binding (Corbett et al. 1997).

In accordance with previous descriptions, multiple and mainly even numbered Cys accumulated in the middle of the exceptionally long CDR3H were identified (Saini et al. 1999a; Kaushik et al. 2009). They are predicted to form intra and inter CDRH disulfide bonds that rigidify the combining site or help to stabilize long CDR3H, as demonstrated in the crystallized human Fab Kol (Schmidt et al. 1983), the camel cAb-Lys3 single domain antibody (Desmyter et al. 1996), and the bovine Fabs BLV1H12 and BLV5B8 (Wang et al. 2013). Former observations of an additional Cys in CDR2H when there were only one or three Cys in CDR3H (Saini et al. 1999a), or at least one Cys in CDR3H regions containing more than 12 amino acid residues and no Cys in CDR3H with less than ten amino acids (Lopez et al. 1998) were verified with some reservations. The numbering system had an influence on the number of additional Cys in CDR2H and some sequences with intermediate CDR3H lengths did not possess any Cys.

The bovine immunoglobulin heavy-chain locus was annotated, the expression of exceptionally long CDR3H in the five bovine immunoglobulin isotypes was demonstrated, and their genomic origin was specified. Thus, this study reviewed the opinion that exceptionally long CDR3H are a unique feature of bovine IgM and IgG1-3.

The new bioinformatics framework enables more detailed analyses of bovine immunoglobulin heavy chains and substantially contributes to the understanding of the development of the transcribed bovine immunoglobulin repertoire (**publication 4**). It is now possible to analyze the unique exceptionally long CDR3H group of bovine immunoglobulin heavy chains. Possible gene conversions within the variable region of bovine heavy chains were examined, too. Further, the dominantly transcribed IGHV, IGHD, and IGHJ gene segments and their combinatorial diversity were investigated. In contrast to previous studies, this investigation extended analyses to breed specific differences in the four cattle breeds Aubrac (A), German Simmental (GS), German Black Pied (GBP), and Holstein Friesian (HF).

Recent studies on bovine immunoglobulin genetics focused on antibody diversification strategies and the junctional diversity of the antibody repertoire. Bovine specific diversification mechanisms were identified such as the generation of exceptionally long CDR3H (Saini et al. 1999b; Walther et al. 2013; Wang et al. 2013), the insertion of conserved short nucleotide sequences (CSNS) at the IGHV-IGHD junction (Koti et al. 2010), the use of pseudogene fragments in lambda light chains, as well as gene conversions (Parng et al. 1996), and somatic hypermutations independent of exposure to external antigens during B-cell development in lambda light chains (Lucier et al. 1998). In addition, new germline gene segments were determined in the recent years (Niku et al. 2012; Walther et al. 2013; Liljavirta et al. 2014). Nevertheless, the current bovine genome assembly is still incomplete and the full germline repertoire remains under active investigation.

Previous analyses of rearranged immunoglobulin germline gene segments applied various software tools for sequence alignments. Due to the difficult and error prone manual assembly of different genes, specialized software tools have been developed such as IMGT/Junction Analysis (Yousfi Monod et al. 2004), IMGT/V-QUEST (Brochet et al. 2008; Giudicelli et al. 2011), IMTG/HighV-QUEST (Alamyar et al. 2012), VBASE2 (Retter et al. 2005), JoinSolver (Souto-Carneiro et al. 2004), iHMMun-align (Gaeta et al. 2007), and IgBLAST (Ye et al. 2013). Only IgBLAST provides different settings for nucleotide and protein sequences, the numbering system of Kabat or the IMGT system (Lefranc et al. 2003; Kabat et al. 2006), and germline gene database searches (Ye et al. 2013). Matching germline IGHV, IGHD, and IGHJ genes, as well as details at rearrangement junctions may be analyzed. Organisms such as human, mouse, rat,

rabbit and rhesus monkey are supported but the representation of livestock is missing or incomplete. In particular, the bovine specific occurrence of exceptionally long CDR3H is not taken into account. For more detailed analyses of the origin of the immunoglobulin diversity considering species-specific diversification mechanisms, a new bioinformatics framework was developed. Three different procedures were applied to improve biological significance. Consequently, the program is based on nucleotide sequences, searches our updated bovine specific immunoglobulin germline gene database, and is able to load other pre-designed databases. The currently recommended and most widely used IMGT nomenclature is used for the delineation of FR and CDR. The adjustment of search parameters for IGHV an IGHJ and especially for the identification of IGHD is also possible.

The analysis of Ig heavy chain variable regions in the four cattle breeds revealed the usage of 15 different IGHV segments, 21 IGHD segments, as well as two IGHJ segments. Within the breeds, statistical analyses showed significant different transcription levels of IGHV, IGHD, and IGHJ segments. IGHV1S39 and IGHV3/33 were used most frequently. The segments IGHV1S26, IGHV1S32, IGHV1S33, IGHV1S37, and IGHV1S38 were rarely used. Both findings were consistent with findings for IGHV1S39, IGHV3/33, IGHV1S26, and IGHV1S38 in bovine fetal bone marrow, ileum, and spleen (Liljavirta et al. 2014). The transcription of IGHV1S32 and S37 was described for the first time. In all breeds IGHDS8, IGHDS5, IGHDS10, and IGHDQ52 (=IGHDS9) were preferred. IGHDS1 to IGHDS8 were transcribed in antisense direction, too. In previous studies 14 IGHD were transcribed, where IGHDS5 was the most frequent one in 42% of the sequences analyzed in bovine fetus (Liljavirta et al. 2014). The germline IGHJ1 was clearly preferred but IGHJ6 (Liljavirta et al. 2014) was transcribed in small proportions. In Aubrac, procedure 3 confirmed IGHV1S39-IGHD5-IGHJ1 as the most common recombination of gene segments which is identical to the most frequent finding in bovine fetus (Liljavirta et al. 2014). This recombination belongs to immunoglobulins possessing a CDR3H region of intermediate length.

Possible gene conversion events were identified. For instance, the pseudogene IGHV4Ψ seem to contribute to gene conversion events by nucleotide substitutions. This gene segment meets the criteria for gene conversion such as the location upstream of the rearranged segment (Reynaud et al. 1989; Becker and Knight 1990) and clusters of nucleotide changes (Parng et al. 1996). The flanking homology of the conversion region also supports the genetic exchange (Becker and Knight 1990) and

the separation from IGHV6 by more than 18 kb on the genome allows looping during rearrangement (Walther et al. 2013). In comparison, in chicken the nearest pseudogene is separated by 7 kb (Reynaud et al. 1989). Nevertheless, it is difficult to consider the order of gene segments to evaluate the plausibility of other gene conversions because of the incomplete annotation of the bovine genome (Niku et al. 2012; Walther et al. 2013). Finally, a rare exchange between the two bovine VH families is indicated. These results might be an influence of breed or method of analysis when compared to previous results (Niku et al. 2012).

In the CDR3H, three different groups of lengths were examined in all four cattle breeds: short CDR3H (group 1, ≤10 aa), intermediate length CDR3H (group 2, 11-47 aa), and exceptionally long CDR3H (group 3, ≥48 aa). The breed GS possessed the highest percentage of group 3 CDR3H. In this study, the longest CDR3H with 65 amino acids were found in GS and HF sequences, whereas the longest ever detected CDR3H in cattle was 67 amino acids long using IMGT numbering (Walther et al. 2013). Four amino acids made up the shortest CDR3H in A animals, GS animals, and HF animals.

The combinatorial diversity of transcribed germline IGHV, IGHD, and IGHJ-segments was represented by 162 different rearrangements. They were expressed with significant differences (procedure 3). Different rearrangements were found in the four breeds, whereas 91 different combinations occurred in A, 74 in animals of GBP, 72 in GS animals, and 85 in the breed HF. Most of these combinations were observed in less than ten sequences but seven occurred in up to 21 sequences in all four cattle breeds examined.

In sequences of group 1 CDR3H, the combination of IGHV3-IGHDQ52 (sense, BTA8)-IGHJ1 (AY158087) dominated over all breeds using procedure 3. As IGHDQ52 possesses only four amino acids, these results explain best the origin of short CDR3H. Sequences of group 3 CDR3H mostly exhibited IGHV10-IGHD2 (sense, (Shojaei et al. 2003))-IGHJ1 (procedure 3) in all breeds. Only results from this procedure identified biological meaningful combinations of germline IGHV, IGHD, and IGHJ, and gave the best explanation for the origin of group 3 CDR3H. IGHV10 contributed solely to those CDR3H of exceptional lengths (Niku et al. 2012; Walther et al. 2013; Liljavirta et al. 2014). The “Thr-Thr-Val-His-Gln” terminal motif of IGHV10 that initiates an ascending β strand in the folded antibody, is assumed to enable the formation of the “stalk and knob” structure together with inserted conserved short nucleotide sequences (CSNS) (Wang et al. 2013; Liljavirta et al. 2014). Further, IGHD2 is the longest IGHD segment

identified so far. In contrast, sequences of group 2 showed a higher number of recombinations than sequences of group 1 and 3. The few preferably rearranged gene segments within group 3 CDR3H may indicate specialized antibodies because group 3 regions are unique in cattle.

Variability plots indicated quite similar features of varying amino acid residues at each position within the variable region in all breeds. From CDR1H to CDR3H variability increased, which was described already as concentrated areas of diversity in equine heavy and light chain CDRs (Tallmadge et al. 2013; Tallmadge et al. 2014). In transition areas between FRHs and CDRHs variability was higher than in the middle of CDRHs. The highest variability within the FRHs was identified at position 96 in FR3H. This residue is located on the outer surface of the variable region of the immunoglobulin molecule (Wang et al. 2013) within the area where the constant region faces the variable region. The high variability at this position may indicate an influence on the position and sterical orientation of variable and constant region. This may affect light chain pairing because heavy chains possessing group 3 CDR3H are connected to a special type of lambda light chains (Saini et al. 2003; Wang et al. 2013).

The highest number of recombinations and variability were observed in the breed A, while GS possessed the lowest number of recombinations and showed less variability except in the middle of the CDR3H region. Therefore, the contribution of insertions and deletions to diversity is indicated in case of few rearrangements (Larsen and Smith 2012). The breeds A and GS were kept under the same management in a mixed herd, whereas GBP and HF were kept at different farms. Consequently, the breeds kept in different areas were exposed to different antigens. Thus, the individual number of rearrangements per breed and differences in variability additionally indicate a specialized immune response as animals on one farm are challenged with the same environment.

Important new results were gained by the application of the newly developed bioinformatics framework. Analyses demonstrated that the bovine heavy chain diversity is not restricted to the use of a limited number of germline genes although there are preferred rearrangements within the three groups of CDR3H lengths. We also found strong evidence for gene conversion using pseudogenes. Despite current advances in the understanding of bovine immunoglobulin diversification, future investigations of the germline repertoire are necessary.

Conclusions

This thesis contributed essentially to the more detailed understanding of the expressed immunoglobulin repertoire in cattle and horses.

Allotypic and allelic variants have been described in equine lambda light chains for the first time. Individual mutations within the breeds are supposed to enable the transcriptional use of pseudogenes. Several amino acid residue substitutions were located at the accessible surface as confirmed by homology-based predicted 3D structural analysis. These modifications are supposed to result in distinct serological properties of the allotypic variants. Further, one of the conserved residues for the main interaction of IGLC and IGHC1 was affected and, therefore, the stable assembly of heavy and light chains might be changed.
Specific investigations on functional differences of allotypic variants are missing in horses, but similar to cattle heterozygotic horses are thought to possess an advantage in their humoral immune response.
Significant differences in transcription frequencies of IGLV and IGLC were observed within and between the two horse breeds examined. As these breeds with different stud book sizes and breeding goals were kept under different conditions, the results could be linked to breeding and an influence of the exposure to antigens is strongly indicated.

The bovine immunoglobulin heavy-chain locus seems to be annotated incorrectly and incompletely because inconsistent germline gene segments were identified on various bovine autosomes. Consequently, future investigations of the germline repertoire are necessary.
The CDR3H length heterogeneity was found in all bovine immunoglobulin isotypes. Hence, exceptionally long CDR3H are not primarily generated to compensate the restricted flexibility of IgM and class switch recombination contributes to the bovine specific long CDR3H in all immunoglobulin isotypes.
Previous analyses of rearranged immunoglobulin germline gene segments applied various software tools for sequence alignments. As underlying databases did not support the bovine specific occurrence of exceptionally long CDR3H and more detailed analyses of the transcribed gene segments was required, a new bioinformatics

framework was developed. Within the breeds, statistical analyses showed significant different transcription levels of IGHV, IGHD, and IGHJ segments as described for equine lambda light chains. In the CDR3H, three different groups of lengths were examined in all four cattle breeds. Thus, exceptionally long CDR3H are neither restricted to isotypes nor to cattle breeds. The combinatorial diversity of transcribed germline IGHV, IGHD, and IGHJ-segments showed significant differences. Only few of these rearrangements were preferred within group 3 CDR3H and may indicate specialized antibodies because group 3 regions are unique in cattle. Although the annotation of the bovine germline repertoire is incomplete, possible gene conversion events were identified within the variable region of bovine heavy chains.

The highest variability within the FRHs was identified at position 96 in FR3H may indicate an influence on the position and sterical orientation of variable and constant region.

The breeds A and GS were kept under the same management in a mixed herd, whereas GBP and HF were kept at different farms. Consequently, the breeds kept in different areas were exposed to different antigens. Breed specific numbers of recombinations and differences in variability were observed and may indicate a specialized immune response as animals on one farm are challenged with the same environment.

Future Prospects

Further research in this field should concentrate on more detailed analyses of both the germline repertoire, as well as the expressed immunoglobulin repertoire in cattle and horses. Allelic or haplotypic differences for both variable and constant region genes should be investigated and linked to individuals or breeds. For instance, *in vitro* analyses of bovine IgG already determined the influences of genetic variations on the immune response (Corbeil et al. 1997; Bastida-Corcuera et al. 1999a; Bastida-Corcuera et al. 1999b; Bastida-Corcuera et al. 2000). Similar *in vitro* studies of the effector functions of recombinant versions of the seven equine IgG subclasses also revealed different effects on the immune response (Lewis et al. 2008). Subsequent studies in genetic, structural, and configurational properties of bovine and equine immunoglobulins, as well as analyses of the immunoglobulin repertoire after infection or immunization might offer opportunities for the development of new antibody-based therapeutics. The results will contribute to the analysis and generation of synthetic recombinant antibodies, which replace the production of recombinant antibodies from animals. In addition, unfavorable side effects of conventional therapeutics may be excluded and higher antigen specificity will be achieved. Consequently, recombinant antibodies and antibody fragment related products are important tools for research, diagnostics and therapy (Hust and Dubel 2004). However, only few engineered species-specific antibodies have been developed against veterinary pathogens (Koti et al. 2014). A very promising outcome from immunoglobulin research might be the construction of tailor-designed antibodies based on the format of single chain fragments variable (scFv). In these antibodies VH and VL domains are linked together. Therefore, naïve or immunized immunoglobulin libraries have to be constructed. Then, phage display can be used to identify highly affine and neutralizing recombinant scFv molecules against any desired viral, bacterial, and parasitic antigenic site, as well as toxins. Through genetic manipulation or the fusion of an appropriate constant domain, a promising scFv-molecules can be enhanced, for example to scFv-FCGRT (FcRn) fusion proteins (Kontermann 2011; Koti et al. 2014) or to complete IgG antibodies from different species as established for human antibodies (Moutel et al. 2009). A further advantage is the elongated half-life time and clearance of scFvs *in vivo*, that also achieves effector functions such as opsonization, complement-dependent cytotoxicity (CDC), and antibody-dependent cellular cytotoxicity (ADCC).

In addition, analyses of antigen/antibody binding using X-ray crystal structures (Saini et al. 1999a; Saini and Kaushik 2002; Wang et al. 2013) and sequencing data from high-throughput next generation sequencing will give valuable information on antigen-/antibody interactions. Subsequent amino acid replacement, especially in the CDR-regions may lead to recombinant antibody molecules with enhanced binding efficiency.

References

Alamyar E, Duroux P, Lefranc MP, Giudicelli V. 2012. IMGT((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol Biol* **882**: 569-604.

Almagro JC, Hernandez I, Ramirez MC, Vargas-Madrazo E. 1998. Structural differences between the repertoires of mouse and human germline genes and their evolutionary implications. *Immunogenetics* **47**: 355-363.

Arakawa H, Buerstedde JM. 2009. Activation-induced cytidine deaminase-mediated hypermutation in the DT40 cell line. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **364**: 639-644.

Arun SS, Breuer W, Hermanns W. 1996. Immunohistochemical examination of light-chain expression (lambda/kappa ratio) in canine, feline, equine, bovine and porcine plasma cells. *Zentralblatt fur Veterinarmedizin Reihe A* **43**: 573-576.

Bastida-Corcuera F, Butler JE, Heyermann H, Thomford JW, Corbeil LB. 2000. Tritrichomonas foetus extracellular cysteine proteinase cleavage of bovine IgG2 allotypes. *The Journal of parasitology* **86**: 328-332.

Bastida-Corcuera FD, Butler JE, Yahiro S, Corbeil LB. 1999a. Differential complement activation by bovine IgG2 allotypes. *Veterinary immunology and immunopathology* **71**: 115-123.

Bastida-Corcuera FD, Nielsen KH, Corbeil LB. 1999b. Binding of bovine IgG2a and IgG2b allotypes to protein A, protein G, and Haemophilus somnus IgBPs. *Veterinary immunology and immunopathology* **71**: 143-149.

Becker RS, Knight KL. 1990. Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. *Cell* **63**: 987-997.

Berens SJ, Wylie DE, Lopez OJ. 1997. Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. *International immunology* **9**: 189-199.

Biedermann G, Clar U, Finke A, Bickel M. 2002. Analyse der Population des Rheinisch-Deutschen Kaltbluts. *Züchtungskunde* **74**: 237-249.

Breathnach CC, Clark HJ, Clark RC, Olsen CW, Townsend HG, Lunn DP. 2006. Immunization with recombinant modified vaccinia Ankara (rMVA) constructs encoding the HA or NP gene protects ponies from equine influenza virus challenge. *Vaccine* **24**: 1180-1190.

Bremond J, Balzer B. 2011. Zentrale Dokumentation Tiergenetischer Ressourcen in Deutschland (TGRDEU).

Brochet X, Lefranc MP, Giudicelli V. 2008. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res* **36**: W503-508.

Butler JE, Sun X, Wertz N, Lager KM, Chaloner K, Urban J, Jr., Francis DL, Nara PL, Tobin GJ. 2011. Antibody repertoire development in fetal and neonatal piglets XXI. Usage of most VH genes remains constant during fetal and postnatal development. *Molecular immunology* **49**: 483-494.

Butler JE, Wertz N, Deschacht N, Kacskovics I. 2009. Porcine IgG: structure, genetics, and evolution. *Immunogenetics* **61**: 209-230.

Constans J, Hazout S, Garruto RM, Gajdusek DC, Spees EK. 1985. Population distribution of the human vitamin D binding protein: anthropological considerations. *Am J Phys Anthropol* **68**: 107-122.

Corbeil LB, Gogolewski RP, Kacskovics I, Nielsen KH, Corbeil RR, Morrill JL, Greenwood R, Butler JE. 1997. Bovine IgG2a antibodies to Haemophilus somnus and allotype expression. *Canadian journal of veterinary research = Revue canadienne de recherche veterinaire* **61**: 207-213.

Corbett SJ, Tomlinson IM, Sonnhammer EL, Buck D, Winter G. 1997. Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, "minor" D segments or D-D recombination. *J Mol Biol* **270**: 587-597.

Desmyter A, Transue TR, Ghahroudi MA, Thi MH, Poortmans F, Hamers R, Muyldermans S, Wyns L. 1996. Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nat Struct Biol* **3**: 803-811.

Eguchi-Ogawa T, Wertz N, Sun XZ, Piumi F, Uenishi H, Wells K, Chardon P, Tobin GJ, Butler JE. 2010. Antibody repertoire development in fetal and neonatal piglets. XI. The relationship of variable heavy chain gene usage and the genomic organization of the variable heavy chain locus. *Journal of immunology* **184**: 3734-3742.

Friedman ML, Tunyaplin C, Zhai SK, Knight KL. 1994. Neonatal VH, D, and JH gene usage in rabbit B lineage cells. *J Immunol* **152**: 632-641.

Gaeta BA, Malming HR, Jackson KJ, Bain ME, Wilson P, Collins AM. 2007. iHMMune-align: hidden Markov model-based alignment and identification of germline genes in rearranged immunoglobulin gene sequences. *Bioinformatics* **23**: 1580-1587.

Giha HA, Nasr A, Iriemenam NC, Arnot D, Troye-Blomberg M, Theander TG, Berzins K, ElGhazali G, Pandey JP. 2009. Antigen-specific influence of GM/KM allotypes on IgG isotypes and association of GM allotypes with susceptibility to Plasmodium falciparum malaria. *Malar J* **8**: 306.

Giudicelli V, Brochet X, Lefranc MP. 2011. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. *Cold Spring Harb Protoc* **2011**: 695-715.

Granoff DM, Pandey JP, Boies E, Squires J, Munson RS, Jr., Suarez B. 1984. Response to immunization with Haemophilus influenzae type b polysaccharide-pertussis vaccine and risk of Haemophilus meningitis in children with the Km(1) immunoglobulin allotype. *The Journal of clinical investigation* **74**: 1708-1714.

Gu F, Chowdhary BP, Andersson L, Harbitz I, Gustavsson I. 1992. Assignment of the bovine immunoglobulin gamma heavy chain (IGHG) gene to chromosome 21q24 by in situ hybridization. *Hereditas* **117**: 237-240.

Guo Y, Bao Y, Meng Q, Hu X, Meng Q, Ren L, Li N, Zhao Y. 2012. Immunoglobulin genomics in the guinea pig (Cavia porcellus). *PloS one* **7**: e39298.

Guo Y, Bao Y, Wang H, Hu X, Zhao Z, Li N, Zhao Y. 2011. A preliminary analysis of the immunoglobulin genes in the African elephant (Loxodonta africana). *PloS one* **6**: e16889.

Hamann H, Distl O. 2008. Genetic variability in Hanoverian warmblood horses using pedigree analysis. *J Anim Sci* **86**: 1503-1513.

Hayes H, Di Meo GP, Gautier M, Laurent P, Eggen A, Iannuzzi L. 2000. Localization by FISH of the 31 Texas nomenclature type I markers to both Q- and R-banded bovine chromosomes. *Cytogenet Cell Genet* **90**: 315-320.

Hayes HC, Petit EJ. 1993. Mapping of the beta-lactoglobulin gene and of an immunoglobulin M heavy chain-like sequence to homoeologous cattle, sheep, and goat chromosomes. *Mamm Genome* **4**: 207-210.

Hosseini A, Campbell G, Prorocic M, Aitken R. 2004. Duplicated copies of the bovine JH locus contribute to the Ig repertoire. *Int Immunol* **16**: 843-852.

Hust M, Dubel S. 2004. Mating antibody phage display with proteomics. *Trends in biotechnology* **22**: 8-14.

Jeong HD, Komisar JL, Kraig E, Teale JM. 1988. Strain-dependent expression of VH gene families. *Journal of immunology* **140**: 2436-2441.

Kaartinen M, Mäkelä O. 1985. Reading of D genes in variable frames as a source of antibody diversity. *Immunol Today* **6**: 324-327.

Kabat EA, Wu TT. 1991. Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. *J Immunol* **147**: 1709-1719.

Kabat EA, Wu TT, Perry HM, Gottesmann KS, Foeller C. 2006. *Sequences of proteins of immunological interest*. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, Bethesda, MD.

Kaushik AK, Kehrli ME, Jr., Kurtz A, Ng S, Koti M, Shojaei F, Saini SS. 2009. Somatic hypermutations and isotype restricted exceptionally long CDR3H contribute to antibody diversification in cattle. *Veterinary immunology and immunopathology* **127**: 106-113.

Keggan A, Freer H, Rollins A, Wagner B. 2013. Production of seven monoclonal equine immunoglobulins isotyped by multiplex analysis. *Veterinary immunology and immunopathology* **153**: 187-193.

Klinman NR, Rockey JH, Karush F. 1965. Equine Antihapten Antibody Ii. The Gamma-G(7s-Gamma) Components and Their Specific Interaction. *International journal of cancer Supplement = Journal international du cancer Supplement* **47**: 51-60.

Kontermann RE. 2011. Strategies for extended serum half-life of protein therapeutics. *Curr Opin Biotechnol* **22**: 868-876.

Korngold L, Lipari R. 1956. Multiple-myeloma proteins. III. The antigenic relationship of Bence Jones proteins to normal gammaglobulin and multiple-myeloma serum proteins. *Cancer* **9**: 262-272.

Koti M, Kataeva G, Kaushik AK. 2010. Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies. *Molecular immunology* **47**: 2119-2128.

Koti M, Saini SS, Sachan A, Kaushik A. 2014. Engineered Bovine Antibodies in the Development of Novel Therapeutics, Immunomodulators and Vaccines. *Antibodies* **3**: 205-214.

Kydd JH, Townsend HG, Hannant D. 2006. The equine immune response to equine herpesvirus-1: the virus and its vaccines. *Veterinary immunology and immunopathology* **111**: 15-30.

Larsen PA, Smith TP. 2012. Application of circular consensus sequencing and network analysis to characterize the bovine IgG repertoire. *BMC immunology* **13**: 52.

Lefranc MP, Pommie C, Kaas Q, Duprat E, Bosc N, Guiraudou D, Jean C, Ruiz M, Da Piedade I, Rouard M et al. 2005. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev Comp Immunol* **29**: 185-203.

Lefranc MP, Pommie 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.

Leleu X, Terriou L, Duhamel A, Moreau AS, Andrieux J, Dupire S, Coiteux V, Berthon C, Micol JB, Guieze R et al. 2006. Long-term outcome in acquired aplastic anemia treated with an intensified dose schedule of horse antilymphocyte globulin in combination with androgens. *Ann Hematol* **85**: 711-716.

Lewis MJ, Wagner B, Woof JM. 2008. The different effector function capabilities of the seven equine IgG subclasses have implications for vaccine strategies. *Molecular immunology* **45**: 818-827.

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**: e99808.

Lopez AM, Hines MT, Palmer GH, Alperin DC, Hines SA. 2002. Identification of pulmonary T-lymphocyte and serum antibody isotype responses associated with protection against Rhodococcus equi. *Clin Diagn Lab Immunol* **9**: 1270-1276.

Lopez O, Perez C, Wylie D. 1998. A single VH family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains. *Immunological reviews* **162**: 55-66.

Lucier MR, Thompson RE, Waire J, Lin AW, Osborne BA, Goldsby RA. 1998. Multiple sites of V lambda diversification in cattle. *Journal of immunology* **161**: 5438-5444.

Matsuda F, Honjo T. 1996. Organization of the human immunoglobulin heavy-chain locus. *Adv Immunol* **62**: 1-29.

Moutel S, El Marjou A, Vielemeyer O, Nizak C, Benaroch P, Dubel S, Perez F. 2009. A multi-Fc-species system for recombinant antibody production. *BMC biotechnology* **9**: 14.

Nadel B, Tang A, Escuro G, Lugo G, Feeney AJ. 1998. Sequence of the spacer in the recombination signal sequence affects V(D)J rearrangement frequency and correlates with nonrandom Vkappa usage in vivo. *J Exp Med* **187**: 1495-1503.

Nelson KM, Schram BR, McGregor MW, Sheoran AS, Olsen CW, Lunn DP. 1998. Local and systemic isotype-specific antibody responses to equine influenza virus infection versus conventional vaccination. *Vaccine* **16**: 1306-1313.

Niku M, Liljavirta J, Durkin K, Schroderus E, Iivanainen A. 2012. The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed. *Developmental and comparative immunology* **37**: 457-461.

Padlan EA, Cohen GH, Davies DR. 1986. Antibody Fab assembly: the interface residues between CH1 and CL. *Molecular immunology* **23**: 951-960.

Pandey JP. 2000. Immunoglobulin GM and KM allotypes and vaccine immunity. *Vaccine* **19**: 613-617.

Pandey JP, Elson LH, Sutherland SE, Guderian RH, Araujo E, Nutman TB. 1995. Immunoglobulin kappa chain allotypes (KM) in onchocerciasis. *The Journal of clinical investigation* **96**: 2732-2734.

Pandey JP, Fudenberg HH, Virella G, Kyong CU, Loadholt CB, Galbraith RM, Gotschlich EC, Parke JC, Jr. 1979. Association between immunoglobulin allotypes and immune responses to Haemophilus influenzae and Meningococcus polysaccharides. *Lancet* **1**: 190-192.

Parng CL, Hansal S, Goldsby RA, Osborne BA. 1996. Gene conversion contributes to Ig light chain diversity in cattle. *Journal of immunology* **157**: 5478-5486.

Ramsland PA, Kaushik A, Marchalonis JJ, Edmundson AB. 2001. Incorporation of long CDR3s into V domains: implications for the structural evolution of the antibody-combining site. *Exp Clin Immunogenet* **18**: 176-198.

Retter I, Althaus HH, Munch R, Muller W. 2005. VBASE2, an integrative V gene database. *Nucleic Acids Res* **33**: D671-674.

Reynaud CA, Anquez V, Dahan A, Weill JC. 1985. A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell* **40**: 283-291.

Reynaud CA, Bertocci B, Dahan A, Weill JC. 1994. Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. *Adv Immunol* **57**: 353-378.

Reynaud CA, Dahan A, Anquez V, Weill JC. 1989. Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell* **59**: 171-183.

Rockey JH. 1967. Equine antihapten antibody. The subunits and fragments of anti-beta-lactoside antibody. *The Journal of experimental medicine* **125**: 249-275.

Rockey JH, Klinman NR, Karush F. 1964. Equine Antihapten Antibody. I. 7s Beta-2a- and 1os Gamma-1- Globulin Components of Purified Anti-Beta-Lactoside Antibody. *The Journal of experimental medicine* **120**: 589-609.

Saini SS, Allore B, Jacobs RM, Kaushik A. 1999a. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *European journal of immunology* **29**: 2420-2426.

Saini SS, Farrugia W, Ramsland PA, Kaushik AK. 2003. Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted VH + Vlambda pairings. *Int Immunol* **15**: 845-853.

Saini SS, Kaushik A. 2002. Extensive CDR3H length heterogeneity exists in bovine foetal VDJ rearrangements. *Scandinavian journal of immunology* **55**: 140-148.

Saini SS, MacGlashan DW, Jr., Sterbinsky SA, Togias A, Adelman DC, Lichtenstein LM, Bochner BS. 1999b. Down-regulation of human basophil IgE and FC epsilon RI alpha surface densities and mediator release by anti-IgE-infusions is reversible in vitro and in vivo. *J Immunol* **162**: 5624-5630.

Sandor G, Korach S, Mattern P. 1964. 7s Globulin, Immunologically Identical to 19s Gamma-1 (Beta-2)-M-Globulin, a New Protein or Horse Serum. *Nature* **204**: 795-796.

Schmidt WE, Jung HD, Palm W, Hilschmann N. 1983. [Three-dimensional structure determination of antibodies. Primary structure of crystallized monoclonal immunoglobulin IgG1 KOL, I]. *Hoppe Seylers Z Physiol Chem* **364**: 713-747.

Sheoran AS, Holmes MA. 1996. Separation of equine IgG subclasses (IgGa, IgGb and IgG(T)) using their differential binding characteristics for staphylococcal protein A and streptococcal protein G. *Veterinary immunology and immunopathology* **55**: 33-43.

Sheoran AS, Sponseller BT, Holmes MA, Timoney JF. 1997. Serum and mucosal antibody isotype responses to M-like protein (SeM) of Streptococcus equi in convalescent and vaccinated horses. *Veterinary immunology and immunopathology* **59**: 239-251.

Shojaei F, Saini SS, Kaushik AK. 2003. Unusually long germline DH genes contribute to large sized CDR3H in bovine antibodies. *Mol Immunol* **40**: 61-67.

Simianer H, Kohn F. 2010. Genetic management of the Gottingen Minipig population. *J Pharmacol Toxicol Methods* **62**: 221-226.

Sinclair MC, Aitken R. 1995. PCR strategies for isolation of the 5' end of an immunoglobulin-encoding bovine cDNA. *Gene* **167**: 285-289.

Sinclair MC, Gilchrist J, Aitken R. 1997. Bovine IgG repertoire is dominated by a single diversified VH gene family. *Journal of immunology* **159**: 3883-3889.

Souto-Carneiro MM, Longo NS, Russ DE, Sun HW, Lipsky PE. 2004. Characterization of the human Ig heavy chain antigen binding complementarity determining

region 3 using a newly developed software algorithm, JOINSOLVER. *J Immunol* **172**: 6790-6802.

Sun Y, Wang C, Wang Y, Zhang T, Ren L, Hu X, Zhang R, Meng Q, Guo Y, Fei J et al. 2010. A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. *Developmental and comparative immunology* **34**: 1009-1020.

Tallmadge RL, McLaughlin K, Secor E, Ruano D, Matychak MB, Flaminio MJ. 2009. Expression of essential B cell genes and immunoglobulin isotypes suggests active development and gene recombination during equine gestation. *Dev Comp Immunol* **33**: 1027-1038.

Tallmadge RL, Tseng CT, Felippe MJ. 2014. Diversity of immunoglobulin lambda light chain gene usage over developmental stages in the horse. *Dev Comp Immunol* **46**: 171-179.

Tallmadge RL, Tseng CT, King RA, Felippe MJ. 2013. Developmental progression of equine immunoglobulin heavy chain variable region diversity. *Dev Comp Immunol* **41**: 33-43.

Tobin-Janzen TC, Womack JE. 1992. Comparative mapping of IGHG1, IGHM, FES, and FOS in domestic cattle. *Immunogenetics* **36**: 157-165.

Tonegawa S. 1983. Somatic generation of antibody diversity. *Nature* **302**: 575-581.

Wagner B, Miller WH, Morgan EE, Hillegas JM, Erb HN, Leibold W, Antczak DF. 2006. IgE and IgG antibodies in skin allergy of the horse. *Vet Res* **37**: 813-825.

Walther S, Czerny CP, Diesterbeck US. 2013. Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins. *PLoS One* **8**: e64234.

Wang F, Ekiert DC, Ahmad I, Yu W, Zhang Y, Bazirgan O, Torkamani A, Raudsepp T, Mwangi W, Criscitiello MF et al. 2013. Reshaping antibody diversity. *Cell* **153**: 1379-1393.

Widders PR, Stokes CR, Bourne FJ. 1986. Investigation of the antigenic relationship between equine IgG and IgGT. *Veterinary immunology and immunopathology* **13**: 255-259.

Winstead CR, Zhai SK, Sethupathi P, Knight KL. 1999. Antigen-induced somatic diversification of rabbit IgH genes: gene conversion and point mutation. *J Immunol* **162**: 6602-6612.

Ye J, Ma N, Madden TL, Ostell JM. 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic acids research* **41**: W34-40.

Yousfi Monod M, Giudicelli V, Chaume D, Lefranc MP. 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**: i379-385.

Zand MS. 2006. B-cell activity of polyclonal antithymocyte globulins. *Transplantation* **82**: 1387-1395.

Zhao Y, Kacskovics I, Rabbani H, Hammarstrom L. 2003. Physical mapping of the bovine immunoglobulin heavy chain constant region gene locus. *The Journal of biological chemistry* **278**: 35024-35032.

Chapter 5:

Appendix

Curriculum Vitae

Personal Details

Name	Stefanie Walther
Date of birth	16 June 1985
Place of birth	Sangerhausen
Nationality	German

Education

Name of school/ college	Subject	Date
Geschwister-Scholl Gymnasium, Sangerhausen	General qualification for university entrance	07/2005
University Bayreuth	Bachelor of Science in Biochemistry Thesis: Analysis of the interactions of the amino acid residues 33, 34, and 37, as well as their influence on the stability of the Gβ1 protein of *Streptococcus*	10/2005 – 09/2008
Georg-August-University Göttingen	Master of Science in Equine Sciences Thesis: Genomic organisation and transcription analysis oft he equine lambda and kappa immunoglobulin genes in the horse breeds Rhenish-German Coldblood and Hanoverian Warmblood	10/2008 – 03/2011
Georg-August-University Göttingen Division of Microbiology and Animal Hygiene, Institute of Veterinary Medicine, Department of Animal Sciences	Ph.D. Program for Agricultural Sciences Thesis: Organization and transcription analyses of the immunoglobulin genes in cattle and horses	10/2011 – 05/2016