BOVINE B CELLS:

ANTIBODY REPERTOIRE DIVERSIFICATION IN FETAL CATTLE

Jenni Liljavirta

Veterinary Anatomy and Developmental Biology
Department of Veterinary Biosciences
Faculty of Veterinary Medicine

and

Doctoral Program in Integrative Life Science (ILS)
Doctoral School in Health Sciences
University of Helsinki

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Veterinary Medicine of the University of Helsinki, for public examination in the Infocenter Korona, Lecture hall 2, Viikinkaari 11, Helsinki, on October 10th 2014, at 12 o'clock noon.

Helsinki 2014

Supervisors: Professor Antti Iivanainen, DVM, PhD
University of Helsinki, Finland

Mikael Niku, PhD
University of Helsinki, Finland

Reviewers: Research Director Pierre Boudinot, PhD
French National Institute for Agricultural Research (INRA), France

Associate Professor Azad K. Kaushik, DVM, DSc
University of Guelph, Canada

Opponent: Docent Petteri Arstila, MD, PhD
University of Helsinki, Finland

Published in Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis

ISBN 978-951-51-0212-6 (paperback)
ISBN 978-951-51-0213-3 (PDF, http://ethesis.helsinki.fi)
ISSN 2342-3161 (Print)
ISSN 2342-317X (Online)

Half an hour old calf Eximia with its dam Ticoria. Eximia was born in 25.03.2009 in the Viikki Research farm and contributed to our κ/ λ light chain studies by donating B cells.

Hansaprint
Vantaa 2014

“Gli esami non finiscono mai”
“Exams never end”

Eduardo De Filippo, 1973

ABSTRACT

The majority of research studies on immunological mechanisms have been conducted in human or rodent models and the results are generalized in respect of other vertebrates. However, the generation of B cells varies considerably between species. B cells are produced in the bone marrow of rodents and humans throughout life. The initial antibody diversity is produced by the random assembly of a great variety of antibody encoding gene segments. The repertoire of these gene segments is limited in many domestic species such as cattle, sheep, chicken, rabbit and pig and *de novo* B lymphopoiesis takes place only during fetal and/ or neonatal life. This results in a basal B cell population but one that is capable of being further expanded by other mechanisms. In this thesis, the specific diversification mechanisms in the bovine are explored.

Knowledge of immunoglobulin genes is essential in order to distinguish various biological mechanisms, which are involved in antibody repertoire diversification. After the bovine genome was sequenced in 2009, the characterization of the immunoglobulin loci became possible. Previously, the bovine immunoglobulin light chain loci were characterized by only a small number of functional gene segments that had been found. Subsequently, we were able to analyse the heavy chain locus and found a total number of 62 heavy chain variable gene segments, of which 10-20 were verified as functional genes.

Many livestock species and chicken rely on gut-associated lymphoid tissue for further expanding their fetal/neonatal antibody repertoire to compensate for the limited effective recombinatorial diversity. In cattle, the ileal Peyer's patch (IPP) is considered to be the major organ for B cell proliferation. In this research, the fingerprints of somatic hypermutation (SHM) in the IPP were analysed. SHM is dependent on activation-induced cytidine deaminase (AID), which contributes to modifying the immunoglobulin variable regions. SHM is conventionally considered to be a secondary diversification mechanism, which is activated with external antigens. AID-mediated SHM was shown to diversify the antibody repertoire in the fetal IPP, before the exposure of exogenous antigen encounters.

Junctional diversity is produced by the random additions and excisions of nucleotides between immunoglobulin gene segments that occur during the somatic recombination. Non-templated nucleotides are added by terminal deoxynucleotidyl transferase (TdT). Sequence investigation in this study indicated that TdT-mediated junctional diversity contributes to the diversification of the antibody repertoire in bovine fetuses. Extensive junctional diversity mainly in the heavy chain sequences in bone marrow, ileum and spleen but also to a lesser extent in the light chains was detected in this study.

The following model for bovine preimmune repertoire diversification was suggested in this study. First, the restricted immunoglobulin germline repertoire is diversified by junctional diversity in fetal bone marrow. Second, a small population of the B cell clones migrates to the fetal/neonatal IPP. In the IPP, further diversification by AID-mediated SHM takes place, which is associated with extensive proliferation. Third, these clones migrate to other peripheral organs where they are subjected to secondary, antigen-induced modifications. The IPP starts to involute in young animals and the animal survives for the rest of its life by proliferating and differentiating its B cell clones from the peripheral repertoire.

ACKNOWLEDGEMENTS

This study was carried out at the Department of Veterinary Biosciences, Faculty of Veterinary Medicine, University of Helsinki. I wish to thank Professors Antti Sukura and Airi Palva for providing splendid research facilities. The following institutions and foundations are kindly acknowledged for funding: The Academy of Finland, Ministry of Agriculture and Forestry, University of Helsinki, Finnish Veterinary Foundation, Finnish Foundation of Veterinary Research, Albin Johanssons Foundation, Orion-Farmos Research Foundation, the Finnish Society for Immunology and Viikki Doctoral Programme in Molecular Biosciences (VGSB).

I wish to express my sincere gratitude to my supervisor Antti Iivanainen who introduced me to the fascinating world of Veterinary Immunology. I want to thank you for letting me pursue my research work independently; however, your door has always been open for discussions. I am indepted to my other supervisor Mikael Niku for always having the time for answering my endless (and not always so intelligent) questions and reviewing my written productions with patience; I have learned so much from you. I also want to thank you for keeping up the spirit in the lab with neverending energy!

I want to thank members of my thesis follow up group Petri Auvinen and Urmas Arumäe for their invaluable comments during my studies. I am grateful to my pre-examiners, Azad K. Kaushik and Pierre Boudinot for reviewing this thesis and making many excellent questions and comments.

I wish to thank my co-authors abroad Jacqueline Knight, Tony Pernthaner and Keith Durkin for sharing their expertise.

I want to thank all my colleagues and co-authors at the Division of Anatomy: Tiina PM. for sharing the scientific (and non-scientific) expertise and advise and also for reviewing this thesis; Tiina S. and Anna, for the endless support in whatever matter, memorable conference trips and most importantly for their friendship during these years; Kirsi for the excellent work in the lab with hundreds (thousands?) of clonings and qPCR samples and discussions during difficult moments; Tuire for the sample collection and preparation as well as helping with all my gardening problems during relaxing coffee breaks; Juha for the memorable moments at the dissection hall; Matti and Joanna for collecting blood samples; Santeri, Thomas and Lea for bringing new enthusiastic atmosphere in the lab and all other colleagues from the Department of Veterinary Biosciences for their friendship; Esa, Simo, Jaana, Maria, Pauli, Pirjo, Jonna, Pikka, Veera, Anja, Tarja, Ritva and others I forgot to mention.

I am most grateful to my favorite fly-friends Jaana and Riitta for your peer support, help and therapeutic discussions during all these years. Eventually we did it! I also

wish to thank my former supervisor Osamu Shimmi for teaching me many essential lab skills. I warmly thank all other *Drosophila* colleagues: Ida, Shinya, Zhao, Jaakko and Mari for helping me taking the first steps in my scientific career. I also wish to thank my co-PhD students from the VGSB as well as the coordinators for the memorable recreation days, meetings, courses and conference trips during my studies.

Last but not least I sincerely thank my family and oldest friends; äiti, iska, Kirsi, Konsta, Marco and Lola, Mine and Sini. Thank you for the gourmet food, maintained car, IT support, compulsory early morning and late evening walks, definitely non-scientific discussions, and always being there for me.

Espoo, September 2014

CONTENTS

1. LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications:

I 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 Jul;37(3-4):457-61.

II **Liljavirta J**, Ekman A, Knight JS, Pernthaner A, Iivanainen A, Niku M. Activation-induced cytidine deaminase (AID) is strongly expressed in ileal Peyer's patches of bovine fetuses and is associated with the expansion of the primary antibody repertoire in the absence of exogenous antigens. *Mucosal Immunol*. 2013 Sep;6(5):942-9.

III **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 Jun 13;9(6):e99808.

2. ABBREVIATIONS

A	adenine
AID	activation-induced cytidine deaminase
APE	purinic/apyrimidinic endonuclease
BAC	bacterial artificial chromosome
BER	base excision repair
BTA	Bos taurus assembly
C	constant gene segment *or* cytosine
CDR	complementarity determining region
CSR	class switch recombination
D	diversity gene segment
DNA-PK	DNA-dependent protein kinase
Exo1	exonuclease1
FISH	fluorescence *in situ* hybridization
FR	framework region
G	guanine
GALT	gut-associated lymphoid tissue
Ig	immunoglobulin
IGH	immunoglobulin heavy chain
IGH-ML	IgM-like locus
IGK	immunoglobulin kappa (κ) light chain
IGL	immunoglobulin lambda (λ) light chain
IMGT	The international immunogenetics information system
IPP	ileal Peyer's patch
J	joining gene segment
JPP	jejunal Peyer's patch
MLH	MutatorL homolog
MMR	mismatch repair
MSH	MutatorS homolog
PCNA	proliferating cell nuclear antigen
PMS	postmeiotic segregation increased
RAG	recombination-activating gene
RIN	RNA integrity number
RSS	recombination signal sequence
RT-qPCR	reverse transcription quantitative PCR
SHM	somatic hypermutation
Sμ	switch region
T	thymine
TdT	terminal deoxynucleotidyl transferase
TdTL	TdT long isoform
TdTS	TdT short isoform
U	uracil
UNG	uracil DNA glycosylase
V	variable gene segment
XRCC	X-ray repair cross-complementing

3. INTRODUCTION

All living animals and plants have some kind of immune system for protection against diseases. This innate immune system presents the most primordial origin of immune defence. The majority of invertebrates rely on these simple defence strategies, which are based on the recognition of pathogen-associated molecular patterns by genome-encoded receptors. In addition to the innate system, vertebrates use the more sophisticated mechanisms of the adaptive immune system. However, during the last decade this concept has been challenged by showing that also invertebrates such as Arthropods and Molluscs are capable of producing diversified antigen recognition molecules and having adaptive-like immune systems (Zhang et al., 2004, Watson et al., 2005). Major parts of the adaptive immune responses are mediated by randomly assembled antibody receptors called antibodies, which are secreted by a class of white blood cells called B lymphocytes.

Millions of different antibodies are needed during an individual's lifetime. However, the human genome consists of between 20 000-25 000 genes (National Human Genome Research Institute, 01.02.2014), which suggests that not all of these antibodies can be encoded by the human genome as such. Instead, antibodies are encoded in small gene segments, which are randomly assembled together in a tightly regulated somatic recombination process. However, even this random recombination activity is not enough to produce the vast amount of pathogen specific antibodies needed. The further expansion of antibody diversity requires vertebrates to have several mechanisms in use. For example, selected segments of the antibodies can be mutated by somatic hypermutation (SHM) or larger pieces of DNA sequence can be exchanged by homologous segments by gene conversion. Comparative immunology has helped researchers to resolve several of these fundamental mechanisms, reviewed for example by Flajnik (2002). Those studies have revealed that B cell development and the generation of the preimmune antibody repertoire varies between species. The preimmune (primary) antibody repertoire consists of antibodies that are present in fetal and newborn animals before exposure to external antigens.

This thesis focuses on the diversification of the preimmune repertoire in cattle. Cattle belong to the *Bovidae* family within the order *Cetartiodactyla,* which consist of even toed-ungulates and whales. *Bos taurus,* or more commonly European cattle, is the most widespread species within the genus *Bos*. In 2009 *Bos taurus* became the first livestock animal whose genome was sequenced due to the efforts of the *Bos taurus* genome consortium. Yet, the immunological characteristics of this globally important species have thus far been largely unknown.

The principles behind innate and adaptive immunity are first reviewed then this is followed by the introduction of immunoglobulin structure and common known

diversification mechanisms. First, the general mechanisms are described. Second, adaptations that cattle, sheep, pig, rabbit and chicken (referred as domestic animals/ species from now on) possess are summarized. Finally; the results of studies that found new bovine immunoglobulin genes are introduced and the mechanisms involved in reshaping of the bovine preimmune antibody repertoire are presented.

4. REVIEW OF THE LITERATURE

4.1. Adaptive immunity provides fine-tuned protection against pathogens

Our immune system protects our bodies from constant attacks by harmful micro-organisms called pathogens, such as bacteria, fungi and viruses. The immune system is conventionally categorized into two subsystems: the innate and the adaptive immunity systems. Innate immunity is the first line of defence and the response initiates within 96 hours of the pathogen encounter, reviewed by Janeway and Medzhitov (2002). The immediate innate response depends on antimicrobial enzymes and peptides, such as lysozyme that can be found in tears and saliva (Fleming, 1922). In the event of a bacterial pathogen invading a host, the innate system of that host responds within minutes of the bacterial invasion by digesting the bacterial cell walls. If the innate system does not kill the invader outright, it attenuates the effect of the bacteria. In addition to these soluble molecules, the innate response relies on germline-encoded pattern-recognition receptors that are able to identify typical pathogen-associated molecular patterns, reviewed by Medzhitov and Janeway (1997).

However, over time many pathogens have evolved mechanisms for circumventing a host's innate immunity defence, reviewed by Hornef et al. (2002). When the innate immune response fails to neutralize the infectious agent, then adaptive immune responses will be induced. In contrast to the innate immunity, the molecular receptors that recognize pathogens in adaptive immune responses are not "ready-made" by the host's genome. They are randomly assembled together from germline encoded segments (Tonegawa, 1983). Adaptive immunity is also called "acquired immunity" due to its empirical nature of first acquiring information for subsequent pathogen recognition, which occurs throughout the whole lifetime of the host. Pathogens bear specific antigens on their surfaces that are recognized by the molecular receptors called antibodies, which are produced by the adaptive immune system of the host. Antibody genes are constantly modified to enable better discrimination recognition between different pathogens, *e.g.* see Eisen and Siskind (1964) and Jerne (1955). Those antibodies with the highest affinity to the pathogen are positively selected and they subsequently proliferate. The adaptive immune responses can take days to develop due to the level of affinity the selected antibodies must attain. The adaptive response is often followed by a long-lasting immunological memory that protects the body in future from the antigen encounters, reviewed by Gourley et al. (2004). This forms the basis for creating effective vaccines.

Adaptive immune responses are mediated by lymphocytes. Lymphocytes are a class of white blood cells that are derived from the common lymphoid progenitor cells.

Antigen receptors on the cell-surface are a characteristic of lymphocytes. There are two main types of lymphocytes, B lymphocytes (B cells) and T lymphocytes (T cells). T cells are responsible for the cell-mediated immunity and B cells mediate the humoral (antibody-mediated) immunity. Antigen receptors on B cells belong to the protein family of immunoglobulins and every single B cell produces unique immunoglobulins of a single specificity for an antigen. Therefore, the immune system consists of a vast range of different B cell clones, reviewed by Janeway (1992). The proliferation of self-tolerant lymphocytes in response to specific antigen is also known as the clonal selection theory which is one of the most important principles in adaptive immunity (Jerne, 1955; Burnet, 1976). The proliferation gives rise to clones of identical lymphocytes with the same antigen specificity and capacity to combat against foreign antigens present at the time.

Lymphocytes originate from lymphoid progenitor cells in the bone marrow. In mammals their development takes place in the primary lymphoid organs, the thymus and bone marrow, reviewed by Cooper (2002). Subsequently, the mature naïve lymphocytes, which have not encountered their specific antigen yet, enter the bloodstream and circulate to the secondary lymphoid organs. The maintenance of these mature naïve lymphocytes and the induction of the immune responses takes place in these secondary organs, which comprise the lymph nodes, the spleen and several mucosal lymphoid tissues. The division between primary and secondary tissues however, is not straightforward and there are species dependent differences (Cooper et al., 1966b; Glick, 1970). This is topic is further dealt with under heading 4.5.

Upon recognition of a specific antigen, the B cell enlarges to a lymphoblast, then it proliferates and differentiates into a plasma cell that produces a secreted form of the receptor called an antibody (Batista and Neuberger, 1998). In addition to antigen stimulation, B cell activation usually requires the presence of helper T cells. In conclusion, the main task of B cells is to secrete a variety of different antibodies that destroy or inhibit the spread of harmful pathogens. Antibodies bound to antigens might neutralize the pathogen or they can induce phagocytosis of the pathogen by a process called opsonization. The antibody coat can also activate an elaborate complement system cascade that plays an important role in activating the innate immune system, reviewed by Tomlinson (1993). As a result, pathogens are destroyed by opsonization, or by phagocytosis or by lysis of the cell membrane. The functions that determine the operative action of an antibody are called effector mechanisms.

4.2. Immunoglobulins are encoded in several gene segments

Immunoglobulins are Y-shaped molecules that consist of two heavy and two light polypeptide chains bound together by disulfide bonds (Figure 1). There are two types

of light chains, kappa (κ) and lambda (λ). Each immunoglobulin is either κ or λ type but never mixed (Putnam et al., 1967; Titani et al., 1967; Putnam, 1969).

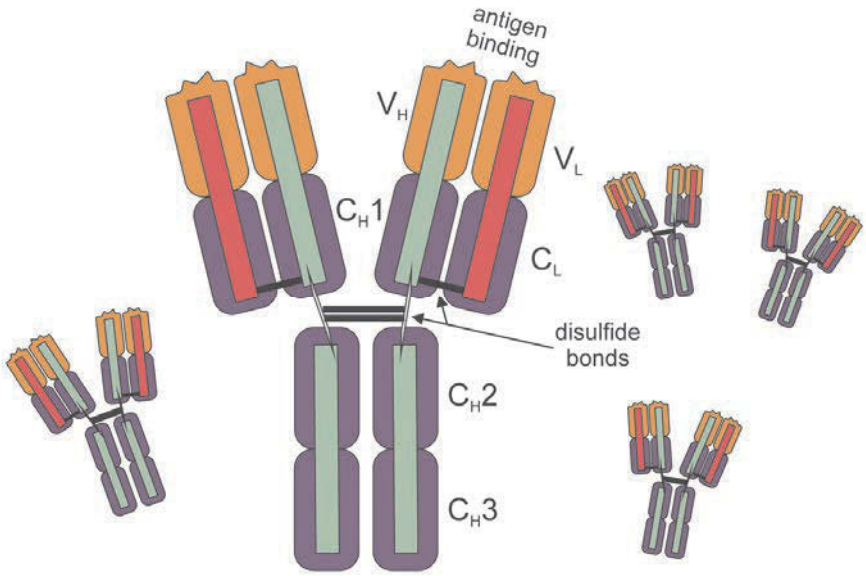

Figure 1. Immunoglobulin molecules consist of heavy and light chains. The light green rectangles illustrate the heavy chains and the fuchsia rectangles indicate the light chains. The orange part of heavy and light chains together forms the variable region (V_H and V_L), which is involved in antigen binding. The violet area illustrates the constant region of the heavy chain (C_H1-3) and the light chain (C_L), which performs the effector functions of the molecule.

Schematically, the heavy chains consist of the variable (V_H) and the constant (C_H1-3) region. The V_H regions consist of the variable (V), diversity (D) and joining (J) segments. The light chain V_L region has the same structure but excludes the D segment. Light chains have only one constant (C_L) region (Figure 1).

Variable regions on both heavy and light chains together create the antigen-binding sites (Hozumi and Tonegawa, 1976). These variable regions are divided into highly variable complementarity determining regions (CDR) and the less variable framework regions (FR) (Figure 2). There are three CDRs, namely CDR1, CDR2 and CDR3 that alternate 4 framework regions (FR1-FR4). The most variable CDR is the CDR3, which contributes the most to the actual antigen binding. For a more complete description see the review by Davies et al. (1990). There are several systems according which CDR/FR boundaries are defined. The three most common numbering schemes used are Kabat (Kabat et al., 1976), Chothia (Chothia et al., 1989) and international immunogenetics information system (IMGT) (Giudicelli et al., 1997). The Kabat system was developed by aligning a small number of antibodies based on the presumption that CDRs include the most variable parts of the

antibodies. The CDR boundaries were determined by these alignments. However, structural information was still unavailable when the Kabat numbering was developed. The Chothia numbering system is similar to the Kabat system except that it is based on structural information. IMGT is the most recent numbering system to be developed and it is based on nucleotide data and also structural information.

During an immune response the CDR regions are actively mutated in order to increase the affinity of an antibody towards a targeted antigenic site of a pathogen. Furthermore, non-silent mutations accumulate in CDR regions more often than in FR regions even in unselected genes (Cowell et al., 1999). FRs serve as a structural framework to the V region and mutations in these regions often destroy the 3D-structure of the protein (Foote and Winter, 1992).

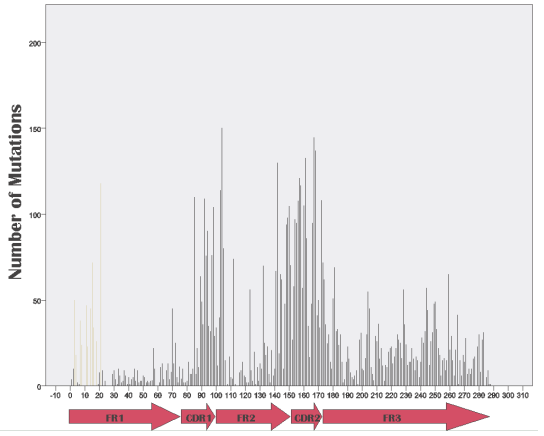

Figure 2. Mutation analysis of *IGHV* sequences. Typical pattern of hypervariability in CDRs. FRs and CDRs according to the IMGT system.

The effector mechanisms, such as distribution and functional activities of antibodies are mediated by the heavy chain constant region, reviewed by Flajnik and Kasahara (2010). Placental mammals have five *IGHC* gene isotypes, namely *IGHA* (α), *IGHD* (δ), *IGHE* (ε), *IGHG* (γ) and *IGHM* (μ). The heavy-chain type is in parenthesis. The isotype defines the mechanism by which a pathogen is eliminated, whether it is neutralized, opsonized or the whole complement cascade is activated against it. Antibodies of a different isotype are also selectively distributed within the body. Antibodies can be monomeric, dimeric or form polymeric complexes depending on the isotype. IGHM is the first immunoglobulin to appear on the B cell surface and to be secreted. It forms pentameric units and due to its large size it is primarily found in the blood. IGHG is the most common isotype found in blood. IGHA resides on epithelial surfaces such as those of the respiratory tract and the salivary glands where it is responsible for the first line of defence in which it neutralizes pathogens. IGHE is present in the monomeric form bound to granule rich mast-cells that are involved in allergenic reactions and parasite defence. Circulating IGHD is not known to

contribute to major effector functions. IGHD is not as well-known as the other isotypes and its role still remains unknown as reviewed by Schroeder and Cavacini (2010).

4.3. Immunoglobulin diversity is generated by various mechanisms

After the B cell has encountered an antigen and differentiated into a plasma cell, each clone produces only one specific type of antibody. Adaptive immune responses, mediated by specific immunoglobulins, form the basis of protecting the body against extracellular pathogens. Thus, there has to be an enormous variety of different clones of B lymphocytes in the mature immune system.

The size of the human antibody repertoire is estimated roughly to be 10^{11} (Glanville et al., 2009). It follows that all of these antibodies cannot be encoded by the human genome, which consist of 20 000-25 000 genes according to the Human Genome Project (National Human Genome Research Institute, 01.02.2014). The antibodies overcome this problem by being encoded by small gene segments, which are brought together in a tightly regulated somatic recombination process (Tonegawa, 1983; Alt et al., 1986). The recombined genes can be further modified by the addition or deletion of individual nucleotides or even multiple nucleotide segments: actions which thus mutate the genes, reviewed by Di Noia and Neuberger (2007) and Motea and Berdis (2010). The somatic recombination process is not only restricted to mammals, it occurs in all jawed vertebrates including cartilaginous fishes and reptiles, reviewed by Hsu (2009).

4.3.1. V(D)J recombination joins immunoglobulin gene segments together

Somatic or V(D)J recombination is a complex process by which different gene segments are correctly assembled together by various enzymes (Tonegawa, 1983). In this way, hundreds of thousands of immunoglobulins can be produced from the original genomic repertoire of V, D and J segments. V, D and J segments are organized in three different genetic loci (λ-chain locus, κ-chain locus and heavy-chain locus) in the genome (Figure 3). These loci consist of different numbers of functional and pseudogene segments *e.g.* see Matsuda et al. (1998).

Figure 3. Germline organization of the immunoglobulin heavy-chain locus. The number of variable, diversity and joining gene segments varies between species. The C region is joined to the VDJ segments by splicing the intron between J_H and C_H genes.

The organization and numbers of V(D)J segments varies greatly between species (Table 1). The V gene segment repertoire is the most variable, and it usually contains tens of V genes. The V segment is followed by multiple D segments and a few J segments.

Table 1. Immunoglobulin diversity between different species, approximate number of functional genes.

Species	V_H	D_H	J_H	V_L	J_L	V_K	J_K
Human[a]	40	25	6	29-33	5	30-35	5
Mouse[b]	101	13	4	3	3	93	5
Cow[c]	10-20	14	3	25	2	8	2
Pig[d]	~20	2	1	9	2	9	1
Chicken[e]	1	15	1	1	1	-	-

References:

a) (Matsuda et al., 1998), (Lefranc, 2001a), (Lefranc, 2001b)

b) (Chevillard et al., 2002), (Sakano et al., 1981; Kurosawa and Tonegawa, 1982; Feeney and Riblet, 1993), (Solin and Kaartinen, 1992), (Eisen and Reilly, 1985) (Martinez-Jean et al., 2001; Thiebe et al., 1999)

c) I, III, (Zhao et al., 2006), (Koti et al., 2008, 2010), (Zhao et al., 2003, Hosseini et al., 2004), (Ekman et al., 2009)

d) (Sun et al., 1994), (Sun and Butler, 1996), (Butler et al., 1996), (Butler et al., 2011), (Schwartz et al., 2012b), (Schwartz et al., 2012a)

e) (Reynaud et al., 1985), (Reynaud et al., 1987), (Reynaud et al., 1991)

In order for the gene segments to rearrange, they must be flanked by a recombination signal sequence (RSS). A functional RSS consists of heptamer and nonamer motives and an intervening spacer of 12 or 23 nucleotides (Sakano et al., 1979; Tonegawa, 1983) (Figure 4).

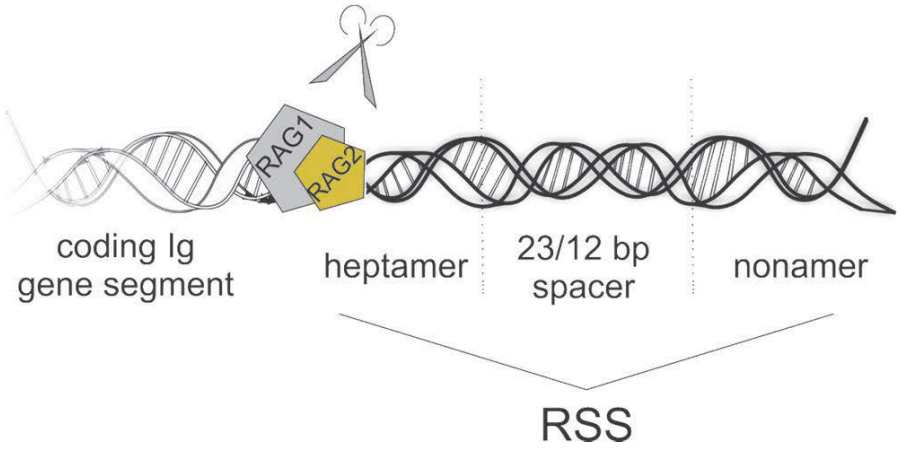

Figure 4. Germline configuration of the structure of recombination signal sequence (RSS). RSS consists of well conserved heptamer and nonamer sequences that are separated by 12 or 23 bp spacer sequences. Immunoglobulin gene rearrangement can proceed among gene segments having 12 bp and 23 bp spacers, with some exceptions. This is called the 12/23 rule. RAG1/2 complex binds the 5' end of the heptamer.

The essential enzymes required to catalyze this recombination process are expressed by the recombination activating genes 1 and 2 (*RAG1/ RAG2*) (Schatz et al., 1989; Oettinger et al., 1990). These well-conserved enzymes induce recombination only when expressed together. The RAG1/2 complex introduces a nick between the coding region and the RSS in order to create a free 3′-hydroxyl group. This 3′-OH group then reacts with the opposing phosphodiester bond of the antiparallel 5' strand resulting in a hairpin structure on the coding end. Furthermore, the signal end remains associated with the RAG1/2 complex. Here, the RAG1/2 complex is assisted by high mobility group proteins of the HMG-box family. The coding and signal ends are processed in a different manner (Figure 5). Coding ends are opened and joined together through the non-homologous end-joining pathway (Daza et al., 1996). The DNA ends are held together by the Ku heterodimer, which consists of Ku70 and Ku80. This complex recruits DNA-dependent protein kinase (DNA-PK) (Lees-Miller and Anderson, 1991). DNA-PK forms a protein complex with Artemis, which usually possesses 5′→3′ exonuclease activity. However, phosphorylation of Artemis by DNA-PK transforms the exonuclease activity into endonuclease activity, which can open the hairpin by single-strand cleavage (Ma et al., 2002). Asymmetric hairpin opening can result in palindromic (P) nucleotides that were initially complementary to each other in the double-stranded DNA (Lafaille et al., 1989; Gauss and Lieber, 1996). Further diversification is produced by terminal deoxynucleotidyl transferase (TdT), which randomly incorporates non-templated (N) nucleotides into the coding ends (Desiderio et al., 1984). Eventually the coding joint is ligated by XRCC4 and DNA ligase IV complex (Li et al., 1995; Wei et al., 1995).

Processing of the signal ends is easier. They remain bound to the RAGs and eventually form a signal joint (Figure 5). It has been suggested that the formation of the signal end has a particular role in maintaining the chromosomal integrity (Agrawal and Schatz, 1997).

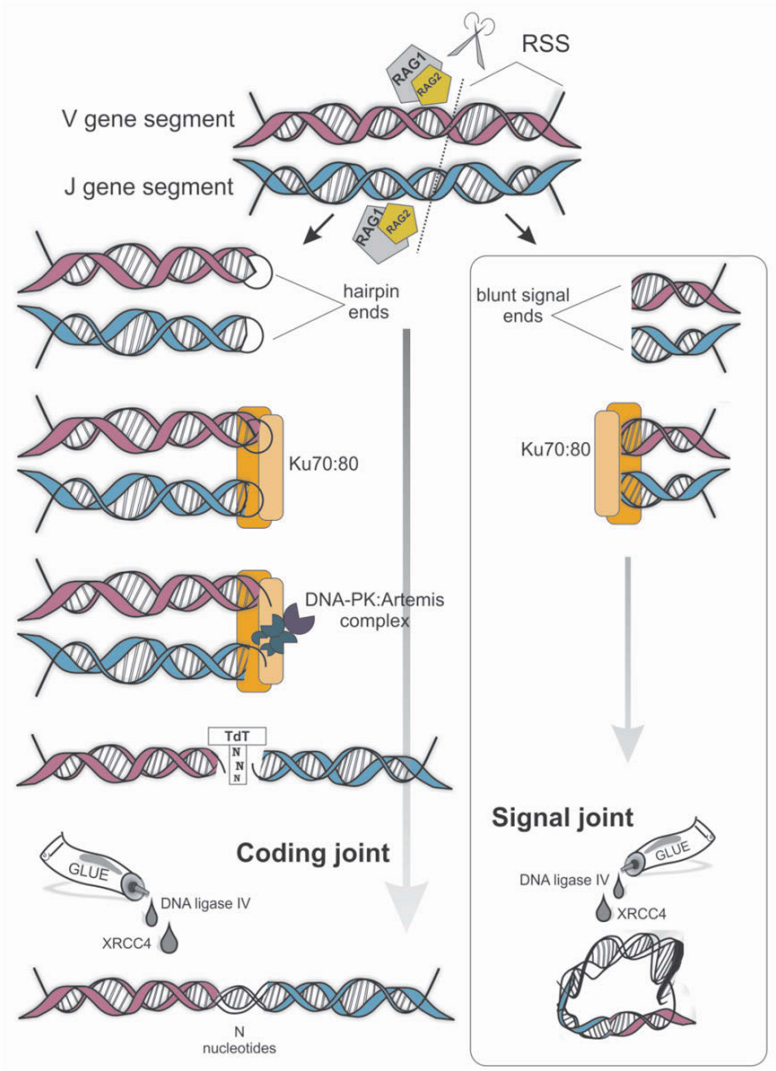

Figure 5. V(D)J recombination. The arrangement of V (purple) and J (turquoise) gene segments is shown here. Two RAG1/2 complexes bind to two RSS sequences. This brings the gene segments together and enables the recombination to occur. The endonuclease activity of RAG1/2 complex results in a hairpin-sealed coding end and a blunt signaling end. Left side: processing of the coding ends. Ku70:80 (orange boxes) binds to the hairpin that holds the two strands together. Subsequently, DNA-PK:Artemis attaches the complex and opens the hairpin at a random site, resulting in single-stranded DNA. The introduction of P nucleotides depends on the site of the cleavage (not shown). Single strand ends are then modified by random addition of N nucleotides by TdT. Coding ends are also diversified by excision of nucleotides. Finally, the sequences are ligated together by DNA ligase IV and XRCC4 (drops) to form a functional coding sequence. Right side: processing of the signal joints. The blunt ends are held together by Ku70:80 complex. Then, DNA ligase IV and XRCC4 ligates the signal ends together, that forms a precise signal joint with no further modifications.

The recombination of the immunoglobulin gene segments is not always achieved in a precise manner. Stop codons can be introduced into the open reading frame or the sequence can be mutated, which leads to non-functional rearrangements. Heavy-chain locus rearrangement initiates the B cell diversification. First, the D segment and the JH segment are joined (Early et al., 1980; Sakano et al., 1980; Alt and Baltimore, 1982). The RSS sequences flank D gene segments at both the 5' and 3' ends and it can be rearranged to the JH segments in direct or indirect manner. The majority of these recombinations are potentially functional because most DHs are translatable in all three reading frames, and some are even translated in reverse frames. The next step is the joining of a VH segment with the DJH segment (Alt et al., 1986). There is a theoretical 0.66 probability that this rearrangement is nonproductive due to incorrect assembly of the reading frames. It has been shown that the actual rearrangement frequencies depend on various factors *e.g.* sequence of the spacer (Montalbano et al., 2003). Thus, this second recombination event takes place initially on one chromosome. Should this fail, the rearrangement takes place on the other chromosome, again with a high probability of failure (Grawunder et al., 1995). Pseudogenes further increase the likelihood of nonproductive rearrangements, because some of them are capable of rearranging but not of producing a functional protein. At this point, the functionality of the heavy chain is tested by the so called surrogate light chains (Löffert et al., 1996; Galler et al., 2004). When the pre-B cell receptor that consists of heavy chains and surrogate light chains is formed the cell can proceed to the light chain rearrangement. Light chains lack the D segments so the rearrangement process is more straightforward than that of the heavy chain.

4.3.2. The combinatorial repertoire is limited in cattle and in several other domestic species

In several domestic species the number of V(D)J germline segments is lower than that found in the human and in the mouse (Table 1). Sequencing of the bovine genome in 2009 allowed a more complete characterization of the light chain reservoir in cattle, including that of the *IGLV* and *IGKV* gene segments (Bovine Genome Sequencing and Analysis Consortium 2009). The light chain gene pools were annotated from the Btau_3.1 assembly (Ekman et al., 2009). λ locus was located in chromosome 17 (BTA17) and 63 IGLV and 3 IGLJ genes were identified. The IGLV genes were grouped into eight phylogenetic subgroups and 25 of these genes were categorized as being potentially functional. The majority of the functional V genes belongs to subgroup 1 and is a characteristic of ruminants. Only two out of the 15 possible J-C pairs could form functional combinations. Furthermore, Pasman et al. (2010) characterized the 5′ end of the λ locus in BTA17 from the Btau_4.0 assembly. Those authors confirmed that the VL1-JL3-CL3 combination encodes most of the light chain repertoire even though higher combinatorial potential would exist in the germline. The characterization of bovine *IGLC* genes has revealed four constant genes (*IGLC1-4*) out of which *IGLC2* and *IGLC3* appeared to be functional (Chen et al., 2008).

The bovine κ locus was located in chromosome 11 (BTA11) with 22 *IGKV*, three *IGKJ* and one *IGKC* genes (Ekman et al., 2009). Those authors also found that the variable genes were assembled in four phylogenetic subgroups and eight genes from these subgroups were potentially functional. All except one functional *IGKV* belong to subgroup 2. Evidently, the κ locus (280 kb) is smaller in size than the λ locus although the relative size of the λ locus could not be reliably estimated from Btau_3.1 due to gaps in the assembly. The size of the locus is often indicative of the ratio by which the light chains are used. Genes in the locus that have more diversity usually dominate expression in the light chains (Table 1).

The ratio of the κ versus λ type immunoglobulins varies between species and the reason for this is still unknown. Many livestock and companion animals including poultry are highly lambda dominant with the extreme example of chicken having only one λ light chain. Previous studies reported that cattle have 91%, dogs 91%, cats 92%, horses 96% (Arun et al., 1996) and rabbits 90% (Hole et al., 1991) of λ chains, whereas pigs have a more balanced ratio of 52% of λ light chains (Arun et al., 1996). Conversely, the mouse is almost completely κ dominant, due to having only 5% λ chains (Almagro et al., 1998). Analysis conducted in the mouse indicates that the kappa/lambda ratio has already been determined in early ontogeny and does not shift during murine B cell development (McGuire and Vitetta, 1981).

The bovine's IGH locus is not as well-characterized as its light chain loci. Furthermore, the knowledge about the *IGHV* genes is very limited. The VH repertoire consists of a maximum of 20 segments as reported in studies of immunoglobulin cDNA sequences (Sinclair et al., 1997; Lopez et al., 1998). The *IGVH* genes all belong to the same family, which is, in turn, homologous to murine Q52 and human VHII families (Berens et al., 1997; Saini et al., 1997). However, Southern blotting has indicated additional subgroups, which were not detected with the expressed *IGHV* genes (Berens et al., 1997; Lopez et al., 1998).

IGHD genes have been studied but they are still incompletely characterized. No intensive, genome wide bioinformatics studies have been performed since the bovine genome sequencing in 2009 was undertaken. The most recent information is based on screening of the bovine genomic library using *DH* specific DNA probes. Until now, a total of 10 *DH* genes had been identified and classified into four families (Shojaei et al., 2003; Zhao et al., 2003, Koti et al., 2008, 2010). From a phylogenetic perspective, these bovine genes have the closest homology with rabbit and chicken *DH* genes. The length of the segments varies from 42 bp to 148 bp with several cysteine residues (Shojaei et al., 2003; Koti et al., 2010). *IGHD2* is currently acknowledged to be the longest *DH* segment and it encodes an exon of 61 amino acids. Approximately 8-10% of peripheral B cells in the bovine express the long *IGHD* variants that contribute directly to the formation of exceptionally long CDR3H regions. These long regions can act as a mechanism in compensation for the modest number of bovine VH segments. A similar finding has also been observed in

camelids but not in the more closely related sheep (Saini et al., 1999; Kaushik et al., 2002; Zhao et al., 2006). At first, the long *IGHD* genes were only detected in IGHM and IGHG isotypes but recently they have also been reported to occur in all bovine immunoglobulin isotypes (Walther et al., 2013).

The bovine JH immunoglobulin arrangement has also been characterized. The *IGHJ* locus resides in BTA21 and comprises six *JH* genes. Only two of these genes have been reported to be functional with the preferential use of one of them (Berens et al., 1997; Zhao et al., 2003; Hosseini et al., 2004). Furthermore, it has been proposed that the sixth *JH* gene would be linked to *IGHML1* assembled to BTA11, although no VH or D segments were characterized there (Hosseini et al., 2004).

The IGHC locus has been located to BTA21q23-q24 (Tobin-Janzen and Womack, 1992; Chowdhary et al., 1996). These loci are extending 150 kb range with seven functional *HC* genes, expressed at different levels. Cattle have all the characteristic mammalian isotypes (IGHM, IGHA, IGHE, IGHD and IGHG1-IGHG3) (Zhao et al., 2003). In addition to this, an IGHM-like locus (*IGHML1*, U63637) has been reported in chromosome BTA11q23 (Hayes and Petit, 1993; Hosseini et al., 2004). Sheep and goats also have an additional *IGHML1* locus but the functionality of this locus has not been validated yet (Hayes and Petit, 1993). Indeed, in 2009 Kuroiwa and co-workers claimed that cattle were the first mammalian species in which multiple fully functional IGH loci were found (Kuroiwa et al., 2009). Those authors claim was based on the discovery that both loci in cattle require inactivation as a prerequisite to producing large amounts of human polyclonal antibodies in hyperimmunized cattle. Although reliable evidence of the functionality of genes residing in both loci was shown, no attempt was made at analysing the actual localization of these loci. Transchromosomal recombination can take place but it is rather uncommon (Knight et al., 1995; Kingzette et al., 1998). A more plausible explanation would be the incorrect assembly of these loci, which Kuroiwa et al. failed to consider.

4.3.3. Terminal deoxynucleotidyl transferase creates junctional diversity

Junctional diversity, coupled with the recombination process, is created by TdT (Desiderio et al., 1984). This crucial enzyme adds N nucleotides at the VDJ junctions in a template independent manner. It uses single-stranded DNA as a substrate and it is the only canonical template independent polymerase known (Yoneda and Bollum, 1965). Junctional diversity often produces nonproductive rearrangements due to the random manner of nucleotide additions, reviewed by Motea and Berdis (2010). Nucleotides are also deleted at the junctions, which creates additional diversity.

The sequence of TdT is highly conserved between species and in addition to mammals its contribution to the antibody diversity has also been shown in cartilaginous fish, turtles (Turchin and Hsu, 1996) and birds (Sharma, 1997).

The template independent characteristics of TdT have appeared simultaneously with the development of V(D)J recombination in mammals (Motea and Berdis, 2010).

TdT is usually expressed in the primary lymphoid organs such as the thymus and bone marrow. In addition to immunoglobulin genes, TdT also diversifies T cell receptor genes. TdT belongs to the X-family of DNA polymerases with the closest homology to the Polμ but is shares only 42% of amino acid identity (Uchiyama et al., 2009).

Initially TdT was purified from the calf thymus (Chang and Bollum, 1971) where it is expressed in large amounts. In cattle, three alternatively spliced variants exist that are highly homologous in sequence as in humans (Figure 6). Both cattle and human species have one short isoform (TdTS) and two long isoforms (TdTL1 and TdTL2) (Takahara et al., 1994; Thai and Kearney, 2004). TdTL1 and TdTL2 contain an extra exon VII and XII respectively. These exons are directly attached to exon VI and XIII without an intron in between them. The short isoform does not contain these exons. During normal human B cell development hTdTS and hTdTL2 are expressed in fetal life (Thai and Kearney, 2004). Almost all human heavy chains contain N additions in contrast to light chains where only occasional additions are seen (Schroeder et al., 1995). *In vitro* studies of the human TdTLs have shown that the long isoforms possess a 3′→5′ exonuclease activity. In fact, the long isoforms may contribute to the repertoire diversity by catalyzing nucleotide deletions in the VDJ junctions (Thai and Kearney, 2005).

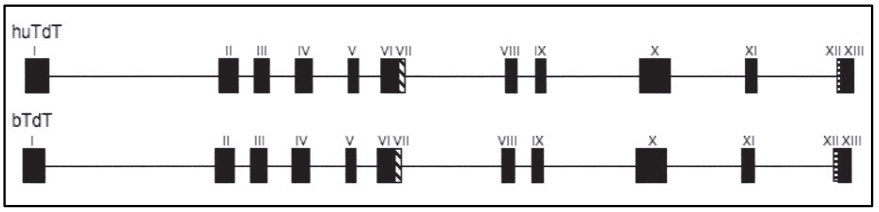

Figure 6. Organization of the respective human and bovine *TdT* gene. The *TdTL2* isoform contains the extra exon VII of 18 (bovine) and 17 (human) amino acids insertion with a homology of 77% between these two species. *TdTL1* contains an extra exon XII which is encoded by a 9 amino acids that shares the similarity of 67% between bovine and human sequences. Figure according to Thai and Kearney; *The Journal of Immunology*, vol. 173, pp. 4009-4019, 2004. *Copyright 2004. The American Association of Immunologists, Inc*.

In the mouse there are two isoforms of the TdT mRNA. *TdTS* encodes 509 amino acids. Murine *TdTS* has some similarities to human and bovine sequences. The murine *TdTL* encodes 529 amino acids and has a 20 amino-acid insertion near the C terminus that is different from the long isoforms of cattle and human (Doyen et al., 1993) (Figure 7). In the mouse, TdT is expressed only after birth, thus no N additions are found in mouse fetuses. In those cases where the adult immune responses are mediated by B cells of fetal or neonatal origin, such as phosphorylcholine, the lower immunoglobulin diversity in fetuses is necessary for

the adequate repertoire diversity in adults. This was shown in *TdTS* transgenic mice (*TdTS* transgene on the $TdT^{-/-}$ background) where N nucleotides were added during fetal life. Premature TdT activity resulted in a defective B cell repertoire and a lack of some important antibodies such as anti-phosphorylcholine. These antibodies required germline-encoded specificity, which was destroyed by premature N additions (Benedict and Kearney, 1999).

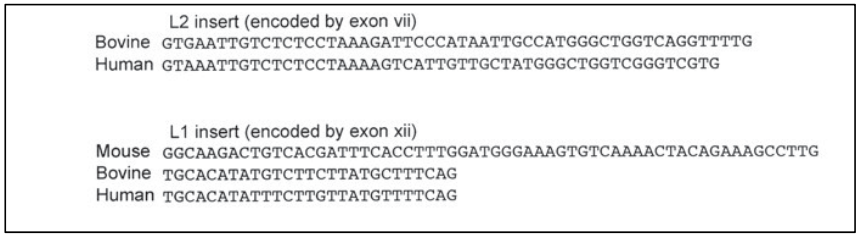

Figure 7. L2 and L1 inserts in the bovine, human and mouse genomes. Figure according to Thai and Kearney; *The Journal of Immunology*, vol. 173, pp. 4009-4019, 2004. *Copyright 2004. The American Association of Immunologists, Inc.*

Controversial data about the function of murine TdTL splice variants exist (Bentolila et al., 1995; Benedict et al., 2000, 2001). At first, it was reported that TdTL was incapable of adding N nucleotides to immunoglobulins due to cytoplasmic localization (Bentolila et al., 1995). This proposal was later challenged by Benedict and co-workers who demonstrated that TdTL was localized in the nuclei of lymphoid cells of transgenic mice. In addition to this, TdTL was visible in the cytoplasm (Benedict et al., 2000). The role of TdTL functioning as an exonuclease was challenged shortly afterwards by Doyen et al. (2004). Those authors did not find any role for TdTL in the V(D)J recombination but suggested that because TdTL is an evolutionary remnant it is not involved in coding or trimming ends. They suggested it might instead recruit an exonuclease already present in the CHO cells that the other group was using. Moreover, exonuclease activity was not seen in NIH 3T3- or in COS-cells (Doyen et al., 2004). Therefore, the actual function of TdT splice variants remains unclear and it requires further investigation.

4.3.4. Post-recombinatorial mechanisms complete the preimmune repertoire

Somatic recombination is complemented by junctional diversity and produces the initial preimmune repertoire. However, secondary diversification is also needed to improve the immunoglobulins' capability of recognition and binding of foreign antigens. This enhanced affinity is accomplished by three post-recombinatorial mechanisms; SHM, gene conversion and class switch recombination (CSR). SHM introduces point mutations to the immunoglobulin V regions, which results in higher affinity antibodies. Gene conversion relies on V region pseudoges, which function as donors to the rearranged V genes. Parts of these pseudogenes are introduced to the V genes and increase the variability (Muramatsu et al., 2000, Flajnik, 2002; Larijani

and Martin, 2012). In CSR, the constant part (IGHμ) of the rearranged immunoglobulin can upon antigen stimulation be exchanged for an alternative *IGHA, IGHE* or *IGHG* heavy chain gene without modifying the specificity for the antigen. Thus, in certain B cell clones the progeny's antigen-specificity always remains the same although the effector function may change depending on the isotype, reviewed by Chaudhuri and Alt (2004).

Traditionally these mechanisms have been considered to be activated upon encountering an external antigen. In various domestic species, however, the combinatorial repertoire is insufficient and the preimmune repertoire is diversified by post-recombinatorial mechanisms directly after somatic rearrangement, independent of encounters with external antigens (Knight and Crane, 1994; Reynaud et al., 1987, 1995). Nonetheless, these mechanisms are crucial for antibody repertoire diversification and are dependent on the same enzyme, activation-induced cytidine deaminase (AID) (Muramatsu et al., 2000; Arakawa et al., 2002).

4.4. Activation-induced cytidine deaminase: a vital mutator enzyme

AID was originally described as an RNA editing enzyme based on its similarity to the RNA editing enzyme APOBEC1 (Muramatsu et al., 2000). Over the course of a decade a lot of effort was made to characterize the function of AID. Nowadays, the concept that AID attacks single-stranded DNA is widely accepted and has been reviewed by Larijani and Martin (2012). Maul and co-workers have recently presented direct evidence for the DNA model by showing the physical existence of the AID mediated reaction intermediates in immunoglobulin genes obtained from B cells *in vitro* (Maul et al., 2011).

AID converts cytosines (C) to uracils (U) and its function is restricted to single-stranded DNA that is exposed during transcription. The resulting U:G mispairing recruits uracil DNA glycosylase (UNG), which removes the uracil residue thereby creating an abasic site on the DNA. Abasic sites can be recognized by base excision repair (BER) pathway that finally inserts a C residue opposite to the G, reviewed by Wilson and Bohr (2007). The uracil can also activate mismatch repair (MMR) pathways even before the UNG is activated. The MMR pathway repairs mismatches especially during recombination (Jiricny, 2006). The abasic site can also be used as a template in the following round of DNA replication. These alternative mechanisms are summarized in Figure 8.

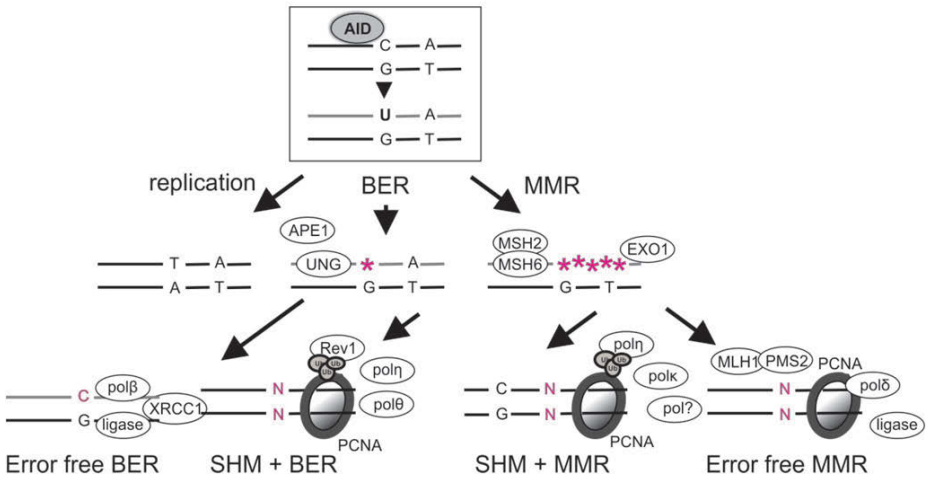

Figure 8. Processing of the AID-induced uracils. In the first step, AID mutates cytosine to uracil which is subsequently used as a template in replication (identified as thymine) when it is not recognized by either the BER or the MMR repair pathway. In the BER pathway, uracils are recognized by UNG, which then removes the uracil bases thus forming abasic sites (denoted by magenta stars). Ape1 excises the ribose, which forms a single-strand nick. This can be faithfully repaired by the BER pathway in the presence of the DNA repair proteins, XRCC1 and polβ. In the SHM, abasic sites are bypassed by polymerase Rev1, which is activated by monoubiquitination of the PCNA clamp. Rev1 also recruits other sloppy polymerases. In MMR pathways, MSH2-MSH6 identifies mismatches, which are excised by EXO1 and forms a larger lesion (denoted by magenta stars). This can be faithfully repaired by MutLα (MLH1-PMS2) complex, which recruits the PCNA clamp that is associated with high fidelity polymerases. In the SHM, monoubiquitinated PCNA recruits low fidelity polymerases and produces mutations. (Reprinted with minor modifications from Seminars in Immunology, 24 (2012), Saribasak & Gearhart, Does DNA repair occur during somatic hypermutation? 287-292, Copyright (2014), with permission from Elsevier).

When the uracil excisions take place in adjacent strands in the switch region (S_μ) the repair process leads to CSR. S_μ regions that are repetitive sequences located in the intron between the JH segment and the CH1 (Maul et al., 2011; Murphy, 2012).

4.4.1. Processing the AID-induced mutations is complex

The SHM targeting of AID takes place within a distance of approximately 2 kb downstream of the immunoglobulin promoter (Lebecque and Gearhart, 1990). Both, the coding and noncoding strand undergo SHM (Milstein et al., 1998). The mutation rate of the SHM is about a million-fold compared that of spontaneous mutations (Peled et al., 2008). Mutations induce error repair pathways, which in other contexts are error free. However, in SHM the antibody diversity is greatly increased by mutations so, how do these error free pathways become so inattentive?

In normal circumstances the BER pathway recruits polβ, which is a high fidelity polymerase and the reparation of the mutation is error free (Wilson and Bohr, 2007). Instead, error prone polymerases are recruited for the AID induced mutations in B

cells. This deviation is suggested to be due to monoubiquitinylation of proliferating cell nuclear antigen (PCNA), which recruits the error prone polymerases such as Rev1. An abasic site created by APE1 can also have a high affinity for the polη, which in turn, creates A:T mutations near to the single strand nick. Polη is dependent on PCNA monoubiquitinylation reaction, as are the other error prone polymerase (Roa et al., 2008).

In non-AID induced MMR the uracil is recognized by MutS protein homolog 2-6 (MSH2-MSH6) heterodimer, which recruits MutL homolog 1/Postmeiotic Segregation increased 2 heterodimer complex (MLH1/PMS2) (Figure 8). The result is a single-strand nick in close proximity to the mismatch. Exonuclease1 (Exo1) then removes the mismatch and also the adjacent bases, thereby creating a larger lesion. High fidelity polδ, which is attached to PCNA complex, fills in the gap (Saribasak and Gearhart, 2012). It has been suggested that the MMR pathway becomes mutagenic only in response to AID-induced mutations but repairs other mutations error-free (Green et al., 2011). This was tested in *MSH2*$^{-/-}$ *BB*$^{+}$ (Big Blue transgenic mouse strain, which lacks the eukaryotic transcription promoters) mice that were deficient in the MMR repair pathway components and also lacking the transcriptional promoters (the lack of which inhibits AID-induced mutations). Those authors showed that MMR machinery is necessary in the germinal center B cells when they compared the mutation numbers, which was 10-fold higher in the transgenic mice than in the WT mice after an immunization. They also proposed that MMR machinery normally repairs non-AID induced mutations even in the germinal center B cells.

The mechanism, of how this conversion from error free to error prone is performed, is still obscure. Some studies found evidence that the UNG from the BER pathway intrudes the MMR pathway into the Exo1 generated path (Frieder et al., 2009; Krijger et al., 2009). AID induces mutations in both leading and lagging strands, as a consequence the MMR controlled Exo1 excision can also take place in the non-mutated strand. This leaves the U residue uncovered in single-stranded DNA, which is then removed by UNG, thus leaving the abasic site in the DNA. Repair of the lesion is initiated by high fidelity polymerases from the MMR pathway, although they are unable to finish the task due to the abasic site. Low fidelity polymerases such as polη or polκ are recruited as they are able to bypass these sites and fill in the gaps in the DNA, reviewed by Saribasak and Gearhart (2012) (Figure 8).

In conclusion, the processing of AID mediated deamination is a very complicated combination of different pathways and their respective proteins. Regardless of the repair mechanism, the result is usually a mutation. After sequential modifications, it is often impossible to determine, which repair mechanism was used. Mutations can be transitions, whereby a purine is changed into other purine or pyrimidine into other pyrimidine or transversions whereby a purine is changed into pyrimidine or *vice versa*. Moreover, additional deletions or insertions near the deaminated site can

occur. Thus B cells have to find a balance between the faithful and the sloppy mutation repair systems. Only in this way can the B cell continue living without accumulating too many mutations in its genome, while being able to produce highly variable antibodies that can combat all possible pathogens.

4.4.2. AID-induced mutations are not haphazard

The targeting of AID at the correct site is crucial for the production of high affinity antibodies by SHM (Peled et al., 2008). AID induced-mutations are preferentially targeted in the CDRs rather than in FRs, the latter of which function as scaffolds to the antibody. Moreover, the mutations in the CDRs are more frequently replacement than silent mutations. Mutations in the FRs have to be silent otherwise they will be selected against. The preferential targeting in the CDRs is not only result of the selection process. This was shown by analysing the mutation pattern of the nonproductively rearranged human *VH* genes. Also in these unselected rearrangements the mutational load was significantly higher in the CDRs than in the FRs, which suggests that mutations are actually targeted at the CDRs (Dörner et al., 1998; Cowell et al., 1999).

DNA sequences surrounding targeted cytosine base, impacts the deamination efficiency (Yu et al., 2004). Within the CDRs and FRs there are 3-4 base pair motives that act as hotspots for AID. However, there is contradictory information about the exact sequence of these hotspots. RGYW/WRCY (R=A/G, Y=C/T, W=A/T) and TAA/TTA were the first motives characteristic of the SHM, which were recognized (Rogozin and Kolchanov, 1992). The first motive was nominated as a target for AID (Dörner et al., 1998) and the second motive, which was later defined as WA/TW motive, is a plausible target for polη (Rogozin et al., 2001). Since then, different motive candidates for AID have also emerged, such as the highly similar WRC or WRCr with the lowercase r having a smaller influence on activity (Pham et al., 2003; Yu et al., 2004). Switch regions (S_{μ}) are especially rich in these motives (Hackney et al., 2009).

4.4.3. Regulation mechanisms of AID are incompletely known

AID is a powerful physiological mutator that can cause chromosomal translocations, thus its exact targeting is crucial. AID activity is not solely restricted to the immunoglobulin locus in the B cells (Yamane et al., 2011). Although AID preferentially binds to the immunoglobulin locus, the association with 5910 genes outside the IGH locus in *in vitro* activated B cells was also shown in a study of Yamane and co-workers that used genome wide chromatin immunoprecipitation coupled with deep sequencing (ChIP-seq). In the IGH loci, the replication protein A was shown to be an important cofactor for AID. It was proposed, that in the switch regions AID can be targeted at the transcription stalling sites of RNA polymerase II via its stalling factor Spt5 (Pavri et al., 2010; Yamane et al., 2011). However, the

targeting of AID has been investigated mostly in CSR. These results may also apply to SHM but further investigation is required to verify this.

Off-target mutations are corrected more efficiently by the abovementioned repair pathways, which function in an error-free fashion. Consequently, if the reparation process fails, then the off-target mutations of AID can cause irreversible changes in the genome that can lead to severe malignancies such as B cell lymphomas (Nussenzweig and Nussenzweig, 2010) and other cancers (Matsumoto et al., 2007; Liu et al., 2008; Mahowald et al., 2008; Klemm et al., 2009; Lin et al., 2009).

New roles for AID have recently been proposed. This deaminase has been suggested to function as DNA demethylase by catalyzing the deamination of methylated cytosines in pluripotency genes. This is possibly followed by DNA repair, which results in the formation of nonmethylated cytosines (Bhutani et al., 2010, 2011; Popp et al., 2010). Recent studies indicate that AID may also be involved in the epigenetic regulation of erythroid differentiation (Niku et al., unpublished data).

4.5. Comparative immunology has revealed various mechanisms of creating antibody diversity

Different species use a variety of strategies to generate a functional immunoglobulin repertoire. The majority of the current knowledge about diversification mechanisms is based on studies in mice and humans. These species have a large number of V, D and J gene segments. For example, humans can create as many as 1.7×10^4 outcomes only through the combination of the immunoglobulin gene segments. In mice and in humans, the B cells are initially differentiated in the fetal liver and after that in bone marrow. This is a continuous process in adults of both species (Alt et al., 1986; Nuñez et al., 1996).

In several species including the chicken, sheep, cattle, rabbit, pig and horse, somatic recombination yields a relatively restricted pool of antibodies, due to the limited germline repertoire of the different gene segments (Table 1). However, these species use additional mechanisms for the expansion of their preimmune repertoires. According to the currently prevailing paradigm, B cell proliferation and differentiation takes place in gut-associated lymphoid tissues (GALT). In birds, this lymphoid organ is called *bursa of Fabricius* (Glick et al., 1956) whereas in pigs, sheep and cattle this structure is considered to be the ileal Peyer's patch (IPP). In contrast, rabbits are thought to utilize an analogous lymphoid organ the *sacculus rotundus* (Cooper et al., 1968) and the appendix as the primary lymphoid organ (Mage et al., 2006).

4.5.1. B as in the *Bursa of Fabricius*

Antibody producing cells were first identified in studies made in chicken in 1956 by Bruce Glick and co-workers. Removal of the *bursa of Fabricius,* which is an

epithelial outgrowth of the cloaca, resulted in failure of the chicken to produce antibodies after an immunization with *Salmonella typhimurium* (Glick et al., 1956; Cooper et al., 1966b). In the 1960s, B cells were characterized for the first time and differentiated from the thymus derived T lymphocytes. Since then, it became obvious, that the *bursa of Fabricius* functions as a primary lymphoid organ and it is crucial for avian B cell development. This lymphoid organ grows most rapidly during the first three weeks after hatching, then reaches its maximal size at eight weeks of age and then starts to involute as do most of other GALT organs in other species (Glick, 1970). The B cell was originally denoted B for "bursa derived cells", it is pure coincidence that this label also fits "bone marrow" (Cooper et al., 1965, 1966b).

The chicken is an extreme example of generating the preimmune antibody repertoire by using only one functional V_L segment that rearranges with a unique J_L segment. The chicken does not have two types of light chains as other species do. In addition to this, there is only minimal junctional diversity between the segments due to the absence of TdT (Reynaud et al., 1987). The heavy chain locus contributes to the diversity with one V_H segment and one J_H segment and about 15 functional D_H segments. Junctional diversity is only created by P nucleotides, as in the light chain (Reynaud et al., 1991).

Additional diversity in the chicken is generated by AID mediated gene conversion. There are families of pseudogenes ΨV_L and ΨV_H found upstream of both light and heavy chain loci. These pseudogenes lack the functional RSS signals, thus they are unable to recombine but they can donate sequences to the functional VJ_L and VDJ_H segments (Reynaud et al., 1987, 1989). This is followed by a template repair whereby pseudogene components are inserted into the nicked DNA, based on homology. In this way the homologous but nonidentical genetic material is exchanged intrachromosomally. Furthermore, the diversity of chicken immunoglobulins is sustained within these pseudogene families (Ratcliffe, 2006).

4.5.2. Preimmune repertoire diversification takes place in the GALT in cattle and other domestic species

After the discovery of the avian *bursa of Fabricius,* intensive studies began in order to find the mammalian equivalent for this important organ. In the 1960s attention was given to rabbit appendix as a B cell producing organ (Archer et al., 1963; Cooper et al., 1966a) and also to Peyer's patches (PP). Originally the PPs were described by Johan Conrad Peyer in 1677 as a mucus secreting accumulation of glands. Later on, Peyer's patch was proposed to be composed of lymphoid tissue and was considered to be sites of intense lymphoid division. The reactive germinal centers in the lymph nodes and the PP follicles were found to be similar.

Peyer's patch is composed of aggregations of lymphoid follicles found in the wall of the small intestine. According to the localization PP is divided into the jejunal (JPP)

and ileal PP (IPP) (Figure 9). In cattle a single IPP extends for up to two meters and differs from the JPP that is composed of multiple discrete aggregations of lymphoid follicles (Griebel and Hein, 1996). Studies in cow, sheep, pig and horse have shown that the JPP persist for the whole of the animal's life whereas the IPP starts to involute soon after birth. In cattle the IPP follicles of the fetus become visible at around 200 days of gestation (Carlens, 1928).

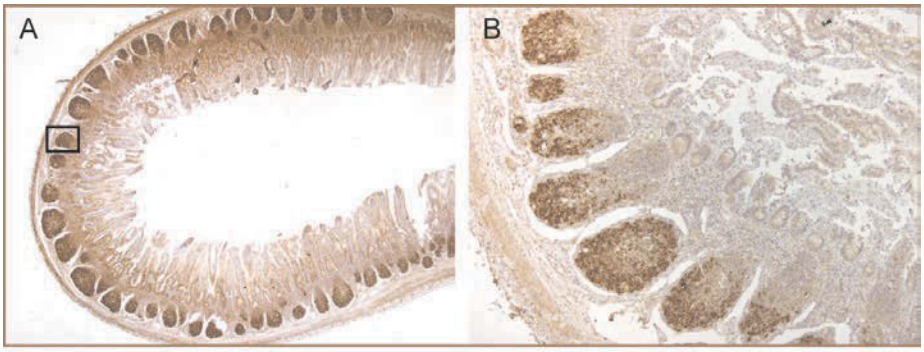

Figure 9. The ileal Peyer's patch follicles. A) A cross-section of the ileum of a 238-gestation days (gd) fetus, showing a large number of single follicles (brown color, one follicle indicated by the black box). B) Higher magnification of the follicles.

Regardless of the intensive studies, fetal liver and bone marrow appeared to be the most important organs for the generation of early B cells. Fetal liver was speculated to be the "bursa-equivalent" by a study on the organ cultures of 11-day murine fetal liver that generated IgM^+ cells *in vitro* (Owen et al., 1977). However, the rudimentary PP in the mouse could not have produced the vast amounts of B cells found in the circulation after birth (Friedberg and Weissman, 1974) and that the rabbit PP accounted more for the mucosal immunology than for the B cell development. In rabbit, the PP was also found to be a rich source of cells that have the capability of becoming IGHA producing immunocytes (Craig and Cebra, 1971). More recent research provided evidence for the rabbit appendix being the mammalian bursal equivalent (Weinstein et al., 1994).

Conversely, there was evidence from sheep that indicated the mouse B cell generation model was not representative of all mammals (Cole and Morris, 1973). The ovine PP has characteristics similar to tissues that are classified as primary lymphoid organs. This was based on the observation that prenatal maturation of the PP take place independent of the external antigens encountered. The PP becomes visible at about 60-days of gestation, lymphoid follicles are present by 75-days and at day-100 of gestation lymphoid cells are actively proliferating. The growth rate of the follicles did not change at birth when external antigens were encountered. In addition to these observations, it was also pointed out that the weight of PP tissue was greater than for any other single lymphoid tissue at six weeks after birth (Reynolds and Morris, 1983). The removal of the PP from fetal lambs caused a long-

standing and severe lymphopenia (Gerber et al., 1986) similar to that found for the removal of the bursa in the chicken.

In sheep, the functions of JPP and IPP were proposed to be different (Mutwiri et al., 1999). This was assessed by using "intestinal-loop" models whereby a 5-6 cm length of jejunal or ileal parts of the ileum were surgically transected to make a blind-ended segment with an intact blood supply. The immune responsiveness of the JPP and IPP were then compared. The JPP was shown to be an efficient site for mucosal and systemic (conventionally known as lymph node and spleen mediated) immune responses. In contrast, the IPP showed no humoral mucosal immune responses after immunization. The authors suggested that the IPP has a distinct role in the development of the B cell pool.

The anatomical similarities and relationship between sheep and cattle allow general inferences to be made about the JPP and IPP of both species. The IPP is thought to be the site for B cell proliferation and diversification in fetal and neonatal cattle, reviewed by Yasuda et al. (2006) and Zhao et al. (2006). As in sheep, *de novo* B cell lymphopoiesis or somatic recombination does not occur in the bovine IPP (Reynolds, 1997; Ekman et al., 2010). In cattle, the pre-B cell markers were shown to exist in fetal bone marrow and lymph nodes, which indicates that B lymphopoiesis occurs in those sites (Ekman et al., 2010, 2012). Cells deriving from these tissues subsequently colonize the IPP, only a few B cells form one follicle (Niku et al., 2002). This small oligoclonal founder population then proliferates and possibly expands the limited combinatorial repertoire by post-recombinatorial mechanisms (Reynolds, 1997; Yasuda et al., 2006). The IPP involutes at about three months of age in lambs hence sheep are thought to rely on the peripheral B cell population for the rest of their lives (Reynolds, 1997). The IPP also involutes in calves; therefore, the initial preimmune antibody repertoire has to be perfectly shaped by then.

5. AIMS OF THE STUDY

The present work aimed at the elucidation and the characterization of the mechanisms involved in diversification of the bovine preimmune antibody repertoire. Particular attention was given to AID and TdT enzymes, which are known to modify immunoglobulin genes. Knowledge of immunoglobulin germline genes is essential to tracing different mechanisms involved. The understanding of the roles of *IGHV* and *IGHD* genes has been especially limited.

In order to understand better how the bovine fetal preimmune repertoire is formed, we aimed at achieving the following specific objectives:

I. Characterize the bovine *IGHV* germline repertoire by using genome assemblies (Btau_4.2 and UMD_3.1), raw genomic sequencing data from trace and high-throughput sequence archives, and from this study's own targeted sequencing data.
II. Analyse the expression and functionality of AID in bovine fetal tissues prior to exposure of external antigens. Compare the mutational profile in fetal *IGHV* sequences to those in post-natal animals that had been generated by secondary, antigen-driven SHM.
III. Examine the significance of TdT-mediated junctional diversity in fetal heavy and light chains and additionally, seek new *IGHD* genes.
IV. Understand the evolution of the bovine immune system, and how it is positioned in respect to those of other species.

6. MATERIALS AND METHODS

Materials and methods used in this study are outlined here. Detailed descriptions, including primer sequences and PCR reactions, are available in the original publications (I-III).

6.1. Ethical statement

No living animals were used in this study. Bovine tissue samples were obtained from cattle that had been slaughtered in local abattoirs (Paimion Teurastomo and Orimattilan Teurastamo) and they were not euthanized for our experiments. Animals were slaughtered according to Finnish legislation and were primarily used for human consumption. Sometimes pregnant cows were also slaughtered because of different reasons (illness of the cow, economic problems encountered by the farm business etc.) and the slaughterhouses kindly donated these fetuses for scientific research purposes.

6.2. Tissues (I-III)

Tissue samples were routinely collected from the ileum, spleen, lymph nodes (*lymphonodus cervicalis superficialis, lymphonodi jejunales*), muscle, liver and bone marrow. Tissue samples for RNA and DNA extraction were snap frozen in liquid nitrogen and stored at -80°C until extraction and analysis, Tissues for immunostainings were processed for paraffin sections by fixing them in 4% phosphate-buffered paraformaldehyde (PFA) for 20-24 hours at room temperature and then embedded in paraffin. The paraffin blocks were then cut into 4 μm sections. Decalcification of the bone marrow was performed by placing a sample of bone marrow in 8% EDTA, at pH 7.0 for 1-3 days, before embedding.

6.3. DNA and RNA (I-III)

6.3.1. Extraction (I-III)

Genomic DNA was extracted from 20-40 mg of muscle tissue using the GenElute Mammalian genomic DNA miniprep kit (Sigma-Aldrich). Frozen tissues were crushed in a mortar and processed according to the manufacturer's instructions.

Total RNA was extracted from 100-300 mg of frozen tissue, which was crushed in a mortar. The tissue crush was lysed with 1 ml Eurozol (EuroClone) or Trizol (Life Technologies) and extraction was carried out according to the relevant manufacturer's instructions. RNA was precipitated with 2.5 M lithium chloride and dissolved in RNase-free water. Turbo DNase (Life Technologies) or DNase I Recombinant (Roche) was used to remove the remaining genomic DNA. Quality of RNA was assessed using RNA 6000 Nano kit and Agilent Bioanalyser (Agilent Technologies).

6.3.2. Preparation and sequencing of IGH, IGL and IGK cDNA libraries (I-III)

IGH cDNA libraries were prepared from the fetal ileum (I-III), spleen and bone marrow (II-III). An ileal cDNA library of a young calf was also prepared (II). IGL and IGK cDNA libraries were prepared from the fetal ileum and bone marrow (III). A 1-2 µg quantity of total RNA was reverse transcribed with SuperScript III First-Strand Synthesis SuperMix (Life Technologies) for all libraries. The first-strand cDNA was primed using target-specific primers (IGH ileum) or equal amounts of oligo $(dT)_{20}$ (Oligomer) and random hexamer primers (rest of the libraries). PCR was then carried out using Phusion High-Fidelity PCR MasterMix and 0.5-2 µl of cDNA as the template.

After electrophoresis and purification 20-100 ng of the PCR product was ligated into the pCR Blunt II-TOPO vector (IGL and IGK) or the pCR 4Blunt-TOPO vector (IGH) (Life Technologies). After 30 min of ligation, vectors were transformed into chemically competent TOP10 *E. coli* (Life Technologies) and grown overnight at +37° C on LB-kanamycin (pCR Blunt II-TOPO vector) or LB-ampicillin plates. A total number of 48-384 single clones were picked up for transformation, purified and sequenced (IGH ileum; Institute of Biotechnology, Helsinki, Finland, rest; GATC Biotech AG, Konstanz, Germany).

6.3.3. Cloning and sequencing of post-immunization IGHV cDNA library (II)

The post-immunization IGHV library was obtained from Dr Tony Pernthaner of the Hopkirk Research Institute, New Zealand. Briefly, total RNA was extracted from a lymph node of a *Streptococcus uberis* vaccinated calf using Trizol (Life Technologies). The first-strand cDNA synthesis was carried out using the iScript Select cDNA synthesis kit (Bio-Rad), which was primed with equal amounts of oligo (dT)20 and random primer mixes. Approximately 20 ng of cDNA was used as the template in the PCR reaction that was carried out using a combination of bovine *IGHV* primers and Platinum *Taq* DNA Polymerase (Life Technologies). DNA was then purified, digested and ligated into the JSC vector (EU 109715), which was then transformed into DH5α-*E. coli* Electromax cells (Life Technologies). A total number of 384 single clones were isolated, purified and the insert was sequenced at the Institute of Biotechnology.

6.3.4. Sequencing of germline *IGHV* genes (II-III)

Genomic DNA was extracted from muscle tissue obtained from three bovine fetuses and one calf. The *IGHV* genes that belonged to subgroup 1 were subsequently amplified using Phusion High-Fidelity PCR Master Mix. PCR products were then purified and ligated to the pCR 4Blunt-TOPO vector as described in section 6.3.2. A total of 384 single colonies were picked up, purified and sequenced by GATC Biotech.

6.3.5. RT-qPCR (II-III)

Total RNA was reverse transcribed by using either RevertAid M-MuLV Reverse transcriptase (Fermentas) or SuperScript III First-Strand Synthesis SuperMix. The first-strand cDNA was primed with oligo $(dT)_{20}$ or with equal amounts of oligo $(dT)_{20}$ and random hexamer primers according to manufacturer's instructions. Reverse transcription-qPCR was completed using the SYBRgreen technology (II, Takara) or probe based method (III, Maxima Probe qPCR Master Mix, Thermo Scientific). Stratagene Mx3005P real-time PCR system (Agilent Technologies) was used for amplification.

Results were calculated by absolute (II) or relative (III) quantification. Briefly, the absolute method involved a standard curve that was prepared from serial dilutions of plasmid templates and the *GAPDH* normalized results were compared to the standard curve. In the relative quantification, the threshold cycles were normalized with 18S RNA. The $2^{-\Delta\Delta Ct}$ method was used in order to compare the relative changes (Livak and Schmittgen, 2001). Expression levels in the adult bovine thymus were used as the calibrator in this method.

6.3.6. Spectratyping (III)

Data consisted of the following tissues of four fetuses: bone marrow, ileum, liver, spleen, lymph node and thymus. First-strand cDNA was synthesized using RevertAid Premium Reverse Transcriptase (Fermentas), which had been primed with equal amounts of oligo $(dT)_{20}$ and random hexamer primers. The first-strand cDNA was used as a template for the nested PCR. In the first round, primers that amplified the sequence from IGHV to IGHC region were used. In the second round, fluorescent labeled primers that covered only the CDR3H region were used. The capillary electrophoresis was carried out in the Institute of Biotechnology and the raw data were analysed in a PeakScanner (ABI).

6.3.7. Bacterial artificial chromosome isolation (I)

The known end sequences of the RPCI bovine BAC library were downloaded from NCBI, repeat masked for repetitive sequences and blasted against the chromosomal sequences to identify the BAC end sequences in the chromosomal regions of interest (BTA7:15445000-15848000 and BTA21:71299096-71599096; UMD_3.1). Two potential BAC clones were recognized and cultured in 2YT media (30 µl/ml chloramphenicol) and plated on LB agar with chloramphenicol to obtain single colonies. Subsequently, PCR using T7/SP6 ends as primers was performed in order to verify that the correct clones had been obtained. Several *IGHV* genes were amplified and sequenced from the BAC clones.

6.3.8. Fluorescence *in situ* hybridization (I)

Slides containing bovine chromosome spreads were generated by standard cytogenetic procedures. The DNA was digoxigenin-labeled (Roche) using the Nick

Translation kit (Abbott Molecular). Labeled DNA (100 ng), 1 μg of bovine COT-1 DNA and 2 μg of bovine genomic DNA were then mixed, precipitated and resuspended in a mixture of water and hybridization buffer (Abbott Molecular). The probes were separately denatured at 73°C for 5 min and deposited on slides with the metaphase spreads. The slides were denatured in 70% formamide in 2×SSC at 75°C for 5 min and dehydrated in ethanol. The separate probes were denatured at 73°C for 5 min and placed on slides containing metaphase spreads. The slides were covered, sealed and incubated at 37°C overnight. Slides were first washed in 0.4×SSC/0.3% Tween-20 at 73°C for 1-3 s and then in 2×SSC/0.1% Tween-20 for 1-3 s. TSA Plus Fluorescein kit (PerkinElmer) was used to detect the probes and the nuclei were counterstained using DAPI II (Abbott Molecular).

6.4. Immunostaining (II-III)

PFA fixed paraffin sections were deparaffinized using xylene, rehydrated and then subjected to heat-induced antigen retrieval at 750W in a microwave oven for 15 min. A 500 ml volume of retrieval solution (10 mM Tris-HCl pH 9.5, 1 mM EDTA pH 8.0) was used. Sections were then permeabilized with Tween-20 in PBS. Endogenous peroxidase was blocked by avidin (10% egg white powder) and biotin with 1 mg/ml D-biotin (Sigma-Aldrich). Furthermore, the slides were incubated in 1% goat serum in order to prevent nonspecific binding.

6.4.1. Immunohistochemistry (II)

The Shandon cover plate system (ThermoShandon) was used. The immunogenic peptide of the anti-human AID monoclonal antibody, EK2-5G9, (Ascenion) is 100% identical to the bovine AID. This primary antibody was used in AID immunohistochemistry determinations and it was detected by using a biotinylated goat anti-rat immunoglobulin secondary antibody (Jackson Immunoresearch) that had been enhanced by tyramide signal amplification (Niku et al., 2006).

6.4.2. Immunofluorescence (II-III)

Double immunofluorescence were performed for the ileum, lymph node and spleen samples (II) using the cross-reactive mouse anti-human CD79α monoclonal antibody HM57 (Dako) and rabbit anti-human Ki67 antigen monoclonal antibody SP6 (Thermo Scientific). Detection of these antibodies was performed by goat anti-mouse Ig secondary antibody conjugated with Alexa Fluor 488 dye and goat anti-rabbit Ig secondary antibody conjugated with Alexa Fluor 546 dye (Life Technologies). Nuclei were counterstained by DAPI and any possible autofluorescence derived from fixation was suppressed by incubation in 0.1% Sudan Black B in 70% ethanol.

Double immunofluorescence staining of the thymus samples (unpublished data) was performed sequentially. First, the AID antibody was detected by goat anti-ratIg DyLight 488 secondary antibody (Jackson Immunoresearch). Then, anti-human

CD79α antibody or the pan-keratin antibody cocktail AE1+AE2 (LabVision) was used and detected by a highly cross-absorbed donkey anti-mouseIg DyLight549 secondary antibody (Jackson Immunoresearch).

In study (III), the following primary antibodies were used for the triple immunofluorescence staining; rabbit anti-bovine TdT polyclonal (Dako), rat anti-CD3 monoclonal CD3-12 (Santa Cruz Biotechnology) and mouse anti-human CD79α monoclonal HM57. Detection was performed using goat anti-rabbit Ig Alexa 647 (Dako), goat anti-ratIg DyLight 488 and donkey anti-mouseIg DyLight549 antibody, which was preabsorbed against rat Igs. DAPI and Sudan Black stainings were performed as above.

6.4.3. Image analysis (II-III)

The stained sections were photographed and documented using a Leica DM4000 microscope (Leica Microsystems) with and Olympus DP70 camera and Cell^P software (Olympus), or Zeiss AxioVision microscope. Counting of positive cells was done after photographing by using Image J software (Bethesda, MD).

6.5. Bioinformatics (I-III)

6.5.1. Identification of new *IGHV* and *IGHD* genes (I, III)

Two bovine genome assembly databases, Btau_4.2 and UMD_3.1, were blasted with all known human and murine genomic *IGHV* sequences found for IMGT, in addition to all known bovine and ovine *IGHV* sequences. Hits for which the E value was lower than 10^{-5} were studied for conserved *IGHV* elements. The available genomic data were then explored using the fuzznuc motif search (Rice et al., 2000) for RSSs. Geneious Pro software (Drummond et al., 2010) was used for gene annotations. The functionality of new *IGHV* genes was assessed by analysing their respective promoter motives, ORFs, consensus splice sites, some well-known conserved residues and the RSS.

Btau_4.6, UMD_3.1 genomic databases and the NCBI were used to search for the new *IGHD* genes (III). Unfinished high throughput genomic sequences and the NCBI trace archive were explored using the fuzznuc motif search for consensus D-RSS sequences.

6.5.2. AID induced mutations in *IGHV* genes (II)

Geneious Pro software version 5.6.6 and R software (R Core Team, 2013) was used for assembling the sequence data obtained from cDNA libraries to the reference *IGHV* genes. Gene annotations were done in Geneious and the detection of the mutation type (replacement/silent) and AID target motives. Mutations in FR/CDR regions and target/non-target nucleotides were analysed in R.

6.5.3. Analysis of V(D)J junctional diversity (III)

Geneious Pro software version 6.0, the EMBOSS package (Rice et al., 2000), MUSCLE version 3.7 (Edgar, 2004) and R software were used for analysing the sequence data. Alignment with previously detected genomic sequences (Berens et al., 1997; Saini et al., 1999; Zhao et al., 2003; Hosseini et al., 2004; Ekman et al., 2009) was performed with Smith-Waterman local alignment algorithm, in Biostrings R-package (Pages et al., 2014) and in blastn (Altschul et al., 1990). The boundaries of V, D and J segments were determined by aligning the heavy chain cDNA sequences against custom *IGHV*, *IGHD* and *IGHJ* gene databases, using the pairwise Alignment function in Biostrings. The corresponding V and J segments were first determined and the intervening sequence was blasted with the *IGHD* database (see III, Table S4). The light chain cDNA sequences were analysed similarly but using the blastn for pairwise alignments.

Exonuclease activity was determined in the V, D and J segments by comparing the nucleotide identities of the germline sequences with the recombined sequences. If the germline sequence end had not been excised, potential P nucleotides were identified that were complementary to the germline sequence. The remaining nucleotides, that did not fulfill the criterion for P nucleotides were categorized as N nucleotides.

6.5.4. Statistics (II-III)

The statistical significances of the RT-qPCR analyses (II-III) data were analysed in R using the non-parametric Friedman two-way ANOVA followed by a pair-wise comparison using Wilcoxon-Nemenyi-McDonald-Thompson test, or the Kruskal-Wallis ANOVA followed by pair-wise comparison using the Nemenyi-Damico-Wolfe-Dunn test.

The cell counts for the immunofluorescence analysis (II) were compared using the non-parametric Mann-Whitney U-test run on the IBM SPSS version 20 (IBM Corporation). The same test was applied when comparing the N nucleotide additions in V(D)J junctions (III).

The statistical significance of the selection (II) was assessed by using Bayesian estimation of antigen driven selection (Uduman et al., 2011; Yaari et al., 2012) and run on the web server at http://selection.med.yale.edu/baseline/

7. RESULTS

7.1. The range of functional *IGHV* gene segments is small in cattle (I-III)

7.1.1. All functional *IGHV* genes belong to the subgroup IGHV1 (I).

A total of 36 *IGHV* genes were identified mostly from the UMD_3.1 bovine genome database. Some genes were also retrieved from the Btau_4.2 database and seven genes were exclusively found in the original sequencing traces or high-throughput raw data. Genes were designated according to the IMGT nomenclature for temporary names (Lefranc, 2001c). Expression of the identified *IGHV* genes was verified using fetal cDNA data obtained from the ileum and EST sequence data. A total of 99% of the cDNA data and 94% of the 1415 EST sequences matched the *IGHV* genes. However, some of these genes may represent allelic variants which are impossible to distinguish from paralogous genes based on this data (Pramanik et al., 2011).

Out of the 36 *IGHV* genes identified, 10 fulfilled the criteria set for functional genes. All of them belong to the subgroup IGHV1. The subgroups IGHV2 and IGHV3 contained only nonfunctional genes. The criteria for functionality included a highly conserved promoter region that contained the promoter octamer (ATGCAAAT) and the TATA box sequence (TTAAATT), a leader sequence, intact exons and functional RSS, which consisted of the canonical heptamer (CACAGTG) and the nonamer (ACAAAAACC) in the *IGHV1* genes (I / Table S4).

Twelve genes were located in BTA21 and 17 were found in BTA7 according to UMD_3.1 derived data. The chromosomal localization was confirmed to BTA21 by fluorescence *in situ* hybridization (FISH). BAC clones 138K11 and 152O19 were identified as representing these chromosomal regions according to the BLAST analysis of the RPCI end sequences. PCR analysis and direct sequencing confirmed the amplification of several IGHV genes from the 152O19 clone but not from the 138K11 clone (I / Table S3). Additional BACs that contained *IGHV*s were also searched for but not found.

The localization of the BAC clones on the chromosome was accomplished by using FISH analysis. The BAC 138K11 clone was located to chromosome pair matching the size of BTA7. The 152O19 BAC clone showed a signal on a pair of chromosomes that matched the size of BTA21 (I / Figure S3). Neither of the BACs showed additional signals on other chromosomes.

Phylogenetic analysis showed remarkable differences in the usage of IGHV clans within *Cetartiodactyla*. IMGT defines clans as a set of subgroups, also from different species, which are related to each other in phylogenetic trees (Giudicelli and Lefranc, 1999). Bovine *IGHV*s belong to mammalian clans I and II. All

functional genes belong to subgroup IGHV1, which belongs to clan II. IGHV1 does not have an apparent human homolog but it is homologous to the murine IGHV2 (Q52) (I / Table S7). Bovine IGHV2 pseudogenes also belong to clan II. The subgroups IGHV1 and IGHV2 were probably expanded by a recent duplication of a large homologous region containing one functional gene from IGHV1 and one pseudogene from IGHV2 (I / Figure S2 and dot plot in Figure S5). Extensive conservation of the pseudogenes indicates that the duplications occurred very recently or that there is a selective pressure. Conserved pseudogenes in chicken function as gene conversion donors but we did not find any evidence of this.

No clan III specific *IGHV* genes were found in cattle, sheep or European elk. Genomic PCR results for white-tailed deer, pig, dog, bottlenose dolphin and mouse were positive for clan III (I / Figure 1 and S1).

7.1.2. Targeted sequencing reveals additional expressed *IGHV* sequences (II, III)

Additional *IGHV* genes were extracted from the animals by genomic sequencing, which were used for constructing cDNA libraries. A total of 26 new germline *IGHV* sequences were identified (GenBank accession KJ491073-KJ491098), with between 17-20 different sequences per animal (III / Table S1). There are a maximum of two alleles per locus, thus an animal that has 20 different IGHV sequences might have 10-20 paralogous genes depending on how many of these sequences are alleles.

7.2. AID-induced somatic hypermutation diversifies the preimmune repertoire as early as the fetus stage (II)

7.2.1. The IPP is the major site for post-recombinatorial diversification

The mRNA expression level of AID was measured by RT-qPCR. AID expression was significantly higher in the ileum and the spleen compared to bone marrow and the lymph nodes. Ileal AID expression was further localized to B cell follicles in the fetal IPP (II / Figure 2a, b) by immunohistochemistry. The follicles were positive from the stage the IPP became visible (approx. 200 days of the gestation) (Figure 10). Only sporadic AID-positive follicles could be seen in fetal lymph nodes (II / Figure 2c, d, Figure 10 in this thesis). The staining in the spleen was weak and diffuse due to non-specific binding of the primary antibody (Figure 10). The expression could not be reliably detected in bone marrow.

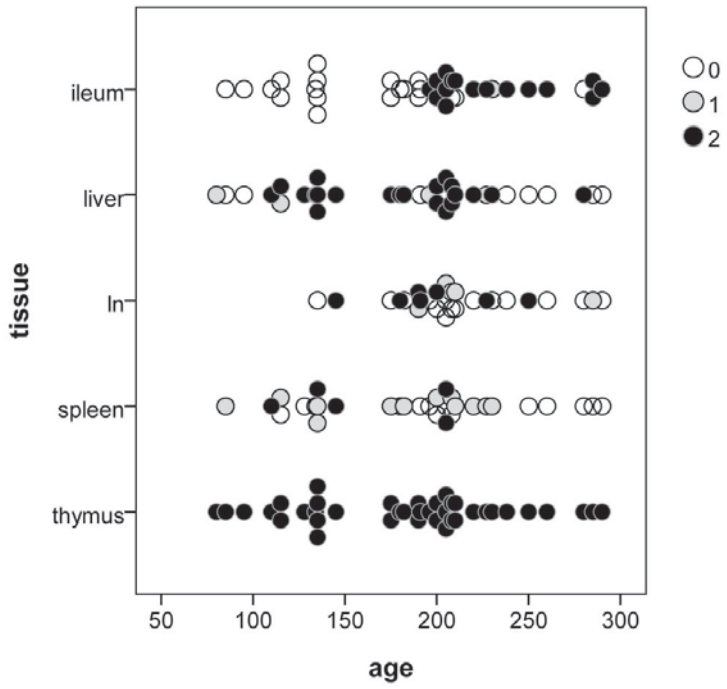

Figure 10. Intensity of AID immunostainings in different tissues. Materials obtained from 40 fetuses, aged from 60 to 290 gd. ln= lymph node. 0=no staining, 1=weak or partial staining of a subset of cells, 2=strong staining of all cells were classified as a positive detection.

The rates of B cell proliferation in the ileum, the spleen and the lymph node were analysed by CD79α/ Ki67 antigen double staining (II / Figure 2e-g). The $CD79\alpha^+$ B cells proliferated at a considerably higher rate ($Ki67^+$ cells, mean 75±9.4%) in the IPP than in the lymph node (9.0±6.7%) or in the spleen (15±10%).

7.2.2. The mutational profiles in the fetal IPP and the spleen are similar to AID-induced SHM in adults

Mutation patterns of fetal *IGHV* sequences (ileum, spleen and bone marrow) was compared to mutation pattern after a secondary antibody response in a 51-day-old calf and the *IGHV* sequences of a 2-year-old immunized cow. The germline *IGHV* sequences that were used as references were obtained from the available bovine genomic sequencing data (I) or by sequencing the germline genes of the same individuals. A total of 35 functional *IGHV* reference genes were used, which include allelic variants in these animals.

Mutation patterns between fetuses and older animals were similar, but there were fewer mutations in the fetuses (II / Figure 3 and Table 1). The mutation load in fetuses was 5-9 mutations/kb (IPP and spleen), in the calf 26 mutations/kb (IPP) and in the immunized cow 80 mutations/kb (lymph node). Mutations were highly

concentrated in CDR regions compared to the FR regions in fetal IPP and spleen, whereas this was not observed in fetal bone marrow, where mutations were rare. Furthermore, mutations in CDRs were 3-fold that of the silent mutations in fetuses and the older animals (II / Figure 3 and Table 1).

There was a positive correlation between AID expression as measured by RT-qPCR and the overall mutation load in fetal samples. AID also has preferential "hotspot" motives in the genome: especially the 3rd C or the 2nd G in WRCY/RGYW are often targeted. The C and G nucleotides within these motives were mutated more often compared to C/G nucleotides outside the motives, both in fetuses and in older animals (II / Figure 4).

7.3. AID is also expressed in fetal thymus and liver

AID expression was also measured in the thymus by RT-qPCR and showed surprisingly high mRNA levels. AID expression as determined by immunohistochemistry was localized to the thymic medulla (Figure 11A and 12).Double immunofluorescence staining was performed using the B-lymphocyte marker CD79α and a pan-keratin antibody cocktail (Figure 12A, B) to investigate the identity of these cells further. A subset of AID^+ $CD79\alpha^+$ double positive cells was found, which suggested that AID was expressed in thymic B lymphocytes. There was also co-localization of AID^+ and pan-$keratin^+$ cells, which suggested that some of the medullary epithelial cells ectopically expressed AID (Figure 12C, D).

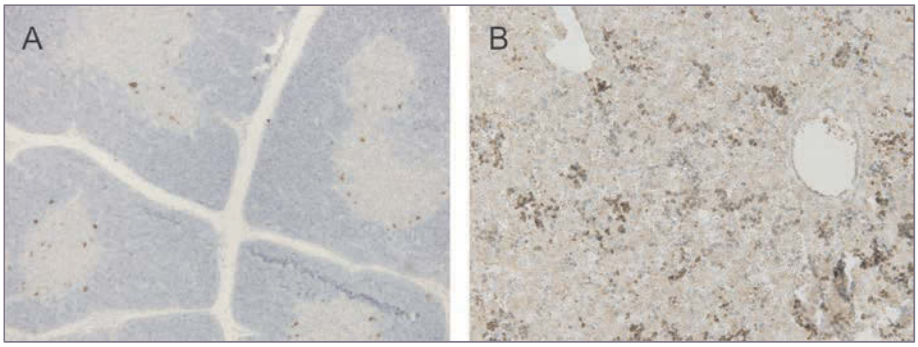

Figure 11. AID immunohistochemistry in the bovine fetus for the A) Thymus at 135 gd. B) Liver at 182 gd.

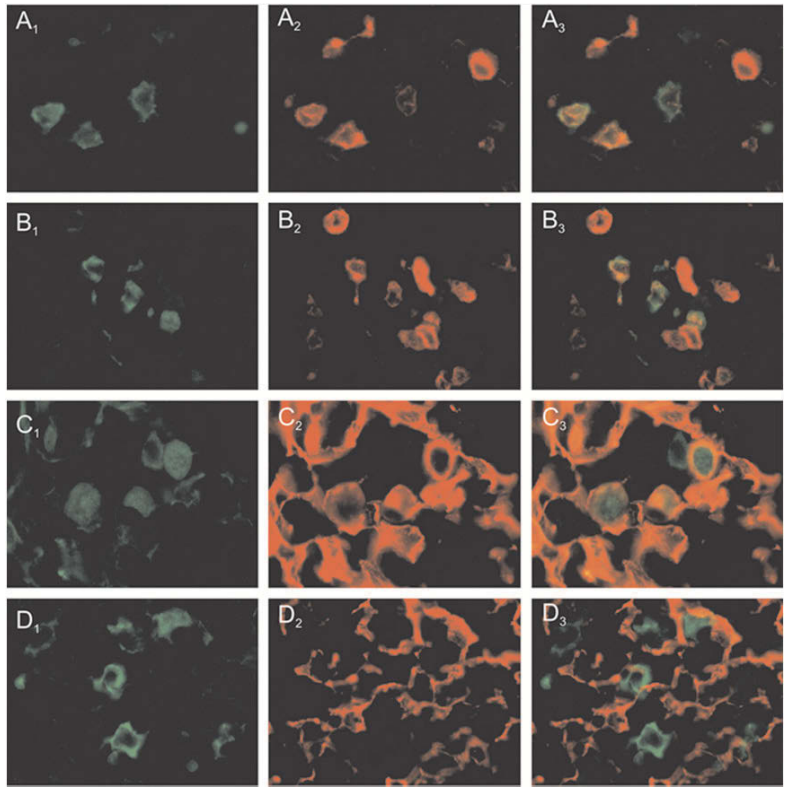

Figure 12. Immunofluorescent double staining for AID, CD79α and pan-keratin in thymus of fetuses aged A) 110 gd 1. AID, 2. CD79α, 3.merge. B) 205 gd 1. AID, 2. CD79α, 3. merge. C) 110gd 1. AID, 2. pan-keratin, 3. merge. D) 182 gd 1. AID, 2. pan-keratin, 3. merge.

Strong AID expression in the non-lymphoid tissues was detected in some liver sections by immunohistochemistry (Figure 10). Further investigation determined that these AID positive clusters were erythroid precursor cells (Figure 11B).

7.4. Several mechanisms and adaptations coupled to somatic recombination compensate the small number V(D)J segments (III)

7.4.1. The exceptionally long CDR3H regions consist of long *IGHD* genes

IGHD genes have been incompletely characterized. We searched for new genes from the available genomic databases and found four new D genes (*IGHDS10* to *IGHDS13*) in addition to the 10 previously characterized *IGHDs* (Hosseini et al., 2004; Koti et al., 2008). One of them, *IGHDS12* comprising 154 bp is the longest that has been reported to date (study III / Table S2). Pairwise alignment between the new IGHD sequences and the IGH cDNAs indicated the existence of a fifth novel *IGHD* sequence, *IGHDS14* that was not uncovered by data mining (III / Figure S1). In total, there was a remarkable size range in the *IGHD* genes, which varied in length

from 31 bp to 154 bp. All of the 14 *IGHD* genes were expressed in gene recombinations in fetal bone marrow, ileum and spleen (III / Table 2).

The CDR3H cDNA sequences of two nearly full term fetuses were analysed (III / Tables S4 and S5). The mean length of the CDR3H region in fetal bone marrow, ileum and spleen (n=645) was 74.9 bp. In 8.4% of these recombinations the CDR3Hs were over 100 bp long and these sequences formed a second subpopulation with exceptionally long CDR3H encoding region (Saini et al., 1999). The CDR3H lengths were further analysed by spectratyping analysis in the thymus, spleen, ileum, lymph node bone marrow and liver. This analysis also clearly revealed two populations of different CDR3H lengths (III / Figure 2).

7.4.2. TdT-mediated N nucleotide additions shape the immunoglobulin junctions in bovine fetal bone marrow

We measured the TdT mRNA expression by RT-qPCR (3 fetuses and 2 adults) in several tissues to characterize where and when TdT is active in cattle. TdT mRNA expression in fetuses, the adult thymus, bone marrow and lymph nodes did not differ statistically and showed higher expression levels than those of the liver, ileum and spleen ($P = 0.003$, $\alpha = 0.05$) (study III / Figure 4). The expression levels of the long isoforms (bovineTdTL1 and bovineTdTL2) were also measured in order to find out if they contributed to the junctional diversity. However, the highest expression levels were seen in the thymus, whereas expressions were barely detected in other tissues.

We performed triple immunofluorescence staining in order to find out in which cell types TdT was active. CD79α was used to identify B lymphocytes and CD3 to identify T lymphocytes. In fetal bone marrow the majority of TdT^+ cells were $CD79\alpha^+$ B lymphocytes (41±13%), whereas only 8.2±2.9% of TdT^+ cells were identified as $CD3^+$ T lymphocytes (study III / Figure 5A). The majority of TdT^+ cells in fetal lymph nodes were T lymphocytes (33±14%) and TdT expressing B lymphocytes were rare (III / Figure 5B).

We analysed the VD and DJ junctions of 645 cDNA clones, derived from the bone marrow, spleen and ileum. N additions were common in both junctions, with 65% of VD and 68% of DJ junctions containing additional nucleotides. In total, 90% of the recombinations contained N nucleotides in at least one of the junctions. The two junctions differed statistically, with a median of one nucleotide in VD and two nucleotides in the DJ junction (III / Figure 3). Some extremely long N additions were also noted which have not been reported in other species so far. In VD junctions the maximum number of N additions was 36 and 16 in the DJ junctions. Palindromic P nucleotides were noted in 16% of VD junctions and 18% of DJ junctions (III / Table 5, Figure 3).

TdT has previously been reported to have biases towards G/C nucleotide additions. Interestingly, the base profile detected in this study was biased towards T nucleotides (33%) followed by A (31%), G (19%) and C (17%) nucleotides.

N nucleotide additions were also detected in the κ and the λ light chains. However, the numbers of additions were lower than those found in the heavy chain, with a median number of zero in λ and one in κ (III / Figure 3). The difference of N additions was statistically significant between the two light chains (Mann-Whitney U test; P=0.006). Only a few P nucleotides could be noted in the light chains. Nucleotide additions in the light chains were biased towards the T nucleotides (36.4%) followed by A (24.3%), C (20.4%) and G (19.0%), which was similar to that found for the heavy chains.

7.4.3. Exonuclease activity produces genetic variability

The nucleotide additions were found to diversify the immunoglobulin genes. Extensive exonuclease activity in the coding segment ends was also detected. In the heavy chains a median of two nucleotides were excised from the 3′ end of the *IGHV* and two nucleotides from the 5′ end of *IGHJ*. Moreover, the *IGHD* genes were extensively trimmed at both ends, particularly those of the long *IGHD*s (III / Table 6). However, the shorter genes were also highly susceptible to trimming with the median values of deleted nucleotides being five in VD junctions and six in DJ junctions.

In the light chains a median of two and half nucleotides were excised from the 3’ end of *IGLV* and *IGKV*. Respectively, median of one nucleotide was excised from 5’end of *IGLJ* and three from *IGKJ*.

8. DISCUSSION

8.1. Working with large animal RNA is difficult

For the most part, this work is based on RT-qPCR analysis and sequence data obtained from cDNA libraries. The starting material used in both analyses was total RNA that had been extracted from different tissues. The specific mRNA component of total RNA is highly susceptible to decay. Degradation pathways ensure that mRNAs do not accumulate in the cell. The mean half-life of the mRNAs in the vertebrate cell is about three hours, though the turnover time varies between RNAs. The general mRNA degradation pathway in eukaryotes begins with the deadenylation of the poly(A) tail, removal of the 5′ cap (decapping) followed by 5′→3′ degradation by diverse exonucleases (Parker and Song, 2004). In addition to this, nucleases of the digestive tract can easily attack mRNAs when the normal balance is lost after death.

Tissue samples were obtained from local slaughterhouses and were snap-frozen in liquid nitrogen or stored in RNAlater immediately after collection. However, the whole fetuses were always transported to our dissection hall for practical reasons. Consequently, the time from the fetuses' deaths until the samples were frozen varied from three to eight hours. The quality of the RNA could, therefore, have been affected during the transportation interval. When taking these logistical impediments into account, it is very difficult to get good quality RNA from the tissue samples of commercially slaughtered cattle than from cell cultures or from murine tissues obtained under laboratory conditions.

The quality of RNA is crucial for accurate gene expression analysis, especially when analysing tissue dependent expression or comparing the expression levels of two or more genes. Freezing and thawing a sample multiple times also adversely affects the quality (own observations), thus after extraction RNA was deep frozen in multiple aliquots and each of which was used once only and subsequently discarded. RNA quantity and integrity were always assessed by using a Nanodrop spectrophotometer and Agilent Bioanalyzer including the RNA integrity number (RIN). The RIN number is a software tool developed by Agilent technologies, which interprets the RNA electropherogram by comparing the ratio of ribosomal RNA peaks and entire electrophoretic trace of the RNA sample. The presence or absence of degradation products can be monitored from these traces. Our own observations indicate that the RIN number for RT-qPCR experiments had to be at least five, preferably over eight in order to convert RNA successfully into cDNA.

The quality control for cDNA after reverse transcription was performed by monitoring the gene expression level of at least one, preferably two reference housekeeping genes. *GAPDH* and *18S* were used for this purpose. Sometimes the

selection of a suitable reference gene is difficult because of the difference in the expression levels of the reference gene compared to those of the gene of interest. It is recommended to use approximately the same amount of RNA for the cDNA conversion (Bustin et al., 2009). Consequently, the Ct values of the reference gene(s) had to be close to each other in samples, which were analysed during the same qPCR run. In this way the efficiency of cDNA reverse transcription could be observed along with the quality of the cDNA.

8.2. Diversification of the preimmune repertoire

8.2.1. Recombinatorial diversity (I, III)

The numbers of functional κ and λ light chains in cattle is moderate compared to the human and the mouse (Ekman et al., 2009) and Table 1. The knowledge of *IGHV* gene selection in ruminants has been limited and the knowledge of *IGHV* and *IGHD* germline genes is crucial for the accurate analysis of the mechanisms that the bovine uses in the preimmune repertoire diversification. Analysis of SHM *inter alia* would have been impossible without the appropriate references. Therefore, we performed intensive searches to complete these gene repertoires.

According to the bioinformatic analyses of UMD_3.1, Btau_4.2, raw sequencing data and targeted genomic sequencing, we could identify 62 *IGHV* sequences out of which 26 were functional. This number contains actual paralogous genes and also possible alleles. The bovine repertoire seems very small by comparison to the 39 functional *IGHV* genes in the human and 101 in the mouse. The genomic databases are based on the shotgun sequence of a Hereford cow, L1 Dominette 01449, which has a 30% inbreeding coefficient. We found 17 to 20 different sequences per animal by targeted sequencing of the *IGHV* germline genes obtained from our samples. There are a maximum of two alleles per locus, therefore an animal that has 20 different IGHV sequences should have between 10 to 20 paralogous genes depending on the level of heterozygocity in its IGH locus. Allelic variants are impossible to differentiate from paralogous genes without knowing each gene's locus in the genome (Pramanik et al., 2011).

IGHV genes were located in two chromosomes as determined by UMD_3.1 genome assembly analysis. This unusual situation has not been reported in any other mammals (Das et al., 2008). Therefore, these suggested chromosomal localizations were analysed further by using two BAC clones (152O19 and 138K11), which were mapped by their end sequences to the relevant chromosomal regions, and used as probes in the FISH analysis. Only clone 152O19 showed evidence of *IGHV* sequences and it was located to the BTA21 region. This suggests that analysis of the UMD_3.1 assembly revealed that a segment of BTA21 had been inaccurately placed to BTA7. The whole *IGH* locus is extremely challenging to assemble due to the highly repetitive and polymorphic sequences it contains (Li et al., 2002), thus targeted sequencing is needed for resolving its correct assembly.

Several attempts have been made to identify all bovine *IGHD* genes. Previously 10 *IGHD* sequences were reported and mapped to BTA21, BTA7 and BTA8 (Shojaei et al., 2003; Koti et al., 2008). In addition to these discoveries, we found four new *IGHD* genes by searching the following databases by using RSS motives as queries: bovine genomic sequencing data, NCBI Unfinished high throughput genomic sequences, and NCBI trace archives. We also found one plausible new *IGHD* among the sequenced cDNAs (*IGHDS14*) but this gene was not detected from the databases. Despite this multi-analytical strategy, we could still have missed some of the *IGHD* germline genes or some of these reported genes could be allelic variants. Expression of all five of the newly identified genes could be detected in our fetal CDR3H cDNA libraries although assigning the correct genomic reference is problematic. *IGHD* genes are highly repetitive and subjected to trimming in order to produce several variable transcripts out of one long genomic sequence. This is possible because many D genes are translatable in all three coding frames.

IGHJ genes are also incompletely characterized. The bovine JH locus has been reported to possess six *IGHJ* genes out of which at least two are functional (Berens et al., 1997; Zhao et al., 2003; Hosseini et al., 2004). However, there is contradictory information about the expressed genes. The locus is highly polymorphic and thus shows great sequence diversity between breeds or between individuals (Zhao et al., 2003). This variability might be one of the reasons why concordance of data has not been found yet. *IGHJ* genes have been reported to locate in two chromosomes, BTA21 and BTA11 by Hosseini et al. (2004). Those authors denoted the former as J_H^{high} locus and the latter a duplicated J_H^{low} according to the frequency of the expressed *IGHJ* sequences. However, if all *IGHV* genes locate to BTA21 as suggested by study (I) data then immunoglobulin gene rearrangements with J_H^{low} would have to exploit uncommon mechanisms such as transchromosomal recombination (Knight et al., 1995; Kingzette et al., 1998). The chromosomal localization conclusion was based on the sequencing data of two BAC clones. The J_H^{high} locus could be anchored to the BTA21 sequence due to the presence of immunoglobulin constant genes known to reside there. The assignment of J_H^{low} locus was deduced from the proximity of duplication of the IGM locus known to reside in BTA11 (Hayes and Petit, 1993).

We reported the use of three different *IGHJ* genes (study III / Tables S3 and S5) using igblast analyses of the bovine fetal bone marrow, ileum and spleen *IGH* sequences. Consistent with the previous reports, *IGHJ1* (AY158087) was predominantly expressed at 90-94% of the recombinations, depending on the analysed tissue. Nevertheless, it cannot be ruled out that we are still missing some paralogous genes or at least some alleles. Almost the entire J segment belongs to the FR4 region in bovine IGH, which is not as susceptible to mutations as the CDR3 regions. This makes the identification of the *IGHJ* genes easier than for the *IGHV* or

the *IGHD* genes. Therefore, it would be relatively straightforward to identify all expressed *IGHJ* genes in one individual by targeted sequencing.

8.2.2. Junctional diversity is coupled to somatic recombination (III)

In cattle, the preimmune repertoire would be substantially smaller than those of the human or the mouse when formed only by V(D)J recombination. This is due to the small number of germline gene segments. When we analysed the mutation pattern in the CDR3H region (II), extensive junctional diversity between the gene segments was detected. TdT has long been known to contribute to the N nucleotide additions (Desiderio et al., 1984), which have also been reported in bovine heavy chains (Berens et al., 1997; Koti et al., 2010). We wanted to determine the expression of TdT and the role of junctional diversity more carefully in bovine fetal tissues, both in heavy and light chains. Therefore, identification of *IGH* germline genes was a prerequisite for N nucleotide identification and for being able to set boundaries between gene segments.

The strongest extra-thymic TdT expression as determined by RT-qPCR was measured in bone marrow in fetuses. Adults were used as a control population and surprisingly an equally strong TdT expression was detected in adult bone marrow. In cattle, somatic recombination is restricted to the fetal and the neonatal periods (Ekman et al., 2012). The TdT^+ cell types were confirmed using immunofluoresce since N additions also occur in the T cell receptor genes (Murphy, 2012). More than 10% of B cells in the fetal bone marrow were TdT^+ whereas the number of T cells was not significant there. A majority of TdT^+ cells in fetal lymph nodes were T lymphocytes. A fraction of TdT^+ cells were negative for both CD markers in bone marrow and lymph nodes. A plausible explanation is that these are early lymphoid progenitor cells (Gore et al., 1991). Extensive studies have shown an association between acute myeloid leukemia and TdT, and proposed that the TdT expression is not restricted to the cells completely destined to the lymphoid lineage (Drexler et al., 1993; Patel et al., 2013). These observations may explain the findings from adult bone marrow in which demonstrated *de novo* lymphopoiesis had already declined (Ekman et al., 2012).

TdT expresses spliceosomal isoforms in cattle (Takahara et al., 1994). One short and two long variants have been detected in the thymus of the bovine. In the long variants, the extra exon forms a new catalytic site, which in some species has been shown to convert the enzyme functioning as an exonuclease (Thai and Kearney, 2004). We also detected the expression of the long variants in adult and fetal thymus of the bovine, though in other tissues the expression levels were below the limits of detection. This finding could indicate that the long variants are fully T cell specific as they function exclusively in the thymus. An alternative explanation is that only a subset of B cells expresses the long variants and we were not able to identify these clones.

We analysed the junctional nucleotide composition in fetal bone marrow, spleen and ileum. These tissues were selected based on previous studies: RAG-dependent gene recombination takes place in fetal bone marrow (Ekman et al., 2010) and immunoglobulin genes are diversified in the IPP and spleen by SHM (study II). Gene segments were identified and boundaries were set by bioinformatics alignment tools, carefully adjusting the parameters for the immunoglobulin genes. In fetal tissues, 90% of VD or DJ junctions included N nucleotides which suggested TdT activity. The median number of nucleotide additions in VD was one and in DJ two. However, approximately 4% of junctions included extremely long additions ranging from 10 to 37 nucleotides. These long junctions had so far not been reported in fetuses in other species, although junctional diversity has been reported as an important mechanism in the pig (Butler et al., 2000), the human (Schroeder et al., 1995), the sheep (Gontier et al., 2005) and possibly in the horse (Sun et al., 2010). In contrast, junctional diversity has only been detected in adults of mice (Feeney, 1990).

N additions did appear in the light chains but to a considerably lesser extent than in the heavy chains. This difference might indicate that the contribution of light chains to the antibody diversity is not as significant as that of the heavy chains. Respective, TdT activity in the human and the mouse is also restricted to the pro-B cell stage and not effective anymore during the rearrangement of the light chains (Li et al., 1993; Galler et al., 2004). Our results indicate that either TdT activity is weaker though not completely inactivated during the rearrangement of the light chain or that there are other polymerases in addition to TdT.

We also analysed the junctional base composition. In both the heavy and the light chains, and detected bias towards T nucleotide additions (33% in VD and 26% in DJ junction and 36.4% in VJ junction respectively). Previously, TdT has been reported to add G residues preferentially, to the extent of 60-70% (Basu et al., 1983). Possible explanations for these findings may arise from the differences in functions *in vivo* and *in vitro* reported for TdTs (Chang and Bollum, 1986; Sandor et al., 2004). If limiting factors are not present *in vitro*, then several kilobases of nucleotides can be added (Chang and Bollum, 1986). Recently, it has been suggested that Ku80, which is a part of the DNA-PK, could be one of these limiting factors (Sandor et al., 2004). The long extensions *in vitro* are less dominant with a G nucleotide bias, which may be due to conformational limitations (Lefler and Bollum, 1969; Mickelsen et al., 1999). The bovine may lack an inhibitor for TdT, which is included in the mouse and in the human genomes. Therefore, TdT in the bovine is able to add tens of nucleotides into the junctions. The hypothesis for this mechanism relies on the above mentioned *in vitro* studies but requires further studies on DNA-PK and Ku80 proteins in the bovine. Studies in human and mouse, found that other PolX family polymerases also have an impact in junctional diversity (Bertocci et al., 2003, 2006). These polymerases are widely uncharacterized in cattle so, it would be interesting to

discover if they make a contribution to the antibody repertoire diversification in the bovine.

In addition to N nucleotide additions, *IGHDs* also diversify the immunoglobulins in several other ways. One of the new *IGHD* genes this study has identified at the time of writing the longest D gene found in the bovine, as it extends up to 154 bp in length. Nevertheless, only the short parts of the long genes were usually expressed in the recombinations, which is probably due to extensive exonuclease activity. This enables the production of several different D segments out of one genomic sequence. This might be a new bovine specific mechanism for repertoire diversification. *IGHDs* are also highly homologous and susceptible to mutations towards cysteines: a phenomenon that permits the formation of new disulfide bonds (Wang et al., 2013).

Results of N nucleotide additions and exonuclease activity studies are completely dependent on the correct set of reference genes and the alignments of the query sequences with them. Consequently, missing germline genes or mutations in the sequence can lead to inaccurate alignment being made. We attempted to avoid this by extensively searching for new *IGHV* and *IGHD* germline genes and adjusted the blast parameters accordingly. We suggest that our data makes a significant contribution to the knowledge of junctional diversity in the fetal bovine.

8.2.3. Post-recombinatorial mechanisms (II)

AID is a well-known mutator enzyme that functions in secondary antibody diversification (Muramatsu et al., 2000; Arakawa et al., 2002). Long before the discovery of AID, post-recombinatorial mechanisms had already been reported to diversify the preimmune antibody repertoire in the sheep IPP (Reynaud et al., 1991; Reynaud et al., 1995). AID mRNA expression in lymphoid tissues was reported in neonatal and in adult cattle but not in the IPP site (Verma et al., 2010).

Strong AID mRNA expression was detected in the fetal IPP, spleen and thymus in study (II). Furthermore, protein expression of AID was localized in the IPP follicles, the thymic medulla, clusters in the liver and occasional cortical follicles in the lymph nodes. In contrast, protein expression of AID in the bone marrow and spleen could not be reliably detected. This was due to nonspecific binding of the primary antibody to non-B lineage cells.

Further investigation of the thymus revealed that AID expressing cells are either $CD79^+$ B lymphocytes or $keratin^+$ medullary epithelial cells. Developing T lymphocytes in the thymus undergo negative selection in order to remove self-reactive clones. The role of thymic B cells has long been unclear but recent studies are beginning to reveal their development and function. Thymic B cells differentiate in the thymus and have different characteristics to those of the peripheral B cells. The expression levels of the MHC class II is high, which suggests that they are ready to present self-antigens efficiently to developing T cells, thus eliminating the

autoreactive clones (Perera et al., 2013). In fact, these B cells often co-localize with the medullary epithelial cells that are also specialized in presenting peripheral antigens, such as AID, for the induction of self-tolerance. Therefore, we hypothesize that AID could be ectopically expressed by the autoimmune regulator AIRE. We suggest that AID is expressed in the thymic medulla by autoreactive thymic B cells for the negative selection of T cells.

In addition to the lymphoid organs, we also detected AID^+ cells in the liver. The AID^+ cells were organized in clusters of differentiating erythroid cells. AID has recently been associated with several major epigenetic reprogramming processes that are known to occur during erythroid differentiation (Bhutani et al., 2010; Popp et al., 2010; Liu et al., 2013) this very interesting finding is definitely worth further investigation.

B lymphocytes that expressed high levels of AID were found in the IPP and the spleen. We constructed and sequenced *IGHV* cDNA libraries of two nearly full term fetuses to characterize the actual mutational profile in these tissues. In both tissues the pattern of the mutations was similar to those of the control animals (two calves) that had AID-induced secondary antibody responses. Mutations were preferentially localized in the CDRs, which resulted in codon replacement. Furthermore, the G/C nucleotides in preferential AID "hotspot" motives were visibly mutated at a higher frequency than other Gs or Cs outside these motives.

SHM produces primarily point mutations, although sometimes larger lesions occur depending on the DNA-repair mechanism (Saribasak and Gearhart, 2012). PCR and sequencing derived errors can lead to misinterpretation of mutations, therefore only the high quality sequences were accepted for the analysis in study III. PCR errors were controlled by sequencing the vector. The error rate was only 0.784 errors/kb: a rate that did not affect the results because it could only account for about 5% of the mean number of mutations. Sometimes PCR chimeras are formed when performing PCR from a pool of highly homologous sequences, such as the *IGHV*s. We adjusted for this by using equal amounts of two *IGHV*s as templates, then cloned and sequenced the PCR products. Out of 109 clones only one resulted the formation of a chimera, which suggested that chimerism was not a prominent occurrence with the polymerase used.

Study (II) provides extensive new data of AID-induced SHM in the fetal IPP and spleen occurring prior to contact with any external antigen. The proliferation rate of B cells in the spleen is one-fifth that which occurs in the IPP. This results in a small splenic B lymphocytes pool that cannot substantially contribute to the repertoire diversification. SHM has generally been classified as a secondary diversification mechanism, which is initiated by external antigen contact (Peled et al., 2008). The regulation of AID targeting during SHM is little understood, reviewed by Zan and Casali (2013) as is the regulation process that occurs during the diversification of the

preimmune repertoire. The presence of intestinal microflora and commensal bacteria in the rabbit are required for the generation of diversified preimmune repertoire in GALT (Rhee et al., 2004, 2005). Recently, it was shown that exposure to exogenous antigens in chicken play an important role in bursal B cell selection after birth (Davani et al., 2014). However, AID induction and regulation in cattle has to be driven by endogeneous antigens or by other factors because no external antigens are present. These very interesting but unanswered questions require further investigation.

After recombination, a small population of cells migrates to the IPP and the spleen. The cells vigorously proliferate in the IPP and are diversified by SHM. How the appropriate cells that enter into the IPP are selected is still unknown. Conformational restrictions may be one mechanism since rearranged antibodies that contain exceptionally long CDR3 regions are reported to form an unusual 3D structure (Wang et al., 2013). Whether this inhibits or facilitates the entrance to the IPP remains to be resolved.

8.3. Comparative and evolutionary aspects

B lymphopoiesis takes place in fetal liver and bone marrow in the human and the mouse. The bone marrow and the thymus are conventionally classified as primary lymphoid organs and the removal of these organs results in deficiency of B and T cells, respectively. The *bursa of Fabricius* in chicken can also be defined as primary lymphoid organ because the removal of the bursa results in B cell deficient chickens (Glick et al., 1956). Another characteristic for the primary organs is that antigen independent repertoire diversification takes place there. Antigen dependent, adaptive immune responses take place in secondary lymphoid tissues.

Cattle, chicken, rabbit, pig and sheep are also called 'GALT species' because their B cell diversification takes place in the IPP, appendix or bursa, according to the prevailing paradigm (Reynaud et al., 1991; Lanning et al., 2000; Ratcliffe, 2002). The bovine IPP cannot be classified as a primary lymphoid organ because *de novo* lymphopoiesis and recombination take place in the bone marrow and the lymph nodes instead. This was shown by the expression of the surrogate light chains and RAG1/2 (Ekman et al., 2010, 2012). However, antigen independent repertoire diversification occurs in the IPP, indicating that it fulfills one of the features of primary organs. Removal of the terminal ileum results in long-standing B lymphopenia in fetal and neonatal lambs (Gerber et al., 1986). Bursectomy, in contrast, results in complete B cell deficiency in chicken (Glick et al., 1956) but somatic recombination does not occur in the bursa (Benatar et al., 1991). These data challenges the conventional division of organs into primary and secondary tissues and perhaps these concepts should be further defined.

Recently, the concept of GALT species in sheep, rabbit and pig has been challenged (Jenne et al., 2003; Sinkora et al., 2011; Butler and Sinkora, 2013). The combinatorial repertoire in the sheep was reported to be more divergent than previously understood, which suggests less emphasis on SHM (Jenne et al., 2003). However, these conclusions were drawn based on data obtained from light chain genes. Novel *IGLV* genes were predicted from the cDNA sequences, without taking into account the possibility for allelic variants or mutated sequences. In fact, data from study III suggest that the light chains may contribute less to the antibody repertoire diversity than the heavy chains. This is also supported by the findings of other investigations, which found minimal junctional diversity and SHM the light chains (Jenne et al., 2003; Wertz et al., 2013). Moreover, some species such as those that belong to the camelids are able to survive completely without light chains (Nguyen et al., 2002; Su et al., 2002).

According to a recent hypothesis, the artiodactyl IPP is not a mammalian bursal analogue: instead its main function is to combat bacterial exposure before T cell dependent mechanisms take over (Butler et al., 2000; Sinkora et al., 2011). These authors also claim that the porcine IPP is not required for B cell lymphopoiesis, maintenance or repertoire development and that SHM-induced repertoire diversification is antigen dependent. Later they also showed that the IPP of the pig is not a significant source of B cells and that removal of the IPP does not affect antibody repertoire diversification in mesenteric lymph nodes, tracheal bronchial lymph nodes or the spleen. More recently they questioned the division of species into GALT and bone marrow groups (Butler and Sinkora, 2013). This evidence may be convincing in the pig but it cannot be directly extrapolated to all other artiodactyls. First, domestic species all have characteristic and distinct mechanisms that they use in the preimmune repertoire development. Second, according to molecular data, the order *Artiodactyla* (the even-toed ungulates) is an incomplete phylogenetic group without the cetaceans (whales, dolphins and porpoises) (Figure 13) (O'Leary, 2001). The grouping of *Artiodactyla* is only based on morphological similarities. Based on the molecular data, whales (*Iniidae*) and dolphins (*Delphinidae*) diverged from cattle more recently than pigs (56 vs. 63 million years ago, Figure 13). These data indicate that these two groups are phylogenetically closer to bovine than pigs.

The immune system of the host has evolved in parallel with the important prokaryotic ecosystem, the microbiota thus; generalizing immunological characteristics based solely on morphology is not feasible (Rhee et al., 2004; Slack et al., 2009; Lee and Mazmanian, 2010). Interactions between the innate immune system with the microbiota are an essential driving force shaping the immune system, reviewed in Thaiss et al. (2014). Furthermore, a recent study showed evidence that microbiota-dependent immune system maturation requires a species-specific composition of the microbiota (Chung et al., 2012).

The germline gene repertoires, the timing of diversification, the location, exact mechanisms and the microbiota are all species dependent. The division between GALT and non-GALT species does not take into account all of these above mentioned facts. Therefore, generalization of development and function of immune systems between species should be done with caution and the immunological particularities of different species should be acknowledged. Studies in comparative immunology have already taught us a lot and can continue to do so also in the future.

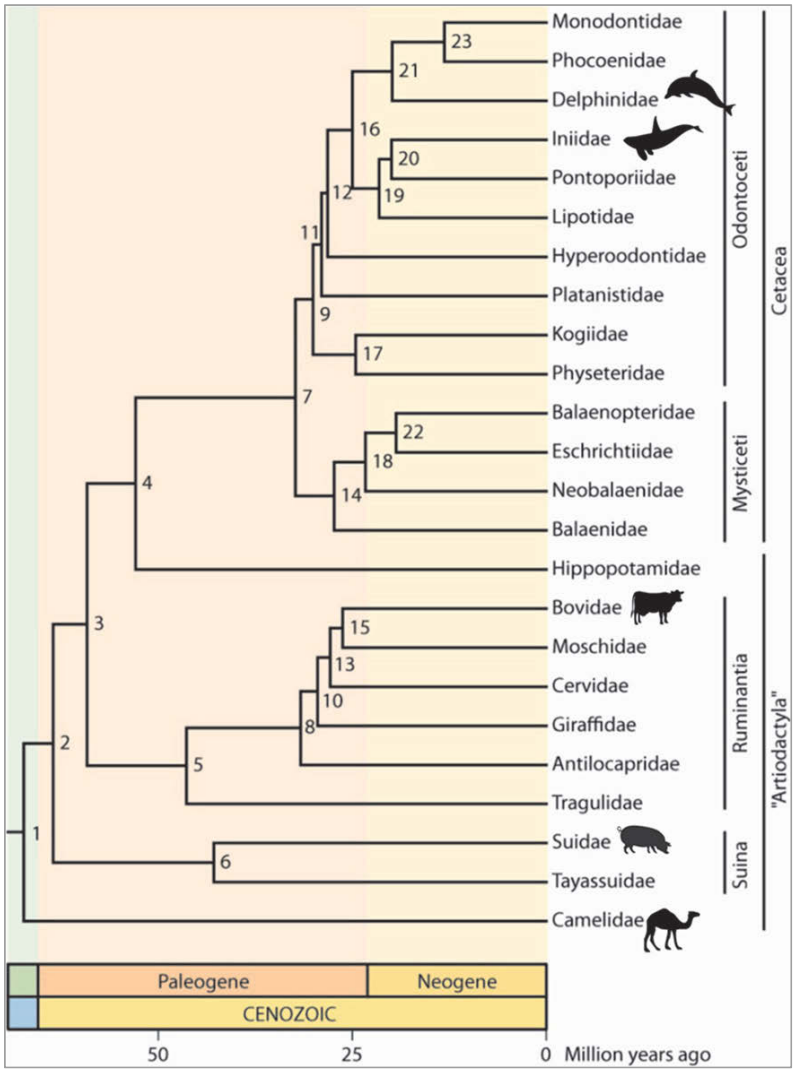

Figure 13. A timetree of Cetartiodactyla (whales, dolphins, porpoises and even toed ungulates). Modified from "The Time tree of Life", S.B. Hedges and S. Kumar. Eds. (Oxford University Press, 2009). Reprinted with permission from OUP.

9. CONCLUSIONS

I) The bovine *IGHV* repertoire is very limited. A total of 62 *IGHV* genes were identified in the genomic databases by targeted sequencing. The *IGHV* gene locus contains 10-20 functional paralogous genes. The *IGHV* locus is located in BTA21.

II) AID-mediated SHM contributes to the diversification of the preimmune repertoire in fetal IPP and spleen of the bovine. The bovine fetal IPP has more impact on diversification due to the extremely rapid proliferation rate in that site. SHM takes place independent of exogenous antigen encounters.

III) TdT-mediated junctional diversity takes place in bovine fetal bone marrow. The V(D)J junctions contain long N nucleotide additions in the heavy chains and to a lesser extent in the light chains, which suggest that heavy chains have more impact on the repertoire diversification process.

IV) *IGHD* genes are susceptible to extensive exonuclease activity. This may be a novel, bovine specific diversification mechanism.

10. REFERENCES

Agrawal, A., Schatz, D.G., 1997. RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. Cell 89, 43–53.

Almagro, J.C., Hernández, I., Ramírez, M.C., Vargas-Madrazo, E., 1998. Structural differences between the repertoires of mouse and human germline genes and their evolutionary implications. Immunogenetics 47, 355–363.

Alt, F.W., 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.

Alt, F.W., Blackwell, T.K., DePinho, R.A., Reth, M.G., Yancopoulos, G.D., 1986. Regulation of genome rearrangement events during lymphocyte differentiation. Immunol. Rev. 89, 5–30.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Arakawa, H., HauschiLd, J., Buerstedde, J.M., 2002. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. Science 295, 1301–1306.

Archer, P.G., Sutherland, D.E., Good, R.A., 1963. Appendix of the rabbit: A homologue of the bursa in the chicken? Nature 200, 337–339.

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. Zentralblatt Für Veterinärmedizin Reihe A 43, 573–576.

Basu, M., Hegde, M.V., Modak, M.J., 1983. Synthesis of compositionally unique DNA by terminal deoxynucleotidyl transferase. Biochem. Biophys. Res. Commun. 111, 1105–1112.

Batista, F.D., Neuberger, M.S., 1998. Affinity dependence of the B cell response to antigen: a threshold, a ceiling, and the importance of off-rate. Immunity 8, 751–759.

Benatar, T., Iacampo, S., Tkalec, L., Ratcliffe, M.J., 1991. Expression of immunoglobulin genes in the avian embryo bone marrow revealed by retroviral transformation. Eur. J. Immunol. 21, 2529–2536.

Benedict, C.L., Gilfillan, S., Kearney, J.F., 2001. The long isoform of terminal deoxynucleotidyl transferase enters the nucleus and, rather than catalyzing nontemplated nucleotide addition, modulates the catalytic activity of the short isoform. J. Exp. Med. 193, 89–99.

Benedict, C.L., Gilfillan, S., Thai, T.-H., Kearney, J.F., 2000. Terminal deoxynucleotidyl transferase and repertoire development. Immunol. Rev. 175, 150–157.

Benedict, C.L., Kearney, J.F., 1999. Increased junctional diversity in fetal B cells results in a loss of protective anti-phosphorylcholine antibodies in adult mice. Immunity 10, 607–617.

Bentolila, L.A., Fanton d'Andon, M., Nguyen, Q.T., Martinez, O., Rougeon, F., Doyen, N., 1995. The two isoforms of mouse terminal deoxynucleotidyl transferase differ in both the ability to add N regions and subcellular localization. EMBO J. 14, 4221–4229.

Berens, S.J., Wylie, D.E., Lopez, O.J., 1997. Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. Int. Immunol. 9, 189–199.

Bertocci, B., De Smet, A., Berek, C., Weill, J.-C., Reynaud, C.-A., 2003. Immunoglobulin kappa light chain gene rearrangement is impaired in mice deficient for DNA polymerase mu. Immunity 19, 203–211.

Bertocci, B., De Smet, A., Weill, J.-C., Reynaud, C.-A., 2006. Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo. Immunity 25, 31–41.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., Blau, H.M., 2010. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature 463, 1042–1047.

Bhutani, N., Burns, D.M., Blau, H.M., 2011. DNA demethylation dynamics. Cell 146, 866–872.

Bovine Genome Sequencing and Analysis Consortium, Elsik, C.G., Tellam, R.L., Worley, K.C., Gibbs, R.A., Muzny, D.M., Weinstock, G.M., Adelson, D.L., Eichler, E.E., Elnitski, L., Guigo, R., Hamernik, D.L., Kappes, S.M., Lewin, H.A., Lynn, D.J., Nicholas, F.W., Reymond, A., Rijnkels, M., Skow, L.C., Zdobnov, E.M., Schook, L., Womack, J., Alioto, T., Antonarakis, S.E., Astashyn, A., Chapple, C.E., Chen, H.C., Chrast, J., Camara, F., Ermolaeva, O., Henrichsen, C.N., Hlavina, W., Kapustin, Y., Kiryutin, B., Kitts, P., Kokocinski, F., Landrum, M., Maglott, D., Pruitt, K., Sapojnikov, V., Searle, S.M., Solovyev, V., Souvorov, A., Ucla, C., Wyss, C., Anzola, J.M., Gerlach, D., Elhaik, E., Graur, D., Reese, J.T., Edgar, R.C., McEwan, J.C., Payne, G.M., Raison, J.M., Junier, T., Kriventseva, E.V., Eyras, E., Plass, M., Donthu, R., Larkin, D.M., Reecy, J., Yang, M.Q., Chen, L., Cheng, Z., Chitko-McKown, C.G., Liu, G.E., Matukumalli, L.K., Song, J., Zhu, B., Bradley, D.G., Brinkman, F.S., Lau, L.P., Whiteside, M.D., Walker, A., Wheeler, T.T., Casey, T., German, J.B., Lemay, D.G., Maqbool, N.J., Molenaar, A.J., Seo, S., Stothard, P., Baldwin, C.L., Baxter, R., Brinkmeyer-Langford, C.L., Brown, W.C., Childers, C.P., Connelley, T., Ellis, S.A., Fritz, K., Glass, E.J., Herzig, C.T., Iivanainen, A., Lahmers, K.K., Bennett, A.K., Dickens, C.M., Gilbert, J.G., Hagen, D.E., Salih, H., Aerts, J., Caetano, A.R., Dalrymple, B., Garcia, J.F., Gill, C.A., Hiendleder, S.G., Memili, E., Spurlock, D., Williams, J.L., Alexander, L., Brownstein, M.J., Guan, L., Holt, R.A., Jones, S.J., Marra, M.A., Moore, R., Moore, S.S., Roberts, A., Taniguchi, M., Waterman, R.C., Chacko, J., Chandrabose, M.M., Cree, A., Dao, M.D., Dinh, H.H., Gabisi, R.A., Hines, S., Hume, J., Jhangiani, S.N., Joshi, V., Kovar, C.L., Lewis, L.R., Liu, Y.S., Lopez, J., Morgan, M.B., Nguyen, N.B., Okwuonu, G.O., Ruiz, S.J., Santibanez, J., Wright, R.A., Buhay, C., Ding, Y., Dugan-Rocha, S., Herdandez, J., Holder, M., Sabo, A., Egan, A., Goodell, J., Wilczek-Boney, K., Fowler, G.R., Hitchens, M.E., Lozado, R.J., Moen, C., Steffen, D., Warren, J.T., Zhang, J., Chiu, R., Schein, J.E., Durbin, K.J., Havlak, P., Jiang, H., Liu, Y., Qin, X., Ren, Y., Shen, Y., Song, H., Bell, S.N., Davis, C., Johnson, A.J., Lee, S., Nazareth, L.V., Patel, B.M., Pu, L.L., Vattathil, S., Williams, R.L., Curry, S., Hamilton, C., Sodergren, E., Wheeler, D.A., Barris, W., Bennett, G.L., Eggen, A., Green, R.D., Harhay, G.P., Hobbs, M., Jann, O., Keele, J.W., Kent, M.P., Lien, S., McKay, S.D., McWilliam, S., Ratnakumar, A., Schnabel, R.D., Smith, T., Snelling, W.M., Sonstegard, T.S., Stone, R.T., Sugimoto, Y., Takasuga, A., Taylor, J.F., Van Tassell, C.P., Macneil, M.D., Abatepaulo, A.R., Abbey, C.A., Ahola, V., Almeida, I.G., Amadio, A.F., Anatriello, E., Bahadue, S.M., Biase, F.H., Boldt, C.R., Carroll, J.A., Carvalho, W.A., Cervelatti, E.P., Chacko, E., Chapin, J.E., Cheng, Y., Choi, J., Colley, A.J., de Campos, T.A., De Donato, M., Santos, I.K., de Oliveira, C.J., Deobald, H., Devinoy, E., Donohue, K.E., Dovc, P., Eberlein, A., Fitzsimmons, C.J., Franzin, A.M., Garcia, G.R., Genini, S., Gladney, C.J., Grant, J.R., Greaser, M.L., Green, J.A., Hadsell, D.L., Hakimov, H.A., Halgren, R., Harrow, J.L., Hart, E.A., Hastings, N., Hernandez, M., Hu, Z.L., Ingham, A., Iso-Touru, T., Jamis, C., Jensen, K., Kapetis, D., Kerr, T., Khalil, S.S., Khatib, H., Kolbehdari, D., Kumar, C.G., Kumar, D., Leach, R., Lee, J.C., Li, C., Logan, K.M., Malinverni, R., Marques, E., Martin, W.F., Martins, N.F., Maruyama, S.R., Mazza, R., McLean, K.L., Medrano, J.F., Moreno, B.T., More, D.D., Muntean, C.T., Nandakumar, H.P., Nogueira, M.F., Olsaker, I., Pant, S.D., Panzitta, F., Pastor, R.C., Poli, M.A., Poslusny, N., Rachagani, S., Ranganathan, S., Razpet, A., Riggs, P.K., Rincon, G., Rodriguez-Osorio, N., Rodriguez-Zas, S.L., Romero, N.E., Rosenwald, A., Sando, L., Schmutz, S.M., Shen, L., Sherman, L., Southey, B.R.,

Lutzow, Y.S., Sweedler, J.V., Tammen, I., Telugu, B.P., Urbanski, J.M., Utsunomiya, Y.T., Verschoor, C.P., Waardenberg, A.J., Wang, Z., Ward, R., Weikard, R., Welsh, T.H., White, S.N., Wilming, L.G., Wunderlich, K.R., Yang, J., Zhao, F.Q., 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324, 522–528.

Burnet, F.M., 1976. A modification of Jerne's theory of antibody production using the concept of clonal selection. CA. Cancer J. Clin. 26, 119–121.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55, 611–622.

Butler, J.E., Sinkora, M., 2013. The enigma of the lower gut-associated lymphoid tissue (GALT). J. Leukoc. Biol. 94, 259–270.

Butler, J.E., Sun, J., Navarro, P., 1996. The swine Ig heavy chain locus has a single JH and no identifiable IgD. Int. Immunol. 8, 1897–1904.

Butler, J.E., Sun, X., Wertz, N., Lager, K.M., Chaloner, K., Urban, J., Jr, 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, 483–494.

Butler, J.E., Weber, P., Sinkora, M., Sun, J., Ford, S.J., Christenson, R.K., 2000. Antibody repertoire development in fetal and neonatal piglets. II. Characterization of heavy chain complementarity-determining region 3 diversity in the developing fetus. J. Immunol. Baltim. Md 1950 165, 6999–7010.

Carlens, O., 1928. Studien über das lymphatische Gewebe des Darmkanals bei einigen Haustieren, mit besonderer Berücksichtigung der embryonalen Entwicklung, der Mengenverhältnisse und der Altersinvolution dieses Gewebes im Dünndarm des Rindes. Z. Für Anat. Entwicklungsgeschichte 86, 393–493.

Chang, L.M., Bollum, F.J., 1971. Deoxynucleotide-polymerizing enzymes of calf thymus gland. V. Homogeneous terminal deoxynucleotidyl transferase. J. Biol. Chem. 246, 909–916.

Chang, L.M., Bollum, F.J., 1986. Molecular biology of terminal transferase. CRC Crit. Rev. Biochem. 21, 27–52.

Chaudhuri, J., Alt, F.W., 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat. Rev. Immunol. 4, 541–552.

Chen, L., Li, M., Li, Q., Yang, X., An, X., Chen, Y., 2008. Characterization of the bovine immunoglobulin lambda light chain constant IGLC genes. Vet. Immunol. Immunopathol. 124, 284–294.

Chevillard, C., Ozaki, J., Herring, C.D., Riblet, R., 2002. A three-megabase yeast artificial chromosome contig spanning the C57BL mouse Igh locus. J. Immunol. Baltim. Md 1950 168, 5659–5666.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., 1989. Conformations of immunoglobulin hypervariable regions. Nature 342, 877–883.

Chowdhary, B.P., Frönicke, L., Gustavsson, I., Scherthan, H., 1996. Comparative analysis of the cattle and human genomes: detection of ZOO-FISH and gene mapping-based chromosomal homologies. Mamm. Genome Off. J. Int. Mamm. Genome Soc. 7, 297–302.

Chung, H., Pamp, S.J., Hill, J.A., Surana, N.K., Edelman, S.M., Troy, E.B., Reading, N.C., Villablanca, E.J., Wang, S., Mora, J.R., Umesaki, Y., Mathis, D., Benoist, C., Relman, D.A., Kasper, D.L., 2012. Gut immune maturation depends on colonization with a host-specific microbiota. Cell 149, 1578–1593.

Cole, G.J., Morris, B., 1973. The lymphoid apparatus of the sheep: its growth, development and significance in immunologic reactions. Adv. Vet. Sci. Comp. Med. 17, 225–263.

Cooper, M.D., 2002. Exploring lymphocyte differentiation pathways. Immunol. Rev. 185, 175–185.

Cooper, M.D., Perey, D.Y., Gabrielsen, A.E., Sutherland, D.E., McKneally, M.F., Good, R.A., 1968. Production of an antibody deficiency syndrome in rabbits by neonatal removal of organized intestinal lymphoid tissues. Int. Arch. Allergy Appl. Immunol. 33, 65–88.

Cooper, M.D., Perey, D.Y., McKneally, M.F., Gabrielsen, A.E., Sutherland, D.E., Good, R.A., 1966a. A mammalian equivalent of the avian bursa of Fabricius. Lancet 1, 1388–1391.

Cooper, M.D., Peterson, R.D., Good, R.A., 1965. Delineation of the thymic and bursal lymphoid systems in the chicken. Nature 205, 143–146.

Cooper, M.D., Raymond, D.A., Peterson, R.D., South, M.A., Good, R.A., 1966b. The functions of the thymus system and the bursa system in the chicken. J. Exp. Med. 123, 75–102.

Cowell, L.G., Kim, H.J., Humaljoki, T., Berek, C., Kepler, T.B., 1999. Enhanced evolvability in immunoglobulin V genes under somatic hypermutation. J. Mol. Evol. 49, 23–26.

Craig, S.W., Cebra, J.J., 1971. Peyer's patches: an enriched source of precursors for IgA-producing immunocytes in the rabbit. J. Exp. Med. 134, 188–200.

Das, S., Nozawa, M., Klein, J., Nei, M., 2008. Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. Immunogenetics 60, 47–55.

Davani, D., Pancer, Z., Ratcliffe, M.J.H., 2014. Ligation of Surface Ig by Gut-Derived Antigen Positively Selects Chicken Bursal and Peripheral B Cells. J. Immunol. Baltim. Md 1950.

Davies, D.R., Padlan, E.A., Sheriff, S., 1990. Antibody-antigen complexes. Annu. Rev. Biochem. 59, 439–473.

Daza, P., Reichenberger, S., Göttlich, B., Hagmann, M., Feldmann, E., Pfeiffer, P., 1996. Mechanisms of nonhomologous DNA end-joining in frogs, mice and men. Biol. Chem. 377, 775–786.

Desiderio, S.V., Yancopoulos, G.D., Paskind, M., Thomas, E., Boss, M.A., Landau, N., Alt, F.W., 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.

Di Noia, J.M., Neuberger, M.S., 2007. Molecular mechanisms of antibody somatic hypermutation. Annu. Rev. Biochem. 76, 1–22.

Dörner, T., Foster, S.J., Farner, N.L., Lipsky, P.E., 1998. Somatic hypermutation of human immunoglobulin heavy chain genes: targeting of RGYW motifs on both DNA strands. Eur. J. Immunol. 28, 3384–3396.

Doyen, N., Boulé, J.-B., Rougeon, F., Papanicolaou, C., 2004. Evidence that the long murine terminal deoxynucleotidyltransferase isoform plays no role in the control of V(D)J junctional diversity. J. Immunol. Baltim. Md 1950 172, 6764–6767.

Doyen, N., d' Andon, M.F., Bentolila, L.A., Nguyen, Q.T., Rougeon, F., 1993. Differential splicing in mouse thymus generates two forms of terminal deoxynucleotidyl transferase. Nucleic Acids Res. 21, 1187–1191.

Drexler, H.G., Sperling, C., Ludwig, W.D., 1993. Terminal deoxynucleotidyl transferase (TdT) expression in acute myeloid leukemia. Leukemia 7, 1142–1150.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A., 2010. Geneious. Biomatters, Aucland.

Early, P., Huang, H., Davis, M., Calame, K., Hood, L., 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. Cell 19, 981–992.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Eisen, H.N., Reilly, E.B., 1985. Lambda chains and genes in inbred mice. Annu. Rev. Immunol. 3, 337–365.
Eisen, H.N., Siskind, G.W., 1964. Variations in affinities of antibodies during the immune response. Biochemistry (Mosc.) 3, 996–1008.
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. Dev. Comp. Immunol. 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 Immunol. 10, 22.
Ekman, A., Pessa-Morikawa, T., Liljavirta, J., Niku, M., Iivanainen, A., 2010. B-cell development in bovine fetuses proceeds via a pre-B like cell in bone marrow and lymph nodes. Dev. Comp. Immunol. 34, 896–903.
Feeney, A.J., 1990. Lack of N regions in fetal and neonatal mouse immunoglobulin V-D-J junctional sequences. J. Exp. Med. 172, 1377–1390.
Feeney, A.J., Riblet, R., 1993. DST4: a new, and probably the last, functional DH gene in the BALB/c mouse. Immunogenetics 37, 217–221.
Flajnik, M.F., 2002. Comparative analyses of immunoglobulin genes: surprises and portents. Nat. Rev. Immunol. 2, 688–698.
Flajnik, M.F., Kasahara, M., 2010. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat. Rev. Genet. 11, 47–59.
Fleming, A., 1922. On a remarkable bacteriolytic element found in tissues and secretions. Proc. R. Soc. B 21, 463–480.
Foote, J., Winter, G., 1992. Antibody framework residues affecting the conformation of the hypervariable loops. J. Mol. Biol. 224, 487–499.
Friedberg, S.H., Weissman, I.L., 1974. Lymphoid tissue architecture. II. Ontogeny of peripheral T and B cells in mice: evidence against Peyer's patches as the site of generation of B cells. J. Immunol. Baltim. Md 1950 113, 1477–1492.
Frieder, D., Larijani, M., Collins, C., Shulman, M., Martin, A., 2009. The concerted action of Msh2 and UNG stimulates somatic hypermutation at A . T base pairs. Mol. Cell. Biol. 29, 5148–5157.
Galler, G.R., Mundt, C., Parker, M., Pelanda, R., Mårtensson, I.-L., Winkler, T.H., 2004. Surface mu heavy chain signals down-regulation of the V(D)J-recombinase machinery in the absence of surrogate light chain components. J. Exp. Med. 199, 1523–1532.
Gauss, G.H., Lieber, M.R., 1996. Mechanistic constraints on diversity in human V(D)J recombination. Mol. Cell. Biol. 16, 258–269.
Gerber, H.A., Morris, B., Trevella, W., 1986. The role of gut-associated lymphoid tissues in the generation of immunoglobulin-bearing lymphocytes in sheep. Aust. J. Exp. Biol. Med. Sci. 64 (Pt 3), 201–213.
Giudicelli, V., Chaume, D., Bodmer, J., Müller, W., Busin, C., Marsh, S., Bontrop, R., Marc, L., Malik, A., Lefranc, M.P., 1997. IMGT, the international ImMunoGeneTics database. Nucleic Acids Res. 25, 206–211.
Giudicelli, V., Lefranc, M.P., 1999. Ontology for immunogenetics: the IMGT-ONTOLOGY. Bioinforma. Oxf. Engl. 15, 1047–1054.
Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G., Mehta, G.R., Ni, I., Mei, L., Sundar, P.D., Day, G.M.R., Cox, D., Rajpal, A., Pons, J., 2009. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc. Natl. Acad. Sci. U. S. A. 106, 20216–20221.
Glick, B., 1970. The Bursa of Fabricius: A Central Issue. BioScience 20, 602–604.

Glick, B., Chang, T.S., Jaap, R.G., 1956. The Bursa of Fabricius and Antibody Production. Poult. Sci. 35, 224–225.
Gontier, E., Ayrault, O., Godet, I., Nau, F., Ladevèze, V., 2005. Developmental progression of immunoglobulin heavy chain diversity in sheep. Vet. Immunol. Immunopathol. 103, 31–51.
Gore, S.D., Kastan, M.B., Civin, C.I., 1991. Normal human bone marrow precursors that express terminal deoxynucleotidyl transferase include T-cell precursors and possible lymphoid stem cells. Blood 77, 1681–1690.
Gourley, T.S., Wherry, E.J., Masopust, D., Ahmed, R., 2004. Generation and maintenance of immunological memory. Semin. Immunol. 16, 323–333.
Grawunder, U., Leu, T.M., Schatz, D.G., Werner, A., Rolink, A.G., Melchers, F., Winkler, T.H., 1995. Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement. Immunity 3, 601–608.
Green, B., Belcheva, A., Nepal, R.M., Boulianne, B., Martin, A., 2011. The mismatch repair pathway functions normally at a non-AID target in germinal center B cells. Blood 118, 3013–3018.
Griebel, P.J., Hein, W.R., 1996. Expanding the role of Peyer's patches in B-cell ontogeny. Immunol. Today 17, 30–39.
Hackney, J.A., Misaghi, S., Senger, K., Garris, C., Sun, Y., Lorenzo, M.N., Zarrin, A.A., 2009. DNA targets of AID evolutionary link between antibody somatic hypermutation and class switch recombination. Adv. Immunol. 101, 163–189.
Hayes, H.C., Petit, E.J., 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 Off. J. Int. Mamm. Genome Soc. 4, 207–210.
Hole, N.J., Young-Cooper, G.O., Mage, R.G., 1991. Mapping of the duplicated rabbit immunoglobulin kappa light chain locus. Eur. J. Immunol. 21, 403–409.
Hornef, M.W., Wick, M.J., Rhen, M., Normark, S., 2002. Bacterial strategies for overcoming host innate and adaptive immune responses. Nat. Immunol. 3, 1033–1040.
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.
Hozumi, N., Tonegawa, S., 1976. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc. Natl. Acad. Sci. U. S. A. 73, 3628–3632.
Hsu, E., 2009. V(D)J recombination: of mice and sharks. Adv. Exp. Med. Biol. 650, 166–179.
Janeway, C.A., Jr, 1992. The immune system evolved to discriminate infectious nonself from noninfectious self. Immunol. Today 13, 11–16.
Janeway, C.A., Jr, Medzhitov, R., 2002. Innate immune recognition. Annu. Rev. Immunol. 20, 197–216.
Jenne, C.N., Kennedy, L.J., McCullagh, P., Reynolds, J.D., 2003. A new model of sheep Ig diversification: shifting the emphasis toward combinatorial mechanisms and away from hypermutation. J. Immunol. Baltim. Md 1950 170, 3739–3750.
Jerne, N.K., 1955. The natural-selection theory of antibody formation. Proc. Natl. Acad. Sci. U. S. A. 41, 849–857.
Jiricny, J., 2006. The multifaceted mismatch-repair system. Nat. Rev. Mol. Cell Biol. 7, 335–346.
Kabat, E.A., Wu, T.T., Bilofsky, H., 1976. Sequences of Proteins of Immunological Interest. Bolt Beranek and Newman Inc.
Kaushik, A., Shojaei, F., Saini, S.S., 2002. Novel insight into antibody diversification from cattle. Vet. Immunol. Immunopathol. 87, 347–350.
Kingzette, M., Spieker-Polet, H., Yam, P.C., Zhai, S.K., Knight, K.L., 1998. Trans-chromosomal recombination within the Ig heavy chain switch region in B lymphocytes. Proc. Natl. Acad. Sci. U. S. A. 95, 11840–11845.

Klemm, L., Duy, C., Iacobucci, I., Kuchen, S., von Levetzow, G., Feldhahn, N., Henke, N., Li, Z., Hoffmann, T.K., Kim, Y., Hofmann, W.-K., Jumaa, H., Groffen, J., Heisterkamp, N., Martinelli, G., Lieber, M.R., Casellas, R., Müschen, M., 2009. The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. Cancer Cell 16, 232–245.

Knight, K.L., Crane, M.A., 1994. Generating the antibody repertoire in rabbit. Adv. Immunol. 56, 179–218.

Knight, K.L., Kingzette, M., Crane, M.A., Zhai, S.K., 1995. Transchromosomally derived Ig heavy chains. J. Immunol. Baltim. Md 1950 155, 684–691.

Koti, M., Kataeva, G., Kaushik, A.K., 2008. Organization of DH-gene locus is distinct in cattle. Dev. Biol. (Basel) 132, 307-313.

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, 2119–2128.

Krijger, P.H.L., Langerak, P., van den Berk, P.C.M., Jacobs, H., 2009. Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. J. Exp. Med. 206, 2603–2611.

Kuroiwa, Y., Kasinathan, P., Sathiyaseelan, T., Jiao, J., Matsushita, H., Sathiyaseelan, J., Wu, H., Mellquist, J., Hammitt, M., Koster, J., Kamoda, S., Tachibana, K., Ishida, I., Robl, J.M., 2009. Antigen-specific human polyclonal antibodies from hyperimmunized cattle. Nat. Biotechnol. 27, 173–181.

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

Lafaille, J.J., DeCloux, A., Bonneville, M., Takagaki, Y., Tonegawa, S., 1989. Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V-(D)-J joining. Cell 59, 859–870.

Lanning, D., Zhu, X., Zhai, S.-K., Knight, K.L., 2000. Development of the antibody repertoire in rabbit: Gut-associated lymphoid tissue, microbes, and selection. Immunol. Rev. 175, 214–228.

Larijani, M., Martin, A., 2012. The biochemistry of activation-induced deaminase and its physiological functions. Semin. Immunol. 24, 255–263.

Lebecque, S.G., Gearhart, P.J., 1990. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. J. Exp. Med. 172, 1717–1727.

Lees-Miller, S.P., Anderson, C.W., 1991. The DNA-activated protein kinase, DNA-PK: a potential coordinator of nuclear events. Cancer Cells Cold Spring Harb. N 1989 3, 341–346.

Lee, Y.K., Mazmanian, S.K., 2010. Has the microbiota played a critical role in the evolution of the adaptive immune system? Science 330, 1768–1773.

Lefler, C.F., Bollum, F.J., 1969. Deoxynucleotide-polymerizing enzymes of calf thymus gland. 3. Preparation of poly N-acetyldeoxyguanylate and polydeoxyguanylate. J. Biol. Chem. 244, 594–601.

Lefranc, M.P., 2001a. Nomenclature of the human immunoglobulin lambda (IGL) genes. Exp. Clin. Immunogenet. 18, 242–254.

Lefranc, M.P., 2001b. Nomenclature of the human immunoglobulin kappa (IGK) genes. Exp. Clin. Immunogenet. 18, 161–174.

Lefranc, M.P., 2001c. Nomenclature of the human immunoglobulin genes. Curr. Protoc. Immunol. Ed. John E Coligan Al Appendix 1, Appendix 1P.

Li, H., Cui, X., Pramanik, S., Chimge, N.-O., 2002. Genetic diversity of the human immunoglobulin heavy chain VH region. Immunol. Rev. 190, 53–68.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B., Ohgi, K.A., Zhang, J., Rose, D.W., Fu, X.-D., Glass, C.K., Rosenfeld, M.G., 2009. Nuclear Receptor-Induced Chromosomal Proximity and DNA Breaks Underlie Specific Translocations in Cancer. Cell 139, 1069–1083.

Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., Schatz, D.G., 2008. Two levels of protection for the B cell genome during somatic hypermutation. Nature 451, 841–845.

Liu, Y., Mukhopadhyay, P., Pisano, M.M., Lu, X., Huang, L., Lu, Q., Dean, D.C., 2013. Repression of Zeb1 and hypoxia cause sequential mesenchymal-to-epithelial transition and induction of aid, Oct4, and Dnmt1, leading to immortalization and multipotential reprogramming of fibroblasts in spheres. Stem Cells Dayt. Ohio 31, 1350–1362.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif 25, 402–408.

Li, Y.S., Hayakawa, K., Hardy, R.R., 1993. The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver. J. Exp. Med. 178, 951–960.

Li, Z., Otevrel, T., Gao, Y., Cheng, H.L., Seed, B., Stamato, T.D., Taccioli, G.E., Alt, F.W., 1995. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. Cell 83, 1079–1089.

Löffert, D., Ehlich, A., Müller, 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.

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.

Mage, R.G., Lanning, D., Knight, K.L., 2006. B cell and antibody repertoire development in rabbits: the requirement of gut-associated lymphoid tissues. Dev. Comp. Immunol. 30, 137–153.

Mahowald, G.K., Baron, J.M., Sleckman, B.P., 2008. Collateral damage from antigen receptor gene diversification. Cell 135, 1009–1012.

Martinez-Jean, C., Folch, G., Lefranc, M.P., 2001. Nomenclature and overview of the mouse (Mus musculus and Mus sp.) immunoglobulin kappa (IGK) genes. Exp. Clin. Immunogenet. 18, 255–279.

Matsuda, F., Ishii, K., Bourvagnet, P., Kuma, K. i, Hayashida, H., Miyata, T., Honjo, T., 1998. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J. Exp. Med. 188, 2151–2162.

Matsumoto, Y., Marusawa, H., Kinoshita, K., Endo, Y., Kou, T., Morisawa, T., Azuma, T., Okazaki, I.-M., Honjo, T., Chiba, T., 2007. Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat. Med. 13, 470–476.

Maul, R.W., Saribasak, H., Martomo, S.A., McClure, R.L., Yang, W., Vaisman, A., Gramlich, H.S., Schatz, D.G., Woodgate, R., Wilson, D.M., 3rd, Gearhart, P.J., 2011. Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. Nat. Immunol. 12, 70–76.

Ma, Y., Pannicke, U., Schwarz, K., Lieber, M.R., 2002. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell 108, 781–794.

McGuire, K.L., Vitetta, E.S., 1981. kappa/lambda Shifts do not occur during maturation of murine B cells. J. Immunol. Baltim. Md 1950 127, 1670–1673.

Medzhitov, R., Janeway, C.A., Jr, 1997. Innate immunity: impact on the adaptive immune response. Curr. Opin. Immunol. 9, 4–9.

Mickelsen, S., Snyder, C., Trujillo, K., Bogue, M., Roth, D.B., Meek, K., 1999. Modulation of terminal deoxynucleotidyltransferase activity by the DNA-dependent protein kinase. J. Immunol. Baltim. Md 1950 163, 834–843.

Milstein, C., Neuberger, M.S., Staden, R., 1998. Both DNA strands of antibody genes are hypermutation targets. Proc. Natl. Acad. Sci. U. S. A. 95, 8791–8794.

Montalbano, A., Ogwaro, K.M., Tang, A., Matthews, A.G.W., Larijani, M., Oettinger, M.A., Feeney, A.J., 2003. V(D)J recombination frequencies can be profoundly affected by changes in the spacer sequence. J. Immunol. Baltim. Md 1950 171, 5296–5304.

Motea, E.A., Berdis, A.J., 2010. Terminal deoxynucleotidyl transferase: The story of a misguided DNA polymerase. Biochim. Biophys. Acta - Proteins Proteomics 1804, 1151–1166.

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, K.M., 2012. Janeway's Immunobiology, 8th Edition. Immunobiology: The Immune System, 8th ed. Garland Science.

Mutwiri, G., Watts, T., Lew, L., Beskorwayne, T., Papp, Z., Baca-Estrada, M.E., Griebel, P., 1999. Ileal and jejunal Peyer's patches play distinct roles in mucosal immunity of sheep. Immunology 97, 455–461.

Nguyen, V.K., Su, C., Muyldermans, S., van der Loo, W., 2002. Heavy-chain antibodies in Camelidae; a case of evolutionary innovation. Immunogenetics 54, 39–47.

Niku, M., Ekman, A., Pessa-Morikawa, T., Iivanainen, A., 2006. Identification of major cell types in paraffin sections of bovine tissues. BMC Vet. Res. 2, 5.

Niku, M., Pessa-Morikawa, T., Andersson, L.C., Iivanainen, A., 2002. Oligoclonal Peyer's patch follicles in the terminal small intestine of cattle. Dev. Comp. Immunol. 26, 689–695.

Nuñez, C., Nishimoto, N., Gartland, G.L., Billips, L.G., Burrows, P.D., Kubagawa, H., Cooper, M.D., 1996. B cells are generated throughout life in humans. J. Immunol. Baltim. Md 1950 156, 866–872.

Nussenzweig, A., Nussenzweig, M.C., 2010. Origin of chromosomal translocations in lymphoid cancer. Cell 141, 27–38.

Oettinger, M.A., Schatz, D.G., Gorka, C., Baltimore, D., 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science 248, 1517–1523.

O'Leary, M.A., 2001. The Phylogenetic Position of Cetaceans: Further Combined Data Analyses, Comparisons with the Stratigraphic Record and a Discussion of Character Optimization. Am. Zool. 41, 487–506.

Owen, J.J., Wright, D.E., Habu, S., Raff, M.C., Cooper, M.D., 1977. Studies on the generation of B lymphocytes in fetal liver and bone marrow. J. Immunol. Baltim. Md 1950 118, 2067–2072.

Pages, H., Aboyoun, P., Gentleman, R., DebRoy, S., 2014. Biostrings: String objects representing biological sequences, and matching algorithms [WWW Document]. R Package.

Parker, R., Song, H., 2004. The enzymes and control of eukaryotic mRNA turnover. Nat. Struct. Mol. Biol. 11, 121–127.

Pasman, Y., Saini, S.S., Smith, E., Kaushik, A.K., 2010. Organization and genomic complexity of bovine lambda-light chain gene locus. Vet. Immunol. Immunopathol. 135, 306–313.

Patel, K.P., Khokhar, F.A., Muzzafar, T., James You, M., Bueso-Ramos, C.E., Ravandi, F., Pierce, S., Medeiros, L.J., 2013. TdT expression in acute myeloid leukemia with minimal differentiation is associated with distinctive clinicopathological features and better overall survival following stem cell transplantation. Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc 26, 195–203.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., Nieland, T.J., Root, D.E., Casellas, R., Nussenzweig, M.C., 2010. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell 143, 122–133.

Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., Scharff, M.D., 2008. The biochemistry of somatic hypermutation. Annu. Rev. Immunol. 26, 481–511.

Perera, J., Meng, L., Meng, F., Huang, H., 2013. Autoreactive thymic B cells are efficient antigen-presenting cells of cognate self-antigens for T cell negative selection. Proc. Natl. Acad. Sci. U. S. A. 110, 17011–17016.

Pham, P., Bransteitter, R., Petruska, J., Goodman, M.F., 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature 424, 103–107.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., Reik, W., 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463, 1101–1105.

Pramanik, S., Cui, X., Wang, H.-Y., Chimge, N.-O., Hu, G., Shen, L., Gao, R., Li, H., 2011. Segmental duplication as one of the driving forces underlying the diversity of the human immunoglobulin heavy chain variable gene region. BMC Genomics 12, 78.

Putnam, F.W., 1969. Immunoglobulin structure: variability and homology. Science 163, 633–644.

Putnam, F.W., Shinoda, T., Titani, K., Wikler, M., 1967. Immunoglobulin structure: variation in amino acid sequence and length of human lambda light chains. Science 157, 1050–1053.

Ratcliffe, M.J.H., 2002. B cell development in gut associated lymphoid tissues. Vet. Immunol. Immunopathol. 87, 337–340.

Ratcliffe, M.J.H., 2006. Antibodies, immunoglobulin genes and the bursa of Fabricius in chicken B cell development. Dev. Comp. Immunol. 30, 101–118.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available: http://www.R-project.org/.

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, 283–291.

Reynaud, C.A., Anquez, V., Grimal, H., Weill, J.C., 1987. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell 48, 379–388.

Reynaud, C.A., Anquez, V., Weill, J.C., 1991. The chicken D locus and its contribution to the immunoglobulin heavy chain repertoire. Eur. J. Immunol. 21, 2661–2670.

Reynaud, C.A., Dahan, A., Anquez, V., Weill, J.C., 1989. Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell 59, 171–183.

Reynaud, C.A., Garcia, C., Hein, W.R., Weill, J.C., 1995. Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. Cell 80, 115–125.

Reynaud, C.A., Mackay, C.R., Müller, R.G., Weill, J.C., 1991. Somatic generation of diversity in a mammalian primary lymphoid organ: the sheep ileal Peyer's patches. Cell 64, 995–1005.

Reynolds, J., 1997. The genesis, tutelage and exodus of B cells in the ileal Peyer's patch of sheep. Int. Rev. Immunol. 15, 265–299.

Reynolds, J.D., Morris, B., 1983. The evolution and involution of Peyer's patches in fetal and postnatal sheep. Eur. J. Immunol. 13, 627–635.

Rhee, K.-J., Jasper, P.J., Sethupathi, P., Shanmugam, M., Lanning, D., Knight, K.L., 2005. Positive selection of the peripheral B cell repertoire in gut-associated lymphoid tissues. J. Exp. Med. 201, 55–62.

Rhee, K.-J., Sethupathi, P., Driks, A., Lanning, D.K., Knight, K.L., 2004. Role of commensal bacteria in development of gut-associated lymphoid tissues and preimmune antibody repertoire. J. Immunol. Baltim. Md 1950 172, 1118–1124.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. TIG 16, 276–277.

Roa, S., Avdievich, E., Peled, J.U., Maccarthy, T., Werling, U., Kuang, F.L., Kan, R., Zhao, C., Bergman, A., Cohen, P.E., Edelmann, W., Scharff, M.D., 2008. Ubiquitylated PCNA plays a role in somatic hypermutation and class-switch recombination and is required for meiotic progression. Proc. Natl. Acad. Sci. U. S. A. 105, 16248–16253.

Rogozin, I.B., Kolchanov, N.A., 1992. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim. Biophys. Acta 1171, 11–18.

Rogozin, I.B., Pavlov, Y.I., Bebenek, K., Matsuda, T., Kunkel, T.A., 2001. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. Nat. Immunol. 2, 530–536.

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, 2420–2426.

Saini, S.S., Hein, W.R., Kaushik, A., 1997. A single predominantly expressed polymorphic immunoglobulin V(H) gene family, related to mammalian group, I, clan, II, is identified in cattle. Mol. Immunol. 34, 641–651.

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

Sakano, H., Kurosawa, Y., Weigert, M., Tonegawa, S., 1981. Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy-chain genes. Nature 290, 562–565.

Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., Tonegawa, S., 1980. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. Nature 286, 676–683.

Sandor, Z., Calicchio, M.L., Sargent, R.G., Roth, D.B., Wilson, J.H., 2004. Distinct requirements for Ku in N nucleotide addition at V(D)J- and non-V(D)J-generated double-strand breaks. Nucleic Acids Res. 32, 1866–1873.

Saribasak, H., Gearhart, P.J., 2012. Does DNA repair occur during somatic hypermutation? Semin. Immunol. 24, 287–292.

Schatz, D.G., Oettinger, M.A., Baltimore, D., 1989. The V(D)J recombination activating gene, RAG-1. Cell 59, 1035–1048.

Schroeder, H.W., Jr, Cavacini, L., 2010. Structure and function of immunoglobulins. J. Allergy Clin. Immunol. 125, S41–52.

Schroeder, H.W., Jr, Mortari, F., Shiokawa, S., Kirkham, P.M., Elgavish, R.A., Bertrand, F.E., 3rd, 1995. Developmental regulation of the human antibody repertoire. Ann. N. Y. Acad. Sci. 764, 242–260.

Schwartz, J.C., Lefranc, M.-P., Murtaugh, M.P., 2012a. Evolution of the porcine (Sus scrofa domestica) immunoglobulin kappa locus through germline gene conversion. Immunogenetics. 64:4, 303–311

Schwartz, J.C., Lefranc, M.-P., Murtaugh, M.P., 2012b. Organization, complexity and allelic diversity of the porcine (Sus scrofa domestica) immunoglobulin lambda locus. Immunogenetics 64:5, 399–407.

Sharma, J.M., 1997. The structure and function of the avian immune system. Acta Vet. Hung. 45, 229–238.

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, 61–67.

Sinclair, M.C., Gilchrist, J., Aitken, R., 1997. Bovine IgG repertoire is dominated by a single diversified VH gene family. J. Immunol. Baltim. Md 1950 159, 3883–3889.

Sinkora, M., Stepanova, K., Butler, J.E., Francis, D., Santiago-Mateo, K., Potockova, H., Karova, K., Sinkorova, J., 2011. Ileal Peyer's patches are not necessary for systemic B cell development and maintenance and do not contribute significantly to the overall B cell pool in swine. J. Immunol. Baltim. Md 1950 187, 5150–5161.

Slack, E., Hapfelmeier, S., Stecher, B., Velykoredko, Y., Stoel, M., Lawson, M.A.E., Geuking, M.B., Beutler, B., Tedder, T.F., Hardt, W.-D., Bercik, P., Verdu, E.F., McCoy, K.D., Macpherson, A.J., 2009. Innate and adaptive immunity cooperate flexibly to maintain host-microbiota mutualism. Science 325, 617–620.

Solin, M.L., Kaartinen, M., 1992. Allelic polymorphism of mouse Igh-J locus, which encodes immunoglobulin heavy chain joining (JH) segments. Immunogenetics 36, 306–313.

Su, C., Nguyen, V.K., Nei, M., 2002. Adaptive evolution of variable region genes encoding an unusual type of immunoglobulin in camelids. Mol. Biol. Evol. 19, 205–215.

Sun, J., Butler, J.E., 1996. Molecular characterization of VDJ transcripts from a newborn piglet. Immunology 88, 331–339.

Sun, J., Kacskovics, I., Brown, W.R., Butler, J.E., 1994. Expressed swine VH genes belong to a small VH gene family homologous to human VHIII. J. Immunol. Baltim. Md 1950 153, 5618–5627.

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.

Takahara, K., Hayashi, N., Fujita-Sagawa, K., Morishita, T., Hashimoto, Y., Noda, A., 1994. Alternative splicing of bovine terminal deoxynucleotidyl transferase cDNA. Biosci. Biotechnol. Biochem. 58, 786–787.

Thaiss, C.A., Levy, M., Suez, J., Elinav, E., 2014. The interplay between the innate immune system and the microbiota. Curr. Opin. Immunol. 26, 41–48.

Thai, T.-H., Kearney, J.F., 2004. Distinct and opposite activities of human terminal deoxynucleotidyltransferase splice variants. J. Immunol. Baltim. Md 1950 173, 4009–4019.

Thai, T.-H., Kearney, J.F., 2005. Isoforms of terminal deoxynucleotidyltransferase: developmental aspects and function. Adv. Immunol. 86, 113–136.

Thiebe, R., Schäble, K.F., Bensch, A., Brensing-Küppers, J., Heim, V., Kirschbaum, T., Mitlöhner, H., Ohnrich, M., Pourrajabi, S., Röschenthaler, F., Schwendinger, J., Wichelhaus, D., Zocher, I., Zachau, H.G., 1999. The variable genes and gene families of the mouse immunoglobulin kappa locus. Eur. J. Immunol. 29, 2072–2081.

Titani, K., Wikler, M., Putnam, F.W., 1967. Evolution of immunoglobulins: structural homology of kappa and lambda Bence Jones proteins. Science 155, 828–835.

Tobin-Janzen, T.C., Womack, J.E., 1992. Comparative mapping of IGHG1, IGHM, FES, and FOS in domestic cattle. Immunogenetics 36, 157–165.

Tomlinson, S., 1993. Complement defense mechanisms. Curr. Opin. Immunol. 5, 83–89.

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

Turchin, A., Hsu, E., 1996. The generation of antibody diversity in the turtle. J. Immunol. Baltim. Md 1950 156, 3797–3805.

Uchiyama, Y., Takeuchi, R., Kodera, H., Sakaguchi, K., 2009. Distribution and roles of X-family DNA polymerases in eukaryotes. Biochimie 91, 165–170.

Uduman, M., Yaari, G., Hershberg, U., Stern, J.A., Shlomchik, M.J., Kleinstein, S.H., 2011. Detecting selection in immunoglobulin sequences. Nucleic Acids Res. 39, W499–504.

Verma, S., Goldammer, T., Aitken, R., 2010. Cloning and expression of activation induced cytidine deaminase from Bos taurus. Vet. Immunol. Immunopathol. 134, 151–159.

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, 1379–1393.

Watson, F.L., Püttmann-Holgado, R., Thomas, F., Lamar, D.L., Hughes, M., Kondo, M., Rebel, V.I., Schmucker, D., 2005. Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science 309, 1874–1878.

Weinstein, P.D., Mage, R.G., Anderson, A.O., 1994. The appendix functions as a mammalian bursal equivalent in the developing rabbit. Adv. Exp. Med. Biol. 355, 249–253.

Wei, Y.F., Robins, P., Carter, K., Caldecott, K., Pappin, D.J., Yu, G.L., Wang, R.P., Shell, B.K., Nash, R.A., Schär, P., 1995. Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. Mol. Cell. Biol. 15, 3206–3216.

Wertz, N., Vazquez, J., Wells, K., Sun, J., Butler, J.E., 2013. Antibody repertoire development in fetal and neonatal piglets. XII. Three IGLV genes comprise 70% of the pre-immune repertoire and there is little junctional diversity. Mol. Immunol. 55, 319–328.

Wilson, D.M., 3rd, Bohr, V.A., 2007. The mechanics of base excision repair, and its relationship to aging and disease. DNA Repair 6, 544–559.

Yaari, G., Uduman, M., Kleinstein, S.H., 2012. Quantifying selection in high-throughput Immunoglobulin sequencing data sets. Nucleic Acids Res. 40, e134.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H., Robbiani, D.F., McBride, K., Nussenzweig, M.C., Casellas, R., 2011. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat Immunol 12, 62–69.

Yasuda, M., Jenne, C.N., Kennedy, L.J., Reynolds, J.D., 2006. The sheep and cattle Peyer's patch as a site of B-cell development. Vet. Res. 37, 401–415.

Yoneda, M., Bollum, F.J., 1965. Deoxynucleotide-polymerizing enzymes of calf thymus gland. I. Large scale purification of terminal and replicative deoxynucleotidyl transferases. J. Biol. Chem. 240, 3385–3391.

Yu, K., Huang, F.-T., Lieber, M.R., 2004. DNA Substrate Length and Surrounding Sequence Affect the Activation-induced Deaminase Activity at Cytidine. J. Biol. Chem. 279, 6496 –6500.

Zan, H., Casali, P., 2013. Regulation of Aicda expression and AID activity. Autoimmunity 46, 83–101.

Zhang, S.-M., Adema, C.M., Kepler, T.B., Loker, E.S., 2004. Diversification of Ig superfamily genes in an invertebrate. Science 305, 251–254.

Zhao, Y., Jackson, S.M., Aitken, R., 2006. The bovine antibody repertoire. Dev. Comp. Immunol. 30, 175–186.

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.