



# The Camel Adaptive Immune Receptors Repertoire as a Singular Example of Structural and Functional Genomics

Salvatrice Ciccarese<sup>1\*</sup>, Pamela A. Burger<sup>2</sup>, Elena Ciani<sup>3</sup>, Vito Castelli<sup>1</sup>,  
Giovanna Linguiti<sup>1</sup>, Martin Plasil<sup>4,5</sup>, Serafina Massari<sup>6</sup>, Petr Horin<sup>4,5</sup> and Rachele Antonacci<sup>1</sup>

<sup>1</sup> Department of Biology, University of Bari "Aldo Moro," Bari, Italy, <sup>2</sup> Research Institute of Wildlife Ecology, Vetmeduni Vienna, Vienna, Austria, <sup>3</sup> Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari "Aldo Moro," Bari, Italy, <sup>4</sup> Department of Animal Genetics, Faculty of Veterinary Medicine, University of Veterinary and Pharmaceutical Sciences, Brno, Czechia, <sup>5</sup> CEITEC-VFU, University of Veterinary and Pharmaceutical Sciences, RG Animal Immunogenomics, Brno, Czechia, <sup>6</sup> Department of Biological and Environmental Science and Technologies, University of Salento, Lecce, Italy

## OPEN ACCESS

### Edited by:

James J. Cai,  
Texas A&M University, United States

### Reviewed by:

David Nicholas Olivieri,  
Universidad de Vgjo, Spain  
Véronique Giudicelli,  
IMGT, the international ImMuno  
GeneTics information system<sup>®</sup>,  
France

### \*Correspondence:

Salvatrice Ciccarese  
salvatricemaria.ciccarese@uniba.it

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Genetics

**Received:** 21 December 2018

**Accepted:** 18 September 2019

**Published:** 17 October 2019

### Citation:

Ciccarese S, Burger PA, Ciani E,  
Castelli V, Linguiti G, Plasil M,  
Massari S, Horin P and Antonacci R  
(2019) The Camel Adaptive Immune  
Receptors Repertoire as a Singular  
Example of Structural and  
Functional Genomics.  
Front. Genet. 10:997.  
doi: 10.3389/fgene.2019.00997

The adaptive immune receptors repertoire is highly plastic, with its ability to produce antigen-binding molecules and select those with high affinity for their antigen. Species have developed diverse genetic and structural strategies to create their respective repertoires required for their survival in the different environments. Camelids, until now, considered as a case of evolutionary innovation because of their only heavy-chain antibodies, represent a new mammalian model particularly useful for understanding the role of diversity in the immune system function. Here, we review the structural and functional characteristics and the current status of the genomic organization of camel immunoglobulins (IG) or antibodies,  $\alpha/\beta$  and  $\gamma/\delta$  T cell receptors (TR), and major histocompatibility complex (MHC). In camelid humoral response, in addition to the conventional antibodies, there are IG with "only-heavy-chain" (no light chain, and two identical heavy gamma chains lacking CH1 and with a VH domain designated as VHH). The unique features of these VHH offer advantages in biotechnology and for clinical applications. The TRG and TRD rearranged variable domains of *Camelus dromedarius* (Arabian camel) display somatic hypermutation (SHM), increasing the intrinsic structural stability in the  $\gamma/\delta$  heterodimer and influencing the affinity maturation to a given antigen similar to immunoglobulin genes. The SHM increases the dromedary  $\gamma/\delta$  repertoire diversity. In *Camelus* genus, the general structural organization of the TRB locus is similar to that of the other artiodactyl species, with a pool of *TRBV* genes positioned at the 5' end of three in tandem D-J-C clusters, followed by a single *TRBV* gene with an inverted transcriptional orientation located at the 3' end. At the difference of TRG and TRD, the diversity of the TRB variable domains is not shaped by SHM and depends from the classical combinatorial and junctional diversity. The MHC locus is located on chromosome 20 in *Camelus dromedarius*. Cytogenetic and comparative whole genome analyses revealed the order of the three major regions "Centromere-ClassII-ClassIII-ClassI". Unexpectedly low extent of polymorphisms and haplotypes was observed in all Old World camels despite different geographic origins.

**Keywords:** Immunome, Old World camelids, *Camelus bactrianus*, *Camelus dromedarius*, *Camelus ferus*, Immunoglobulins, T cell receptors, major histocompatibility complex

## INTRODUCTION

In vertebrates, B and T lymphocytes together with the antigen-presenting cells play central roles in the adaptive immune system. They respond to a large variety of antigens that are specifically recognized through highly specialized proteins: immunoglobulins (IG) or antibodies in B cells, and T cell receptors (TR) in T cells (Lefranc, 2014a). The common shape of IG is the tetrameric structure, two identical dimers each made up of an IG heavy (H) chain, and an IG light (L) chain (either IG light kappa (IGK) or IG light lambda (IGL) chains). The TR is a heterodimeric receptor that may occur in two types:  $\alpha\beta$  (TR-Alpha\_Beta, composed of a T cell receptor alpha (TRA) and a T cell receptor beta (TRB) chain) and  $\gamma\delta$  (TR-Gamma\_Delta, composed of a T cell receptor gamma (TRG) and a T cell receptor delta (TRD) chain).

Each chain contains a variable domain and a constant region (Lefranc, 2014a). The variable domain forms the antigen-binding site and it is generated during B or T lymphocyte development by a sequential gene rearrangement at the DNA level of the *variable (V)* and *joining (J)* genes of the IGK or IGL, TRG, and TRA loci, and *V*, *diversity (D)*, and *J* genes of the IGH, TRB and TRD loci. After transcription, the rearranged V-(D)-J sequence is spliced to the *constant (C)* gene (Lefranc and Lefranc, 2001; Lefranc and Lefranc, 2001b; Jung and Alt, 2004). The resulting IG and TR chains are proteins with a variable (V) domain at the N-terminal end. Each V domain comprises nine beta sheets forming four framework regions or FR, which support three hypervariable loops (complementarity determining regions or CDR) (Lefranc 2014; Lefranc and Lefranc 2019). CDR1 and CDR2 are encoded by the germline V gene; the third, CDR3, results from the V-(D)-J rearrangement. The six CDR loops of the paired V domains of the IG (VH and VL) and those of the TR gamma/delta (V-gamma and V-delta) contribute to the antigen-binding site. In contrast in the TR alpha/beta, only the two CDR3 principally recognize and bind the antigenic peptide bound to major histocompatibility (MH) proteins of class I (MH1) or class II (MH2), whereas the germline-encoded CDR1 and CDR2 loops mainly contact the helices of the MH proteins (Lefranc, 2014a).

For IGH chains, the rearranged variable domain VH will initially be expressed together with IGHM, the most J-proximal *IGHC* gene, leading to the IgM class synthesis. After the encounter with the antigen and B cell activation and with T cell cooperation, a further DNA recombination event, referred to as class switch recombination, can take place in B cells, resulting in replacement of the IGHM by one of the gene of the other *IGHC* gene subgroups, namely, *IGHG*, *IGHE*, or *IGHA*. This process leads to the expression of a new H chain with different effector functions, thereby shifting the IG from the IgM class to one of the IgG or IgA subclasses or to IgE class (Lefranc and Lefranc, 2001)

The genes encoded for each IG or TR chain are located in different loci. There are three IG loci (IGK, IGL, and IGH) and four TR loci (TRA, TRB, TRG, and TRD) (<http://www.imgt.org/IMGTrepertoire/LocusGenes>). The TRA and TRD loci occupy the same chromosome location, being the TRD inserted into the TRA locus. The number of the *V*, *D*, and *J* genes within loci as well as their genomic organization can vary significantly

among species. This implies that the gene content is an important element in generating the full extent of the IG and TR repertoires, providing the species with the ability to adapt to its own habitat to defend against infections from a large variety of pathogens.

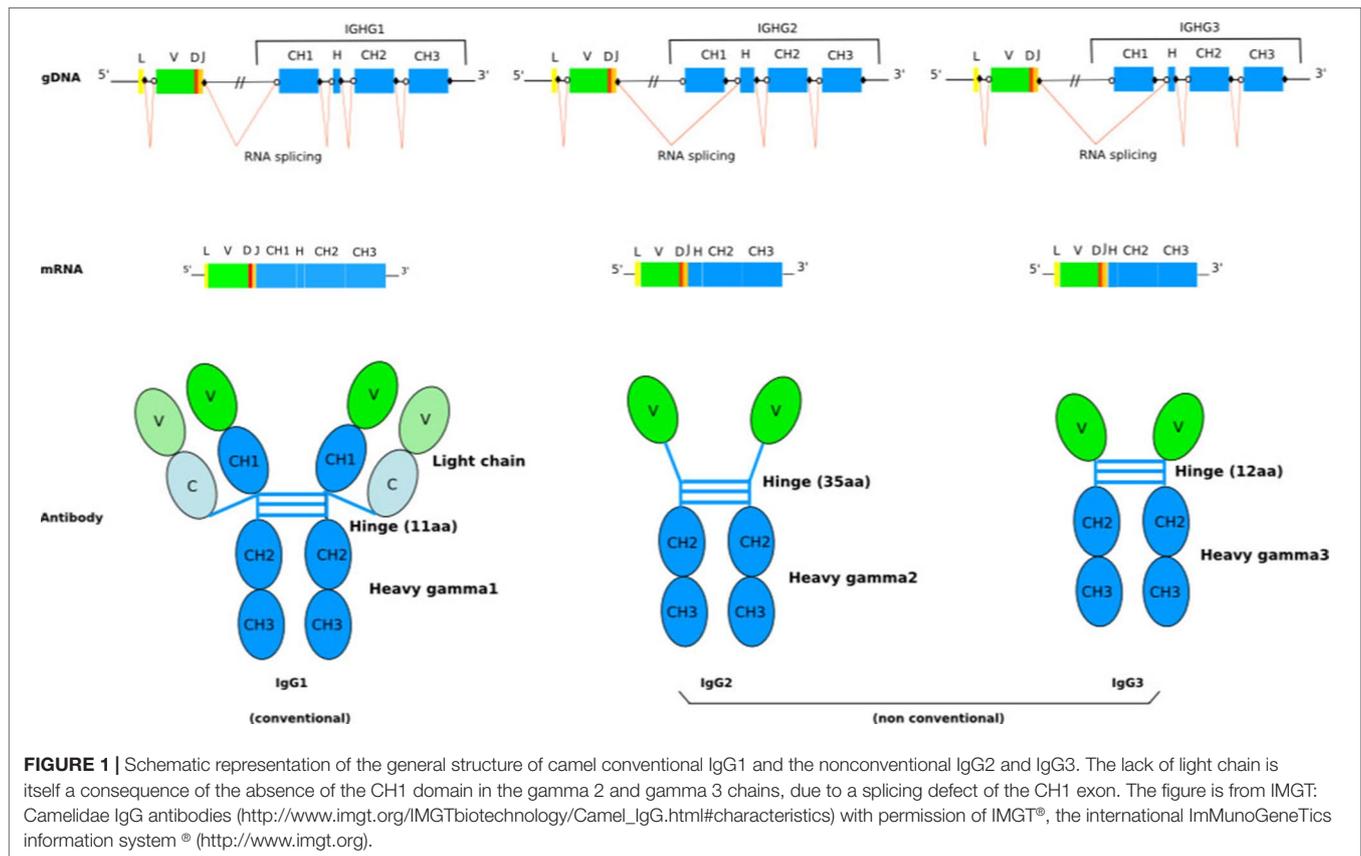
The complex response of camelids to different pathogens has been investigated over nearly three decades. In this focused review, we provide a comprehensive overview based on the search of key publications from the more recent literature on the genomic and functional characteristics of the IG, TR, and MH molecules in camelids.

## THE CAMEL IMMUNOGLOBULIN: A DICHOTOMOUS ADAPTIVE HUMORAL IMMUNE SYSTEM

The humoral immune system of camelids (i.e., *Camelus bactrianus* (Bactrian camel), *C. dromedarius* (Arabian camel), *C. ferus* (Wild Bactrian camel), *Lama glama* (llama), *L. guanaco* (guanaco), *Vicugna pacos* (alpaca), and *V. vicugna* (vicugna) have the particularity of including, in addition to the conventional tetrameric IgG (IgG1 subclass) composed of two identical heavy (H) and two identical light (L) chains connected by disulphide bonds, functional homodimeric IgG (IgG2 and IgG3 subclasses) lacking L chains and, therefore, comprising only two identical H chains (only-heavy-chain antibodies hcAb, or hcIG) (**Figure 1**) (Hamers-Casterman et al., 1993). Sequence and structure analysis revealed a number of characteristic features of camelid hcAb (Muyldermans and Lauwereys, 1999) to make the H chain functional in antigen binding in absence of the L chain. Besides their dissimilar IG chain content, tetrameric, and homodimeric IgG display distinct H chains, with the H chain within hcAb composed of three instead of four globular domains. Biochemical and cDNA sequence analyses have shown that the C region (CHH) of the homodimeric IgG lacks the first domain, CH1, which normally binds to the L chain. This region is eliminated by splicing during mRNA processing due to a point mutation on the donor-splicing site present in the first C exon/intron boundary (Nguyen et al., 1999; Woolven et al., 1999). Hence, the variable domain is joined directly to the hinge region in hcAb (Lefranc, 2014b) (**Figure 1**). The hinge region itself can be different from conventional IG heavy chains (Henry et al., 2019)

### Structural Features and Binding Properties of the VHH Domain

Since the hcAb do not contain L chains, the antigen-binding site is reduced to a single domain (referred to as VHH) that resembles the structure of the H chain variable domain (VH) of the tetrameric IgG. However, the VHH domains display remarkable amino acid differences in positions that are conserved in the conventional VH domains. In FR2-IMGT, amino acids highly conserved across species, located at positions 42, 49, 50, and 52, according to the IMGT unique numbering (Lefranc et al., 2003; Lefranc, 2011a; Lefranc, 2011b), which, in conventional VH domains, form the hydrophobic surface associating with VL (Chothia et al., 1985), are changed to more hydrophilic amino acids (Hamers-Casterman et al., 1993; Muyldermans et al., 1994;



Vu et al., 1997; Tillib et al., 2014; Li et al., 2016; Brooks et al., 2018). The hydrophobic to hydrophilic amino acid changes at these positions make impossible the association of the VHH with a conventional VL domain and undoubtedly help in the solubility behaviour of the hcAb (Davies and Riechmann, 1994; Muyldermans et al., 1994). Moreover, the amino acid change, in most of the Bactrian and Arabian camel V-REGION, of Leu (L) 12 (IMGT numbering) > Ser (S) in FR1, which is seen as an adaptation to the absence of the CH1 domain, helps the solubility of the hcAb. However, the VHH are notably more conserved in sequence and structure across their FR regions than conventional VH domains (Mitchell and Colwell, 2018b).

CDR1-IMGT and CDR2-IMGT of the IGHV encoding VH or VHH domains are highly similar in term of sequences and lengths (IMGT Repertoire (IG and ITR) > Protein displays Arabian camel (*Camelus dromedarius*)).<sup>1</sup>

Conversely, the CDR3 loop of the hcAb tends to be longer and more variable than that of the conventional IgG (Muyldermans, 2013; Mitchell and Colwell, 2018a; Mitchell and Colwell, 2018b). This difference is much higher in *Camelus bactrianus* than that reported for llama and dromedary (Henry et al., 2019). A longer CDR3 in the hcAb could potentially greatly increase both the sequence and shape diversity of the paratope. As recently reported (Mitchell and Colwell, 2018a), the CDR3 loop is more frequently in contact with the antigen than CDR1 and CDR2 loops. Therefore,

CDR3 plays a more dominant role in determining interaction specificity. Moreover, in Bactrian as well as in dromedary camel, the long rearranged CDR3 of VHH in most cases harbours a cysteine that forms a disulphide bond with another additional cysteine located either in CDR1, or on position 50 (*Camelus dromedarius* (Camdro) IGHV) or 55 (*Lama glama* (Lamgla) IGHV) in FR2-IMGT (IMGT Repertoire > IMGT Protein display)<sup>1</sup> (Hamers-Casterman et al., 1993; Muyldermans et al., 1994; Vu et al., 1997). This second disulphide bond stabilizes the VHH domain and fixes the long CDR3 loop into an optimal conformation, increasing the affinity for the antigen (Govaert et al., 2012).

The VHH domain of the hcAb is fully capable of antigen binding. It recognizes a broad range of epitopes with high affinity. Moreover, the hcAb repertoire, largely diversified by extensive somatic hypermutation (SHM) involving the variable domains, results in novel and unusual paratopes different from those of conventional IgG (Nguyen et al., 2000; Nguyen et al., 2001). Many VHH domains are competitive enzyme inhibitors since they interact specifically with the active site of enzymes that, in general, is of low antigenicity for the conventional VH-VL domains (Lauwereys et al., 1998).

The unique characteristics of the VHH and their straightforward bacterial expression have made them of particular interest in biotechnological and pharmaceutical applications. In recent years, the industrialization of camel VHH domains (designated as single-domain antibodies or “nanobodies” for their format) has produced a great expansion of their use

<sup>1</sup> <http://www.imgt.org>

(Muyldermans et al., 2009; Hassanzadeh-Ghassabeh et al., 2013; Helma et al., 2015; Fernandes, 2018), and in 2018, a diabody of two VHH humanized from *Lama glama*, caplaximab, an anti-von Willebrand factor (VWF) A1 domain, has been approved by the European Medicine Agency (EMA) in Europe and by the Food and Drug Administration (FDA) in 2019 for treatment of acquired thrombotic thrombocytopenia purpura (TTP) [IMGT/mAb-DB, (Lefranc et al., 2015)]. Beyond their application as therapeutics to treat human diseases (Muyldermans, 2013; Rissiek et al., 2014; Steeland et al., 2016), nanobodies have become a valuable research tool. For example, they are used as affinity reagents to assist the crystallization process, to detect antigen trafficking inside living cells, to interfere with protein-protein interactions, and to direct proteins to degradation (Loris et al., 2003; Hassanzadeh-Ghassabeh et al., 2013; Helma et al., 2015; Beghein and Gettemans, 2017; Baudisch et al., 2018; Schumacher et al., 2018).

## The Camel IGH Locus

To understand the molecular mechanisms governing the formation of tetrameric and homodimeric IgG in camelids, the characterization of the organization of the genes that encode them is an essential step. Although hcAb have been extensively investigated, to date, there has not been a comprehensive analysis of the repertoire based on high-throughput sequencing (Jirimutu et al., 2012; Li et al., 2016; Ali et al., 2019; Henry et al., 2019), but most efforts have been based upon low throughput sequence analysis, and the reports trying to analyse and describe the complete immune repertoire of camel hcAb are limited. Although the annotated data in public databases are limited, the available sequences show that the camelid *IGHV* genes, which encode VH and VHH, belong to the IGHV3 subgroup (IMGT Repertoire (IG and TR) 2. Proteins and alleles > Protein displays Arabian camel (*Camelus dromedarius*); ibid:alpaca (*Vicugna pacos*) IGHV; and ibid:llama (*Lama glama*) IGHV)<sup>1</sup>. The high percentage of identity between IGHV encoding VH or VHH classifies them in the same IGHV3 subgroup, the differences between them being the characteristic amino acid changes at the four IMGT positions 42, 49, 50, and 52. It is, therefore, the IGHV, which is involved in the rearrangement which determines if the expressed domain is VH or VHH. The constant region of the *Camelus dromedarius* H-gamma1 chains is encoded by the *IGHG1* gene, whereas the constant region of the H-gamma2 and H-gamma3 chains are encoded by *IGHG2* and *IGHG3* genes, which both have a splicing defect of the CH1 DONOR-SPLICE leading to the absence of the CH1 in the transcript, although this CH1 sequence is present in the genomic DNA (**Figure 1**) (Lefranc, 2014b) (IMGT Biotechnology > Antibody camelization > Characteristics of the camelidae (camel, llama) antibody synthesis)<sup>1</sup>. The presence of a point mutation (G to A) in the putative donor splicing site flanking the first C exon and the specific hinge region makes it possible to distinguish the *IGHG* genes encoding the constant region of hcAb chain and the *IGHG* genes encoding the constant region of conventional antibody heavy chain (Vu et al., 1997; Nguyen et al., 1999; Woolven et al., 1999). In Arabian camel,

three *IGHG* genes have been identified (IMGT Repertoire (IG and TR) > Gene table > Gene table: Arabian camel (*Camelus dromedarius*) IGHG)<sup>1</sup>. Four *IGHG* genes (*IGHG1A*, *IGHG1B*, *IGHG2B*, and *IGHG2C*, these last two genes without CH1 in the alpaca genome) are present in the alpaca and llama genomes (Achour et al., 2008; Henry et al., 2019) (IMGT Repertoire (IG and TR) > Gene table > Gene table: Alpaca (*Vicugna pacos*) IGHG)<sup>1</sup>, even if other studies have also identified *IGHG2A* and *IGHG3* genes in the expressed IgG repertoire of llama (De Genst et al., 2006; Saccodossi et al., 2012).

*IGHV* genes encoding VH and VHH domains as well as the *IGHG* genes encoding the C-REGION of conventional IG (*IGHG1*) and only-heavy-chain IG (*IGHG2* and *IGHG3*) are in an intermixed conformation, in V and C clusters, respectively, in a single IGH locus (**Figure 1**). *IGHV* genes encoding VH and VHH domains recombine with the same *IGHD* and *IGHJ* genes. The identification of identical *IGHD* and *IGHJ* genes in VH and VHH cDNAs suggests the common use of the *IGHD* and *IGHJ* genes (Harmsen et al., 2000; Nguyen et al., 2000; Achour et al., 2008). The *IGHV* genes are designed in three or four subgroups based on their degree of homology with human IGHV subgroups. There are IGHV subgroups that contain exclusively classical *IGHV* genes, whereas one subgroup, IGHV3, contains both classical *IGHV* genes and *IGHV* genes with the FR2-IMGT camelid hydrophilic amino acid in 50 (together with 42, 49, and/or 52) (Lefranc and Lefranc, 2019), which are rearranged and expressed in VH and VHH domains, respectively (Harmsen et al., 2000; Achour et al., 2008; Deschacht et al., 2010; Griffin et al., 2014). The transcription of a VHH domain (V-D-J-REGION) with the *IGHG2* or *IGHG3* constant region leads to the synthesis of a heavy chain without CH1, which is the characteristic feature of the hcAb repertoire (Hamers-Casterman et al., 1993) (**Figure 1**). However, it has been reported that also classical *IGHV* genes can contribute to the hcAb pool. In this case, the V domain has a classical short junction without the additional cysteine (compared to the VHH) and the conserved anchor Trp 118 is substituted mostly by an arginine codon (Deschacht et al., 2010). It was shown that also the IGHV4 subgroup contributes to produce both classical IgG and hcAb. Interestingly, a same IGHV4-IGHD-IGHJ rearrangement has been shown to be shared between a classical tetrameric IgG and a dimeric hcAb (heavy chain with no CH1, and no light chain).

From a limited number of germline VHH genes, camelids can generate a large and diversified repertoire by extensive SHM (Nguyen et al., 2000). The pattern of variability particularly within CDR loops but also in framework regions is larger in VHH than in VH cDNAs. As for classical VH, crystallographic studies of the VHH-antigen complexes demonstrated that amino acids located in the CDR1 and CDR2 loops and, particularly, the long CDR3 interact with the antigen (Desmyter et al., 1996; Decanniere et al., 1999). Contact analysis between the VHH and the ligand is provided in IMGT/3Dstructure-DB, in 38 entries of *Camelus dromedarius* and 58 for *Lama glama* (August 2019). Thus, the introductions of mutations together with a long CDR3 increase the VHH potential repertoire for antigen binding.

## DROMEDARY TRG AND TRD GENES

### The Genomic Organization of the Dromedary TRG Locus

$\gamma\delta$  T cells have unique features when compared with the more abundant  $\alpha\beta$  T cells, e.g., a preferential distribution in both epithelial and mucosal sites, and an immunoglobulin like antigen recognition mechanism in addition to the MH-restricted one. In the immune response during inflammatory processes,  $\gamma\delta$  T cells release cytokines and kill infected macrophages; they combine the characteristics of an innate-like immune response with those of an adaptive response to inflammation (Allison et al., 2001; Allison and Garboczi, 2002; Adams et al., 2005). Their percentage in peripheral blood cells, depending on age and species, differs strikingly from that of  $\alpha\beta$  T cells (Carding and Egan, 2002). Artiodactyls (sheep, cows and pigs) are referred to as “ $\gamma\delta$ -high species” since they exhibit a higher frequency and a wider physiological distribution of  $\gamma\delta$  T cells with respect to other mammalian species, including humans and mice, which are referred to as “ $\gamma\delta$ -low species” (Hein and Dudler, 1993; Ciccarese et al., 1997).

Recent studies have shown the presence of SHM in the  $\gamma$  chain of gamma/delta in shark T cells and in both chains  $\gamma$  and  $\delta$  of the dromedary camel (Chen et al., 2009; Antonacci et al., 2011; Chen et al., 2012; Vaccarelli et al., 2012). In each work, it was shown that SHM followed the characteristics of the mutational profiles detected in the B cells that undergo affinity maturation.  $\gamma\delta$  T cells, unlike  $\alpha\beta$  T cells, interact with nonclassical major histocompatibility complex (MHC) and have a small number of genes, so they have a limited diversity (Allison et al., 2001; Adams et al., 2005). The SHM seems to be used as a mechanism for further diversification of the  $\gamma\delta$  receptor and for an optimal recognition of the ligand. Consequently, this would allow the evolutionary changes in the loci, which, in turn, allow the receptor to evolve more rapidly in mutant environments (Adams et al., 2005; Kazen and Adams, 2011).

In “ $\gamma\delta$ -high” species, the TRG and TRD expressed repertoire is mainly affected by a large number of genes distributed in reiterated duplications of functional TRG cassettes (Vaccarelli et al., 2005; Conrad et al., 2007; Vaccarelli et al., 2008) and by a marked expansion and preferential usage of the TRDV1 multigene subgroup (Ishiguro et al., 1993; Yang et al., 1995; Hein and Dudler, 1997; Massari et al., 2000; Antonacci et al., 2005). Usually, in mammals, less than a few exceptions such as in human (Lefranc and Rabbitts, 1989) and in dolphin (Linguini et al., 2016), TRG loci are quite complicated, containing numerous V, J, and C genes, sometimes located in different chromosomal bands (Massari et al., 1998; Miccoli et al., 2003), or spanning hundreds of kb (Massari et al., 2009).

In *Camelus dromedarius*, the TRG locus spans approximately 45 kb and it maps in a homology region established between bovids chromosome 4, human chromosome 7, and pig chromosome 9, where orthologous TRG loci have been mapped (Vaccarelli et al., 2012).

The dromedary locus consists of two TRGV genes (TRGV1 and TRGV2), four TRGJ genes, and two TRGC genes (TRGC1 and TRGC2), all in the same transcriptional orientation, organized in two functional cassettes (5'-TRGV1-TRGJ1-TRGJ1-2-TRGC1

and TRGV2-TRGJ2-1-TRGJ2-2-TRGC2 -3'). Considering the exon organization of the ovine and human C regions, we inferred that both the dromedary C regions keep a connecting region encoded by three different exons, as is observed in the sheep TRGC2, TRGC4, and TRGC6 genes (Miccoli et al., 2001) and in the polymorphic human TRGC2 gene (Lefranc and Lefranc, 2001b). The dromedary locus organization in two (V-J-J-C) cassettes potentially limits the combinatorial usage of its genes. However, cDNA sequencing clearly revealed that, besides the combinatorial diversity and the introduction of N region diversity typical of all known IG and TR genes (Lefranc and Lefranc, 2001a; Lefranc and Lefranc, 2001b), the SHM mechanism enhances the TRG and TRD repertoire diversity in *Camelus dromedarius* (Ciccarese et al., 2014).

### SHM in TRG and TRD V Domains and Nature of AA Changes

Among mammals, SHM occurs primarily in germinal center B cells, it introduces point mutations into the variable domains of IG, and it is the driving force for antibody affinity maturation (Li et al., 2004). During SHM, the dgyw/wrch motif (where d = a or g or t, y = c or t, w = t or a, r = a or g, and h = c or t or a) has been found to be the principal hotspot for activation cytidine deaminase (AID) inducing g:u lesions in rearranged IG genes (Rogozin and Diaz, 2004; Liu and Schatz, 2009). These changes are dominated by point mutations and biased toward transitions (G:A and C:T). Moreover, the principal site for a/t mutations has been identified in the dinucleotide target wa/tw. This secondary mutator has allowed to define the roles of the error-prone polymerases in mismatch repair (Pavlov et al., 2002; Zhao et al., 2013).

The features of mutations were evaluated comparing the genomic sequence of the single TRDV4 gene with the TRDV4 cDNA clones derived from spleen and blood, excluding CDR3 and including the TRDJ4 genomic sequence. Four related sets of TRDV4 clones with common CDR3 sequences and unrelated sequences derived from independent rearrangements in blood and spleen were deduced (Antonacci et al., 2011). The analysis of mutations identified a clonal genealogy among related sequences, showing that tandem mutations resulted from sequential point mutations and, given the TRDV4 gene uniqueness, the nucleotide substitutions are not the result of gene conversion events (Antonacci et al., 2011). AID-dependent deamination of cytidine to uracil produces mutations at c/g nucleotides activates the repair proteins MSH2-MSH6 to bind u:g mismatches and recruits the lowfidelity DNA polymerase  $\eta$  (Pol  $\eta$ ) (Wilson et al., 2005; Schanz et al., 2009; Zhao et al., 2013). The analysis of substitutions in the TRDV1 and TRDV4 mutated sequences show a bias for transition changes in blood and spleen. Therefore, the nature of V domain resulting from dromedary TRG and TRD rearranged genes is the result of a combined action of AID, uracil-DNA glycosylase (UNG), and mismatch (MSH) repair pathways (**Supplementary Figure 1**) (Vaccarelli et al. 2012; Ciccarese et al., 2014).

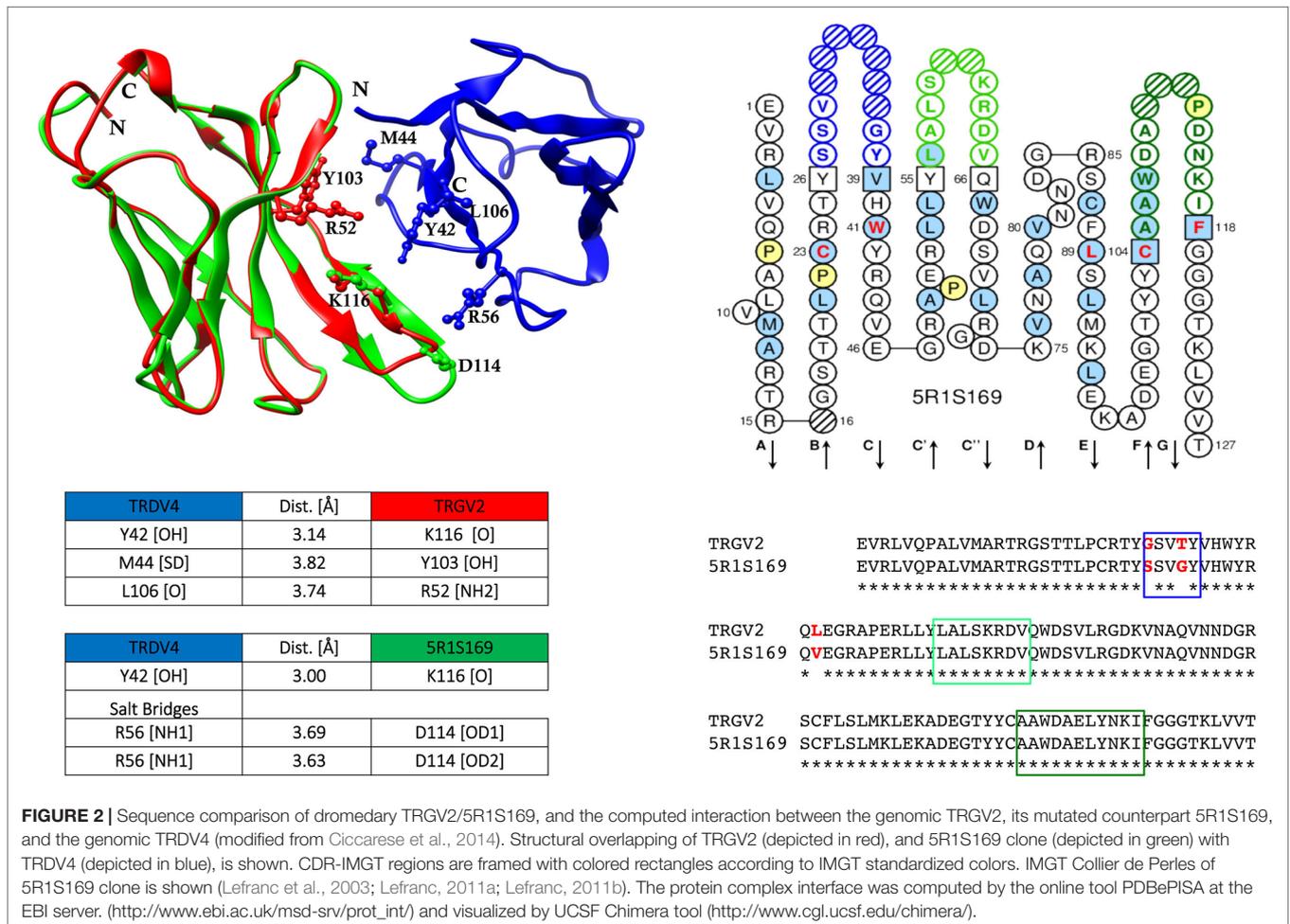
In mutated TRGV1 and TRGV2 clones, replacement mutations occur preferentially in bases inside the [dgyw] and [wrch] AID motifs, whereas neutral mutations are favoured

outside the motifs. Chi-square analysis used to compare observed and expected numbers of replacement mutations between FR and CDR regions highlighted no significant difference in TRGV1 and in TRDV4 clones. On the contrary, in TRGV2 and TRDV1 clones, the difference in the R/S ratio between CDR and FRs was significant. This feature is consistent with a selection pressure acting on dromedary  $\gamma\delta$  T cells and with the affinity maturation during clonal expansion.

When shared mutations were used as position set-points to construct putative lineages from TRG and TRD cDNA clone sets, each of them sharing an identical CDR3, the inferred lineages harbouring multiple radiations highlighted that the substitutions that arose first (progenitor mutations) were replacements changes. These progenitor mutations occurred starting from the CDR3/FR4 region and proceeding toward the leader region along the variable domain with additional mutations in FR2 and FR3. Moreover, it was found that progenitor mutations are selected and transmitted during clonal expansion and they are nonconservative of the amino acid physicochemical properties, i.e., change nonhydrophobic amino acid residues to hydrophobic ones. A computed model was constructed with the TRGV2 translated sequences of the corresponding mutated cDNAs; these sequences were visualized

in their two-dimensional structure with the IMGT tool Collier de Perles (Ehrenmann et al., 2011; Lefranc, 2011a; Lefranc, 2011b). The comparative modelling procedure was applied using the counterpart  $\gamma\delta$  human T cell receptor subunits (Ehrenmann et al., 2010; Xu et al, 2011) and a notable difference between the genomic paired TRGV2/TRDV4 and the mutated cDNA TRGV2 paired to genomic TRDV4 was observed (Figure 2). Only for the last paired V domains, the occurrence of putative hydrogen bonds and salt bridges confirmed that the changes alter the conformation of the variable domains of the  $\gamma\delta$  receptor with consequent effects on its stability.

If the replacements of hydrophilic amino acid residues with hydrophobic ones are maintained and positively selected during the proliferation of T cells, it follows that they stabilize the structure of the receptor whether they fall into the CDR or into the FR. Therefore, conclusions are consistent both with the acquisition of new antigenic specificities and repertoire diversification and simultaneously with selection for changes in paratopes in the manner similar to that of immunoglobulin gene during the B cells affinity maturation to a given antigen. The same conclusions were recently reached by Ott et al., in a paper where the authors propose that the SHM in TRA chain contributes to selection of  $\alpha\beta$  T cells in nurse “couch potato” shark



**FIGURE 2 |** Sequence comparison of dromedary TRGV2/5R1S169, and the computed interaction between the genomic TRGV2, its mutated counterpart 5R1S169, and the genomic TRDV4 (modified from Ciccarese et al., 2014). Structural overlapping of TRGV2 (depicted in red), and 5R1S169 clone (depicted in green) with TRDV4 (depicted in blue), is shown. CDR-IMGT regions are framed with colored rectangles according to IMGT standardized colors. IMGT Collier de Perles of 5R1S169 clone is shown (Lefranc et al., 2003; Lefranc, 2011a; Lefranc, 2011b). The protein complex interface was computed by the online tool PDBePISA at the EBI server. ([http://www.ebi.ac.uk/msd-srv/prot\\_int/](http://www.ebi.ac.uk/msd-srv/prot_int/)) and visualized by UCSF Chimera tool (<http://www.cgl.ucsf.edu/chimera/>).

thymus (Ott et al., 2018). The absence of SHM for the TRB chain reported recently in expression assays of spleen in dromedary (Antonacci et al., 2017a) could constitute the watershed between the cartilaginous fish that are the most divergent jawed vertebrate group relative to mammals, and mammals in the scenario of the thymic selection of  $\alpha\beta$  and  $\gamma\delta$  T cells.

In the dromedary TRG and TRD loci, evolution allowed the SHM to increase the receptor repertoire of cell-mediated immunity. Previously, we have proposed that requirements related to immunoprotective functions, including the first defensive barrier in the epithelia of the digestive tract, are likely to have induced in TRG and TRD loci of ruminants a sort of genome functional fluidity resulting in duplications of TRG gene cassettes and in a marked expansion of the TRDV1 multigene subgroup (Vaccarelli et al., 2012).

In this review, we point out that, in dromedary, TRG and TRD evolution was favoured by mutation in the productively rearranged TRG V-J and TRD V-D-J genes, so that a large and diversified TRG and TRD repertoire could be generated even in absence of functional reiterated gene duplications. Because SHM has not been shown to occur in any mammalian organism, we can hypothesize that Camelidae by themselves might occupy a peculiar immunological niche, proposing the camel lineage as a fascinating model in the evolution of immune systems.

## THE ORGANIZATION AND EVOLUTION OF CAMELUS TRB LOCUS IS SHARED IN TYLOPODA, SUINA, AND RUMINANTIA

The organization of the TRB locus has been extensively investigated in different mammalian species and it consists of a general structure with a group of TRBV genes located at the 5' end of the locus followed by in tandem TRB D-J-C clusters. A TRBV gene, with an inverted transcriptional orientation, lies at the 3' end of the region. A common aspect to most species, such as humans, rabbits, and dogs (Mineccia et al., 2012; Antonacci et al., 2014; Lefranc et al., 2015), is the presence of two TRB D-J-C clusters, each composed of one TRBD, several TRBJ, and one TRBC genes. Instead, three TRB D-J-C clusters composed the TRB locus in artiodactyl species, i.e., sheep, cattle, and pig (Antonacci et al., 2008; Connelley et al., 2009; Eguchi-Ogawa et al., 2009; Massari et al., 2018). The additional TRB D-J-C cluster 3 is located between the conserved TRB D-J-C cluster 1 and 2 (Figure 3). The sequence analysis revealed that the new TRB D-J cluster is correlated to the last one, whereas the TRBC3 gene is more similar to the TRBC2 gene in the first part and to the TRBC1 gene sequence in the last part.

The structure of the TRB locus has been recently investigated in *Camelus dromedarius* (Antonacci et al., 2017a; Antonacci et al., 2017b) and in its wild and domestic Bactrian camel congeners, *Camelus ferus* and *Camelus bactrianus* (Antonacci et al., 2019). The analysis showed that the camel TRB organization is similar to that of the other artiodactyl species, with the presence of three TRB D-J-C clusters (Figure 3). This outcome suggests that the TRB genomic organization with three TRB D-J-C clusters was established, prior the Tylopoda/Ruminantia/Suina divergence,

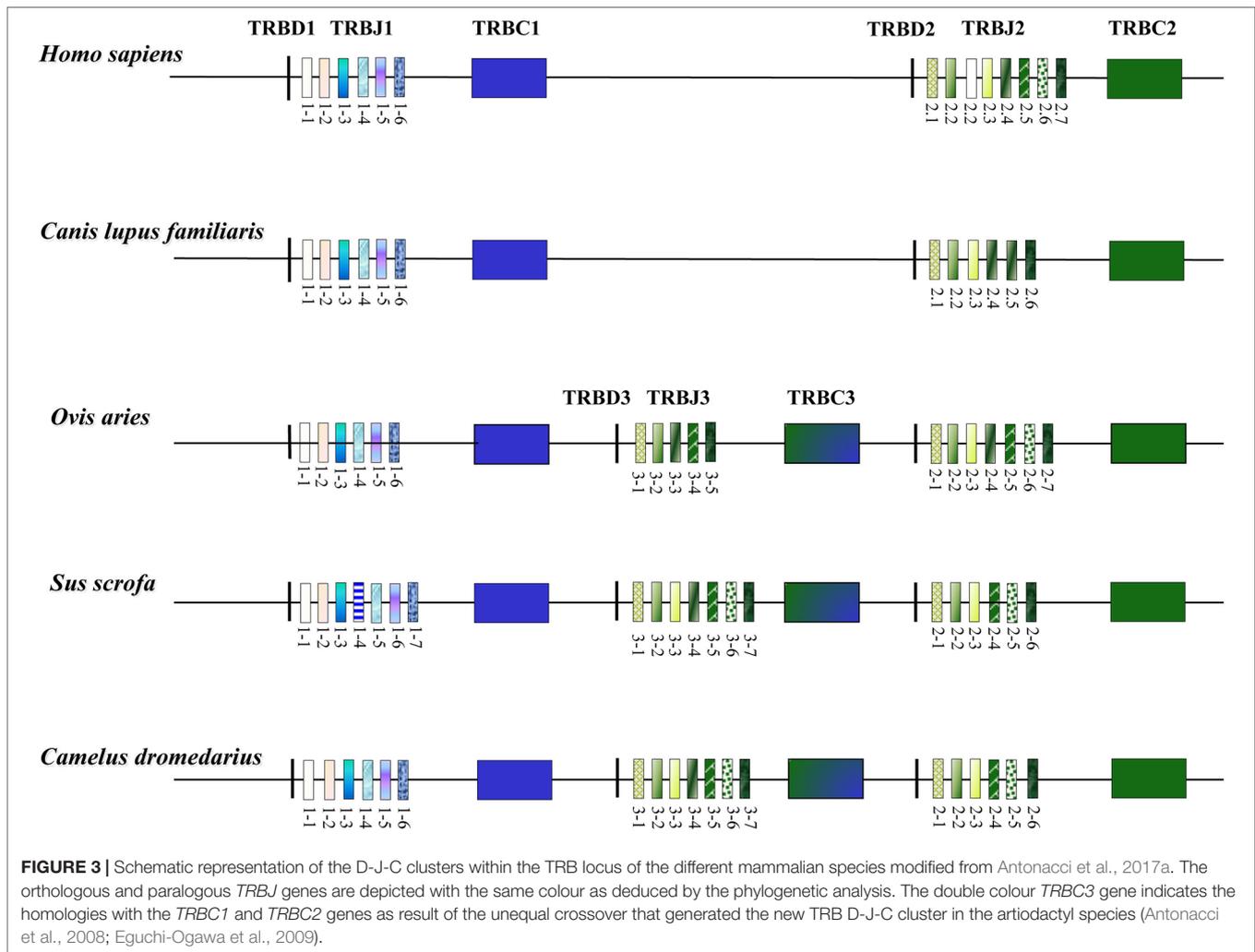
through a duplication event due to an unequal crossing-over between the ancestral TRBC1 and TRBC2 genes. Following duplication, subsequent species-specific diversifications were made that led to the current genomic organization of the 3' end of the TRB locus in the different artiodactyl species.

As in all mammalian species, MOXD2 and EPHB6 genes border the camel locus at the 5' and 3' end, respectively, whereas TRY genes are interspersed among TRBV genes and arranged in two distinct genomic positions (Antonacci et al., 2017a; Antonacci et al., 2017b; Antonacci et al., 2019).

An expression assay conducted on dromedary spleen T cells (Antonacci et al., 2017a) has demonstrated that all the three TRBD-J-C clusters are used to generate a functional TR $\beta$  chain increasing the combinational and junctional diversity of the CDR3 domain. Moreover, the analysis of the cDNA collection shows a preferential usage of the TRBD1 gene followed by the TRBD3 and TRBD2. This may result in a greater efficiency of the PD $\beta$ 1 promoter with respect to the PD $\beta$ 3 and PD $\beta$ 2, whereas the activity of two similar PD $\beta$ 3 and PD $\beta$ 2 could be correlated with their position from 5' to 3' within the locus. Furthermore, a prominent utilization of the TRBJ3 gene set with respect to the TRBJ2 and TRBJ1 clusters was also observed, probably depending on the number of genes that lie in the genomic region. Probably, multiple 12 bp spacer-recombination signal sequence (12-RS) located in a restricted region may increase the local concentration of the RAG protein that mediates the recombinant process (Di Tommaso et al., 2010). Beside the number of TRBD and TRBJ genes, other mechanisms seem to increase the dromedary TRB chain functional repertoire, including the incorporation of two TRBD genes in the rearrangement process, the intercluster recombination, and the trans-rearrangement (Antonacci et al., 2017a).

While the structure of the TRB D-J-C clusters is similar to the other artiodactyl species, the 5' end of the camel TRB locus appears to be different with a contraction of the total number of the TRBV genes link to a reduction of duplicated events within the TRBV cluster (Antonacci et al., 2017a; Antonacci et al., 2017b; Antonacci et al., 2019). 30 genes in *Camelus ferus* and 33 in *Camelus dromedarius* as in *Camelus bactrianus*, in all cases assigned to 26 different subgroups, are a low number when compared to the 134 TRBV genes for bovine, 67 for human and 74 for rabbit, but only slightly lower than that of pig (38 TRBV genes) and dog (37 TRBV genes) (Connelley et al., 2009; Mineccia et al., 2012; Antonacci et al., 2014; Lefranc et al., 2015; Massari et al., 2018).

The phylogenetic analysis of the TRBV genes (Figure 4) shows that each of *Camelus ferus*, *Camelus bactrianus*, and *Camelus dromedarius* subgroups come together and form a monophyletic group with a corresponding TRBV gene in human and, if present, in dog, sheep, and pig. This is consistent with the occurrence of distinct subgroups prior to the divergence of the different mammalian species. Three human TRBV subgroups (TRBV4, TRBV17, and TRBV18) are lacking in all three camel species, indicating that these subgroups have been lost in these species (i.e., TRBV4) or alternatively they might have originated after the separation of Camelidae (i.e., TRBV18) or Artiodactyla (i.e., TRBV17) from the other mammalian species.



## THE MHC IN CAMELS

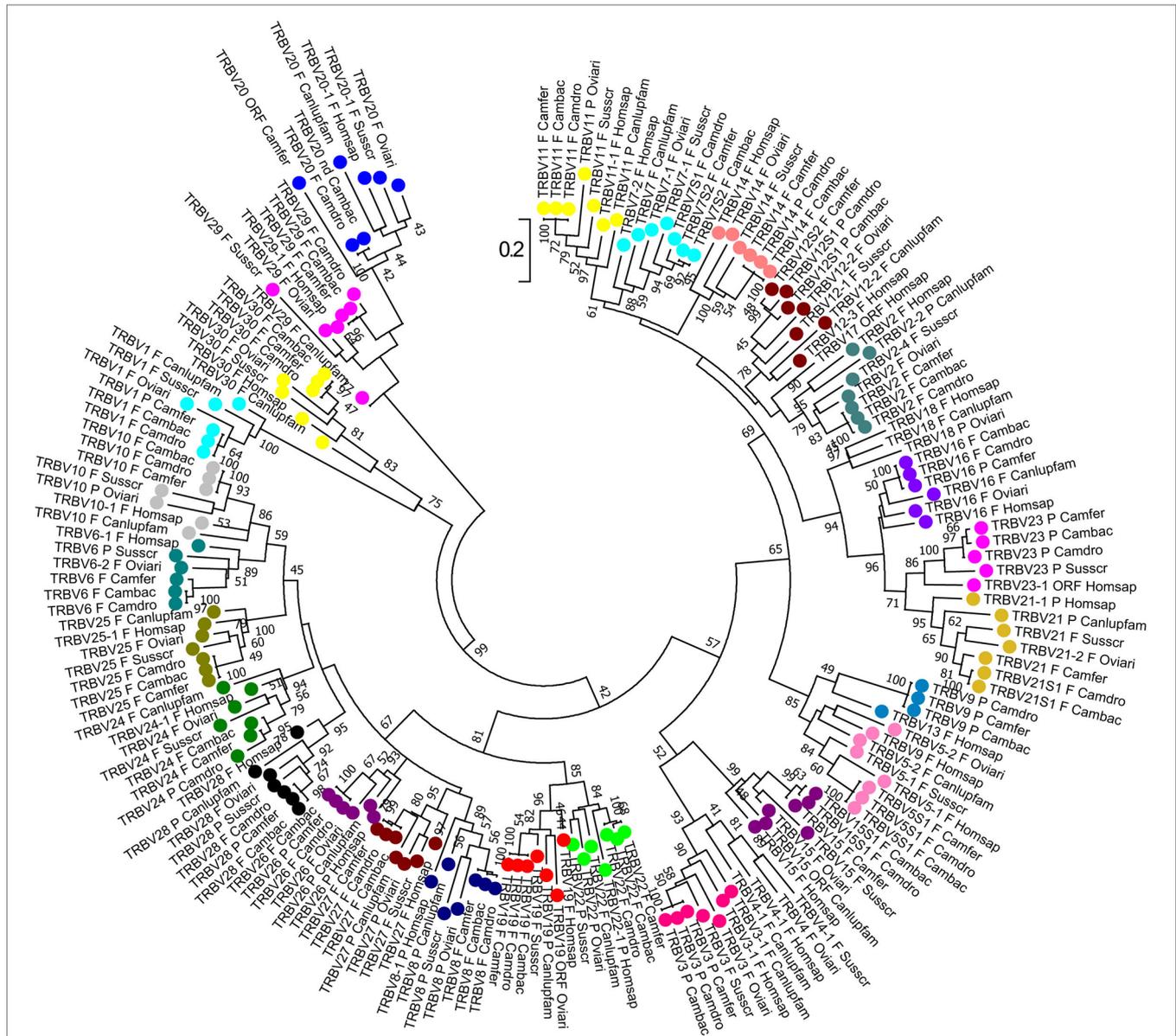
### The Overall Organization and Genetic Diversity

The MHC has evolved as part of adaptive immunity in vertebrates. In mammals, it spans approximately 4 Mbp and harbours genes, encoding hundreds of proteins with different immune as well as nonimmune functions. The mammalian MHC locus is a complex genomic region that evolved from an ancestral MHC locus, encoding primarily antigen presenting molecules, which are expressed on the surface of the cell and are involved in the immune system's defence to recognize foreign ("nonself") substances.

Two classes of antigen-presenting molecules and their genes can be distinguished. MH class I molecules present antigenic peptides originating from self as well as nonself (e.g., virus-encoded) intracellular proteins. MH class II proteins present peptides derived from extracellular proteins (e.g., bacterial products), which were internalized by specialized cells of the immune system. While molecules encoded by MH class I and II genes are mainly responsible for antigen presentation to T

lymphocytes (adaptive immunity), the MH class III region includes multiple genes involved among others in the innate immune system, such as tumour necrosis factor alpha (*TNFA*) and members of the complement cascade, which help to eliminate invading pathogens (Trowsdale and Knight, 2013).

In Old and New World camelids, the MHC is positioned on the long arm of chromosome 20 (Avila et al., 2014; Plasil et al., 2016). In Old World camels, available genome sequence data (Wu et al., 2014; Fitak et al., 2016a; Fitak et al., 2016b) were combined with initial next generation sequencing to characterize the overall MHC organization (Figure 5). The most common general structure of the mammalian MHC region, represented by the order of genes MH class II – MH class III – MH class I, was also confirmed in camels (Plasil et al., 2016; Plasil et al., 2019). Based on the nucleotide sequences retrieved, the MHC of camels seems to be, at least in some of its subregions, more similar to the human and pig MHC rather than to cattle, a phylogenetically closely related species. Besides differences in the location of the *TAP1* gene, phylogenetic analyses of individual MHC genes showed that certain MHC subregions of camels are also closer to pigs than to cattle. Detailed phylogenetic analyses can be found

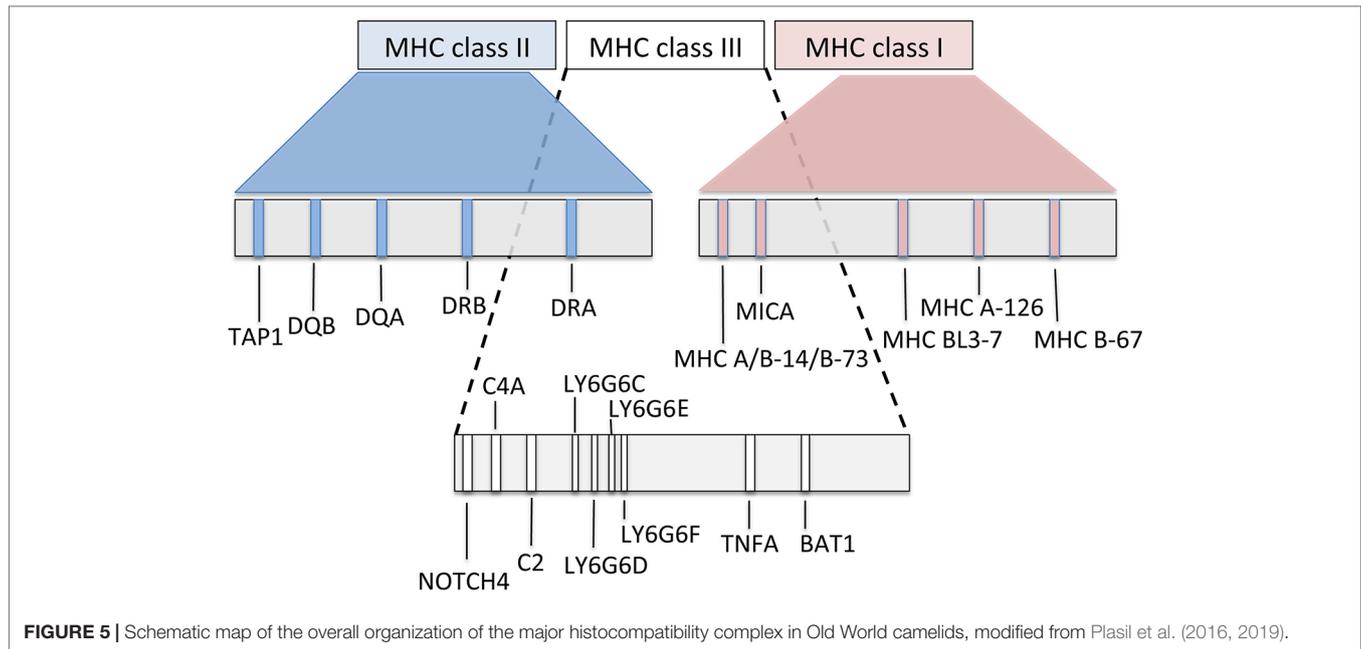


**FIGURE 4 |** Phylogenetic tree of the dromedary, wild, and domestic Bactrian camel, human, dog, sheep, and pig *TRBV* gene sequences. The sequences of the human, dog, sheep, and pig *TRBV* genes were retrieved from the IMGT database (<http://www.imgt.org/>), whereas the camel *TRBV* gene sequences derived from Antonacci et al., 2019. The nucleotide sequences of the V-REGION were combined by adopting two selection criteria: (1) only potential functional and in-frame pseudogenes (excepted for human *TRBV1*) were included; and (2) only one gene for each of the subgroups was selected for each species. The evolutionary analysis was conducted in MEGA7 (Kumar et al., 2016). The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 181 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. The different colours highlight the distribution of the phylogenetic groups corresponding to the 26 camel *TRBV* subgroups. The gene functionality according to IMGT rules (F, functional; ORF, open reading frame; P, pseudogene) is indicated; nd indicates that the nucleotide sequence of the gene is incomplete and its functionality cannot be defined. The IMGT 6-letter for species (Homsap, Susscr, Oviari, Camdro, Camfer, and Cambac) and 9-letter for subspecies (Canlupfam) standardized abbreviation for taxon is used.

in Plasil et al., 2016 for MH class II and in Plasil et al., 2019 for MH class I.

MH class I and II molecules are heterodimers consisting of two polypeptidic chains ( $\alpha$  and  $\beta$ ). While both  $\alpha$  and  $\beta$  polypeptides of the MH class II heterodimers are encoded within the MHC, the

MH class I  $\beta$ -2-microglobulin is encoded by a gene located on a different chromosome. In camels, it is on chromosome 6. MH class I loci can be distinguished into “classical” and “nonclassical” MH class I genes, the former coding for antigen-presenting molecules, thus being highly polymorphic, whereas the latter



less polymorphic genes code for a group of structurally related proteins with a variety of immune functions. Some of them may even be considered as part of the innate immune system (Allen and Hogan, 2013). MH class II genes are represented by groups of often duplicated loci (e.g., DR, DB, DM, DO, DP, and DQ). Within such a group, *A* genes (e.g., *DRA*) encode the alpha chain and *B* genes (e.g., *DRB*) the beta chain of the class II heterodimer. Most of functionally relevant polymorphisms are concentrated in the antigen binding sites (ABS) of the molecule, encoded by MH class I exons 2 and 3 and by exon 2 of class II genes.

## Overview of the MHC Genes Polymorphism

The most characteristic feature of the MHC genes is their high polymorphism, i.e., high numbers (up to hundreds) of allelic variants, especially for genes encoding antigen-presenting molecules. In fact, the MHC is one of the most polymorphic regions in the genome (Janeway et al., 2001). This is important for the immune system to respond fast to rapidly evolving pathogens, a mechanism also described as an evolutionary “arms race” (Sironi et al., 2015).

The genetic diversity of the MHC of camels was first studied in exon 2 sequences of selected class II genes. A surprisingly low level of polymorphism of the *DRA*, *DRB*, and *DQB* genes was observed in all three Old World camel species (Plasil et al., 2016). For the *DRA* locus, *DRA* exon 2 spanning 246 bp contains one synonymous and one nonsynonymous single nucleotide polymorphism (SNP) combined in three different alleles shared between dromedaries, domestic and wild Bactrian camels. Successful amplification of this gene in ancient (13<sup>th</sup> - 16<sup>th</sup> century, common era) dromedary specimens resulted in three additional substitutions when compared to the reference sequence (Plasil et al., 2016). The *DRB* exon 2 (270 bp) contains five polymorphisms

shared between the three Old World camel species. In *DQA* exon 2, 11 SNPs, 4 of them synonymous, were identified in *Camelus bactrianus*, of which nine were shared with *Camelus ferus*. In total, three haplotypes were detected, one of them common to all three species and another one shared only between domestic and wild Bactrian camels. The remaining allele was found only in domestic Bactrian camels. The *DQB* exon 2 locus was the most polymorphic with 21 polymorphic sites identified across the Old World camels (Plasil et al., 2016). The *DQB* exon 2 harbours a 12-bp long potentially functional insertion not observed in other mammalian species. Since a complete *DQB* exon 2 sequence was not retrieved, the overall extent of polymorphism in this locus remains undefined. However, data available so far suggest that only limited numbers of haplotypes may exist, similarly to the *DQA* locus containing comparable numbers of SNPs. The *DQB* locus is therefore still under investigation.

Similar observations of low diversity were made for MH class I and related loci. In the classical locus *B-67*, only one synonymous polymorphism was found in the entire exon 2 - 3 region. This SNP is shared between dromedaries and Bactrian camels. The *BL3-7* gene is a locus of unclear status, highly similar to the annotated sequence of the *BL3-6* in alpacas (Avila et al., 2014). Interestingly, it is also closely related to the locus *SLA-11* in pig, one of MHC loci with unknown function and unusual structure. In the Old World camels, this gene contains four SNPs (Plasil et al., 2019). The MH class I related locus *MR1* is an antigen-presenting molecule contributing to the regulation of the microbiome in the intestinal tract (McWilliam et al., 2016). Over the total 22 kbp long *MR1* sequence (located on chromosome 21), 170 polymorphic sites were identified, 5 of them located in the coding sequence and partially shared between dromedaries and Bactrian camels (Plasil et al., 2019). The MH class I related locus *MICA* functions as a stress signalling molecule recognized by the NKG2-D type II receptor on natural killer (NK) cells,  $\alpha\beta$

T-cells, and  $\gamma\delta$  T-cells (Shafi et al., 2011; Xiao et al., 2015). A total of 40 SNPs were observed in this sequence, of which eight were found in the coding region (Plasil et al., 2019). It is rather unusual that these MH class I related loci are more polymorphic than a classical MH class I locus, *B-67*. However, knowledge on MH class I classical genes is rather limited and a better-supported conclusion on their diversity can be done only after an extensive analysis of the entire subregion. The first step toward such an analysis and toward an analysis of the MH class III subregion will be made by annotating further genes in a new genome assembly of the dromedary.

## CONCLUSION

The Camelidae species occupy an important immunological niche within the humoral as well as cell mediated immune response. In addition to the conventional IG, the serum contains a significant amount of IgG composed solely of paired H chains, which are largely diversified by extensive SHM, resulting in novel paratopes different from those of conventional IgG. The antigen binding fragment of these unique hcAbs comprises only one single domain. When produced by microbial expression system, these recombinant miniature antigen binding fragments possess beneficial biophysical properties useful as research tools and for *in vivo* pharmacological applications as candidate drugs for the treatment of human diseases (Muyldermans, 2013; Rissiek et al., 2014; Steeland et al., 2016).

Moreover, an SHM mechanism in productively rearranged *TRD* and *TRG* genes never identified in mammalian species so far, increases the repertoire diversity of the dromedary  $\gamma\delta$  T cells that recognize the antigen in a manner antibody like. In this contest, the structural changes within the  $\gamma\delta$  heterodimer, which is stabilized by mutations both in FR and in CDR in genealogical related clones, could enable the acquisition of new antigenic specificity and, at the same time, could influence the affinity maturation to a given antigen in a manner similar to that of *IG* genes (Ciccarese et al., 2014).

As single-chain antibodies and SHM in *TR* genes were described also in cartilaginous fish (Chen et al., 2009; Flajnik et al., 2011; Chen et al., 2012; Ott et al., 2018). It can be argued that a molecular convergence of the adaptive immune response between these species does exist.

Conversely, in  $\alpha\beta$  T cells, the limited germline TRBV repertoire in *Camelus dromedarius* as in *Camelus ferus* and *Camelus bactrianus* with a great sequence identity between orthologous genes (Antonacci et al., 2017a; Antonacci et al., 2017b; Antonacci et al., 2019) is not shaped by SHM and it might be related to the constraint imposed on  $\alpha\beta$  CDR1 and CDR2 domains by the requirements for binding to MH molecules, which, in turn, show a low level of genetic diversity in all three camel species (Plasil et al., 2016).

In Old World camels, the MHC region is structured similarly to a number of other mammalian species. Phylogenetic relationships of camel *MHC* genes do not always follow relationships based on other (neutral) nuclear genes.

The diversity of the MH class I, II, and class I – related genes is generally lower than expected. This observation is consistent with a low genome-wide nuclear diversity in dromedaries, wild and domestic Bactrian camels (Fitak et al., 2016a; Fitak et al., 2016b). Several bottlenecks, in the evolutionary history of the three species, but also in recent times due to domestication of dromedaries and Bactrian camels or hunting and habitat decline of the wild camels, respectively, might be responsible for the reduced immunogenetic and genome-wide variability in Old World camels. However, experimental evidence for answering the question whether the low MHC diversity is really due to low diversity of the camel genomes is still lacking.

As the camel is a useful and promising model for therapeutic applications and for phylogenetic and evolution studies about the humoral and cell mediated immunity in jawed vertebrates, implementation of the camelid genomic sequences of IG, TR, and MHC loci is necessary to encourage progress for improvement of the global knowledge of the adaptive immune responses of these animal models.

## AUTHOR CONTRIBUTIONS

SC, PB, SM, and RA designed and wrote the review; MP and PH contributed to manuscript writing; EC, VC, and GL contributed to searching of genomic literature. All authors have read and approved the final manuscript.

## FUNDING

Austrian Science Fund (FWF): P29623-B25.

## ACKNOWLEDGMENTS

We deeply acknowledge the overall support of camel breeders and the wild camel protection foundation for providing samples and in-depth knowledge about camels. PB acknowledges funding from the Austrian Science Fund (FWF): P29623-B25. The financial support of the University of Bari and of the University of Salento is gratefully acknowledged.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00997/full#supplementary-material>

**SUPPLEMENTARY FIGURE 1|** Percentages of replacement (R) and silent (S) mutations inside and outside the activation-induced cytidine deaminase (AID) target motifs in TRGV1 and TRGV2 cDNA. Replacement mutations [(R); blue areas of bars] occur preferentially in bases inside the (a/g/t)g(c/t)(a/t) [or dgyw] and (a/t)(a/g)c(c/t/a) [or wrch] AID motifs, whereas neutral mutations [silent (S); red areas of bars] are favoured outside the motifs. Strikingly, in the coding strand of the wrch hot spot, most of the cytidine mutations are replacements (Ciccarese et al., 2014). This figure is reproduced with permission from Elsevier (License number 4515441285009).

## REFERENCES

- Achour, I., Cavelier, P., Tichit, M., Bouchier, C., Lafaye, P., and Rougeon, F. (2008). Tetrameric and homodimeric camelid IgGs originate from the same IgH locus. *J. Immunol.* 181, 2001–2009. doi: 10.4049/jimmunol.181.3.2001
- Adams, E. J., Chien, Y. H., and Garcia, K. C. (2005). Structure of a gammadelta T cell receptor in complex with the nonclassical MHC T22. *Science* 308, 227–231. doi: 10.1126/science.1106885
- Ali, A., Baby, B., and Vijayan, R. (2019). From desert to medicine: a review of camel genomics and therapeutic products. *Front. Genet.* 10, 17. doi: 10.3389/fgene.2019.00017
- Allen, R. L., and Hogan, L. (2013). “Non-Classical MHC Class I Molecules (MHC-Ib),” in *eLS* (Chichester, UK: John Wiley & Sons Ltd). doi: 10.1002/9780470015902.a0024246
- Allison, T. J., and Garboczi, D. N. (2002). Structure of gammadelta T cell receptors and their recognition of non-peptide antigens. *Mol. Immunol.* 8, 1051–1061. doi: 10.1016/S0161-5890(02)00034-2
- Allison, T. J., Winter, C. C., Fournié, J. J., Bonneville, M., and Garboczi, D. N. (2001). Structure of a human gammadelta T-cell antigen receptor. *Nature* 411, 820–824. doi: 10.1038/35081115
- Antonacci, R., Lanave, C., Del Faro, L., Vaccarelli, G., Ciccarese, S., and Massari, S. (2005). Artiodactyl emergence is accompanied by the birth of an extensive pool of diverse germline TRDV1 genes. *Immunogenetics* 57, 254–266. doi: 10.1007/s00251-005-0773-7
- Antonacci, R., Di Tommaso, S., Lanave, C., Cribiu, E. P., Ciccarese, S., and Massari, S. (2008). Organization, structure and evolution of 41 kb of genomic DNA spanning the D-J-C region of the sheep TRB locus. *Mol. Immunol.* 45, 493–509. doi: 10.1016/j.molimm.2007.05.023
- Antonacci, R., Mineccia, M., Lefranc, M. P., Ashmaoui, H. M. E., Lanave, C., Piccinni, B., et al. (2011). Expression and genomic analyses of *Camelus dromedarius* T cell receptor delta (TRD) genes reveal a variable domain repertoire enlargement due to CDR3 diversification and somatic mutation. *Mol. Immunol.* 48, 1384–1396. doi: 10.1016/j.molimm.2011.03.011
- Antonacci, R., Giannico, F., Ciccarese, S., and Massari, S. (2014). Genomic characteristics of the T cell receptor (TRB) locus in the rabbit (*Oryctolagus cuniculus*) revealed by comparative and phylogenetic analyses. *Immunogenetics* 66, 255–266. doi: 10.1007/s00251-013-0754-1
- Antonacci, R., Bellini, M., Pala, A., Mineccia, M., Hassanane, M. S., Ciccarese, S., et al. (2017a). The occurrence of three D-J-C clusters within the dromedary TRB locus highlights a shared evolution in Tylopoda, Ruminantia and Suina. *Dev. Comp. Immunol.* 76, 105–119. doi: 10.1016/j.dci.2017.05.021
- Antonacci, R., Bellini, M., Castelli, V., Ciccarese, S., and Massari, S. (2017b). Data characterizing the genomic structure of the T cell receptor (TRB) locus in *Camelus dromedarius*. *Data Brief* 14, 507–514. doi: 10.1016/j.dib.2017.08.002
- Antonacci, R., Bellini, M., Ciccarese, S., and Massari, S. (2019). Comparative analysis of the TRB locus in the *Camelus* genus. *Front. Genet.* 10, 482. doi: 10.3389/fgene.2019.00482
- Avila, F., Baily, M. P., Perelman, P., Das, P. J., Pontius, J., Chowdhary, R., et al. (2014). A comprehensive whole-genome integrated cytogenetic map for the alpaca (*Lama pacos*). *Cytogenet. Genome Res.* 144, 196–207. doi: 10.1159/000370329
- Baudisch, B., Pfort, I., Sorge, E., and Conrad, U. (2018). Nanobody-Directed Specific Degradation of Proteins by the 26S-Proteasome in Plants. *Front. Plant Sci.* 9, 130. doi: 10.3389/fpls.2018.00130
- Beghein, E., and Gettemans, J. (2017). Nanobody Technology: A versatile toolkit for microscopic imaging, protein-protein interaction analysis, and protein function exploration. *Front. Immunol.* 8, 771. doi: 10.3389/fimmu.2017.00771
- Brooks, C. L., Rossotti, M. A., and Henry, K. A. (2018). Immunological functions and evolutionary emergence of heavy-chain antibodies. *Trends Immunol.* 39, 956–960. doi: 10.1016/j.it.2018.09.008
- Carding, S. R., and Egan, P. J. (2002). Gammadelta T cells: functional plasticity and heterogeneity. *Nat. Rev. Immunol.* 2, 336–345. doi: 10.1038/nri797
- Chen, H., Kshirsagar, S., Jensen, I., Lau, K., Covarrubias, R., Schluter, S. F., et al. (2009). Characterization of arrangement and expression of the T cell receptor gamma locus in the sandbar shark. *Proc. Natl. Acad. Sci. U. S. A.* 106, 8591–8596. doi: 10.1073/pnas.0811283106
- Chen, H., Bernstein, H., Ranganathan, P., and Schluter, S. F. (2012). Somatic hypermutation of TCR  $\gamma$  V genes in the sandbar shark. *Dev. Comp. Immunol.* 37, 176–183. doi: 10.1016/j.dci.2011.08.018
- Chothia, C., Novotný, J., Brucocoleri, R., and Karplus, M. (1985). Domain association in immunoglobulin molecules. The packing of variable domains. *J. Mol. Biol.* 186, 651–663. doi: 10.1016/0022-2836(85)90137-8
- Ciccarese, S., Lanave, C., and Saccone, C. (1997). Evolution of T-cell receptors gamma and delta constant region and other T-cell related proteins in the human-rodent-artiodactyl triplet. *Genetics* 145, 409–419.
- Ciccarese, S., Vaccarelli, G., Lefranc, M. P., Tasco, G., Consiglio, A., Casadio, R., et al. (2014). Characteristics of the somatic hypermutation in the *Camelus dromedarius* T cell receptor gamma (TRG) and delta (TRD) variable domains. *Dev. Comp. Immunol.* 46, 300–313. doi: 10.1016/j.dci.2014.05.001
- Connelley, T., Aerts, J., Law, A., and Morrison, W. I. (2009). Genomic analysis reveals extensive gene duplication within the bovine TRB locus. *BMC Genomics* 10. doi: 10.1186/1471-2164-10-192
- Conrad, M. L., Mawer, M. A., Lefranc, M. P., McKinnell, L., Whitehead, J., Davis, S. K., et al. (2007). The genomic sequence of the bovine T cell receptor gamma TRG loci and localization of the TRGC5 cassette. *Vet. Immunol. Immunopathol.* 115, 346–356. doi: 10.1016/j.vetimm.2006.10.019
- Davies, J., and Riechmann, L. (1994). Camelising’ human antibody fragments: NMR studies on VH domains. *FEBS Lett.* 339, 285–290. doi: 10.1016/0014-5793(94)80432-X
- Decanniere, K., Desmyter, A., Lauwereys, M., Ghahroudi, M. A., Muyldermans, S., and Wyns, L. (1999). A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. *Structure* 7, 361–370. doi: 10.1016/S0969-2126(99)80049-5
- De Genst, E., Saerens, D., Muyldermans, S., and Conrath, K. (2006). Antibody repertoire development in camelids. *Dev. Comp. Immunol.* 30, 187–198. doi: 10.1016/j.dci.2005.06.010
- Deschacht, N., De Groeve, K., Vincke, C., Raes, G., De Baetselier, P., and Muyldermans, S. (2010). A novel promiscuous class of camelid single-domain antibody contributes to the antigen-binding repertoire. *J. Immunol.* 184, 5696–5704. doi: 10.4049/jimmunol.0903722
- Desmyter, A., Transue, T. R., Ghahroudi, M. A., Thi, M. H., Poortmans, F., Hamers, R., et al. (1996). Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nat. Struct. Biol.* 3, 803–811. doi: 10.1038/nsb0996-803
- Di Tommaso, S., Antonacci, R., Ciccarese, S., and Massari, S. (2010). Extensive analysis of D-J-C arrangements allows the identification of different mechanisms enhancing the diversity in sheep T cell receptor beta-chain repertoire. *BMC Genomics* 11. doi: 10.1186/1471-2164-11-3
- Eguchi-Ogawa, T., Toki, D., and Uenishi, H. (2009). Genomic structure of the whole D-J-C clusters and the upstream region coding V segments of the TRB locus in pig. *Dev. Comp. Immunol.* 33, 1111–1119. doi: 10.1016/j.dci.2009.06.006
- Ehrenmann, F., Kaas, Q., and Lefranc, M. P. (2010). IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhSF. *Nucleic Acids Res.* 38, D301–D307. doi: 10.1093/nar/gkp946
- Ehrenmann, F., Giudicelli, V., Duroux, P., and Lefranc, M. P. (2011). IMGT/Collier de Perles: IMGT standardized representation of domains (IG, TR, and IgSF variable and constant domains, MH and MhSF groove domains). *Cold Spring Harb. Protoc.* 6, 726–736. doi: 10.1101/pdb.prot5635
- Fernandes, J. C. (2018). Therapeutic application of antibody fragments in autoimmune diseases: current state and prospects. *Drug Discov. Today* 23, 1996–2002. doi: 10.1016/j.drudis.2018.06.003
- Fitak, R. R., Mohandesan, E., Corander, J., and Burger, P. A. (2016a). The de novo genome assembly and annotation of a female domestic dromedary of North African origin. *Mol. Ecol. Resour.* 16, 314–324. doi: 10.1111/1755-0998.12443
- Fitak, R., Mohandesan, E., Corander, J., Yadamuren, A., Chuluunbat, B., Abdelhadi, O., et al. (2016b). Genomic Footprints of Selection Under Domestication in Old World Camelids. *Plant Animal Genomic Conf. XXIV*. San Diego.
- Flajnik, M. F., Deschacht, N., and Muyldermans, S. (2011). A case of convergence: why did a simple alternative to canonical antibodies arise in sharks and camels? *PLoS Biol.* 9, e1001120. doi: 10.1371/journal.pbio.1001120
- Govaert, J., Pellis, M., Deschacht, N., Vincke, C., Conrath, K., Muyldermans, S., et al. (2012). Dual beneficial effect of interloop disulfide bond for single domain antibody fragments. *J. Biol. Chem.* 287, 1970–1979. doi: 10.1074/jbc.M111.242818
- Griffin, L. M., Snowden, J. R., Lawson, A. D., Wernery, U., Kinne, J., and Baker, T. S. (2014). Analysis of heavy and light chain sequences of conventional camelid antibodies from *Camelus dromedarius* and *Camelus bactrianus* species. *J. Immunol. Methods* 405, 35–46. doi: 10.1016/j.jim.2014.01.003

- Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E. B., et al. (1993). Naturally occurring antibodies devoid of light chains. *Nature* 363, 446–448. doi: 10.1038/363446a0
- Harmen, M. M., Ruuls, R. C., Nijman, I. J., Niewold, T. A., Frenken, L. G., and de Geus, B. (2000). Llama heavy-chain V regions consist of at least four distinct subfamilies revealing novel sequence features. *Mol. Immunol.* 37, 579–590. doi: 10.1016/S0161-5890(00)00081-X
- Hassanzadeh-Ghassabeh, G., Devoogdt, N., De Pauw, P., Vincke, C., and Muyldermans, S. (2013). Nanobodies and their potential applications. *Nanomedicine* 8, 1013–1026. doi: 10.2217/nmm.13.86
- Hein, W. R., and Dudler, L. (1993). Divergent evolution of T cell repertoires: extensive diversity and developmentally regulated expression of the sheep gamma delta T cell receptor. *EMBO J.* 12, 715–724. doi: 10.1002/j.1460-2075.1993.tb05705.x
- Hein, W. R., and Dudler, L. (1997). TCR gamma delta cells are prominent in normal bovine skin and express a diverse repertoire of antigen receptors. *Immunology* 91, 58–64. doi: 10.1046/j.1365-2567.1997.00224.x
- Helma, J., Cardoso, M. C., Muyldermans, S., and Leonhardt, H. (2015). Nanobodies and recombinant binders in cell biology. *J. Cell Biol.* 209, 633–644. doi: 10.1083/jcb.201409074
- Henry, K. A., van Faassen, H., Marcus, D., Marciel, A., Hill, J. J., Muyldermans, S., et al. (2019). Llama peripheral B-cell populations producing conventional and heavy chain-only IgG subtypes are phenotypically indistinguishable but immunogenetically distinct. *Immunogenetics* 71, 307–320. doi: 10.1007/s00251-018-01102-9
- Ishiguro, N., Aida, Y., Shinagawa, T., and Shinagawa, M. (1993). Molecular structures of cattle T-cell receptor gamma and delta chains predominantly expressed on peripheral blood lymphocytes. *Immunogenetics* 38, 437–443. doi: 10.1007/BF00184524
- Janeway, C. A. Jr, Travers, P., Walport, M., et al. (2001). “Immunobiology: The Immune System in Health and Disease,” in *The major histocompatibility complex and its functions*, (New York: Garland Science). Available from: <https://www.ncbi.nlm.nih.gov/books/NBK27156/>.
- Jirimutu, Wang, Z., Ding, G., Chen, G., Sun, Y., Sun, Z., et al. (2012). Genome sequences of wild and domestic bactrian camels. *Nat. Commun.* 3, 1202. doi: 10.1038/ncomms2192
- Jung, D., and Alt, F. W. (2004). Unraveling V(D)J recombination; insights into gene regulation. *Cell* 116, 299–311. doi: 10.1016/S0092-8674(04)00039-X
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Kazen, A. R., and Adams, E. J. (2011). Evolution of the V, D, and J gene segments used in the primate gammadelta T-cell receptor reveals a dichotomy of conservation and diversity. *Proc. Natl. Acad. Sci. U. S. A.* 108, 332–340. doi: 10.1073/pnas.1105105108
- Lauwereys, M., Arbabi Ghahroudi, M., Desmyter, A., Kinne, J., Hölzer, W., De Genst, E., et al. (1998). Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. *EMBO J.* 17, 3512–3520. doi: 10.1093/emboj/17.13.3512
- Lefranc, M. P., and Rabbitts, T. H. (1989). The human T-cell receptor gamma (TRG) genes. *Trends Biochem. Sci.* 14, 214–218. doi: 10.1016/0968-0004(89)90029-7
- Lefranc, M. P., and Lefranc, G. (2001a). *The Immunoglobulin Facts-Book*. Academic, New York 1–457.
- Lefranc, M. P., and Lefranc, G. (2001b). *The T cell Receptor Facts-Book*. Academic, New York 1–398.
- Lefranc, M.-P., Pommie, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L., et al. (2003). IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.* 27, 55–77. doi: 10.1016/S0145-305X(02)00039-3
- Lefranc, M. P., Pommie, C., Kaas, Q., Duprat, E., Bosc, N., Guiraudou, D., et al. (2005). IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev. Comp. Immunol.* 29, 185–203. doi: 10.1016/j.dci.2004.07.003
- Lefranc, M. P. (2011a). IMGT unique numbering for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. *Cold Spring Harb. Protoc.* 6, 633–642. doi: 10.1101/pdb.ip85
- Lefranc, M. P. (2011b). IMGT Collier de Perles for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. *Cold Spring Harb. Protoc.* 6, 643–651. doi: 10.1101/pdb.ip86
- Lefranc, M. P. (2014a). Immunoglobulin (IG) and T cell receptor genes (TR): IMGT® and the birth and rise of immunoinformatics. *Front. Immunol.* 5, 22. doi: 10.3389/fimmu.2014.00022
- Lefranc, M. P. (2014b). “IMGT® immunoglobulin repertoire analysis and antibody humanization,” in *Molecular Biology of B cells*, Eds. Alt, F. W., Honjo, T., Radbruch, A., and Reth, M. (London, UK: Academic Press, Elsevier Ltd), 481–514. doi: 10.1016/B978-0-12-397933-9.00026-6
- Lefranc, M. P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., et al. (2015). IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res.* 43, D413–D422. doi: 10.1093/nar/gku1056
- Lefranc, M. P., and Lefranc, G. (2019). “IMGT® and 30 years of Immunoinformatics Insight in antibody V and C domain structure and function,” in *Antibodies*, vol. 8. Eds. Jefferis, R., Strohl, W. R., and Kato, K., 29. doi: 10.3390/antib8020029
- Li, X., Duan, X., Yang, K., Zhang, W., Zhang, C., Fu, L., et al. (2016). Comparative analysis of immune repertoires between bactrian camel’s conventional and heavy-chain antibodies. *PLoS One* 11, e0161801. doi: 10.1371/journal.pone.0161801
- Li, Z., Woo, C. J., Iglesias-Ussel, M. D., Ronai, D., and Scharff, M. D. (2004). The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev.* 18, 1–11. doi: 10.1101/gad.1161904
- Linguiti, G., Antonacci, R., Tasco, G., Grande, F., Casadio, R., Massari, S., et al. (2016). Genomic and expression analyses of Tursiops truncatus T cell receptor gamma (TRG) and alpha/delta (TRA/TRD) loci reveal a similar basic public yδ repertoire in dolphin and human. *BMC Genomics* 17, 634–651. doi: 10.1186/s12864-016-2841-9
- Liu, M., and Schatz, D. G. (2009). Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol.* 30, 173–181. doi: 10.1016/j.it.2009.01.007
- Loris, R., Marianovsky, I., Lah, J., Laeremans, T., Engelberg-Kulka, H., Glaser, G., et al. (2003). Crystal structure of the intrinsically flexible addiction antidote MazE. *J. Biol. Chem.* 278, 28252–28257. doi: 10.1074/jbc.M302336200
- Massari, S., Lipsi, M. R., Vonghia, G., Antonacci, R., and Ciccarese, S. (1998). T-cell receptor TRG1 and TRG2 clusters map separately in two different regions of sheep chromosome 4. *Chromosome Res.* 6, 419–420. doi: 10.1023/A:1009245830804
- Massari, S., Antonacci, R., Lanave, C., and Ciccarese, S. (2000). Genomic organization of sheep TRDJ segments and their expression in the delta chain repertoire in thymus. *Immunogenetics* 52, 1–8. doi: 10.1007/s002510000243
- Massari, S., Bellahcene, F., Vaccarelli, G., Carelli, G., Mineccia, M., Lefranc, M. P., et al. (2009). The deduced structure of the T cell receptor gamma locus in *Canis lupus familiaris*. *Mol. Immunol.* 46, 2728–2736. doi: 10.1016/j.molimm.2009.05.008
- Massari, S., Bellini, M., Ciccarese, S., and Antonacci, R. (2018). Overview of the germline and expressed repertoires of the TRB genes in *Sus scrofa*. *Front. Immunol.* 9, 2526. doi: 10.3389/fimmu.2018.02526
- McWilliam, H. E., Eckle, S. B., Theodossis, A., Liu, L., Chen, Z., Wubben, J. M., et al. (2016). The intracellular pathway for the presentation of vitamin B-related antigens by the antigen-presenting molecule MR1. *Nat. Immunol.* 17, 531–537. doi: 10.1038/ni.3416
- Miccoli, M. C., Lipsi, M. R., Massari, S., Lanave, C., and Ciccarese, S. (2001). Exon-intron organization of TRGC genes in sheep. *Immunogenetics* 53, 416–422. doi: 10.1007/s002510100340
- Miccoli, M. C., Antonacci, R., Vaccarelli, G., Lanave, C., Massari, S., Cribiu, E. P., et al. (2003). Evolution of TRG clusters in cattle and sheep genomes as drawn from the structural analysis of the ovine TRG2@ locus. *J. Mol. Evol.* 57, 52–62. doi: 10.1007/s00239-002-2451-9
- Mineccia, M., Massari, S., Linguiti, G., Ceci, L., Ciccarese, S., and Antonacci, R. (2012). New insight into the genomic structure of dog T cell receptor beta (TRB) locus inferred from expression analysis. *Dev. Comp. Immunol.* 37, 279–293. doi: 10.1016/j.dci.2012.03.010
- Mitchell, L. S., and Colwell, L. J. (2018a). Analysis of nanobody paratopes reveals greater diversity than classical antibodies. *Protein Eng. Des. Sel.* 31, 267–275. doi: 10.1093/protein/gzy017
- Mitchell, L. S., and Colwell, L. J. (2018b). Comparative analysis of nanobody sequence and structure data. *Proteins* 86, 697–706. doi: 10.1002/prot.25497
- Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. A., and Hamers, R. (1994). Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng.* 7, 1129–1135. doi: 10.1093/protein/7.9.1129

- Muyldermans, S., Baral, T. N., Retamozzo, V. C., De Baetselier, P., De Genst, E., Kinne, J., et al. (2009). Camelid immunoglobulins and nanobody technology. *Vet. Immunol. Immunopathol.* 128, 178–183. doi: 10.1016/j.vetimm.2008.10.299
- Muyldermans, S., and Lauwereys, M. (1999). Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies. *J. Mol. Recogn.* 12, 131–140. doi: 10.1002/(SICI)1099-1352(199903/04)12:2<131::AID-JMR454>3.0.CO;2-M
- Muyldermans, S. (2013). Nanobodies: natural single-domain antibodies. *Annu. Rev. Biochem.* 82, 775–797. doi: 10.1146/annurev-biochem-063011-092449
- Nguyen, V. K., Hamers, R., Wyns, L., and Muyldermans, S. (1999). Loss of splice consensus signal is responsible for the removal of the entire C(H)1 domain of the functional camel IGG2A heavy-chain antibodies. *Mol. Immunol.* 36, 515–524. doi: 10.1016/S0161-5890(99)00067-X
- Nguyen, V. K., Hamers, R., Wyns, L., and Muyldermans, S. (2000). Camel heavy-chain antibodies: diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire. *EMBO J.* 19, 921–930. doi: 10.1093/emboj/19.5.921
- Nguyen, V. K., Desmyter, A., and Muyldermans, S. (2001). Functional heavy-chain antibodies in Camelidae. *Adv. Immunol.* 79, 261–296. doi: 10.1016/S0065-2776(01)79006-2
- Ott, J. A., Castro, C. D., Deiss, T. C., Ohta, Y., Flajnik, M. F., and Criscitiello, M. F. (2018). Somatic hypermutation of T cell receptor  $\alpha$  chain contributes to selection in nurse shark thymus. *Elife* 17, 7. doi: 10.7554/eLife.28477
- Pavlov, Y. I., Rogozin, I. B., Galkin, A. P., Akshenova, A. Y., Hanaoka, F., Rada, C., et al. (2002). Correlation of somatic hypermutation specificity and A-T base pair substitution errors by DNA polymerase  $\delta$  during copying of a mouse immunoglobulin  $\mu$  light chain transgene. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9954–9959. doi: 10.1073/pnas.152126799
- Plasil, M., Mohandesan, E., Fitak, R. R., Musilova, P., Kubickova, S., Burger, P. A., et al. (2016). The major histocompatibility complex in Old World camelids and low polymorphism of its class II genes. *BMC Genomics* 17, 167. doi: 10.1186/s12864-016-2500-1
- Plasil, M., Wijkmark, S., Elbers, J. P., Oppelt, J., Burger, P. A., and Horin, P. (2019). The major histocompatibility complex of Old World camelids: class I and class I-related genes. *HLA* 93, 203–215. doi: 10.1111/tan.13510
- Rissiek, B., Koch-Nolte, F., and Magnus, T. (2014). Nanobodies as modulators of inflammation: potential applications for acute brain injury. *Front. Cell Neurosci.* 8, 344. doi: 10.3389/fncel.2014.00344
- Rogozin, I. B., and Diaz, M. (2004). Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J. Immunol.* 172, 3382–3384. doi: 10.4049/jimmunol.172.6.3382
- Saccodossi, N., De Simone, E. A., and Leoni, J. (2012). Structural analysis of effector functions related motifs, complement activation and hemagglutinating activities in Lama glama heavy chain antibodies. *Vet. Immunol. Immunopathol.* 145, 323–331. doi: 10.1016/j.vetimm.2011.12.001
- Schanz, S., Castor, D., Fischer, F., and Jiricny, J. (2009). Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5593–5598. doi: 10.1073/pnas.0901726106
- Schumacher, D., Helma, J., Schneider, A. F. L., Leonhardt, H., and Hackenberger, C. P. R. (2018). Nanobodies: chemical functionalization strategies and intracellular applications. *Angew. Chem. Int. Ed. Engl.* 57, 2314–2333. doi: 10.1002/anie.201708459
- Shafi, S., Vantourout, P., Wallace, G., Antoun, A., Vaughan, R., Stanford, M., et al. (2011). An NKG2D-mediated human lymphoid stress surveillance response with high interindividual variation. *Sci. Transl. Med.* 3, 113ra124. doi: 10.1126/scitranslmed.3002922
- Sironi, M., Cagliani, R., Forni, D., and Clerici, M. (2015). Evolutionary insights into host–pathogen interactions from mammalian sequence data. *Nat. Rev. Gen.* 16, 224. doi: 10.1038/nrg3905
- Steeland, S., Vandenbroucke, R. E., and Libert, C. (2016). Nanobodies as therapeutics: big opportunities for small antibodies. *Drug Discov. Today* 21, 1076–1113. doi: 10.1016/j.drudis.2016.04.003
- Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526. doi: 10.1093/oxfordjournals.molbev.a040023
- Tillib, S. V., Vyatchanin, A. S., and Muyldermans, S. (2014). Molecular analysis of heavy chain-only antibodies of *Camelus bactrianus*. *Biochemistry* 79, 1382–1390. doi: 10.1134/S000629791412013X
- Trowsdale, J., and Knight, J. C. (2013). Major histocompatibility complex genomics and human disease. *Ann. Rev. Gen. Hum. Gen.* 14, 301–323. doi: 10.1146/annurev-genom-091212-153455
- Vaccarelli, G., Miccoli, M. C., Lanave, C., Massari, S., Cribiu, E. P., and Ciccarese, S. (2005). Genomic organization of the sheep TRG1 locus and comparative analyses of Bovidae and human variable genes. *Gene* 357, 103–114. doi: 10.1016/j.gene.2005.05.033
- Vaccarelli, G., Miccoli, M. C., Antonacci, R., Pesole, G., and Ciccarese, S. (2008). Genomic organization and recombinational unit duplication-driven evolution of ovine and bovine T cell receptor gamma loci. *BMC Genomics* 9, 81. doi: 10.1186/1471-2164-9-81
- Vaccarelli, G., Antonacci, R., Tasco, G., Yang, F. T., Giordano, L., El Ashmaoui, H. M., et al. (2012). Generation of diversity by somatic mutation in the *Camelus dromedarius* T-cell receptor gamma variable domains. *Eur. J. Immunol.* 42, 3416–3428. doi: 10.1002/eji.201142176
- Vu, K. B., Ghahroudi, M. A., Wyns, L., and Muyldermans, S. (1997). Comparison of llama VH sequences from conventional and heavy chain antibodies. *Mol. Immunol.* 34, 1121–1131. doi: 10.1016/S0161-5890(97)00146-6
- Wilson, T. M., Vaisman, A., Martomo, S. A., Sullivan, P., Lan, L., Hanaoka, F., et al. (2005). MSH2-MSH6 stimulates DNA polymerase  $\epsilon$ , suggesting a role for A: T mutations in antibody genes. *J. Exp. Med.* 201, 637–645. doi: 10.1084/jem.20042066
- Woolven, B. P., Frenken, L. G., van der Logt, P., and Nicholls, P. J. (1999). The structure of the llama heavy chain constant genes reveals a mechanism for heavy-chain antibody formation. *Immunogenetics* 50, 98–101. doi: 10.1007/s002510050694
- Wu, H., Guang, X., Al-Fageeh, M. B., Cao, J., Pan, S., Zhou, H., et al. (2014). Camelid genomes reveal evolution and adaptation to desert environments. *Nat. Commun.* 5, 5188. doi: 10.1038/ncomms6188
- Xiao, G., Wang, X., Sheng, J., Lu, S., Yu, X., and Wu, J. D. (2015). Soluble NKG2D ligand promotes MDSC expansion and skews macrophage to the alternatively activated phenotype. *J. Hematol. Oncol.* 8, 13. doi: 10.1186/s13045-015-0110-z
- Xu, B., Pizarro, J. C., Holmes, M. A., McBeth, C., Groh, V., Spies, T., et al. (2011). Crystal structure of a gammadelta T-cell receptor specific for the human MHC  $\chi$ 1008 I homolog MICA. *Proc. Natl. Acad. Sci. U. S. A.* 108, 2414–2419. doi: 10.1073/pnas.1015433108
- Yang, Y. G., Ohta, S., Yamada, S., Shimizu, M., and Takagaki, Y. (1995). Diversity of T cell receptor delta-chain cDNA in the thymus of a one-month-old pig. *J. Immunol.* 155, 1981–1993.
- Zhao, Y., Gregory, M. T., Biertmpfel, C., Hua, Y.-J., Hanaoka, F., and Yangb, W. (2013). Mechanism of somatic hypermutation at the WA motif by human DNA polymerase  $\delta$ . *Proc. Natl. Acad. Sci. U. S. A.* 110, 8146–8151. doi: 10.1073/pnas.1303126110

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ciccarese, Burger, Ciani, Castelli, Linguiti, Plasil, Massari, Horin and Antonacci. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.