Developmental and Comparative Immunology 76 (2017) 105–119

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

The occurrence of three D-J-C clusters within the dromedary TRB locus highlights a shared evolution in Tylopoda, Ruminantia and Suina

Rachele Antonacci [a], Mariagrazia Bellini [a], Angela Pala [a], Micaela Mineccia [a], Mohamed S. Hassanane [b], Salvatrice Ciccarese [a], Serafina Massari [c, *]

[a] Department of Biology, University "Aldo Moro" of Bari, Bari, Italy
[b] Cell Biology Department National Research Center, Dokki, Giza, Egypt
[c] Department of Biological and Environmental Science e Technologies, University of Salento, Lecce, Italy

ARTICLE INFO

Article history:
Received 7 April 2017
Received in revised form
26 May 2017
Accepted 26 May 2017
Available online 31 May 2017

Keywords:
T cell receptor
TRB locus
Dromedary genome
Beta chain repertoire
Comparative genomics
IMGT

ABSTRACT

The αβ T cells are important components of the adaptive immune system and can recognize a vast array of peptides presented by MHC molecules. The ability of these T cells to recognize the complex depends on the diversity of the αβ TR, which is generated by a recombination of specific Variable, Diversity and Joining genes for the β chain, and Variable and Joining genes for the α chain. In this study, we analysed the genomic structure and the gene content of the TRB locus in *Camelus dromedarius*, which is a species belonging to the Tylopoda suborder. The most noteworthy result is the presence of three in tandem TRBD-J-C clusters in the dromedary TRB locus, which is similar to clusters found in sheep, cattle and pigs and suggests a common duplication event occurred prior to the Tylopoda/Ruminantia/Suina divergence. Conversely, a significant contraction of the dromedary TRBV genes, which was previously found in the TRG and TRD loci, was observed with respect to the other artiodactyl species.

1. Introduction

Camelids can be classified into two groups: Old World Camelids, which include dromedary and Bactrian camels, and the New World camelids, which are represented by four closely related Lamini species, specifically llama (*Lama glama*), alpaca (*Lama pacos*), guanaco (*Lama guanicoe*) and vicuna (*Vicugna vicugna*). In classical systematics, the Camelidae family belongs to Tylopoda, which represents the three suborders of Artiodactyla along with Ruminantia and Suina. The single humped dromedary (*Camelus dromedarius*) and the two-humped Bactrian camel (*Camelus bactrianus*) are the key domestic species in semi-arid and arid areas of Arabian, Northern African and Asian countries. In contrast, Lamini, which have no hump, live in the high altitudes of South America and do not exhibit similar adaptations to hot desert environments.

The unique tolerance for extremely hot and arid conditions makes the dromedary camel an economically and logistically strategic animal considering the observed changes in climate along with the projected increase in global warming and desertification. The dromedary is a comparatively hardy animal, and it is less susceptible to many of the diseases that affect other livestock species, such as trypanosomiasis (Njiru et al., 2004) and brucellosis (Abbas and Agab, 2002). Furthermore, reports have uncovered several camel products that can be applied to human medicine, including milk that may contain beneficial properties for the treatment of diabetes (Agrawal et al., 2011). Studies of the humoral immune response have indicated that dromedary camels, as well as other camelid species, are important models for immunological research since dromedary camels have special antibodies with two heavy chains and no light chains (Hamers-Casterman et al., 1993) in addition to conventional tetrameric immunoglobulins. These features have been found to be particularly useful for biotechnological applications (Muyldermans et al., 2009; Vincke et al., 2009; Deschacht et al., 2010). Additionally, the heavy-chain antibody repertoire is largely diversified by extensive somatic hypermutation (SHM) involving the variable domains and resulting novel

* Corresponding author. Department of Biological and Environmental Science e Technologies, University of Salento, Via Monteroni 165, Centro Ecotekne, 73100, Lecce, Italy.
E-mail addresses: rachele.antonacci@uniba.it (R. Antonacci), mg1192@hotmail.it (M. Bellini), angela.pala@uniba.it (A. Pala), mineccia14@alice.it (M. Mineccia), hmsaber@yahoo.com (M.S. Hassanane), salvatricemaria.ciccarese@uniba.it (S. Ciccarese), sara.massari@unisalento.it (S. Massari).

http://dx.doi.org/10.1016/j.dci.2017.05.021
0145-305X/

paratopes that are different from paratopes of conventional antibodies (Nguyen et al., 2001, 2000).

SHM has been shown to occur also in T cell receptor (TR) chains in salmonids (Yazawa et al., 2008) and sharks (Chen et al., 2012, 2009) but not in mammalian species, including artiodactyl species. Nevertheless, more recent investigations have shown that SHM occurs in the dromedary TR gamma (TRG) and delta (TRD) rearranged genes (Antonacci et al., 2011; Vaccarelli et al., 2012; Ciccarese et al., 2014) to generate a large and diversified repertoire of the T cell delta and gamma chains, how, as an alternative, the expansion of the TRD and TRG genes does in the other species of artiodactyls.

Generally, artiodactyls (i.e., sheep, cattle and pigs) with chickens and rabbits are mentioned to as "γδ-high species" since they exhibit a higher frequency and a wider physiological distribution of γδ T cells with respect to other mammalian species, including humans, mice and dogs, which are referred to as "γδ-low species". In sheep, cattle and pigs, the "γδ-high" condition seems to be correlated with a complex genomic organization of the TRG and TRD loci affected by a large number of functional germline genes. Particularly, within the TRD locus, the presence of a high number of germline Variable (TRDV) genes can be noted with a marked expansion and preferential usage of the TRDV1 multigene subgroup (Antonacci et al., 2005; Eguchi-Ogawa et al., 2009; Connelley et al., 2014; Piccinni et al., 2015).

Similarly, the TRG locus is constituted by reiterated duplication of functional V-J-C cassettes with many TRGV and TRGJ genes (Conrad et al., 2007; Vaccarelli et al., 2008) spanning hundreds of Kb and located in two different chromosomal bands in Bovidae species (Massari et al., 1998; Miccoli et al., 2003; Antonacci et al., 2007).

In contrast, the dromedary TRG locus maps in a unique chromosomal region and it is composed of only a few genes arranged in two juxtaposed V-J-C cassettes, which are distributed over only 45 Kb and each contain one TRGV, two TRGJ, and one TRGC gene (Vaccarelli et al., 2012). Likewise, the dromedary TRD locus seems to be composed of few distinct TRDV genomic germline sequences with only six TRDV1 subgroup genes (Antonacci et al., 2011).

To complete the analysis of the TR loci in dromedary camels, we analysed the identified and annotated (see companion data paper by Antonacci et al., 2017) locus encoding for the TR beta chain (TRB) using the first sequence of the dromedary genome provided by Wu et al. (2014), which was improved upon by Long-PCR experiments on genomic DNA. The analysis of the genomic organization of the dromedary TRB locus showed that Tylopoda, Ruminantia and Suina share a similar structure defined by a pool of TRBV genes positioned upstream of three in tandem aligned D-J-C clusters. Finally, if SHM mechanisms appear to compensate the limited repertoire of the TRGV and TRDV genes (Ciccarese et al., 2014), this does not seem to be applicable for the TRB genes. In fact, the analysis of our cDNA collection excludes the presence of SHM in the dromedary TRB rearranged genes.

2. Materials and methods

2.1. Genome analyses

To analyse the dromedary TRB locus, a sequence of 457 871 pb (gaps included) was retrieved directly from the PRJNA234474_Ca_dromedarius_V1.0 assembly of the whole genome shotgun sequence available at GenBank (ID: GCA_000767585.1). Particularly, the analysed region comprises eight unplaced genomic not continuous scaffolds: NW_011591622, NW_011593440, NW_011591151, NW_011620189, NW_011616084, NW_011607149, NW_011601111, NW_011623391.

MOXD2 (monooxygenase, dopamine-beta-hydroxylase-like 2, DBH-like2) and EPHB6 (ephrin type-B receptor 6) genes, flanking, respectively, the 5′ and 3' ends of TRB locus, were included in the analysis. In particular, the MOXD2 sequence is located within the NW_011591622, whereas the EPHB6 gene is in the NW_011593440 scaffold.

The human genomic sequence and the sheep genomic D-J-C region (Lefranc et al., 2015; Antonacci et al., 2008) were used against the *Camelus dromedarius* genome sequence to identify genes, based on homology by the BLAST program, the corresponding genomic TRBV, TRBD, TRBJ and TRBC genes. Moreover, the homology-based method was used, aligning the dromedary retrieved sequence against itself with the PipMaker program (Schwartz et al., 2000) (data not shown). The presence in the pip of superimposed lines clearly indicates the occurrence of redundant matches between dromedary genes (especially TRBV genes), due to the homology among the different groups of genes. The beginning and end of each coding exon of TRBV, TRBD and TRBJ genes were identified by the presence of splice sites or flanking recombination signal sequences (RSs).

Computational analysis of the dromedary D-J-C region was conducted using the following programs: RepeatMasker for the identification of genome-wide repeats and low complexity regions (Smit, A.F.A. Hubley, R. Green, P. RepeatMasker open-4.0. at http://www.repeatmasker.org) and Pipmaker (Schwartz et al., 2000; http://www.pipmaker.bx.psu.edu/pipmaker/) for the alignment of the dromedary sequence with the sheep counterpart.

2.2. Classification of the dromedary TRB genes

Considering the percentage of nucleotide identity of the genes with respect to human and the other mammalian species and based on the genomic position within the locus, each TRB gene was classified and the nomenclature established according to IMGT at http://www.imgt.org/IMGTScientificChart/SequenceDescription/IMGTfunctionality.html (Lefranc, 2014) (see Table 1 in the companion data paper by Antonacci et al., 2017).

TRBV genes were assigned to 26 different subgroups on the basis of the percentage of nucleotide identity by using Clustal Omega alignment tool available at EMBL-EBI website (http://www.ebi.ac.uk/), adopting the criterion that sequences with nucleotide identity of more than 75% in the V-region belong to the same subgroup. Due to the fragmented and incomplete nature of the genomic assembly, a temporary designation is used for multigene subgroups in which the Arabic number (for the subgroup) is followed by the letter S, themselves followed by the number of the gene in the subgroup.

The TRBD, the TRBJ and the TRBC genes were annotated, according to the similarity with the sheep species (Antonacci et al., 2008). Each TRBJ1, TRBJ2 and TRBJ3 gene was designed by a hyphen and a number corresponding to their position in the cluster. They were all predicted to be functional.

2.3. Phylogenetic analyses

The TRBV and TRBJ genes used for the phylogenetic analysis were retrieved from the following sequences deposited in the GEDI (for GenBank/ENA/DDBJ/IMGT/LIGM-DB) databases: NG_001333 (human TRB locus contig), NW_003726086 (dog TRB locus contig as characterized by Mineccia et al., 2012); NW_003159384 (rabbit TRB locus contig as characterized by Antonacci et al., 2014); AM420900 (sheep D-J-C region, Antonacci et al., 2008); AB476299 and AB476300 (pig TRB locus contig, Eguchi-Ogawa et al., 2009); NW_011591622, NW_011593440, NW_011591151, NW_011620189, NW_011616084, NW_011607149, NW_011601111 and LT837971

(this work) (dromedary TRB locus contig).

The phylogenetic analysis was performed to classify the dromedary TRBV genes. We combined the nucleotide sequences of the V-REGION of the dromedary TRBV genes with the corresponding gene sequences of humans, dogs and rabbits.

The phylogenetic relationship of the TRBJ genes were investigated by aligning the nucleotide sequences of all dromedary TRBJ genes (coding region plus RSS) with those of humans, dogs, sheep and pigs.

Multiple alignments of the gene sequences under analysis were carried out with the MUSCLE program (Edgar, 2004). Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011). We used the neighbour-joining (NJ) method to reconstruct the phylogenetic tree (Saitou and Nei, 1987). The evolutionary distances were computed using the p-distance method (Nei and Kumar 2000) and are in the units of the number of base differences per site.

2.4. Long PCR

Dromedary lung genomic DNA was prepared from a single animal using the EuroGold Tissue DNA Mini kit (EuroClone) according to the manufacturer's recommended protocol. DNA was quantified spectrophotometrically at 260 nm, aliquoted and stored at −20 °C until it was assayed. PCRs were performed following the manufacturer's instructions for the High Fidelity DNA polymerase (Expand Long Template PCR system, Roche). The PCR products were cloned and sequenced by a commercial service. The list of the clones with the primer pairs that were used and the PCR conditions are in the companion data paper (Antonacci et al., 2017).

2.5. 5' rapid amplification of cDNA ends (RACE) PCR

Total RNA was extracted from the spleen of two adult unrelated animals using the Trizol method according to the manufacturer's protocol (Invitrogen, Carlsbad, CA). Two RNA samples were obtained and two different experiments were set up (5R1B and 5R2B series clones). Approximately 5 μg of RNA was reverse transcribed with Superscript II (Invitrogen, Carlsbad, CA) by using a specific primer, CB2L (5′ TGGTTGCGGGGGTTGTGC 3′), designed on the conserved sequence of the first exon of the TRBC genes. After linking a poly-C tail at the 5'end of the cDNAss, the cDNAds was performed with Platinum Taq polymerase (Invitrogen) by using CBL2 as the lower primer and an anchor oligonucleotide as the upper primer (AAP) provided by the supplier (Invitrogen). The PCR conditions were as follows: 30 s at 94 °C, 45 s at 55 °C, and 1 min at 72 °C for 35 cycles. The products were then amplified in a subsequent nested PCR experiment by using B8L/AUAP primer pair. B8L (5′ CGTGGTCGGGGTAGAAGC 3′) is designed on a conserved sequence upstream of CB2L while the AUAP oligonucleotide was provided by the supplier (Invitrogen). The PCR conditions were as follows: 30 s at 94 °C, 40 s at 57 °C, 30 s at 72 °C for 30 cycles. The final cycle was extension for 30 min at 72 °C.

The RACE products were then gel-purified and cloned using the StrataClone PCR Cloning Kit (Statagene). Random selected positive clones for each cloning were sequenced by a commercial service. cDNA sequence data were processed and analysed using the Blast program (http://www.blast.ncbi.nlm.nih.gov/Blast.cgi), the Clustal Omega alignment tool (http://www.ebi.ac.uk/) and IMGT tools [IMGT/V-QUEST (Brochet et al., 2008; Giudicelli et al., 2011) with integrated IMGT/JunctionAnalysis tools (Yousfi Monod et al., 2004; Giudicelli and Lefranc, 2011)] and the IMGT unique numbering for the V domain (Lefranc et al., 2003) (http://www.imgt.org/).

All cDNA clones were registered in EMBL database with the Accession numbers from LT837972 to LT838006.

2.6. Genomic PCR

Genomic DNA (1 μg) sample derived from the 5R2B animal, was used as the template in one (V21U 5′ CTCTGCTGTCTGGCCCTTTT 3'/ V21L 5′ CTGCCTGAGTGGAGTTGATT 3' primer pair) TRBV21 PCR reaction with 30 cycles annealing at 58 °C for 30s and extending at 68 °C for 45s. The final cycle was a 15-min extension at 68°. Platinum Taq Polymerase (Invitrogen) was used in the PCR reaction. The PCR product was gel-purified and cloned using the StrataClone PCR Cloning Kit (Statagene). Random selected positive clones were sequenced by a commercial service and analysed using the Blast program (http://www.blast.ncbi.nlm.nih.gov/Blast.cgi). We recovered three different sequences registered in EMBL database with the Accession numbers from LT837968 to LT837970.

3. Results

3.1. Genomic organization of the Camelus dromedarius TRB locus

The recent submission to NCBI (BioProject PRJNA234474) of a draft genome sequence from the Arabian camel (Wu et al., 2014) allowed us to infer the genomic structure of the TRB locus for the first time in a mammalian species belonging to the *Camelus* genus. A standard BLAST search of the dromedary genomic resource was then performed by using human and sheep TRB gene sequences to assess their physical location in the dromedary genome (see Section 2.1 in the companion data paper by Antonacci et al., 2017). The general structural organization of the dromedary TRB locus follows that of the other mammalian species, with a library of TRBV genes positioned at the 5′ end of D-J-C clusters, followed by a single TRBV gene with an inverted transcriptional orientation located at the 3′ end (Fig. 1). It is noteworthy that, despite the fragmented and incomplete nature of the dromedary genomic assembly, especially in the 3′ part of the locus, the analysis of the sequence revealed the potential presence of three D-J-C clusters similar to clusters found in sheep (Antonacci et al., 2008). D-J-C cluster 1 is the only complete cluster in the current assembly, and it consists of one TRBD, six TRBJ and one TRBC functional genes. The second cluster contains at least four TRBJ3 genes and a putative incomplete TRBC3 gene, and the third D-J-C cluster is composed of at least three TRBJ2 genes and the TRBC2 gene positioned at the 5' of the TRBV30 gene (Fig. 1; see also Table 1 in Antonacci et al., 2017).

The comparison of the entire dromedary sequence to the human sequence allowed us to identify and annotate the unrelated TRB genes consisting of a group of trypsin-like serine protease (TRY) genes that are typically interspersed among mammalian TRB genes, and of the MOXD2 and EPHB6 genes, which delimit the TRB locus. Their classification, location and predicted functionality are reported in Section 2.1 and Table 3 in the companion data paper by Antonacci et al. (2017).

3.2. Classification of the dromedary TRBV genes

We annotated 33 TRBV germline genes, which could be assigned to 26 distinct subgroups. Their functionality was defined based on the IMGT rules as described in Antonacci et al. (2017). Five subgroups are multimembers (TRBV5, TRBV7, TRBV12, TRBV15 and TRBV21) with a limited number of genes (from 2 to 3). Twenty-five genes were predicted to be functional (approximately 76%) and eight are pseudogenes (see Tables 1 and 2 in Antonacci et al., 2017.). Six pseudogenes (TRBV3, TRBV9, TRBV12, TRBV14, TRBV23 and TRBV24) are the only members of their own subgroup. Therefore, the dromedary putative functional germline TRBV repertoire consists of at least 20 subgroups with a limited expansion.

To classify the dromedary TRBV gene subgroups, the

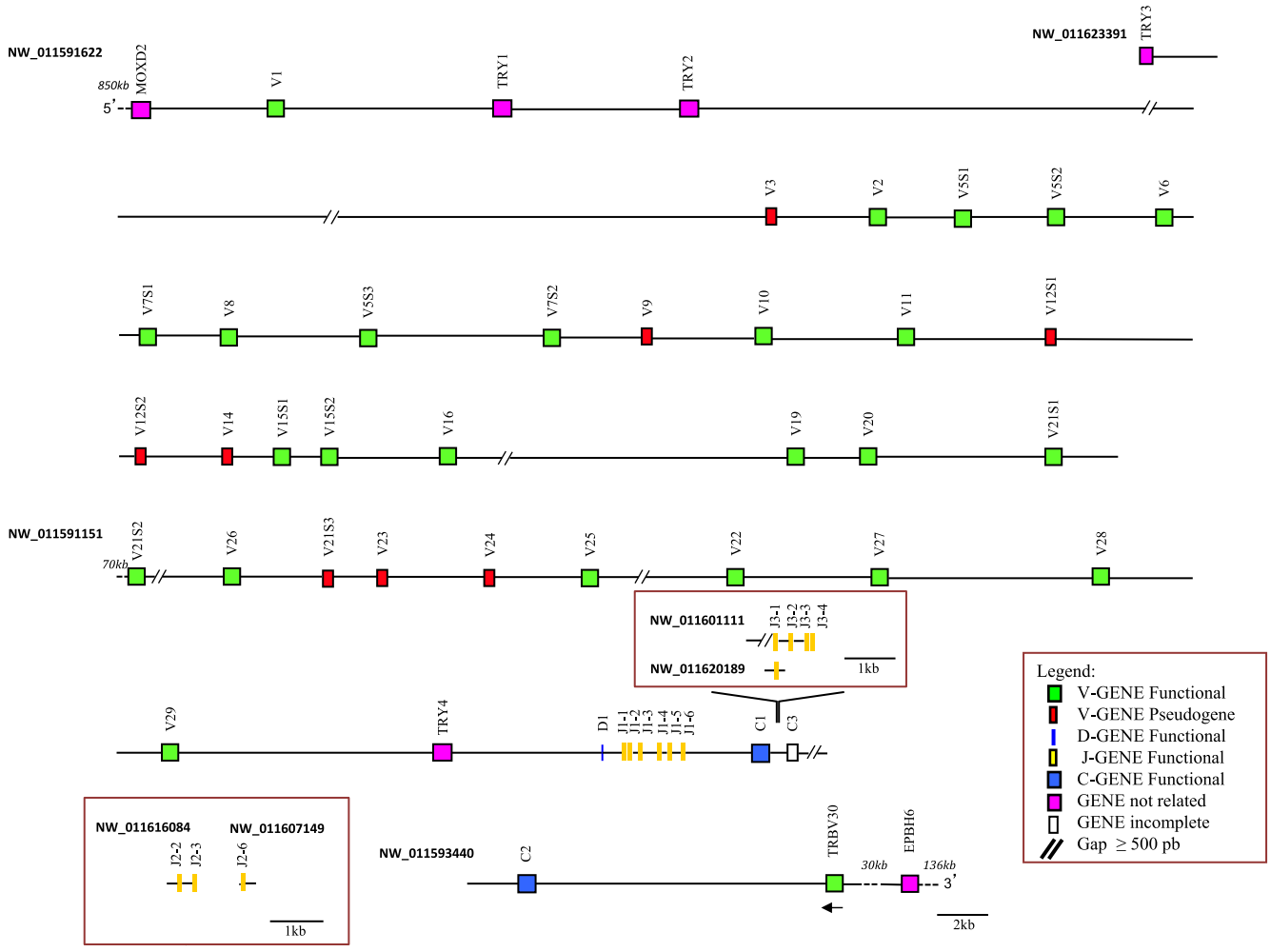

Fig. 1. Schematic representation of the genomic organization of the dromedary TRB locus deduced from the genome assembly Ca_dromedarius_V1.0. The unplaced genomic scaffolds are inserted in the map in a potential localization (see text). The diagram shows the position of all related and unrelated TRB genes according to nomenclature. Boxes representing genes are not to scale. Exons are not shown. Arrows indicate transcriptional orientation of the TRBV genes.

evolutionary relationship of these genes was investigated by comparing all dromedary with available genes corresponding to humans, dogs and rabbits by adopting two selection criteria: (1) only potential functional genes and in-frame pseudogenes (except for human TRBV1) were included; (2) only one gene for each of the subgroups was selected. Thus, the V-REGION nucleotide sequences of all selected TRBV genes, were combined in the same alignment and an unrooted phylogenetic tree was made using the NJ method (Saitou and Nei, 1987) (Fig. 2). The tree shows that each of 26 dromedary subgroups form a monophyletic group with a corresponding human and, if present, with a dog and rabbit gene, consistent with the occurrence of distinct subgroups prior to the divergence of the different mammalian species. Therefore, according to phylogenetic clustering, we classified each dromedary TRBV subgroup as orthologous to their corresponding human subgroups. The only exception is the dromedary TRBV9 gene, which was named as the rabbit gene on the basis of its genomic position within the dromedary (Fig. 1) as well as the rabbit (Antonacci et al., 2014) TRB locus, even if it is related to the human TRBV13 gene. Four human TRBV subgroups were not found in dromedary, including TRBV13 and TRBV17, which are also missing in the rabbit and dog genomes (Mineccia et al., 2012; Antonacci et al., 2014), and TRBV4 and TRBV18 that lack only in the dromedary genome.

It should be noted that the dromedary TRBV1 gene groups together with the corresponding dog and rabbit genes and lack a human orthologous.

3.3. Molecular characterization of the D-J-C region and analysis of the gene content

To confirm the presence of three D-J-C clusters in the dromedary TRB locus, we isolated the 3′ region of the locus performing long-PCR experiments from genomic DNA (see Section 2.3 in the companion data paper by Antonacci et al., 2017). We obtained a 22 722 bp region (Acc. N° LT837971) encompassing the 3′ end of the dromedary TRB locus that was annotated based on the sheep D-J-C region (Antonacci et al., 2008). The gene content of the sequenced region is schematically illustrated in Fig. 3. To reconstruct and analyses the entire D-J-C-V region, we have completed the sequence by adding the TRBD1 to the 5′ and the TRBV30 gene to the 3' end (in red in Fig. 3) retrieved from the genomic assembly. In summary, the region contains three in tandem D-J-C clusters over approximately 34 Kb. D-J-C cluster 1 spans 6488 bp and contains one TRBD, six TRBJ and one TRBC genes. D-J-C cluster 3 is located at about 2.2 Kb downstream of the cluster 1 with a total length of 6698 bp and includes one TRBD, seven TRBJ and one TRBC genes. Finally, D-J-C

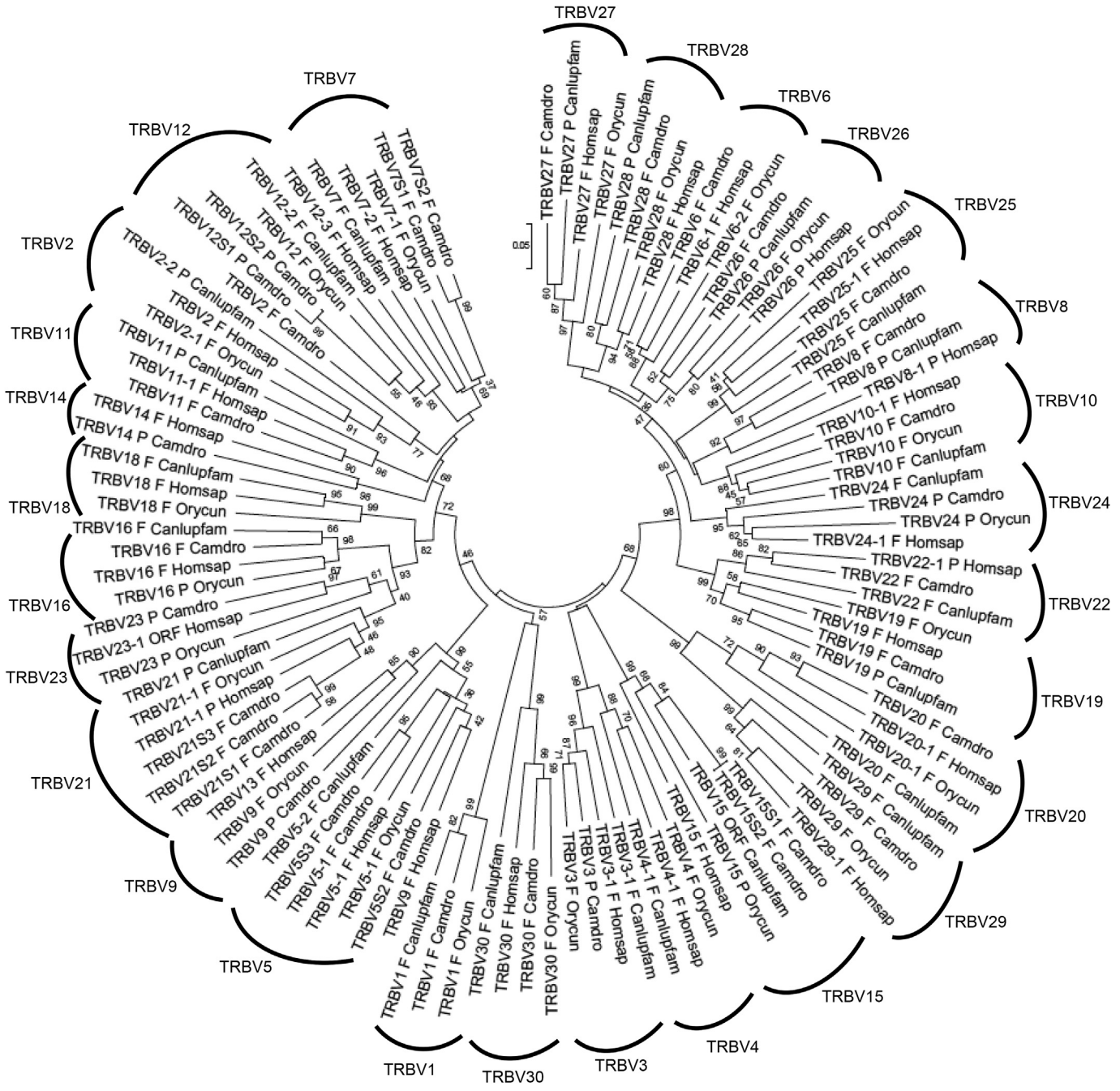

Fig. 2. The NJ tree inferred from the dromedary, human, dog and rabbit TRBV gene sequences. The evolutionary analysis was conducted in MEGA5 (Tamura et al., 2011). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). The trees are drawn to scale with branch lengths in the same units as those of the evolutionary distances used to infer phylogenetic trees. The evolutionary distances were computed using the p-distance method (Nei and Kumar, 2000) and are in the units of the number of base differences per site. The analysis involved 109 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were 149 positions in the final dataset. The dromedary TRBV subgroup classification is performed according to the clustering with the orthologous mammalian TRBV subgroups. The gene functionality according to IMGT rules (F: functional, ORF: open reading frame, P: pseudogene) is indicated. The IMGT 6-letter for species (Homsap, Camdro, Orycun) and 9-letter for subspecies (Canlupfam) standardized abbreviation for taxon is used.

cluster 2 is positioned at approximately 2.2 Kb downstream of the cluster 3 and extends over 5577 bp with one TRBD, six TRBJ and one TRBC genes. At approximately 10 Kb, separated from the TRBC2 gene, lies the TRBV30 gene.

The nucleotide and deduced amino acid sequences of the three TRBD genes are shown in Fig. 2a in Antonacci et al. (2017). They consist of a 12 bp (TRBD1), 16 bp (TRBD3) and extraordinarily 32 bp (TRBD2) G-rich stretches. The TRBD1 and TRBD3 genes can be productively read through their three coding phases and they encode 2–3 glycine residues, depending on the phase, whereas the TRBD2 gene can be read only in two phases encoding 4 glycine residues in both the phases. The RSs that flank the 5′ and 3′ sides of the coding region are well conserved with respect to the consensus.

The nucleotides and deduced amino acid sequences of all the TRBJ genes identified in the region were reported in Fig. 2b in Antonacci et al. (2017). They are classified according to the

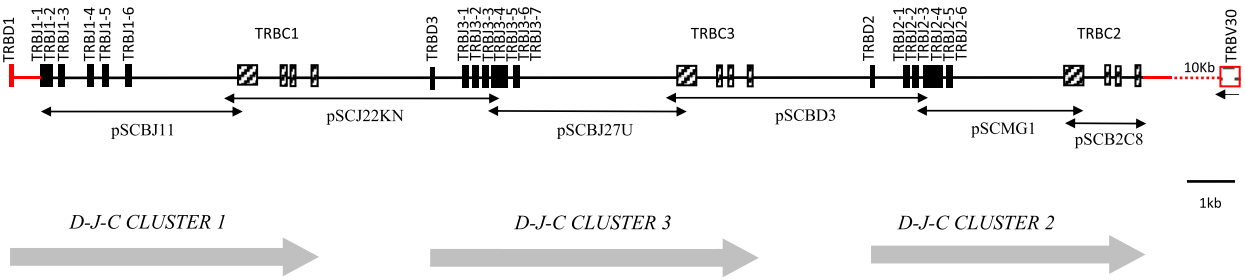

Fig. 3. Schematic map of the 3'-end of the dromedary TRB locus inferred from the analysis of the clones obtained by long PCR and the analysis of the genomic assembly (in red). The name of each PCR clone is also indicated. Boxes indicate the positions of the TRBD, TRBJ, TRBC and TRBV genes identified in the region with the cluster names indicated below. The underlying arrow indicates the transcriptional orientation of the TRBV30 gene. The genes are not to scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

following TRBC gene and numbered in agreement with the genomic position within the respective cluster. All genes are typically 44–53 bp long and conserve the canonical FGXG amino acid motif, whose presence define the functionally of J genes. The only exception is the TRBJ3-6 gene, which presents the FGAR motif. Each TRBJ is flanking by a 12 RS at the 5′ end and by a donor splice site at the 3′ end. All the RSs are well conserved with respect to the consensus. The identified TRBJs match perfectly with those found in the genome assembly and in the cDNAs (see next section).

As widely recognized in all known mammalian species, the three dromedary TRBC genes are composed of four exons and three introns. The four exons measure 393, 18, 107 and 21 bp, respectively. The 3′-UTR region that is 217 bp long in TRBC1 and TRBC3 and 195 bp long in TRBC2 is also present in the fourth exon. All three TRBC genes encode a similar protein of 178 amino acid with the extracellular domain encoded by the exon 1, exon 2 and by the first codon of exon 3. The transmembrane region is encoded by the remaining part of exon 3; while the cytoplasmic portion is encoded by exon 4 (Fig. 2c in Antonacci et al., 2017).

When we compared the nucleotide sequences of the TRBC1, TRBC2 and TRBC3 coding regions, we observed that TRBC1 and TRBC3 differ for seven nucleotides, which are all located in the first exon with residue changes at the amino acid level (Fig. 2c in Antonacci et al., 2017). TRBC3 and TRBC2 genes change for five nucleotides (one is localized both in the first and second, and three in the fourth exon). Overall, TRBC1 and TRBC2 differed for twelve nucleotides, and eight of these were located in the first exon, including one in the second and three in the fourth exon. These differences result in nine amino acid variations between TRBC1 and TRBC2, seven of which are clustered in the extracellular domain. The remaining two variations are localized in the cytoplasmic portion. Therefore, the TRBC3 isotype is identical to the TRBC2 gene at the amino acid level in the extracellular domain, but it produces results similar to the TRBC1 gene in the cytoplasmic portion. The transmembrane domain is completely conserved in all three gene products.

The 3′UTR of TRBC3 is identical to TRBC1, whereas it differs from TRBC2. Since there are no nucleotide differences between the sequences of the TRBC1 and TRBC2 first exons identified in the assembly, the variations that we found might be considered polymorphisms. For the same reason, we did not confirm the variation in the second exon in the TRBC2 of the assembly while only two of the three variations found in the genomic assembly, are present in exon 4.

Minimal differences in intron length were detected for the three TRBC genes. The first intron is 444 bp long for TRBC3 and TRBC2 compared to 472 bp long in the TRBC1 gene. The intron 2 is 160 nucleotides long for all three genes, while little difference has been identified in the third intron (314, 312 and 283 bp long for TRBC1, TRBC3 and TRBC2, respectively). In contrast, large nucleotide composition differences have been observed in the first intron of TRBC1 with respect to TRBC2 and TRBC3, while the third intron of TRBC2 differs significantly from TRBC1 and TRBC3. Table 1 summarizes the percentage of nucleotide identity for each exon and intron calculated in pair-wise combinations between the three TRBC genes. In this respect, it can be noticed (in bold in Table 1) that TRBC3 shows a high level of nucleotide identity in its first half part with TRBC2, while from exon 2 to 3′UTR, the results are almost identical to the TRBC1 gene. The most peculiar feature of the dromedary Cβ protein is the presence of one or two additional amino acids located in the FG loop (position 105–117 as in IMGT definition) with respect to sheep and cattle (Antonacci et al., 2008; Connelley et al., 2014), or humans, dogs and pigs (Eguchi-Ogawa et al., 2009; Mineccia et al., 2012; Lefranc et al., 2015), respectively. Moreover, the alignment of the amino acids of the FG loop of selected TRBC2 sequences across mammalian shows diversity in length as well as in amino acid composition, with only nine residues conserved (Supplementary Figure S1).

3.4. Phylogenetic analysis of the dromedary TRBJ genes

To explore the phylogenetic relationship of the TRBJ genes with respect to the TRBC genes in each cluster, the nucleotide sequences of all dromedary TRBJ genes (coding region plus RS) were aligned with human, dog, sheep and pig sequences where the genomic location is known and the relative nomenclature was adopted. The alignment was then used to construct a phylogenetic tree with the NJ method (Fig. 4). The results of the analysis support a distribution of the 19 dromedary TRBJ genes into 13 phylogenetic groups. If the TRBJ1 genes were distributed into six groups, with each one containing orthologous TRBJ genes indicated by their conserved germline order within the different TRB loci, the TRBJ2 genes were distributed both with the respective genes of humans and dogs and with the TRBJ3 paralogous genes of sheep, pigs and dromedary as well. The only exception is the closer phylogenetic relationship of the dog TRBJ2-4 and TRBJ2-5 genes with respect to the other orthologues (Mineccia et al., 2012). This result indicates the close phylogenetic relationship of the TRBJ2 and TRBJ3 genes as a consequence of the duplication event occurring in the artiodactyl species. Fig. 5 is a schematic representation of the orthologous and paralogous TRBJ genes as deduced by the phylogenetic tree. Prior to the mammalian radiation, the TRBJ2 cluster made up of seven genes, which are then duplicated to form the third TRBJ cluster. It should be noted that the evolution within the different mammalian species has preserved the tight relationship between TRBJ genes, which indicates the existence of a strong functional constraint. Alternatively, the time spent on the duplication event may not be sufficient to generate changes in gene function within each lineage.

Table 1
Percentage of nucleotide identity between each exon/intron of the dromedary TRBC genes.

	EX1 (393 bp)	INT1	EX2 (18 bp)	INT2	EX3 (107 bp)	INT3	EX4 (CDS) (21 bp)	EX4 (3′-UTR)
C1 vs C2	97.9	73.8	94.4	97.5	100	51.2	85.7	48.3
C1 vs C3	98.2	74.9	**100**	**100**	**100**	**98.7**	**100**	**100**
C2 vs C3	**99.7**	**98.8**	94.4	97.5	100	51.6	85.7	48.3

The highest level of nucleotide identity between genes is indicated in bold.

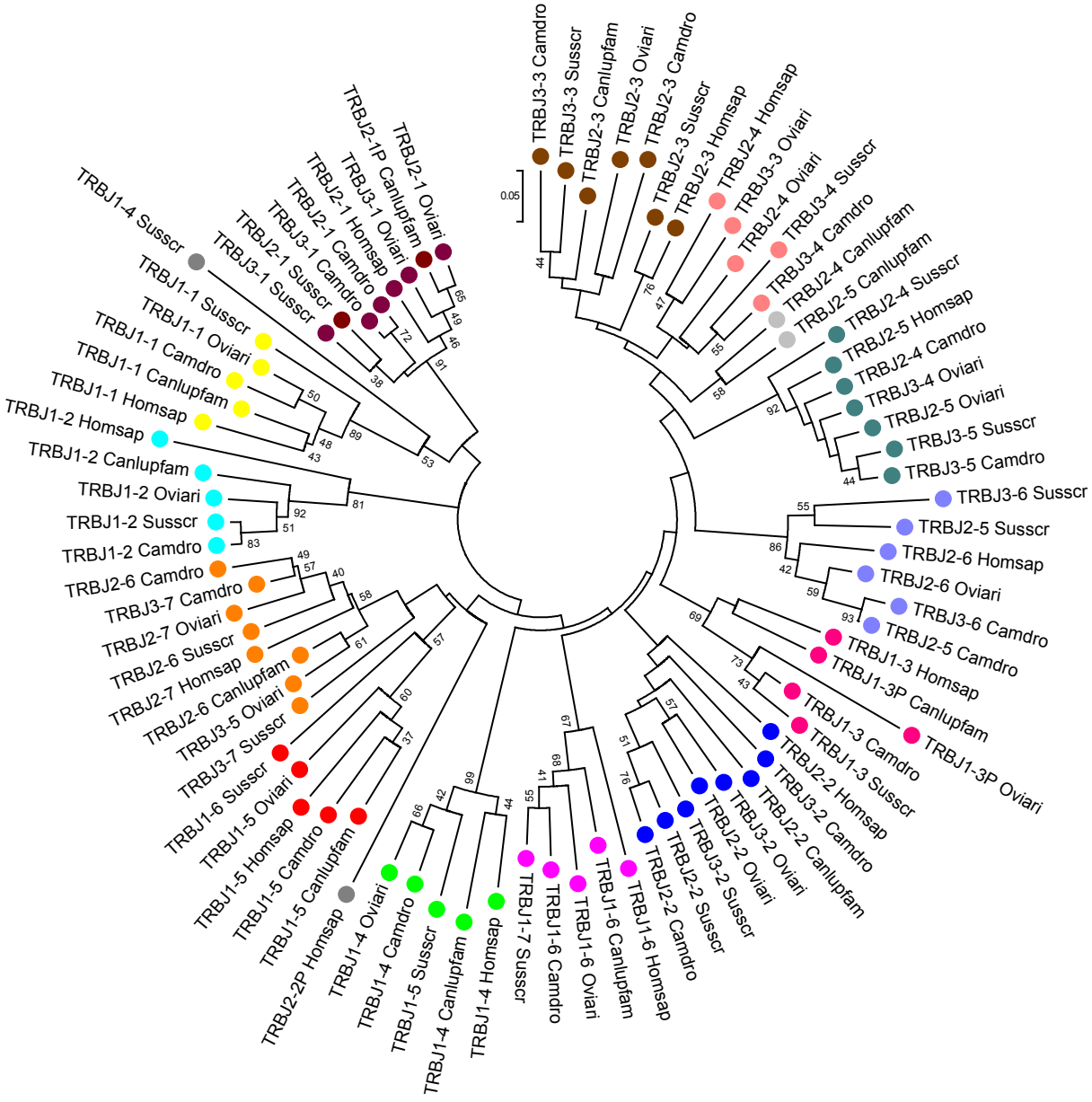

Fig. 4. The NJ tree inferred from the dromedary, human, dog, sheep and pig TRBJ gene sequences. The different colours highlight the distribution of the TRBJ genes into phylogenetic groups. The evolutionary analysis was conducted in MEGA5 (Tamura et al., 2011). The optimal tree with the sum of branch length = 8.06768288 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei and Kumar, 2000) and are in the units of the number of base differences per site. The analysis involved 83 nucleotide sequences. The codon positions included 1st+2nd+3rd + Non-coding. All positions containing gaps and missing data were eliminated. There were 48 positions in the final dataset. The IMGT 6-letter for species (Homsap, Camdro, Oviari, Susscr) and 9-letter for subspecies (Canlupfam) standardized abbreviation for taxon is used.

3.5. 5' RACE PCR assay

To evaluate the functional competency of the three predicted D-J-C clusters from the genomic analysis we performed a 5′ RACE assay. For this purpose, we set up two 5' RACE experiments on total RNA isolated from the spleen of two unrelated animals using a TRBC-specific primer. The two RACE products were cloned into the TA-vector and randomly selected positive clones for each cloning

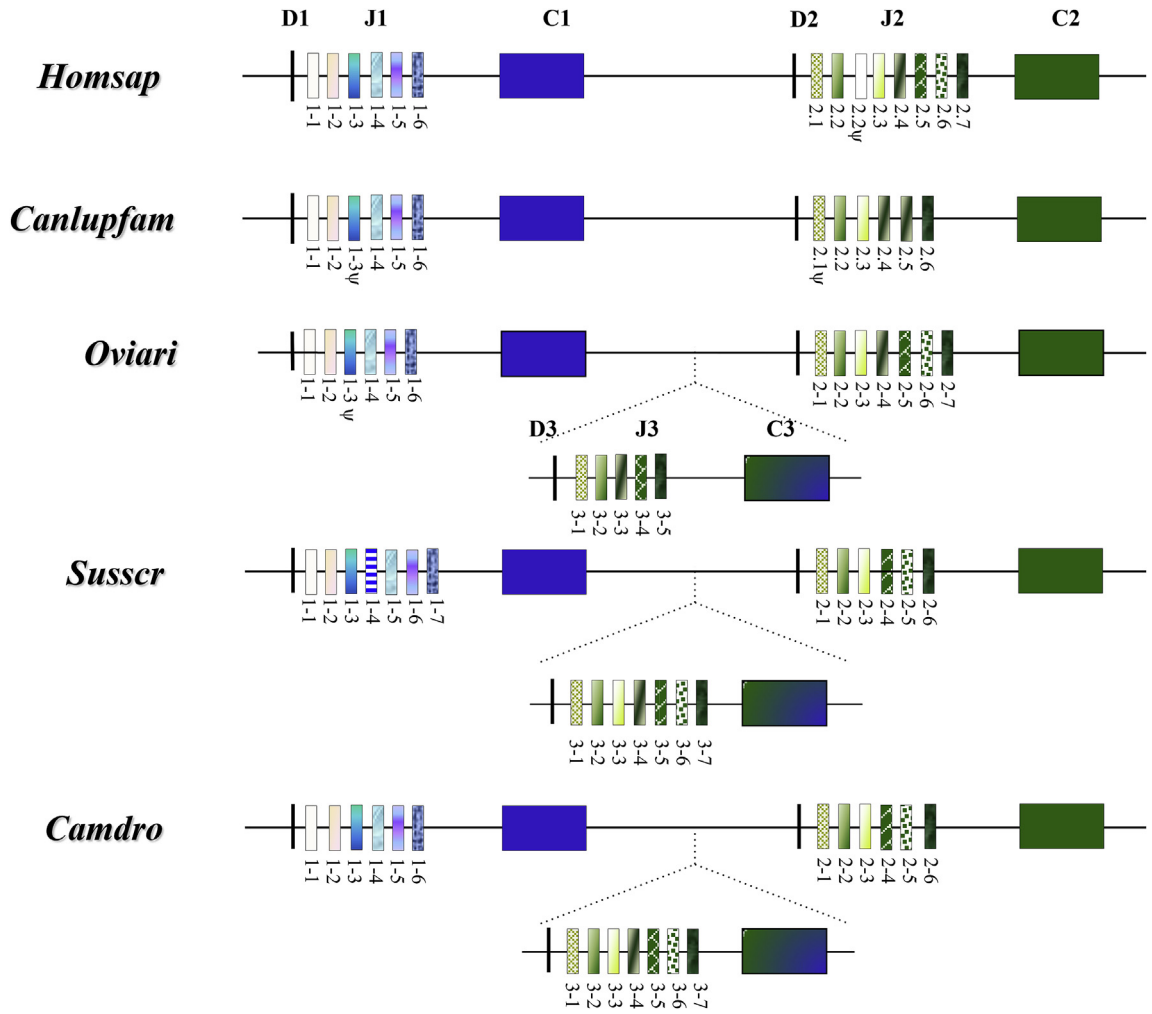

Fig. 5. Schematic representation of the D-J-C clusters within the TRB locus of the different mammalian species. The orthologous and paralogous TRBJ genes are depicted with the same colour as deduced by the phylogenetic tree. The double colour TRBC3 gene indicates the homologies with the TRBC1 and TRBC2 genes as result of the unequal crossover that generated the new TRBD–J–C cluster in the artiodactyl species (Antonacci et al., 2008; Eguchi-Ogawa et al., 2009; this study).

were sequenced. A total of 35 diverse clones of different lengths containing rearranged V–(D)–J–C transcripts with a correct open reading frame were obtained and named according to the cDNAs sample through which they were derived (5R1 and 5R2). Each sequence was manually analysed to identify the TRBV, TRBD and TRBJ genes through alignment with the germline dromedary TR genes. Fig. 6 shows the deduced amino acid sequences of the V-D-J region of all cDNA clones according to IMGT unique numbering for the V-REGION and V-DOMAIN (Lefranc et al., 2003).

In the cDNA clones, we identified 14 different TRBV genes belonging to 11 different subgroups. Only one clone was found for each of six (TRBV7, TRBV16, TRBV20, TRBV26, TRBV28 and TRBV29) subgroups and only the 5R2B55 clone sequence perfectly matched the corresponding TRBV26 germline gene sequence. The remaining five clone sequences showed a nucleotide identity ranging from 97% to 99% with respect to the reference TRBV germline gene sequence (data not shown). Based on the assumption that sequences sharing 97–99% nucleotide identity represent the same gene (Arden et al., 1995a, 1995b), we referred to these sequences as new alleles. Two different clones (5R1B3 and 5R2B2) were found for the TRBV5 subgroup and both showed a nucleotide identity of 99.7% with respect to the germlines TRBV5S2 and TRBV5S3, respectively. In agreement with this result, we considered the two clones as alleles of the two germline TRBV5 genes. Additionally, the two clones of the TRBV27 subgroup (5R1B28 and 5R2B12) can be considered alleles because the nucleotide identity is 98.8 and 99.1% (which was consistent in one amino acid variation), respectively compared to the corresponding germline gene (Fig. 6).

The number of clones assigned to the TRBV15, TRBV21 and TRBV25 subgroups is significant. Seven TRBV15 clones can be divided into three groups, which each have the same TRBV gene, showing a nucleotide identity ranging from 98.2 to 98.8% with respect to the two genomic identical TRBV15 sequences (Supplementary Figure S2). The nucleotide identity is still above the 97% in a comparison between the three types of clones. The differences among the TRBV correspond to 4 or 5 amino acid changes. Based on these results, we have classified these new TRBV15 genes as alleles. We cannot determine which of the two germline TRBV15 genes the alleles match because of the identical sequence of the two genes. For this reason, we classify the three alleles with consecutive numbers of the TRBV15S1 gene (Fig. 6).

Three groups of clones based on the differences of the TRBV21 gene sequences were also obtained. The first group with a nucleotide identity of 98.5% that has been classified as an allele of the germline TRBV21S1 gene, whereas the other two groups were both considered to be alleles of the TRBV21S2 gene since they exhibit a

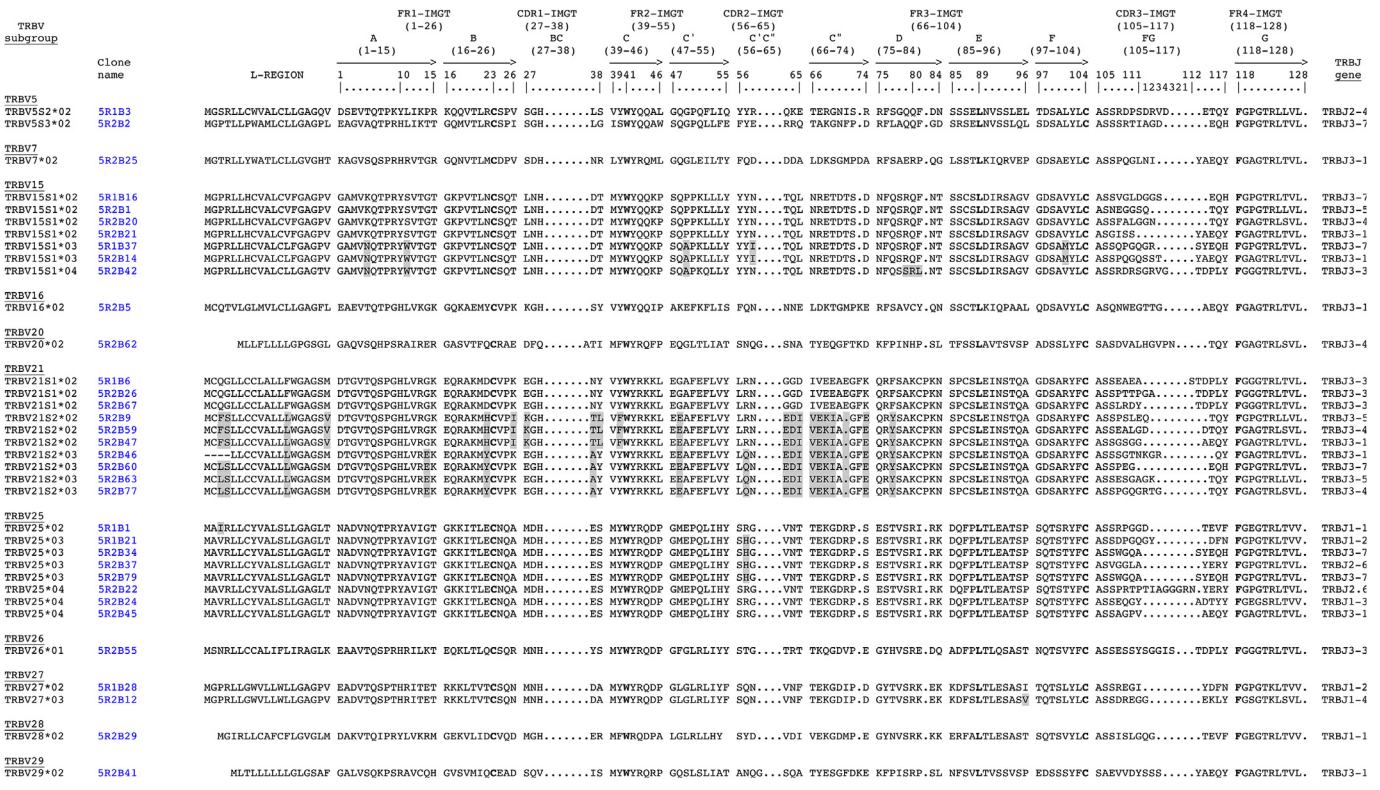

Fig. 6. Protein display of the cDNA clones. The TRBV and TRBJ genes, named according to the criteria specified in the text, are listed, respectively, at the left and the right of the figure. Leader region (L-REGION), complementarity determining regions (CDR-IMGT) and framework regions (FR-IMGT) are also indicated according to the IMGT unique numbering for the V-DOMAIN (Lefranc et al., 2003). The five conserved amino acids of the V-DOMAIN (1st-CYS 23, CONSERVED-TRP 41, hydrophobic AA 89, 2nd-CYS 104 and J-PHE 118) are indicated in bold. The TRBV allele amino acid changes, if any, are boxed. The name of the clones is also reported. The J3-1 may be also J2-1, and the J3-5 may be also J2-4.

nucleotide identity of 97.3% and 99.4% with respect to the germline TRBV21S2 gene (Supplementary Figure S2). However, since the percentage of nucleotide identity between the TRBV21S2*02 and TRBV21S2*03 alleles is 96.7% (Supplementary Figure S2) with nine amino acid changes, we could not exclude the presence of a third functional TRBV21 gene. To verify the presence of an additional functional TRBV21 gene in the dromedary genome, we performed a PCR experiment on DNA samples derived from one of the same unrelated animals, using a primer pair designed on the first and second exon in a region of homology among the different TRBV21 genes and alleles. When the PCR product was cloned and randomly selected positive clones were sequenced, we recovered three groups of sequences whose coding regions perfectly match with each of the three cDNA groups. The comparison of the entire sequence, including the introns, of these genomic clones with the germline TRBV21 genes derived from the assembly might confirm the presence of only two different TRBV21 genes since the gV21-15 sequence that corresponds to the TRBV21S2*02 allele (5R2B9 cDNA group), showed a nucleotide identity >97% with the gV21-13 genomic sequence whose coding region corresponds to the TRBV21S2*03 allele (5R2B60 cDNA group) (lower part of Supplementary Figure S2).

Finally, eight TRBV25 cDNA clones can be divided into three groups with the same TRBV gene, with a nucleotide identity of approximately 99% (consistent in one amino acid variation) among them and with respect to the only germline TRBV25 gene. Along with these results, we classified the TRBV genes identified in the cDNA clones as alleles of the TRBV25 subgroup (upper part of Supplementary Figure S2).

The TRBJ genes were unambiguously classified with respect to the corresponding genomic sequences, in 26 out of 35 cDNA clones for a total of 10 different TRBJ genes (Fig. 6). The TRBJ3 cluster seems to be preferentially expressed (17/26), while six clones retain a member of the TRBJ1 cluster and two clones retain the TRBJ2 gene set. The remaining nine cDNA clones contain a TRBJ gene not discernible between TRBJ3-1 and TRJB2-1 or between TRBJ3-5 and TRBJ2-4 because of their 100% sequence identity.

More complex is the determination of the contribution of the TRBD genes involved in the CDR3 formation. For a close inspection, the nucleotide sequences corresponding to the CDR3 have been excised from each cDNA and analysed in detail (Fig. 7). By the comparing with the TRBD genomic sequences, the nucleotides located in the CDR3 region were considered to belong to a TRBD gene if they constituted a stretch of at least five consecutive nucleotides corresponding to the TRBD1, TRBD3 and TRBD2 germline sequences. In this way, the TRBD was unambiguously identified in 21 out of 35 sequences (60%) with the TRBD1 present in 16, the TRBD3 in three and TRBD2 in two clones, respectively. The remaining 14 sequences either do not have an identifiable TRBD gene (seven clones), or it is not possible to distinguish between TRBD3 and TRBD2 genes because of their similar germline sequence (eight clones). The absence of a TRBD region could be interpreted as direct V-J junction. However, it is also possible that nucleotide trimming masked the initial participation of the TRBD gene during the rearrangement. In one case (5R2B46), there is the presence of two TRBD genes in the CDR3 region. A particular feature of the CDR3 region of the clones are the presence within the D region of nucleotide substitutions (5R2B5, 5R2B37, 5R2B63 and 5R2B67) or insertions (5R2B22) with respect to the germline sequences. The corresponding amino acid sequence of the CDR3 loop deduced from the nucleotide sequences reveals that it is heterogeneous for amino acid composition (Figs. 6 and 7). The mean

clone	TRBV	CDR3	TRBD	TRBJ
5R1B1	25	GCCAGCAGTagaccGGGGGGAGACACTGAAGTT A S S R P G G D T E V	D3°	1-1
5R1B3	5	GCCAGCAGCCGTGAcccatcGGACAGGGttgatGAGACCCAGTAC A S S R D P S D R V D E T Q Y	D1	2-4
5R1B6	21	GCCAGCAGTGAAgccgaggcgAGCACAGATCCTCTGTAT A S S E A E A S T D P L Y	-	3-3
5R1B16	15	GCCAGCAGTgtcgggctcgacGGGGGGAGTGAGCAGCAT A S S V G L D G G S E Q H	D3°	3-7
5R1B21	25	GCCAGTAGTGAcccaGGACAGGGTTATGACTTTAAC A S S D P G Q G Y D D F N	D1	1-2
5R1B28	27	GCCAGCAGCcgcgAGGGGatTTATGACTTTAAC A S S R E G I Y D F N	D1	1-2
5R1B37	15	GCCAGCAGccagcccGGACAGGGaagGTCCTATGAGCAGCAT A S S Q P G Q G R S Y E Q H	D1	3-7
5R2B1	15	GCCAGCAatgaGGGGGGATCACAGACCCAGTAC A S N E G G S Q T Q Y	D3°	3-5
5R2B2	5	GCCAGCAGctccaGGACTATagcgggtgaTGAGCAGCAT A S S S R T I A G D E Q H	D3°	3-7
5R2B5	16	GCCAGCcaaaatTGGGAGGGAaccacgggtGCTGAGCAGTAC A S Q N W E G T T G A E Q Y	D3	3-1*
5R2B9	21	GCCAGCAGTccaagcctggagCAGACCCAGTAC A S S P S L E Q T Q Y	-	3-5
5R2B12	27	GCCAGCAGCGaccgcgagggtggTGAAAAACTCTAT A S S D R E G G E K L Y	-	1-4
5R2B14	15	GCCAGCAGccctCAGGGccaatcatccaCCTACGCTGAGCAGTAC A S S P Q G Q S S T Y A E Q Y	D1	3-1*
5R2B20	15	GCCAGCAGTttcgctcTGGGGGGGAACACTCAGTAC A S S F A L G G N T Q Y	D3	3-4
5R2B21	15	GCCAGCggtataagCTCCTACGCTGAGCAGTAC A S G I S S Y A E Q Y	-	3-1*
5R2B22	25	GCCAGCAGTcccaggacccaACTATAGCGGGGGGAGgtaggaaCTATGAGAGGTAT A S S P R T P T I A G G G R N Y E R Y	D3	2-6
5R2B24	25	GCCAGTAGTGAACAGGGatatgCAGACACCTACTAT A S S E Q G Y A D T Y Y	D1	1-3
5R2B25	7	GCCAGCAGCCcACAGGGactgaacatCTACGCTGAGCAGTAC A S S P Q G L N I Y A E Q Y	D1	3-1*
5R2B26	21	GCCAGCAGcccgaccaccccGGGGGCCACAGATCCTCTGTAT A S S P T T P G A T D P L Y	D1	3-3
5R2B29	28	GCCAGCAGTATATCgctaGGACAGGGaaCTGAAGTTTTC A S S I S L G Q G T E V F	D1	1-1
5R2B34	25	GCCAGtagctgGGGACAGGcGTCCTATGAGCAGCAT A S S W G Q A S Y E Q H	D1	3-7
5R2B37	25	GCCAGCGTCGGGGGgctcgCCTATGAGAGGTAT A S V G G L A Y E R Y	D2	2-6
5R2B41	29	AGCGCTGAAGTTGTtgactatagctcCTCCTACGCTGAGCAGTAC S A E V V D Y S S S Y A E Q Y	D2	3-1*
5R2B42	15	GCCAGCAGTAGAGAtcggtccGGGCGGGtcggaACAGATCCTCTGTAT A S S R D R S G R V G T D P L Y	D3	3-3
5R2B45	25	GCCAGCAGTGccggtcccgtcgccGAGCAGTAC A S S A G P V A E Q Y	-	3-1*
5R2B46	21	GCCAGCAGTtccGGGACAaacaaGGGGAGGCAGTAC A S S S G T N K G R Q Y	D1/D3°	3-1*
5R2B47	21	GCCAGCAGTGggtCAGGGGGCGCTGAGCAGTAC A S S G S G G A E Q Y	D1	3-1*
5R2B55	26	GCCAGCAGTGAATCgtcctatagcGGGGGGAttAGCACAGATCCTCTGTAT A S S E S S Y S G G I S T D P L Y	D3°	3-3
5R2B59	21	GCCAGCAGTGAggccctAGGGGacgACACTCAGTAC A S S E A L G D D T Q Y	D1	3-4
5R2B60	21	GCCAGCAGccctgagggaGAGCAGCAT A S S P E G E Q H	-	3-7
5R2B62	20	AGTGCTAGTGAtgttgccctgcacggagttccGAACACTCAGTAC S A S D V A L H G V P N T Q Y	-	3-4
5R2B63	21	GCCAGCAGTGAAtcgGGGGCAGGGaAGACCCAGTAC A S S E S G A G K T Q Y	D1	3-5§
5R2B67	21	GCCAGCAGTCTAAGGGattaCACAGATCCTCTGTAT A S S L R D Y T D P L Y	D3°	3-3
5R2B77	21	GCCAGCAGTcccGGGCAGGGccggacgggCACTCAGTAC A S S P G Q G R T G T Q Y	D1	3-4
5R2B79	25	GCCAGtagctgGGGACAGGcGTCCTATGAGCAGCAT A S S W G Q A S Y E Q H	D1	3-7

*TRBJ3.1 or TRBJ2.1 §TRBJ3.5 or TRBJ2.4 ° TRBD3 or TRBD2

Fig. 7. CDR3 nucleotide and predicted amino acid sequences retrieved from the TRB cDNA clones. CDR3-IMGT sequences are shown from codon 105 (the codon after the 2nd-CYS 104 of the V-REGION) to codon 117 (the codon before J-PHE 118 of the J-REGION) according to the unique numbering (Lefranc et al., 2003). The CDR3 nucleotide/amino acid sequence, and the classification of the TRBV and TRBJ genes of each clone are also listed. Nucleotides of the 3′V-REGION and of the 5′J-REGION are indicated in uppercase letters. The sequences considered to present recognizable TRBD genes are indicated in red uppercase letters and the nucleotide substitutions or insertions are underlined. Nucleotides that cannot be attributed to any V, D or J region (*N*-nucleotides), are indicated in lower cases on the left and on the right sides of the TRBD regions. The name of the clones is reported. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

length of the CDR3 loop was 12.8 amino acids (range 9–19 amino acids). A similar CDR3 length and size range were reported in the sheep spleen (means 12.3 amino acids, range 10–16 amino acids) (Di Tommaso et al., 2010).

Since the genomic organization of the D-J-C is known (Fig. 3), a formal interpretation of the DJC arrangement is also possible. The

intracluster rearrangements are present in seven clones with four TRBD1-TRBJ1, two TRBD3-TRBJ3 and one TRBD2-TRBJ2 rearrangements. Eight rearrangements can be interpreted by direct 5'- to- 3′ joining across the clusters (intercluster rearrangements), with seven TRBD1-TRBJ3 and one TRBD1-TRBJ2 (Fig. 7). Interestingly, we also observed one TRBD3 or TRBD2-TRBJ1 joining (5R1B1 clone). Since the D-J-C cluster 3 or D-J-C cluster 2 are located downstream of D-J-C cluster 1 within the TRB locus, this junction may be explained by chromosomal inversion, or by *trans*-rearrangement with more probability, which occurs during TRB locus recombination.

3.6. Architecture and identification of the regulatory elements in the dromedary D-J-C region

To delineate the genomic structure of the D-J-C region of the dromedary TRB locus as determined in this study, we conducted a comparative analysis with the corresponding region in sheep. For this purpose, a 24 Kb dromedary sequence was screened from TRBD1 to TRBC2 genes with the RepeatMasker program to highlight the compositional properties and to identify the interspersed repeats. For comparison, the corresponding region in sheep (26 Kb) was tested with the same program. The results are summarized in Supplementary Table S1. If the GC content is similar in both species, the total number of interspersed repeats are lower in dromedary compared to sheep (10.7% vs 17.5%). The difference is mainly due to the presence of a lower percentage of short interspersed elements (SINEs) in dromedary (seven elements, 4.8%) compared to sheep (14 elements, 9.2%). Mammalian-wide interspersed repeat (MIR) elements are the only type of SINE present in dromedary, while SINE/tRNA in sheep are the most abundant (10 elements, 7.1%). Long interspersed elements (LINEs) cover less than 4% of the whole region in both species with LINE2 being the most representative. This result is consistent with the general assumption that SINEs are predominantly found in GC- and gene-rich regions, whereas LINEs are most prevalent in low-GC and gene-poor regions. Two long terminal repeat (LTR) elements (1.6%) are present only within the region of the sheep sequence.

The dromedary masked sequence was then aligned with the sheep D-J-C sequence using the PipMaker program (Schwartz et al., 2000) and the alignment expressed as a percentage identity plot (pip) (Supplementary Figure S3). The pip shows the position of all genes identified in the region as well as the location and orientation of all repetitive sequences. It is noticeable there are the seven MIRs distributed along the entire region. Two of them are located upstream and downstream of the TRBJ3 and the TRBJ2 genes. The other three elements are positioned upstream each of the TRBC genes. A block of two LINE2 are located upstream of the TRBD3 and the TRBD2 genes. The comparison of the dromedary and sheep sequences showed extensive matches along the entire region, especially corresponding to the orthologous genes for which the number of coding exons, their size and sequence are well conserved. The occurrence of redundant matches that cause lines that appear to be superimposed in the pip is consistent with the presence of duplicate regions. The duplications are shown to be more clear by the dot-plot in Fig. 8, which is a complementary display of exactly the same alignment shown in the pip. In fact, besides the main diagonal line which indicates the high structure identity of the dromedary sequence with the sheep sequence, two parallel lines are evident and represent two repeated regions within the sequence. The smaller region of duplication involves TRBC1 versus TRBC2, while the longer repeated unit concerns TRBC1 plus TRBD-TRBJ-TRBC cluster 3 versus TRBC3 plus TRBD-TRBJ-TRBC cluster 2. However, a more consistent inspection of the dot-plot shows that all the three lines are interrupted. This is the result of the presence of a block of no-syntenic regions between the two species due to the presence of two LTR/MaLR elements in the first intron of the TRBC1 and TRBC2 sheep (Antonacci et al., 2008), which are missing in the dromedary genes (grey arrows in Fig. 8). Much smaller no-syntenic regions are distributed along the entire sequence with the presence of SINE/tRNA repeats in the sheep region (Antonacci et al., 2008) lacking in dromedary region.

The fully annotated sequences of the 3′ end of the TRB locus available for humans, mice and sheep allowed us to conduct a detailed comparative analysis to increase the genome annotation of the last part of the dromedary TRB locus and to identify functional sites such as regulatory elements. For this purpose, 35 Kb of the dromedary genomic sequence corresponding to the D-J-C-V region was aligned with the human, mouse and sheep counterparts using the PipMarker program (data not shown). In addition to the coding regions, we found three regions sharing 70% minimum of sequence identity. The first of these conserved non-coding regions (CNSs) spans 88 bp upstream of the TRBD1 gene (pos. 326–413 in the dromedary sequence). This region has been experimentally determined in mice as the Dβ1 promoter (PDβ1), which together with the beta enhancer element (Eβ), regulate the germline transcription of the D-J-C portion of the TRB locus (Doty et al., 1999; Yang et al., 2003). A four-species alignment shows conservation in the dromedary sequence of the TATAAA box as well as the binding sequence for the Ikaros/Lyf-1 GC-rich motif and the GATA binding sequence (Supplementary Figure S4a). All these sites were found to be essential for the murine promoter activity. We also identified a conserved CpG site in the dromedary heptamer of the 23 RS. Analysis of methylation levels surrounding RSs before, during and after recombination have demonstrated that demethylation of this CpG site appears to be essential for TRBD1 cleavage by the V(D)J recombinase in mice (Whitehurst et al., 2000).

A second similar CNS was identified proximal to the TRBD2 gene. In this case, the conserved region spans 290 bp upstream of the dromedary TRBD2 gene (pos. 17 934-18 2233). This region corresponds to the 5′PDβ2 promoter that was experimentally revealed in mice when the germline sequence at the 3′ of TRBD2 gene was deleted (McMillan and Sikes, 2009), which suggested there was promoter activity in TRB assembly after the Dβ2Jβ2 joining for this region. In the alignment, the dromedary and sheep corresponding regions proximal to the TRBD3 gene were also inserted (pos. 8924–9213 in the dromedary sequence) due to the high nucleotide identity with the dromedary and sheep TRBD2 region, respectively. A binding site for E protein (E box) as well as overlapping RNA polymerase II initiator (inr) elements and a TATAAA box site were conserved in the dromedary sequence (Supplementary Figure S4b).

The third CNS (pos 30 651-30 8388 in the dromedary sequence) is located approximately 7 Kb from the TRBC2 gene. This region corresponds to the Eβ element that is fully characterized in humans and mice and is located approximately 5 Kb away from the TRBC2 gene (Gottschalk and Leiden, 1990; Takeda et al., 1990). Seven protein binding motifs called βE1-βE7 were identified within the mouse Eβ, and a systematic analysis of the elements has demonstrated that a fragment containing the βE1–βE6 regions is required for full enhancer activity in pro-T cells as well as in mature cells (Carvajal and Sen, 2000). Inspection of the four-species alignment of the region shows a more consistent conservation in the dromedary sequence of the βE1, βE2, βE4 as well as its 3′ flanking sequence and βE6 corresponding regions (Supplementary Figure S4c).

4. Discussion

The extremely diversified and peculiar features of camels have

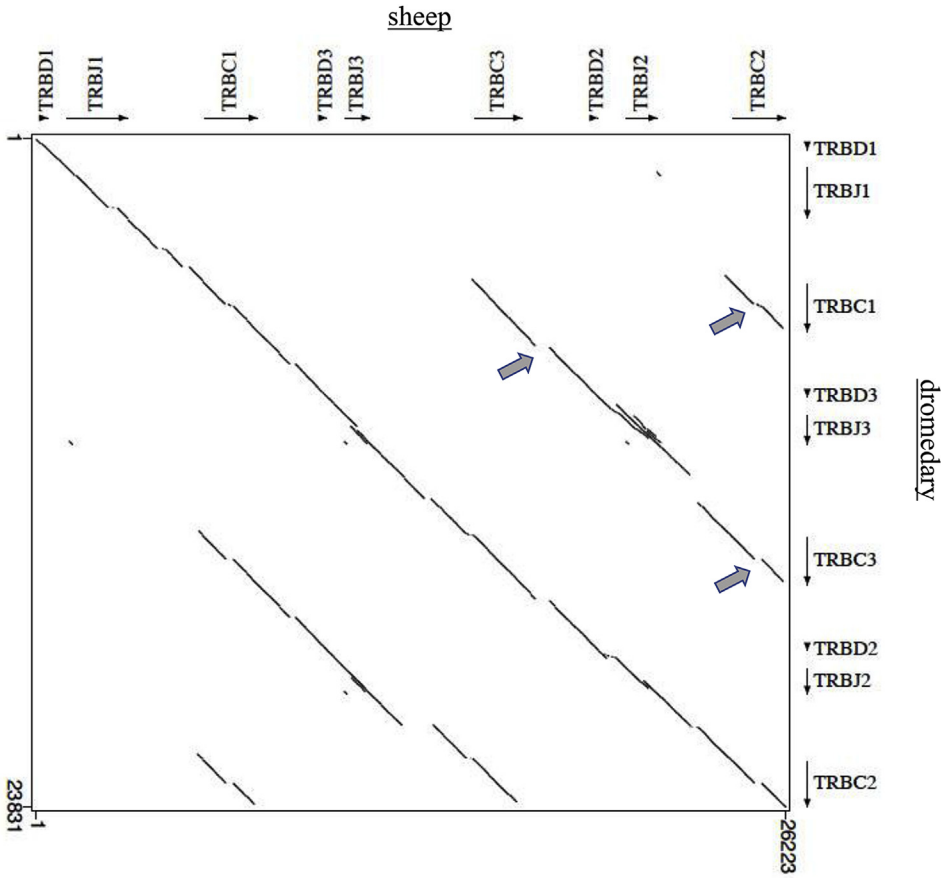

Fig. 8. Dot-plot matrix of dromedary/sheep TRBD-J-C sequence genomic comparison. The transcriptional orientation of each gene is indicated by arrows and arrowheads. The main diagonal line shows the high structure identity of the entire dromedary sequence with the sheep sequence. The pattern of parallel lines indicates the internal homology units between the TRBD-J-C clusters. The grey arrows indicate gaps (insertions or deletions) between the homology units.

prompted several research groups to investigate the genetic basis of certain relevant traits.

In this study, we investigated the nature of the dromedary TRB gene repertoire and its genomic organization and evolution (see also companion data paper by Antonacci et al., 2017). We have achieved this aim by integrating the eight unplaced and not continuous sequence scaffolds, which were isolated from genomic assembly (Wu et al., 2014) with sequences obtained by PCR experiments conducted in our laboratory and taking into account the general organization of the locus in the other mammalian species. No additional TRB genes were revealed from our expression analysis to reinforce the genomic data. This results in the first, mostly complete, map of the TRB locus in a species of the Tylopoda suborder.

Similar to the other species of mammals studied to date, the general genomic organization of the dromedary TRB locus consists of a pool of TRBV genes located upstream of in tandem TRBD-J-C clusters, followed by a TRBV gene with an inverted transcriptional orientation. MOXD2 and EPHB6 genes border the locus at the 5′ and 3′ end, respectively, while TRY genes are interspersed among TRBV genes and arranged in two distinct genomic positions (Fig. 1). A unique aspect of the dromedary TRB locus structure is the presence of three TRBD-J-C clusters (Fig. 3), which is a common feature of sheep, cattle and pig sequences (Antonacci et al., 2008; Eguchi-Ogawa et al., 2009; Connelley et al., 2014) and each cluster is typically composed of one TRBD, several TRBJ and one TRBC genes. Instead, two TRBD-J-C clusters appear in the other mammalian species, such as humans, rabbits and dogs (Mineccia et al., 2012; Antonacci et al., 2014; Lefranc et al., 2015). The analysis of the dromedary sequence revealed that the single TRBD3 and the seven TRBJ3 are correlated to the corresponding genes of the TRBD-J-C cluster 2 (Fig. 4); 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 (Table 1). This outcome suggests that the TRB genomic structure with three TRBD-J-C clusters was established through a duplication event due to an unequal crossing-over between the ancestral TRBC1 and TRBC2 genes that occurred prior the Tylopoda/Ruminantia/Suina divergence. Following duplication, subsequent species-specific diversifications have taken place and led to the current genomic organization of the 3′ end of the TRB locus in the different artiodactyl species.

The extra TRBD-J-C cluster in dromedary, when compared with humans, mice and dogs, does not seem to interfere with the general regulatory architecture of the dromedary TRB locus given the presence of a consistent conserved promoter region located upstream the TRBD-J-C cluster 1, and of an enhancer element, which is located downstream of the last cluster (Supplementary Figure S4a, c). Both the elements were demonstrated to govern the germline transcription of the TRB locus in mice (Doty et al., 1999; Yang et al., 2003). In addition, a conserved region has also been found upstream of the TRBD3 as well as TRBD2 genes (Supplementary Figure S4b), where promoter activity in the TRB assembly, after the Dβ2Jβ2 joining, was experimentally demonstrated in mice (McMillan and Sikes, 2009).

The repeat sequence content of the dromedary TRBD-J-C region that is similar to the content analysed for the entire genome (Wu

et al., 2014), is lower in sheep due to the small number of SINEs (Supplementary Table S1). Moreover, the only type of SINEs present in the dromedary region, are MIR elements, which are part of an ancient family of tRNA-derived SINEs (Jurka et al., 1995) whose anomalous conservation among mammalian genomes were initially taken as evidence that they possess some unknown regulatory function (Silva et al., 2003). Recent studies have demonstrated that in human cell lines, MIRs are the only Transposon Element-derived sequences whose presence shows a positive correlation to tissue-specific gene expression (Jjingo et al., 2011) and that this correlation is likely to be due to the propensity of MIRs to operate as enhancers (Jjingo et al., 2014). The seven MIR elements in the dromedary TRBD-J-C region are located upstream and downstream of the TRBJ3 and the TRBJ2 genes, and upstream each of TRBC gene (Supplementary Figure S3). The sheep corresponding region has lost the MIR elements positioned upstream each of TRBC genes, replaced by SINE/tRNA, but retains those located upstream and downstream of the TRBJ3 and the TRBJ2 genes. The maintenance of the MIR in the vicinity of the TRBJ regions between the two species could be related with the role in gene expression regulation that is assumed for these elements.

Our expression analyses have demonstrated that all three TRBD-J-C clusters can be used to generate a functional TR β chain with the possibility to enhance the combinational and junctional diversity of the CDR3 domains for the presence of overall three TRBD and 19 TRBJ genes (Figs. 6 and 7). However, the analysis conducted on 35 unique TRBD-J-C rearrangements has shown that mechanisms for generating diversity appear to adhere to the paradigms of established thought for the study of the other mammalian species. Despite the presence of a longer nucleotide sequence in TRBD genes (especially TRBD2) (Fig. 2a in Antonacci et al., 2017) compared to humans (Lefranc et al., 2015) and dogs (Mineccia et al., 2012), the overall size of the dromedary CDR3 is similar to other species, and it is not related with the size of the TRBD gene used in the rearrangement. This result seems to be achieved by a more or less consistent deletion at the 5′ and 3′ of the TRBD genes (Fig. 7). This suggests, once again, that the length of CDR3 in TR β chain is essential for T cell function. Anyhow, the incorporation of TRBD genes in the rearrangement is translated into a high number of glycine residues in the CDR3 region. McCormack et al. (1991). have suggested that the presence of glycine residue(s) seems to be important for the structure of the antigen-binding CDR3 loop in the beta chain.

Although the number of cDNAs is too low to be statistically significant, a dissimilar usage of the TRBD and TRBJ genes can be identified and it could depend on the dromedary TRB genomic organization. Similar to sheep (Antonacci et al., 2008; Di Tommaso et al., 2010), TRBD1 accounts for the majority of total clones, followed by TRBD3 and TRBD2 (Fig. 7). This may result in greater efficiency of the PDβ1 promoter with respect to PDβ3 and PDβ2, whereas the activity of two similar PDβ3 and PDβ2 could be correlated with their position from 5′ to 3′ within the locus.

The analysis of the cDNA collection of this study also shows a prominent utilization of the genes of the TRBJ3 with respect to the TRBJ2 and TRBJ1 sets (Figs. 6 and 7). Our idea is that the preferential usage of the TRBJ3 gene set depends on the number of genes that lie in the genomic region (Fig. 3). In this regard, it is notable that the six TRBJ1 genes lie in about 1.8 Kb, the seven TRBJ3 in about 1 Kb and the six TRBJ2 in about 900 bp. Similarly, the prominently expressed gene set regards the TRBJ2 in sheep with seven TRBJ genes in 1 Kb (Di Tommaso et al., 2010). Thus, it is possible that multiple 12-RSs in a more limited area are important for increasing the local concentration of the RAG protein that mediates the recombinant process (Curry et al., 2005). Beside the elevated number of dromedary TRBD and TRBJ genes, several other mechanisms, including the incorporation of two TRBD genes in the rearrangement process, the inter-cluster recombination and the *trans*-rearrangement seem to increase the dromedary β chain functional repertoire (Fig. 7).

The two TRBC gene isotypes found in all mammalian species typically present minimal differences in the coding region and a very high divergence in the 3′ UTR. These structural constraints are so strong that similarity between the TRBC isotypes is likely to be maintained by gene conversion (Rudikoff et al., 1992). Also in dromedary, both the nucleotide and protein sequences of all three TRBC genes are very similar with nine amino acid differences between TRBC1 and TRBC2 genes, and only two between TRBC2 and TRBC3 (Fig. 2c in Antonacci et al., 2017; Table 1). The pattern of amino acid replacement between TRBC1 and TRBC2 or TRBC3 genes was located in the extracellular domain, including the A, D and G β-strands (one amino acid for each) and in DE (one amino acid) and FG (two amino acids) loops. Particularly, the dromedary FG loop is two amino acids longer with respect to humans, pigs and dogs, and there was one amino acid longer compared to sheep and cattle (Supplementary Fig. S1). The FG loop in TR Cβ is not involved in the CD3 complex binding in the inactivated state but it is able to interact with CD3γε after the conformation changes produced by antigen ligation. (Das et al., 2015). Therefore, the Cβ FG loop seems to be involved in the activation of the αβ T cells, which influences the TR-pMHC bond.

If the genomic organization of the 3′ end of the dromedary TRB is similar to the other artiodactyl species, the 5′ end appears to be different. It spans approximately 450 Kb, which makes it smaller in size compared to the sequence for cattle (700 Kb), humans (650 Kb) and rabbits (600 Kb) (Connelley et al., 2014; Antonacci et al., 2014; Lefranc et al., 2015), while it is higher compared to the dog region (300 Kb) (Mineccia et al., 2012). The dromedary has 33 TRBV genes and this is 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 dog (37 TRBV genes). Therefore, in general, a relationship between the size of the locus with the extent and complexity of the TRBV gene duplications could exist. Conversely, the dromedary germline repertoire is only slightly reduced considering the presence of only eight TRBV pseudogenes (24%) (Table 2 in Antonacci et al., 2017). Although there are differences in the number of genes, the phylogenetic clustering of the TRBV genes in the different species (Fig. 2) confirms the birth of distinct TRBV subgroups occurred prior to the radiation of mammals, while duplications or deletions within each subgroup have occurred post-speciation. TRBV21 and TRBV25 represent the most expressed subgroups (Fig. 6). Interestingly, TRBV21 was found to be well expressed in pigs as well (Eguchi-Ogawa et al., 2009), although its human counterpart TRBV21-1 has a frameshift in exon 1 and it is therefore considered to be a pseudogene (Lefranc et al., 2015). Thus, the TRBV21 gene appears to be commonly functional in artiodactyl species.

Although it cannot rule out that the published genomic sequences may be incomplete, and the results of the studies partial, a reduction of duplicated events observed in cattle and human species can be assumed for the dromedary TRB locus. Significant contraction of gene families was also observed in the dromedary genome within the TRG and TRD loci (Antonacci et al., 2011; Vaccarelli et al., 2012) with respect to the other artiodactyl species (Antonacci et al., 2005; Eguchi-Ogawa et al., 2009; Connelley et al., 2014; Piccinni et al., 2015). However, if SHM mechanisms appear to compensate for the limited repertoire of TRGV and TRDV genes, this does not seem to be applicable to TRB genes. In fact, the analysis of our cDNA collection excludes the presence of SHM in the TRB-rearranged genes. All 14 different TRBV genes belonging to 11 different subgroups, were found expressed in the cDNA clones and mostly matched the corresponding germline sequences. The slight

variations (range 97–99%) that we observed can be attributed only to allelic polymorphisms.

The lack of SHM for the beta chains, even in the presence of a limited repertoire, may depend on the different antigenic recognition mechanism provided for the two types of receptor. In fact, while the γδ TRs can bind antigens directly in the same way as immunoglobulins do (Hayday, 2000), the αβ TRs recognize processed immunogenic peptides presented to them by MHC molecules. Structural studies and the proposed two-step model of αβ TR antigen recognition are consistent with the idea of separable MHC and peptide recognition domains within the TR, with the V(D)J junction (CDR3) principally involved in peptide recognition, and CDR1 and CDR2, which are located in the TRBV germline sequence, predominantly implicated in the contact with the peptide-binding groove of MHC molecules (Wang and Reinherz, 2012). Thus, there is a constraint imposed on αβ CDR1 and CDR2 domains by the requirements for binding to MHC antigens, while γδ CDR1/2 sequences are free to diverge. It remains to determine the reasons for the presence of the wide diversified TRBD, TRBJ and TRBC repertoire and a more restricted group of TRBV genes in dromedary.

5. Conclusions

In this paper we analysed the genomic structure and the gene content of the TRB locus in *Camelus dromedary* through a combination of different techniques, including a species belonging to the Tylopoda suborder. The most noteworthy results were the presence of three in tandem TRBD-J-C clusters in the dromedary TRB locus similar to sheep, cattle and pigs, and each cluster is composed of one TRBD, six-seven TRBJ genes and one TRBC gene located to the 3′ end of a limited set of TRBV genes. This suggests that the TRB genomic structure with three TRBD-J-C clusters was established through a duplication event that occurred prior to the Tylopoda/Ruminantia/Suina divergence. Our data indicate phylogenetic significance in the evolution of Camelidae and contribute to solving the relative placement of this species within the Artiodactyla order. We hope that the knowledge of the genomic organization of the TRB locus may be useful for subsequent functional studies, that consider the extremely diversified and peculiar features of dromedary with respect to their tolerance of a variety of climatic conditions and pathogens.

Acknowledgements

The “Bilateral agreement of scientific cooperation between CNR and ASRT” for the years 2009–10 is gratefully acknowledged as well as the Italian Ministry of Foreign Affairs and Egyptian Academia of Science for supporting the “Programme of scientific and technological cooperation between Italy and Egypt”. The financial support of the University of Bari “Aldo Moro” (ex 60% delivered to RA) and University of Salento is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.dci.2017.05.021.

Conflict of interest

The authors declare no financial or commercial conflict of interest.

Authors' contributions

RA, SC and SM designed research and wrote the paper; MB contributed to the genomic study; AP provided technical assistance in cloning experiments; MM contributed to the expression study; MSH participated in the design of the study and provided dromedary tissue samples. All authors have read and approved the final manuscript.

References

Abbas, B., Agab, H., 2002. A review of camel brucellosis. Prev. Vet. Med. 55, 47–56.
Agrawal, R.P., Jain, S., Shah, S., Chopra, A., Agarwal, V., 2011. Effect of camel milk on glycemic control and insulin requirement in patients with type 1 diabetes: 2-years randomized controlled trial. Eur. J. Clin. Nutr. 65, 1048–1052.
Antonacci, R., Lanave, C., Del Faro, L., Vaccarelli, G., Ciccarese, S., Massari, S., 2005. Artiodactyl emergence is accompanied by the birth of an extensive pool of diverse germline TRDV1 genes. Immunogenetics 57, 254–266.
Antonacci, R., Vaccarelli, G., Di Meo, G.P., Piccinni, B., Miccoli, M.C., Cribiu, E.P., Perucatti, A., Iannuzzi, L., Ciccarese, S., 2007. Molecular in situ hybridization analysis of sheep and goat BAC clones identifies the transcriptional orientation of T cell receptor gamma genes on chromosome 4 in bovids. Vet. Res. Commun. 31, 977–983.
Antonacci, R., Di Tommaso, S., Lanave, C., Cribiu, E.P., Ciccarese, S., 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.
Antonacci, R., Mineccia, M., Lefranc, M.-P., Ashmaoui, H.M., Lanave, C., Piccinni, B., Pesole, G., Hassanane, M.S., Massari, S., Ciccarese, S., 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.
Antonacci, R., Giannico, F., Ciccarese, S., 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.
Antonacci, R., Bellini, M., Castelli, V., Ciccarese, S., Massari, S., 2017. Data characterizing the genomic structure of the T cell receptor (TRB) locus in Camelus dromedarius. Data in Brief: submitted for publication.
Arden, B., Clark, S.P., Kabelitz, D., Mak, T.W., 1995a. Mouse T-cell receptor variable gene segment families. Immunogenetics 42, 501–530.
Arden, B., Clark, S.P., Kabelitz, D., Mak, T.W., 1995b. Human T-cell receptor variable gene segment families. Immunogenetics 42, 455–500.
Brochet, X., Lefranc, M.-P., Giudicelli, V., 2008. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V–J and V–D–J sequence analysis. Nucleic Acids Res. 36, W503–W508.
Carvajal, I.M., Sen, R., 2000. Functional analysis of the murine TCR beta-chain gene enhancer. J. Immunol. 164, 6332–6339.
Conrad, M.L., Mawer, M.A., Lefranc, M.-P., McKinnell, L., Whitehead, J., Davis, S.K., Pettman, R., Koop, B.F., 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.
Chen, H., Kshirsagar, S., Jensen, I., Lau, K., Covarrubias, R., Schluter, S.F., Marchalonis, J.J., 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.
Chen, H., Bernstein, H., Ranganathan, P., Schluter, S.F., 2012. Somatic hypermutation of TCR γ V genes in the sandbar shark. Dev. Comp. Immunol. 37, 176–183.
Ciccarese, S., Vaccarelli, G., Lefranc, M.-P., Tasco, G., Consiglio, A., Casadio, R., Linguiti, G., Antonacci, R., 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.
Connelley, T.K., Degnan, K., Longhi, C.W., Morrison, W.I., 2014. Genomic analysis offers insights into the evolution of the bovine TRA/TRD locus. BMC Genomics 15, 994.
Curry, J.D., Geier, J.K., Schlissel, M.S., 2005. Single-strand recombination signal sequence nicks in vivo: evidence for a capture model of synapsis. Nat. Immunol. 6, 1272–1279.
Das, D.K., Feng, Y., Mallis, R.J., Li, X., Keskin, D.B., Hussey, R.E., Brady, S.K., Wang, J.H., Wagner, G., Reinherz, E.L., Lang, M.J., 2015. Force-dependent transition in the T-cell receptor β-subunit allosterically regulates peptide discrimination and pMHC bond lifetime. Proc. Natl. Acad. Sci. U. S. A. 112, 1517–1522.
Deschacht, N., De Groeve, K., Vincke, C., Raes, G., De Baetselier, P., Muyldermans, S., 2010. A novel promiscuous class of camelid single-domain antibody contributes to the antigen-binding repertoire. J. Immunol. 184, 5696–5704.
Di Tommaso, S., Antonacci, R., Ciccarese, S., 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, 3.
Doty, R.T., Xia, D., Nguyen, S.P., Hathaway, T.R., Willerford, D.M., 1999. Promoter element for transcription of unrearranged T-cell receptor beta-chain gene in pro-T cells. Blood 93, 3017–3025.
Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.
Eguchi-Ogawa, T., Toki, D., 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.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Gottschalk, L.R., Leiden, J.M., 1990. Identification and functional characterization of the human T-cell receptor beta gene transcriptional enhancer: common nuclear proteins interact with the transcriptional regulatory elements of the T-cell receptor alpha and beta genes. Mol. Cell. Biol. 14, 4286–4295.
Giudicelli, V., Brochet, X., Lefranc, M.-P., 2011. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. Cold Spring Harb. Protoc. 6, 695–715.
Giudicelli, V., Lefranc, M.-P., 2011. IMGT/JunctionAnalysis: IMGT standardized analysis of the V–J and V–D–J junctions of the rearranged immunoglobulins (IG) and T cell receptors (TR). Cold Spring Harb. Protoc. 6, 716–725.
Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E.B., Bendahman, N., Hamers, R., 1993. Naturally occurring antibodies devoid of light chains. Nature 363, 446–448.
Hayday, A.C., 2000. [gamma][delta] cells: a right time and a right place for a conserved third way of protection. Annu. Rev. Immunol. 18, 975–1026.
Jjingo, D., Huda, A., Gundapuneni, M., Mariño-Ramírez, L., Jordan, I.K., 2011. Effect of the transposable element environment of human genes on gene length and expression. Genome Biol. Evol. 3, 259–271.
Jjingo, D., Conley, A.B., Wang, J., Mariño-Ramírez, L., Lunyak, V.V., Jordan, I.K., 2014. Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. Mob. DNA 5, 14.
Jurka, J., Zietkiewicz, E., Labuda, D., 1995. Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. Nucleic Acids Res. 23, 170–175.
Lefranc, M.-P., Pommié, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L., Thouvenin-Contet, V., Lefranc, G., 2003. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev. Comp. Immunol. 27, 55–77.
Lefranc, M.-P., 2014. Immunoglobulin (IG) and T cell receptor genes (TR): IMGT® and the birth and rise of immunoinformatics. Front. Immunol. 5, 22.
Lefranc, M.-P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., Hadi-Saljoqi, S., Sasorith, S., Lefranc, G., Kossida, S., 2015. IMGT®, the international ImMunoGeneTics information system® 25 years on. Nucleic Acids Res. 43, D413–D422.
Massari, S., Lipsi, M.R., Vonghia, G., Antonacci, R., Ciccarese, S., 1998. T-cell receptor TCRG1 and TCRG2 clusters map separately in two different regions of sheep chromosome 4. Chromosome Res. 6, 419–420.
McCormack, W.T., Tjoelker, L.W., Stella, G., Postema, C.E., Thompson, C.B., 1991. Chicken T-cell receptor beta-chain diversity: an evolutionarily conserved D beta-encoded glycine turn within the hypervariable CDR3 domain. Proc. Natl. Acad. Sci. U. S. A. 88, 7699–7703.
McMillan, R.E., Sikes, M.L., 2009. Promoter activity 5_ of D_2 is coordinated by E47, Runx1, and GATA-3. Mol. Immunol. 46, 3009–3017.
Miccoli, M.C., Antonacci, R., Vaccarelli, G., Lanave, C., Massari, S., Cribiu, E.P., Ciccarese, S., 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.
Mineccia, M., Massari, S., Linguiti, G., Ceci, L., Ciccarese, S., 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.
Muyldermans, S., Baral, T.N., Retamozzo, V.C., De Baetselier, P., De Genst, E., Kinne, J., Leonhardt, H., Magez, S., Nguyen, V.K., Revets, H., Rothbauer, U., Stijlemans, B., Tillib, S., Wernery, U., Wyns, L., Hassanzadeh-Ghassabeh, G., Saerens, D., 2009. Camelid immunoglobulins and nanobody technology. Vet. Immunol. Immunopathol. 128, 178–183.
Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.
Nguyen, V.K., Hamers, R., Wyns, L., 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.
Nguyen, V.K., Desmyter, A., Muyldermans, S., 2001. Functional heavychain antibodies in Camelidae. Adv. Immunol. 79, 261–296.
Njiru, Z.K., Constantine, C.C., Ndung'u, J.M., Robertson, I., Okaye, S., Thompson, R.C., Reid, S.A., 2004. Detection of Trypanosoma evansi in camels using PCR and CATT/T. evansi tests in Kenya. Vet. Parasitol. 124, 187–199.
Piccinni, B., Massari, S., Caputi Jambrenghi, A., Giannico, F., Lefranc, M.-P., Ciccarese, S., Antonacci, R., 2015. Sheep (Ovis aries) T cell receptor alpha (TRA) and delta (TRD) genes and genomic organization of the TRA/TRD locus. BMC Genomics 16, 709.
Rudikoff, S., Fitch, W.M., Heller, M., 1992. Exon-specific gene correction (conversion) during short evolutionary periods: homogenization in a two-gene family encoding the beta-chain constant region of the T-lymphocyte antigen receptor. Mol. Biol. Evol. 9, 14–26.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., Miller, W., 2000. PipMaker – a web server for aligning two genomic DNA sequences. Genome Res. 10, 577–586.
Silva, J.C., Shabalina, S.A., Harris, D.G., Spouge, J.L., Kondrashovi, A.S., 2003. Conserved fragments of transposable elements in intergenic regions: evidence for widespread recruitment of MIR- and L2-derived sequences within the mouse and human genomes. Genet. Res. 82, 1–18.
Takeda, J., Cheng, A., Mauxion, F., Nelson, C.A., Newberry, R.D., Sha, W.C., Sen, R., Loh, D.Y., 1990. Functional analysis of the murine T-cell receptor beta enhancer and characteristics of its DNA-binding proteins. Mol. Cell. Biol. 10, 5027–5035.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Vaccarelli, G., Miccoli, M.C., Antonacci, R., Pesole, G., Ciccarese, S., 2008. Genomic organization and recombinational unit duplication-driven evolution of ovine and bovine T cell receptor gamma loci. BMC Genomics 9, 81.
Vaccarelli, G., Antonacci, R., Tasco, G., Yang, F., Giordano, L., El Ashmaoui, H.M., Hassanane, M.S., Massari, S., Casadio, R., Ciccarese, S., 2012. Generation of diversity by somatic mutation in the Camelus dromedarius T-cell receptor gamma variable domains. Eur. J. Immunol. 42, 3416–3428.
Vincke, C., Loris, R., Saerens, D., Martinez-Rodriguez, S., Muyldermans, S., Conrath, K., 2009. General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. J. Biol. Chem. 284, 3273–3284.
Wang, J.H., Reinherz, E.L., 2012. The structural basis of αβ T-lineage immune recognition: TCR docking topologies, mechanotransduction, and co-receptor function. Immunol. Rev. 250, 102–119.
Whitehurst, C.E., Schlissel, M.S., Chen, J., 2000. Deletion of germline promoter PDb 1 from the TCR b locus causes hypermethylation that impairs Db1 recombination by multiple mechanisms. Immunity 13, 703–714.
Wu, H., Guang, X., Al-Fageeh, M.B., Cao, J., Pan, S., Zhou, H., Zhang, L., Abutarboush, M.H., Xing, Y., Xie, Z., Alshanqeeti, A.S., Zhang, Y., Yao, Q., Al-Shomrani, B.M., Zhang, D., Li, J., Manee, M.M., Yang, Z., Yang, L., Liu, Y., Zhang, J., Altammami, M.A., Wang, S., Yu, L., Zhang, W., Liu, S., Ba, L., Liu, C., Yang, X., Meng, F., Wang, S., Li, L., Li, E., Li, X., Wu, K., Zhang, S., Wang, J., Yin, Y., Yang, H., Al-Swailem, A.M., Wang, J., 2014. Camelid genomes reveal evolution and adaptation to desert environments. Nat. Commun. 2015 (6), 6107, 5, 5188. Erratum in: Nat. Commun.
Yang, X.O., Doty, R.T., Hicks, J.S., Willerford, D.M., 2003. Regulation of T-cell receptor D beta 1 promoter by KLF5 through reiterated GC-rich motifs. Blood 101, 4492–4499.
Yazawa, R., Cooper, G.A., Beetz-Sargent, M., Robb, A., McKinnel, L., Davidson, W.S., Koop, B.F., 2008. Functional adaptive diversity of the Atlantic salmon T-cell receptor gamma locus. Mol. Immunol. 45, 2150–2157.
Yousfi Monod, M., Giudicelli, V., Chaume, D., Lefranc, M.-P., 2004. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V–J and V–D–J JUNCTIONs. Bioinformatics 20 (Suppl. 1), i379–i385.