Developmental and Comparative Immunology 106 (2020) 103614

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

Short communication

Comprehensive genomic analysis of the dromedary T cell receptor gamma (TRG) locus and identification of a functional TRGC5 cassette

R. Antonacci[a,*], G. Linguiti[a], P.A. Burger[b], V. Castelli[a], A. Pala[a], R. Fitak[b,c], S. Massari[d], S. Ciccarese[a]

[a] Department of Biology, University of Bari "Aldo Moro", Bari, Italy
[b] Research Institute of Wildlife Ecology, Vetmeduni Vienna, Vienna, Austria
[c] Department of Biology, Genomics and Bioinformatics Cluster, University of Central Florida, Orlando, FL, 32816, USA
[d] Department of Biological and Environmental Science and Technologies, University of Salento, Lecce, Italy

ARTICLE INFO

Keywords:
T cell receptor
γδ T cell
TRG locus
Camelus dromedarius
Somatic hypermutation

ABSTRACT

The emergent availability in public databases of more complete genome assemblies allows us to improve genomic data obtained by classical molecular cloning. The main goal of this study was to refine the genomic map of the dromedary TRG locus by integrating our previous genomic data with the analysis of recent genomic assemblies. We identified an additional TRGC cassette, defined as a V-J-C recombination unit, located at the 5' of the locus and made up of five TRGV genes followed by three TRGJ genes and one TRGC gene. Hence, the complete dromedary TRG locus spans about 105 Kb and consists of three in tandem TRGC cassettes delimited by AMPH and STARD3NL genes at the 5′ and 3′ end, respectively. An expression assay carried out on peripheral blood showed the functional competency for the dromedary TRGC5 cassette and confirmed the presence of the somatic hypermutation mechanism able to enlarge the repertoire diversity of the dromedary γδ T cells.

1. Introduction

The locus encoding for the T cell receptor (TR) γ chain (TRG) is made up of separate Variable (V) and Joining (J) genes that rearrange together and join to the Constant (C) gene to express a functional γ chain on the γδ T cells. The genomic arrangement of these genes at the TRG locus is considerably different across the so far examined species, including carnivores (*Canis lupus familiaris*), lagomorphs (*Oryctolagus cuniculus*) and cetaceans (*Tursiops truncatus*) (Massari et al., 2009, 2012; Linguiti et al., 2016). Unique is their organization in Bovidae. In *Ovis aries*, two TRG loci, TRG1 and TRG2, exist that map in two different positions of the same chromosome (Massari et al., 1998). In each locus, the TRG genes are arranged in reiterated cassettes, each containing the basic V-J-J-C recombination unit (Miccoli et al., 2003; Vaccarelli et al., 2008). The TRG1 locus maps on 4q3.1 (Antonacci et al., 2007), syntenic to the human TRG locus at 7p14, and consists of three cassettes in sheep, TRGC5, TRGC3 and TRGC4; while in *Bos taurus*, an extra cassette, TRGC7, is present in the TRG1 locus (Conrad et al., 2007). The TRG2 locus at 4q2.2 position contains the TRGC1, TRGC2 and TRGC6 cassettes both in sheep and bovine genome. Variations in the number of the TRGV genes can be observed among the cassettes while the enhancer-like sequences are present downstream all the TRGC genes to define each recombination unit. Phylogenetic approaches demonstrated that the ruminant TRGC5 cassette resembles the ancestral unit that gave rise to the reiterated duplications (Vaccarelli et al., 2008).

More recently, a genomic study, by using both PCR and Genome Walker DNA strategies, revealed that also the dromedary TRG genes are arranged in tandem cassettes localized at a single chromosome region (7q11-12) (Vaccarelli et al., 2012). Two in tandem cassettes, distributed over 45 kb, were characterized, each consisting of one TRGV, two TRGJ and one TRGC genes. The locus organization in two cassettes potentially limits the combinatorial usage of its genes. However, cDNA sequencing clearly revealed, for the first time in a mammalian organism, that besides the combinatorial diversity, a further mechanism of somatic hypermutation (SHM) enhances the TR δ (Antonacci et al., 2011) and γ (Vaccarelli et al., 2012) chain repertoire in *Camelus dromedarius* in a manner similar to the process of affinity maturation of immunoglobulin genes (Ciccarese et al., 2014, 2019). Thus, the finding of SHM in dromedary genes differentiates this camelid species from the other artiodactyls, where no indication of SHM was found.

However, since the genomic approach used to understand the organization of the dromedary TRG locus was not able to exclude the presence of other TRG genes, we have analysed the recent genomic assemblies to produce a comprehensive genomic map of the TRG locus

* Corresponding author. Department of Biology, University of Bari "Aldo Moro", 70124, Bari, Italy.
E-mail address: rachele.antonacci@uniba.it (R. Antonacci).

https://doi.org/10.1016/j.dci.2020.103614
Received 29 November 2019; Received in revised form 10 January 2020; Accepted 10 January 2020
Available online 18 January 2020
0145-305X/

in this species.

2. Materials and methods

2.1. Genome analysis

With the aim of completing the genomic region comprising the TRG locus in dromedary, we searched the Cdrom64K assembly available at GenBank (ID: GCA_000803125.1, Fitak et al., 2016). The human AMPH gene was used as a probe. We recovered a sequence of about 132 kb (gaps included) from the 8666367_scaffold containing the (dromedary) AMPH gene, the TRGV, TRGJ genes and only the first exon of the C gene belonging to the TRGC5 gene cassette. A sequence of approximately 32 kb (gaps included) containing the entire TRGC5 gene was then isolated from the scaffold ScvgHr6_24370. The latter came from a further dromedary assembly (Elbers et al., 2019). This 32 Kb sequence was combined with that of the 8666367_scaffold to constitute a new region of about 162 Kb (Supplementary Fig. 1). Again, this new region has been further combined with the previous genomic sequence of around 46 Kb obtained by Vaccarelli et al., (2012). Then, the reconstructed dromedary sequence containing the AMPH gene at its 5′ and the last exon of the TRGC2 gene at its 3', of about 208 kb (gaps included) is presented in Supplementary Fig. 1.

To complete the locus and in order to perform genomic comparison with human and sheep TRG loci, a sequence of 25666 pb, containing the STARD3NL gene, was retrieved from the *Camelus ferus* genomic scaffold NW_006210979 (positions 3960730–3986395) and joined to the 3' of the 208 Kb long sequence above described.

Computational analysis of dromedary TRG locus was conducted with the following programs: RepeatMasker for the identification of genome-wide repeats and low complexity regions (Smit, A. F. A., Hubley R., Green P., RepeatMasker at http://repeatmasker.org), Pipmaker (Schwartz et al., 2000; http://pipmaker.bx.psu.edu/pipmaker/) for the alignment of the dromedary sequence with the human (GenBank accession number NT_007819: positions 37690000–38240000) and sheep (GenBank accession numbers DQ992075, TRG1, and DQ992074, TRG2) counterparts. RepeatMasker screens DNA sequences for interspersed repeats and low complexity DNA sequences.

2.2. Classification of variable (V), joining (J) and constant (C) genes belonging to the TRG5 cassette

Each TRG gene was classified according to the similarity with the sheep species and based on the genomic position within the locus, and the nomenclature established on the basis of IMGT at http://www.imgt.org/IMGTScientificChart/SequenceDescription/IMGTfunctionality.html (Lefranc, 2014).

The functionality of V, J and C genes was predicted through the manual alignment of sequences adopting the following parameters: (a) identification of the leader sequence at the 5′ of the TRGV genes; (b) determination of proper recombination signal (RS) sequences located at 3′ of the TRGV and 5' of the TRGJ genes, respectively; (c) determination of correct acceptor and donor splicing sites; (d) estimation of the expected length of the coding regions; (e) absence of frameshifts and stop signals in the coding regions of the genes.

2.3. 5' rapid amplification of cDNA ends (RACE) PCR

Total RNA was extracted from blood of an adult animal using the Trizol method according to the manufacturer's protocol (Invitrogen, Carlsbad, CA). Approximately 5 μg of RNA was reverse transcribed with Superscript II (Invitrogen, Carlsbad, CA) by using an oligo-dT adapter primer.

After linking a poly-C tail at the 5'end of the ss cDNA, the ds cDNA was performed with Platinum Taq polymerase (Invitrogen) by using C5drom1R (5′ TCCAGAACTGTATTGCTGTCC 3′), designed on the sequence of the first exon of the TRBC5 gene, 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 C5drom1RN/AUAP primer pair. C5drom1RN (5′ GCCATGTCTGCATC AAGCTTC 3′) is designed on a sequence upstream of C5drom1R, 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, 40 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 (Stratagene). 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 LR735241 to LR735257.

2.4. Phylogenetic analyses

The human and sheep TRGV, TRGJ and TRGC genes used for the phylogenetic analysis were retrieved from IMGT ®(IMGT Repertoire, http://www.imgt.org, IMGT/GENE-DB, Giudicelli et al., 2005).

Multiple alignments of the gene sequences under analysis were carried out with the MUSCLE program (Edgar, 2004). The evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). 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).

3. Results and discussion

3.1. Update of the genomic organization of the TRG locus

In a previous work, a genomic region of about 45 kb was identified using the DNA walking strategies (Vaccarelli et al., 2012). It contained two V-J-J-C cassettes, each consisting of one TRGV, two TRGJ and one TRGC gene and lacked the AMPH and STARD3NL genes, found in all mammalian species so far studied flanking the TRG locus at 5′ and 3′, respectively. Therefore, this result did not exclude the existence of one or more V-J-C cassettes located upstream of the first and perhaps also downstream of the second cassette. The availability of genomic assemblies allowed us to explore the possible presence of other genes and to complete and define the organization of the TRG locus in dromedary. For this purpose, the sequences of human AMPH and STARD3NL genes were used as probe to analyse the dromedary Cdrom64K genomic assembly (Fitak et al., 2016). A sequence of about 132 kb was recovered (Fig. 1A). It contained the entire AMPH gene, 5 TRGV and 3 TRGJ new genes plus the first exon of one TRGC gene never identified before, indicating the presence of a third gene cassette. A genomic region containing the last three exons of the TRGC gene and following the intronic region up to overlapping with the sequence already published (Vaccarelli et al., 2012) was then recovered in the scaffold ScvgHr6_24370 (Elbers et al., 2019) (Fig. 1A). Instead, no genomic scaffold overlapping at the 3′ end of the locus and containing the STARD3NL gene was recovered from the available dromedary assemblies. To get information on the part containing the missing 3' end, considering also the high sequence identity previously detected between *Camelus dromedarius* and *Camelus ferus* in the TRB locus (Antonacci et al., 2019), we analysed the *Camelus ferus* genome

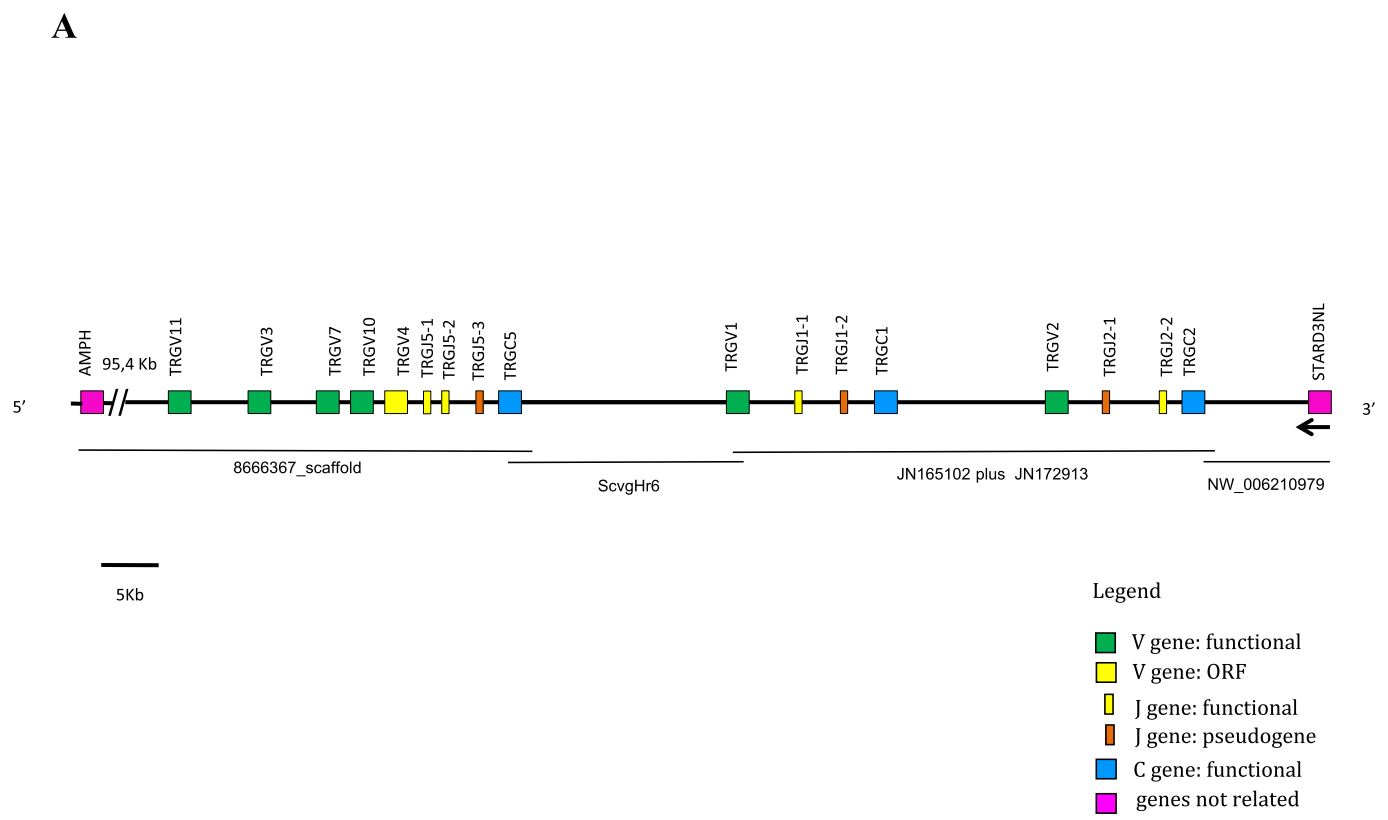

Fig. 1. (A) Schematic representation of the genomic organization of the dromedary TRG locus as deduced from the genome sequences. The genomic NW_006210979 scaffold sequence has been retrieved from the *Camelus ferus* assembly (Jirimutu et al., 2012). The diagram shows the position of all the related and unrelated TRG genes according to nomenclature. The boxes representing the genes are not to scale. The exons are not shown. **(B)** Schematic representation of the genomic organization of the TRG locus in human (HSA), dromedary (CDR) and sheep (OAR). Red, bleu and green circles show the TRGV genes that form respectively the A, B and C groups in the phylogenetic tree. Rectangular boxes indicate the cluster formed by three TRGJ genes. The enhancer-like elements are also inserted (pale blue circles).

assembly CB1 and a single contig (GenBank ID: NW_006210979) (Fig. 1A) representative of the entire TRG locus was recovered, with both AMPH flanking the 5 ′and STARD3NL flanking the 3′ genes, respectively. The deduced sequence data indicate that the distance between the STARD3NL gene and the last exon of the TRGC2 gene is 8725 bp and in this portion no genes are present.

Ultimately, the identification of the reference (AMPH and STARD3NL) genes allowed us to establish that the TRG locus is now complete in all its parts in camel species. Starting from the first TRGV gene at its 5′ end, and ending to TRGC2 gene at its 3' end, it spans about 105 kb and similarly to each of the two TRG loci in sheep, it is organized in three V-J-C cassettes (Vaccarelli et al., 2008). The additional cassette was classified as TRGC5 for the homology to the corresponding sheep one (see below). The updated map of the region is shown schematically in Fig. 1A.

3.2. Classification of the variable genes belonging to the TRGC5 cassette

All TRGV genes where identified using human and sheep sequences available in public databases and their functionality was defined based on the IMGT rules (IMGT®,http://www.imgt.org) (Supplementary Table 1). We annotated five new TRGV germline genes, which were assigned to distinct five subgroups adopting the criterion that sequences with nucleotide identity of more than 75% in the V-region belong to the same subgroup (Arden et al., 1995a,b). Hence, a total of seven TRGV subgroups are present overall within the dromedary TRG locus and each consists of only one member germline gene (Fig. 1).

To classify the new dromedary TRGV gene subgroups, the evolutionary relationship of these genes was investigated by comparing all the dromedary genes with the available human and sheep corresponding ones. Thus, the V-REGION nucleotide sequences of all TRGV genes were combined in the same alignment and an unrooted phylogenetic tree was made using the NJ method (Saitou and Nei, 1987) (Supplementary Fig. 2). The tree shows that each of the dromedary gene subgroups form a monophyletic group with a corresponding sheep and, if present, with a human gene, consistent with the occurrence of distinct subgroups prior to the divergence of these different mammalian species. Therefore, according to phylogenetic clustering, we classified the new dromedary TRGV subgroups as orthologous to their corresponding sheep subgroups. In particular, on the basis of the dromedary TRGV genes distribution, the tree can be subdivided in three groups labelled A to C (named independently from the human TRGVA and TRGVB genes).

The number of groups corresponds to the number of the dromedary cassettes and each group contains dromedary TRGV genes belonging to its own TRGC cassette. In group A, TRGV1 gene of the second cassette (in its genomic position) is closely related to the sheep TRGV genes belonging to TRGC1 (TRGV5-1), TRGC2 (TRGV5-2), and TRGC4 (TRGV1) cassettes. In addition, the TRGV2-1, TRGV8-1 and TRGV9-1 genes belonging to the sheep cassette TRGC3 are grouped in A together with the nine (from TRGV1 to TRGV8) genes of the human locus that are classified as subgroup 1 (IMGT Repertoire, http://www.imgt.org; red circles in Fig. 1B). It would seem that the common ancestral TRGV gene has remained unique in the dromedary (TRGV1) while it is duplicated in humans and sheep to give rise to new functions related to the variable domains in these two species. Conversely, in B, the dromedary TRGV2 is grouped to the sheep TRGV6-1 gene belonging to the TRGC6 cassette and to the human pseudogene TRGVA located

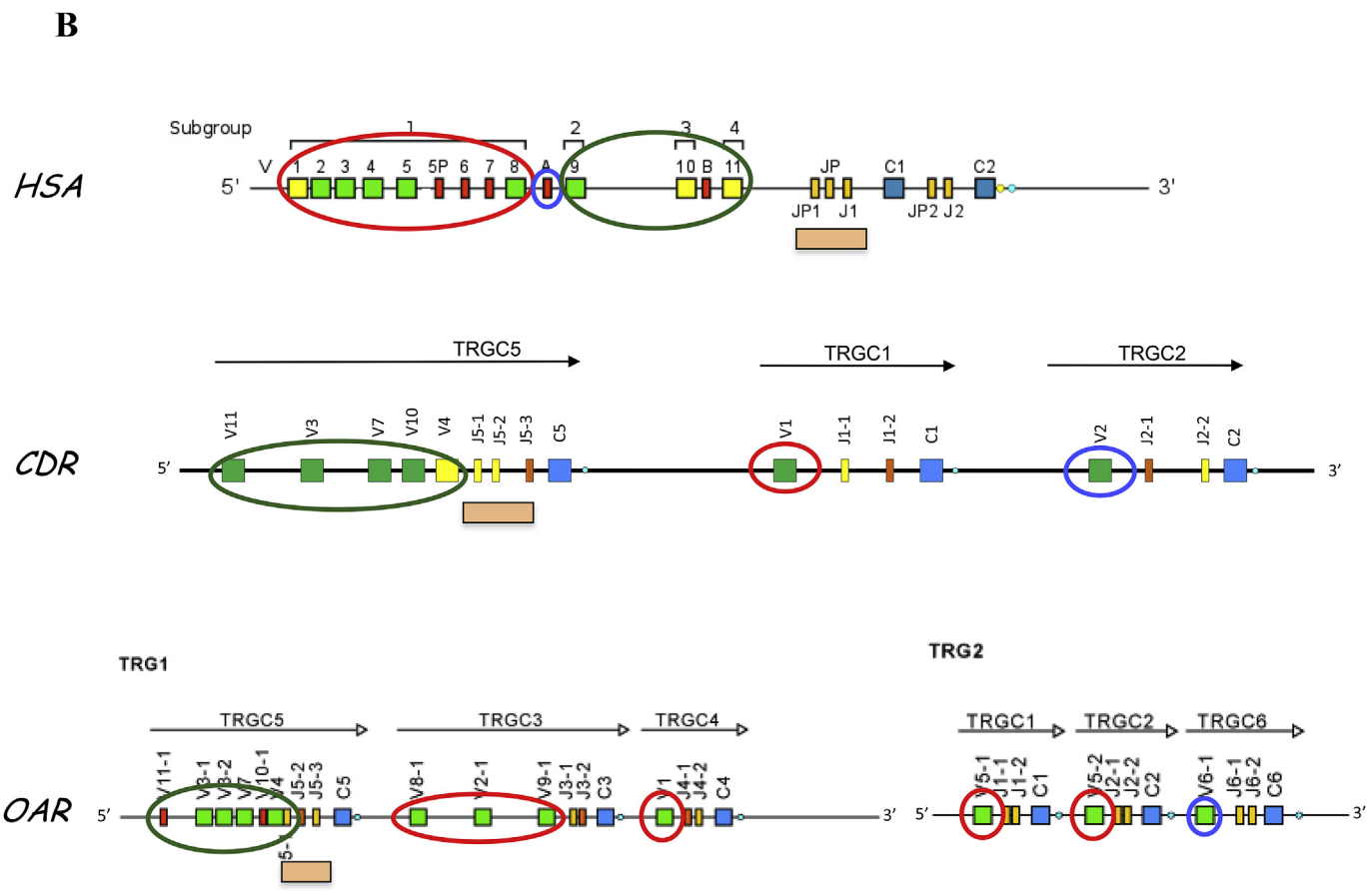

Fig. 1. (*continued*)

immediately after the genes of the subgroup 1 (blue circles in Fig. 1B). In C, it is possible to notice a close phylogenetic relationship between the five dromedary TRGV genes of the TRGC5 cassette located at 5′ of the locus and the six TRGV genes belonging to the sheep TRGC5 cassette, which we have previously defined as the "ancient cassette" (Vaccarelli et al., 2008). The four human TRGV (TRGV9, TRGV10, TRGVB and TRGV11) genes, located closer to the TRGJ gene clusters, are interspersed in C. It is interesting to note that the corresponding orthologous genes grouped in C have maintained their genomic order within the TRG locus of each of the three species throughout evolution (green circles in Fig. 1B).

The deduced amino acid sequences of the dromedary new germline TRGV genes were manually aligned according to IMGT unique numbering for the V-REGION (Lefranc et al., 2003) to maximize the percentage of identity (Fig. 2A, upper part). All sequences exhibit the typical framework regions (FR) and complementarity determining regions (CDR) as well as four amino acids: cysteine 23 (1st-CYS) in FR1-IMGT (except for the TRGV4), tryptophan 41 (CONSERVED-TRP) in FR2-IMGT (except for the TRGV4), hydrophobic amino acid 89, and cysteine 104 (2nd-CYS) in FR3-IMGT (Lefranc et al., 2003). Conversely, CDR-IMGT varies in amino acid composition and length.

All the dromedary new germline TRGV genes have been classified as potential functional genes with the exception of TRGV4, which has been defined as ORF due to the absence of canonical AA in FR1-IMGT and FR2-IMGT.

3.3. Description and classification of the TRGJ and TRGC genes belonging to the TRGC5 cassette

The new TRGJ genes have been named according to the criteria for membership to the cassette and numbered for their genomics position. Of the TRGJ5 genes, only TRGJ5-3 gene is a pseudogene since it has an abnormal heptamer and its coding region has stop codon. Also the TRGJ1-2 is a pseudogene, whereas the TRGJ2-1 gene is classified as ORF because of no canonical J-HEPTAMER. Hence, of the 7 TRGJ genes present within the dromedary TRG locus, 4 are functional genes (Fig. 2A, middle part; Supplementary Table 1).

The phylogenetic tree obtained by aligning dromedary TRGJ genes with the corresponding human and ovine TRGJ genes has highlighted three groups named in Supplementary Fig. 3 as A, B and C. Each of the three groupings contains TRGJ genes that occupy the same relative position in the J-(J)-J-C genomic cassette of the three TRG loci (Fig. 1B). Particularly, the eleven TRGJs, whether they are functional or pseudogenes, that occupy a distal position with respect to the TRGC gene, are grouped in A; the TRGJs that are in the middle of the cluster formed by three genes (rectangular boxes in Fig. 1B), are grouped in B; finally, the eleven TRGJ genes located in proximity to the TRGC gene of each cassette, are grouped in C. These data highlight an evident functional constraint due to the role that TRGJ genes play in the recombination process, already described in a previous work (Miccoli et al., 2005), where the similarity between orthologous genes of different species is highlighted, rather than that between paralogous genes within the same species.

The TRGC gene of the new identified cassette has been named TRGC5 since the exon organization is similar to the sheep TRGC5. As the ovine gene, it consists of three exons, with only one EX2 encoding for the connecting region (Fig. 2A, lower part); whereas, the TRGC1 and TRGC2, encoded by five exons, consist of three EX2 as observed in the sheep TRGC2, TRGC4 and TRGC6 genes (Vaccarelli et al., 2012). To investigate the evolutionary relationship of the dromedary TRGC5 gene, we compared its coding nucleotide sequence with those of the dromedary TRGC1 and TRGC2 as well as with the human and sheep

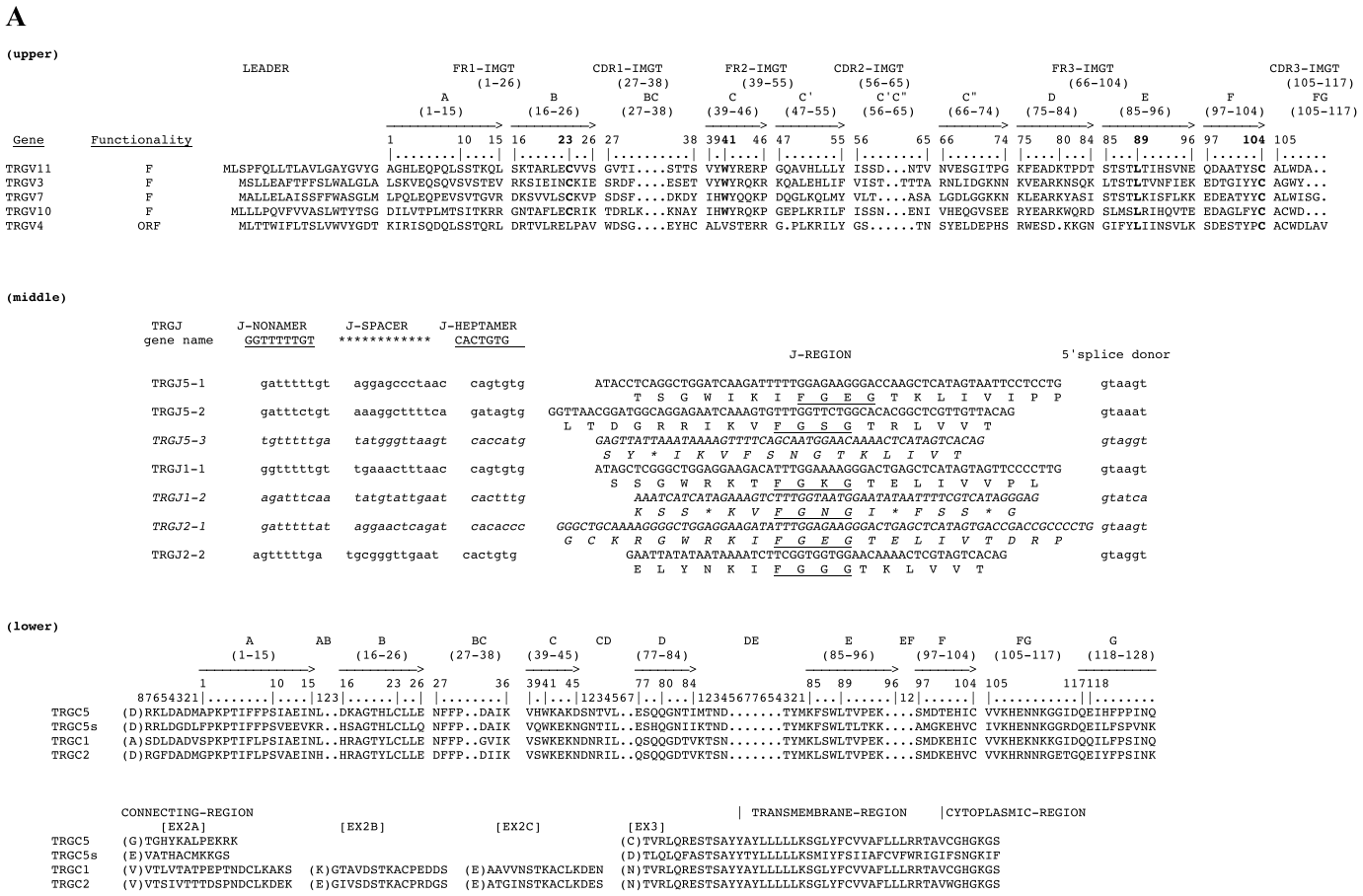

Fig. 2. (A) (Upper) The IMGT Protein display of the dromedary TRGV genes belonging to the TRGC5 cassette. The description of the strands and loops and of the FR-IMGT and CDR-IMGT is according to the IMGT unique numbering for V-REGION (Lefranc et al., 2003). The five conserved amino acids of the V-DOMAIN (1st-CYS 23, CONSERVED-TRP 41, hydrophobic AA 89, 2nd-CYS 104 and J-PHE 118) are indicated in bold. (Middle) Nucleotide and deduced amino acid sequences of the dromedary TRGJ genes. The consensus sequence of the heptamer and nonamer is provided at the top of the figure and is underlined. The numbering adopted for the gene classification is reported on the left of each gene. The donor splice site for each TRGJ is shown. The canonical FGXG amino acid motifs are underlined. No functional TRGJ genes are indicated in italics. (Lower) IMGT Protein display of the dromedary TRGC genes together with the sheep TRGC5 (TRGC5s). The descriptions of the strands and loops were collected according to the IMGT unique numbering for the C-DOMAIN (Lefranc et al., 2005). **(B)** Protein display of the in frame and out of frame cDNA clones. The TRGV and the TRGJ genes are listed, respectively, at the left and the right of the figure. Leader region (L-REGION), CDR-IMGT and FR-IMGT are also indicated according to the IMGT unique numbering for V-DOMAIN. 1st-CYS 23, CONSERVED-TRP 41 and 2nd-CYS 104 are highlighted. The name of the clones is also reported. The amino acid changes are boxed.

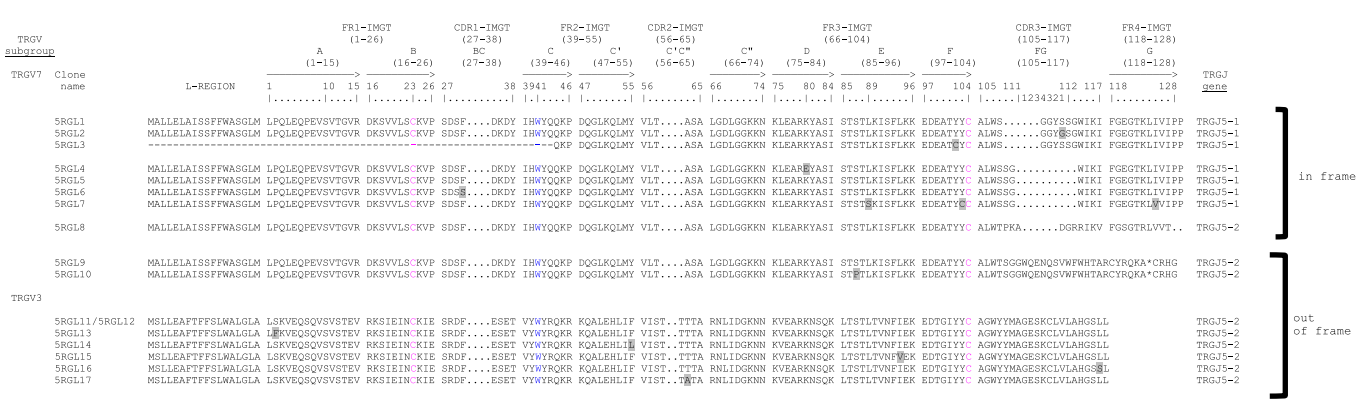

Fig. 2. (*continued*)

TRGC genes. A phylogenetic tree was constructed using NJ method (Saitou and Nei, 1987) (Supplementary Fig. 4).

In the phylogenetic tree, the dromedary TRGC5 gene is grouped together with its paralogues TRGC1 and TRGC2 genes as well as in humans the TRGC1 groups with TRGC2 (A), in both cases in a species-specific manner, indicating that the TRGC duplications occurred after the divergence of humans and artiodactyls. This is also true for all TRG genes in *Ovis aries* (in B) with the exception of the TRGC5 gene that does not group together with the corresponding paralogues but is more related to the TRGC genes of the other two species (in A), confirming its ancestral origin.

3.4. 5' RACE assay

In a previous work, Vaccarelli et al. (2012) performed 5' RACE using degenerate primers based on a conserved C domain amino acid sequence to isolate putative dromedary TRG chain cDNA clones in spleen. Only clones containing TRGC1 and TRGC2 genes were selected and identified. We ascribe this result to a lower or absent level of expression of the new cassette (in the present study), when compared to the other two cassettes.

Hence, to evaluate the functional competency of the TRGC5 cassette, a 5′RACE assay was performed on total RNA derived from peripheral blood using TRGC5 gene specific primers on dT-primed single-stranded cDNA. The RACE product was cloned into TA-vector and 33 randomly selected positive clones were sequenced. A total of 17 different sequences were obtained containing rearranged V-J-C transcripts. The remaining 16 were redundant sequences and therefore not considered for further analyses.

The sequence analysis revealed that only eight clones are productive with a correct open reading frame, and all containing the TRGV7 gene (Fig. 2B). Particularly, seven clones are the result of a TRGV7/TRGJ5-1 recombination and one clone derives from a TRGV7/TRGJ5-2 rearrangement. Moreover, the TRGV7/TRGJ5-1 clones can be divided in two groups on the basis of the amino acid sequence identity; this implies that the 5RGL1, 5RGL2 and 5RGL3 as well as the 5RGL4, 5RGL5, 5RGL6 and 5RGL7 cDNAs could derive from a same V-J rearrangement. When the variable coding region (L-PART1 plus V-EXON till the amino acid at position 104) of all the productive cDNAs that contain the TRGV7 gene have been aligned for a nucleotide sequence comparison, we found that only the 5RGL5 and 5RGL8 clones, which do not share CDR3, have a 100% nucleotide identity. For the remaining clones, the identity of the TRGV gene sequence varies from 98.32 to 99.71% due to 9 nucleotide substitutions, of which five variations changed the corresponding amino acid in CDR1 (5RGL6) and in FR3 (5RGL3, 5RGL4 and 5RGL7) (Fig. 2B).

9 out of 17 rearranged V-J-C transcripts are out of frame probably caused by frameshift mutations during the recombination between TRGV7/TRGJ5-2 (2 clones) or TRGV3/TRGJ5-2 (7 clones) genes (Fig. 2B). The sequences are attributable to only two types of cDNA and amino acid variations can be observed also among these clones excepted for 5RGL11 and 5RGL12 clones that differ for only one nucleotide.

We analysed in detail, as previously described (Ciccarese et al., 2014), the nucleotide variations within all the cDNA sequences (from FR1 to FR4 with the exclusion of the L-region and the CDR3) with respect to the corresponding germline genes (TRGV and TRGJ). We found an overall mutation frequency of 0.32% (11 mutations in 3431 bp) for the TRGV7 clones and 0.39% (9 mutations in 2268 bp) for the TRGV3 cDNAs, which is about twice as much as observed for a not relevant gene under the same experimental conditions (Vaccarelli et al., 2012). All mutations are transitions and 12 out of 20 variations changed the corresponding amino acid in CDRs as well as in FRs (Fig. 2B and Supplementary Table 2). Furthermore, the mutations are detected in the AID hot spot motifs context (Supplementary Table 2). In particular, five mutations exactly match the classic AID hot spot motifs (dgyw/wrch and tw), eight mutations have a single mismatch in the consensus, while the remaining mutations differ from the consensus by two or more mismatches (Supplementary Table 2). Overall, the characteristics of the mutation profile of the TRGC5 cassette cDNAs remind the SHM of the γ and δ genes that, as already highlighted (Antonacci et al., 2011; Vaccarelli et al., 2012; Ciccarese et al., 2014), contributes to increase the repertoire diversity of the dromedary γδ T cells.

Finally, the reduced number of different cDNAs, confined to recombination with only the TRGV7 gene, in association with the redundancy, would lead to think of a reduced expression of the TRGC5 cassette in peripheral blood of adult individuals. The feature of the near absence of expression of the variable genes rearranged with TRGC5 gene in blood is consistent with the data previously found in the spleen of adult dromedary, where 5′RACE assays have only allowed the selection of cDNA clones derived from productive rearrangements of variable genes with the TRGC1 and TRGC2 cassettes (Vaccarelli et al., 2012). Furthermore, a comparison of gene expression in the peripheral blood of γδ T cells at different stages of development revealed in sheep (Hein and Dudler, 1993) distinct patterns of TRG gene repertoire, confining the expression of the TRGC5 cassette to foetal ontogeny and in neonatal animals.

In this regard, recently, using the Next Generation Sequencing (NGS) of cDNA libraries, Damani-Yokota and coworkers (Damani-Yokota et al., 2018) have demonstrated that the TRGC5 cassette is the most transcribed one in calves peripheral blood at birth and it remains the most expressed at one month after birth, even if at a lower level of expression. It is even to be noted that although there are four functional TRGV genes in the bovine TRGC5 cassette, the TRGV3 and TRGV7 are the only expressed genes (Damani-Yokota et al., 2018).

Therefore, we hypothesize that, as in sheep and cow, also in dromedary the expression of the genes of TRGC5 cassette might occur during foetal and/or neonatal developmental stage and might be reduced then in the adult stage.

3.5. Comparison of the genomic structure of the dromedary with human and sheep TRG loci

We analysed the genomic structure of the dromedary TRG locus aligning the entire sequence (from the first TRGV gene to the STARD3NL gene located to the 3′ end) with the human TRG and sheep TRG1 and TRG2 counterparts, using the PipMaker program, and we evaluated the alignment expressed as a dot-plot sequence comparison graph. The resulting human/dromedary dot plot matrix shows two parallel lines of similarity in correspondence with each of the three camel J-C regions (yellow boxes in Supplementary Fig. 5). Furthermore, according to the phylogenetic results (see A in Supplementary Fig. 2), nine dashes represent the homology between the genomic region containing the nine TRGV genes belonging to subgroup 1 located at 5' of the human locus and that with the only dromedary TRGV1 gene (blue box in Supplementary Fig. 5). Similarly, again according to the phylogenetic tree, lines in the orange box correspond to the similarity between all the dromedary TRGV genes belonging to the cassette TRGC5, and the human from TRGV9 to TRGV11 genes that are close to the JP1-JP-J1-C1 cluster. Finally, the genomic region containing the human TRGVA gene is matched by that of the dromedary TRGV2 gene (light blue box in Supplementary Fig. 5).

Small lines of homology are located downstream to each dromedary TRGC gene, corresponding to the single human enhancer sequence (green circles in Supplementary Fig. 5). The presence of the enhancer-like sequences at 3 ′of each TRGC gene in the dromedary is congruent with the data referring to the genomic organization of *Ovis aries* and *Canis lupus familiaris* TRG loci, which result from reiterated duplications of the V-J-J-C cassette (Vaccarelli et al., 2008; Massari et al., 2009). Moreover, similarity spots with the STARD3NL gene are evident at the 3 ′end of the dromedary TRGC5 and TRGC2 cassettes (red square in Supplementary Fig. 5). The STRARD3NL gene duplications suggest an

evolutionary model of the dromedary TRG locus according to which a minimal ancestral cassette consisting of one TRGV, two TRGJ and one TRGC genes underwent duplication event including an enhancer-like region and part of the STRARD3NL gene. This event would have generated the TRGC5 and TRGC2 cassettes. A similar evolutionary model was proposed to explain the origin of the TRG locus in dog where the duplications also involved enhancer-like elements and part of the STRARD3NL gene (Massari et al., 2009) We could think of an initial evolutionary event, occurred before mammalian radiation, which has been shared between dromedary and dog.

Comparison of the genomic structure of the dromedary with sheep TRG loci shows a more evident similarity confirmed by the length of the lines that are continuous and with few interruptions (Supplementary Fig. 6). Supplementary Fig. 6a highlights extensive collinearity shared among the entire TRGC5 cassette in *Camelus dromedarius* and the corresponding cassette of the ovine TRG1 locus (red box) further confirming the ancestral nature of the latter. Moreover, comparison between dromedary and sheep TRG loci enhances an almost perfect colinearity with the TRG2 ovine locus and, as expected, the exclusion of the dromedary TRGV genes of TRGC5 cassette (Supplementary Fig. 6b).

4. Conclusion

This research has broadened our previous work to produce a comprehensive genomic map of the dromedary TRG locus and to focus on its evolutionary history as compared to the single human locus and to the two sheep TRG1 and TRG2 loci.

The identification of the AMPH and STARD3NL genes flanking the dromedary TRG locus as all mammalian species so far studied, would indicate that also the dromedary locus is now fully characterized and it consists of three in tandem TRGC5, TRGC1, and TRGC2 cassettes.

Dromedary and sheep species belong respectively to Tylopoda and Ruminantia suborders, both of which are part of Artiodactyla order. The comparative genomic analysis together with the TRGV phylogenetic tree point how the entire structure of the dromedary TRG locus would resemble the ancestral locus arose before sheep and dromedary divergence. As a matter of fact, Vaccarelli and coworkers (Vaccarelli et al., 2008) described an hypothetical evolutionary model for the ovine TRG loci where, starting from an ancestral TRGC cassette, the first duplicative events have led to the emergence of an intermediate locus consisting of the TRGC5, TRGC3 and TRGC6. This transitional locus is similar to the today's dromedary locus. Conversely, looking at the TRGC tree, the duplications could have been independent in the two species starting from the ancestral TRGC5 cassette.

The expression assay conducted on peripheral blood of one adult animal has revealed that the repertoire relative to the productive TRGC5 cassette is poorly represented and that it is also the result of the mechanism of SMH already highlighted in spleen productive cDNA clones from TRGC1 and TRGC2 cassettes (Vaccarelli et al., 2012).

Ethics approval and consent to participate

Blood samples were provided by Zoo Safari, Fasano, Italy and were collected from three animals. The availability of blood of this study was a byproduct of standard health checks. The blood samples were taken from the jugular vein. Training of medical behaviours brings the animals to collaborate completely, assuming and maintaining the positions that allow the veterinarian, to perform the necessary procedures to monitor their welfare.

Declaration of competing 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.

Acknowledgments

We deeply acknowledge the wild camel protection foundation (WCPF) for providing samples and in-depth knowledge about camels. We thank Dr. Daniele Laguardia for providing us with adult dromedary blood samples from the Zoo Safari in Fasano (Italy). The valuable advice of Prof. Elena Ciani is also acknowledged. The financial support of University of Bari and University of Salento is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dci.2020.103614.

References

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., 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., Bellini, M.G., Linguiti, G., Ciccarese, S., Massari, S., 2019. Comparative analysis of the TRB locus in the *Camelus* genus. Front. Genet. 10, 482. https://doi.org/10.3389/fgene.2019.00482.

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 Web Server issue.

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.

Ciccarese, S., Burger, P.A., Ciani, E., Castelli, V., Linguiti, G., Plasil, M., Massari, S., Horin, P., Antonacci, R., 2019. The camel adaptive immune receptors repertoire as a singular example of structural and functional genomics. Front. Genet. 10, 997. https://doi.org/10.3389/fgene.2019.00997.

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.

Damani-Yokota, P., Gillespie, A., Pasman, Y., Merico, D., Connelley, T.K., Kaushik, A., Baldwin, C.L., 2018. Bovine T cell receptors and γδ WC1 co-receptor transcriptome analysis during the first month of life. Dev. Comp. Immunol. 88, 190–199. https://doi.org/10.1016/j.dci.2018.07.023.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment with reduced time and space complexity. BMC Bioinf. 5, 113.

Elbers, J.P., Rogers, M.F., Perelman, P.L., Proskuryakova, A.A., Serdyukova, N.A., Johnson, W.E., Horin, P., Corander, J., Murphy, D., Burger, P.A., 2019. Improving Illumina assemblies with Hi-C and long reads: an example with the North African dromedary. Mol. Ecol. Resour. 19, 1015–1026.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Fitak, R.R., Mohandesan, E., Corander, J., Burger, P.A., 2016. The de novo genome assembly and annotation of a female domestic dromedary of North African origin. Mol. Ecol. Resour. 16 (1), 314–324 Jan.

Giudicelli, V., Chaume, D., Lefranc, M.-P., 2005. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res. 33, D256–D261. https://doi.org/10.1093/nar/gki010.

Giudicelli, V., Brochet, X., Lefranc, M.-P., 2011. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T celle 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 junction of the rearranged immunoglobulins (IG) and T cell receptors(TR). Cold Spring Harb. Protoc. 6, 716–725.

Hein, W.R., 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.

Jirimutu, Wang, Z., Ding, G., 2012. Genome sequences of wild and domestic Bactrian camels. Nat. Commun. 3, 1202. https://doi.org/10.1038/ncomms2192.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

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., Pommié, C., Kaas, Q., Duprat, E., Bosc, N., Guiraudou, D., Jean, C., Ruiz, M., Da Piédade, I., Rouard, M., Foulquier, E., Thouvenin, V., Lefranc, G., 2005. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Dev. Comp. Immunol. 29, 185–203.
Lefranc, M.-P., 2014. Immunoglobulin (IG) and T cell receptor genes (TR): IMGT and the birth and rise of immunoinformatic. Front. Immunol. 5, 22.
Linguiti, G., Antonacci, R., Tasco, G., Grande, F., Casadio, R., Massari, S., Castelli, V., Consiglio, A., Lefranc, M.P., Ciccarese, S., 2016. Genomic and expression analyses of Tursiops truncatus T cell receptor gamma (TRG) and alpha/delta (TRA/TRD) loci reveal a similar basic public γδ repertoire in dolphin and human. BMC Genomics 17, 634 Erratum in: BMC Genomics 2016, 17, 778.
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.
Massari, S., Bellahcene, F., Vaccarelli, G., Carelli, G., Mineccia, M., Lefranc, M.-P., Antonacci, R., Ciccarese, S., 2009. The deduced structure of the T cell receptor gamma locus in *Canis lupus familiaris*. Mol. Immunol. 46, 2728–2736.
Massari, S., Ciccarese, S., Antonacci, R., 2012. Structural and comparative analysis of the T cell receptor gamma (TRG) locus in Oryctolagus cuniculus. Immunogenetics 64, 773–779. https://doi.org/10.1007/s00251-012-0634-0.
Miccoli, M.C., Antonacci, R., Vaccarelli, G., Lanave, C., Massari, S., Cribiu, E., 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.
Miccoli, M.C., Vaccarelli, G., Lanave, C., Cribiu, E.P., Ciccarese, S., 2005. Comparative analyses of sheep and human TRG joining regions: evolution of J genes in Bovidae is driven by sequence conservation in their promoters for germline transcription. Gene 355, 67–78.
Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.
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.
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., 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 (TCRG) variable domains. Eur. J. Immunol. 42, 1–13.
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.