

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Data characterizing the genomic structure of the T cell receptor (TRB) locus in *Camelus dromedarius*



Rachele Antonacci^{a,}, Mariagrazia Bellini^a, Vito Castelli^a, Salvatrice Ciccarese^a, Serafina Massari^b

^a Department of Biology, University "Aldo Moro" of Bari, Bari, Italy

^b Department of Biological and Environmental Science e Technologies, University of Salento, Lecce, Italy

ARTICLE INFO

Article history: Received 30 May 2017 Received in revised form 21 July 2017 Accepted 1 August 2017 Available online 3 August 2017

Keywords: T cell receptor TRB locus Dromedary genome Camelus dromedarius IMGT

ABSTRACT

These data are presented in support of structural and evolutionary analysis of the published article entitled "The occurrence of three D-J-C clusters within the dromedary TRB locus highlights a shared evolution in Tylopoda, Ruminantia and Suina" (Antonacci et al., 2017) [1]. Here we describe the genomic structure and the gene content of the T cell receptor beta chain (TRB) locus in *Camelus dromedarius*. As in the other species of mammals, 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. A peculiarity of the dromedary TRB locus structure is the presence of three TRBD-J-C clusters, which is a common feature of sheep, cattle and pig sequences.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

http://dx.doi.org/10.1016/j.dib.2017.08.002

DOI of original article: http://dx.doi.org/10.1016/j.dci.2017.05.021

^{*} Correspondence to: Dept. of Biology, University of Bari, Via Orabona, 4, 70126 Bari, Italy.

E-mail addresses: rachele.antonacci@uniba.it (R. Antonacci), mg1192@hotmail.it (M. Bellini), vtcast@gmail.com (V. Castelli), salvatricemaria.ciccarese@uniba.it (S. Ciccarese), sara.massari@unisalento.it (S. Massari).

^{2352-3409/© 2017} The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Subject area More specific subject area	Biology, genetics, genomics Genetics, Genomics and Molecular Biology
Type of data	Tables and figures
How data was acquired	A standard BLAST search (Basic Local Alignment Search Tool. http://blast.ncbi. nlm.nih.gov/Blast.cgi.) of the public dromedary genomic assembly, Long PCR on genomic DNA and cloning
Data format	Analyzed
Experimental factors	Sequence analysis and dromedary DNA extraction
Experimental features	Dromedary lung genomic DNA was prepared from a single healthy animal. PCRs were performed by High Fidelity DNA polymerase. The PCR products were purified and cloned into the TA-vector system.
Data source location	Bari and Lecce, Italy
Data accessibility	The whole dromedary genome shotgun sequence is available at GenBank (ID: GCA_000767585.1). Sequence data published with this article were registered in EMBL database with the Accession number LT837971

Specifications Table

Value of the data

- These data insight into the genomic structure of the T cell receptor (TRB) locus in *Camelus dromedaries*. This results in the first, mostly complete, map of the TRB locus in a species of the Tylopoda suborder.
- The dromedary TRB locus characterization can be used to increase the understanding in the evolution of Camelidae and to contribute to solving the relative placement of this species within the Artiodactyla order.
- The availability of the sequence of the dromedary TRB locus allows researchers to concentrate on functional study and provides a tool to use this specie as a valuable model for immunological research.

1. Data

Data presented in the text include tables and figures giving information on the genomic structure and the gene content of the dromedary TRB locus, a mammalian species belonging to the *Camelus* genus. This information was obtained by integrating the sequence data deduced from the public genomic assembly [2] with sequences obtained by PCR experiments conducted in our laboratory. Table 1 describes position, classification and functionality of the TRB genes retrieved from the dromedary public genome assembly. Table 2 shows the description of the dromedary TRBV pseudogenes. Table 3 describes position, classification and functionality of the unrelated TRB genes recovered from the dromedary TRBV genes according to IMGT unique numbering for the V-REGION [6]. Table 4 provides the list of the genomic clones of the dromedary TRBD-J-C region with the primer pairs used and the PCR conditions. Fig. 2 shows the TRBD, the TRBJ and the TRBC gene sequences.

Table 1

Description of the TRB genes in the Camelus dromedarius genome assembly. The position of all genes and their classification and functionality are reported.

Gene classification	Functionality ^a	NCBI Reference Sequence	Position ^b
TRBV1	F	NW_011591622	861263-861886
TRBV2	F	NW_011591622	932263-932714
TRBV3	Р	NW_011591622	927952-928412
TRBV5S1	F	NW_011591622	937384-937843
TRBV5S2	F	NW_011591622	940879-941358
TRBV5S3	F	NW_011591622	955293-955748
TRBV6	F	NW_011591622	944809-945237
TRBV7S1	F	NW_011591622	947134-947581
TRBV7S2	F	NW_011591622	962228-962689
TRBV8	F	NW_011591622	950124-950593
TRBV9	Р	NW_011591622	965923-966346
TRBV10	F	NW_011591622	970368-970809
TRBV11	F	NW_011591622	975860-976308
TRBV12S1	Р	NW_011591622	981727-982197
TRBV12S2	Р	NW_011591622	992125-992569
TRBV14	Р	NW_011591622	995472-995906
TRBV15S1	F	NW_011591622	997569-998023
TRBV15S2	F	NW_011591622	999129-999583
TRBV16	F	NW_011591622	1003645-1004098
TRBV19	F	NW_011591622	1018094-1018641
TRBV20	F	NW_011591622	1020910-1021565
TRBV21S1	F	NW_011591622	1028337-1028797
TRBV21S2	F	NW_011591151	70843-70731
TRBV2153	P	NW_011591151	62/38-62511
TRBV22	F	NW_011591151	40518-40381
	P	NW 011501151	56428 56106
TRBV25	F	NW/ 011501151	57347-57210
TRBV26	F	NW/ 011591151	66428-66297
TRBV27	F	NW 011591151	41158-41032
TRBV28	F	NW 011591151	32762-32640
TRBV29	F	NW 011591151	27109-26837
TRBD1	F	NW_011591151	9932-9943
TRBJ1-1	F	NW_011591151	9247-9294
TRBJ1-2	F	NW_011591151	9116-9159
TRBJ1-3	F	NW_011591151	8861-8910
TRBJ1-4	F	NW_011591151	8258-8308
TRBJ1-5	F	NW_011591151	7982-8031
TRBJ1-6	F	NW_011591151	7491-7543
TRBC1	F	NW_011591151	EX1 4773-5166
			EX2 4311-4328
			EX3 4044-4150
			EX4 3711-3731
TRBC3	nd	NW_011591151	EX2 2866-2883
			EX3 2599-2705
TPRI3_1	F	NW 011620189	EA4 2200-2280 653-702
	r E	NW 011601111	202-702
TRBI3_2	F	NW 011601111	2234-2203
TPRI3_3	F	NW 011601111	2420-2470
TRBI3_4	nd	NW 011601111	2042-2030
TRBI2-2	F	NW 011616084	215-265
TRBI2-3	nd	NW 011616084	2-46
TRBI2-6	nd	NW 011607149	185-231
TRBC2	nd	NW 011593440	EX1 1911-2149
		011353110	EX2 2622-2639
			EX3 2800-2906
			EX4 3190-3210
TRBV30	F	NW_011593440	14509-14160

^a nd: not defined (indicates that the nt sequence of the gene is incomplete and its functionality cannot be defined). ^b L-PART1/ V-exon for TRBV genes and coding sequence for TRBD and TRBJ.

TRBV genes	Defective Leader	Frameshift	Stop codon	Defective splice sites	Defective RSS
TRBV3 TRBV9 TRBV12S1 TRBV12S2 TRBV14 TRBV21S3 TRBV23 TRBV24	•	• • •	•	•	•

Table 2Description of the Camdro TRBV pseudogenes.

2. Experimental design, materials and methods

2.1. Analysis of the dromedary TRB locus retrieved from the genome assembly: identification of the related and unrelated TRB genes

We employed the recent submission to NCBI (BioProject PRINA234474) of a draft genome sequence from the Arabian camel [2] to identify the TRB locus in this species. A standard BLAST search (Basic Local Alignment Search Tool. http://blast.ncbi.nlm.nih.gov/Blast.cgi.) 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. We directly retrieved a sequence of 457871 pb (gaps included) from the PRJNA234474_Ca_dromedarius_V1.0 assembly that corresponds to eight distinct unplaced and not continuous scaffolds (Fig. 1 in [1]). The sequence comprises the MOXD2 and the EPHB6 genes that flank the 5' and 3' ends, respectively, of all mammalian TRB loci studied to date. All dromedary TRB genes have been recognized and annotated while taking into account both the human sequence and the sheep genomic D-I-C region as a reference [3-5] (Table 1). The functionality of V, I 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 TRBV genes; (b) determination of proper recombination signal (RS) sequences located at 3' of the TRBV, 5' of the TRBJ, and 3' and 5' ends of the TRBD 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. We annotated 33 TRBV germline genes (twenty-five functional genes and eight pseudogenes) (Table 2), one TRBD, 13 TRBJ and two complete and one incomplete TRBC genes. The analysis of the 3' part of the locus revealed the potential presence of three D-J-C clusters similar to clusters found in sheep [4,5].

We also identified and annotated four trypsin-like serine protease (TRY) genes (Table 3). In this context, downstream of the TRBV1 gene, proceeding from 5' to 3', we found as in humans two protease genes that we recognized tentatively, according to their genomic position, as TRY1 (alias PRSS58 or TRYX3) and TRY2 (alias TRY2P), respectively. A third TRY gene, named TRY3, was homologous to a gene located after the TRY2P gene in humans that was found within the NW_011623391 unplaced scaffold. Extrapolation of the synteny with the human sequence predicts that the NW_011623391 scaffold should be juxtaposed within the dromedary TRB locus, upstream of the TRBV3 gene (Fig. 1 in [1]). An additional TRY gene, classified as TRY4, was found before the D-J-C region. Thus, unlike humans, only one TRY gene encompasses the array of the TRBV genes. All dromedary TRY genes appear putatively functional with the presence of correct acceptor and donor splicing site and an absence of frameshifts and stop codon in their coding regions. The genomic structure of the MOXD2 and EPHB6 genes, which delimit the TRB locus, was also defined (Table 3).

Table 3

Description of the unrelated TRB genes in the *Camelus dromedarius* genome assembly. The position of all genes and their classification and functionality are reported.

Gene classification	Functionality ^a	NCBI reference sequence	Position
MOXD2	F	NW_011591622	850155-856730
TRY1	F	NW_011591622	870036-876394
TRY2	F	NW_011591622	882909-888072
TRY3	nd	NW_011623391	1-2387
TRY4	F	NW_011591151	13974-17714
EPBH6	F	NW_011593440	46466-60647

^a nd: not defined (indicates that the nt sequence of the gene is incomplete and its functionality cannot be defined).

	L-PART1+L-PART2		FR1-II (1-26)	IGT	CDR1-IMGT (27-38)	FR2-1 (39-55	EMGT	CDR2-IMGT		FR3- (66-10	IMGT		CDR3-IMGT (105-117)
			A	В	BC	C	C'	C'C"	C"	D	E	F	FG
			(1-15)	(16-26)	(27-38)	(39-46)	(47-55)	(56-65)	(66-74)	(75 - 84)	(85-96)	(97 - 104)	(105 - 117)
			>	>		>	>		>	>	>	>	
Gene Fun	ctio	nality	1 10 15	16 23 26	27 38	3941 46	47 55	56 65	66 74	75 80 84	85 89 96	97 104	105
				1	1	1.1							1
TRBV1	F	MVLFLILWGWVLMAAVSTDAT	ASFVEQRPRWVLVPR	GQAETLRCILR	NSQYPW	MSWYQQDL	QGQLQMLAS	LRSPGD	KEVISLP.G	ADYQATRVNN	TELRLHVANVTQ	GRTLLCT C	SK[6.6.2]
TRBV2	F	MHSRLFCWVIFGLLGVGHT	ELGVSQSPRHYVTAM	GQNVTLRCDPI	PGHLY	LSWYRQTP	GKGMEFLLS	FYNKAP	SEKADFLKD	HFLAEMP.NG	LYLTLKIQPVQP	GDSAMILC	ASSV[5.6.4]
TRBV3	Ρ	MGSRLLCCVVLYFLGAGAL	DTGVFQMPKYLITQM	GSKMSLK C EQK	LGYDS	MYWYKQGS	KQLLKIMFI	YNNKEL	ILNETVP.S	RFSPEST.DK	AHLNLHVDSLEP	DDSAMYFC	ASS*[5.6.4]
TRBV5S1	F	MGSCLLCCVVLCLLGAGPV	DSGVTQTPRHLIKAR	GQQVTLSCSPI	SGHTS	VSWYQQAP	SEGPQFLFE	FYEKMQ	RAKGNFP.D	RFSGQQF.DN	YSSELIVNFLEV	TDSALYFC	ASSL[5.6.4]
TRBV5S2	F	MGSRLLCWVALCLLGAGQV	DSEVTQTPKYLIKPR	KQQVTLRCSPV	SGHLS	VYWYQQAL	GQGPQFLIQ	YYRQKE	TERGNIS.R	RFSGQQF.DN	SSSELNVSSLEL	TDSALYLC	ASSR[5.6.4]
TRBV5S3	F	MGPTLLPWAMLCLLGAGPL	EAGVTQTPRHLIKTT	GQMVTLRCSPI	SGHLG	ISWYQQAW	SQGPQLLFE	FYERRQ	TAKGNFP.D	RFLAQQF.GD	SRSELNVSSLQL	SDSALYLC	ASSL[5.6.4]
TRBV6	F	MCTCLLCCVAFCLLQAGPV	NAGVIQTPKFQVLKM	GQGVTLKCAQD	LKHDY	MYWYRQDP	GHGLRLIHY	SAGSGT	TNKGDIP.D	GYSVSRS.DV	ENFTLKLEYATP	NQTSVYLC	ASSY[5.6.4]
TRBV7S1	F	MGTRLLYWATLYLLGVGHT	EAGISQSPRHRVTGR	GQNVTLRCDPV	SDHNR	LYWYRQTL	GQGLEILTL	FQNKDA	VDTSGMPDA	RFSAKRP.QG	LSSTLEIQRVEP	GDSAVYLC	ASSL[5.6.4]
TRBV7S2	F	MGTRLLYWATLCLLGVGHT	KAGVSQSPRHRVTGR	GQNVTLMCDPV	SDHNR	LYWYRQML	GQGLEILTY	FQDDDA	LDKSGMPDA	RFSAERP.QG	LSSTLKIQRVEP	GDSAEYLC	ASSL[5.6.4]
TRBV8	F	MSPSLFCVILCLLGAGPT	DAEITQTPRYRIVST	GHKTVLECSQH	MNHFG	MFWYQQDP	GQGPGLIHY	SNDIGS	TAEGEVT.E	GYRVSRP.EK	AHFPLTLESAST	NQTSLYLC	ASSV[5.6.4]
TRBV10	F	MVTRLFFCAASCLLWTGQV	EAGITQSPRYKIAVI	GETLTLQ C HQT	DNHDY	MYWYRQDL	GHGLKLIYY	TYDVNK	TDKGEIP.D	GYSVSRS.NI	GDFPLTLESVVP	SRTSAYLC	ASSN[5.6.4]
TRBV11	F	MGTRLLCWATLCLLGAELT	EGGVAQSPRHKITEE	RQAVALWCDPI	SGHSF	LYWYRQTL	EQDLELLIL	FEETAG	INVLQSPKN	RFSAERP.SG	ANSTLKIQSAEL	GDSAVYL C	ASSL[5.6.4]
TRBV15S1	F	MGPRLLHCVALCLFGAGPV	GAMVNQTPRYWVTGT	GKPVTLN C SQT	LNHDT	MY W YQQKP	SQAPKQLLY	YYNTQL	NRETDTS.D	NFQSRQF.NT	SSCSLDIRSAGV	GDSAVYLC	ASSR[5.6.4]
TRBV15S2	F	MGPRLLHCVALCLFGAGPV	GAMVNQTPRYWVTGT	GKPVTLNCSQT	LNHDT	MYWYQQKP	SQAPKQLLY	YYNTQL	NRETDTS.D	NFQSRQF.NT	SSCSLDIRSAGV	GDSAVYLC	ASSR[5.6.4]
TRBV16	F	MCQTVLGLMVLCLLGAGFL	EAEVTQTPGHLVKGT	GQKAEMY C VPR	KGHNY	VCWYQQIP	AKEFKFLIS	FQNNNE	LDKTGMPKE	RFSAVCY.QN	SSCTLKIQPAAL	QDSAVYL C	ASSE[5.6.4]
TRBV19	F	MGSSVIYWVALCLIGAGTV	DGGITQTPKYVLKEE	GRDVTLECEQD	FNYDT	MYWYRQDP	GQRPRLIYF	SLSAKD	VQKGDMA.E	GYSASRE.KK	PLFPLTVTLTQK	NQTAVYLC	ASSI[5.6.4]
TRBV20	F	MLLFLLLLGPGSGL	GAQVSQHPSRAIRER	GASVTFQCRAE	DFQATI	MFWYRQFP	EQGLTLIAT	SNQGSNA	TYEQGFTKD	KFPINHP.SL	TFSSLTVTSVSP	ADSSLYFC	SAS[6.7.3]
TRBV21S1	F	MCQGLLCCLALLFWGAGST	DTGVTQSPGHLVREK	EQRAKMDCVPK	EGHNY	VYWYRKKL	EEAFEFLVY	LRNGGD	IMEEAEGFK	QRFSAKCPKN	SPCSLEINSTQA	GDSARYF C	ASSE[5.6.4]
TRBV21S2	F	MCLSLLCCVALLLWGAGSV	DTGVTQSPGHLVREK	EQRAKMYCVPK	EGHAY	VYWYRKKL	EEAFEFLVY	LRNEDI	VEKIA.GFE	QRYSAKCPKN	SPCSLEINSTQA	GDSARYF C	ASSE[5.6.4]
TRBV22	F	MCSWALCYVVLCLLGAGP	TDAEIYQRRFLLAGA	GRDVTLE C EQS	LRYNA	MYWYRQDP	GLGLRLIYY	SKVEKD	VQRGDIS.E	GYSVSRE.EK	GLFPLTVKLAHI	NQTALYVC	SGSA[5.6.4]
TRBV23	P	MGTWVLCSVALVFWEQA	L*MPVSPRLQDPWGK	GQTANMDCIPP	KGHTF	FNWNQQIQ	KKELKFLIS	FQNERV	LDQIELVKK	RVSAECP.PN	SPCSLEIQSSEP	GDSALYF C	ARSQ[5.6.4]
TRBV25	F	MAVRLLCYVALSLLGAGLT	NADVNQTPRYAVIGT	GKKITLECNQA	MDHES	MYWYRQDP	GMEPQLIHY	SHGVNT	TEKGDRP.S	ESTVSRI.RK	DOFPLTLEATSP	SQTSRYFC	ASSE[5.6.4]
TRBV26	F	MSNRLLCCALIFLIRAGLK	EAAVTQSPRHRILKT	EQKLTLQCSQR	MNHYS	MYWYRQDP	GFGLRLIYY	STGTRT	TKQGDVP.E	GYHVSRE.DQ	ADFPLTLQSAST	NQTSVYFC	ASSE[5.6.4]
TRBV27	F	MGPRLLGWVLLWLLGAGPV	EADVTQSPTHRITET	RKKLTVTCSQN	MNHEA	MYWYRQDP	GLGLRLIYF	SRNVNF	TEKGDIP.D	GYTVSRK.EK	KDFSLTLESASI	TQTSLYLC	ASSV[5.6.4]
TRBV28	F	MGIRLLCAFCFLGVGLM	DAKVTQIPRYLVKRM	GEKVLIDCVQD	MGHER	MFWYRQDP	ALGLQLLHY	SYDVDI	VEKGDMP.E	GYNVSRK.KK	ERFALTLESAST	SQTSVYLC	ASSI[5.6.4]
TRBV29	F	MLTLLLLLGLGSAF	GALVSQKPSRAICQH	GVSVMIQ C EAD	SQVIS	MYWYRQRP	GQSLSLIAT	ANQGSQA	TYESGFAKE	KFPISRP.SL	NFSVLTVSSVSP	EDSSSYFC	SA[5.7.2]
TRBV30	F	MPCCLLALLLGTFLGVR	AQTIHORPPTMVQPM	GSSLSLECTVK	GTSNPS	LYWYRQAA	GGALQLLFY	SFSAD	QVDSEEP.Q	NFEASVS.QD	GSFILSSKKLLL	NNSGFYL C	AWS[6.5.3]

Fig. 1. The IMGT Protein display of the dromedary TRBV genes. Only functional genes and in-frame pseudogenes are shown. 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 [6]. The amino acid length of the CDR-IMGT AA is also indicated in square brackets.

Table 4

Camelus dromedarius D-J-C region genomic clones. The primer sequences, the PCR conditions and the size of each clone are reported.

Clone	Primer pairs sequence (5'-3')	Primer location	T annealing	Product length (bp)
pSCBJ11	JB11U: CTTTGGAGAAGGCACCAG	TRBJ1-1 gene	55/58	4396
pSCJ22KN		TRBC gene exon 1	53/55	5000
pSCBJ27U	JB34U: AACACTCAGTACTTTGGC	TRBJ3-4 gene TRBJ3-4 gene	56/58	4077
pSCBD3	CB2U: GCACAACCCCCGCAACCA	TRBC gene exon 1	55/56	4848
pSCMG1	JB23L: CCGCCGAAAAACAGTGTC JB23U: GACACTGTTTTTCGGCGG	TRBJ2-3 gene	55/58	3160
pSCB2C8	CB2L: TGGTTGCGGGGGTTGTGC CB2U: GCACAACCCCCGGCAACCA 3UTR:GTTGAGCTCACTTTGCAGGG	TRBC gene exon 1 TRBC gene exon 1 TRBC2 gene 3UTR	62	1331

2.2. Protein display of the dromedary TRBV genes

The deduced amino acid sequences of the germline TRBV genes were manually aligned according to IMGT unique numbering for the V-REGION [6] to maximize the percentage of identity (Fig. 1). Only

А

TRBD gene name	5' D-NONAMER <u>GGTTTTTGT</u>	5'D-SPACER **********	5' D-HEPTAM CACTGTG	ER D-REGION 3'D	-HEPTAMER CACAGTG	3'D-SPACER ************************************	3'D-NONAMER ACAAAAACC
TRBD1	cgtttttgt	ataaagctgtaa	cgttgtg	GGGACAGGGGGC G T G G G Q G D R G	cacggtg	attcaaccctatgggaaatcttt	acaaaaacc
TRBD3	catttttgt	atcttggcttaa	cattgtg	GGACTATGGGGGGGAG G L W G G D Y G G E T M G G	cacaatg	attcaggtagagaaggtcctttt	acaaaaagc
TRBD2	catttttgt	atcttggcttaa	cattgtg	GGACTATGGGGGGGGACACTAGTTGTGGGGGGGGG G L W G D T S C G G	cacaatg	attcaggtagagaaggtcctttt	acaaaaagc

DYGGTLVVGG TMGGH*LWGE

В

TRBJ gene name	J-NONAMER GGTTTTTGT	J-SPACER *********	J-HEPTAMER CACTGTG		
gone name	001111101		01101010	J-REGION 5'splice	donor
TRBJ1-1	gttttcctt	tttgctccgtgt	cactgtg	TGAACACTGAAGTTTTCTTTGGAGAAGGCACCAGACTCACGGTTGTAG N T E V F <u>F G E G</u> T R L T V V	gtaaga
TRBJ1-2	cgttgtagg	gtggctgtattc	tgatgtg	TTATGACTTTAACTTCGGCCCAGGGACCAAGCTGACGGTCGTAG Y D F N <u>F G P G</u> T K L T V V	gtaagg
TRBJ1-3	ggttttgaa	gtggatctggga	ggctgtg	CTTTGCAGACACCTACTATTTTGGGGAGGGAAGCCGGCTCACTGTTGTAG F A D T Y Y <u>F G E G</u> S R L T V V	gtaagc
TRBJ1-4	agttttcct	accaggetttag	tgttgtg	tgactaatgaaaact ctatticgcagc gcggacga agct ticcgtcttgg T N E K L Y \underline{F} G S G T K L S V L	gtaagt
TRBJ1-5	gggtctgcc	acactcgtgtgt	tactgtg	TAGCAACCGGGTGCAGCATTTTGGAGAAGGGACTCGACTCTCTGTCCTAG S N R V Q H \underline{F} G E G T R L S V L	gtaagg
TRBJ1-6	ggttttacc	acagctgtctgc	agctgtg	TTCCTATAATTCACCCCCCCACTTTGGGCTCGGCTCCAGGCTCACGTGACAG S Y N S P L H \underline{F} G L G S R L T V T	gtatgg
TRBJ3-1	gaattcttg	gcagccccttct	cactgtg	CTCCTACGCTGAGCAGTACTTCGGGGCAGGGACTCGGCTCACTGTGCTAG S Y A E Q Y <u>F G A G</u> T R L T V L	gtaaga
TRBJ3-2	agtttgctt	gtggatctccag	ggctgtg	TAAACAGCCAGCAGCTGTACTTTGGGGAGGGTTCCAAGCTGATTGTCCTGG N S Q Q L Y $\underline{F~G~E~G}$ S K L I V L	gtaagg
TRBJ3-3	ggtttttgg	cctgggttccag	ggctgtg	agcacagatcctctgtatttcggcggggggcacccgtctgaccgtgctag S T D P L Y \underline{F} G G \underline{G} T R L T V L	gtaagc
TRBJ3-4	tgttttgt	gctgcgccccgg	ggctgtg	AGCCAGAACACTCAGTACTTTGGCGCAGGCACCCGGCTGTCGGTGCTAG S Q N T Q Y \underline{F} G A G T R L S V L	gtgagc
TRBJ3-5	ggtttttgc	acgggggctgggg	ggccgtg	ACTCACAGACCCAGTACTTCGGGCCGGGCACGCGGCTCCTGGTGCTAG S Q T Q Y $\underline{F~G~P~G}$ T R L L V L	gtgagt
TRBJ3-6	ggtttctgg	ddddadcccddd	ggctgtg	CTTCCAGCAGCCGCCCGCAGCCGGCCGGCCGGCCGGCCGG	gtgagt
TRBJ3-7	ggtttgcgt	gcggggctcctc	ctctgtg	GTCCTATGAGCAGCATTTCGGCCCAGGCACCAGGCTCACGGTCTTAG S Y E Q H <u>F G P G</u> T R L T V L	gtaaga
TRBJ2-1	gaattettg	gcagccccttct	cactgtg	CTCCTACGCTGAGCAGTACTTCGGAGCAGGGACTCGGCTCACTGTGCTAG S Y A E Q Y <u>F G A G</u> T R L T V L	gtaaga
TRBJ2-2	agtcggtgc	cttagtccccag	ggctgtg	TGAACACCGGACAGCTGTATTTTGGGGAAGGTTCCAAGCTGACTGTCCCGG N T G Q L Y $\underline{F~G~E~G}$ S K L T V P	gtaagg
TRBJ2-3	ggtttttgt	cctgggcctcgg	ggcggtg	AGCAGACACTGTTTTTCGGCGGGGGAACCAGGCTGTCTGT	gtgagc
TRBJ2-4	ggtttctgt	gcggggttgggg	ggctgtg	AACAAGAGACCCAGTACTTCGGGCCCGGGCACGCGGCTCCTGGTGCTAG Q E T Q Y $\underline{F \ G \ P \ G}$ T R L L V L	gtgagt
TRBJ2-5	ggtttctgg	ggggageeeggg	ggctgtg	CTTCCAGCAGCCGCCCTGACCTTCGGGGCGGCAGCCGGCTGGCCGG L P A A A L T <u>F G A G</u> S R L A V P	gtgagt
TRBJ2-6	ggtttgcgt	gcggggctcctc	ctttgtg	CTCCTATGAGAGGTATTTCGGCCCAGGCACCAGGCTCACGGTCTAG SYERY <u>FGPG</u> TRLTVL	gtaaga

Fig. 2. Nucleotide and deduced amino acid sequences of the dromedary TRBD (a), TRBJ (b) and TRDC (c) genes. The consensus sequence of the heptamer and nonamer are provided at the top of the figure and underlined. The numbering adopted for the gene classification is reported on the left of each gene. The gene sequence retrieved from the Ca_dromedarius_VI.0 genomic assembly is highlighted in red. In (a), the inferred amino acid sequence of the TRBD genes in the three coding frames are reported. In (b), the donor splice site for each TRBJ is shown. The canonical FGXG amino acid motifs are underlined. The unusual TRBJ3.6 gene motif is in italics. In (c), IMGT Protein display of the dromedary TRBC genes. Descriptions of the strands and loops were collected according to the IMGT unique numbering for C-DOMAIN [7].

C														
TRBC1 TRBC3 TRBC2		A (1-15) PRVMVFEPSEAEI PRVAVFEPSEAEI PRVAVFEPSEAEI	AB B (16-26) 5 16 23 26 (123) SRTQKATLVCLAT SRTQKATLVCLAT	BC (27-38) 27 36 11 GFYPDHVE GFYPDHVE GFYPDHVE	C (39-45) 3941 45 1.111 LSWWVNGH LSWWVNGH	CD L23456 KEVQS. KEVQS. KEVQS.	D (77-84) 77-80-84 7[]1 .GVSEDPQPY .GVSTDPQPY	DE 23456776543. REDPNL.SDST REDPSL.SDST REDPSL.SDST	E (85-96) 85 89 211 YCLSSRLRVSAT YCLSSRLRVSAT YCLSSRLRVSAT	EF -> - 96 S . 12 FWH.N FWH.N	F (97-104) (7 104) (97 104) (97 104) (97 104) (97 104) (97 104) (97 104) (97 104) (97 -	FG (105-117) 105 11234567876543 QVHFHGLTDKDRDQWKEG.W QVHFHGLTDKDRDQWKED.W		G (118-128) > 118121 128
	CONNECTING-	REGION T	RANSMEMBRANE-RE	GION CYTOPL	ASMIC-REC	GION								
	[EX2]		[EX3]		[EX4]									
TRBC1 TRBC3 TRBC2	(Ⅲ) CGFTS (D) CGFTS (D) CGFTS	(V) SYQQGVLSA (V) SYQQGVLSA (V) SYQQGVLSA	TLLYEILLGKATLYA TLLYEILLGKATLYA TLLYEILLGKATLYA	VLVSALVLMAM VLVSALVLMAM VLVSALVLMAM	VKRKDS VKRKDS V RK KDS									
Fig. 2. (continued)														

potential functional genes and in-frame pseudogenes are shown. All sequences exhibit the typical framework regions (FR) and complementarity determining regions (CDR) as well as the four amino acids: cysteine 23 (1st-CYS) in FR1-IMGT, tryptophan 41 (CONSERVED-TRP) in FR2-IMGT, hydrophobic amino acid 89, and cysteine 104 (2nd-CYS) in FR3-IMGT [6]. Conversely, CDR-IMGT varies in amino acid composition and length. It should be noted that the TRBV21 genes show a difference in length of one amino acid in the FR3 that corresponds to a C^{''} strand that is shorter and has a diverse amino acid sequence for TRBV21S2 compared to the TRBV21S1 gene.

2.3. Isolation of the dromedary TRBD-J-C region and analysis of the gene content

To isolate the entire TRBD-J-C region, we set up six different PCRs to produce six consecutive amplicons that cover the region between the first TRBJ and the last TRBC gene. Mostly, for each amplification, we used a primer pair, a gene-specific primer designed on the sequence of the TRBJ genes identified within the cDNA clones (see [1]), and a conserved primer constructed on the first exon of the TRBC genes. For the isolation of the TRBC2 gene, a 3'UTR lower primer derived from the sequence of the genomic assembly was used. Amplification consisted of an initial denaturation step at 93 °C for 2 min followed by 10 amplification cycles that each comprised a denaturation step at 93 °C for 10 s, an annealing step with a low temperature (53–56 °C, according to the melting temperature of the primers) for 30 s, an extension step at 68 °C for 7 min, followed by 25 cycles with a higher annealing temperature (55–58 °C, according to the melting temperature of the primers) and a gradually increasing extension time of 20 s as well as a final incubation at 68 °C for 7 min. A 30-deoxyadenosine overhang was added to blunt-ended amplicons by incubation with 1.0 unit of Platinum Taq DNA Polymerase (Invitrogen) at 72 °C for 10 min. These products were purified and cloned into the StrataClone TA-vector per the manufacturer's instructions. For each sample, 6 to 10 colonies were propagated and bi-directionally sequenced using M13 and T7 vector-specific primers. All plasmid sequence data were manually analysed. For the list of the clones with the primer pairs used and the PCR conditions see Table 4. All the obtained amplicons were sequenced (Acc. no. LT837971). The sequenced region is schematically illustrated in Fig. 3 in [1].

The nucleotide and deduced amino acid sequences of the TRBD, TRBJ and TRBC genes classified according to the similarity to the sheep sequence are shown in Fig. 2.

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.

Transparency document. Supplementary material

Transparency document associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.08.002.

References

- R. Antonacci, M. Bellini, A. Pala, M. Mineccia, M.S. Hassanane, S. Ciccarese, S. Massari, The occurrence of three D-J-C clusters within the dromedary TRB locus highlights a shared evolution in Tylopoda, Ruminantia and Suina, Dev. Comp. Immunol. 76 (2017) 105–119.
- [2] (a) H. Wu, X. Guang, M.B. Al-Fageeh, J. Cao, S. Pan, H. Zhou, L. Zhang, M.H. Abutarboush, Y. Xing, Z. Xie, A.S. Alshanqeeti, Y. Zhang, Q. Yao, B.M. Al-Shomrani, D. Zhang, J. Li, M.M. Manee, Z. Yang, L. Yang, Y. Liu, J. Zhang, M.A. Altammami, S. Wang, L. Yu, W. Zhang, S. Liu, L. Ba, C. Liu, X. Yang, F. Meng, S. Wang, L. Li, E. Li, X. Li, K. Wu, S. Zhang, J. Wang, Y. Yin, H. Yang, A.M. Al-Swailem, J. Wang, Camelid genomes reveal evolution and adaptation to desert environments, Nat. Commun. 5 (2014) 5188;
 - (b) H. Wu, X. Guang, M.B. Al-Fageeh, J. Cao, S. Pan, H. Zhou, L. Zhang, M.H. Abutarboush, Y. Xing, Z. Xie, A.S. Alshanqeeti, Y. Zhang, Q. Yao, B.M. Al-Shomrani, D. Zhang, J. Li, M.M. Manee, Z. Yang, L. Yang, Y. Liu, J. Zhang, M.A. Altammami, S. Wang, L. Yu, W. Zhang, S. Liu, L. Ba, C. Liu, X. Yang, F. Meng, S. Wang, L. Li, E. Li, X. Li, K. Wu, S. Zhang, J. Wang, Y. Yin, H. Yang, A.M. Al-Swailem, J. Wang, Erratum in: Nat. Commun. 6 (2015) 6107.
- [3] M.-P. Lefranc, V. Giudicelli, P. Duroux, J. Jabado-Michaloud, G. Folch, S. Aouinti, E. Carillon, H. Duvergey, A. Houles, T. Paysan-Lafosse, S. Hadi-Saljoqi, S. Sasorith, G. Lefranc, S. Kossida, IMGT[®], the international ImMunoGeneTics information system[®] 25 years on, Nucleic Acids Res. 43 (2015) D413–D422.
- [4] S. Di Tommaso, R. Antonacci, S. Ciccarese, S. Massari, 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 Genom. 11 (2010) 3.
- [5] R. Antonacci, S. Di Tommaso, C. Lanave, E.P. Cribiu, S. Ciccarese, S. Massari, Organization, structure and evolution of 41 Kb of genomic DNA spanning the D-J-C region of the sheep TRB locus, Mol. Immunol. 45 (2008) 493–509.
- [6] M.-P. Lefranc, C. Pommié, M. Ruiz, V. Giudicelli, E. Foulquier, L. Truong, V. Thouvenin-Contet, G. Lefranc, IMCT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains, Dev. Comp. Immunol. 27 (2003) 55–77.
- [7] M.-P. Lefranc, C. Pommié, Q. Kaas, E. Duprat, N. Bosc, D. Guiraudou, C. Jean, M. Ruiz, I. Da Piédade, M. Rouard, E. Foulquier, V. Thouvenin, G. Lefranc, IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains, Dev. Comp. Immunol. 29 (2005) 185–203.