

Research article

Open Access

## Extensive analysis of D-J-C arrangements allows the identification of different mechanisms enhancing the diversity in sheep T cell receptor $\beta$ -chain repertoire

Silvia Di Tommaso<sup>1</sup>, Rachele Antonacci\*<sup>2</sup>, Salvatrice Ciccarese<sup>2</sup> and Serafina Massari<sup>1</sup>

Address: <sup>1</sup>Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Universita' del Salento, Lecce, Italy and <sup>2</sup>Dipartimento di Genetica e Microbiologia, Universita' degli Studi di Bari, Bari, Italy

Email: Silvia Di Tommaso - [silvia.dt@libero.it](mailto:silvia.dt@libero.it); Rachele Antonacci\* - [r.antonacci@biologia.uniba.it](mailto:r.antonacci@biologia.uniba.it); Salvatrice Ciccarese - [ciccarese@biologia.uniba.it](mailto:ciccarese@biologia.uniba.it); Serafina Massari - [sara.massari@unisalento.it](mailto:sara.massari@unisalento.it)

\* Corresponding author

Published: 4 January 2010

Received: 17 June 2009

BMC Genomics 2010, 11:3 doi:10.1186/1471-2164-11-3

Accepted: 4 January 2010

This article is available from: <http://www.biomedcentral.com/1471-2164/11/3>

© 2010 Di Tommaso et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** In most species of mammals, the *TRB* locus has the common feature of a library of *TRBV* genes positioned at the 5'- end of two in tandem aligned D-J-C gene clusters, each composed of a single *TRBD* gene, 6-7 *TRBJ* genes and one *TRBC* gene. An enhancer located at the 3'end of the last *TRBC* and a well-defined promoter situated at the 5'end of the *TRBD* gene and/or a undefined promoter situated at the 5'end of the *TRBD2* are sufficient to generate the full recombinase accessibility at the locus. In ruminant species, the 3'end of the *TRB* locus is characterized by the presence of three D-J-C clusters, each constituted by a single *TRBD*, 5-7 *TRBJ* and one *TRBC* genes with the center cluster showing a structure combined with the clusters upstream and downstream, suggesting that a unequal crossover occurred in the duplication. An enhancer downstream the last *TRBC*, and a promoter at the 5'-end of each *TRBD* gene are also present.

**Results:** In this paper we focused our attention on the analysis of a large number of sheep TR  $\beta$ -chain transcripts derived from four different lymphoid tissues of three diverse sheep breed animals to certify the use and frequency of the three gene clusters in the  $\beta$ -chain repertoire. As the sheep *TRB* locus genomic organization is known, the exact interpretation of the V-D-J rearrangements was fully determined. Our results clearly demonstrate that sheep  $\beta$ -chain constitutes a level of variability that is substantially larger than that described in other mammalian species. This is due not only to the increase of the number of D and J genes available to the somatic recombination, but also to the presence of the trans-rearrangement process. Moreover, the functional complexity of  $\beta$ -chain repertoire is resolved by other mechanisms such as alternative cis- and trans-splicing and recombinational diversification that seems to affect the variety of the constant region.

**Conclusion:** All together our data demonstrate that a disparate set of molecular mechanisms operate to perform a diversified repertoire in the sheep  $\beta$ -chain and this could confer some special biological properties to the corresponding  $\alpha\beta$  T cells in the ruminant lineage.

## Background

Mature T lymphocytes must express heterodimeric  $\alpha$  and  $\beta$  or  $\gamma$  and  $\delta$  chain T cell receptors (TRs) on its surface in order to provide protection from pathogens. The diversity of the TR repertoire derives in large part from the random somatic rearrangements of Variable (V), Diversity (D) and Joining (J) genes in the case of  $\delta$  and  $\beta$  chain, and Variable (V) and Joining (J) genes in the case of  $\gamma$ , and  $\alpha$  chain encoding the variable portion of these molecules during the T-cell differentiation.

The V(D)J process requires the binding of the lymphocyte-specific recombination activating gene 1 and 2 (RAG1/2) protein complex to recombination signal sequences (RSs) flanking the rearranging sides of the individual V, D and J genes [1]. Upon binding, the RAG1/2 recombinases introduce a nick at the border between the RS heptamer and the adjacent coding sequence. The DNA repair factors of the nonhomologous end-joining (NHEJ) machinery join the nicked genes [2]. The RSs consist of conserved heptamer and nonamer sequences, separated by a spacer of 12 or 23 bp of relatively non-conserved DNA. Efficient recombination involves pairs of genes flanked by dissimilar 12- and 23RSs (the 12/23 rule) [3]. However, at the locus encoding for the  $\beta$ -chain (*TRB*), despite the 12/23 compatibility, the *TRBD* 12RSs, but not the *TRBJ* 12RSs efficiently target V $\beta$  23RSs. This phenomenon termed "beyond 12/23 rule" [4], preserving the *TRBD* gene utilization, ensures an ordered V(D)J recombination at the *TRB* locus with the *TRBD*-to-*TRBJ* joining which occurs before the *TRBV*-to-*TRBD* gene assembly.

Diversity at the recombination level is further enhanced by other processes that include the exonuclease digestion (trimming) of 3'-V, 5'- and 3'-D, and 5'-J genes, the imprecise joining of nicked genes, and the addition of non germline nucleotides (N nucleotides) at the V-J, V-D and D-J junctions. For this reason the product of the V(D)J joining, corresponding to the CDR3 region in the chain, is markedly polymorphic and is dominant in the recognition of peptide. After transcription, the V(D)J sequence is spliced to the constant (C) gene.

The resources available to generate the potential repertoires and to establish the regulation are described by the genomic organization of the TR loci. In most species of mammals, the *TRB* locus has the common feature of a library of *TRBV* genes positioned at the 5'- end of two in tandem aligned D-J-C gene clusters, each composed of a single *TRBD*, 6-7 *TRBJ* and one *TRBC* genes, followed by a single *TRBV* gene with an inverted transcriptional orientation located at the 3'-end. This genomic organization is reported well conserved from human [5], mouse [6,7], rat [8], chimpanzee [9], rhesus monkey [10], and horse [11]. A peculiar feature of the mammalian *TRB* locus is the presence of two very similar *TRBC* genes, since they differ by

only a few residues in the coding region; conversely, they are different in their own 3'-UTR regions.

In the artiodactyls lineage, i.e., in sheep [12] as well as in cattle [13] and in pig [14], a duplication event within the 3'-end of the *TRB* locus has led to the generation of a third D-J-C cluster. The presence of an additional cluster produces an increase in the number of D and J genes available to partake in somatic recombination, but also expand the distance between the enhancer (E $\beta$ ) and the promoter (PD $\beta$ 1) elements within the locus. Surprisingly also, in presence of three D-J-C clusters, both the nucleotide and protein sequences of all three *TRBC* genes are highly similar. Only four amino acid residues have undergone replacement in the *TRBC1* gene with respect to the *TRBC2* and *TRBC3* genes, while the *TRBC3* 3'-UTR region is identical to that of *TRBC1* gene [12]. The amino acid replacements were located, two in the N- terminus and one in the E  $\beta$ -strand and in the FG loop of well-defined regions of the extracellular domain of the TRBC molecule [15].

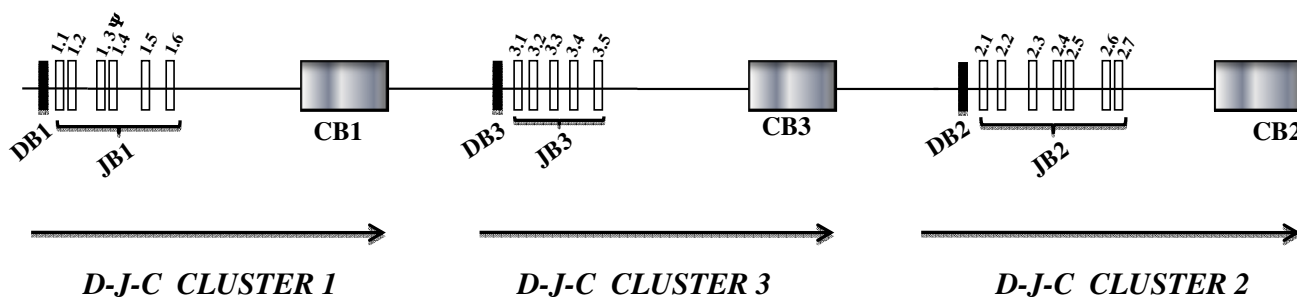
To know if the altered genomic architecture of the ruminant *TRB* locus can modify the mechanisms of recombination, we investigated on the  $\beta$ -chain repertoire in sheep. For this purpose we produced a collection of cDNAs derived from four different tissues belonging to four different adult animals of three diverse sheep breeds. As the genomic organization is known, the exact interpretation of the  $\beta$  chain transcripts was determined. The results of the analyses clearly demonstrate that sheep possess a repertoire of functional TR $\beta$  genes that is substantially larger than that described for other mammalian species, but also that other mechanisms as trans-rearrangement, intrallelic trans-splicing and DNA recombinational diversification involving the constant regions seem to shape the  $\beta$ -chain repertoire in a consistent way. However, the general paradigms of the mammalian *TRB* regulation seem to be preserved.

## Results

### Analysis of $\beta$ -chain transcripts

A previous study on cloning and sequencing of the sheep *TRB* locus revealed that the D-J-C region is organized in three independent clusters tandem aligned, with D-J-C cluster 3 additional with respect to the other mammalian *TRB* loci [12]. D-J-C cluster 1 contains one *TRBD*, six *TRBJ* and one *TRBC* gene. D-J-C cluster 3, located at 2.4 Kb downstream cluster 1, includes one *TRBD*, five *TRBJ* and one *TRBC* gene. Finally, D-J-C cluster 2 is positioned at 2.6 Kb downstream cluster 3 with one *TRBD*, seven *TRBJ* and one *TRBC* gene (fig. 1).

To evaluate the contribution of each gene cluster in the formation of the  $\beta$ -chain repertoire, a total of 72 clones containing rearranged V-D-J-C transcripts with a correct open reading frame were analyzed. All cDNA clones were



**Figure 1**  
**Schematic representation of the genomic organization of the 3'-end of the sheep TRB locus from TRBD1 to TRBC2 genes (modified from fig. 1 by Antonacci et al. [12]).**

registered in EMBL database with the Accession numbers from FM993913 to FM993984. 21 of these clones were derived from perinatal thymus (pSTMOS series) of a *Moscia Leccese* breed sheep, 15 from adult thymus (pSTA series) and 19 from spleen (pSMA series) of a *Gentile di Puglia* breed sheep, 17 from peripheral blood (pSSAR series) of a *Sarda Ionica* breed sheep. The clones were obtained by RT-PCR. The 5' primer was chosen on the YLCASS amino acid motif of the TRBV genes as members of the TRBV subgroups with this motif which seem to be the most frequently used [16] while the 3'-primer was designed on a conserved region of the three TRBC genes [12].

The deduced amino acid sequences of the V-D-J regions of all 72 cDNA clones are reported in the Table 1 together with the corresponding TRBC genes, according to the tissue of origin. Among the clones only one sequence is shared between blood (pSSAR25) and adult thymus (pSTA03). No tissue-specific expression of the genes was found. A total of 16 TRBJ genes were recovered within the different cDNAs. Thus, only one out of 17 functional TRBJ genes present in the genomic sequence was completely absent (TRBJ2.6). Besides, all TRBJ sequences match well with the corresponding genomic ones, and the high level of sequence similarity observed among the different animals is consistent with a close phylogeny of sheep breeds. The TRBJ2 cluster seems to be preferentially used (38/72 = 52.7%) and, although the numbers are too low to be statistically relevant, a slight increase in the use of TRBJ2.3 (14/38 = 36.8%) and TRBJ2.7 (10/38 = 26.3%) genes can be observed. Moreover, 20 clones retain a member of the TRBJ3 cluster, with the TRBJ3.4 gene (9/20 = 45%) more frequently used, while 14 clones retain the TRBJ1 gene set, without any preferential usage.

Three nucleotide differences at the N-terminus allow to distinguish the three TRBC gene isotypes: TRBC1 differs with respect to TRBC2 and TRBC3 genes for two nucle-

otide substitutions in the third and fourth codons; TRBC3 (as well as TRBC1 gene) is distinguishable from TRBC2 because of a silent nucleotide substitution at the third position of the first codon [12]. On the basis of these criteria, the N-terminus of the TRBC portions within the cDNA sequences was analyzed and a significant group of cDNAs with the TRBC3 gene (35/72 = 48.6%) identified. Moreover, 25 clones retain the TRBC2 (34.7%) and 12 clones are with the TRBC1 (16.6%) gene (Table 1).

More complex is the determination and the contribution of the genes involved in the CDR3 formation. The CDR3  $\beta$  region is defined as a stretch of nucleotides running after the codon encoding the cysteine in position 104 of the TRBV gene to the codon before that which encodes the phenylalanine of the FGXG motif of the TRBJ gene <http://imgt.cines.fr/> [17]. The corresponding amino acid sequence of the CDR3 loop deduced from the nucleotide sequences reveals that it is heterogeneous for amino acid composition (Table 1). The mean length of the CDR3 loop was approximately the same in spleen (mean 12.3 aa, range 10-16 aa) and adult thymus (mean 12.6 aa, range 9-16 aa), but larger in blood (mean 13.9 aa, range 10-15 aa) and young thymus (mean 13.7 aa, range 10-20 aa). For comparison, human peripheral blood CDR3 $\beta$  loop is about 12.7 residue long [18] and mouse is 11.9 residue long [19]. A similar CDR3 length and size range was reported in thymus and peripheral blood lymphocytes of piglets (mean 13.1 aa, range 10-17 aa) [20].

For a close inspection of the CDR3 s, the nucleotide sequences have been excised from each cDNA sequence and analyzed in detail. In the absence of the TRBV germline sequences, the deletions at the 3'-end of the TRBV and the nucleotide addition at the V-D junctions cannot be accurately estimated. However, the comparison of the 72 V-D-J junctions after the ASS motif allowed the determination of the probable 3'-end of the TRBV gene that has not been trimmed by exonuclease during rearrangement

**Table 1: Predicted amino acid sequences and length of the junctional diversity of the cDNAs. The classification of the TRBD, TRBJ and TRBC genes is indicated.**

CLONE	CDR3 sequence				Ac.N	D segment	J segment	C segment	CDR3 length
	V	N(D)N	J						
pSTMos.01	ASSQ	A(GT)VI	SYEQYFGPGTKLTVV	FM993913	D1	J2.7	C2	14	
pSTMos.02	ASSPT	NIAY	SYEQYFGPGTKLTVV	FM993914	--	J2.7	C2	14	
pSTMos.03	ASSQ	SD(G)	EQYFGPGTKLTVV	FM993915	D3	<b>J2.7</b>	<b>C3</b>	10	
pSTMos.04	ASSP	PP(GQ)V	NTEVFFGKGTTRLTVV	FM993916	D1	J1.1	C1	14	
pSTMos.05	ASSQG	R(D)L	SNNPLYFGGGTRLLVL	FM993917	D1	<b>J2.3</b>	<b>C3</b>	14	
pSTMos.06	ASSQG	P	NTDPLYFGAGSKLTVL	FM993918	--	J2.2	C2	12	
pSTMos.07	ASSR	VGTME(GQ)I	YNSOLQFGIGTRLTVT	FM993919	D1	J1.6	C1	18	
pSTMos.08	ASSPT	E(GLWGG)R	YEYFGPGTKLTVV	FM993920	D2	<b>J2.7</b>	<b>C3</b>	16	
pSTMos.09	ASSQE	AR(LG)AR	YFGAGTRLSVL	FM993921	D3	J3.3	C3	12	
pSTMos.11	ASSR	TALRP(LGE)HAL	YGELHFGPGTRLTVL	FM993922	D2	J2.1	C2	20	
pSTMos.12	ASSPT	PT	YDERHFGPGTRLTVL	FM993923	--	J3.1	C3	12	
pSTMos.13	ASSR	P(WGQG)D	GELHFGPGTRLTVL	FM993924	D3	J2.1	C2	14	
pSTMos.14	ASS	G	SNNPLYFGGGTRLLVL	FM993925	--	J2.3	C2	10	
pSTMos.17	ASSQD	M(G)ASA	AQLYFGAGSKLTVL	FM993926	D3	J3.2	C3	14	
pSTMos.18	ASSQE	(GA)DH	NPLYFGGGTRLLVL	FM993927	D3	<b>J2.3</b>	<b>C3</b>	13	
pSTMos.19	ASSQ	W	SNQAQHFHGHGTRLAVL	FM993928	--	J1.5	C1	11	
pSTMos.20	ASSPT	LR(G)IIPT	YEYFGPGTKLTVV	FM993929	D1	<b>J2.7</b>	<b>C3</b>	16	
pSTMos.21	ASS	RSR(WG)QD	SERYFGAGTRLTVL	FM993930	D3	J3.5	C3	14	
pSTMos.22	ASSR	(WG)QD	SERYFGAGTRLTVL	FM993931	D3	J3.5	C3	12	
pSTMos.23	ASSQ	APYG(TF)	SETQYFGPGTRLTVL	FM993932	D2	J3.4	C3	15	
pSTMos.25	ASSPT	R(Q)GG	DPLYFGAGSKLTVL	FM993933	D1	<b>J2.2</b>	<b>C3</b>	13	
pSSAR.01	ASSQ	SRR(D)VS	QTQYFGPGTRLTVL	FM993934	D1	J2.5	C2	14	
pSSAR.02	ASSQG	HR(TA)K	NERLYFGNGTKLSVL	FM993935	D1	J1.4	C1	15	
pSSAR.04	ASSQ	(AGGW)AL	SETQYFGPGTRLTVL	FM993936	D3	J3.4	C3	15	
pSSAR.05	ASSK	LG(R)DILN	EQYFGPGTKLTVV	FM993937	D1	<b>J2.7</b>	<b>C3</b>	14	
pSSAR.07	ASSQD	S(GTA)D	ERLYFGNGTKLSVL	FM993938	D1	J1.4	C1	14	
pSSAR.08	ASS	(LG)	NERLYFGNGTKLSVL	FM993939	D3	J1.4	C1	10	
pSSAR.10	ASSL	DI(R)PN	GELHFGPGTRLTVL	FM993940	D2	<b>J2.1</b>	<b>C3</b>	13	
pSSAR.11	ASSP	K(RD)GY	NPLYFGGGTRLLVL	FM993941	D1	J2.3	C2	13	
pSSAR.16	ASSQE	Q(Q)SRF	NNPLYFGGGTRLLVL	FM993942	D1	<b>J2.3</b>	<b>C1</b>	15	
pSSAR.17	ASSQ	SRRD	SNQAQHFHGHGTRLAIL	FM993943	--	J1.5	C1	14	
pSSAR.19	ASSP	S(DFG)IG	NNPLYFGGGTRLLVL	FM993944	D2	J2.3	C2	15	
pSSAR.23	ASS	LSTVD	SQSTQYFGAGTRLSVL	FM993945	--	J3.3	C3	14	
pSSAR.24	ASSQD	RKQGG	NSPLQFGIGTRLTVT	FM993946	--	J1.6	C1	15	
pSSAR.25	ASS	FFST(G)E	ETQYFGPGTRLTVL	FM993947	D3	J3.4	C3	13	
pSSAR.28	ASSQ	(GQ)DRI	NPOLYFGGGTRLLVL	FM993948	D1	J2.3	C2	14	
pSSAR.31	ASSP	DP(DSG)A	AQLYFGAGSKLTVL	FM993949	D1	J3.2	C3	14	
pSSAR.32	ASSQD	IS(QR)A	TDPLYFGAGSKLTVL	FM993950	D1	<b>J2.2</b>	<b>C3</b>	15	
pSMA.09	ASSP	(LWGG)D	NPLYFGGGTRLLVL	FM993951	D2	J2.3	C2	13	
pSMA.10	ASSQD	(AG)F	NPLYFGGGTRLLVL	FM993952	D3	J2.3	C2	12	
pSMA.41	ASSQ	KR(TAG)	ERHFGPGTRLTVL	FM993953	D1	J3.1	C3	12	
pSMA.42	ASS	L(GQRG)G	YEYFGPGTKLTVV	FM993954	D1	J2.7	C2	13	
pSMA.46	ASSQD	I	ETQYFGPGTRLTVL	FM993955	--	J3.4	C2	10	
pSMA.48	ASSQE	LT	YEYFGPGTKLTVV	FM993956	--	<b>J2.7</b>	<b>C3</b>	11	
pSMA.55	ASSR	(DL)C	NNPLYFGGGTRLLVL	FM993957	D2	<b>J2.3</b>	<b>C3</b>	12	
pSMA.59	ASSS	V(S)D	GELHFGPGTRLTVL	FM993958	D1	J2.1	C2	11	
pSMA.60	ASSP	AV(G)SD	NPLYFGGGTRLLVL	FM993959	D1	<b>J2.3</b>	<b>C3</b>	13	
pSMA.62	ASSP	Q(TAG)E	DPLYFGAGSKLTVL	FM993960	D1	J2.2	C2	13	
pSMA.65	ASSK	GR(TAG)P	SNNPLYFGGGTRLLVL	FM993961	D1	<b>J2.3</b>	<b>C3</b>	16	
pSMA.66	ASSS	DR(GW)S	QTQYFGPGTRLTVL	FM993962	D3	J2.5	C2	13	
pSMA.67	ASSK	TDF	YEYFGPGTKLTVV	FM993963	--	J2.7	C2	11	
pSMA.68	ASS	W(T)LNA	AQLYFGAGSKLTVL	FM993964	D1	J3.2	C2	12	
pSMA.70	ASSR	E(GTG)L	YEYFGPGTKLTVV	FM993965	D1	<b>J2.7</b>	<b>C3</b>	13	
pSMA.71	ASSP	(GP)T	NTEVFFGKGTTRLTVV	FM993966	D1	J1.1	C1	12	
pSMA.73	ASSP	FL(DS)V	DERHFGPGTRLTVL	FM993967	D1	J3.1	C3	13	
pSMA.74	ASS	R(H)QNI	TDTQYFGPGTRLSVL	FM993968	D1	J2.4	C2	13	

**Table 1: Predicted amino acid sequences and length of the junctional diversity of the cDNAs. The classification of the TRBD, TRBJ and TRBC genes is indicated. (Continued)**

pSMA.76	ASSP	(S)	SNNPLYFGGGTRLLVL	FM993969	D3	J2.3	C2	11
pSTA.01	ASSR	Y(SE)GD	EYHFGPGTKLTVV	FM993970	D1	J1.2	C3	12
pSTA.02	ASSQD	LV(GTA)R	YEHFGPGTKLTVV	FM993971	D1	J1.2	C3	15
pSTA.03	ASS	FFST(G)E	ETQYFGPGTRLLVL	FM993972	D3	J3.4	C3	13
pSTA.04	ASS	DS(W)DVQS	TQYFGPGTRLLVL	FM993973	D3	J3.4	C3	13
pSTA.06	ASSQD	(R)D	YEHFGPGTKLTVV	FM993974	D1	J1.2	C3	11
pSTA.08	ASSP	S(R)D	TEVFFGKGTRLTVV	FM993975	D1	J1.1	C3	11
pSTA.09	ASS	(LR)	ETQYFGPGTRLLVL	FM993976	D2	J3.4	C3	9
pSTA.11	ASSP	(GQR)PP	DTQYFGPGTRLSVL	FM993977	D1	J2.4	C2	13
pSTA.12	ASS	LS(GTRG)D	TQTQYFGPGTRLLVL	FM993978	D1	J2.5	C2	15
pSTA.13	ASSR	K(RG)H	SETQYFGPGTRLLVL	FM993979	D1	J3.4	C3	13
pSTA.15	ASS	IEKA	NPLYFGGGTRLLVL	FM993980	--	J2.3	C2	11
pSTA.24	ASSK	DL(AGG)VS	SETQYFGPGTRLLVL	FM993981	D3	J3.4	C2	16
pSTA.25	ASSL	ER(QR)	DERHFGPGTRLTVL	FM993982	D1	J3.1	C3	12
pSTA.26	ASSP	R(S)DK	GELHFGPGTRLTVL	FM993983	D1	<b>J2.1</b>	<b>C1</b>	12
pSTA.29	ASSP	RQ(T)GPFQ	EYHFGPGTKLTVV	FM993984	D1	J1.2	C1	14

in a significant proportion of sequences (Table 1). By the comparison of the TRBD genomic sequences, the nucleotides located in the CDR3 regions were considered to belong to a TRBD gene if they constituted a stretch of at least four consecutive residues corresponding to the TRBD1, TRBD3 or TRBD2 germline sequences. In this way the 72 sequences were grouped according to the TRBD1 (fig. 2a, 36 sequences), TRBD3 (fig. 2b, 16 sequences) or TRBD2 (fig. 2c, 8 sequences) gene usage. 12 sequences with no recognizable TRBD genes were grouped separately (fig. 2d). These last sequences could be interpreted as direct V-J junctions. However, it is also possible that nucleotide trimming masked the initial participation of D gene during the rearrangement. In the other cases the degree of germline nucleotide trimming in the 3'-V and 5'-J as well as the 5' and 3' D region is similar in all groups (fig. 2). Nucleotides that could not be attributed to any template sequence are considered N-elements. The mean length for N-D-N addition, including D region, is 15 nt (range 6-23 bases) for the first group (fig. 2a), 13.8 nt (range 4-22 bases) for the second group (fig. 2b) and 16 nt (range 6-33 bases) for the group with TRBD2 participation (fig. 2c). The mean of N addition in the clones without TRBD sequence (fig. 2d) is 8.3 nt (range 2-16 bases). Particular features of the CDR3 region of the clones are the presence within the D region of nucleotide substitutions as well as the presence of insertion (psTMOs 13 in fig. 2b) and deletion (psTA12 in fig. 2a) with respect to the germline sequences. Although the numbers are too low to be statistically relevant, a trend towards longer CDR3 length in TRBD2 (mean 42.3 bp, range 27-60) with respect to TRBD1 (mean 40.3 bp range 33-54) and TRBD3 (mean 38.5 bp, range 30-48), or with no apparent TRBD (mean 36.2 bp, range 30-42) transcripts was evident.

These data together suggest that all three TRB D-J-C clusters are used to generate in sheep functional TR  $\beta$ -chain with no specific influence of any clusters.

### Analysis of the D-J-C rearrangements

Since the genomic organization of the 3' region of the sheep TRB locus is known (fig. 1) [12], the formal interpretation of the D-J-C arrangements is possible. The intra-cluster rearrangements represent a consistent portion of the repertoire (41.6%), with 10 TRBD1-TRBJ1, 9 TRBD3-TRBJ3 and 6 TRBD2-TRBJ2 rearrangements (Table 1). A similar number of rearrangements (53.3%) can be interpreted by direct 5'- to- 3' joining across the clusters (inter-cluster rearrangements) with 20 TRBD1-TRBJ2, 6 TRBD1-TRBJ3 and 6 TRBD3-TRBJ2 rearrangements (Table 1). Interestingly, we also observed two TRBD2-TRBJ3 (psTMOs23 and psTA09, italics in Table 1) and one TRBD3-TRBJ1 (psSAR08, italics in Table 1) joining. Since the D- J-C cluster 2 is located downstream D- J-C cluster 3 as well as D- J-C cluster 3 is downstream D- J-C cluster 1 within the TRB locus, both these junctions may only be explained by chromosomal inversion, or with more probability, by trans-rearrangement occurring during TRB locus recombination.

A systematic analysis of the constant region of the transcripts also revealed that multiple splice variants are present. In fact, the canonical splicing is present in 49/72 (68%) clones with 10 TRBJ1-TRBC1, 17 TRBJ3-TRBC3 and 22 TRBJ2-TRBC2 transcripts (Table 1). A group of 7 clones (4 TRBJ1-TRBC3 and 3 TRBJ3-TRBC2) comes from an alternative or cis-splicing mechanism (9.7%). Finally, it is noteworthy that 16 clones (22.2%, bold in Table 1) with TRBJ2 genes showed TRBC3 or TRBC1 instead of the expected TRBC2 gene. Since TRBC3 as well as TRBC1 genes are located upstream TRBJ2 cluster in the germline DNA, TRBJ2 joined to TRBC1 or TRBC3 sequences cannot be a cis-spliced product of a single precursor RNA. Consequently, they must be the product of a trans-splicing between a transcript with TRBJ2-TRBC2 genes and a transcript containing TRBC1 or TRBC3 genes.

**a**

TRBD1 <u>gggacagcggggc</u>			
CLONE	CDR3 SEQUENCE (NT)	-nt J	J element
pSTMos.01	GCCAGCAGCCAGGAgggacagTAATATCCTATGAGCAGTAT	0	J2.7
pSTMos.04	GCCAGCAGCCACCGCCgggacagGTCATACTGAGGTT	-6	J1.1
pSTMos.05	GCCAGCAGCCAGGGAGgggacCTTAGCAATAACCCCTGTAT	0	J2.3
pSTMos.07	GCCAGCAGCAGTAGGTACAATGGAgggacagATCTATAATTGGCCCTCCAG	-3	J1.6
pSTMos.20	GCCAGCAGCCCAACCTCAGAggggATATTCCTACTATGAGCAGTAT	-3	J2.7
pSTMos.25	GCCAGCAGCCCAACTCgacagggAGGAGATCCCTTTAC	-7	J2.2
pSSAR.01	GCCAGCAGCCAGTCGAGAAggggGTCCTCCAGACCCAGTAC	-3	J2.5
pSSAR.02	GCCAGCAGCCAGGGACATAggacagcagAAGAAATGAAGACTCTAT	-3	J1.4
pSSAR.05	GCCAGCAGCAACTCGgggATATCCTTAATGAGCAGTAT	-4	J2.7
pSSAR.07	GCCAGCAGCAAGACTCCgggacagcagATGAAGACTCTAT	-4	J1.4
pSSAR.11	GCCAGCAGCCCAAgggggcGATATAACCCCTGTAT	-4	J2.3
pSSAR.16	GCCAGCAGCAAGAACAgacagTCGAGATTCAATAACCCCTGTAT	-2	J2.3
pSSAR.28	GCCAGCAGCCAAggcagGATCGGATTAACCCCTGTAT	-5	J2.3
pSSAR.31	GCCAGCAGCCAGATCCgacagggcCCGCGCAGCTGTAC	-6	J3.2
pSSAR.32	GCCAGCAGCAAGACATCTCCgagcgggCACAGATCCCTTTAC	-4	J2.2
pSMA.41	GCCAGCAGCCAGAGAggacagcggTGAAGCCGAC	-9	J3.1
pSMA.42	GCCAGCAGCCTgggacagcggAGCCCTATGAGCAGTAT	-3	J2.7
pSMA.59	GCCAGCAGCTCACTgagcAGGGGAGCTGCAC	-7	J2.1
pSMA.60	GCCAGCAGCCCGCGGTAggggGTATAACCCCTGTAT	-4	J2.3
pSMA.62	GCCAGCAGCCCTAgacaggggGAGATCCCTTTAC	-7	J2.2
pSMA.65	GCCAGCAGCAAGGCAAggacaggggCTAGCAATAACCCCTGTAT	0	J2.3
pSMA.68	GCCAGCAGCTggacCTGAACCCCGGCGAGCTGTAC	-6	J3.2
pSMA.70	GCCAGCAGCAGAGAgggacagggTTGTATGAGCAGTAT	-3	J2.7
pSMA.71	GCCAGCAGCCAggacagACGAACACTGAGGTTTTC	-3	J1.1
pSMA.73	GCCAGCAGCCGTTTTCgacagcGTCATGAGCCGAC	-7	J3.1
pSMA.74	GCCAGCAGCCgacatAGAACATTACAGACCCAGTAC	-3	J2.4
pSTA.01	GCCAGCAGCAGATcagcagggGGGATGAATATCAC	-2	J1.2
pSTA.02	GCCAGCAGCAAGACCTGGTgggacagcCGTTATGAATATCAC	0	J1.2
pSTA.06	GCCAGCAGCCAAAGATcgggATTATGAATATCAC	0	J1.2
pSTA.08	GCCAGCAGCCATCCgggACACTGAGGTTTTC	-4	J1.1
pSTA.11	GCCAGCAGCCCGgggacagcTCTCCAGACAGCAGTAC	-4	J2.4
pSTA.12	GCCAGCAGCCTAAGCgggacggggGATACTCAGACCCAGTAC	0	J2.5
pSTA.13	GCCAGCAGCAGAAAgcgggTCACTCAGAGACCGAGTAC	-2	J3.4
pSTA.25	GCCAGCAGCTTGGAAGagcagcAGATGAGCCGAC	-7	J3.1
pSTA.26	GCCAGCAGCCCAAGGagcagATAAAGGGGAGCTGCAC	-7	J2.1
pSTA.29	GCCAGCAGCCAGAGAgacagGACCCTTCGGTGAATATCAC	-3	J1.2

**b**

TRBD3 <u>gggctggggggggg</u>			
CLONE	CDR3 SEQUENCE	-nt J	J element
pSTMos.03	GCCAGCAGCCAGTCGGATgggAGCAGTAT	-7	J2.7
pSTMos.09	GCCAGCAGCCAAAGAGCCCGcgtggggCGAGATAC	-15	J3.3
pSTMos.13	GCCAGCAGCAGACCTgggacagggATGGGGAGCTGCAC	-5	J2.1
pSTMos.17	GCCAGCAGCCAAAGATgggggCCAGCCGCGCAGCTGTAC	-6	J3.2
pSTMos.18	GCCAGCAGCCAAAGAgggcctgACCATAACCCCTGTATT	-4	J2.3
pSTMos.21	GCCAGCAGCCGCTCAAGATggggACAGGATAGTGAGCGGTAT	-4	J3.5
pSTMos.22	GCCAGCAGCAGATgggACAGGATAGTGAGCGGTAT	-4	J3.5
pSSAR.04	GCCAGCAGCCAAgctggggggtggcCTTATCAGAGACCGCAGTAC	-2	J3.4
pSSAR.08	GCCAGCAGCCctgggTAATGAAGACTCTAT	-2	J1.4
pSSAR.25	GCCAGCAGCTTCTTTAGTActgggAGAGACCGCAGTAC	-4	J3.4
pSMA.10	GCCAGCAGCCAAAGAcctggCTTTAACCCCTGTAT	-5	J2.3
pSMA.66	GCCAGCAGCTCAGACCgggctggAGTCAGACCCAGTAC	-2	J2.5
pSMA.76	GCCAGCAGCCGctTAGCAATAACCCCTGTAT	0	J2.3
pSTA.03	GCCAGCAGCTTCTTTAGTActgggAGAGACCGCAGTAC	-4	J3.4
pSTA.04	GCCAGCAGCGATAgctgggACGTCCTCAAGCAGCTCAGTAC	-2	J3.3
pSTA.24	GCCAGCAGCAAGATCgctgggggGTCTCCTCAGAGACCGCAGTAC	-2	J3.4

**c**

TRBD2 <u>ggacttggggggggc</u>			
CLONE	CDR3 SEQUENCE (NT)	-nt J	J element
pSTMos.08	GCCAGCAGCCCAACGGAagacttggggggggcGTTATGAGCAGTAT	-3	J2.7
pSTMos.11	GCCAGCAGCAGAACTGCGCTCCGCTtgggggggAGGCTGTATGGGAGCTGCAC	-4	J2.1
pSTMos.23	GCCAGCAGCCAAAGCTCCCTATgggacttCTCAGAGACCGCAGTAC	-2	J3.4
pSSAR.10	GCCAGCAGCTGTGATtggcCTAATGGGAGCTGCAC	-5	J2.1
pSSAR.19	GCCAGCAGCCATCggacttgggATCGGCAATAACCCCTGTAT	-1	J2.3
pSMA.09	GCCAGCAGCCCTtgggggggGATATAACCCCTGTAT	-4	J2.3
pSMA.55	GCCAGCAGCAGacttTAGCAATAACCCCTGTAT	-1	J2.3
pSTA.09	GCCAGCAGCcttggGAGACCGCAGTAC	-5	J3.4

**d**

TRBD?			
CLONE	CDR3 SEQUENCE	-nt J	J element
pSTMos.02	GCCAGCAGCCCAACGAATATCGCGTATTCCTATGAGCAGTAT	0	J2.7
pSTMos.06	GCCAGCAGCCAGGGACCCAAATACAGATCCCTTTAC	-2	J2.2
pSTMos.12	GCCAGCAGCCCAACTCCGACCTATGATGAGCGGCAC	-2	J3.1
pSTMos.14	GCCAGCAGCCGGGAGCAATAACCCCTGTAT	0	J2.3
pSTMos.19	GCCAGCAGCCAGTGGAGCAACAGGCACAGCAC	0	J1.5
pSSAR.17	GCCAGCAGCCAGTCAGCAGGGATAGCAACCCAGGCACAGCAC	-1	J1.5
pSSAR.23	GCCAGCAGCCATCGACTGTCTGATAGCCAAAGCACTCAGTAC	0	J3.3
pSSAR.24	GCCAGCAGCCAAAGATCGGAACAGGGAGGCAATTCGCCCTCCAG	-7	J1.6
pSMA.46	GCCAGCAGCCAAAGACATAGAGACCGCAGTAC	-4	J3.4
pSMA.48	GCCAGCAGCCAAAGACTCACTTATGAGCAGTAT	-3	J2.7
pSMA.67	GCCAGCAGCAAGACGGATTCTATGAGCAGTAT	-2	J2.7
pSTA.15	GCCAGCAGCATAGGAAGGCTAACCCCTGTAT	-5	J2.3

**Figure 2**  
**CDR3 nucleotide sequences retrieved from the cDNA clones.** Sequences are shown from the codon after the *cys-94* of the *TRBV* gene to the codon before the *phe-104* of the *TRBJ* gene and grouped on the basis of the *TRBD1* (a), *TRBD3* (b), *TRBD2* (c) or no *TRBD* usage (d). Nucleotides that are retrieved in the 3' end of the V portion are considered of *TRBV* genomic origin and indicated in bold upper cases. Residues belonging to the different *TRBJ* genes, on the right, are indicated also in bold upper case at the 3' end of each sequence. The germline sequences of *TRBD1*, *TRBD3* and *TRBD2* gene are indicated at the top of each figure. The sequences considered to present recognizable *TRBD* genes (see text) are indicated in lower cases and nucleotide substitutions or insertions are underlined. Nucleotides that cannot be attributed to any coding elements (N-nucleotides) are indicated in capital letters on the left and on the right sides of the *TRBD* regions. Numbers in the right column indicate the level of 5'- *TRBJ* nucleotide trimming.

We excluded that all these non canonical sequences may be the result of PCR artifacts since the crossover points have not as expected a random distribution, but they always lie at the D-J or/and J-C junction, giving rise to products of the appropriate length and sequence.

The presence of splice variants may suggest the involvement of the *TRBC* gene in generating the TR  $\beta$ -chain functional diversity.

### Structure of the *TRBC* region

To complete the analysis of the TRBC domain in the cDNA clones, the whole constant portion of the transcripts was retrieved from the sequences and aligned according to the three *TRBC* isotypes for each animal in the different tissues.

The comparison of the 72 cDNAs showed the presence of different sequences that can be identified for the nucleotide variability in 14 different positions, 12 located in the first and two in the third exon, resulting in six amino acid substitutions all grouped in the first exon, and as a consequence, in the extracellular domain of the chain (fig. 3). By means of these variations, we observed a number of different sequences in excess. For example, five different groups of sequences were assigned to the *TRBC3* gene in the young thymus of the *Moscia Leccese* breed individual. This number is certainly higher than the expected two allelic forms, at the most, of the gene. In order to understand the origin of the additional sequences, we have isolated by PCR the allelic variants of all three *TRBC* genes from the young thymus genomic DNA of the *Moscia Leccese* individual, used as a reference model with respect to the others. The specificity of the PCR reactions was achieved by using a reverse primer which binds to either *TRBC1* and *TRBC3* (B40) or *TRBC2* 3'-UTR (B42) sequences, and completely TRBC specific forward primers complementary to a specific region upstream the *TRBC1* (CC1), *TRBC3* (CC3) and *TRBC2* (CC2) coding regions (see Methods). The three different PCR products were sequenced, and in every case, two allelic forms for each *TRBC* gene were obtained (data not shown). The comparison of the genomic with the corresponding sequences within the young thymus cDNAs allows us to establish that the first two more abundant groups of *TRBC3* sequences represent the two allelic forms of the *TRBC3* genes (pink and lilac in fig. 3), while alternative splicing of the third exon and DNA recombinational diversification process with the *TRBC2* gene can have generated the other three groups of *TRBC3* sequences (mixed color in fig. 3). Moreover, the two groups of *TRBC2* cDNA sequences (green and yellow in fig. 3) perfectly matched with the two allelic forms (data not shown). Only one allelic form was recovered for the *TRBC1* gene (italics in fig. 3), while the other *TRBC1* sequence can have been

generated by a mechanism of DNA recombinational diversification with the allele of *TRBC3* gene (mixed color).

After deducing the allelic variants of the three constant genes in the other tissues, alternative splicing and recombinational diversification can explain the excess of the sequences also in those cases.

### Discussion

To validate the real participation of the third additional D-J-C cluster and compare its usage with respect to the others in the formation of the TR  $\beta$ -chain repertoire, we analyzed transcripts of 72 unique D-J-C rearrangements recovered from four different tissues of four different animals, belonging to three different ovine breeds. Although the analyzed sequences lacked the *TRBV* genes, the presence of the CDR3  $\beta$  region, the *TRBJ* gene as well as most of the *TRBC* gene sequence was sufficient to permit a comprehensive analysis of the expressed TR  $\beta$  chain. Data presented here show that the mechanisms for generating diversity in sheep  $\beta$  chain polypeptides appear to adhere to the paradigms established through the study of humans and rodents. However, the diversity is enhanced by somatic rearrangement of 3 *TRBD* and 17 *TRBJ* genes that, by virtue of the expected recombination imprecision and N-region addition, maximizes diversity in the CDR3 region, thus expanding the potential repertoire of antigen specificities (Table 1). In spite of the presence of a longer coding nucleotide sequence in *TRBD* genes if compared with the human and mouse counterpart [12], the overall size of the CDR3 region is conserved in all tissues among the different mammalian species (Table 1). This conservation was archived by a greater deletion at the 5' end of *TRBJ* genes and a concomitant increase in N-nucleotide addition at the V-D-J junction during rearrangement (fig. 2). This suggests that the length of CDR3 in TR  $\beta$  chain is essential for T-cell function.

While there is not a specific influence of any cluster in the formation of the sheep  $\beta$ -chain in the different tissues, a dissimilar usage of the genes can be identified and it could depend on the sheep *TRB* genomic organization. Consistent with a promoter-enhancer facilitated recombination model, in human and mouse, assembly of the DJ $\beta$ 1 cassette is dependent on the interaction of the enhancer with the PD $\beta$ 1 promoter positioned immediately 5' of the *TRBD1* gene. Assembly of DJ $\beta$ 2 proceeds independent from that of DJ $\beta$ 1, albeit with less efficiency. Also in this case, an undefined PD $\beta$ 2 region continues to associate with the enhancer [21]. Our analyses suggest that also in sheep the mechanisms selectively alter D usage, so that the "privileged" *TRBD1* gene can account for the 60% of the total clones with respect to 26.6% of *TRBD3* and 13.3% of *TRBD2*. This may reside in the greater efficiency of the

		EXON I												EXON III			
Aminoacid		D	R/S	Q/R	T/M	D	T/A	Q/E	A/D	P	V	S	A	A	Y		
Codon number		1	4	5	68	71	85	106	115	117	118	123	124	137	160		
YOUNG THYMUS	CB3	4	GAC	AGA	CAG	ACG	GAT	ACC	CAG	GCC	CCT	GTC	AGT	GCT	GCA	TAC	
		4	---	---	---	-T-	--C	G--	---	-A-	---	---	--C	--C	--G	--C	
		1	---	---	---	---	---	---	---	---	---	---	---	---	--G	--T	
		2	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--T
		1	---	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	--G	--T	
	CB2	5	--T	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	--G	--T	
		1	--T	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	--G	---	
		2	---	--C	-G-	---	--C	---	G--	-A-	---	---	--C	--C	---	---	
		1	---	--C	-G-	---	---	---	---	---	---	---	--C	--C	---	---	
		1	---	--C	-G-	---	---	---	---	---	---	---	--C	--C	---	---	
BLOOD	CB3	3	---	---	---	---	---	---	---	---	---	---	---	---	---		
		1	---	---	---	---	---	---	---	--C	---	--C	--C	--G	--T		
		1	---	---	---	---	---	---	---	---	---	---	---	---	---	--T	
		2	---	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	--G	--T	
		3	--T	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	--G	--T	
	CB2	1	--T	---	---	-T-	--C	G--	---	-A-	---	---	--C	--C	--G	---	
		2	---	--C	-G-	---	--C	---	---	-A-	---	---	--C	--C	---	---	
		2	---	--C	-G-	---	--C	G--	---	---	---	---	--C	--C	--G	--T	
		1	---	--C	-G-	---	--C	---	---	-A-	---	---	--C	--C	---	--T	
		1	---	--C	-G-	-T-	--C	G--	---	-A-	---	---	--C	--C	--G	---	
SPLEEN	CB3	3	---	---	---	-T-	--C	G--	---	-A-	---	--G	--C	--C	--G	---	
		2	---	---	---	-T-	--C	G--	---	---	---	--C	--C	--G	--T		
		1	---	---	---	-T-	--C	G--	---	-A-	--C	---	--C	--C	--G	---	
		1	---	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	--G	--T	
		7	--T	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	--G	--T	
	CB2	1	--T	---	---	---	--C	G--	---	-A-	---	---	--C	--C	---	---	
		1	--T	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	---	---	
		1	--T	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	--G	---	
		1	--T	---	---	---	--C	G--	---	-A-	---	---	--C	--C	--G	---	
		1	--T	---	---	---	--C	G--	---	-A-	---	---	--C	--C	--G	---	
CB1	1	---	--C	-G-	---	--C	---	---	-A-	---	---	--C	--C	---	--T		
	1	---	--C	-G-	---	--C	---	---	-A-	---	---	--C	--C	---	--T		
	1	---	--C	-G-	---	--C	---	---	-A-	---	---	--C	--C	---	--T		
	1	---	--C	-G-	---	--C	---	---	-A-	---	---	--C	--C	---	--T		
	1	---	--C	-G-	---	--C	---	---	-A-	---	---	--C	--C	---	--T		
ADULT THYMUS	CB3	4	---	---	---	---	--C	---	---	-A-	---	---	--C	--C	---	---	
		2	---	---	---	---	--C	G--	---	---	---	---	--C	--C	--G	--T	
		1	---	---	---	---	--C	G--	---	---	---	---	--C	--C	---	---	
		1	---	---	---	---	--C	---	---	-A-	---	---	--C	--C	---	--T	
		1	---	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	--G	--T	
	CB2	4	--T	---	---	---	--C	G--	---	-A-	--C	---	--C	--C	--G	--T	
		1	---	--C	-G-	---	--C	---	---	-A-	---	---	--C	--C	---	---	
		1	---	--C	-G-	---	--C	---	---	-A-	---	---	--C	--C	---	---	
		1	---	--C	-G-	---	--C	---	---	-A-	---	---	--C	--C	---	---	
		1	---	--C	-G-	---	--C	G--	---	-A-	--C	---	--C	--C	--G	--T	

LEGEND

- TRBC1
- TRBC2
- TRBC3

Figure 3

The nucleotide sequences of the TRBC isotypes derived from the cDNA clones. Only the 14 variable nucleotide codons (12 in the first and two in the third exons numbered from the first position of the constant region in the cDNA) are depicted. The amino acids specified by the corresponding codons and those due to the nucleotide substitutions are given at the top of each codon, using the single letter code. The sequences are organized with respect to the one allelic TRBC3 sequence isolated from *Moscia Leccese* breed young thymus. Identities of the other allelic form of the same gene or of the other TRBC isotypes in the other tissues with respect to the reference sequence are indicated by dashes, while nucleotide substitutions are shown. The number on the left indicates the clones with the corresponding sequences. All the allelic forms of the TRBC isotypes are identified by a color. Color changes indicate recombinational diversification or alternative splicing.



PD $\beta$ 1 promoter activity with respect to the PD $\beta$ 3 or PD $\beta$ 2. A striking conservation of the PD $\beta$ 1 and PD $\beta$ 2 (as well as PD $\beta$ 3) regions among sheep, human and mouse [12] can support this observation, whereas the activity of the two similar PD $\beta$ 3 and PD $\beta$ 2 promoters could be correlated with their position from 5' to 3' within the locus.

The prominent utilization of the members of the *TRBJ*2 with respect to the *TRBJ*3 and *TRBJ*1 sets, as deduced from our cDNA collection, results from inter cluster or trans-rearrangements. It is possible that the preferential usage of the *TRBJ*2 set could depend on the number of genes that lie in the genomic region, if multiple J $\beta$  12-RSs are important for increasing the local concentration of the RAG proteins that first bind a 12-RS and then capture a 23-RS to form a synaptic complex [22]. In this regard, it is notable that the six sheep *TRBJ*1 genes lie in about 2.1 Kb, the five *TRBJ*3 genes in about 900 bp, while the seven *TRBJ*2 genes are grouped in about 1 Kb. Recently, Franchini et al. [23] have demonstrated, by means of an *in vitro* RAG1/2 mediated DNA coupled cleavage assay using various pair-wise RS combinations, that in mouse, the coupled cleavage of D $\beta$ 1-J $\beta$ 1 and D $\beta$ 2-J $\beta$ 2 substrates are similar and are both weak if compared to D $\beta$ 1-J $\beta$ 2 substrates, suggesting that J $\beta$ 2 RSs are better partners than J $\beta$ 1 RSs. In the same way, in sheep there could be the presence of a hierarchy efficiency of coupled cleavage with the D $\beta$ 1-J $\beta$ 2 > D $\beta$ 1-J $\beta$ 3 > D $\beta$ 1-J $\beta$ 1.

As the increment of the number of *TRBD* and *TRBJ* genes produces larger variation in TR  $\beta$  chain, particularly in CDR3 region as expected, similarly, the presence of an additional *TRBC* gene seems to affect the variety of the  $\beta$  chain repertoire. In fact careful analysis of the cDNA constant regions obtained from the different animals showed a level of unexpected variability in the first exon of the *TRBC* genes (fig. 3) if compared with that established in the genomic sequence [12]. By using the single nucleotide variations present in the first and third exon of the *TRBC* genes as hallmarks, we demonstrated that alternative splicing concerning the first and/or the third exon and/or somatic recombinatorial processes are involved in the diversification of the constant region of the sheep  $\beta$ -chain. The alternative splicing can occur either in cis or in trans. The presence of a cis-splicing mechanism comes from the analysis of six clones with *TRBJ*1-*TRBC*3 and *TRBJ*3-*TRBC*2 arrangement, while the presence of a trans-splicing process derives from the analysis of 16 clones with *TRBJ*2 spliced to *TRBC*3 or *TRBC*1 instead of the expected *TRBC*2 gene (Table 1). *TRBJ*2 to *TRBC*1 or *TRBC*3 splicing could be possible only when TRBV-TRBD-TRBJ transcripts are spliced with a transcript of the other allele. As a consequence, trans-splicing of two RNA separate precursors is the only logical possibility. The involvement of interallelic trans-splicing has already been documented in IgH

chains [24]. Beyond this case the presence of interallelic trans-splicing in vertebrates is problematical to demonstrate. It has been documented to be an essential process for the expression of the *lola* Drosophila gene. *Lola* encodes 20 protein isoforms belonging to a family of BTB zinc-finger transcriptional factor [25]. Genetic tests have demonstrated that some isoforms were generating thought intrallelic trans-splicing [26]. No particular sequences for trans-splicing have been identified around the exon-intron boundary in the *lola* gene; therefore, the basic mechanism of trans-splicing is likely to be shared with those of cis-splicing and occur co-transcriptionally where nascent pre mRNA are produced in close proximity, as is the case for cis-splicing [26]. It is possible that also in sheep *TRB* locus, the cis and trans-splicing shared the same mechanism.

Investigation of the constant domain of the sheep cDNAs led us to deduce that a minimal set of sequences are also generated by a somatic recombinatorial process (fig. 3). Somatic recombinatorial diversification occurs in vertebrates, yeast and plants [27-29], and such a modification of germline sequences can generate individuals with different starting gene repertoires in different tissues.

The precise effect and significance of the variability in the constant region of TR  $\beta$ -chain remain to be determined. It might create diversity in the T cell function. The extracellular domain of the TRBC molecule consists of well-defined regions [15]. The pattern of amino acid replacements in the sheep cDNA was located, beyond the N-terminus, one both in the TRBC E  $\beta$ -strand and in the DE loop and two in the FG loop. This last is TR  $\beta$ -chain specific loop in all mammalian species and contains 12 residues that are conserved between the two TRBC isotypes in human and mouse. In sheep sequences, the FG loop is one amino acid longer and underwent replacement among the three *TRBC* genes. So the Gln in position 106 in the first half part of the loop can be replaced by Glu; while the Asp in position 115 of the second part of the loop can be substituted by Ala (fig. 3). Three-dimensional structures of the TR [30] have shown that the FG loop of the TR  $\beta$  chain exists as an elongated, rigid element forming a sidewall of a cavity created by the asymmetric disposition of C $\alpha$  and C $\beta$  domains that receive the  $\epsilon$  subunit of the CD3 complex [31]. Therefore a primary function of the C $\beta$  FG loop in the thymus is to facilitate negative selection, while following maturation,  $\alpha\beta$  T cells are dependent on the C $\beta$  FG loop to their activation [32]. Our hypothesis is that amino acid replacement in the FG loop of the sheep *TRBC* genes can be modified by the sensitivity of  $\alpha\beta$  T cell for cognate peptide recognition, and this can be correlated with the function of the  $\alpha\beta$  T cell in sheep.

## Conclusions

All together our results show that in sheep the presence of an additional D-J-C cluster enhances the  $\beta$ -chain repertoire. These findings, together with the evidence of the expansion of gene repertoires for other TR loci in ruminants [33-35], suggest that strong evolutionary pressures have driven a generic enlargement of TR gene numbers, thus generating a greater potential TR diversity in this lineage.

## Methods

### Animals (source of tissue)

Thymus, spleen and blood were obtained from animals of three different autochthonous breeds. One thymus was collected from one neonatal *Moscia Leccese* sheep; spleen and the other thymus from one adult *Gentile di Puglia*; and blood from one young *Sarda Ionica* sheep. All animals were conventionally reared outbred sheep and were healthy at the time of sample collection. All animal manipulations were carried out with the approval of the Bari Animal Ethics Committee and in compliance with Institutional Animal Care and Use Committee (IACUC) requirements.

### RT-PCR

The different organs were removed from the animals, immediately frozen in liquid nitrogen and stored at  $-70^{\circ}\text{C}$  until preparation of RNA. In the case of blood, RNA was prepared before freezing.

Total RNA was extracted from tissues under the protocol approved by the manufacturer (Trizol reagent, Invitrogen). First-strand cDNA synthesis was performed by reverse transcription of 5  $\mu\text{g}$  of total RNA primed with 2.5  $\mu\text{l}$  of oligodT (0.5  $\mu\text{g}/\mu\text{l}$ ) using 2  $\mu\text{l}$  dNTP (10 mM), 2  $\mu\text{l}$  DTT (100 mM) and 1  $\mu\text{l}$  PowerScript<sup>™</sup> Reverse Transcriptase (Clontech) in the recommended buffer in a total of 20  $\mu\text{l}$ . The genes of interest were amplified from 10% of cDNA preparations using a sense V primer (VB3; 5'-TATCTCTGTGCCAGCAGC-3') complementary to a conserved region in the 3'-end of sheep *TRBV* genes [16] and an antisense CB3 primer (5'-CACCAGGGCGCTGACCAG-3'; AM420900; 8,222-8,239 positions) (5'-CACCAGGGCGCTGACCAG-3'; AM420900; 17,442-17,459 positions) (5'-CACCAGGGCGCTGACCAG-3'; AM420900; 26,702-26,719 positions) located in the third exon of the sheep *TRBC* genes. All the PCR were performed in a 50  $\mu\text{l}$  volume with 5  $\mu\text{l}$  10 $\times$  buffer, 2  $\mu\text{l}$  MgCl<sub>2</sub> (50 mM), 1  $\mu\text{l}$  dNTP (10 mM), 0.5  $\mu\text{l}$  of Taq Platinum 5 U/ $\mu\text{l}$  (Invitrogen) and 2  $\mu\text{l}$  of the sense and antisense primers (10 mM). After 2 min of initial denaturation at  $94^{\circ}\text{C}$ , the samples were subjected to 35 cycles of amplification (30 s at  $94^{\circ}\text{C}$ , 30 s at  $58^{\circ}\text{C}$ , 30 s at  $72^{\circ}\text{C}$ ). The final cycle was extended for 10 min at  $72^{\circ}\text{C}$ . Amplified cDNA fragments were puri-

fied by using the PureLink PCR Purification Kit (Invitrogen-Life Technologies), ligated into StrataClone PCR Cloning Vector and transformed into StrataClone Competent Cells (Stratagene).

### DNA amplification

Genomic DNA was isolated from the young *Moscia Leccese* thymus by standard techniques. For the DNA amplifications, 50-200  $\mu\text{g}$  of thymus DNA was used with the TaKaRa LA Taq in 50  $\mu\text{l}$  reactions, according to the recommendations (TAKARA BIO INC.). The cycling conditions were as follows:  $94^{\circ}\text{C}$  for 1 min; 35 cycles of 30 s denaturation at  $95^{\circ}\text{C}$ , 1 min annealing at  $58^{\circ}\text{C}$ , 2 min polymerization at  $68^{\circ}\text{C}$ ; and  $68^{\circ}\text{C}$  for 10 min. The primer combinations used were CC1 and B40 for the *TRBC1* gene, CC3 and B40 for the *TRBC3* gene and CC2 and B42 for the *TRBC2* gene. The CC1 (5'-CTGTGGCCCCCTTTCCTTGTT-3'; AM420900, 6,805-6,824 positions), CC3 (5'-ACACACACAGCCCCTACCA-3'; AM420900, 16,324-16,342 positions) and CC2 (5'-AGAGATGGGTTGTCGTAGG-3'; AM420900, 25,117-25,136 positions) are designed on the 5'- end specific of the *TRBC1*, *TRBC3* and *TRBC2* genes respectively. B40 (5'-TCAGGGCAGTAACAGGCT-3'; AM420900, 8587-8569 positions) (5'-TCAGGGCAGTAACAGGCT-3'; AM420900; 17832-17815 positions) is complementary to the 3'UTR of *TRBC1* as well as *TRBC3* genes, while B42 (5'-ATGACTCGGGACGCACCT-3'; AM420900, 27,040-27,057 positions) is complementary to the 3'UTR of the *TRBC2* genes. Amplified DNA fragments were purified by PureLink PCR Purification Kit (Invitrogen-Life Technologies) and used directly for DNA sequencing.

### Determination of CDR3 length and sequence analyses

The CDR3 size was calculated by the number of amino acids between the amino acid after the conserved 2<sup>nd</sup> cysteine in the V gene (pos.104), and the amino acid before the phenylalanina of the FGXG motif in the J gene [http://imgt.cines.fr/\[17\]](http://imgt.cines.fr/[17]). This method gives the CDR3 length with three amino acids more than that done in Kabat et al [36].

Nucleotide sequences were determined by a commercial service. DNA sequence data were processed and analyzed using the blasta program <http://www.ncbi.nlm.nih.gov/BLAST>, Clustal W <http://www.ebi.ac.uk/clustalw/index.html>[37] and IMGT database <http://imgt.cines.fr/> [17].

### Abbreviations

TR: T cell receptor; TRB: T cell receptor beta; TRBV: T cell receptor beta variable gene; TRBJ: T cell receptor beta joining gene; TRBD: T cell receptor beta diversity gene; TRBC: T cell receptor beta constant gene.

## Authors' contributions

SDT carried out the molecular studies and analyzed data. RA participated in the design of the study and helped to draft the manuscript. SC revised the manuscript critically. SM conceived of the study, participated in its design and coordination and wrote the manuscript. All authors have read and approved the final manuscript.

## Acknowledgements

This work was supported by grants from the University of Bari and University of Salento (Progetti di Ricerca Scientifica d'Ateneo).

## References

- Jung D, Alt FW: **Unraveling V(D)J recombination; insights into gene regulation.** *Cell* 2004, **116**:299-311.
- Bassing CH, Swat W, Alt FW: **The mechanism and regulation of chromosomal V(D)J recombination.** *Cell* 2002, **109(Suppl)**:S45-S55.
- Tonegawa S: **Somatic generation of antibody diversity.** *Nature* 1983, **302**:575-581.
- Bassing CH, Alt FW, Hughes MM, D'Auteuil M, Wehrly TD, Woodman BB, Gartner F, White JM, Davidson L, Sleckman BP: **Recombination signal sequences restrict chromosomal V(D)J recombination beyond the 12/23 rule.** *Nature* 2000, **405**:583-586.
- Rowen L, Koop BF, Hood L: **The complete 685-kilobase DNA sequence of the human beta T cell receptor locus.** *Science* 1996, **272**:1755-1762.
- Malissen M, Minard K, Mjolsness S, Kronenberg M, Goverman J, Hunkapillar T, Prystowsky MB, Yoshikai Y, Fitch F, Mak TW, Hood L: **Mouse T cell antigen receptor: Structure and organization of constant and joining gene segments encoding the beta polypeptide.** *Cell* 1984, **37**:1101-1110.
- Siu G, Kronenberg M, Strauss E, Haars R, Mak TW, Hood L: **The structure, rearrangement and expression of D beta gene segments of the murine T-cell antigen receptor.** *Nature* 1984, **311**:344-350.
- Williams CB, Blankenhorn EP, Byrd KE, Levinson G, Gutman GA: **Organization and nucleotide sequence of the rat T cell receptor beta-chain complex.** *J Immunol* 1991, **146**:4406-4413.
- Jaeger EE, Bontrop RE, Lanchbury JS: **Nucleotide sequences, polymorphism and gene deletion of T cell receptor beta-chain constant regions of *Pan troglodytes* and *Macaca mulatta*.** *J Immunol* 1993, **151**:5301-5309.
- Cheyrier R, Henrichswark S, Wain-Hobson S: **Sequence of the Rhesus monkey T-cell receptor beta chain diversity and joining loci.** *Immunogenetics* 1996, **43**:83-87.
- Schrenzel MD, Watson JL, Ferrick DA: **Characterization of horse (*Equus caballus*) T-cell receptor beta chain genes.** *Immunogenetics* 1994, **40**:135-144.
- Antonacci R, Di Tommaso S, Lanave C, Cribiu EP, Ciccarese S, Massari S: **Organization, structure and evolution of 41 kb of genomic DNA spanning the D-J-C region of the sheep TRB locus.** *Mol Immunol* 2008, **45**:493-509.
- Connelley T, Aerts J, Law A, Morrison WL: **Genomic analysis reveals extensive gene duplication within the bovine TRB locus.** *BMC Genomics* 2009, **10**:192.
- Eguchi-Ogawa T, Toki D, Uenishi H: **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* 2009, **33**:1111-1119.
- Lefranc MP, Pommie C, Kaas Q, Duprat E, Bosc N, Guiraudou D, Jean A C, Ruiz M, Da Piedade I, Rouard M, Foulquier E, Thouvenin V, Lefranc G: **IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains.** *Dev Comp Immunol* 2005, **29**:185-203.
- Halsey WA Jr, Palmer BE, DeMartini JC, Howell MD: **Analysis of sheep T-cell receptor beta-chain heterogeneity.** *Immunogenetics* 1999, **49**:206-214.
- Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G: **IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains.** *Dev Comp Immunol* 2003, **27**:55-77.
- Hall MA, Lanchbury JS: **Healthy human T-cell receptor beta-chain repertoire. Quantitative analysis and evidence for J beta-related effects on CDR3 structure and diversity.** *Hum Immunol* 1995, **43**:207-18.
- Rock EP, Sibbald PR, Davis MM, Chien YH: **CDR3 length in antigen-specific immune receptors.** *J Exp Med* 1994, **179**:323-328.
- Watanabe M, Iwasaki Y, Mita Y, Ota S, Yamada S, Shimizu M, Takagaki Y: **Porcine T-cell receptor beta-chain: a genomic sequence covering Db1.1 to Cbeta2 gene segments and the diversity of cDNA expressed in piglets including novel alternative splicing products.** *Mol Immunol* 2007, **44**:2332-2343.
- Oestreich KJ, Cobb RM, Pierce S, Chen J, Ferrier P, Oltz EM: **Regulation of TCRbeta gene assembly by a promoter/enhancer holocomplex.** *Immunity* 2006, **24**:381-391.
- Curry JD, Geier JK, Schlissel MS: **Single-strand recombination signal sequence nicks in vivo: evidence for a capture model of synapsis.** *Nat Immunol* 2005, **6**:1272-1279.
- Franchini DM, Benoukraf T, Jaeger S, Ferrier P, Payet-Bornet D: **Initiation of V(D)J recombination by Dbeta-associated recombination signal sequences: a critical control point in TCRbeta gene assembly.** *PLoS ONE* 2009, **4**:e4575.
- Shimizu A, Honjo T: **Synthesis and regulation of trans-mRNA encoding the immunoglobulin epsilon heavy chain.** *FASEB J* 1993, **7**:149-154.
- Ohsako T, Horiuchi T, Matsuo T, Komaya S, Aigaki T: ***Drosophila lola* encodes a family of BTB-transcription regulators with highly variable C-terminal domains containing zinc finger motifs.** *Gene* 2003, **311**:59-69.
- Horiuchi T, Giniger E, Aigaki T: **Alternative trans-splicing of constant and variable exons of a *Drosophila* axon guidance gene, *lola*.** *Genes Dev* 2003, **17**:2496-2501.
- Chamnanpant J, Shan WX, Tyler BM: **High frequency mitotic gene conversion in genetic hybrids of the oomycete *Phytophthora sojae*.** *Proc Natl Acad Sci USA* 2001, **98**:14530-14535.
- Lucht JM, Mauch-Mani B, Steiner HY, Metraux JP, Ryals J, Hohn B: **Pathogen stress increases somatic recombination frequency in *Arabidopsis*.** *Nat Genet* 2002, **30**:311-314.
- Diaz M, Flajnik MF: **Evolution of somatic hypermutation and gene conversion in adaptive immunity.** *Immunol Rev* 1998, **162**:13-24.
- Wang J, Lim K, Smolyar A, Teng M, Liu J, Tse AG, Liu J, Hussey RE, Chishti Y, Thomson CT, Sweet RM, Nathenson SG, Chang HC, Sacchetti JC, Reinherz EL: **Atomic structure of an alphabeta T cell receptor (TCR) heterodimer in complex with an anti-TCR Fab fragment derived from a mitogenic antibody.** *EMBO J* 1998, **17**:10-26.
- Ghendler Y, Smolyar A, Chang HC, Reinherz EL: **One of the CD3epsilon subunits within a T cell receptor complex lies in close proximity to the Cbeta FG loop.** *J Exp Med* 1998, **187**:1529-1536.
- Touma M, Chang HC, Sasada T, Handley M, Clayton LK, Reinherz EL: **The TCR C beta FG loop regulates alpha beta T cell development.** *J Immunol* 2006, **176**:6812-6823.
- Miccoli MC, Antonacci R, Vaccarelli G, Lanave C, Massari S, Cribiu EP, Ciccarese S: **Evolution of TRG clusters in cattle and sheep genomes as drawn from the structural analysis of the ovine TRG2@ locus.** *J Mol Evol* 2003, **57**:52-62.
- Conrad ML, Mawer MA, Lefranc MP, McKinnell L, Whitehead J, Davis SK, Pettman R, Koop BF: **The genomic sequence of the bovine T cell receptor gamma TRG loci and localization of the TRGC5 cassette.** *Vet Immunol Immunopathol* 2007, **115**:346-356.
- Vaccarelli G, Miccoli MC, Lanave C, Massari S, Cribiu EP, Ciccarese S: **Genomic organization of the sheep TRG1@ locus and comparative analyses of Bovidae and human variable genes.** *Gene* 2005, **357**:103-114.
- Kabat A, Wu TT: **Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites.** *J Immunol* 1991, **147**:1709-1719.
- Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucl Acids Res* 1994, **22**:4673-4680.