Available online at www.sciencedirect.com
ScienceDirect
Molecular Immunology 45 (2008) 493–509
Molecular Immunology
www.elsevier.com/locate/molimm

Organization, structure and evolution of 41 kb of genomic DNA spanning the D-J-C region of the sheep TRB locus

R. Antonacci [a,*], S. Di Tommaso [b], C. Lanave [c], E.P. Cribiu [d], S. Ciccarese [a], S. Massari [b]

[a] *Dipartimento di Genetica e Microbiologia, Universita' degli Studi di Bari, Italy*
[b] *Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento, Lecce, Italy*
[c] *Istituto di Tecnologie Biomediche Sezione di Bari, CNR, Milano, Italy*
[d] *Departement de Genetique Animale, Centre de Recherche INRA de Jouy-en-Josas, France*

Received 30 March 2007; accepted 21 May 2007
Available online 27 July 2007

Abstract

A genomic region of 41,045 bp encompassing the 3′-end of the sheep T cell receptor beta chain was sequenced. Extensive molecular analysis has revealed that this region retains a unique structural feature for the presence of a third D-J-C cluster, never detected in any other mammalian species examined so far. A total of 3 TRBD, 18 TRBJ and 3 substantially identical TRBC genes were identified in about 28 kb. At 13 kb, downstream from the last TRBC gene, in an inverted transcriptional orientation, lies a TRBV gene. Sequence comparison and phylogenetic analyses have demonstrated that the extra D-J-C cluster originated from an unequal crossing over between the two ancestral TRBC genes. Interspersed repeats spanning 22.2% of the sequence, contribute to the wider size of the sheep TRB locus with respect to the other mammalian counterparts, without modifying the general genomic architecture. The nucleotide and predicted amino acid sequences from peripheral T cells cDNA clones indicated that the genes from cluster 3 are fully implicated in the beta chain recombination machinery. Closer inspections of the transcripts have also shown that inter-cluster rearrangements and splice variants, involving the additional cluster, increase the functional diversity of the sheep beta chain repertoire.

Keywords: T-cell receptor; β-Chain; TRB locus; D-J-C-region; Sheep

1. Introduction

T- cell mediated immunity is based on antigen recognition by a disulphide-linked heterodimer T-cell receptor (TR) that may occur in two types: αβ or γδ. The diversity displayed by the α and β (or γ and δ) chains that make up the heterodimer, is determined by multiple variable (V), diversity (D, β- and δ-chain) and joining (J) genes that associate more or less randomly by somatic recombination during T-cell differentiation to generate a variable exon. The process of somatic recombination is understood to involve the site-specific cleavage of DNA by RAG-1 and RAG-2 proteins at conserved recombination signal sequences (RSSs), flanking the borders of the V, D and J genes. RSSs are composed of conserved heptamers and nonamers with intervening non-conserved 12- or 23-bp spacers. Recombination only occurs between variable gene segments flanked by RSSs of dissimilar spacer lengths (Tonegawa, 1983). This restriction is known as the "12/23 rule". However, at the locus encoding for β chain (TRB), despite the 12/23 compatibility, V genes do not rearrange directly to the Js (skipping over the D segment), a phenomenon termed "beyond 12/23 rule" (Sleckman et al., 2000; Bassing et al., 2000).

After transcription, the V(D)J sequence is spliced to the constant (C) gene segment. The resulting chain is a protein with the variable domain composed of three complementary-determining regions (CDRs). Two of these CDRs, CDR1 and CDR2, are encoded by the germline V gene sequence. The third CDR domain, CDR3, reflects the ability of the V gene to rearrange to any (D)J gene. This implies that the number of V, D and J genes

* Corresponding author at: Department of Genetics and Microbiology, University of Bari, Via Amendola 165/A, 70126 Bari, Italy. Tel.: +39 080 5443338; fax: +39 080 5443386.
E-mail address: r.antonacci@biologia.uniba.it (R. Antonacci).

0161-5890/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.molimm.2007.05.023

in the germline DNA is an important determinant of the extent of the TR repertoire. Moreover, the diversity of CDR3 is further increased by additional mechanisms that involve: imprecise joining of the gene segments resulting in small deletions from the coding ends; the templated (P region diversification) and non-templated (N region diversification) addition of nucleotides at the junctions of the gene segments. The αβ T cells recognize processed immunogenic peptides presented to them by MHC molecules. Structural studies (Garcia et al., 1999) and the proposed two-step model of αβ TR antigen recognition (Wu et al., 2003) are consistent with the idea of separable MHC and peptide recognition domains within the TR, with the V(D)J junctions (CDR3) principally involved in peptide recognition, whereas CDR1 and CDR2 predominantly form contacts with the α-helices bordering the peptide-binding groove of MHC molecules (Kjer-Nielsen et al., 2003).

αβ TR has been identified in all major phylogenetic groups of jawed vertebrates and the genomic organization of the TRB locus has been determined in representative species of several orders of mammals as well as in bony fish (De Guerra and Charlemagne, 1997; Zhou et al., 2003) and chicken (Shigeta et al., 2004).

A common feature of the genomic structure of TRB locus in mammalian species (i.e. human and mouse) (IMGT database; Lefranc et al., 2003) is the presence of a semi-cluster organization with a pool of TRBV genes positioned upstream from two clusters of genes, each made of one TRBD, several TRBJ and a single TRBC genes (D-J-C cluster). At the 3′-end of the last TRBC gene is a single inverted TRBV gene. Few differences in the number of TRBV and TRBJ genes between the two species can be observed. There are 12 and 13 functional TRBJ genes divided between the two D-J-C clusters of human and mouse, respectively. In human, 32 subgroups of TRBV genes have been defined which contain a total of 47 functional genes. In mouse 25 subgroups with 22 functional TRBV genes have been identified. A peculiar feature of the mammalian TRB locus is also 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.

αβ T cells have demonstrated to be the most important T cells also in artiodactyl species (pig, cattle and sheep) despite their high level of γδ T cells (Binns et al., 1992; Hein and MacKay, 1991), even if little information about the genomic organization of TRB locus are available in these species. In the D-J-C region of porcine TRB locus, 2 TRBD genes and 12 functional TRBJ genes divided between two D-J-C clusters have been identified (Watanabe et al., 2007). Moreover, analyses of transcripts of more than 300 unique V-D-J rearrangements recovered from porcine thymocytes and peripheral T cells, have demonstrated the presence of 19 subgroups of TRBV genes (Butler et al., 2005). Expression studies in cattle (Houston et al., 2005) and sheep (Halsey et al., 1999) identified more elevated numbers of TRBD and TRBJ genes. Sixteen TRBV subgroups, eight of them were found to contain more than one member, form the basis for the TRB repertoire in cattle. In addition, Southern blot analysis in sheep (Grossberger et al., 1993) and the sequence analysis of a bovine BAC clone corresponding to the D-J-C region (Conrad et al., 2002) suggest the presence of three TRBC genes in ruminants.

In order to evaluate the exact number of TRBD, TRBJ and TRBC genes in ruminant lineage, we started to analyze appropriate genomic clones in sheep. In this report, we describe the organization, structure and evolution of 41 kb of genomic DNA encompassing the 3′-end of the sheep TRB locus. The sequence revealed for the first time that the sheep TRB locus contains, unlike the other mammalian species, an additional D-J-C cluster, comprised between D-J-C cluster 1 and 2, having five TRBJ genes with an upstream TRBD and a downstream TRBC gene.

Our phylogenetic analysis, carried out on the J and C genes, established the tight relationship between sheep D-J-C cluster 3 and D-J-C cluster 2 as a consequence of a duplication event. Moreover, the comparative analysis of the genomic structure of the entire TRB region among sheep, human, mouse and pig revealed that the presence of reiterated D-J-C clusters and the large quantity of interspersed repeats make the sheep region wider in size with respect to the other species. However, the general regulatory architecture of the sheep TRB region seems to be preserved. Finally, since RT-PCR based assays have demonstrated that sheep D-J-C cluster 3 represents a fully functional cluster, we discuss the hypothesis that the presence of an additional group of D and J genes could represent a specific advantage for the species to generate a diverse repertoire of TR specificities.

2. Materials and methods

2.1. Screening of genomic libraries

The phage genomic library was constructed in λDASHII (Stratagene) using DNA obtained from sheep (Altamurana breed) lung. The screening of this genomic library and the isolation of λB1 and λB2 clones were previously described (Antonacci et al., 2001). The BAC ovine library, used in this study, has been constructed by Vaiman et al. (1999) using DNA from ram (Romanov breed) brain. In this library, the clones have been arrayed in superpools for rapid PCR-based screening. The library was thus screened with a primer set (B1 and B2 in Table 1), complementary to the first exon of TRBC1 gene according to the protocol suggested by the manufacturers. One positive clone, denominated G12, was recovered.

2.2. Southern blot analysis

λB1, λB2 and G12 genomic clones were characterized by Southern blot analysis. Phage DNA was digested using the restriction enzymes BamHI, XbaI, EcoRI and HindIII in single and double digestions. Digested DNA was electrophoresed through 1% agarose gel and transferred by capillary action to Hybond-N charged nylon membranes for hybridization. Hybridizations were performed at 42 °C in 50% formamide. Final wash conditions were 65 °C in 2 × SSC. The probe was prepared from a human truncated cDNA containing the TRBJ1-TRBC1 sequence and radioactively labeled.

Table 1
PCR primers

Primer sets	Sequence 5′-3′	Accession no.	Primer location
B1 forward	AGAGATCTCCCGGACCCA	AJ309004	114–131
B2 reverse	TGGTTGCGGGGGTTGTGC		335–318
F forward	CCAATTCTCATGTTTGACAGCT		651–672
FR reverse	CGGATCTCTACGATAATGGGA	pBeloBAC11	2,131–2,112
R forward	AGTTATTAGCGATGAGCTCGGA		7,472–7,451
RR reverse	CTGCCCCGATGGTGGATTCGT	pBeloBAC11	5,350–5,370
B1 forward	AGAGATCTCCCGGACCCA	AM420900	6,961–6,978
B9 reverse	GTGCCGCTCATCATAGGA	AM420900	11,849–11,832
VB1 forward	TTGTTACCTCCCCTTCCTCA	AM420900	34,799–34,818
VB2 reverse	ATGCCCTGCTGTCTGCTG	AF035472	1–18
B7 forward	CGGTGTTGGCTTCTCCCTTC	AM420900	15,668–15,687
J13 reverse	GCCTGGGCCGAAATACTG	AM420900	22,720–22,702
CB40 reverse	TCAGGGCAGTAACAGGCT[a]	AM420900	8,587–8,569
VB3 forward	TATCTCTGTGCCAGCAGC[b]	AF030005	322–339
CB3 reverse	CACCAGGGCGCTGACCAG[c]	AM420900	8,239–8,222

[a] This primer sequence is also recognizable at position 17,832–17,815.
[b] Primer sequence detectable in different cDNA clones as reported by Halsey et al. (1999).
[c] This primer sequence is also recognizable at position 17,442–17,459 and 26,719–26,702.

For G12 clone analysis, restriction-digested XhoI and NotI/XhoI BAC DNA was subjected to contour clamped homogeneous electric field (CHEF) electrophoresis in 1.2% agarose gel for 11 h at 14 °C, 6 V/cm, 120° with ramped switching time of 2 s. This separated restriction fragments ranging in size from 5 to 75 kb. The gel was transferred to Hybond-N charged nylon membrane, which was then hybridized by using the DIG High Prime DNA Labeling and Detection Kit (Roche Diagnostic GmbH) to various probes. Hybridizations were performed at 65 °C and membrane was washed at high stringency (65 °C with 0.2 × SSC and 0.1% SDS) and subjected to autoradiography.

The probes used for mapping the T7 and SP6 ends of BAC clone were generated by PCR of the pBeloBAC11 vector DNA, with the primer pairs F/FR and R/RR (see Table 1). Approximately 50 ng of BAC DNA were used in each amplification reaction in a 50 μl final volume with primers at 0.05 μM and 0.5 μl of Taq polymerase (Fermentas). The thermal cycling protocol was: 5 min at 95 °C, 30 s at 95 °C, 40 s at 58 °C, 40 s at 68 °C for 30 cycles. The final cycle was extended for 10 min at 68 °C. The amplification produces a 1.4 and 2.1 kb fragment, respectively.

The Jb1 probe, used for TRBJ1 gene mapping, was prepared as DNA fragment excised from the pzB3.07 plasmid; while Cb probe, for TRBC gene mapping, was obtained using the B1/B2 primer pair (Table 1) on phage DNA in a standard PCR.

2.3. Subcloning of G12 BAC clone

The G12 BAC clone was grown overnight in 100 ml LB medium with 25 mgr/ml of chloramphenicol at 37 °C by shaking. The BAC clone DNA was extracted and purified using NucleoBond BAC 100 Plasmid DNA Purification (Macherey-Nagel). The extracted DNA was digested with BamHI or HindIII restriction enzymes, ligated into pZero cloning vector (Invitrogen-Life Technologies) and transformed into TOP10 electrocompetent cells. The plasmid library was then screened with the B1/B2 primer set (Table 1). The positive pzSD1.7, pzSB6.4 and pzSC2.35 clones were isolated and characterized.

2.4. Long-PCR

For the amplifications, 50–200 ng of purified G12 BAC DNA was used with the TaKaRa LA Taq in 50 μl reactions, according to the recommendations (TAKARA BIO INC.). The cycling conditions were as follows: 94 °C for 1 min; 30 cycles of 30 s denaturation at 95 °C, 2 min annealing at 58 °C, and 15 min polymerization at 68 °C; 72 °C for 10 min. Primer combinations used were B1 and B9 for the pXL4 clone, B7 and J13 for the pXLB4J13 clone, and VB1 and VB2 for the pXLVB clone. The nucleotide sequences and the positions of all the primers are listed in Table 1. Amplified DNA fragments were purified using PureLink PCR Purification Kit (Invitrogen-Life Technologies), ligated into pCR-XL-TOPO vector (Invitrogen-Life Technologies) and transformed into TOP10 competent cells.

2.5. Drawing of dendrograms

For the phylogenetic analysis we designed the various genes from different organisms by the gene notation as proposed by IMGT (Lefranc et al., 2003) plus the species' Latin names.

The GenBank accession numbers of the TRBC genes used are the following:

- *Ambystoma mexicanus* (axolotl): TRBC1 (X70168), TRBC2 (L08498), TRBC3 (AF334472), TRBC4 (AF324473).
- *Bos taurus* (cattle): TRBC1, TRBC2 and TRBC3 (AF453325).

- *Canis familiaris* (dog): TRBC (M97510).
- *Equus caballus* (horse): TRBC1 (L27845), TRBC2 (L27844).
- *Gadus morhua* (Atlantic cod): TRBC1 (AJ133848), TRBC2 (AJ133850), TRBC3 (AJ133849), TRBC4 (AJ133851).
- *Gallus gallus* (chicken): TRBC (M37802).
- *Heterodontus francisci* (Horn shark): TRBC (U07624).
- *Homo sapiens* (human): TRBC1 (M12887), TRBC2 (M12886).
- *Ictalurus punctatus* (Channel catfish): TRBC1 (U39193), TRBC2 (U58508).
- *Mus musculus* (mouse): TRBC1 (M11456), TRBC2 (K02803).
- *Oryctolagus cuniculus* (rabbit): TRBC1 (M12889), TRBC2 (M13895).
- *Ovis aries* (sheep): TRBC1, TRBC2, TRBC3 (AM420900).
- *Raja eglanteria* (skate): TRBC (U75769).
- *Rat norvegicus* (rat): TRBC1 (M63793), TRBC2 (M63795).
- *Salmon salar* (salmon): TRBC (X97435).
- *Salmon trutta* (trout): TRBC (U18122).
- *Stegastes partitus* (bicolour damselfish): TRBC1 (AF324813), TRBC2 (AF324823).
- *Sus scrofa* (pig): TRBC1 and TRBC2 (AB079894).
- *Xenopus laevis* (African clawed frog): TRBC (U60424).

For the phylogenetic analysis of TRBC genes within Vertebrates, we used both the amino acid and the nucleotide sequences of the coding regions and the nucleotide sequences of the 3′-UTR portions (from the stop codon to the poly-A site) when they were complete.

Multiple alignments of the sequences under analysis were carried out with the Clustal W program (Thompson et al., 1994) and manual adjustments were made with the SEAVIEW program (Galtier et al., 1996) and the GeneDoc program (Nicholas et al., 1997). The analysis was carried out on gap-free alignments. The phylogenetic tree based on the aligned amino acid sequences of TRBC genes was constructed by applying ProtML (protein maximum likelihood, a program in MOLPHY package) using JTT-f for distances (Jones et al., 1992) and the neighbor-joining (NJ) method for phylogeny reconstruction (Saitou and Nei, 1987) as the starting tree for the maximum likelihood (ML) analysis. A phylogenetic tree was calculated also on the non-synonymous (first plus second) codon positions of TRBC nucleotide sequences and constructed by the NJ method using the stationary Markov distances (Lanave et al., 1984).

The analysis on the nucleotide sequences of the 3′-UTR portions of TRBC genes was performed using the maximum parsimony (MP) method. We used the NJ method to reconstruct the phylogenetic tree.

For the sheep TRBJ genes analysis, we aligned the nucleotide sequences (coding plus RSS regions) with the corresponding sequences of human, mouse and pig. The nucleotide alignment length of TRBJ genes was 86, and the number of the analyzed nucleotide sites was 42 in full positions. Phylogenetic analysis was carried out by the MrBayes v3.0b4 program (Huelsenbeck and Ronquist, 2001), by using the general-time-reversible (GTR) substitution model (Lanave et al., 1984; Saccone et al., 1990) for nucleotide sequences with the invariant site plus Gamma options (four categories). For the Bayesian analysis, one cold and three incrementally heated chains were run for 1,000,000 generations and trees were sampled every 100 generations. For the analyses of nucleotide sequences, the results of the initial 200,000 generations were discarded (burnin) after stationarity of the lnL had been reached. The PAUP*4.04b program was then used to construct a majority role consensus tree with the results of the remaining 800,000 generations (8000 reconstructed trees), thus allowing the calculation of the Bayesian posterior probabilities at the different nodes.

2.6. Sequence analysis

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; Thompson et al., 1994) and IMGT database (http://imgt.cines.fr/; Lefranc et al., 2003). WebLogo (http://weblogo.berkeley.edu/; Crooks et al., 2004) was used for generating TRBJ RSSs logos.

The analysis of the genomic structure of the sheep TRB locus was conducted with the following analyses 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 (http://pipmaker.bx.psu.edu/pipmaker/; Schwartz et al., 2000) for the alignment of the sheep sequence (this study) with the human (GenBank accession no. NG_001333: pos. 638,850–667,340), mouse (GenBank accession no. AE000665: pos. 151,874–178,377) and pig (GenBank accession no. AB079894: pos. 1–32,138) counterparts.

2.7. RT-PCR

Total RNA was extracted from sheep-isolated peripheral blood lymphocytes under the protocol approved by the manufacturer (Trizol reagent, Gibco BRL). About 5 μg of RNA was reverse transcribed into cDNA using the oligonucleotide CB40, which is complementary to a region in the 3′-UTR shared by TRBC1 and TRBC3 genes, followed by primer extension with Superscript RT II (Invitrogen-Life Technologies). PCR amplification of cDNA was performed with Platinum Taq Polymerase (Invitrogen-Life Technologies), using a sense V primer (VB3) complementary to a conserved region in the 3′-end of sheep TRBV genes (Halsey et al., 1999) and an antisense C primer (CB3) located in the third exon of the sheep TRBC genes. The final volume was of 50 μl. The PCR conditions were the following: 30 s at 94 °C, 30 s at 58 °C, 30 s at 72 °C for 40 cycles. The final cycle was extended for 10 min at 72 °C. The nucleotide sequences and the positions of primers are listed in Table 1.

Amplified cDNA fragments were purified using the PureLink PCR Purification Kit (Invitrogen-Life Technologies), ligated into StrataClone PCR Cloning Vector and transformed into StrataClone Competent Cells (Stratagene).

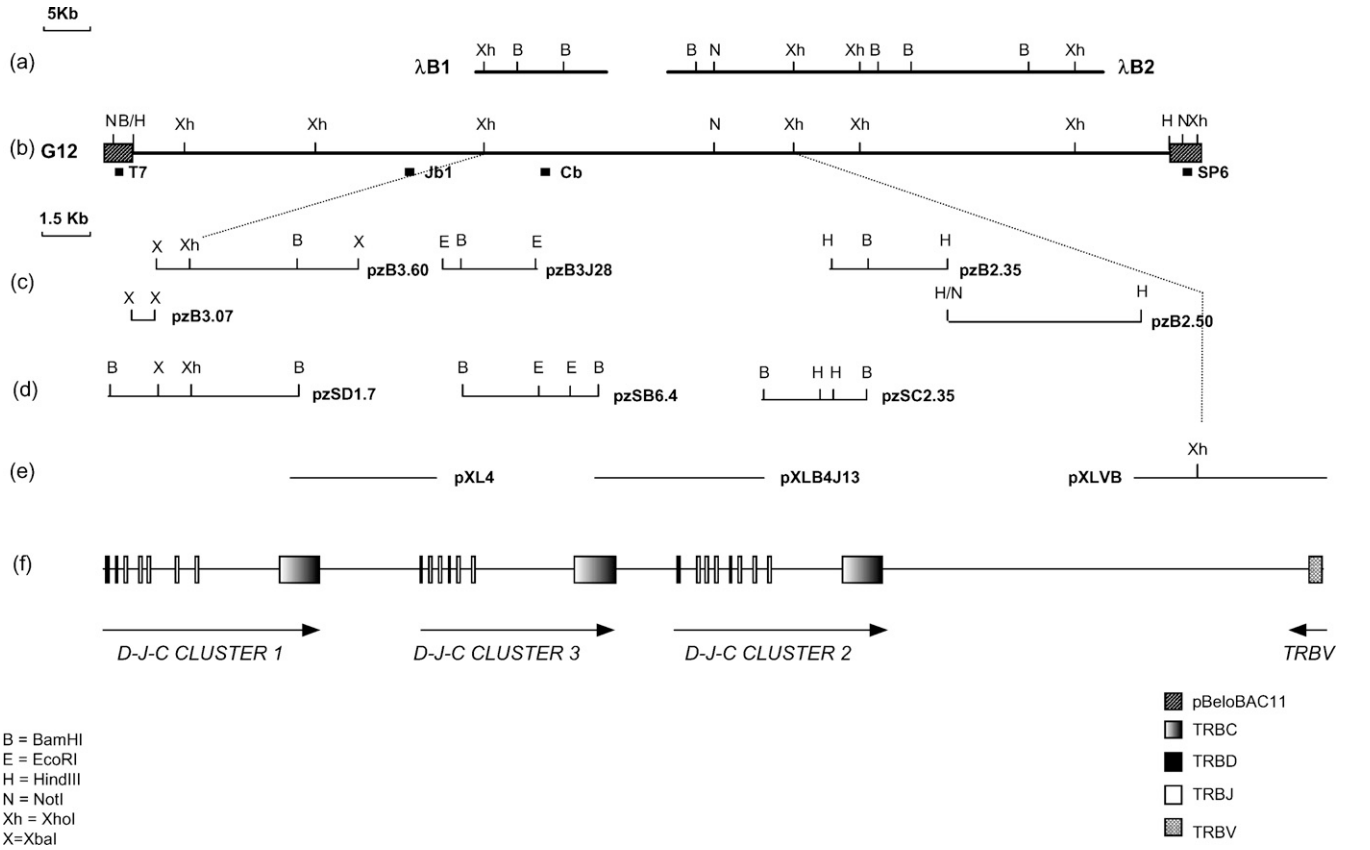

Fig. 1. Schematic maps of λB1, λB2 phage (a) and G12 BAC (b) clones and derived subclones (c–e), representative for the 3′-end of sheep TRB locus. Solid boxes in (b) indicate the probes used in Southern blot analysis (see text). In (c), plasmid subclones derived from phages; in (d), the plasmid subclones obtained from G12 DNA BamHI digest; in (e), plasmid subclones obtained from G12 clone by Long-PCR (see text for details) are represented. (f) Genomic organization of the 3′-end of TRB locus as inferred from the analyses of clones and subclones. Boxes indicate the positions of the TRBD, TRBJ, TRBC and TRBV genes identified within this region, with the cluster names indicated immediately below. The underlying arrows indicate transcriptional orientations. The genes are not to scale.

3. Results

3.1. Isolation of sheep genomic clones

In order to know the genomic organization of TRB locus in sheep, overlapping genomic clones containing the sheep D-J-C region were isolated.

In a first moment we screened a sheep λDASHII genomic library using a radioactive labeled human truncated cDNA probe containing the TRBJ1-TRBC1 sequence. Two phage clones (λB1 and λB2) were recovered (Antonacci et al., 2001). The restriction maps of λB1 and λB2 with BamHI, XhoI and NotI endonucleases showed that the clones are different, non-overlapping, with genomic insert of 13.7 and 44 kb, respectively (Fig. 1a).

Southern blot analysis were carried out on λB1 DNA digested with EcoRI and XbaI and on λB2 DNA digested with HindIII restriction enzymes and hybridized by using the same probe in low stringency conditions. The hybridization revealed the presence in λB1 of two XbaI (0.7 and 6 kb long) and one EcoRI (2.8 kb long) positive fragments, whereas in λB2 only a 3.8 kb HindIII fragment was identified. All positive fragments were subcloned and sequenced (Fig. 1c). Sequencing analysis, against database entries, identified both in pzB3.07 clone (0.7 kb XbaI fragment) and in the first portion of pzB3.60 clone (6 kb XbaI fragment), two TRBJ genes showing a high degree of homology to the human TRBJ1 genes. In the same way, the sequence analysis of the last part of pzB3.60 clone identified a region of nucleotide homology with the human TRBC1 gene. Based on these results, we recognized the TRBC sequence in λB1 clone as TRBC1 gene (Antonacci et al., 2001). In the pzB3J28 clone (2.8 kb EcoRI fragment) we recognized four more TRBJ genes that have shown a sequence homology with the human TRBJ2 genes. The sequence analysis of the pzB2.35 clone (3.8 kb HindIII fragment) of λB2 showed the presence of one additional TRBC gene that we identified by sequence homology, as TRBC2 gene. To confirm it, we also subcloned and sequenced the 6 kb HindIII fragment (pzB2.50), following the TRBC2 fragment (Fig. 1c). In this clone we identified a region showing a high degree of nucleotide similarity with the human TRB locus enhancer region that is well known to be localized just after the TRBC2 gene.

To isolate the intervening region between the two phage clones, and to complete the entire D-J-C region of the sheep TRB locus, in a second moment, a sheep BAC library in pBeloBAC11 vector was analyzed. The library was screened by PCR using B1/B2 primers (Table1) designed on the first exon of the sheep TRBC genes. One clone (G12) was isolated (Fig. 1b). The insert

size of G12 clone was estimated after digestion of the extracted BAC DNA with XhoI and XhoI/NotI restriction enzymes that have recognized sites in pBeloBAC11 vector. The digest products were analyzed by CHEF gel electrophoresis. Estimated BAC size is 112 kb. For the Southern blot analysis, following electrophoresis, the digested BAC clone DNA was transferred onto nylon membrane and hybridized with radiolabeled PCR products or plasmid inserts (data not shown). Particularly, the G12 insert ends were determined by hybridization on XhoI and XhoI/NotI digested DNA with PCR products generated with appropriate vector primers (Table 1). Probe F/FR-1.4 (SP6 probe in Fig. 1b) singled out the SP6 end in an 11 kb XhoI fragment and probe R/RR-2.1 (T7 probe in Fig. 1b) in the 6 kb XhoI fragment. To map TRBC genes, we hybridized the digested BAC DNA with the radiolabeled PCR product of B1/B2 amplification (Table 1) (Cb probe in Fig. 1b). The hybridization was detected in a 33 kb XhoI and in a 25 kb XhoI/NotI fragments. When we used the insert of the pzB3.07 clone as probe (Jb1 probe in Fig. 1b) in high stringency condition of hybridization, we identified an 18 XhoI fragment. This result implies that the TRBJ1 cluster is contained in this fragment and it is upstream from the 33 XhoI fragment. The XhoI digest pattern of λB2 phage clone was used for the identification of the 3′-end of the BAC insert. Restriction analysis and Southern blot results are schematically reported in Fig. 1.

3.2. Characterization of G12 BAC clone

To scrutinize the hypothesis of the presence of three TRBC genes in the ovine genome, the Cb-probe was hybridized on BamHI and HindIII digested G12 DNA. We detected three bands on both BamHI (6, 4.3 and 3.3 kb) and HindIII (9, 4.2 and 3.4 kb) digest (data not shown). The G12 DNA was thus digested by the BamHI restriction enzyme and the resulting fragments were subcloned in the pZero plasmid vector and screened with the same Cb probe. Three positive plasmid clones were found (pzSD1.7, pzSC2.35 and pzSB6.4), characterized by restriction fragment analysis and subjected to nucleotide sequencing analysis (Fig. 1d).

In the pzSD1.7 clone, with the 6 kb BamHI insert, we identified the first exon of TRBC1 gene as well as the TRBD1 gene and six TRBJ1 genes. As shown in Fig. 1, this clone overlaps the pzB3.60 phage clone and completes D-J-C cluster 1. The pzSC2.35 clone, with the 3.3 kb BamHI insert, overlaps the pzB2.35 phage clone and contains the first exon of TRBC2 gene. The clone also retains a TRBJ gene surely belonging to D-J-C cluster 2. Finally the sequencing of the pzSB6.4 clone, with the 4.3 kb BamHI insert, showed the presence of the first exon of a third TRBC gene and of a TRBJ gene, which is identical to the last one already found in the pzB3J28 clone. Therefore this TRBJ gene, together with the other TRBJ genes of the pzB3J28 clone and with the exon of the TRBC gene, identified a third D-J-C cluster. We named the new cluster as D-J-C cluster 3 for the chronological order of its discovery.

In order to complete the sequencing of the three clusters, we cloned the intervening region between pzSD1.7 and pzSB6.4 clones, and between pzSB6.4 and pzSC2.35 clones by two Long-PCRs (see Section 2). Primers designed on the sequences of the 3′-end of the pzSD1.7 and 5′-end of pzSB6.4 clones were used for amplification of the region comprised between D-J-C cluster 1 and D-J-C cluster 3, using the BAC DNA as a template. In the same way, primers designed on the sequences of the 3′-end of pzSB6.4 and the 5′-end of pzSC2.35 clones were used for amplification of the region comprised between D-J-C cluster 3 and D-J-C cluster 2. The PCR products were cloned in pXL-TOPO TA vector and the obtained clones were denominated pXL4 and pXLB4J13, respectively (Fig. 1e). The sequence analysis showed the presence in pXL4 of the remaining portion of TRBC1 gene, a TRBD and four TRBJ genes, belonging to D-J-C cluster 3. Moreover, in pXLB4J13 we isolated the residual portion of the third TRBC gene, the TRBD2 and six TRBJ2 genes.

We completed the TRBC2 sequence in the BAC clone by recovering from HindIII digested BAC DNA the 3.4 kb fragment positive to Cb probe. Since the sequence analysis of the corresponding clone revealed that the TRBC2 gene was identical to that found in λB2 phage clone, the BAC subclone is not reported in Fig. 1.

Finally, to investigate for the presence also in sheep of an inverted TRBV gene at the 3′-end of the TRB locus, we performed a third Long-PCR amplification on BAC DNA with a sense primer designed on the known sequence of the 3′-end of the pzB2.50 clone (Fig. 1c), while the antisense primer was constructed on the leader sequence of the goat TRBV3S1 gene present within a cDNA clone (GenBank accession no. AF035472) that resulted closely related to the unique inverted TRBV gene described in artiodactyls (GenBank accession no. AB079894). A PCR product of about 6 kb was obtained. The fragment was cloned in pXL-TOPO TA vector (pXLVB) and sequenced (Fig. 1e).

When the sequences of all clones were assembled, we obtained a 41,045 bp region (GenBank accession no. AM420900) encompassing the 3′-end of the sheep TRB locus from the TRBD1 to the TRBV gene.

Contents of the sequenced region are schematically illustrated in Fig. 1f. In summary, the region contains three D-J-C clusters distributed over about 28 kb which lie in the same transcriptional orientation. D-J-C cluster 1 spans about 8.7 kb and contains one TRBD, six TRBJ and one TRBC gene. D-J-C cluster 3 is located at 2.4 kb downstream cluster 1 with a total length of about 6.8 kb and includes one TRBD, five TRBJ and one TRBC gene. Finally D-J-C cluster 2 is positioned at 2.6 kb downstream cluster 3 and extends over 6.6 kb with one TRBD, seven TRBJ and one TRBC gene. At 13 kb separated from the TRBC2 gene, in an inverted transcriptional orientation, lies a TRBV gene.

3.3. Gene organization

3.3.1. TRBC genes

The most striking feature of the ovine 3′-end sequence of the TRB locus is the presence of three nearly identical TRBC genes. It is widely recognized that the structure of all known mammalian TRBC genes is conserved and composed of four exons

Table 2
Percentage nucleotide identities of each exon/intron between the sheep TRBC genes

	EX 1	INT 1	EX 2	INT 2	EX 3	INT 3	EX 4 (CDS)	EX 4 (3′-UTR)
C1 vs. C2	98.5	98	100	80	99.1	<20	95.2	<20
C1 vs. C3	99.0	79	100	78	**99.1**	**97**	**100**	**98.4**
C2 vs. C3	**99.5**	79	100	**94**	98.1	<20	95.2	<20

and three introns. The analysis of our sequence showed that also in sheep, all three genes present the same organization (GenBank accession no. AM420900). The four exons measure 390, 18, 107 and 21 bp, respectively. The 3′-UTR region that is 199 bp long in TRBC1 and TRBC3 and 202 bp long in TRBC2 is also present in the fourth exon. When we compared the nucleotide sequences of TRBC1, TRBC2 and TRBC3 coding region, we observed that TRBC1 and TRBC3 differ for only five residues, four in the first and one in the third exon. Also TRBC3 and TRBC2 genes change for five residues, two of which were localized both in the first and third, and one in the fourth exon. Diversely, TRBC1 and TRBC2 differ for eight nucleotides, six of these were located in the first exon and one both in the third and fourth exon. The 3′-UTR of TRBC3 is almost identical to TRBC1, while it differs from TRBC2.

Substantial differences in length and nucleotide composition in introns were detected. The first intron is 675 bp long for TRBC1 and TRBC2 versus 406 bp long in the TRBC3 gene. Minimal length differences were identified in the second (156, 158, 161 bp for TRBC1, TRBC2 and TRBC3, respectively) and third (286, 311, 275 bp for TRBC1, TRBC2 and TRBC3, respectively) introns. In contrast, large nucleotide composition differences have been observed in the second intron of TRBC1 with respect to TRBC2 and TRBC3, while the third intron of TRBC2 differs significantly from TRBC1 and TRBC3. Table 2 summarizes the percentage of nucleotide homology for each exon and intron calculated in pair-wise combinations between the three TRBC genes. In this respect, it can be noticed (in bold in Table 2) that TRBC3 shows a high level of nucleotide homology in its first half part with TRBC2, while from exon 3 to 3′-UTR it results almost identical to the TRBC1 gene.

All of the three genes encode a similar protein of 178 aa. Particularly, the eight nucleotide differences between TRBC1 and TRBC2 result in four amino acid variations, while TRBC2 and TRBC3 proteins are identical (Fig. 2). All the variable residues are clustered in the extracellular domain that is encoded by exon 1, exon 2 and by the first codon of exon 3. In the extracellular domain, a nine amino acid connecting peptide (hinge region) containing the cysteine involved in interchain disulfide binding, can also be identified. The transmembrane domain, encoded by the remaining part of exon 3 and the cytoplasmatic portion encoded by exon 4, are completely conserved in all three of the gene products. Therefore, the TRBC3 isotype is identical to the TRBC2 gene on the amino acid level, but it results similar to the TRBC1 gene in the 3′-UTR region. If we also include in the alignment the amino acid sequence of the TRBC1 gene isolated from the phage clone λB1 (GenBank accession no. AJ309004), we can deduce that only the fourth and fifth amino acid residues can be considered specific of the TRBC1 genes, while the other amino acid variations are polymorphisms.

3.3.2. *TRBJ genes*

The nucleotide and deduced amino acid sequences of all the TRBJ genes identified in the region are reported in Fig. 3a and b. They are classified in accordance with the following TRBC gene and numbered in agreement with the genomic position within the respective cluster. All genes are typically 44–53 bp long and conserve the canonical FGXG amino acid motif, whose pres-

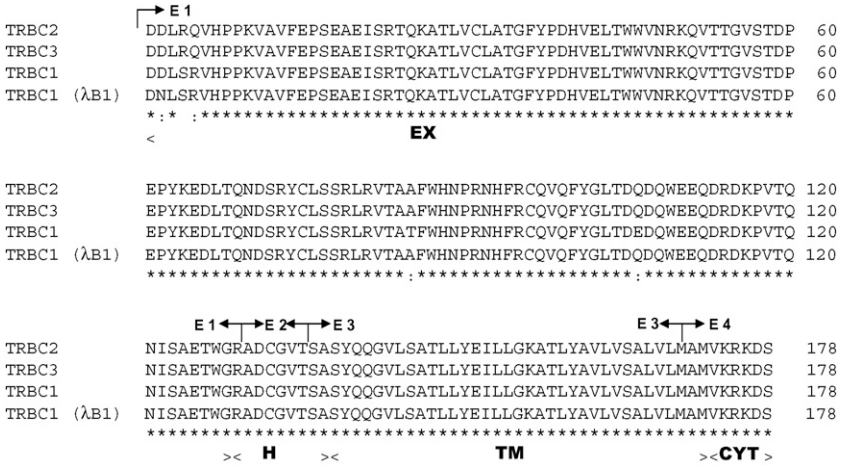

Fig. 2. The output of the Web-based multiple alignment program Clustal W comparing the inferred amino acid sequences of sheep TRBC1, TRBC2, TRBC3 genes, derived from G12 clone, and the TRBC1 gene from the λB1 clone. The lengths of the sequences are indicated at the right of the figure. Asterisks or dots below the alignment mark identical or functionality similar amino acid residues conserved, respectively, whereas arrows mark the boundaries between the exons. The boundaries between the predicted extracellular (EX), hinge (H), transmembrane (TM) and cytoplasmatic (CYT) domains are also indicated below the alignment.

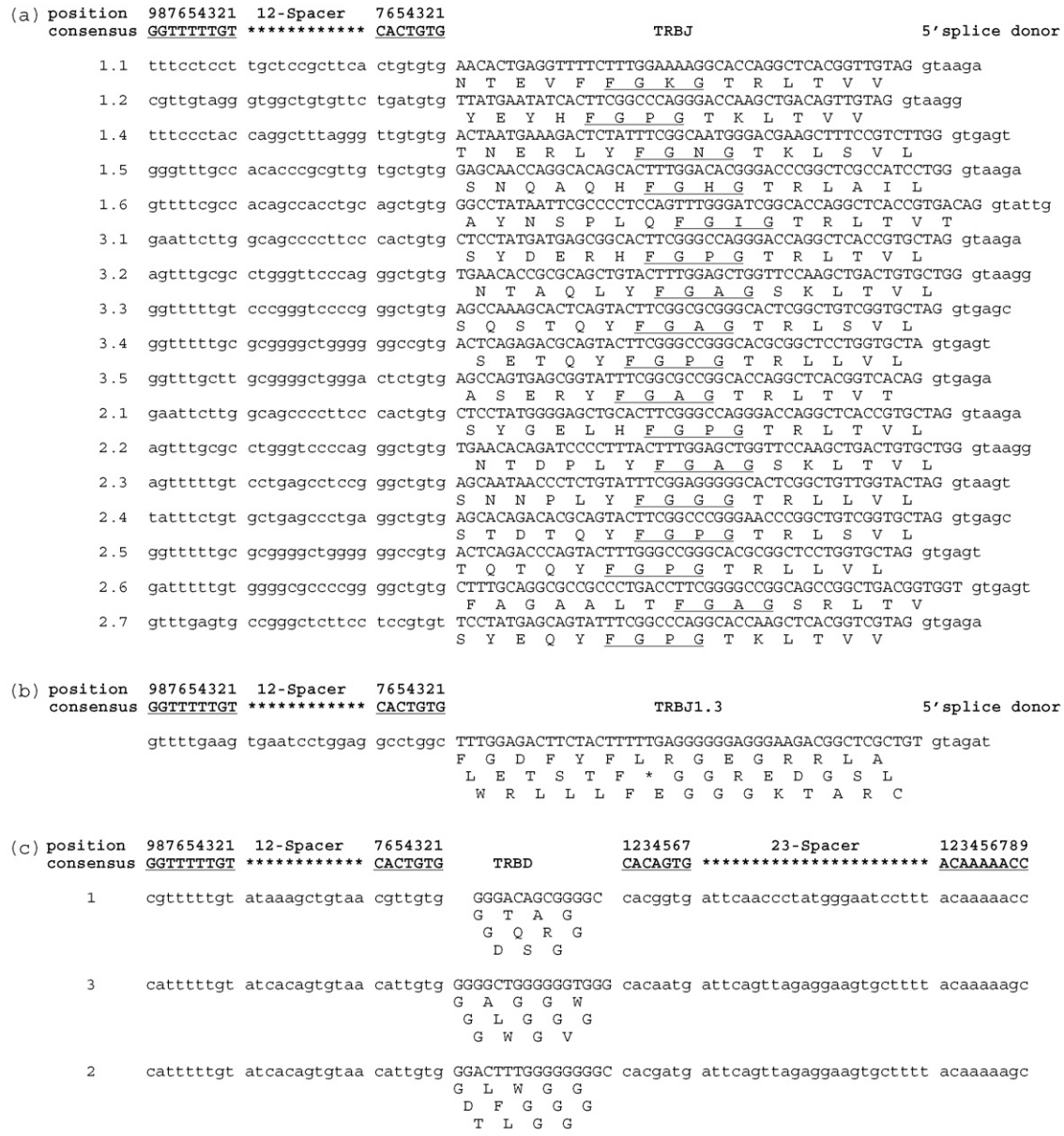

Fig. 3. Nucleotide and deduced amino acid sequences of sheep TRBJ (a and b) and TRBD genes (c). The consensus sequences of the heptamer and nonamer (Hesse et al., 1989) are provided at the top of the figure and underlined. The numbering adopted for the gene classification is reported on the left of each gene. In (a) and (b), the donor splice site for each TRBJ is shown. The canonical FGXG amino acid motifs are underlined. The TRBJ1.3 gene is a pseudogene since it lacks the FGXG amino acid motif in all the frames and shows an imperfect RSS. In (c), the inferred amino acid sequences of the TRBD gene in the three coding frames are reported.

ence defines the functionality of J segments. The only exception is TRBJ1.3, which lacks the FGXG amino acid motif in all the frames (Fig. 3b). Each TRBJ gene is flanked by a 12 RSS at the 5′-end and by a donor splice site at the 3′-end. Particularly, TRBJ1.3 exhibits alterated RSS. For this reason, together with the absence of the canonical FGXG motif, the TRBJ1.3 element could be considered a pseudogene. To evaluate the degree of conservation of the RSS sequences, the heptamers and nonamers of

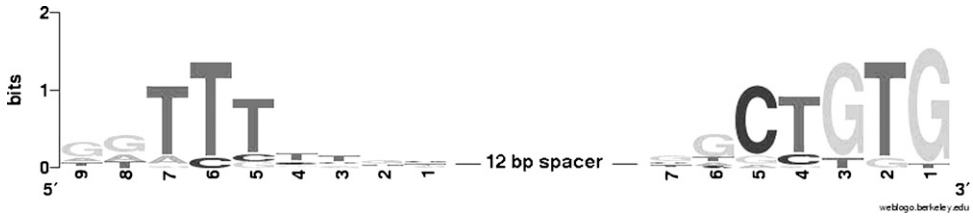

Fig. 4. Recombination signal sequence logos of the heptamers and nonamers of the sheep TRBJ genes, generated using Weblogo (Crooks et al., 2004). The taller letters indicate the high relative frequency of each nucleotide at that position.

all the TRBJ genes were aligned to obtain a consensus sequence by the program WebLogo3 (Crooks et al., 2004) (Fig. 4). Evidently, TRBJ1.3 was excluded from the analysis. The WebLogo results show that most of the positions of the heptamer matches with the canonical consensus sequence (Hesse et al., 1989). In particular, the first 3 bp, GTG, which flank the coding end, are highly conserved. These three nucleotides have demonstrated to be essential for RSS recognition, and changes in those positions preclude recombination (Akamatsu et al., 1994; Hesse et al., 1989). The following two positions are 82% conserved with nucleotides identical to the canonic heptamer consensus sequence, while the sixth and seventh positions are much less conserved. Within the nonamer most individual positions are less conserved. However, a stretch of 3 Ts, that appears to be a key feature of the sequence (Akamatsu et al., 1994), is almost perfectly conserved.

3.3.3. TRBD genes

Upstream of each TRBJ gene cluster, a TRBD gene has been identified and classified in accordance with the number of the J-C cluster. Fig. 3c shows the nucleotide and deduced amino acid sequences of the three TRBD genes. They consist of a 13 bp (TRBD1) or 16 bp (TRBD2 and TRBD3) G-rich stretch that can be productively read in its three coding phases and encoded 1–4 glycine residues, depending on the phase. The RSSs that flank the 5′ and 3′ sides of the coding region are well conserved. Both the 5′ and the 3′ nonamers are almost well conserved with respect to the consensus. The heptamers are also well conserved, and there is even similarity among the 12 as well as the 23 spacers.

3.3.4. TRBV gene

In the last portion of sheep TRB locus, as in the other mammals, we have also identified a TRBV gene in an inverted transcriptional orientation (GenBank accession no. AM420900 at position 40,386–41,045). This gene is typically made up of two exons: the first one (44 bp long), which encodes the initial part of the leader region, is followed by an intron measuring 325 bp. The second exon (295 bp in length) encodes the remaining portion of the leader and the entire V region. Characteristic 23 RSS, showing only slight differences with the consensus sequences (Ramsden et al., 1994), is visible immediately downstream of the 3′-end of the gene. This gene belongs to the TRBV3 subgroup on the basis of the high level of homology with the goat TRBV3S1 gene (GenBank accession no. AF035472).

3.4. Phylogenetic analysis of sheep TRBC and TRBJ genes

To gain insight into the evolution of the three sheep TRBC genes, we combined in the same alignment all the TRBC gene sequences included in GenBank database either as genomic or cDNA sequence from species belonging to different classes of vertebrates. In particular, we used within mammalian species three TRBC sequences isolated from sheep (OA) (this study) and cow (BT); two derived from human (HS), mouse (MM), rat (RN), pig (SS), horse (EC), and rabbit (OC) and only one TRBC sequence from dog (CF); three axolotl (AM) and one *Xenopus* (XL) TRBC genes for amphibians; a single chicken (GG) sequence for avians; four from cod (GM), two from catfish (IP) and damselfish (SP), one from salmon (SS) and trout (ST) for teleost fish; one TRBC sequence derived both from skate (RE) and shark (HF) for cartilaginous fish.

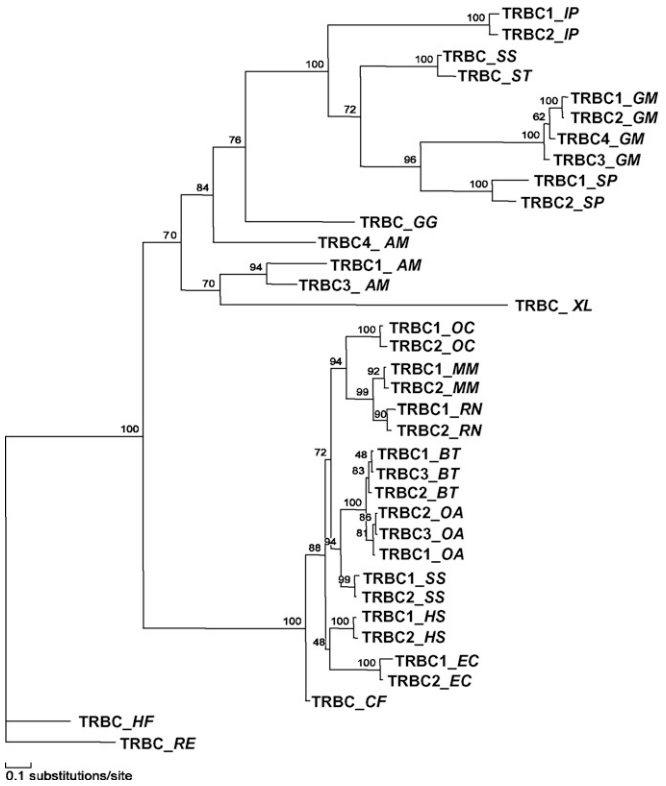

Fig. 5. The ML unrooted tree inferred from the TRBC gene sequences within vertebrate species (36 sequences) based on the JTT-F model of amino acid substitution and NJ reconstruction tree. Numbers along the branches show the percentage occurrence of nodes in 100 replicates of simulation. The ML tree was constructed using PROTML. The gene sequences from different organisms were listed by the gene notation as proposed by IMGT (Lefranc et al., 2003) plus the *two-letter code* derived from the species' Latin names.

A molecular phylogenetic tree of the coding region of TRBC genes was constructed using the ML and NJ methods (see Section 2) (Fig. 5). The tree resolved the TRBC genes into three separate groups with the mammalian TRBC genes markedly divergent from those of the non-mammals (amphibians, avians and teleost fish). Cartilaginous fish (skate and shark) are a breed apart from the sequences of the vertebrates examined. The tree recapitulates the accepted phylogeny of these species. It is noteworthy that, within each species, the phylogenetic tree presents a branching order that places the different TRBC isotypes together, except for axolotl TRBC4. To examine the reliability of the tree topologies generated by the amino acid analysis, a multiple alignment including the nucleotide sequences of the same TRBC sequences was also carried out. A phylogenetic tree drawing by using the Markov model (see Section 2), showed the same group distribution as in Fig. 5 (data not shown).

Previous studies have identified, in the 3′-UTR of human and mouse TRBC genes, specific isotypic sequence conservation. The TRBC sheep sequences have shown how TRBC1

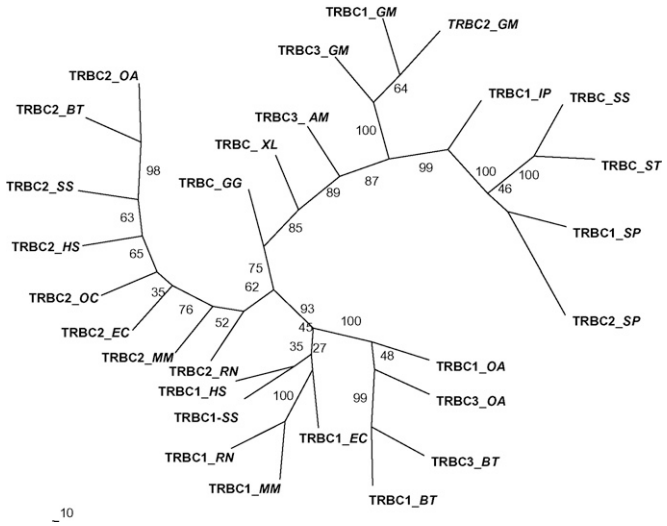

Fig. 6. The MP unrooted tree inferred from the 3′-UTR nucleotide sequences of the TRBC genes within different species of vertebrates (28 sequences), based on the NJ reconstruction tree. Numbers along the branches show the percentage occurrence of nodes in 100 replicates of simulation. The tree was constructed using MP method. The gene sequences from different organisms were listed by the gene notation as proposed by IMGT (Lefranc et al., 2003) plus the *two-letter code* derived from the species' Latin names.

and TRBC3 genes shared the same 3′-UTR. To investigate the evolutionary behaviour of this region with respect to their coding region, we performed a further phylogenetic analysis based on the nucleotide sequences of the 3′-UTR portions of TRBC genes of the different species of vertebrates. For this purpose, all the sequence portions situated downstream of the stop codon to the poly-A site were aligned. Consequently, the TRBC sequences lacking this portion were discarded from the analysis. A tree was constructed by using MP and NJ methods (see Section 2) (Fig. 6). In this phylogenetic tree, significant bootstrap values support a clear-cut subdivision of mammalian TRBC genes (but not of the non-mammals) into two main groups, each containing a different TRBC isotype. As expected, in the case of sheep as well as of cattle, the TRBC3 gene is grouped together with the TRBC1 gene. This different branching order with respect to the above tree revealed that a substantial different mechanism of evolution has occurred on the coding and non-coding regions of the mammalian TRBC genes.

To explore the phylogenetic relationship of TRBJ genes with respect to TRBC genes in each cluster, the nucleotide sequences of all sheep TRBJ genes (coding region plus RSS) were aligned with those of human, mouse and pig where the genomic location is known and the relative nomenclature was adopted (see Section 2) The alignment was then used to construct a phylogenetic tree by using Bayesian method (see Section 2) (Fig. 7). The results appear to emphasize the fundamental role of J genes and related RSS in the recombination event since the tree resolved a TRBJ subdivision of all species into 13 groups each containing orthologous TRBJ genes as indicated by their conserved germline order within the different TRB loci. If sheep TRBJ sequences are considered, a significant distribution can be noticed in the tree, that

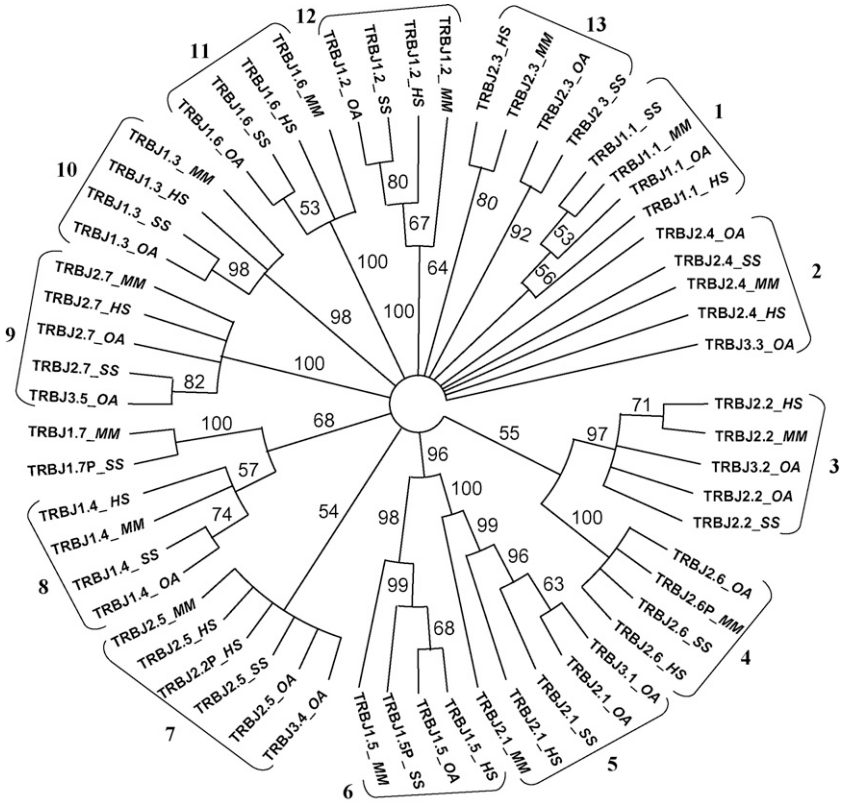

Fig. 7. The Bayesian phylogenetic tree of mammalian TRBJ genes, reconstructed from nucleotide sequences. Branch lengths are proportional to the number of substitutions per site. Numbers close to the nodes represent Bayesian posterior probabilities. The gene sequences from different organisms were listed by the gene notation as proposed by IMGT (Lefranc et al., 2003) plus the *two-letter code* derived from the species' Latin names.

is, that, the TRBJ3 genes are in the same group the TRBJ2 elements are, indicating the close phylogenetic relationship of the TRBJ2 and TRBJ3 genes.

3.5. Architecture of the genomic D-J-C region

To delineate the genomic structure of the entire 3′-end of the sheep TRB locus, the 41 kb sequence, determined in this study, was screened, at first, with the RepeatMasker program to highlight interspersed repeats. The density of total interspersed repeats results in 22.2% (supplementary Fig. 1Sa). The most abundant repeat elements are SINEs (13.39%), with half of them belonging to the SINE/BovA family (supplementary Fig. 1Sb). This result is consistent with the general assumption that regions that are rich in SINE have a high content in GC since the GC content is of 51.86%.

The masked sequence was then aligned against itself, using the PipMaker program (Schwartz et al., 2000), and the alignment expressed as a percentage identity plot (pip) (Fig. 8a). The pip shows the position of all gene segments identified in the region and the location and orientation of all repetitive sequences. It is noticeable the presence of three LTR/MaLR elements, two of them (210 bp long) are located in the first intron of the TRBC1 and TRBC2 genes and exhibit 100% sequence similarity. The other LTR (430 bp long) is located in the long, intergenic region comprised between the TRBC2 and TRBV genes. This region appears particularly rich in repetitive sequences. The display in Fig. 8a also shows putative CpG

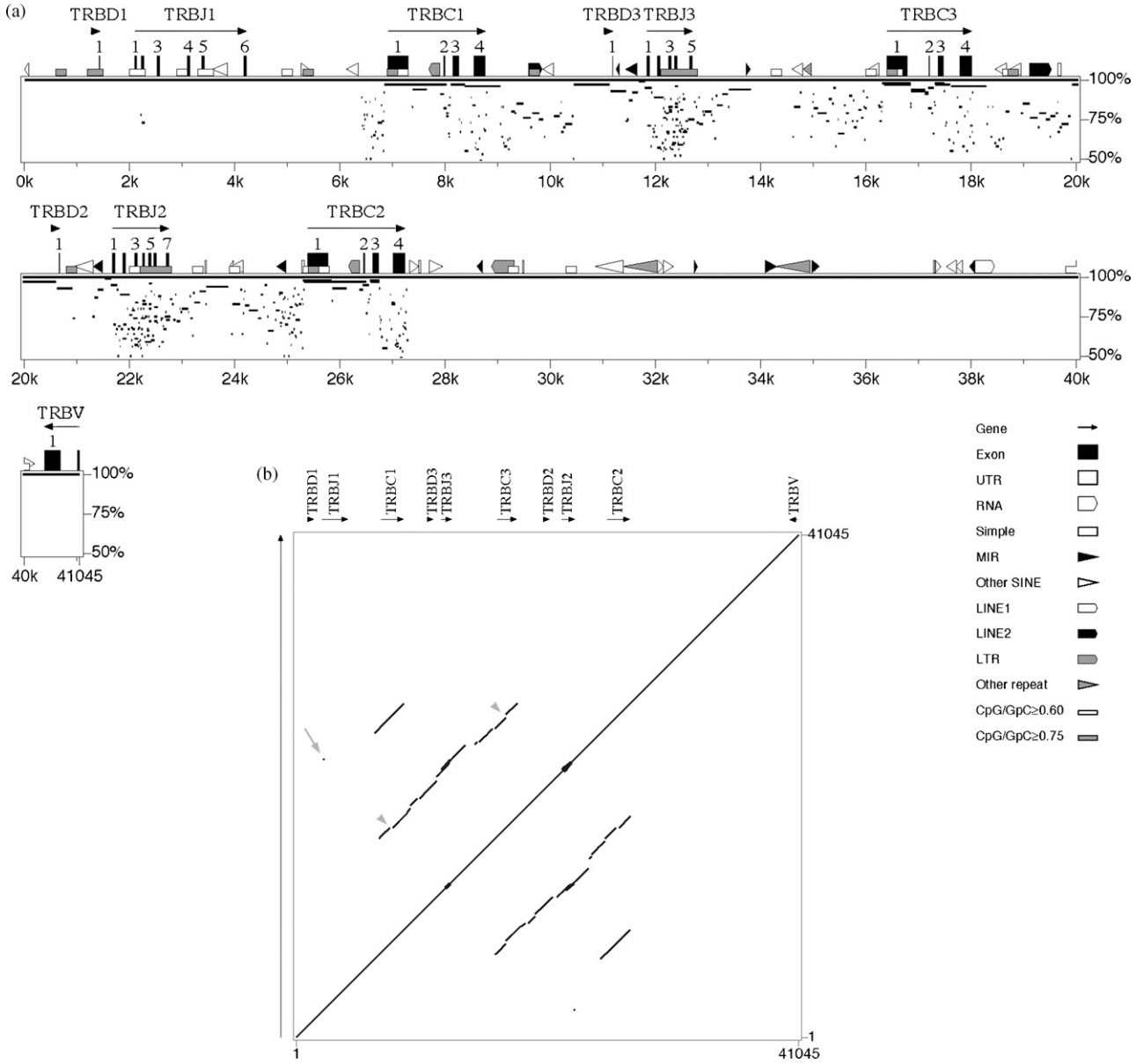

Fig. 8. Analysis of the genomic structure of the sheep D-J-C region by PipMaker program (Schwartz et al., 2000). (a) Pip of the 41 kb sequence versus itself. The position and orientation of all genes and repetitive sequences are indicated. Horizontal lines represent ungapped alignments at the percentage similarity corresponding to the scale on the right. (b) Dot-plot analysis of the same alignment. With the exception of the main diagonal, diagonal lines indicate internal homology units in the sequence.

islands, with prominent CpG islands centered at 1.5, 12.5 and 22.5 kb position.

Inspection of the pip allows us to identify portions of the sequence that align with more regions within the sequence itself. The occurrence of redundant matches cause lines, which appear superimposed in the pip. This aspect is consistent with the presence of duplicate regions in the sequence. The duplications are shown more clearly by the dot-plot in Fig. 8b, which is a complementary display of exactly the same alignments shown in the pip of Fig. 8a. Typically, this dot-plot matrix has identical halves above and below the main diagonal. Particularly, it can be observed, besides the perfect main diagonal line for the match of each base with itself, two parallel lines representing two repeated regions within the sequence. The smaller region of duplication regards the constant gene, TRBC1 versus TRBC2, while the longest repeated unit concerns TRBC1 plus D-J-C cluster 3 versus TRBC3 plus D-J-C cluster 2. In this second repeat unit, the lines are interrupted. This is the result of random deletion of individual duplicate regions subsequent to the initial duplication of the whole region. Among other things, the deletion within intron 1 of TRBC3 with respect to TRBC1 and TRBC2 genes is evident (grey head-arrows in Fig. 8b).

Furthermore, the 5′-end of the sequence, containing the D-J cluster 1 segments, as well as the 3′-end, from D-J-C cluster 2 to the TRBV gene, does not show similarity with any other region of the entire sequence except for the homology between the TRBJ1.2 and TRBJ2.7 genes (grey arrow in Fig. 8b).

In addition to sheep sequence determined in this study, fully annotated sequences of the 3′-end of TRB locus are currently available for humans, mice and pigs. We used these sequences to conduct a comparative analysis of the last part of TRB locus among these species in order to increase the genome data analysis and annotation of our sequence and identify functional sites such as regulatory elements.

For this purpose, the 41 kb of the sheep genomic sequence were aligned with the human, mouse and pig counterparts, using the same PipMaker program (Fig. 9).

Comparison of the sequences showed extensive matches along the entire region. Redundant matches from the TRBC1 to TRBC2 region are also evident due to the presence of the third duplicated D-J-C cluster in the sheep genome. More detailed comparison of the sequences showed that the clearest matches occurred within the syntenic region composed of orthologous

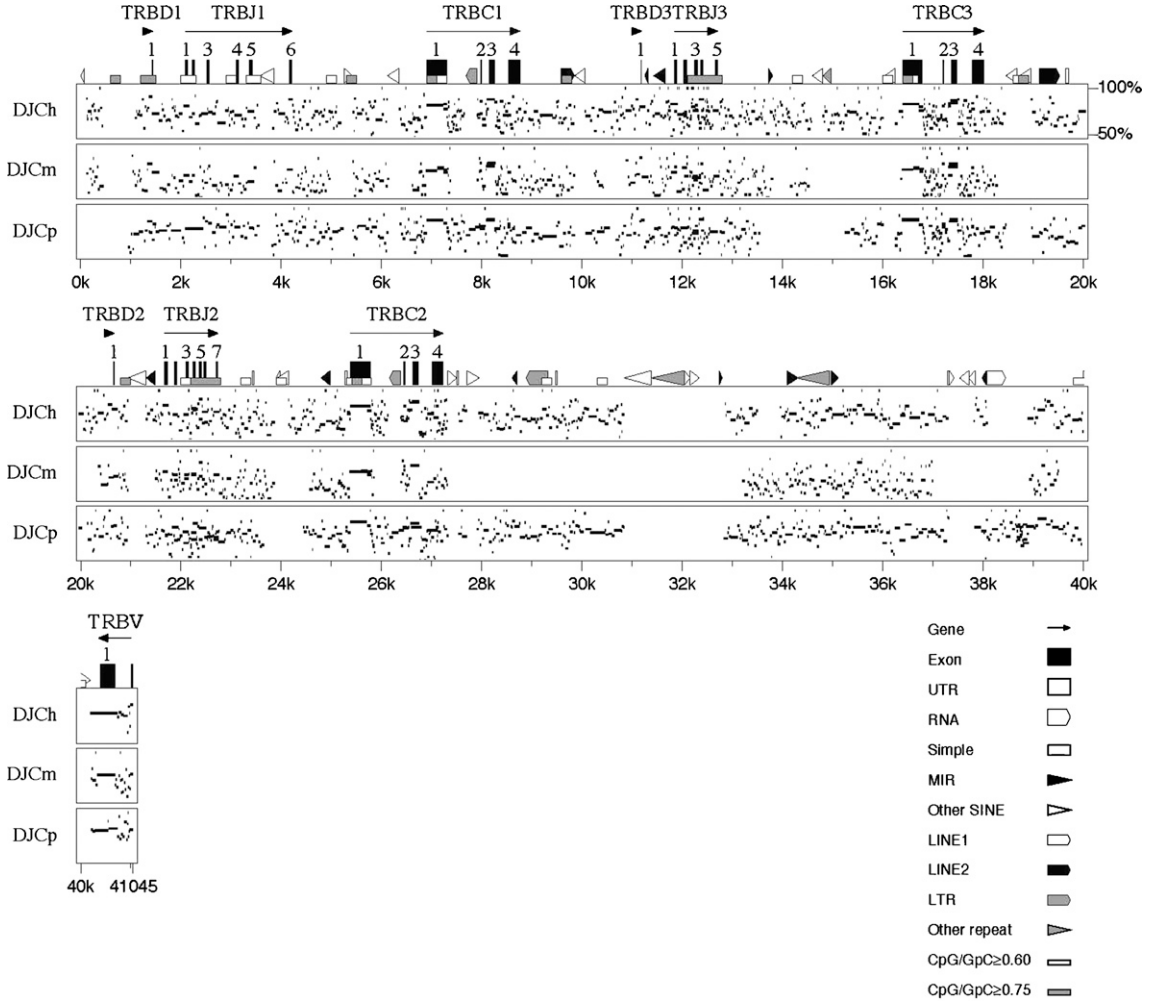

Fig. 9. Pip of the 3′-end of TRB locus between sheep, human, mouse and pig. The sheep sequence is shown on the horizontal axis and the percentage identity to the human (top plot), mouse (middle plot) and pig (lower plot) are shown on the vertical axis.

Table 3
Percentage nucleotide identities for the PDβ1 and PDβ2 regions

PDβ1% \ PDβ2%	sheep	human	mouse	pig
sheep		86	79	85
human	84		80	85
mouse	85	83		79
pig	85	86	83	

Numbers above the diagonal indicate the percentage nucleotide identity of the TRBD2 region, whereas numbers below the diagonal indicate the percentage nucleotide identity of the TRBD1 region.

genes for which the number of coding exons, their size and sequence are well conserved.

One block of large non-syntenic region between coordinate 30,846–32,845 in the sheep locus was also recognized. This block was found to be composed of three SINE/BovA and two LINE repeats. A further downstream block of five SINE and one LINE/L1 repeat elements (coordinate 37,310–38,900 in the sheep locus) was shared with pig but not with human and mouse. Much smaller non-syntenic regions are distributed along the entire sequence with SINE repeats.

Detailed comparison of the sequences allows us to note also three regions with high percent identities that do not coincide with any known coding regions. These regions can be identified as conserved non-coding sequences (CNS), that is, orthologous regions sharing at least 70% sequence identity across at least 100 bp.

The first of these CNS (coordinate 1321–1492 in the sheep sequence) spans from 98 bp upstream through 61 bp downstream of the TRBD1 gene. This region has been experimentally determined in mouse as the Dβ1 promoter (PDβ1), which together with the beta enhancer element (Eβ) regulate the germline transcription of the D-J-C portion of the locus (Doty et al., 1999; Yang et al., 2003). The identity percentages in pair-wise comparison of this region between the different species vary from 83 to 86% (Table 3). The four-species alignment, provided as supplementary Fig. 2Sa, shows the conservation in the sheep sequence of the TATA box, located in the 5′ RSS flanking the TRBD1 gene, the binding sequence for Ikaros/Lyf-1 at the -35 GC-rich motif and the GATA binding sequence at -74, essential for the murine promoter activity. Intriguingly, we also found a CpG site in the sheep heptamer of the 23 RSS shared with mouse and pig but not human. Analysis of methylation levels surrounding RSS before, during and after recombination have demonstrated that demethylation of this CpG site appears to be essential for TRBD1 cleavage by the V(D)J recombinase in mouse (Whitehurst et al., 2000).

A second similar CNS was identified proximal to the TRBD2 gene. In this case the conserved region (coordinate 20,561–20,707 in the sheep sequence) spans from 90 bp upstream through 39 downstream of TRBD2 gene, and the identity percentages in pair-wise comparison vary from 79 to 86% (Table 3). Sequence analysis indicates that this region is structurally different from that of the TRBD1 region because of the lack of the TATA box, the −35 GC-rich motif and the GATA binding sequence, while a high sequence conservation is present at −90 and −60 among all four species (supplementary Fig. 2Sb). Although a second Dβ promoter has not yet been characterized, germline transcriptions of D-J-C portion can also be recognized near the TRBD2 gene (Doty et al., 1999; Yang et al., 2003). In the alignment, the corresponding sheep region proximal to TRBD3 gene might be inserted since it presents an identity percentage with the sheep TRBD2 region of 94% (data not shown).

Table 4
Percentage nucleotide identity of the enhancer region

Eβ%	Sheep	Human	Mouse
Human	82		
Mouse	80	84	
Pig	82	76	75

The third CNS, coordinate 35,441–35,627 in the sheep sequence, is located about 8 kb from the TRBC2 gene with identity percentages in pair-wise comparison varying from 75 to 84% (Table 4). This region corresponds to the Eβ element fully characterized in human and mouse and located about 5 kb away from the TRBC2 gene (Gottschalk and Leiden, 1990; Takeda et al., 1990). Seven protein binding motifs termed βE1-βE7 were identified within the mouse Eβ, and a systematic analysis of the elements has demonstrated that a fragment containing the βE1–βE6 regions is required for full enhancer activity in pro-T cells as well as in mature cells (Carvajal and Sen, 2000). The human Eβ five regions (Tβ1–Tβ5) of a 480 bp fragment, were shown to bind nuclear factors. Alignment between the murine and human enhancer sequences show that the murine βE1 and βE2 motif fall within Tβ2, βE4 falls within Tβ3, βE5 and βE6 fall within Tβ4, and βE7 falls within Tβ5 (Carvajal and Sen, 2000). Inspection of the four-species alignment of the region shows a more consistent conservation of the βE1, βE4 with its 3′ flanking sequence, and βE6 corresponding region (supplementary Fig. 2Sc).

3.6. Transcription assay of newly isolated D-J-C cluster 3

To examine whether the newly identified D-J-C cluster 3 is used in the formation of the TR β-chain repertoire, we developed a rapid RT-PCR-based assay. A sense V primer (VB3) complementary to a conserved region in the 3′-end of sheep TRBV genes and an antisense C primer (CB3) located in the third exon of sheep TRBC genes, were used in RT-PCR reactions from adult peripheral blood mRNA. The first-strand cDNA was prepared with an antisense C primer (CB40) designed on the 3′-UTR specific of the TRBC1 and TRBC3 genes. The nucleotide sequences and the positions of all primers are listed in Table 1. PCR products were cloned by using the StrataClone PCR Cloning Kit and random selected positive clones were sequenced.

Sequence information was obtained from a total of eight clones containing rearranged TRBV-D-J-C transcripts with a correct open reading frame.

clone	TRBV	N (TRBD) N	TRBJ	TRBC	
psSP40.1	YLCASS	GE(R)IN	HFGPGTKLTVV	DDLSR	**DB1-JB1.2-CB1**
psSP40.6	YLCASSQ	DSPA(GV)	SERYFGAGTRLTVT	DDLRQ	**DB3-JB3.5-CB3**
psSP40.18	YLCASSPT	E(TA)R	TQYFGAGTRLSVL	DDLRQ	**DB1-JB3.3-CB3**
psSP40.20	YLCASSPT	S(Q)L	ERHFGPGTRLTVL	DDLRQ	**DB1-JB3.1-CB3**
psSP40.22	YLCASSQ	A(GAG)	ETQYFGPGTRLLVL	DDLRQ	**DB3-JB3.4-CB3**
psSP40.23	YLCASS	S(GG)V	SQSTQYFGAGTRLSVL	DDLRQ	**DB3-JB3.3-CB3**
psSP40.30	YLCASSQ	GS(G)RRS	ERLYFGNGTKLSVL	DDLRQ	**DB1-JB1.4-CB3**
psSP40.33	YLCASSQ	VE(TA)RH	LYFGAGSKLTVL	DDLRQ	**DB1-JB3.2-CB3**

CDR3

Fig. 10. Predicted amino acid sequences of the CDR3 region of sheep β chain cDNA clones, segregated according to TRBJ and TRBC genes utilization. CDR3 region was defined according to Kabat and Wu (1991).

The nucleotide sequences of the cDNA clones analyzed in this work were registered in GenBank with the accession nos. from AM492538 to AM492545.

Fig. 10 shows the amino acid sequences of the clones starting from the V region to the first five amino acids of the C portion. The V-D-J junction or CDR3 region as defined by Kabat and Wu (1991) and the J sequences are indicated. Since the genomic nucleotide sequence is known, the contribution of each TRBD, TRBJ and TRBC genes becomes discernible in the cDNAs. It was also possible to determine the N-nucleotide contribution to the D-J junction. Nucleotide trimmings at the 5′- and 3′-end of the different TRBD and at the 5′-end of the TRBJ gene were also observed. Among these clones, only one contains the TRBC1 gene; half of the clones with the TRBC3 gene involve the TRBD1 gene, the other half, the TRBD3 gene. All the cDNAs match well with the genomic TRBJ segments. Overall, all the TRBJ genes belonging to cluster 3 are present, irrespective of which TRBD gene is used. We note the presence of the TRBC3 gene with a TRBD1-TRBJ1.4 rearrangement, indicating an alternative splicing product of the TRBC3 gene joining directly to the D1-J1 portion.

Although the number of clones analyzed in the present work is limited, the usage of all components of D-J-C cluster 3 appears to be demonstrated.

4. Discussion

By isolation and characterization of overlapping genomic clones containing the last portion of the sheep TRB locus, we obtained a nucleotide sequence covering a 41 kb stretch, from the TRBD1 to the downstream TRBV gene. The identification of the region allowed us a comparative assessment of the sheep genomic structure with respect to the previously characterized species.

Among the mammalian species studied so far, the genomic organization of TRB locus is highly conserved with two in tandem D-J-C gene clusters positioned downstream from a single large region of TRBV genes. The analysis of the sheep sequence revealed for the first time the presence in mammalian genome of an additional complete D-J-C cluster, in tandem aligned (Fig. 1). The structure of D-J-C cluster 1 and D-J-C cluster 2 are reminiscent of their mammalian equivalent. D-J-C cluster 3, located in between, clearly presents the same structure with the TRBC3 sequence identical to the TRBC2 gene in the coding region and to the TRBC1 gene in its 3′-UTR. Moreover, the single TRBD3 and the five TRBJ3 genes are correlated to the corresponding elements in cluster 2. Thus, the organization of this additional cluster is consistent with the idea that it has arisen through a duplication event due to an unequal crossing-over between the ancestral TRBC1 and TRBC2 genes.

The presence of a third cluster has also been established in cattle (Conrad et al., 2002), even if the sequence has not been fully annotated, but not in pig (Watanabe et al., 2007). These species are all members of the Cetardiodactyla phylum, with sheep and cattle belonging to the Bovidae family. Therefore, the birth of the third cluster must have occurred after the divergence between Suina and Ruminantia lineages. A more precise picture of the timing of this event could be obtained by analyzing the status of the TRB locus in other Ruminantia families (i.e. Cervidae).

Sequence comparison of the entire region among sheep, human, mouse and pig, revealed that the rate of occupation of interspersed repeats is significantly more abundant in sheep (22.2% in supplementary Fig. 1Sa) than in the other species (15.9% in human, 4.6% in mouse and 21.9% in pig) (data not shown). The most predominant repeat elements types are SINE, belonging to the BovA family, and LINE (supplementary Fig. 1Sb). The insertion of repetitive sequences, together with the presence of the additional D-J-C cluster, contributes to making this TRB region larger in size in the sheep genome (41 kb) with respect to those of human, mouse and pig (28, 26 and 32 kb, respectively). The origin and the role of these elements remain a matter for speculation, but it is possible that the large number of repeats may cause genetic instability in the locus. Particularly, the inspection of the sheep sequence shows the presence of two perfectly conserved LTR elements localized in the first intron of the TRBC1 and TRBC2 genes (Fig. 8a) that may be considered the substrate for the illegitimate crossing-over. Also the GC content is greater in sheep (51.9% in supplementary Fig. 1Sa) than in the other species (48.3% in human, 43.4% in mouse and 50.9% in pig) (data not shown), with prominent CpG sites associated with each functional region (Fig. 8a). Their association with the structure of the genomic region and/or with the regulation of gene expression should be evaluated.

The presence of the extra cluster and the numerous repetitive elements in the sheep region do not seem to interfere with the

locus arrangement. In fact, besides the coding regions, several intergenic spaces show a high degree of conservation between the four species (Fig. 9). One of these regions resides immediately at the 5′ of D-J-C cluster 1, encompassing the TRBD1 gene. This region has been characterized in human and mouse where it corresponds to the PDβ1 promoter region (Doty et al., 1999; Yang et al., 2003). Closer inspection of the homologous region in sheep revealed a distribution of putative transcriptional binding sites similar to the other species (supplementary Fig. 2Sa). Interestingly, the presence of a TATA box located at the 5′ of the sheep TRBD1 gene, like in mouse, suggests that a D-J germline transcript may also be produced in sheep. Another related finding with the mouse region is the presence of a CpG sequence within the 3′ TRBD1-RSS heptamer, of which methylation can be correlated to the local recombinational accessibility (Whitehurst et al., 2000). A striking conservation between species is also present in the region around the sheep TRBD2, where there resides an undefined promoter region in human and mouse, as well as the TRBD3 genes (Fig. 9 and supplementary Fig. 2Sb). In the sheep sequence the CpG site in the 3′ TRBD2-RSS heptamer is also conserved, and this would represent another level of control for the rearrangement. Another region of consistent conservation is located between the TRBC2 and the TRBV gene (Fig. 9 and supplementary Fig. 2Sc). This region corresponds to the beta enhancer, well characterized in human and mouse (Gottschalk and Leiden, 1990; Takeda et al., 1990). Also in this case, sequences for the major binding sites are shared among the four species analyzed, even if, because of the larger size of the entire TRB region, the sheep enhancer lies more distant from the PDβ1 promoter.

All together these findings support the idea that in sheep, the regulatory architecture of the TRB locus, with a promoter for germline transcription for each D-J-C region, that controls the local recombinational accessibility and an enhancer separated by a large distance that governs the general recombination, seems to be preserved. Moreover, the same enhancer element would control reiterated D-J-C regions in sheep.

The two TRBC gene isotypes found in all the mammalian species present a set of divergence gradient, with minimal differences in the coding region, extending to a very high divergence in the 3′-UTR, both within and between the species. The extent of permitted variation in the constant portion of the beta chain is constrained by both structural and functional requirements of the molecule. In fact, it does not bind antigen, while it must interact with monomorphic structures, like the extracellular part of the TR alfa chain and some members of the CD3 complex; therefore, it plays an important role in signal transduction. These structural constraints are so strong that similarity between the TRBC protein isotypes is likely to be maintained by gene conversion rather than by selection at the protein level (Rudikoff et al., 1992). Also in sheep, both the nucleotide and protein sequences of all three TRBC genes are highly similar to each other. Only two out of four amino acid residues, located in the N-terminal region, have undergone replacement in the TRBC1 gene with respect to the TRBC2 and TRBC3 genes (Fig. 2), while the other two variable residues can be related to the presence of different alleles or breed polymorphisms, as assessed by the sequence comparison between the TRBC1 genes isolated from the λB1 phage and G12 BAC clones, which derived from two libraries of two different sheep breeds. No differences were identified in the TRBC2 genes of the two breeds. With regard to the sheep TRBC3 gene, the maintenance of the high degree of homology in the coding region with respect to the TRBC1 and TRBC2 genes might have been due to a too short divergence time from the duplication event. Alternatively, this conservation may reflect, also in this case, the strong functional constraint.

The concerted evolution of the TRBC genes that seems to be an absolute necessity in mammalian species is not present in other vertebrate species. The channel catfish, a teleost in which the D-J-C sequences have been reported, contains two in tandem TRBC genes, which are substantially different, with only 36% identity at the amino acid level (Zhou et al., 2003). The situation is reminiscent of the recently reported TRBC genes in bicolor damselfish (Kamper and McKinney, 2002). TRBC isotypic variants have also been described in the Atlantic cod where four different TRBC genes were sequenced (Wermenstam and Pilstrom, 2001). The Mexican axolotl also appears to have multiple divergent TRBC genes as assessed by cDNA analyses (Kerfourn et al., 1996; Fellah et al., 2001). The presence of TRBC genes that seem to evolve more freely in non-mammalian than in mammalian species, may confer different biological properties on the corresponding beta chain.

Furthermore, the sequence comparison combined with a molecular phylogenetic analysis of the TRBC genes from mammalian and non-mammalian species revealed that if the coding region is reflective of the phylogenetic history in all species (Fig. 5), substantial variations are accumulated during mammalian evolution in the 3′-UTR sequence of each specific TRBC isotype, and they are conserved among all mammalian species (Fig. 6). Such inter-species conservation implies a strong selective pressure to preserve the sequence of the 3′-UTR of the different mammalian TRBC isotypes. Thus, we can infer that the 3′-UTRs are involved in post-transcriptional mechanisms of regulation. The same selective pressure may oblige the strong conservation of the 3′-UTR in sheep TRBC1 and TRBC3 genes with respect to TRBC2 gene.

If the presence of more than two TRBC genes, organized in three different D-J-C clusters, imposes constrains, it could also give biological advantages as the possibility to enhance the combinational and junctional diversity of the CDR3 domains increasing the number of TRBD and TRBJ genes.

In human, mouse and pig two TRBD genes and 12 (for human) or 13 (for mouse and pig) functional TRBJ genes take into account for the expression of two TRBC genes. In sheep, three TRBD and 18 TRBJ genes, distributed in the three clusters, are identified.

Furthermore, each of the sheep TRBD genes is composed of longer coding nucleotide sequences, comprised in two well-conserved RSSs, when compared with the human and mouse counterparts. It is noteworthy how these increments, observed also in pig, are translated into a major number of glycine residues useable in the CDR3 region (Fig. 3c). McCormack et al. (1991) suggested that the presence of glycine residue(s) seems to be

important for the structure of the antigen-binding CDR3 loop in beta chain.

Except for one, all the TRBJ genes appear to be functional; they are in frame with the FGXG motif and they have conserved splice signal sequences for a correct splicing with a TRBC gene (Fig. 3a and b) Moreover, each of them has a recognizable RSS, with conserved GTG nucleotides at the 3′-end of the heptamer, and a stretch of three Ts in the nonamer, which have been demonstrated to be critical for recombination by recombinase substrate assays (Ramsden et al., 1994; Hesse et al., 1989).

Our phylogenetic studies based on the TRBJ sequences allowed us to discern a classification of the sheep TRBJ genes (Fig. 7). The majority rule tree shows a distribution of the TRBJ genes in groups, each characterized by homologous genes of the different species, with an identical genomic position within its own cluster. The tree also confirmed the tight relationship between the sheep TRBJ3 and TRBJ2 genes as a consequence of the duplication event.

Our present knowledge on the genomic sequence of the sheep D-J-C region makes it easier to recognize the contribution of each D, J and C gene, particularly within cluster 3, in the sheep beta chain repertoire. Preliminary studies carried out on a limited panel of rearranged TR beta sequences, show, first of all, that the genes of the additional cluster 3 are relevant for the production of the β chain transcripts at least in peripheral T cells (Fig. 10). Moreover, a careful examination of the CDR3 region shows that a TRBD gene is always included in all V(D)J junctions according to the so called "beyond 12/23 rule" (Bassing et al., 2000). In spite of the presence of a germline TRBD gene with at least one glycine encoded in all three reading frames, we observed rearranged sequences without any glycine in the CDR3 region. We also noted the presence of TRBJ3 genes rearranged with either TRBD1 or TRBD3 genes, and a TRBD1-TRBJ1 rearrangement spliced to the TRBC3 gene. All together our results demonstrate that if the elevated number of sheep TRBD and TRBJ genes is able to create a major set of different D-J combinations, several other different mechanisms, including inter cluster rearrangements and splice variants, can be used to increase once more the sheep beta chain repertoire functional diversity.

In summary, the study reported here provides insights into the organization, evolution and the general architecture of the sheep TRB locus. Our finding emphasizes the presence of species-specific differences in the conserved mammalian TRB locus structure in spite of the same general mechanism for TR diversification. Our purpose will be to carry out a more intensive study in different immune tissues to conclusively determine the role and the possible advantage of cluster 3 in the sheep beta chain repertoire (work in preparation). Understanding the meaning of the differences among the species could contribute to the knowledge of the T cell mediated immunity.

Acknowledgments

This work was supported by grants from the University of Bari and University of Salento (Progetti di Ricerca Scientifica d'Ateneo). We thank Prof. M. Bozzetti for critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molimm.2007.05.023.

References

Akamatsu, Y., Tsurushita, N., Nagawa, F., Matsuoka, M., Okazaki, K., Imai, M., Sakano, H., 1994. Essential residues in V(D)J recombination signals. J. Immunol. 153, 4520–4529.

Antonacci, R., Massari, S., De Iaco, R., Ciccarese, S., 2001. Assignment of the TRB@ locus encoding the T-cell receptor beta chain to sheep, cattle, goat and river buffalo chromosomes by in situ hybridization. Cytogenet. Cell Genet. 94, 82–83.

Bassing, C.H., Alt, F.W., Hughes, M.M., D'Auteuil, M., Wehrly, T.D., Woodman, B.B., Gartner, F., White, J.M., Davidson, L., Sleckman, B.P., 2000. Recombination signal sequences restrict chromosomal V(D)J recombination beyond the 12/23 rule. Nature 405, 583–586.

Binns, R.M., Duncan, I.A., Powis, S.J., Hutchings, A., Butcher, G.W., 1992. T lymphocytes in the blood of young pigs identified by specific monoclonal antibodies. Immunology 77, 219–227.

Butler, J.E., Wertz, N., Sun, J., R.E, S., 2005. Comparison of the expressed porcine Vβ and Jβ repertoire of thymocytes and peripheral T cells. Immunology 114, 184–193.

Carvajal, I.M., Sen, R., 2000. Functional analysis of the murine TCR beta-chain gene enhancer. J. Immunol. 164, 6332–6339.

Conrad, M.L., Pettman, R., Whitehead, J., McKinnel, L., Davis, S.K., Koop, B.F., 2002. Genomic sequencing of the bovine T cell receptor beta locus. Vet. Immunol. Immunopathol. 87, 439–441.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.

De Guerra, A., Charlemagne, J., 1997. Genomic organization of the TcR beta-chain diversity (Dbeta) and joining (Jbeta) segments in the rainbow trout: presence of many repeated sequences. Mol. Immunol. 34, 653–662.

Doty, R.T., Xia, D., Nguyen, S.P., Hathaway, T.R., Willerford, D.M., 1999. Promoter element for transcription of unrearranged T-cell receptor beta-chain gene in pro-T cells. Blood 93, 3017–3025.

Fellah, J.S., Durand, C., Kerfourn, F., Charlemagne, J., 2001. Complexity of the T cell receptor Cβ isotype in Mexican axolotl: structure and diversity of the VDJCβ3 and VDJCβ4 chains. Eur. J. Immunol. 31, 403–411.

Galtier, N., Gouy, M., Gautier, C., 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. 12, 543–548.

Garcia, K.C., Teyton, L., Wilson, I.A., 1999. Structural basis of T cell recognition. Annu. Rev. Immunol. 17, 369–397.

Gottschalk, L.R., Leiden, J.M., 1990. Identification and functional characterization of the human T-cell receptor beta gene transcriptional enhancer: common nuclear proteins interact with the transcriptional regulatory elements of the T-cell receptor alpha and beta genes. Mol. Cell. Biol. 14, 4286–4295.

Grossberger, D., Marcuz, A., Fichtel, A., Dudler, L., Hein, W.R., 1993. Sequence analysis of sheep T-cell receptor beta chains. Immunogenetics 37, 222–226.

Halsey, W.A.J., Palmer, B.E., DeMartini, J.C., Howell, M.D., 1999. Analysis of sheep T-cell receptor beta-chain heterogeneity. Immunogenetics 49, 206–214.

Hein, W.R., MacKay, C.R., 1991. Prominence of γδ T cell in the ruminant immune system. Immunol. Today 12, 30–34.

Hesse, J.E., Lieber, M.R., Mizuuchi, K., Gellert, M., 1989. V(D)J recombination: a functional definition of the joining signals. Genes Dev. 3, 1053–10761.

Houston, E.F., Connelley, T., Parsons, K., MacHugh, N.D., Morrison, W.I., 2005. Analysis of T-cell receptor BV gene sequences in cattle reveals extensive duplication within the BV9 and BV20 subgroups. Immunogenetics 57, 674–681.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 7, 754–755.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.

Kabat, E.A., Wu, T.T., 1991. 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. 147, 1709–1719.

Kamper, S.M., McKinney, C.E., 2002. Polymorphism and evolution in the constant region of the T-cell receptor β chain in an advanced teleost fish. Immunogenetics 53, 1047–1054.

Kerfourn, F., Charlemagne, J., Fellah, J.S., 1996. The structure, rearrangement, and ontogenic expression of DB and JB segments of the Mexican axolotl T-cell antigen receptor β chain (TCRB). Immunogenetics 44, 275–285.

Kjer-Nielsen, L., Clements, C.S., Purcell, A.W., Brooks, A.G., Whisstock, J.C., Burrows, S.R., McCluskey, J., Rossjohn, J., 2003. A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. Immunity 18, 53–64.

Lanave, C., Preparata, G., Saccone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20, 86–93.

Lefranc, M.P., Pommie, 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.

McCormack, W.T., Tjoelker, L.W., Stella, G., Postema, C.E., Thompson, C.B., 1991. Chicken T-cell receptor β-chain diversity: an evolutionarily conserved Dβ-encoded glycine turn within the hypervariable CDR3 domain. Proc Natl Acad Sci USA 88, 7699–7703.

Nicholas, K.B., Nicholas, H.B., Deerfield, D.W., 1997. GeneDoc: analysis and visualization of genetic variation. EMBNEW.NEWS 4, 14.

Ramsden, D.A., Baetz, K., Wu, G.E., 1994. Conservation of sequence in recombination signal sequence spacers. Nucl. Acids Res. 137, 5–33.

Rudikoff, S., Fitch, W.M., Heller, M., 1992. Exon-specific gene correction (conversion) during short evolutionary periods: homogenization in a two-gene family encoding the beta-chain constant region of the T-lymphocyte antigen receptor. Mol. Biol. Evol. 9, 14–26.

Saccone, C., Lanave, C., Pesole, G., Preparata, G., 1990. Influence of base composition on quantitative estimates of gene evolution. Meth. Enzymol. 183, 570–583.

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.

Shigeta, A., Sato, M., Kawashima, T., Horiuchi, H., Matsuda, H., Furusawa, S., 2004. Genomic organization of the chicken T-cell receptor beta chain D-J-C region. J. Vet. Med. Sci. 66, 1509–1515.

Sleckman, B.P., Bassing, C.H., Hughes, M.M., Okada, A., D'Auteuil, M., Wehrly, T.D., Woodman, B.B., Davidson, L., Chen, J., Alt, F.W., 2000. Mechanisms that direct ordered assembly of T cell receptor beta locus V, D, and J gene segments. Proc. Natl. Acad. Sci. USA 97, 7975–7980.

Takeda, J., Cheng, A., Mauxion, F., Nelson, C.A., Newberry, R.D., Sha, W.C., Sen, R., Loh, D.Y., 1990. Functional analysis of the murine T-cell receptor beta enhancer and characteristics of its DNA-binding proteins. Mol. Cell. Biol. 10, 5027–5035.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22, 4673–4680.

Tonegawa, S., 1983. Somatic generation of antibody diversity. Nature 302, 575–581.

Vaiman, D., Billault, A., Tabet-Aoul, K., Schibler, L., Vilette, D., Oustry-Vaiman, A., Soravito, C., Cribiu, E.P., 1999. Construction and characterization of a sheep BAC library of three genome equivalents. Mamm. Genome 10, 585–587.

Watanabe, M., Iwasaki, Y., Mita, Y., Ota, S., Yamada, S., Shimizu, M., Takagaki, Y., 2007. Porcine T-cell receptor beta-chain: a genomic sequence covering Dbeta1 1 to Cbeta2 gene segments and the diversity of cDNA expressed in piglets including novel alternative splicing products. Mol. Immunol. 44, 2332–2343.

Wermenstam, N.E., Pilstrom, L., 2001. T-cell antigen receptor in Atlantic cod (*Gadus morhua* L.): structure, organisation and expression of TCR α and β genes. Dev. Comp. Immunol. 25, 117–135.

Whitehurst, C.E., Schlissel, M.S., Chen, J., 2000. Deletion of germline promoter PDβ 1 from the TCR β locus causes hypermethylation that impairs Dβ1 recombination by multiple mechanisms. Immunity 13, 703–714.

Wu, C., Bassing, C.H., Jung, D., Woodman, B.B., Foy, D., Alt, F.W., 2003. Dramatically increased rearrangement and peripheral representation of Vbeta14 driven by the 3′Dbeta1 recombination signal sequence. Immunity 18, 75–85.

Yang, X.O., Doty, R.T., Hicks, J.S., Willerford, D.M., 2003. Regulation of T-cell receptor D beta 1 promoter by KLF5 through reiterated GC-rich motifs. Blood 101, 4492–4499.

Zhou, H., Bengten, E., Miller, N.W., Clem, L.W., Wilson, M., 2003. The T cell receptor beta locus of the channel catfish, *Ictalurus punctatus*, reveals unique features. J. Immunol. 170, 2573–2581.