Immunogenetics (2005) 57: 254–266
DOI 10.1007/s00251-005-0773-7

ORIGINAL PAPER

R. Antonacci · C. Lanave · L. Del Faro · G. Vaccarelli · S. Ciccarese · S. Massari

Artiodactyl emergence is accompanied by the birth of an extensive pool of diverse germline *TRDV1* genes

Received: 16 October 2004 / Revised: 21 December 2004 / Published online: 9 March 2005

Abstract Molecular cloning of cDNA from γ/δ T cells has shown that in sheep, the variable domain of the δ chain is chiefly determined by the expression of the *TRDV1* subgroup, apparently composed of a large number of genes. There are three other *TRDV* subgroups, but these include only one gene each. To evaluate the extent and the complexity of the genomic *TRDV* repertoire, we screened a sheep liver genomic library from a single individual of the Altamurana breed and sheep fibroblast genomic DNA from a single individual of the Gentile di Puglia breed. We identified a total of 22 *TRDV1* genes and the *TRDV4* gene. A sequence comparison between germline and the rearranged genes indicates that, in sheep, the *TRDV* repertoire is generated by the *VDJ* rearrangement of at least 40 distinct *TRDV1* genes. All germline *TRDV1* genes present a high degree of similarity in their coding as well as in 5′ and 3′ flanking regions. However, a systematic analysis of the translation products reveals that these genes present a broadly different and specific repertoire in the complementarity-determining regions or recognition loops, allowing us to organize the *TRDV* genes into sets. We assume that selection processes operating at the level of ligand recognition have shaped the sheep *TRDV* germline repertoire. A phylogenetic study based on a sequence analysis of the *TRDV* genes from different mammalian species shows that the diversification level of these genes is higher in artiodactyl species compared to humans and mice.

R. Antonacci · L. Del Faro · G. Vaccarelli · S. Ciccarese
Dipartimento di Anatomia Patologica e di Genetica,
University of Bari,
Bari, Italy

C. Lanave
Istituto di Tecnologie Biomediche Sezione di Bari, CNR,
Milan, Italy

S. Massari (✉)
Dipartimento di Scienze e Tecnologie Biologiche ed
Ambientali, University of Lecce,
Via per Monteroni - Centro Ecotekne,
73100 Lecce, Italy
e-mail: sara.massari@unile.it
Tel.: +39-832-298950

Keywords T-cell receptor · *TRDV* genes · Sheep · Evolution · Artiodactyl

Introduction

T lymphocytes can be divided into two subsets according to the specificity of their heterodimer antigen receptor. The first subset of lymphocytes expresses receptors made of α and β polypeptide chains, while the second subset expresses receptors made of γ and δ chains. Each of the T-cell receptor polypeptides is encoded by a specific set of separate genes: α and γ chains are encoded by variable (*V*), joining (*J*) and constant (*C*) genes, whereas β and δ polypeptides are encoded by the same type of genes plus diversity (*D*) genes (Lefranc and Lefranc 2001). For each chain, there are only a few copies of *D* and *J* genes, while the number of *V* genes is much larger. Based on nucleotide sequence similarity, the *V* genes at each locus can be classified into subgroups. Therefore, the *V* genes represent a typical example of multigene family. The *V* and *J* or *V*, *D*, and *J* genes are juxtaposed by somatic rearrangement during T-cell differentiation. After transcription, the *V*(*D*)*J* sequence is spliced to the constant region. The resulting chain is a protein with the variable domain composed of seven distinguishable regions: three antigen-recognition loops or complementarity-determining regions (CDRs) and four framework regions (FRs). Two of the CDR loops, CDR1 and CDR2, are encoded by the *V* region. The third CDR loop (CDR3) reflects the ability of the *V* gene to rearrange to any (*D*)*J* gene. In all mammals studied so far, the large arrays of genes that code for β and γ chains are localized at two distinct loci (*TRB* and *TRG*), whereas the genes that code for α and δ chains are found in a single chromosomal region (*TRA/TRD*). In particular, the genes coding for δ chain (*TRDV*, *TRDD*, *TRDJ*, and *TRDC* genes) are embedded between the *V* and *J* genes coding for α chain (*TRAV* and *TRAJ* genes). In humans, there are three *TRDV* subgroups (*TRDV1*, *TRDV2*, and *TRDV3*), each of which contains only one member gene (Lefranc and Lefranc 2001; Scaviner and Lefranc 2000). The *TRDV1* gene is interspersed among

TRAV genes, while the *TRDV2* gene is located at the 3′ end of the array of *TRAV* genes, upstream of the three *TRDD*, the four *TRDJ*, and the single *TRDC* genes. The *TRDV3* gene is located downstream from the *TRDC* gene in an opposite transcriptional orientation. Moreover, in humans, five *TRAV* genes belonging to five different subgroups have been found rearranged to either *TRAJ* or *TRDD* genes and can therefore be used in the synthesis of α and δ chains. For this reason, they are designated *TRAV/DV* genes (Scaviner and Lefranc 2000). In mice, four specific functional *TRDV* subgroups (*TRDV1*, *TRDV2*, *TRDV4*, and *TRDV5*) have been found. Each subgroup contains one member gene except for *TRDV2* with two closely related genes (*TRDV2-1* and *TRDV2-2*). There is also a *TRDV3* subgroup consisting of a pseudogene (Bosc and Lefranc 2003). As in humans, the typical *TRDV* genes are predominantly located at the 3′ end of the array of variable genes, while the *TRDV5* gene is located downstream from the respective C gene in an opposite transcriptional orientation. Ten *TRAV* genes belonging to seven subgroups have been found rearranged to either *TRAJ* or *TRDD* genes. Among these *TRAV/DV* genes, two genes (*TRAV15-2/DV6-2* and *TRAV15D-2/DV6D-2*) belonging to *TRAV15/DV6* subgroup have been assigned to this set on the basis of sequence alignment, although they have been found so far rearranged only to *TRDD* genes (Bosc and Lefranc 2003).

The γ/δ T cells show different frequencies and physiological distributions in the diverse species. In humans and mice, γ/δ T lymphocytes account for less than 5% of the peripheral T-cell pool (“ γ/δ low” species). In contrast, in artiodactyls (Binns et al. 1992; Hein and MacKay 1991) and chickens (Cooper et al. 1991), γ/δ T cells correspond to one half of the peripheral T-cell pool (“γ/δ high” species).

Molecular cloning of cDNA from T cells of γ/δ high species has shown that the variable domain of the δ chain is formed from the preferential utilization of one subgroup (*TRDV1*), consisting of a large number of genes with different sequences (Hein and Dudler 1993, 1997; Kubota et al. 1999; Ishiguro et al. 1993; Massari et al. 2000; Yang et al. 1995). In artiodactyls, this predominant subgroup is related to the unique gene of the human *TRDV1* subgroup and to the mouse *TRAV15-2/DV6-2* and *TRAV15D-2/DV6D-2* genes. Three more subgroups, *TRDV2*, *TRDV3*, and *TRDV4*, are also present in sheep, but these include only one gene each (Hein and Dudler 1993). In the same way, besides the large *TRDV1* subgroup, four more subgroups, *TRDV2*, *TRDV3*, *TRDV4*, and *TRDV5*, were found in pigs, and these consist of a limited number of genes: two close genes were identified for the *TRDV3* and *TRDV4* subgroups and only one for both the *TRDV2* and *TRDV4* subgroups (Yang et al. 1995). At this time, the genomic organization of the *TRD* locus in artiodactyls is not well known, and the features of the germline *TRDV* repertoire have not yet been determined. We have begun to investigate the germline *TRDV* repertoire in a representative γ/δ high species, the sheep. Using various methods, we obtained a representative pool of sheep germline *TRDV1* genes as well as the germline *TRDV4* gene. We performed structural and phylogenetic analysis of the isolated *TRDV* genes to provide a genomic perspective for understanding *TRDV* gene evolution in γ/δ high species.

Materials and methods

Isolation of germline *TRDV1* genes by screening of a genomic library

A liver genomic library from adult sheep (Altamurana breed), consisting of 5.4×10^6 pfu with insert size of 9–23 kb, was screened by plaque hybridization, under low stringency condition, probing with the 205-bp *Eco*RI–*Sca*I fragment of pBVCD8 clone corresponding to the *TRDV1S1* gene (Massari et al. 2000). After three rounds of screening, 14 positive clones were recovered and further characterized by Southern blot analysis with the same probe on phage DNA digested with *Eco*RI, *Hin*dIII, *Bam*HI, and *Xba*I restriction enzymes. Two different approaches were used for the isolation of *TRDV* genes from the different phage clones. For λD10, λD13, λD14, λD15, λD21, λD26, λD29, λD30, λD33, λD34, and λD37 clones, the fragments that hybridized to the pBVCD8 probe were subcloned in a pZero cloning vector (Invitrogen) and sequenced. The variable genes of λD1, λD2, and λD4 phage clones were obtained by PCR amplification with primers derived from conserved regions of the previously isolated *TRDV* genomic sequences. In particular, one primer (LC1) of 20 nucleotides corresponds to the upstream conserved sequence around the ATG codon (5′-CTCATGCCGCTCTCCAGTCT-3′, positions 343–362 from Fig. 2, accession no. AJ786827). The other primer (LC2) of the same size anneals to the downstream conserved nonamer sequence (5′-GGGTTTTTGTACAGCTTTTC-3′, positions 917–936 from Fig. 2, accession no. AJ786827). PCRs were performed with 50 ng phage DNA with 2.5 U of *Taq* DNA polymerase (Applied Biosystems). We used the following conditions: 30 cycles of 30 s at 94°C, 30 s at 55°C, and 30 s at 72°C with a final incubation of 10 min at 72°C. PCR products were sequenced directly.

Isolation of *TRDV1* germline genes by DNA amplification

Genomic DNA was extracted from sheep fibroblasts (Gentile di Puglia breed) by standard procedure and amplified with the primer pairs from variable region described in the previous paragraph. To clone PCR products, an *Xba*I-recognized sequence was added at the 5′ end of each primer. We used PCR experimental conditions, already described, with 100 ng genomic DNA. PCR products were gel-purified, digested with *Xba*I restriction enzyme, ligated into the appropriate digested pZero cloning vector, and transformed into TOP10 cells. Randomly chosen clones with inserts of the expected size were sequenced. Two sets of PCR and cloning were carried out independently to exclude possible errors.

Table 1 Classification of *TRDV1* genes

gDNA		cDNA		*TRDV Gene (e)*
Altamurana breed (a)	Gentile di Puglia breed (b)	Blood (c)	Thymus (d)	
λD29	pZVD91	Z12970		*TRDV1S1*
		Z12981		*TRDV1S2*
		Z12988		*TRDV1S3*
		Z12989		*TRDV1S4*
λD26	pZVD61	Z12990		*TRDV1S5*
		Z12991		*TRDV1S6*
λD4	pZVD14	Z12992		*TRDV1S7*
λD30	pZVD84	Z12993		*TRDV1S8*
λD14(B)	pZVD46	Z12994		*TRDV1S9*
λD37	pZVD100	Z12971		*TRDV1S10*
	pZVD86	Z12972		*TRDV1S11*
		Z12973		*TRDV1S12*
λD15(B)		Z12974		*TRDV1S13*
		Z12975		*TRDV1S14*
		Z12976		*TRDV1S15*
λD33	pZVD65	Z12977		*TRDV1S16*
		Z12978		*TRDV1S17*
λD13		Z12979		*TRDV1S18*
		Z12980		*TRDV1S19*
		Z12982		*TRDV1S20*
		Z12983		*TRDV1S21*
		Z12984		*TRDV1S22*
		Z12985		*TRDV1S23*
		Z12986		*TRDV1S24*
		Z12987		*TRDV1S25*
			AJ290092	*TRDV1S26*
			AJ290081	*TRDV1S27*
			AJ290093	*TRDV1S28*
λD2			AJ290085	*TRDV1S29*
λD1				***TRDV1S30***
λD15(A)	pZVD63			***TRDV1S31***
λD14(A)				***TRDV1S32***
λD21	pZVD57			***TRDV1S33***
λD34				***TRDV1S34***
	pZVD5			***TRDV1S35***
	pZVD47			***TRDV1S36***
	pZVD60			***TRDV1S37***
	pZVD92			***TRDV1S38***
	pZVD105			***TRDV1S39***
	pZVD110			***TRDV1S40***

[a]Genomic clones from Altamurana breed
[b]Genomic clones from Gentile di Puglia breed
[c]Accession nos. of cDNA derived from blood
[d]Accession nos. of cDNA derived from thymus
[e]Name of *TRDV* genes; new genes are indicated in *boldface*

Isolation of *TRDV4* gene

To isolate the sheep genomic *TRDV4* gene located downstream from the *TRDC* gene, we subcloned in pZero cloning vector, the terminal *Eco*RI fragment of λMCD phage clone previously isolated from the same genomic library (Massari et al. 2000). The obtained plasmid clone was sequenced.

Table 2 Classification of the *TRDV1* sets

Set	CDR1 (a)	CDR2 (b)	CRD3 (c)	*TRDV* genes
1A	9	Y or basic	+/−	*-1S9*, *-1S18*, *-1S30*, *-1S32*, *-1S38*, *-1S39*, *-1S35*
1B	7	Y	−	*-1S5*, *-1S7*, *-1S8*, *-1S13*, *-1S36*
1C	7	G or hydroxyl	−	*-1S1*, *-1S10*, *-1S11*, *-1S37*, *-1S40*
1D	9	Acidic	+	*-1S16*, *-1S33*
1E	7	Amide	+	*-1S29*[d], *-1S31*, *-1S34*

[a]Amino acid length of CDR1
[b]Class of amino acid at position 57
[c]Presence of a Trp at the 107 position
[d]CDR2 is absent

Sequence analysis and drawing of dendrograms

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).

We used a total of 115 *TRDV* genes for the phylogenetic analysis. Since we used various genes from different organisms, we designated each of them by the gene notation as proposed by IMGT, the international ImMunoGeneTics information system, http://www.imgt.cine.fr (Lefranc 2003) plus the specie's Latin names.

The accession numbers of the *TRDV* genes used are the following:

- *Homo sapiens* (human): *TRDV1* (M22198), *TRDV2* (X15207), and *TRDV3* (M23326).
- *Mus musculus* (mouse): *TRDV1* (AE008686), *TRDV2-1* (AE008686), *TRDV4* (AE008686), *TRDV5* (AE008686), and *TRAV15/DV6-2* (AF085010).
- *Sus scrofa* (pig): *TRDV1S1–1S13* (D49564–76), *TRDV2* (D49579), *TRDV3S1* (D49580), *TRDV4* (D49582), and *TRDV5S1* (D49583).
- *Bos taurus* (cattle) *TRDV1S1–S2* (D13655–56), *TRDV1S3* (D16112), *TRDV1S4* (D13660), *TRDV1S5* (D13657), *TRDV1S6* (D16115), *TRDV1S7* (D16114), *TRDV1S8* (D16116), *TRDV1S9* (D13658), *TRDV1S10* (D13661), *TRDV1S11* (D16113), *TRDV1S12* (D13659), *TRDV1S13–1S17* (U73380–84) and *TRDV1S19–1S22* (U73386–89).
- *Ovis aries* (sheep): *TRDV1S1* (AJ786827), *TRDV1S2* (Z12981), *TRDV1S3* (Z12988), *TRDV1S4* (Z12989), *TRDV1S5* (AJ786828), *TRDV1S7-1S10* (AJ786829–AJ786832), *TRDV1S11* (AJ809501), *TRDV1S12* (Z12973), *TRDV1S13* (AJ786833), *TRDV1S14–1S15* (Z12975–76), *TRDV1S16* (AJ786834), *TRDV1S18* (AJ786835), *TRDV1S19–1S21* (Z12980–Z12982–Z12983), *TRDV1S24–1S25* (Z12986–87), *TRDV1S29–*

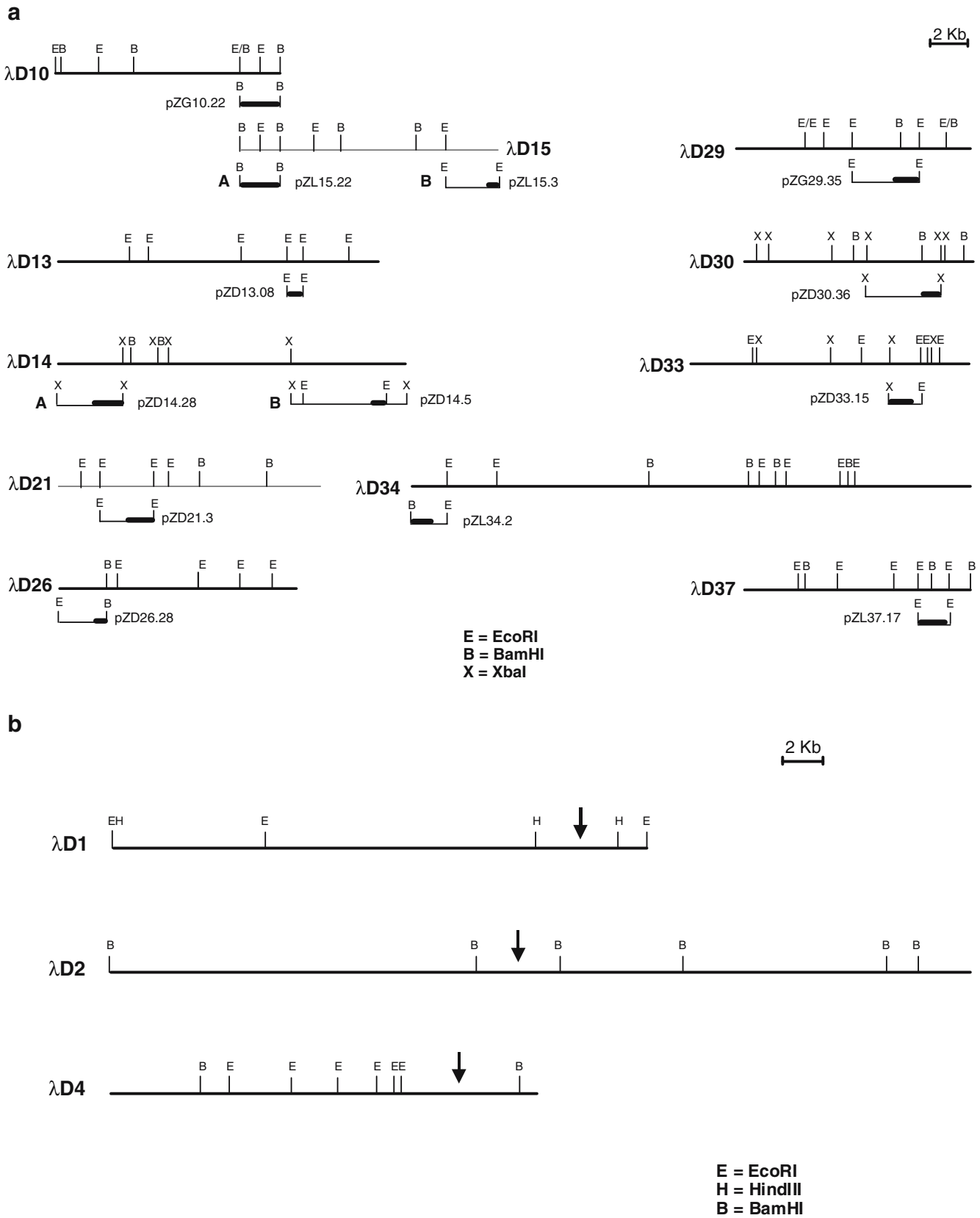


Fig. 1 Names and restriction maps of the phage clone inserts isolated from the sheep genomic library (Altamurana breed) and their derived plasmid subclones. **a** Phage clones in which fragments with the *TRDV* gene were subcloned into plasmids. **b** Phage clones in which the *TRDV* genes were isolated by PCR. *Arrows* indicate the *TRDV* regions. **c** Phage clone containing *TRDV4* (this study), *TRDJ2*, and *TRDC* genes (Massari et al. 2000). The direction of transcription is indicated by *arrows*. *Thick bars* indicate the sequenced regions of the subclones

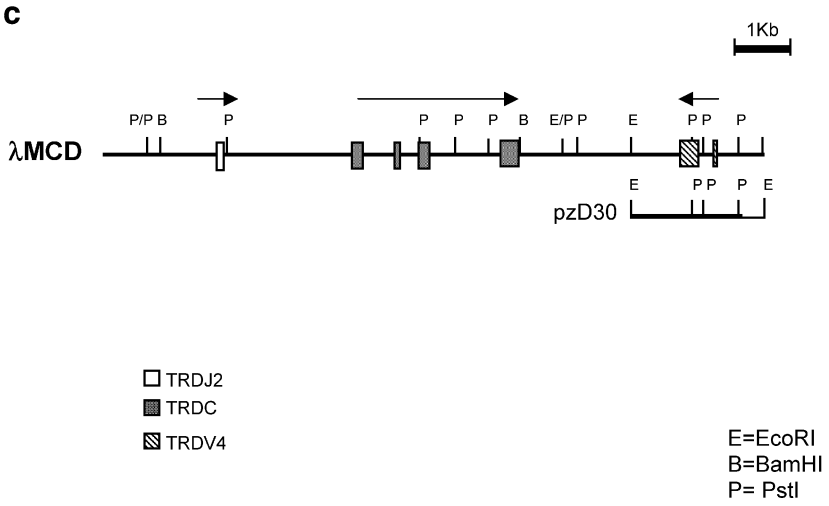


Fig. 1 (continued)

1S34 (AJ786836–AJ786841), *TRDV1S35–TRDV1S40* (AJ809502–AJ809507), *TRDV2* (Z12995), *TRDV3* (Z12996), and *TRDV4* (AJ810117).

– *Gallus gallus* (chicken): *TRDV1S1–1S29* (AF175435–AF175463).

For the phylogenetic analysis of the sheep *TRDV1* subgroup, we used only the amino acid sequences of *TRDV1* genes whose available coding regions were complete. For this reason, the total numbers of sequences analyzed were 33 instead of 40 as described in Table 1. In the analysis, we included the coding regions of human *TRDV1* and mouse *TRAV15/DV6* genes to root the phylogenetic tree.

In the evolutionary analysis of *TRDV* genes within mammals, we used both the amino acid and the nucleotide sequences of the functional genes. In the tree, we have included all available sequences of *TRDV* genes from humans, mice, cattle, and pigs, after excluding those that are closely related and using only some representative sheep *TRDV1* genes. In fact, our previous analysis showed that all sheep *TRDV1* genes can be classified into five sets (Table 2). We have therefore used 15 representative germline sequences covering the five sets (Table 1, lane a). For the purpose of rooting the tree, we included 29 *TRDV* chicken genes.

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 analyses were carried out on gap-free alignments. The phylogenetic tree based on the aligned amino acid sequences 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 analysis. Other models were applied to draw the phylogenetic tree: maximum parsimony and minimum evolution based on the distances with Poisson correction and with the gamma model in MEGA, version 2.1 (Kumar et al. 2001). The values reported on the branch of the phylogenetic tree are the bootstrap numbers obtained by 100 replicates of simulation. Only the first plus second codon positions were taken into account in nucleotide analyses, and analyzed using the Markov model, also known as the GTR model in PAUP* package (Swofford 1998), after checking the stationarity in base composition (Saccone et al. 1990).

Results

Isolation of sheep germline *TRDV1* genes

To evaluate the germline *TRDV1* gene repertoire in sheep, we screened a genomic library from the Altamurana breed with pBVCD8 cDNA probe containing the *TRDV1S1* gene (see "Materials and methods"). A total of 14 positive phage clones were isolated and further characterized. Restriction maps of all phage clones were produced by digesting clonal DNA with *Eco*RI, *Bam*HI, *Hin*dIII, and *Xba*I restriction enzymes, followed by a Southern blot hybridization with the pBVCD8 probe. The maps showed that the clones are different with genomic inserts ranging from 12 kb to 40 kb (Fig. 1a, b). Each clone contains only one *TRDV* gene, except for λD14 and λD15, each of which contains two genes. These regions were designated λD14(A), λD14(B), λD15(A), and λD15(B), respectively. λD10 and λD15 are overlapping clones, with the 3′ end of λD10 in common with the λD15(A) region. To recover the *TRDV* genes from the clones, we used two different approaches. For λD10, λD13, λD14, λD15, λD21, λD26, λD29, λD30, λD33, λD34, and λD37 (Fig. 1a), all fragments which hybridized to the probe were subcloned and sequenced. These sequences were scanned for genes that might qualify as *TRDV* genes on the basis of length, presence of the 3′ recombination signal sequences (RSSs) and homology to published *TRDV* genes. As an example, the 1,449 bp of the pZG29.35 fragment of the λD29 clone is shown in Fig. 2. The *TRDV* coding region is 49+295 bp long and begins with an ATG initiation codon. This ATG starts an open reading frame that is interrupted, after 49 bp, by an intron of 207 bp. Characteristic RSSs located downstream from the coding

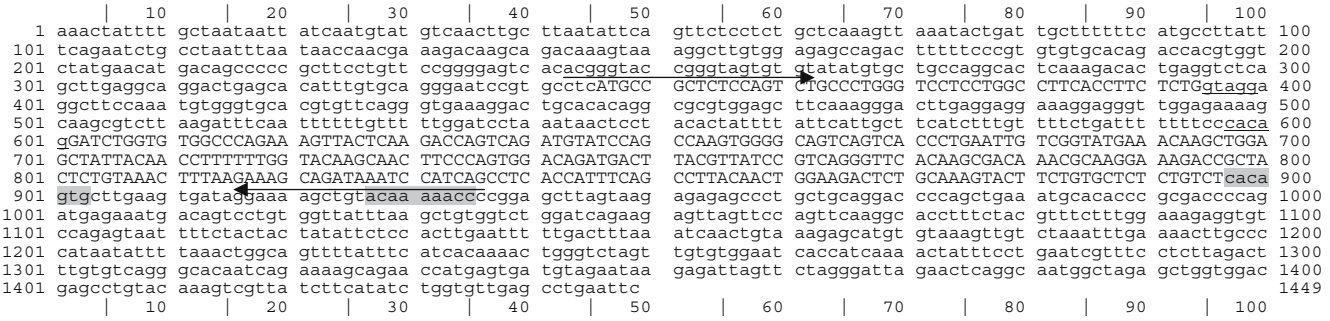


Fig. 2 Nucleotide sequence of the *TRDV1S1* gene (accession no. AJ786827) found in the λD29 genomic clone. The coding region is in *capital letters*. The donor and acceptor splice sites are *underlined*. The heptamer and nonamer within the 3′ recombination signal sequence are *highlighted*. The *arrows* indicate the nucleotide sequences corresponding to the LC1 and LC2 primer pair used for PCR

region are identical to the consensus sequences CACAG TG/ACAAAAACC (Hess et al. 1989). The spacer region between the heptamer and nonamer is 23 nucleotides long. This gene organization is similar for all the *TRDV* genes identified in every phage clone. With this method we collected 12 different *TRDV* genes; the gene of λD10 turned out to be identical to that of λD15(A). The *TRDV* genes of λD1, λD2 and λD4 phage clones shown in Fig. 1b were recovered by direct sequencing of PCR fragments constructed with a primer pair derived from conserved regions of the *TRDV* genomic sequences (see "Materials and methods"). By using this alternative approach, we picked up the coding regions of three more different *TRDV* genes identified against database entries. In total, we retrieved 15 different *TRDV* genes (Table 1, lane a). When we compared the coding sequence of all the *TRDV* genes with each other, the level of DNA identity ranged from 78% to 97% (data not shown). The maximum level of similarity was between λD15(A) and λD26, with only ten nucleotide differences over 346 bp (97.1%). According to the criteria that establish a nucleotide similarity of more than 75% within members of the same subgroup, all the genes were grouped in the same *TRDV1* subgroup. The nucleotide sequences of all 15 *TRDV1* genes have been submitted to the EMBL databank (accession nos. AJ786827–AJ786841).

To expand the ovine *TRDV1* database, germline variable genes were also obtained by cloning and sequencing PCR fragments of *TRDV* genes derived from genomic DNA of a single sheep belonging to the Gentile di Puglia breed (see "Materials and methods"). With this method we obtained 54 analyzable sequences. Eight of them, containing a stop codon, could be pseudogenes and were excluded from further analyses. When we compared the remaining 46 sequences with each other, the level of DNA similarity ranged from 79% to 99%. Only the sequences with a level of similarity <97% were considered to represent distinct genes. In this way we picked up 16 new germline *TRDV1* genes (Table 1, lane b). Comparing the two germline collections from the Altamurana and Gentile di Puglia breeds, we found nine sequences in common, with a level of nucleotide similarity from 99% to 98.6% (Table 1, lanes a, b). This minimal level of nucleotide substitution between common sequences is consistent with allelic variation. The nucleotide sequences of the 16 new germline *TRDV1* genes derived from Gentile di Puglia breed have also been submitted to the EMBL databank (accession nos. AJ809501–AJ809507, AJ868217–AJ868225).

Classification of *TRDV1* genes

To determine the sequence similarity, the 22 distinct germline *TRDV* genes were compared with a cDNA collection for δ chains derived from thymus (Massari et al. 2000) and peripheral blood (Hein and Dudler 1993). The genes from the λD2, λD4, λD13, λD14(B), λD15(B), λD26, λD30, λD33, λD37, and pzVD86 clones exhibited a 99% nucleotide similarity with the variable region of the corresponding cDNA, called, respectively, TRDV1S29, *TRDV1S7*, *TRDV1S18*, *TRDV1S9*, *TRDV1S13*, *TRDV1S5*, *TRDV1S8*, *TRDV1S16*, *TRDV1S10*, and *TRDV1S11*, according to the IMGT provisional nomenclature (Table 1). For λD29, we found a 97.6% nucleotide similarity with the cDNA corresponding to the *TRDV1S1* gene (eight nucleotide differences over 341 bp). Thus, 11 out of 22 germline sequences were assigned to their corresponding cDNA. The minimal number of nucleotide substitutions observed between the genomic and cDNA sequences could be due to allelic variation. The remaining five genes of the phage clones λD1, λD14(A), λD15(A), λD21, and λD34, and six of the plasmid clones pzVD5, pzVD47, pzVD60, pzVD92, pzVD105, and pzVD110, showing a nucleotide homology with cDNA <97%, were considered to represent new genes of the *TRDV1* subgroup and were numbered consecutively *TRDV1S30–TRDV1S40* according to the IMGT provisional nomenclature (Table 1). A total of 22 distinct germline *TRDV1* genes were isolated in this study, and only one half were found to be common to lymphocyte and thymocyte RNA. Moreover, 18 *TRDV1* genes, previously isolated as cDNA, have not yet been matched to a germline sequence (Table 1). Therefore, we inferred that the sheep genome is characterized by the presence of a large number of diverse and specific *TRDV1* genes.

Isolation and characterization of sheep germline *TRDV4* gene

In all mammalian species studied so far a variable gene, belonging to a *TRDV* subgroup different from *TRDV1*, has been found at the 3′ end of the *TRDC* gene in an inverse transcriptional orientation relative to the other *TRD* genes (Bosc and Lefranc 2003; Lefranc and Lefranc 2001). To establish if this is also present in the sheep genome, we subcloned and sequenced the 3′ end *Eco*RI fragment of the phage clone λMCD (Fig. 1c), isolated previously from the same genomic library and containing the *TRDJ2* and *TRDC* regions (Massari et al. 2000). In the first 2,139 bp of the pzD30 subclone, we identified a region showing a high degree of nucleotide similarity with the sheep *TRDV4* gene already isolated as cDNA (Hein and Dudler 1993). The comparison revealed that the gene lies in an inverse transcriptional orientation within a region of 585 bp, with the first exon (*L-PART1*) of 37-bp long separated from the *V-EXON* by an intron of 239 bp. In the 3′ region of the gene, the heptamer and nonamer recognition signals are present, separated by 23 nucleotides. The nucleotide sequence of the *TRDV4* gene has been submitted to the EMBL databank (accession no. AJ810117).

Fig. 3 Alignment of the deduced amino acid sequences of the germline *TRDV1* genes. The names of the genomic sequences, in accordance with the international ImMunoGeneTics information system (IMGT) nomenclature, are listed at the *left*. Leader sequence, CDR-IMGT complementarity determining regions (*CDR-IMGT*) and framework regions (*FR-IMGT*) are also indicated. *Dashes* indicate amino acids identical to the *TRDV1S1* gene (*first line*). Gaps were inserted to align all sequences and are indicated by *dots*. The consensus sequence for the *TRDV1* subgroup is shown in the *bottom line*. The amino acids with 9% (two of 22) or fewer variations were considered consensus; those considered non-consensus are indicated by an *asterisk*

```
TRDV                                        FR1-IMGT             CDR1-IMGT
gene             leader                     (1-26)                (27-38)
                                                                  
                               1.......10........20...... ..30........ 40...
                               |        |         |          |           |
1S1         MPLSSLPWVLLAFTFSGSGVAQKVTQDQSDVSSQVGQSVTLNCRYE TSWSYYN..... LFWYK
1S5         -Q----F---------A-C-------N-P-I--------------- ------Y..... -V---
1S7         ------L--F---S--A-----E---NRP-I------PIA------ ------Y..... -V---
1S8         ------L--F----------------------G------------- ---GA-Y..... -----
1S9         ------L---------------S-----P-IL-E--KT-------- ---DMFYYH... I----
1S10        ------L----------------------------------S---- ----V-Y..... -Y---
1S11        ----------------------------------L------S---- ----V-Y..... -Y---
1S13        -R----L---------A-C-------N-P-I------------W-- ----D-Y..... FV---
1S16        ------L--F------------------PYIT--I----I---Q-- V---GYMHY... -----
1S18        ------L---------------S-----PVII-E--KTI------- ---DMYTYW... I----
1S29        ------L--F--------S-----I---P-IP-R--E--------- --R-S-S..... I----
1S30        ------L--F--S------A--------P-ITR--E-------L-- V--YMDAYS... I--F-
1S31        ------L---------------------P-I-----E-A-M--Q-- ---RS--..... S-S--
1S32        ------L--F--------S-S-----Y-P-IT---RNV-I------ ---NNYTYW... I----
1S33        ------L---------------------PYIT--T----I------ VR--GYTHY... -----
1S34        -L----F--F------------------P-I-----E-AAM--Q-- ---RS--..... S----
1S35        ------L--F--------S---Q---V-TA-------T-------- V--TMDYYY... I----
1S36        ------L--F------------------PHI--------------- ----L-Y..... -----
1S37        ----------------------------------L----------- ---R---..... --C--
1S38        ------L--F--------S-------Y-P-IT---WNV-I---Q-- ---NVYTYW... I----
1S39        ---------------------------------------------- ---NMLSYY... I----
1S40        ------------------C--------------------------- ----A-Y..... -Y---

consensus   M*LSSL*WV*LAFTFS*S*VAQ*VTQ*Q****S*******LNC*YE *SW**Y****** **WYK
```

```
               FR2-IMGT     CDR2-IMGT                FR3-IMGT                  CDR3-IMGT
               (39-55)       (56-65)                 (66-104)                  (105-109)

            .....50..... ...60..... ...70........80........90.......100.... .....
                 |          |          |         |         |         |
1S1         QLPSGQMTYVIR QGS....... QATNARK.DRYSVNFKKADKSISLTISALQLEDSAKYFC ALC.. [7.3.3]
1S5         ------I--L-Q -Y-....... ENS---N.G------Q--------I--S----------- --RD. [7.3.4]
1S7         ---------L-Q -Y-....... GYS---N.G---------------I--S----------- ---D. [7.3.4]
1S8         ---------L-Q -Y-....... GNS---N.G------Q-----------S----------- --RE. [7.3.4]
1S9         ---------L-- -Y-....... DDG---D.G------Q--H----------E-----M--- V-RV. [9.3.4]
1S10        ------------ -S-....... E---E--.E-------------K---------------- --SLS [7.3.5]
1S11        -----------H ---....... -V-----.------------------------------- --SV. [7.3.4]
1S13        ---------L-Q -Y-....... EYS---D.G---I--Q--------I--S----------- ---D. [7.3.4]
1S16        -----EI-FL-- -E-....... SGP---N.G------QR-QN----------P----M--- -VWE. [9.3.4]
1S18        --S------L-- -Y-....... EGG---D.G------QE-H----------E-----M--- V-RA. [9.3.4]
1S29        H----E-IFLT. .......... .....-D.G---I-SARSR--S-----T----------- V-WE. [7.0.4]
1S30        ---R---I-L-- -Y-....... EDG---D.G------Q-----------------T----- V-VN. [9.3.4]
1S31        -----E-I-L-G -N-....... YSP---D.G---I--QRSR-A--------K--------- --WE. [7.3.4]
1S32        -----E---L-P -Y-....... EDG-E-D.GH-----QN-H-F------S-K--H-G---- --WE. [9.3.4]
1S33        -----E--FL-- -E-....... SGL---N.G------QR-QNT------------------ -VWE. [9.3.4]
1S34        -F---E-I-L-G -N-....... YSP---D.G---IHCQSQG-A---I----K--------- --WE. [7.3.4]
1S35        ---R-E--LL-H -Y-....... EYS---N.G------Q--H-----I------K------- --WE. [9.3.4]
1S36        ---------L-Q -Y-....... EYS---D.G------Q--------I--S----------- ---D. [7.3.4]
1S37        ------------ ---....... -V-----.E------------------------------ --SV. [7.3.4]
1S38        ---R-EI--L-H -Y-....... EDG-E-D.GH-----QE-H-F------S-K--H-G---- --WE. [9.3.4]
1S39        ---------L-- -H-....... DDG---D.G------Q--H----------E--------- V-RV. [9.3.4]
1S40        -----------H ---....... EVI---E.------------------------------- ---D. [7.3.4]

consensus   QLP*G*****I* Q*S******* ***N*R***RYS*NF****K*ISL*IS*L*LEDSA*YFC *L***
```

Structural analysis of the *TRDV* coding regions

The deduced amino acid sequences of all sheep germline *TRDV1* genes were manually aligned according to the IMGT unique numbering for the *V-REGION* and *V-DOMAIN* (Lefranc et al. 2003) to maximize homology (Fig. 3). On the basis of the comparison, we established a consensus sequence for the *TRDV1* subgroup in which the amino acids with about 9% (2 of 22) or fewer variations were considered consensus. All the sequences exhibit typical features, i.e., a leader region of 20 amino acids with most conserved residues, the preserved Cys23 in FR1-IMGT and CONSERVED-TRP 41 in FR2-IMGT. In all cases, the FR3-IMGT interval is 38 amino acids long and contains the 3′ YFC motif. Conversely, there is variation in amino acid sequence and length in CDR1-IMGT and in CDR3-IMGT. In fact, *TRDV1* genes can be divided into two groups on the basis of the length of CDR1: seven or nine amino acids. The diversity of *TRDV1* genes within the subgroup is also due to the variable length of CDR3-IMGT, with four amino acids for 20 of them, three amino acids for *TRDV1S1*, and five for *TRDV1S10*. This region shows also a high content of tryptophan in position 107 (8 of 22 sequences). The CDR2-IMGT is three amino acids long in all the *TRDV* genes except *TRDV1S29* and shows a QXS motif (positions 56–58). According to the chemical characteristic of the amino acid at position 57, the *TRDV1* genes can be divided into six groups, with ten of them belonging to Y, four to G, two to acidic, two to amide, one to basic, and one belonging to hydroxyl class (Pommié et al. 2004). Based on hallmarks of CDRs, we classified the *TRDV1* genes into five sets named 1A–1E (Table 2). The structure of *TRDV1S29* is peculiar with a deletion of 27 nucleotides encompassing the last amino acids of FR2-IMGT (position 55), the CDR2-IMGT (positions 56–58), and the first five residues of FR3-IMGT (66–70). The deduced amino acid sequences of all germline *TRDV1* genes were then aligned with the remaining genes of the *TRDV1* subgroup listed in Table 1 and available as complete cDNA (Hein and Dudler 1993). A neighbor-joining phylogenetic tree with a total of 33 sequences was constructed (Fig. 4). The tree clearly shows a grouping of all the sequences except for *TRDV1S35*. The groups appear to correspond to the *TRDV1*

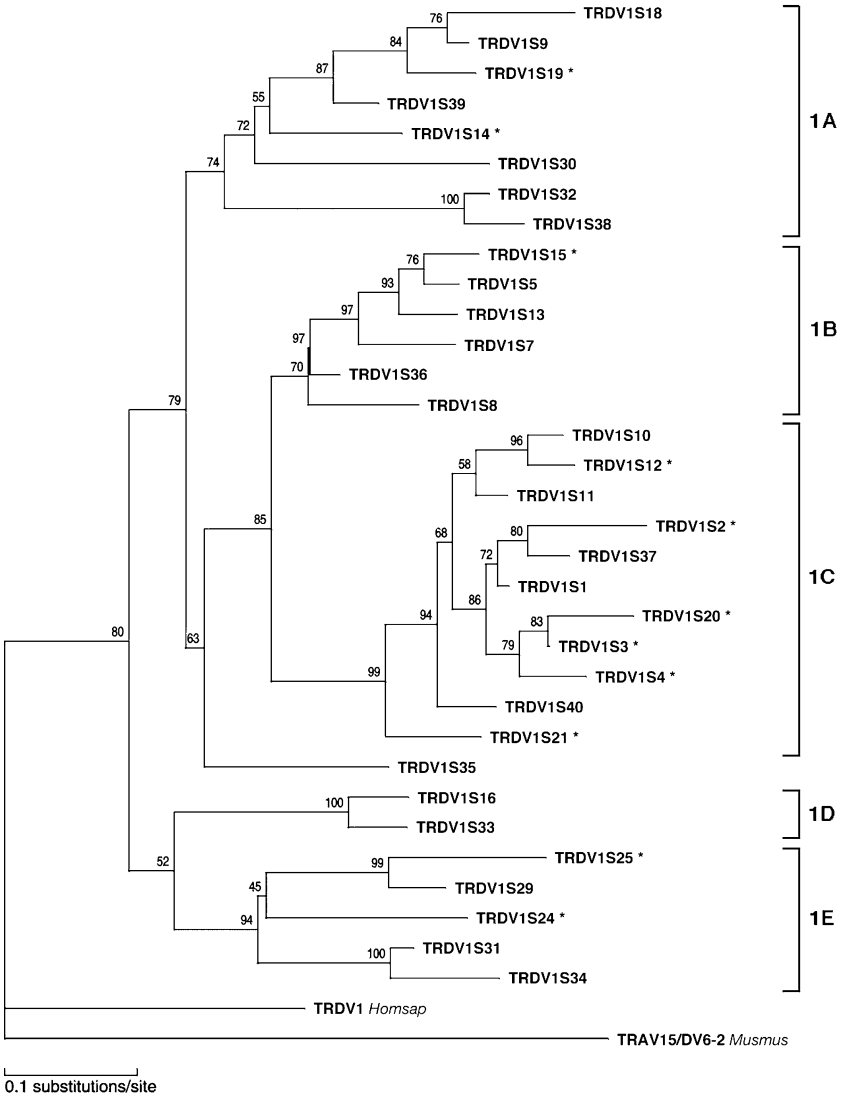


Fig. 4 The maximum-likelihood rooted tree (ln*L*: −2,726.19±188.12) inferred from the sheep *TRDV1* subgroup, human *TRDV1*, and mouse *TRAV15/DV6* genes (35 sequences, 119 sites), based on the JTT-F model of amino acid substitution and neighbor-joining (NJ) reconstruction tree (see "Materials and methods"). An *asterisk* indicates the *TRDV1* genes provided as cDNA. The five *TRDV1* sets (1A–1E) drawn by the tree are indicated by *brackets*. *Numbers along the branches* show the bootstrap numbers observed among 100 replicates of simulation. The maximum likelihood (ML) tree was constructed using ProtML (see "Materials and methods")

sets defined in Table 2. We also compared the deduced amino acid sequence of *TRDV1S1* gene, as representative for *TRDV1* subgroup, with those of the sheep variable genes belonging to the other *TRDV* subgroups, i.e., the germline *TRDV4* gene (this study) and *TRDV2* and *TRDV3* cDNA (Hein and Dudler 1993) (Fig. 5). A careful analysis of the substitutions in the other *TRDV* subgroups with respect to the *TRDV1S1* gene reveals that modifications are distributed along the entire sequence except for Cys23, Trp41, and some positions in the 3′ portion of FR3. The CDR1 is seven amino acids long in all the other subgroups, while the length and amino acid composition of CDR2 are most predictive of the *TRDV* subgroup identity. Briefly, CDR2 is three amino acids long in *TRDV1*, *TRDV2*, and *TRDV3*, with different amino acid composition, while it has five amino acids in *TRDV4*. The CDR3 is three amino acids long in *TRDV4*, while for *TRDV2* and *TRDV3* the CDR3 has still to be established.

Structural analysis of the 3′ and 5′ flanking regions

We extended the structural analysis of the *TRDV1* and the *TRDV4* genes to their 3′ and 5′ flanking regions. For the 3′ region, we performed multiple sequence alignment with the flanking RSSs of the *TRDV1* genes from the phage clones (Fig. 1a) and the *TRDV4* gene. The entire 23-mer (39 bp, including heptamer, nonamer, and the spacer region) of each RSS was subjected to Clustal W (Thompson et al. 1994) analysis to determine the degree of sequence similarity among them. The comparison revealed a striking grouping of the sequences, where each group consists of RSSs belonging to the genes previously classified into the same set (1A–1E, Table 2; Fig. 6). To assess the significance of the short RSS alignment, we also evaluated the overall sequence conservation within the *TRDV1* set by analyzing the sequence of the 39 nucleotides immediately adjacent to the 3′ end of each RSS. No known function is associated with these sequences. While a high degree of similarity was observed between genes of the same set in the RSSs, no reproducible similarity was seen in the immediately adjacent regions (data not shown).

Our results show that the *TRDV1* genes are closely related to each other within a set both in protein and RSS motifs.

Clustal W multiple sequence alignment of the 235 bp upstream from the initiation codon of representative *TRDV1* subgroup clearly revealed also in this case, a grouping of the sequences in agreement with the previous proposed set division (Fig. 7). In contrast, the 5′ flanking region of *TRDV4* gene does not show a remarkable homology with any of the *TRDV1* genes (data not shown).

Phylogenetic analysis of the *TRDV* genes

To establish the evolutionary relationship of the sheep *TRDV* genes with those of other mammals, we combined *TRDV* genes from humans, mice, sheep, cattle, and pigs in the same alignment. In particular, we used all the available sheep, cattle, and pig TRDV genes, and the human and mouse *TRDV* genes that have been found rearranged only to *TRDD* and not to *TRAJ* genes. In this analysis, amino

```
TRDV                                           FR1-IMGT            CDR1-IMGT
gene             leader                         (1-26)              (27-38)

                               1.......10........20...... ..30........ 40...
                               |        |         |          |          |
TRDV1S1    MPLSSLPWVLLAFTFSGSGVAQKVTQDQSDVSSQVGQSVTLNCRYE TSWSYYN..... LFWYK
TRDV2      MLCPGLLWVFMATFGFGSSMADKVTQAQTTVTGQEREAVTIGCTYE TSQTYYT..... LFWYR
TRDV3      MQSGPPALLCAVVALICLGSDNVESADVPTVYKKEGESVTVECKFS VSYTYYM..... MYWYR
TRDV4      MFLPVGFSLLL.FYK.G.VLCNQVTQSSPEERVASGREVTLQCTFQ TTYSNPD..... LYWYR
           *                                     **  *                   **


           FR2-IMGT    CDR2-IMGT            FR3-IMGT                   CDR3-IMGT
           (39-55)      (56-65)             (66-104)                   (105-109)

           .....50..... ...60..... ...70........80....A....90.......100.... .....
                 |          |          |         |          |         |
TRDV1S1    QLPSGQMTYVIR QGS....... QATNARK.DRYSVNFKKAD.KSISLTISALQLEDSAKYFC ALC.. [7.3.3]
TRDV2      QFPGGRMEFLIY QDS....... NQANARR.DRYSVNFQKGKKIIISLTISSLHLADSAKYFC ALW   [7.3. ]
TRDV3      QPSSGEMIYMIN IYS....... QSKQTRE.GRYSVEFYKPN.QMLKLTISALKLSDSAVYFC AVR   [7.3. ]
TRDV4      KTPDAVFQFVLY RDNTR..... SRDADFARGRFTVQHSVRS.KTFHLVISSVRPEDTATYYC ALEL. [7.5.4]
                                                           * ****   * * * * *
```

Fig. 5 Alignment of the deduced amino acid sequences of sheep *TRDV* subgroups. The names of the genes, in accordance with IMGT nomenclature, are listed at the *left*. The *TRDV1S1* gene (*first line*) was selected for comparison with the *TRDV* genes belonging to the other subgroups. *TRDV2* and *TRDV3* amino acid sequences were obtained from cDNA (Hein and Dudler 1993), while the amino acid sequence of *TRDV4* was deduced from the genomic sequence of the current study. Gaps were inserted to align all sequences and are indicated by *dots*. *Blank spaces* in the CDR3 of *TRDV2* and *TRDV3* genes indicate that this region is still to be established. Leader sequence, CDR-IMGT (complementarity determining regions), and FR-IMGT (framework regions) are also indicated in accordance with IMGT nomenclature (Lefranc et al. 2003). CDR lengths are indicated at the end of each sequence. *Asterisks* in the *bottom line* indicate identical amino acids

Fig. 6 Nucleotide sequence alignment of the recombination signal sequences (RSSs) found downstream of selected *TRDV1* and *TRDV4* genes. The name of each *TRDV1* gene and the corresponding set are indicated on the *left* and on the *right*, respectively. Nucleotides of the spacer region conserved within sets are *highlighted*. The consensus sequences of the heptamer and nonamer (Hess et al. 1989) are provided in the *last line*. The nucleotide variations of *TRDV4* RSS with respect to the consensus are *highlighted*

acid sequences were used, and various phylogenetic trees were constructed by different methods (see "Materials and methods"), introducing *TRDV* chicken genes as an outgroup. In all cases, we obtained the same arrangement of the genes. As an example, Fig. 8 shows the phylogenetic tree drawn by the maximum likelihood and NJ methods.

The tree shows that the *TRDV* genes in mammals can be classified into three major groups, A–C, with the exception of the mouse *TRAV15/DV6*, which did not cluster reliably with any other groups. This classification is supported by bootstrap probability (Pb) values ranging from 60% to 80%. As expected, all the chicken sequences are divergent from *TRDV* genes of mammals and form in the tree a reliable group, D, with a Pb value of 98%. Group A contains the artiodactyl *TRDV1* multigene gene family with the related human *TRDV1* gene. It is noteworthy that sheep and cattle genes do not form two separate clusters while the pig genes are grouped separately. Moreover, in the ruminant branch, some sheep and cattle genes are tightly related; we tentatively assumed that they were orthologous pairs of genes. However, the presence of the bovine genes does not alter the grouping of the sheep genes. Within group A there are two more clusters that are highly significant. One cluster includes sheep *TRDV2* and pig *TRDV3* genes without any human or mouse tightly related genes, so these genes might be a peculiarity of artiodactyls; the second cluster consists of pig and mouse *TRDV2* genes. In the tree, the location of the mouse *TRAV15/DV6* gene proximally to the group A genes, is consistent with the relationship between this mouse gene and the *TRDV1* subgroup (Bosc and Lefranc 2003). Group B includes sheep *TRDV3* and pig TRDV4 genes without any human and mouse counterparts. Group C is composed of human *TRDV3*, mouse *TRDV5*, sheep *TRDV4*, and pig *TRDV5* genes. The phylogenetic location of these genes is correlated to their genomic position within the *TRD* locus in the different species. In fact, they are located in all cases in an inverted transcriptional direction in the 3′ end next to the

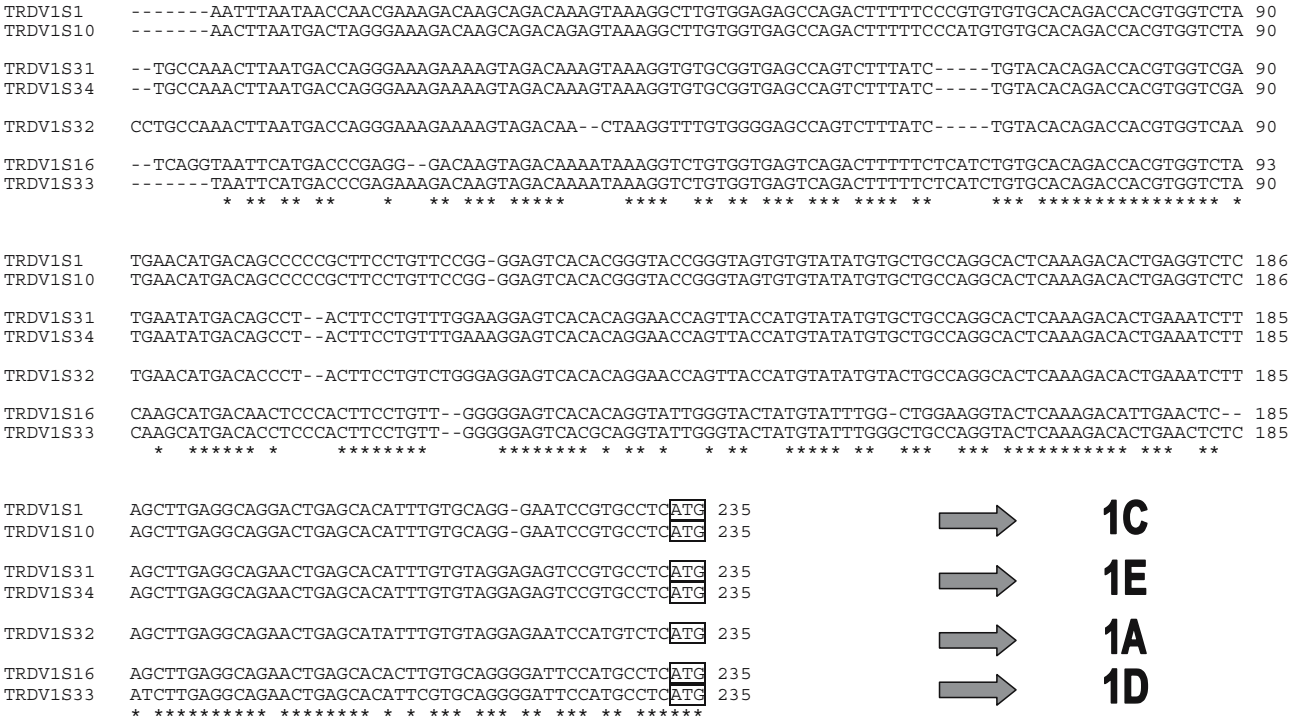
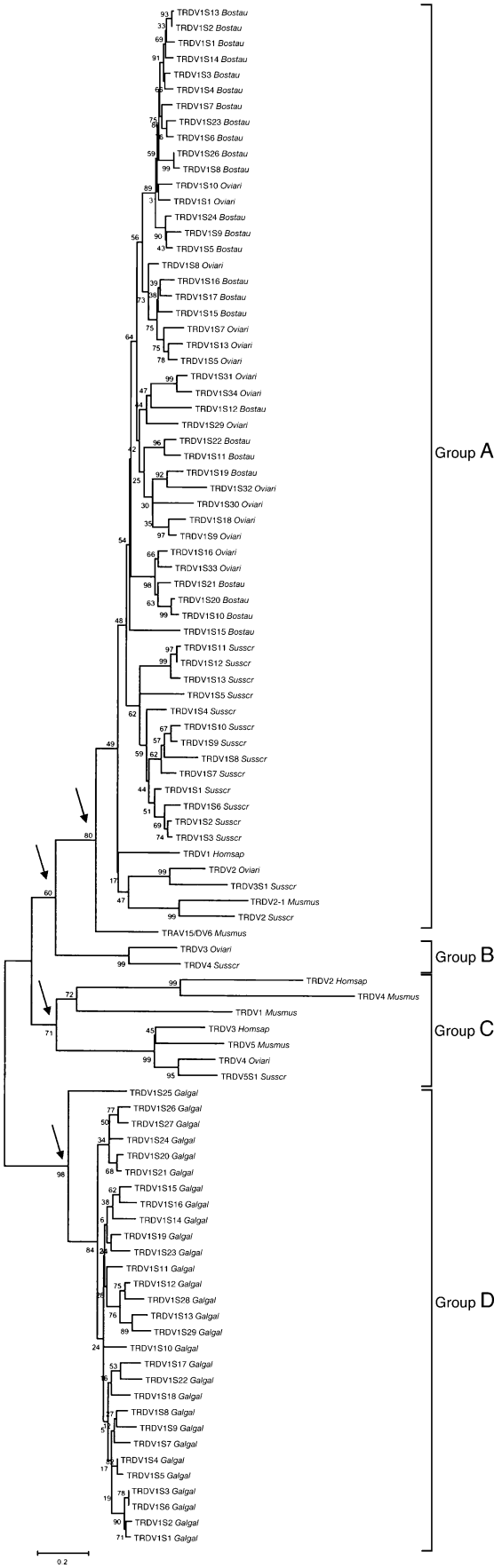


Fig. 7 Alignment of nucleotide sequences of the 5′ flanking regions of selected *TRDV1* genes. The name of each *TRDV1* gene and the corresponding set are indicated on the *left* and on the *right*, respectively. The ATG codon is *boxed*



◀**Fig. 8** The ML rooted tree (ln*L*: −8191.93±409.52) inferred from the *TRDV* gene sequences within mammals (97 sequences, 77 sites) based on the JTT-F model of amino acid substitution and NJ reconstruction tree, rooted on the chicken *TRDV* genes (see "Materials and methods"). The four groups (A–D) of *TRDV* genes drawn by the tree are indicated by *brackets*. *Numbers along the branches* show the bootstrap numbers observed among 100 replicates of simulation. The *arrows* indicate the bootstrap values that separate groups. The ML tree was constructed using ProtML (see "Materials and methods"). The species are indicated by the standardized *six-letter code* derived from Latin names

constant gene, which may have promoted their remarkably long evolutionary maintenance. Group C comprises also the human *TRDV2* and the mouse *TRDV4* and *TRDV1* genes without any artiodactyl counterpart. Also in this case, the closely related human *TRDV2* and mouse *TRDV4* genes represent an orthologous pair of genes since they lie in the 5′ end close to *TRDC* gene.

To examine the reliability of the tree topologies generated by the amino acid analysis, a multiple alignment including the nucleotide sequences of the same *TRDV* genes was also carried out. A phylogenetic tree drawing by using the Markov model (see "Materials and methods"), shows the same cluster distribution as in Fig. 8 (data not shown). The *TRDV* chicken genes, human *TRDV2*, and mouse *TRDV1* and *TRDV4* genes were excluded from the analysis because they lost the "stationary condition." This reveals that in the various lineages there are different patterns of base composition thus preventing a complete analysis of sequences at nucleotide level.

Discussion

By using two different methods, we obtained an adequately representative germline repertoire of *TRDV1* genes from single individuals of two sheep breeds. In total 22 unique *TRDV1* genes were identified. One half of them are *TRDV1* genes already identified as cDNA; the remaining genes can be classified as new. In most cases, the germline sequence is identical to the corresponding cDNA, with occasional differences of one to three nucleotides. Additionally, few nucleotide differences were observed between coincident genes in the different sheep breeds, indicating that the strain polymorphism for *TRDV1* genes in sheep is slight. Although the number of genomic *TRDV1* genes obtained in this study is not large enough to cover all the genes isolated so far as cDNA, the presence of genomic sequences identical to cDNA excludes the involvement of somatic mechanisms, such as hypermutation, in generating the sequence diversity. Thus, the *TRDV1* subgroup in sheep is a multimember gene family, composed of at least 40 distinctive variable genes. It is possible that the total number of different *TRDV* genes varies slightly in different strains, mostly due to different haplotypes. In any case the presence of many distinct sequences in the two different breeds suggest that a large number of *TRDV* genes can be used to form a TR δ chain in sheep. In pigs, in 31 distinct *TRDV1* genes obtained partly from thymocyte RNA and partly from the liver genomic DNA, only five genes were found to be in common

between cDNA and germline sequences (Yang et al. 1995). This is consistent with the presence of a large *TRDV1* germline repertoire in all artiodactyl species. It would seem that the diversification of the variable regions of TR δ chain in artiodactyls relies primarily on the presence of a large number of diverse *TRDV1* genes.

In contrast to α/β TR, which recognizes as antigen only peptide fragments bound to major histocompatibility complex (MHC), γ/δ TR appears to perceive proteins directly (Lefranc and Lefranc 2001). Therefore, the CDR regions are directly involved in antigen recognition.

The sheep *TRDV1* subgroup presents a high level of structural diversity due mainly to differences in length and amino acid composition of CDR1, CDR2, and CDR3. Li et al. (1998) maintained that longer CDR3 could protrude from the surface of the variable domain, creating prominence, which may facilitate the binding of γ/δ T cells to a wide variety of antigenic surfaces. This seems equally plausible for CDR1. We postulate that *TRDV1* genes used longer CDR1 and CDR3 to create a large accessible surface area available for antigenic binding. At the same time, the *TRDV1* genes accumulated in these regions amino acid variations that led to subtle surface modifications to improve the fit of the variable domain with the antigen. Fewer variations among the *TRDV1* genes are evident in the CDR2 interval that is involved in the formation of CDR2 (C′–C″) loop (Allison et al. 2001; Lefranc et al. 2003). In all cases, the *TRDV1* genes, with the exception of *TRDV1S25* (Hein and Dudler 1993) and *TRDV1S29* (Fig. 3), are three amino acids long in this region with a specific set pattern of amino acid composition. Therefore, the germline *TRDV1* genes are largely diversified by a structural variability that includes the increased area of CDRs and a high degree of amino acid changes. Taken together, these results suggest that sheep δ variable germline repertoire has been shaped by selection operating at the level of ligand recognition. As predicted by the classic model of evolution, we infer that diversification at the CDRs has guaranteed the maintenance of functional multiple copies of the genes. The preservation of such a high number of genes is also assured by the conservation of the 3′ and 5′ flanking regions involved in the control of *TRDV* gene rearrangement and expression, respectively. There is evidence that even single nucleotide changes in the heptamer, spacer or nonamer motifs, respect to the consensus sequences, can alter recombination frequency (Feeney et al. 2000). Therefore, the perfect sequence conservation of the heptamer and nonamer in all the *TRDV1* genes might secure the involvement of all genes in the formation of the primary combinatorial antigen receptor repertoire. We also observed a substantial conservation for the promoter regions and this can be related to the markedly similar expression of all the genes in adult animals (Hein and Dudler 1993; Massari et al. 2000).

If in adult sheep the variable δ repertoire is determined by the marked expression of *TRDV1* subgroup genes, the profile of fetal repertoire is characterized by the predominant presence of the other three *TRDV* subgroups (Hein and Dudler 1993). The *TRDV2* and *TRDV3* subgroups consist of single genes with sequences relatively similar to *TRDV1*, especially those *TRDV1* genes with minimum CDR1 lengths. In contrast, the *TRDV4* gene presents differences in the structural organization as well as in the regulatory regions compared to *TRDV1* genes; the CDR1 and CDR3 regions are seven and three amino acids long, respectively, the minimum length of the respective *TRDV1* genes. However, the CDR2-FR3 interval is three amino acids longer, signifying a different conformation. The *TRDV4* gene also presents a mismatch in both its heptamer and nonamer sequences that could result in a slight decrease of the recombination substrate activity. The 5′ flanking region of the *TRDV4* gene does not show homology with the corresponding region in *TRDV1* genes. These last results parallel the markedly different expression of the *TRDV4* and *TRDV1* subgroups between fetus and adult (Hein and Dudler 1993). The mechanisms that influence the different repertoire in the fetus and the adult will in part depend on the ligands that are reactive in the two periods. In the fetal period, there are few ligands, and the repertoire of variable δ chains programmed by the *TRDV2*, *TRDV3*, and *TRDV4* genes results in receptors that have minimal sequence diversity. As ontogeny proceeds, the number of ligands increases, and a more extensive sequence diversity is produced by using the different genes of the *TRDV1* subgroup. The preferential use of the *TRDJ1* and *TRDJ3* genes in adult sheep with respect to *TRDJ2* (Massari et al. 2000), and a more extensive junctional diversity, reflecting an increased level of *N*-nucleotide addition and/or the usage of multiple D genes (Hein and Dudler 1993), complete the transition from fetal to adult stage.

The molecular events that might have governed the evolution of the coordinate expansion of the *TRDV* subgroups in γ/δ high species compared to γ/δ low species can be deduced from the analysis of a phylogenetic tree. Our phylogenetic approach classified the *TRDV* genes from artiodactyl species into three major groups, A, B, and C, where only A and C are shared with humans and mice. Therefore, we can assume that the ancestor of mammalian species had two variable genes and that subsequently the progenitor of artiodactyls acquired one more *TRDV* gene. The monophyletic cluster of the *TRDV1* genes of mammalian species within group A indicates that these genes were derived from a single common ancestral gene existing prior to mammalian radiation. Moreover, within the artiodactyl species the presence of two separate clusters for ruminants and pigs attests that gene duplication events that affected the birth of the expanded *TRDV1* subgroup occurred after the separation of ruminants (sheep and cattle) from pigs. The tree also shows that sheep and cattle genes do not form two separate clusters. This implies that many gene duplications occurred before and after ruminant divergence and that duplicated *TRDV* genes have not been subject to any significant interlocus homogenization of sequences within either of the two species. Deviation from the clustering pattern for other sheep and pig *TRDV* genes are also evident. For instance, pigs possess a *TRDV* gene belonging to the *TRDV2* subgroup not represented in sheep. The most reasonable explanation is that the conservation and expansion of a particular group of genes in the genome reflects the adaptation of the

repertoire to certain types of antigens. Sheep and cattle are likely to encounter similar types of antigens, which are different from the antigens to which swine are exposed. Therefore, the evolution of *TRDV* is assumed to be expanded or contracted, depending on the need to protect the host from ever-changing groups of parasites. Different life environments may have determined the greater diversity of the artiodactyl *TRDV* repertoire compared to humans and mice.

Acknowledgements The financial support of Progetti di Interesse Nazionale (PRIN) and Ministero per l'Università e la Ricerca Scientifica e Tecnologica (MIUR) is gratefully acknowledged. We thank Prof. Bozzetti Maria for critical reading of the manuscript

References

Allison TJ, Winter CC, Fournie JJ, Bonneville M, Garboczi DN (2001) Structure of a human gammadelta T-cell antigen receptor. Nature 411:820–824

Binns RM, Duncan IA, Powis SJ, Hutchings A, Butcher GW (1992) T lymphocytes in the blood of young pigs identified by specific monoclonal antibodies. Immunology 77:219–227

Bosc N, Lefranc, MP (2003) The mouse (*Mus musculus*) T cell receptor alpha (TRA) and delta (TRD) variable genes. Dev Comp Immunol 27:465–497

Cooper MD, Chen CLH, Bucy RP, Thompson CB (1991) Avian T cell ontogeny. Adv Immunol 50:87–117

Feeney AJ, Tang A, Ogwaro KM (2000) B-cell repertoire formation: role of the recombination signal sequence in non-random V segment utilization. Immunol Rev 175:59–69

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

Hein WR, Dudler L (1993) Divergent evolution of T cell repertoires: extensive diversity and developmentally regulated expression of the sheep γδ T cell receptor. EMBO J 12:715–724

Hein WR, Dudler L (1997) TCR γδ+ cells are prominent in normal bovine skin and express a diverse repertoire of antigen receptors. Immunology 91:58–64

Hein WR, MacKay CR (1991) Prominence of γδ T cell in the ruminant immune system. Immunol Today 12:30–34

Hess JE, Lieber MR, Mizuuchi K, Gellert M (1989) V(D)J recombination: a functional definition of the joining signals. Genes Dev 3:1053–1061

Ishiguro N, Aida Y, Shinagawa T, Shinagawa M (1993) Molecular structures of cattle T-cell receptor gamma and delta chains predominantly expressed on peripheral blood lymphocytes. Immunogenetics 38:437–443

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Kubota T, Wang J, Gobel TW, Hockett RD, Cooper MD, Chen CLH (1999) Characterization of an avian (*Gallus gallus domesticus*) TCR alpha delta gene locus. J Immunol 163:3858–3866

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17: 1244–1245

Lefranc MP (2003) IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 31:307–312

Lefranc M-P, Lefranc G (2001) The T cell receptor facts book. Academic, London, pp 389

Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol 27:55–77

Li H, Lebedeva MI, Llera AS, Fields BA, Brenner MB, Mariuzza, RA (1998) Structure of the Vδ domain of a human γδ T-cell antigen receptor. Nature 391:502–506

Massari S, Antonacci R, Lanave C, Ciccarese S (2000) Genomic organization of sheep *TRDJ* segments and their expression in the δ-chain repertoire in thymus. Immunogenetics 52:1–8

Nicholas KB, Nicholas HB, Deerfield DW (1997) GeneDoc: analysis and visualization of genetic variation. Emb News 4:14

Pommié C, Levadoux S, Sabatier R, Lefranc G, Lefranc M-P (2004) IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. J Mol Recog 17: 17–32

Saccone C, Lanave C, Pesole G, and Preparata G (1990) Influence of base composition on quantitative estimates of gene evolution. Methods 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

Scaviner D, Lefranc MP (2000) The human T cell receptor alpha variable (TRAV) genes. Exp Clin Immunogenet 17:83–96

Swofford D (1998) PAUP*: phylogetic analysis using parsimony (* and other methods). Sinauer, Sunderland

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680

Yang YG, Ohta S, Yamada S, Shimizu M, Takagaki Y (1995) Diversity of T cell receptor δ-chain in the thymus of a one-month-old pig. J Immunol 155:1981–1993