Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 355 (2005) 67 – 78
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Comparative analyses of sheep and human *TRG* joining regions: Evolution of *J* genes in *Bovidae* is driven by sequence conservation in their promoters for germline transcription

M.C. Miccoli [a,1], G. Vaccarelli [a,1], C. Lanave [b], E.P. Cribiu [c], S. Ciccarese [a,*]

[a] *Dipartimento di Anatomia Patologica e di Genetica, University of Bari, Italy*
[b] *Istituto di Tecnologie Biomediche- Sezione di Bari, CNR- Milano, Italy*
[c] *Departement de Genetique Animale, Centre de Recherche INRA de Jouy-en-Josas, France*

Received 8 November 2004; received in revised form 23 February 2005; accepted 17 May 2005
Available online 21 July 2005

Abstract

The availability of genomic clones representative of the T cell receptor gamma (*TRG1@* and *TRG2@*) ovine loci enabled us to compare the germline genomic organization and nucleotide diversity of joining (*J*) segments and reconstruct their evolutionary history by phylogenetic analysis of cattle, sheep and human expressed sequences. Expression profiling (RT-PCR data) in fetus and adult indicated that only the ovine *J* genes in which two or more of the key sequence features, such as recombination signal sequences (RSS), 3′ splice sites, and core sequences, are missing or severely altered fail to be transcribed. Comparative genomic examination of the two human with the six sheep germline transcription promoters located at 5′ of the relevant constant (*C*)-distal *J* segments showed a strong conservation of the redundant STAT consensus motifs, indicating that *TRG1@* and *TRG2@* loci are under the influence of IL-7 and STAT signalling. These findings support the phylogenetic analysis of human and *Bovidae* (cattle and sheep) that revealed a different grouping pattern of *C*-distal compared to *C*-proximal *J* segments. Likewise, the phylogenetic behaviour of either *C*-distal and *C*-proximal *J* segments is in accordance with the *Bovidae TRG* clusters evolution. Comparison of sheep and human structures of recombination signal sequences (RSS) has highlighted a greater conservation in sheep 12 RSS rather than 23 RSS thus suggesting that the initial recruitment of recombination activating genes (RAG) products requires at least one relatively high-affinity RSS per recombination event.

Keywords: T cell receptor gamma joining genes; Cattle and sheep; Recombinational accessibility; STAT proteins motifs; Genome evolution

1. Introduction

γ/δ T cells represent only a small percentage (1–5%) of circulating T lymphocytes in most adult organisms, as human and mouse ("γδ low" species), but this percentage increases in artiodactyls and chicken ("γδ high" species) to more than 70% in young subjects (Hein and MacKay, 1991; Six et al., 1996). However, although various immunoprotective and immunoregulating functions have been proposed for γ/δ T cells, we still lack a compelling explanation of how these cells developed and were conserved along the evolutionary pathway. We know a great deal about the important biological properties of γ/δ T cells which differentiate them from α/β T cells and about the differences existing between

Abbreviations: $P_{B,}$, bootstrap probability; IMGT Database, immunogenetics information system; IgH, immunoglobulin heavy chain; IL-7, interleukin-7; IL-7R, interleukin-7 receptor; *J*, joining; MP, maximum parsimony method; NJ, neighbor-joining method; NMD, Nonsense-Mediated mRNA Decay; PTC, premature termination codon; RAG, recombination activating genes; RSS, recombination signal sequences; RT-PCR, reverse transcription PCR assay; STAT proteins, sequences target activating transcription proteins; *TRG@*, T-cell receptor gamma locus.

* Corresponding author. DAPEG-Genetics Section-Faculty of Biotechnological Sciences-University of Bari-via Amendola 165/A, 70126 Bari, Italy. Tel.: +39 080 5443384; fax: +39 080 5443386.
E-mail address: ciccarese@biologia.uniba.it (S. Ciccarese).
[1] M.C. Miccoli and G. Vaccarelli contributed equally to this work.

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.05.032

species and strains with regard to the function of this specific lymphocyte population. Still much remains to be learned about the stage of the lymphocyte development in the "γδ high" species where the different regulation takes place, which results in a preferential differentiation of γδ T cells (Ciccarese et al., 1997). The early TCR γδ expression may be a direct determinant of cell fate toward γδ maturation; likewise, the pre-TCR may be a dominant determinant of the commitment to the αβ lineage (Instructive Model). This hypothesis is substantiated by the correlation between the complexity of the TCR γδ gene loci and the abundance of γδ-expressing T cells in the organism (Dudley et al., 1995; Hoffman et al., 1998). In short, the larger the number of the γδ genes (as is the case for *Bovidae* and chicken) the higher the likelihood of γδ TCR being expressed in early thymocytes and, consequently, the larger the number of γδ T that undertake differentiation (Hayday, 2000).

The γδ TCR expression is closely linked to the rearrangement of the involved genes during lymphocyte maturation. According to the accessibility hypothesis, *V(D)J* recombination could be regulated by developmental changes in the accessibility to recombinases of chromatin-embedded recombination signal sequences (RSSs), even if recent evidence shows that beyond acting as a chromatin barrier, RSS possess specific features that could remarkably influence the efficiency and specificity of *V(D)J* recombinational events (Krangel, 2003). IL-7 interacting with IL-7R is an essential cytokine for early lymphocyte development, when *V(D)J* recombination takes place. It promotes recombination in IgH and TCRγ loci as it induces germline transcription and DNA rearrangement in *V(D)J* segments (He and Malek, 1996; Corcoran et al., 1998; Ye et al., 1999). IL-7 interacting with IL-7R basically regulates TCRγ locus accessibility to recombinases and trascriptional machinery by triggering the activation of STAT proteins (Lee et al., 2001; Huang et al., 2001). Activated STAT proteins translocate into the nucleus and bind to specific consensus motifs (TCCNNNGAA) found, among others, in promoters at 5′ joining segments but also in enhancers and Hsa elements of human and mouse TCRγ loci (Lee et al., 2001). In cattle and in sheep the *TRG* locus has two different locations on chromosome 4, one at position 4q31 (locus *TRG1@*), and the other one (*TRG2@*) at 4q22 (Miccoli et al., 2003; in preparation). In companion paper we have established that *TRG1@* locus spans about 140 kb and consists of three clusters named *TRG5*, *TRG3*, and *TRG1* according to the constant (*C*) genes. We have demonstrated that in sheep, *TRG1@* locus *V*, *J*, and *C* segments are arranged in the same transcriptional orientation, with a *V–J–J–C* structural scheme having, in the case of *TRG5* and *TRG3* clusters, repeated structures within *V* gene regions.

In this paper, our aims were: i) to determine the genomic organization, sequence analysis and transcription assay of sheep *TRG1@* and *TRG2@* loci *J* gene segments; ii) to perform a detailed evolutionary analysis of the *J* region in *Bovidae* and human; iii) to identify the promoters for germline transcription in both sheep loci and comparing them with the corresponding human sequences; iv) to compare sheep and human structures of the 23 and of the 12 RSS. The comparative analysis of the germline promoters located at the 5′ of each *C*-distal *J* gene strongly supports the key role of such promoters in determining the local accessibility to the site-specific recombination machinery. Phylogenetic analysis of joining genes shows a correlation between the distribution of *J* segments of *Bovidae* in the phylogenetic trees and their position in the *TRG* clusters genomic organization. *Bovidae J* segments have evolved both as elements belonging to the evolutionary relationship existing between *V* and *C* genes of each *TRG* cluster, and as *C*-proximal or *C*-distal placed elements within the cluster itself. This data provides a genomic perspective for the understanding of the "γδ high" speciation. The important immunoprotective role of a large γδ T cell population in ruminant animals is likely to have caused during the evolution a series of functional duplications of a minimal *TRG* structure. The reiterations have involved also STAT proteins nucleotide motifs responding to IL-7 signalling and thereby regulating the recombination accessibility to the *TRG@* loci.

2. Materials and methods

2.1. Joining genes mapping and subcloning

Isolated BAC clone, address 201E2 (see Section 3.1 and Table 1b companion paper) DNA was digested with various rare-cutting restriction enzymes. DNA samples separation was performed on 1.2% agarose gel. Gels were run on a CHEF for 15 h, at 14 °C, 6V/cm, 120° and 2 s switching time in 0.5X TBE. The gels were blotted onto a nylon membrane (Amersham), which was then hybridized by using the DIG High Prime DNA Labeling and Detection Kit (Roche Diagnostic GmbH) to various *J* specific cDNA probes. The cDNA clones were a gift from Dr. WR Hein (Basel Institute for Immunology, Basel, Switzerland). Amplified DNA fragments and subclone inserts were purified using the High Pure PCR Product Purification Kit (Roche Diagnostic GmbH). Nucleotide sequences were determined using a commercial service and the resulting sequences were compared to existing ones using the Basic BLASTN research program and VISTA tools for comparative genomics (website: http://gsd.1bl.gov/vista/index.shtml).

The distances between *J* and *C* elements were determined by PCR on BAC DNAs with primers based on published cDNA sequences. Approximately 0.7 μg of each BAC DNA template were used in the PCR reactions. The standard buffer reaction consisted of 10× PCR buffer (Roche Diagnostic GmbH); 0.2 mM of each dNTP; primers at 0.05 μM; 2.5 U Expand Long Template Polymerase (Roche Diagnostic GmbH) was added in a 50 μL final

volume. The thermal cycling protocol was: 4 min 94 °C, 30 s 94 °C, 40 s 58 °C, 5 min 68 °C for 30 cycles. The final cycle was extended for 15 min at 68 °C.

2.2. RT-PCRs (J transcription assay)

To test for the presence of correctly spliced transcription products, 10 *J*-specific 18- to 21-mer primers were designed. Five *C*-specific primers were also synthesized. Total RNAs were extracted from adult and fetus ovine spleen (Trizol method-Invitrogen) and polyA(+)RNAs were purified (Dynabeads oligo(dT)$_{25}$-Dynal). First strand cDNA synthesis was performed using *C* third exon specific primers (cDNA synthesis kit-Invitrogen). Ten microliters of the purified cDNA (50 μL) were used together with two primers, one for *J* or *V* and one for the relevant *C* first exon, in a standard 50 μL-PCR reaction. PCR reactions were stopped during the exponential phase (20 cycles) and at the end point (35 cycles). PCR products were run on a 2.0% agarose gel, stained and analysed by ChemiDoc (Bio-Rad). Quantitative analyses of PCR products using the program Quantity-one (Bio-Rad) indicated in all cases that the relative ratio of the various transcripts at 20 and 35 cycles was substantially the same. The following primers were used, for *J1.1*: 5′-TTC-AACGTTGGAACAAAACT-3′, for *J1.2*: 5′-AAGCTCATA-ATAGCTCCCCA-3′, for *J3.1*: 5′-AAACCAA-CTGAGCTCATAG-3′, for *J3.2*: 5′-CAGGGTTAATTGG-CAGGGA-3′, for *J5.1*: 5′-GATCAAGGTATTTGGTGAAG-3′, for *J5.2*: 5′-TCGGCGATGGGACAAAACT-3′, for *J2.1/J6.1*: 5′-GAAGATATTTGGAGAAGGAG-3′, for *J2.2/J6.2*: 5′-CAAAACTCATTGTCACAGG-3′, for *J4.1*: 5′-ATCTT-TGGTGGAGGAACA-3′, for *J4.2*: 5′-CAGGTTGGAA-GATGTTTGG-3′.

Total RNAs were extracted from young adult bovine spleen (Trizol method-Invitrogen) and polyA(+)RNAs were purified (Dynabeads oligo(dT)$_{25}$-Dynal). The treated RNA was used to generate full-length cDNA. First strand cDNA synthesis (cDNA synthesis kit-Invitrogen) was performed using bovine *C4* first exon specific primer (C4BR1: 5′-GGTCTTCATGGTATTTCCCT-3′). The 5′ RACE PCR were performed with a second *C4* specific primer (C4BR3: 5′-GTTGCAGCCTGTTGATGGTT-3′) and an ovine *V6* specific primer (V6: 5′-CAACTTGAGCCCCAC-CACC-3′). PCR products obtained were purified with High Pure PCR Product Purification Kit (Roche Diagnostic GmbH) and cloned into pCR-XL-TOPO vector (TOPO XL PCR cloning kit; Invitrogen).

2.3. Sequence alignment and computational analysis

The nucleotide Joining sequences were multialigned on the basis of amino acids alignment using CLUSTALW (Thompson et al., 1994; website: http://www.ebi.ac.uk/clustalw/), with manually optimization by using GeneDoc program (Nicholas et al., 1997; website: http://www.psc.edu/biomed/genedoc/). Multiple sequence alignments of human (NG 001336.2) and sheep (see accession numbers in Table 1b, companion paper) 5′ Joining sequences and of human (NG 001336.2) and sheep (see accession numbers in Table 1b, companion paper) RSS were constructed using the program CLUSTAL W, with GeneDoc manually optimization. Sequence logos of 23RSSs extracted from the 3′ ends of functional *V* elements from human and sheep, and of 12RSSs extracted from the 5′ ends of functional *J* elements from human and sheep, were by WebLogo (website: http://weblogo.berkeley.edu/logo.cgi). Repetitive elements AluSx (275 bp), SINE Bov-tA3 (126 bp), SINE Bov-A2 (272 bp) and LINE Bov-B (62 bp) were identified by Repeat Masker (website: http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) (A.F.A. Smit and P. Green, unpublished data. Current Version: 3.0.2).

2.4. Drawing of joining segments dendrograms

TRGJ sequences used in the analysis were: human (*Homo sapiens*): *J1*, *J2*, *JP*, *JP1*, *JP2* (NG001336.2), *Jδ1* (M20289), *Jδ2* (L36386), *Jδ4* (AJ249814); cattle (*Bos taurus*): *J1S1* (D16131), *J1S2* (D16122), *J1S3* (D16129), *J2S1* (D13648), *J2S2* (D13653), *J2S3* (D16117), *J4S1* (AY560834), *J5S1* (U73186); sheep (*Ovis aries*): *J1S1* (AY362774), *J1S2* (AY362778), *J3S1* (AY362776), *J5S1, J5P* and *J5S2* (AY147900), *J2S1* (AF540883), *J2S2* (AY362771), *J6S1* (AF540882), *J6S2* (AY362773), *J4S1* (AF540884), *J4S2* (AY362772); platyrrhina (*Aotus nancymaae*): *JP1* (AF333732), *J2* (AF378747), *JP* (AF378749); chicken (*Gallus gallus*): *J1S1* (U78238), *J2S1* (U78240), *J3S1* (U78241). Phylogenetic analysis was done on amino-acidic aligned sequences using PROTML from Molphy (Adachi and Hasegawa, 1992). ML reconstructions utilized PROTML program with the Jones–Taylor–Thornton (JTT-f). The analysis on nucleotide aligned sequences was performed using maximum parsimony method (MP) on non-synonymous positions (1+2 codon position). We used the neighbor-joining (NJ; Saitou and Nei, 1987) method to reconstruct the phylogenetic trees. MP as implemented in PAUP*4.04 b (Swofford, 2000) was used to examine the reliability of topologies generated by the NJ method. The reliability of trees was examined by the bootstrap test (Felsestein, 1985) and the interior-branch test (Rzhetsky and Nei, 1992; Sitnikova, 1996), which produced the bootstrap probability (P_B) values, for each interior branch in the tree.

3. Results

3.1. Phylogenetic analysis of J segments highlights a C-related evolutionary pattern and confirms a cattle-and sheep-specific TRG cluster evolution

Fig. 1 shows the location of *J* segments on a physical map, relative to their own *C* and to the other *J*, as well as

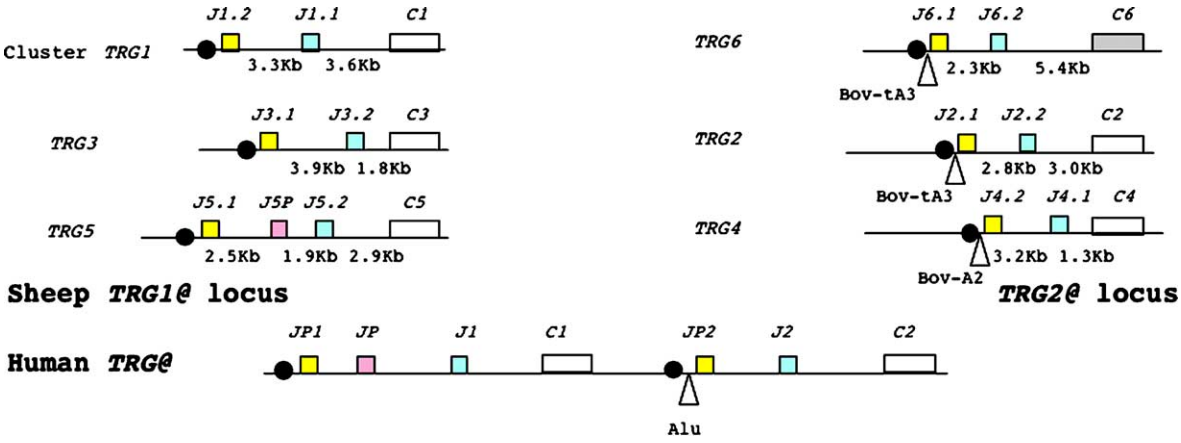

Fig. 1. Schematic illustration of sheep and human *TRG* loci *J* and *C* genes. *J* genes are depicted as boxes filled with different colors (in online version) on the basis of their location with respect to *C*; *C* genes are white with the exception of the pseudogene *C6* (grey) (Miccoli et al., 2001). The map of *TRG2@* locus scheme is an update containing newly identified *J2.2*, *J6.2*, and *J4.2* gene segments. The distances between gene segments, assessed by PCR, are indicated. (△) Indicates SINE-LINE insertion and a (●) is a schematic representation of STAT consensus motif (Lee et al., 2001).

their respective distances in both sheep *TRG1@* and *TRG2@* loci. All *J* gene segments are in the same transcriptional orientation as the relevant *C*. The *J* segments (2 for each cluster, except for the *TRG5* which has three of them) play a critical role in the recombination event and occupy a genomic position that places them right in the central part of each cluster. Newly isolated *J* segments were identified by examining the sequences for *J* core motifs, DNA recombination signals, RNA splice sites and comparative analysis of genomic sequences by mVISTA.

The nucleotide and deduced amino acid sequences of the seven *J* segments belonging to *TRG1*, *TRG3* and *TRG5* clusters, respectively are reported in Fig. 2a. *J* gene segments are typically 49–60 bases in length and are flanked by DNA recombination signal sequences (RSSs) at the 5′ end and by an RNA splice site at the 3′ end. *J1.2* may be considered a pseudogene (IMGT Database) because it is defective in the RSS heptamer upstream of the open reading frame and contains a premature stop codon. Fig. 2b shows the alignment of the *J2.1–J6.1* and *J2.2–J6.2* segments, plus the nucleotide and deduced amino acid sequences of *J4.1* and *J4.2*. In spite of the fact that the *J4.2* RSS nonamer does not conform to the 90% consensus recombination features, this *J* is considered as a transcribed pseudogene (see Table 1).

In all sheep *J* genes the centrally located conserved Phe-Gly (or Asn)-X-Gly amino acid motif, whose presence defines the functionality of *J* genes (IMGT Database), seems expanded to a Lys-X-Phe-Gly-X-Gly-X-Lys-Leu amino acid motif, whose conservation reaffirms the importance of these aminoacids in generating the correct tertiary conformation of the G strand in the upper β sheet of the variable domain. In this respect, the only exceptions are *J3.2* and *J5P* since both of them can be considered vestigial. *J5P* has no donor splicing site at 3′ end and altered RSS; *J3.2* exhibits a canonical RSS upstream of the open reading frame, but has no donor splicing site at 3′ end and presents significant amino acid sequence variations from functional *J* segments. To investigate the evolutionary behaviour of *J* with respect to *C* and *V* gene segments of each *TRG* cluster, the amino acid sequences putatively encoded by all sheep germline *J* elements, with the exception of *J3.2*, were aligned with the homologous complete sequences expressed in cattle, human, platyrrhina and chicken available in the public databases. The human *Jδ* locus segments *Jδ1* (M20289), *Jδ2* (L36386) and *Jδ4* (AJ249814), were included in the alignment. Fig. 3 shows the phylogenetic tree obtained by the maximum parsimony (MP) method (as implemented in PAUP*4.04b, used to examine the reliability of topologies generated by the NJ method, on non-synonymous positions). The results of phylogenetic analyses appear to emphasize the fundamental role of *J* genes in the recombination event since the bootstrap value close to 85% supports a clear-cut subdivision of human and sheep *J* sequences into two main groups. Group I consists of all the *J*s occupying the position closest to the constant gene (*C*-proximal) while group II consists of all the *J*s occupying the farthest position from the constant gene (*C*-distal). At present we have knowledge of the position of the genes on the physical map only for human and sheep *TRG* loci.

If cattle and sheep *J* sequences are considered, a significant distribution can be noticed in the tree, which is closely related to the clusters from *TRG1* to *TRG6* in relation to both group I and group II. Group III consists of chicken *J* genes only, with the exception of human *JP* and sheep *J5P,* these latter having in the physical map an intermediate position with respect to the other two *J* segments (Fig. 1) and platyrrhina *JP.* The same topology, as well as the same cluster distribution, was observed in a tree constructed by ML utilizing the PROTML program with the JTT-f method on amino acid sequences (not shown).

a

```
J1.1 ggccatcataattccactagaatttgaaggaatcagtggagctaaattagtctaaagttttgatacaggctgaatcactgtg

        N   S   Y   I   K   N   F   N   V   G   T   K   L   V   V   T
J1.1 T AAT AGT TAC ATA AAA AAC TTC AAC GTT GGA ACA AAA CTT GTC GTC ACA G gtaggtatcagagaa

J1.2  tgccagtacaaagtttttaactgctagtcatgtcatgaaggtaggtgaaggctgagattattgtaggagcttcaaccagggta

          G   S   E   *   K   K   I   F   G   K   G   P   K   L   I   I   A   P   H
J1.2  AT GGC TCA GAA TGA AAG AAG ATA TTT GGA AAA GGA CCT AAG CTC ATA ATA GCT CCC CAT G gt

J1.2  aagtaaatttttttccattt

J3.1 ttccagtacaaaatttttaacttccagtcatgacatgaagttaggagaaggctaagattattgtaggagcttcaaccagtgtg

        S   S   G   W   K   K   I   F   G   K   A   T   E   L   I   V   A   S   P
J3.1 AC AGT TCA GGC TGG AAG AAG ATA TTT GGA AAA GCA ACT GAG CTC ATA GTA GCT TCC CCT G gt

J3.1 gagtaatgtttttctctttttgccttagttatgg

J3.2 taattgcctatctgtttggtttgtgaaaggcatgtttcctgtaatgatggagttgaggatttcagacagttgagagcacagtg

        N   C   N   H   R   N   Q   C   P   G   L   I   G   R   D   S
J3.2 G AAT TGC AAT CAC AGG AAC CAG TGT CCA GGG TTA ATT GGC AGG GAT TCA gttgcaaccaaaaatact

J5.1 ccaccacaaagtgttttgaacttcgaatcacatgttgaagcaaagagaaggctgagatttttgtagaagctccaatcattgtg

        Q   G   I   K   V   F   G   E   G   T   K   L   V   V   I   P   P
J5.1 A CAA GGG ATC AAG GTA TTT GGT GAA GGA ACT AAG CTC GTA GTA ATA CCT CCT G gtaagtaactt

J5.2 gactgtcacagttacattaggatttgaagaaatcagaggtgctaaattagtctacagttttgatatgggttggatcactgtg

        S   Y   Y   V   K   I   F   G   D   G   T   K   L   V   V   T
J5.2 G AGT TAT TAT GTA AAA ATC TTC GGC GAT GGG ACA AAA CTT GTA GTC ACA G gtaggtatcagagga

J5P gtatatttctacaaatattgaaaagatatgggcacaatttgagagcaggcacagagggagatttctataaaagctttttaggtgt

        Q   T   G   G   Q   S   I   E   V   F   G   S   R   T   Q   L   I   V   T

J5P  G CAA ACA GGT GGC CAG AGC ATC GAA GTG TTT GGT TCC AGA ACA CAA CTT ATT GTT ACA G gc

J5P  gaagtttctttaaattctgtagtgaggtaggga
```

b

```
J2.1 agtatgaagttttttgacttctagtcactttataaaggtagaaaaaggctggattattgtaggagcatttatcattgtg
J6.1 cagtatg-ag-----a--------------------------g------------------------t------------

        S   S   G   W   K   K   I   F   G   E   G   A   N   I   I   V   T   P   P
J2.1 AC AGC TCG GGC TGG AAG AAG ATA TTT GGA GAA GGA GCT AAC ATC ATA GTA ACT CCT CCC G
J6.1 -- --- GA- --- --- --- --- --- --- --- --- --- --- --- --- A-- --- --- --- -

J2.1 gtaagtatcgttttttctattcttactgtagtcactaaaatactgatagatgctttgtagaaacaacttgggatctccct
J6.1 ---------a-ag-----g---------atagtca---------------------------------------------

J2.2 caaagtcatcgttattctagaatttgaagaaatcagagaagctaattagtctaaagttttgatatgacttgaatcactgtg
J6.2 agtca----a---c-a----------g-------------------------------------ca---------------

        N   N   Y   I   K   N   F   N   V   G   A   K   L   I   V   T
J2.2 G AAT AAT TAC ATA AAA AAC TTC AAT GTT GGA GCA AAA CTC ATT GTC ACA G gtaggtatcagaa
J6.2 - --- G-- --- --- --- --- --- --- --- --G --- --- --- --- --- --- - -------------
           D

J4.1 tgtcatgattccactaggatttaaagaactcaggcaagctagattagtctaaagttttgccacaggttgaatcactgtg

        N   L   N   V   M   I   F   G   G   G   T   K   V   F   V   Q
J4.1 G AAT TTA AAC TAC ATG ATC TTT GGT GGA GGA ACA AAA GTT TTC GTC CAA G gtaagtacctga

J4.2 ccagcccaaagttgttaaactgctagtcacttcatgaaggtaggagaaggctgagctcactgtaggagcctgaaccagtgtg

        S   S   G   W   K   M   F   G   E   G   T   K   L   I   V   A   P   C
J4.2 AC AGC TCA GGT TGG AAG ATG TTT GGA GAA GGA ACT AAG CTC ATA GTA GCT CCC TGT G gtaagt
```

Fig. 2. Nucleotide and amino acid sequences of *J* genes belonging to sheep *TRG1@* (a) and *TRG2@* (b) loci. *J2.1*, *J6.1* and *J4.1* genes are placed based on previous report (Miccoli et al., 2003). Uppercase coding sequences are indicated in triplet form. Deduced amino acid sequences are presented on the top and the conserved K-X-F-G-X-G-X-K-L motifs are boxed. Heptamers and nonamers and the donor splice sites in the 5′ and 3′, respectively are underlined. The translation stop codon in the amino acid sequence is indicated (*).

Table 1
Analysis of the expression of sheep *J* genes

Gene name	Adult	Fetus
J5.1[a]	+	+
J5.2	+	+
J5P[b]	−	−
J3.1[a]	+	+
J3.2[b]	−	−
J1.1[a]	+	+
J1.2[c]	+	+
J6.1[a]	+	+
J6.2	+	+
J2.1[a]	+	+
J2.2	+	+
J4.1[a]	+	+
J4.2	+	+

The results of RT-PCR performed on polyA+RNA extracted from fetus (F) and adult (A) spleen, meant do determine whether they are represented in TCR γ chain repertoire, are shown. (+) To indicate that a PCR product was observed. (−) To indicate that a PCR product was not observed. The gene-specific oligonucleotide primers used are described in Materials and methods; their efficiency was determined on genomic clones and negative or low results were repeated twice.

[a] Control (Hein and Dudler, 1997).

[b] Vestigal form.

[c] Coding region has stop codon.

3.2. Transcription assay of J segments

To examine whether the newly identified *J5.2*, *JP*, *J3.2* and *J1.2* from *TRG1@* and *J6.2*, *J2.2* and *J4.2* from *TRG2@* are transcribed and spliced, we performed RT-PCR reactions on first-strand cDNA prepared from adult and fetus spleen polyA(+)RNA. Reactions included *J5.1*, *J3.1*, *J1.1*, *J6.1*, *J2.1* and *J1.1* as control for each cluster, since they were previously reported as a cDNA isolated from a sheep library (Hein and Dudler, 1993).

The RT-PCR data on *TRG1@ J* segments indicates that the only segments that, accordingly to the definition of vestigial, does not appear to be transcribed either in the fetus or in the adult, are *J3.2* and *J5P* (Table 1). A premature stop codon does not affect transcription as is the case for *J1.2*. Now from the RT-PCR data presented in Table 1, it is clear that there is an mRNA species containing the *J1.2* exon, a finding very similar to that of the *VP* exon (see also Section 3.3 companion paper). In this case however it is conceivable that the stop codon is too close to the *J1.2* segment 3′ splice site and could thus escape the Nonsense-Mediated mRNA Decay (NMD) mechanism. Taken together, this data indicates that only *J* genes in which two or more of the key sequence features (RSS, 3′ splice sites, and core sequences) are missing or severely altered fail to be transcribed. Minor differences in the RSS, 3′ splice sites and core sequences are assumed to represent normal variation and not lethal mutations.

3.3. Comparison of STAT consensus motifs in the promoters for germline transcription of sheep and human J segments

One of the objectives of the present research was to study the 5′ regions of the *J* segments of the sheep genome which play a key role to determine the "recombinational accessibility" of the gamma locus and compare them with the already identified human ones. While the enhancer elements govern the general accessibility of each locus in the *V(D)J* recombination, the germline transcription promoters play a crucial role in gene recombination in that they control local accessibility (Sikes et al., 1999). In particular the IL-7 and STAT5 proteins are required for accessibility to the *TRG* locus. IL-7 binding to IL-7R activates the STAT5A and STAT5B proteins; these latter dimerize forming both homodimers and heterodimers, traslocate into the nucleus, bind to the consensus binding motif (TTCNNNGAA) in the 5′ regions of *J* segments and induce germline transcription in these regions (Ye et al., 1999; Schlissel et al., 2000). We aligned the two human with the six ovine germline transcription promoters located at 5′ of the relevant *C*-distal *J* segments (Fig. 4). Sheep and human 5′ *J* germline promoters show a relatively high sequence similarity, i.e., 75% for the first 200 bp and 72% for the last 150 bp, that include RSSs. Two STAT consensus motifs were found about 300 bp upstream of sheep *J1.2*, *J3.1*, *J5.1*, and human *J1.1* (Fig. 4a), and are conserved in both location and sequence. The only difference observed concerns sheep *J3.1* second motif (TTANNNGAA) and human *J1.1* and *J2.1* (Fig. 4a,b) second motif (TTCNNNGTA). This latter is considered atypical (Lee et al., 2001). It was impossible to align the three 5′ *J* regions belonging to the sheep *TRG2@* locus either with human 5′ *J1.1* or with the 5′ regions belonging to the Joining segments of *TRG1@* due to two reasons: first, they are too divergent to be aligned properly and, second, the 5′ *J* regions of *TRG2@* contain a copy of different repetitive elements inserted between the STAT motifs and the Joining segment. Strikingly, an AluSx (275 bp) sequence is inserted in human *J2.1*, a Bov-tA3 SINE (126 bp) in sheep 5′ *J2.1* and 5′ *J6.1* and a Bov-A2 SINE (272 bp) with a truncated Bov-B LINE (62 bp) in *J4.2* (Fig. 4b).

Looking at the alignments, it is evident that there is high sequence conservation in the regions immediately upstream of the STAT motifs and the RSSs; however, the conservation becomes less clear in the region that includes the initiation sites of germline transcription. The phylogenetic conservation of STAT consensus motifs in all of the six sheep 5′ *J* germline promoters belonging to both *TRG@* loci indicates that these are under the strong influence of IL-7 and STAT5 signaling.

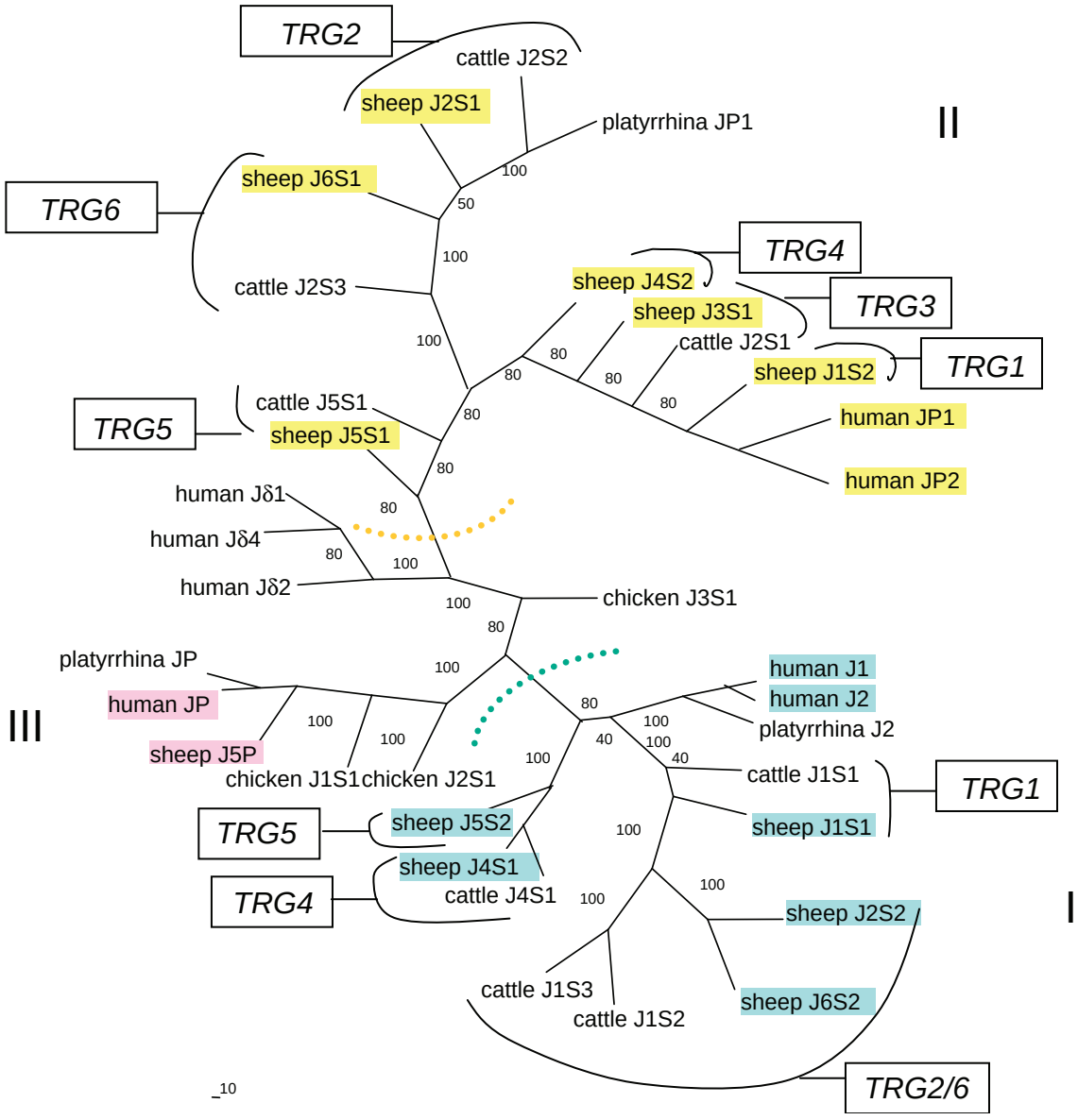

Fig. 3. Phylogenetic tree calculated on the nonsynonymous (first plus second) codon positions of *J* genes from cattle (this paper, Ishiguro et al., 1993; Hein and Dudler, 1997; Takeuchi et al., 1992), sheep (this paper; Miccoli et al., 2003), human (IMGT Database, 2003), platyrrhina (Daubenberger et al., 2001) and chicken (Six et al., 1996), constructed by the neighbor-joining method using Maximum Parsimony (MP) method as implemented in PAUP*4.04 b. Numbers at nodes indicate the bootstrap values (100 replicates were performed). The recommended IMGT nomenclature for the cattle, sheep, human, and chicken is used. The nomenclature for platyrrhina genes is from the literature. Accession numbers for the sequences are reported in Materials and methods. Rectangular boxes indicate the recombination clusters (*TRG*) to which cattle and sheep *J* genes belong. Human and sheep *J* genes are colored differently: light-blue (*C*-proximal-I), yellow (*C*-distal-II) and pink (intermediate-III).

3.4. Comparison of sheep and human recombination signal sequences

Differences in RSS heptamer and nonamer elements, as well as in the less well conserved spacer and coding flanking sequence, can influence recombination frequency and, consequently, the relative representation of different gene segments in the primary antigen receptor repertoire. RSS are composed of a conserved palindromic heptamer, whose consensus is CACAGTG, and an AT-rich nonamer (ACAAAAACC) separated by less conserved 12(±1) or 23(±1) bp spacers. *V–J* recombination occurs only between pairs of gene segments flanked, respectively, by RSS containing 23 (23RSS) and 12 (12RSS) bp spacers, which is referred to as the 12/23 rule. To relate sequence conservation and function (i.e., recombinational capability), human RSS sequences of rearranged and transcribed *V* and *J* gamma segments were retrieved from the IMGT Database and compared to those of sheep genes found transcribed in our RT-PCR assays, including the *J1.2* and *VP* pseudogenes, but excluding *J3.2* and *J5P*. 12 RSS and 23 RSS sequences were aligned with ClustalW. The alignment

introduces gaps to promote optimal alignment, as spacers may vary in length from 11 to 13 bp for 12 RSS and from 22 to 24 bp for 23 RSS. The alignment data have been used to obtain a consensus by the program WebLogo (Crooks et al., 2004).

The majority of human (7/9) and sheep (7/12) 23 RSS show a single base pair gap in the spacer (24 bp), while there are no sheep or human sequences with an 11 or 13 bp spacer separating the heptamer and the nonamer in 12 RSS. Studies using recombination substrates confirmed that 11 or 13 bp spacers, or 22 or 24 bp spacers, recombine less frequently than the wild type 12 or 23 bp spacers (Feeney et al., 2000). As expected, the first 3 bp of the heptamer, CAC, which flank the coding end, are highly (>99%) conserved (Fig. 5); exceptions are human *JP* (not shown), and *V3.2* and *J1.2*. The last four positions are well conserved only in sheep 12RSS (Fig. 5c). Within the consensus nonamer most individual positions are less conserved, particularly in sheep 23 RSS. However, as indicated by Sakano, a core of three As flanked by non-As appears to be a key feature of the sequence (Akamatsu et al., 1994). The A at the fifth position from the heptamer (A^5) has been shown to be significantly conserved in immunoglobulin (Ramsden et al., 1994) as well as in TCR (Glusman et al., 2001b) RSSs. We found A^5 to be conserved in human 23 RSS, sheep 12 RSS, and human 12 RSS (with the exception of *JP*), but not in sheep 23 RSS. In accord with Ramsden et al., we have found that at the second position of the spacer a T (T^2) is present in 7/9 human 23 RSS and in 7/

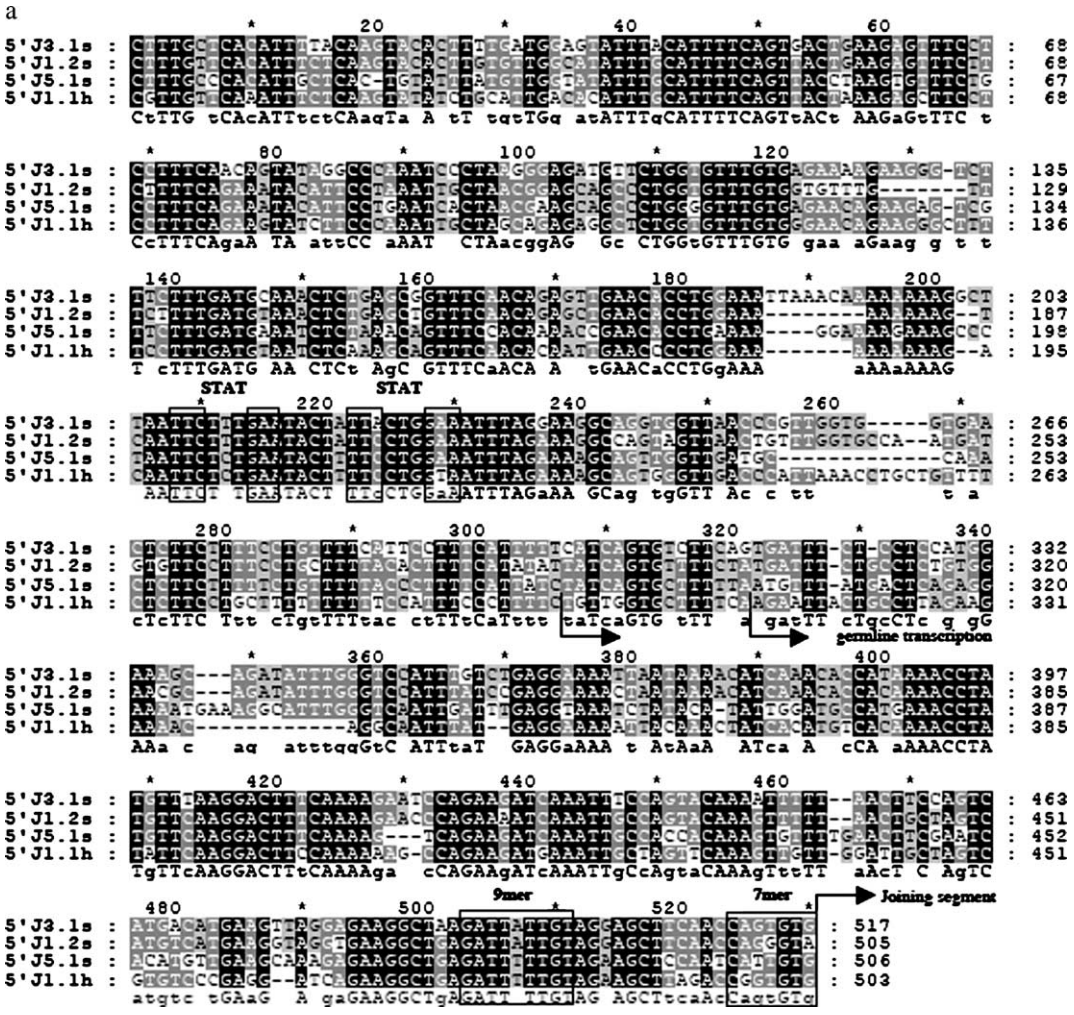

Fig. 4. Nucleotide sequence alignment of the 5′ *J* regions containing germline promoters of (a) sheep *TRG1@* locus and human *J1.1*, and (b) sheep *TRG2@* locus and human *J2.1*. The alignments were generated using ClustalW. Identical nucleotide positions in all sequences are shaded in black, highly conserved positions are shaded in grey. Frames indicate characteristic STAT consensus motifs and the nonamer and heptamer recombination signals. The initiation sites of germline transcription (Lee et al., 2001) and boundaries of *J* segments coding regions are shown by arrows (⟶). An Alu repetitive sequence (⇢), a SINE Bov-A2 (⟶), and a SINE Bov-tA3 (⟶), inserted in the human 5′ *J2.1*, and in the sheep 5′ *J4.2*, 5′*J2.1* and 5′ *J6.1*, respectively are indicated. Bov-A2 together with Bov-tA and Bov-B represent three SINE families in the Ruminantia (Nijman et al., 2002). Human *J1.1* and *J2.1* correspond to *JP1* and *JP2* from the IMGT Database (2003).

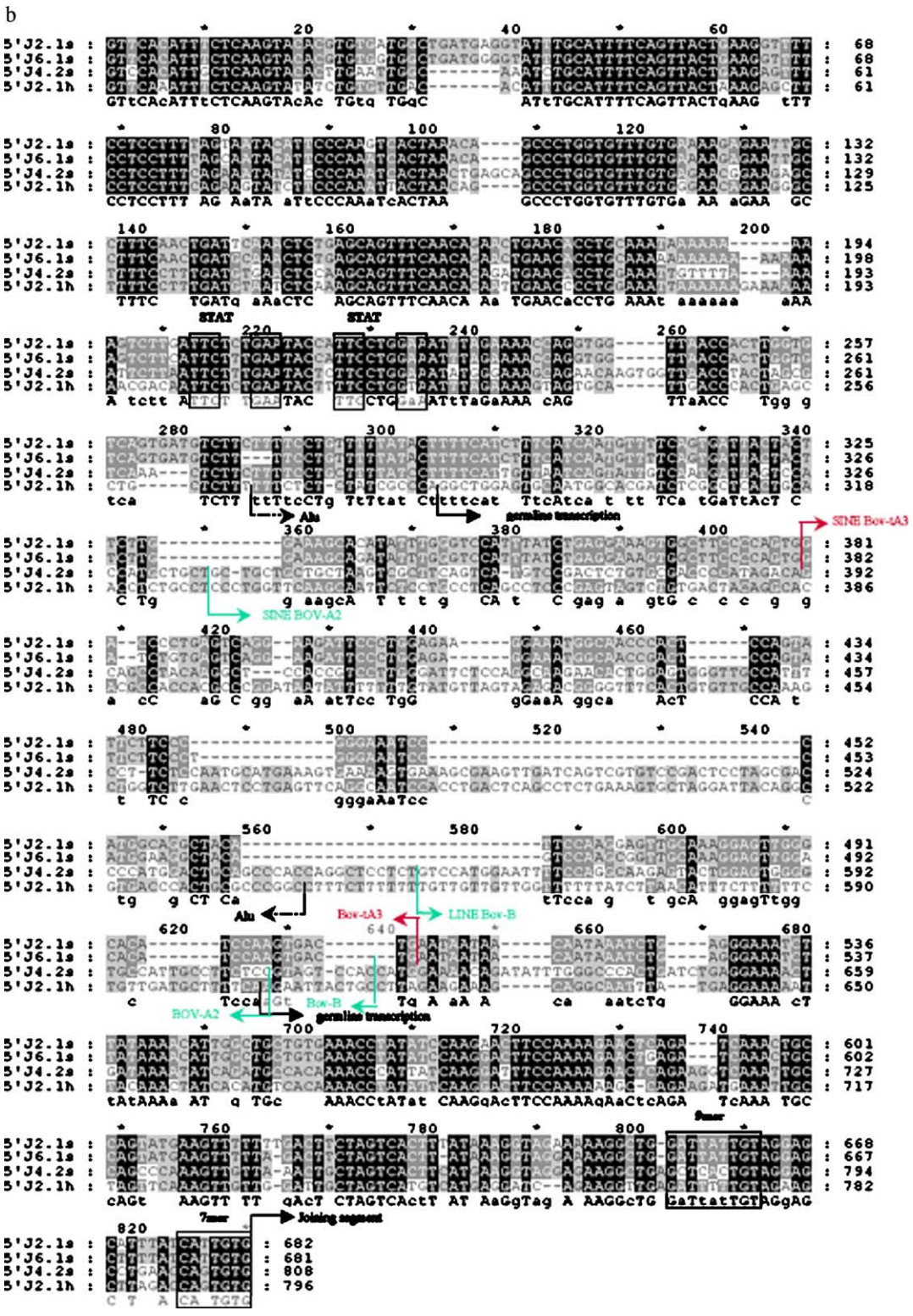

Fig. 4 (*continued*).

12 sheep 23 RSS, and in all sheep and human 12 RSS. Feeney et al. predicted also that the presence of a G at the second position was the change most likely to originate a dramatic decrease in recombination.

4. Discussion

The objective of this study was to conduct a detailed analysis of the joining region standing between the *V* and *C*

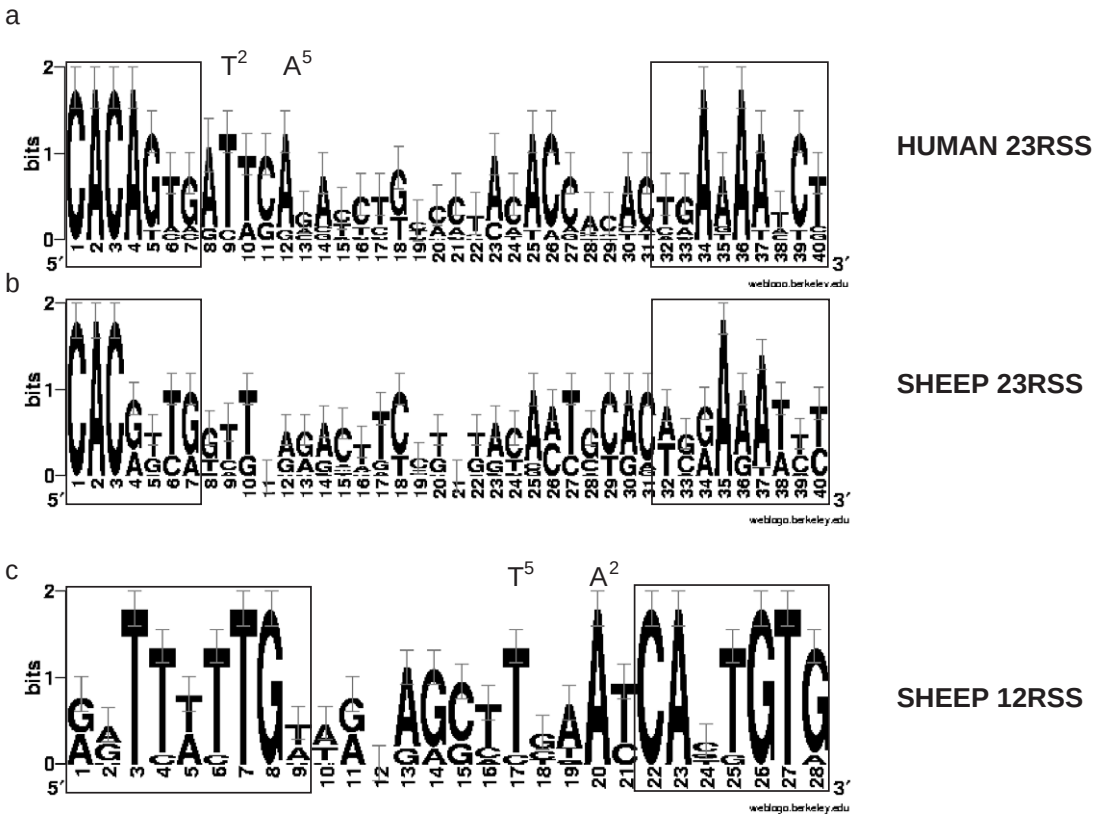

Fig. 5. Sequence Logos of human (a) and sheep (b) 23RSS and sheep 12RSS (c) of all rearranged and transcribed *TRG V* and *J* gene segments, obtained by WebLogo. The heptamer and the nonamer signals are boxed. Taller letters indicate more highly conserved nucleotides. Consensus RSS on the 3′ side of *IgV* and *TRV* consists of a highly conserved heptamer (CACAGTG) and a slightly less well conserved nonamer (ACAAAAACC) separated by a 23(±1) bp spacer. Consensus RSS on the 5′ side of *J* segments is the complement (5′-GGTTTTTGT, 12 bp spacer, CACTGTG- coding region).

genomic regions in *TRG1@* and *TRG2@* loci; this analysis includes the germline structure identification and the profile expression detection of *J* segments in fetus and adult.

The evolutionary relationship between the *C* and *V* gene segments within each *TRG* cluster, demonstrated that in cattle and sheep the *V* phylogenetic topology correlates to the rearrangement process of *V* segments with the relevant *C*. The evolutive analysis conducted here indicates how the groupings of the phylogenetic tree place each of the sheep *J* gene segments within the cluster they belong to. Moreover, we argue that many still unknown features of the genomic organization of *J* genes in cattle, e.g., their belonging to a given cluster and their distance from *C*, can be inferred on the basis of the distribution of the known cattle and sheep sequences in the tree. Thereafter we propose that the different evolutive behaviour of *C*-distal *J*s as compared to the behaviour of *C*-proximal *J*s is strictly correlated to the physical proximity of the germline transcription promoters and then to the recombinational machinery.

The peculiar genomic organization of *Bovidae TRG* loci in clusters is likely to imply a biased association of *V* and *J* gene segments. This is in agreement with the increased variability in sheep 23RSS since it could further favor specific *V–J* rearrangements by preferentially recruiting rearrangement partners. Although the RSS associated with each *V*, and *J* element are phylogenetically conserved, they usually deviate considerably from the consensus, especially in the nonamer and spacer (Cowell et al., 2003). With the availability of complete mouse and human TCR locus sequences, it became possible to study the TCR-specific constraints on heptamer, spacer, and nonamer sequences (Glusman et al., 2001b); however, no analysis of the *TRG* locus RSS has been conducted so far. The species- and locus-specific comparative analysis of the 12 and 23 RSS of sheep and human *TRG* loci here presented, emphasizes a greater conservation in sheep 12 RSS rather than in 23 RSS; this result is not surprising since the initial recruitment of recombination activating genes (RAG) products is supposed to require at least one relatively high-affinity RSS per recombination event. In fact, the RAG1/2 multimer has been shown to assemble on an isolated RSS (12 RSS); then the complementary RSS (23 RSS) enters the complex as naked DNA (Jones and Gellert, 2002). At this step, pairwise RSS compatibility may have an impact on *V(D)J* recombination efficiency and specificity.

To investigate the “recombinational accessibility” of gamma locus in the *V–J* recombination process in *Bovidae*, we have identified, in sheep, the promoters for germline transcription located upstream of the six *C*-distal *J* segments and compared them with the corresponding two

human sequences. These promoters contain typical binding sites for activated STAT proteins. It has been proposed that STAT5 controls the accessibility to the *TRG* locus by inducing germline transcription and therefore making chromatin accessible not only to the transcriptional, but also to the recombinational machinery. The identification of these transcriptional regulatory elements may be important for understanding the expression of γ genes in the "γδ high" animals. Our findings show a strong phylogenetic conservation of STAT consensus motifs in all of the six sheep 5′ *J* germline promoters of both *TRG* loci, indicating they are under the strong influence of IL-7 and STAT signaling. After entering the thymus, T cell precursors receive a signal from IL-7R to induce the rearrangement and transcription of the *TRG* locus. This will help them to commit and maintain themselves into the γδ T cell lineage. This is in accord with the hypothesis that during the evolution of the immune system, γδ T cells may have emerged first by a simple mechanism whereby IL-7, which is produced by the epithelial cells of the skin and the intestine, induces *V(D)J* recombination and clonal cell expansion.

The conservation of the six STAT motifs found in the sheep, might be due to the following reasons: i) *TRG* genes respond to IL-7 produced by epithelia where γδ T are the most represented T lymphocytes, this being the simplest mechanism to induce recombination; ii) functional needs such as immunoprotective functions, including the first defensive barrier in the epithelia of the digestive tract of ruminants, are likely to have induced genome modifications with reiterations of genomic structures resulting in functional duplications into gamma loci; iii) consequently a larger number of gamma genes, each containing its 5′ *J* promoter germline, led to redundant recombinational events, which in turn produced functional γ transcripts with highly diversified variable genes. The above mentioned sequence strongly indicates that γδ T cells differentiation dominated over the αβ T cells in the epithelia, this consequently creating a constitutive prevalence of γδ T also in peripheral blood.

Acknowledgments

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 P. Barsanti for critical review in the preparation of this manuscript. We also thank anonymous referees for critical comments on a previous version of this article.

References

Adachi, J., Hasegawa, M., 1992. Amino acid substitution of proteins coded for mitochondrial DNA during mammalian evolution. Jpn. J. Genet. 67, 187–197.

Akamatsu, Y., et al., 1994. Essential residues in V(D)J recombination signals. J. Immunol. 153, 4520–4529.

Ciccarese, S., Lanave, C., Saccone, C., 1997. Evolution of T-cell receptor γ and δ constant region and other T-cell related proteins in the human–rodent–artiodactyl triplet. Genetics 145, 409–419.

Corcoran, A.E., Riddell, A., Krooshoop, D., Venkitaraman, A.R., 1998. Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor. Nature 391, 904–907.

Cowell, L.G., Davila, M., Yang, K., Kepler, T.B., Kelsoe, G., 2003. Prospective estimation of recombination signal efficiency and identification of functional cryptic signals in the genome by statistical modeling. J. Exp. Med. 197, 207–220.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14 (6), 1188–1190.

Daubenberger, C.A., et al., 2001. Functional and structural similarity of V gamma 9V delta 2 T cells in humans and *Aotus* monkeys, a primate infection model for *Plasmodium falciparum* malaria. J. Immunol. 167 (11), 6421–6430.

Dudley, E.C., Girardi, M., Owen, M.J., Hayday, A.C., 1995. αβ T cells and γδ T cells can share a late common precursor. Curr. Biol. 5, 659–669.

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

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

Glusman, G., et al., 2001. Comparative genomics of the human and mouse T cell receptor loci. Immunity 15, 337–349.

Hayday, A.C., 2000. γδ cells: a right time and a right place for a conserved third way of protection. Annu. Rev. Immunol. 18, 975–1026.

He, Y.-W., Malek, T.R., 1996. Interleukin-7 receptor alpha is essential for the development of γδ T cells but not natural killer cells. J. Exp. Med. 184, 289–293.

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

Hoffman, E.S., Passoni, L., Dudley, E.C., Girardi, M., Hayday, A.C., 1998. The αβ/γδ lineage decision: the molecular biology of B-cell and T-cell development. In: Monroe, J.G., Rothenberg, E. (Eds.), Humana, Totowa, N.J, pp. 367–396.

Huang, J., Durum, S.K., Muegge, K., 2001. Cutting edge: histone acetylation and recombination at the TCR γ locus follows IL-7 induction. J. Immunol. 167, 6073–6077.

Jones, M.J., Gellert, M., 2002. Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. EMBO J. 21, 4162–4171.

Krangel, M.S., 2003. Gene segment selection in *V(D)J* recombination: accessibility and beyond. Nat. Immunol. 4, 624–630.

IMGT Database: http://imgt.cines.fr:8104/.

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

Lee, H.-C., Ye, S.-K., Honjo, T., Ikuta, K., 2001. Induction of germline transcription in the human TCR γ locus by STAT5. J. Immunol. 167, 320–326.

Miccoli, M.C., Lipsi, M.R., Massari, S., Lanave, C., Ciccarese, S., 2001. Exon-intron organization of TRGC genes in sheep. Immunogenetics 53, 416–422.

Miccoli, M.C., et al., 2003. Evolution of TRG clusters in cattle and sheep genomes as drawn from the structural analysis of the ovine TRG2 locus. J. Mol. Evol. 57, 52–62.

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

Nijman, I.J., van Tessel, P., Lenstra, J.A., 2002. SINE Retrotransposition during the evolution of the pecoran ruminants. J. Mol. Evol. 54, 9–16.

Ramsden, D.A., Baetz, K., Wu, G.E., 1994. Conservation of sequence in recombination signal sequence spacers. Nucleic Acids Res. 22, 1785–1796.

Rzhetsky, A., Nei, M., 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. 9, 945–967.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Schlissel, M.S., Durum, S.D., Muegge, K., 2000. The interleukin 7 receptor is required for T cell receptor gamma locus accessibility to the *V(D)J* recombinase. J. Exp. Med. 191 (6), 1045–1050.

Sikes, M.L., Suarez, C.C, Oltz, M., 1999. Regulation of *V(D)J* recombination by transcriptional promoters. Mol. Cell. Biol. 19, 2773–2781.

Sitnikova, T., 1996. Bootstrap method of interior-branch test for phylogenetic trees. Mol. Biol. Evol. 13, 605–611.

Six, A., et al., 1996. Characterization of avian T-cell receptor γ genes. Proc. Natl. Acad. Sci. 93, 15329–15334.

Swofford, D.L., 2000. PAUP. Phylogenetic Analysis Using parsimony (*and other methods), Version 4. Sinauer Associates, Sunderland, MA.

Takeuchi, N., Ishiguro, N., Shinagawa, M., 1992. Molecular cloning and sequence analysis of bovine T-cell receptor γ and δ chain genes. Immunogenetics 35, 89–96.

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. Nucleic Acids Res. 22 (22), 4673–4680.

Ye, S.K., et al., 1999. Induction of germline transcription in the TCRgamma locus by Stat5: implications for accessibility control by the IL-7 receptor. Immunity 11, 213–223.