

Research article

Open Access

## Genomic analysis reveals extensive gene duplication within the bovine TRB locus

Timothy Connelley\*<sup>1</sup>, Jan Aerts<sup>2</sup>, Andy Law<sup>1</sup> and W Ivan Morrison<sup>1</sup>

Address: <sup>1</sup>The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Roslin, EH25 9RG, UK and <sup>2</sup>Genome Dynamics and Evolution Group, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK

Email: Timothy Connelley\* - [timothy.connelley@ed.ac.uk](mailto:timothy.connelley@ed.ac.uk); Jan Aerts - [jan.aerts@sanger.ac.uk](mailto:jan.aerts@sanger.ac.uk); Andy Law - [andy.law@ed.ac.uk](mailto:andy.law@ed.ac.uk); W Ivan Morrison - [Ivan.Morrison@ed.ac.uk](mailto:Ivan.Morrison@ed.ac.uk)

\* Corresponding author

Published: 24 April 2009

Received: 28 August 2008

BMC Genomics 2009, 10:192 doi:10.1186/1471-2164-10-192

Accepted: 24 April 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/192>

© 2009 Connelley et al; licensee BioMed Central Ltd.

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

### Abstract

**Background:** Diverse TR and IG repertoires are generated by V(D)J somatic recombination. Genomic studies have been pivotal in cataloguing the V, D, J and C genes present in the various TR/IG loci and describing how duplication events have expanded the number of these genes. Such studies have also provided insights into the evolution of these loci and the complex mechanisms that regulate TR/IG expression. In this study we analyze the sequence of the third bovine genome assembly to characterize the germline repertoire of bovine TRB genes and compare the organization, evolution and regulatory structure of the bovine TRB locus with that of humans and mice.

**Results:** The TRB locus in the third bovine genome assembly is distributed over 5 scaffolds, extending to ~730 Kb. The available sequence contains 134 TRBV genes, assigned to 24 subgroups, and 3 clusters of DJC genes, each comprising a single TRBD gene, 5–7 TRBJ genes and a single TRBC gene. Seventy-nine of the TRBV genes are predicted to be functional. Comparison with the human and murine TRB loci shows that the gene order, as well as the sequences of non-coding elements that regulate TRB expression, are highly conserved in the bovine. Dot-plot analyses demonstrate that expansion of the genomic TRBV repertoire has occurred via a complex and extensive series of duplications, predominantly involving DNA blocks containing multiple genes. These duplication events have resulted in massive expansion of several TRBV subgroups, most notably TRBV6, 9 and 21 which contain 40, 35 and 16 members respectively. Similarly, duplication has led to the generation of a third DJC cluster. Analyses of cDNA data confirms the diversity of the TRBV genes and, in addition, identifies a substantial number of TRBV genes, predominantly from the larger subgroups, which are still absent from the genome assembly. The observed gene duplication within the bovine TRB locus has created a repertoire of phylogenetically diverse functional TRBV genes, which is substantially larger than that described for humans and mice.

**Conclusion:** The analyses completed in this study reveal that, although the gene content and organization of the bovine TRB locus are broadly similar to that of humans and mice, multiple duplication events have led to a marked expansion in the number of TRB genes. Similar expansions in other ruminant TR loci suggest strong evolutionary pressures in this lineage have selected for the development of enlarged sets of TR genes that can contribute to diverse TR repertoires.

## Background

Diverse  $\alpha\beta$ TR repertoires are crucial to the maintenance of effective T cell-mediated immunity [1]. Estimates based on direct measurement indicate that in humans and mice individuals express a repertoire of approximately  $2 \times 10^7$  [2] and  $2 \times 10^6$  [3] unique  $\alpha\beta$ TRs respectively. As with the other antigen-specific receptors (IG of B cells and  $\gamma\delta$ TRs of  $\gamma\delta$ T cells) diversity is generated in lymphocytic precursors by somatic recombination of discontinuous variable (V), diversity (D – TRB chains but not TRA chains) and joining (J) genes to form the membrane-distal variable domains. Diversity is derived from both the different permutations of V(D)J genes used to form the TRA and TRB chains expressed by individual thymocytes (combinatorial diversity) and also by the activity of terminal deoxynucleotide transferase and exonuclease at the V(D)J junction during recombination (junctional diversity). Consequently, much of the diversity is focused in the third complementarity determining region (CDR3), which is encoded by the V(D)J junction and forms the most intimate association with the antigenic peptide component of the peptide-MHC (pMHC) ligand of  $\alpha\beta$ TRs, whereas the CDR1 and CDR2 of the TRA and TRB chains, that predominantly interact with the MHC, are encoded within the germline V genes [4,5].

TRB chain genes are located in the TRB locus, which in humans is ~620 Kb long and situated on chromosome 7 and in mice is ~700 Kb and located on chromosome 6 [6-8]. In both species, the organisation of TRB genes is similar, with a library of TRBV genes positioned at the 5' end and 2 DJC clusters (each composed of a single TRBD, 6–7 TRBJ and a single TRBC gene) followed by a single TRBV gene with an inverted transcriptional orientation located at the 3' end [9,10]. The germline repertoire of TRBV genes in humans is composed of 65 genes belonging to 30 subgroups (genes with > 75% nucleotide identity), whilst in mice the repertoire comprises 35 genes belonging to 31 subgroups [10-12]. The disparity between the number of TRBV genes in the 2 species is the result of multiple duplication events within the human TRB locus, most of which have involved tandem duplication of blocks of DNA (homology units) containing genes from more than one subgroup [10,13].

V(D)J recombination is initiated by site-specific DNA cleavage at recombination signal sequences (RSs) mediated by enzymes encoded by recombination activating genes (RAG) 1 and 2 [14]. RSs comprise conserved heptamer and nonamer sequences separated by spacers of either 12 bp (12-RS – located 5' to TRBD and TRBJ genes) or 23 bp (23-RS – located 3' to TRBV and TRBD genes). Correct V(D)J assembly is achieved as recombination can only occur between genes flanked with RS of dissimilar length (the '12/23 rule') and direct TRBV/TRBJ recombi-

nation is prohibited by the 'beyond 12/23' phenomenon [15-17]. As with other antigen-specific receptor loci, recombination in the TRB locus is under strict lineage-, stage- and allele-specific regulation associated with control of RAG accessibility to RSs mediated through alterations in chromatin structure (the 'accessibility hypothesis') [18-20]. Numerous studies have shown that both the TRB enhancer (E $\beta$ ) and transcriptional promoters within the TRB locus serve as RAG accessibility control elements, playing a critical role in regulating chromatin structure and therefore recombination of TRB genes [21-27].

Current knowledge of the TRB gene repertoires of agriculturally important artiodactyl species (e.g. pigs, cattle and sheep) is limited. Published analyses of rearranged TRB transcripts have demonstrated the expression of 19 TRBV subgroups in pigs [28,29], 13 subgroups in sheep [30] and 17 subgroups in cattle, some of which have undergone extensive duplication [31-34]. Information on the genomic organisation of the TRB loci is predominantly restricted to the DJC region, which in the pig was found to be composed of 2 tandemly arranged DJC clusters [35] but in sheep contained 3 tandemly arranged DJC clusters [36]. Preliminary analysis of a BAC clone corresponding to part of the DJC region indicates that in cattle the DJC region may also consist of 3 DJC clusters [37].

Sequencing of the complete TRB loci in human and mice allowed the repertoire of TRB genes in these species to be fully characterised and also permitted analysis of the organisation, regulation and evolution of this immunologically important locus [9,10]. In this study we have used the sequence of the third bovine genome assembly (Btau\_3.1) to further study the bovine TRB repertoire and TRB locus. Although the sequence of the TRB locus is incomplete, the results reveal that duplication within the locus has been prolific leading to a massive expansion of TRBV gene numbers and the generation of a third DJC cluster. Furthermore, the analysis shows that the genomic organisation of the TRB locus and the non-coding elements that regulate TRB expression are highly conserved in cattle when compared to that of humans and mice.

## Results

### **Extensive duplication has generated a large germline repertoire of bovine TRBV genes**

A total of 134 TRBV genes, distributed over 5 scaffolds was identified in Btau\_3.1 (Additional File 1). Consistent with data from fluorescent *in situ* hybridisation studies [38], the majority of the TRBV genes were located on 2 scaffolds (Chr4.003.105 [91 TRBV] and Chr4.003.108 [21 TRBV]) mapped to chromosome 4, whilst the remaining genes were located on 3 scaffolds (ChrUn.003.1717 [18 TRBV], ChrUn.003.4367 [3 TRBV] and ChrUn.003.12588 [1

TRBV]) which have not been assigned a chromosomal location. Within the scaffolds are several regions of undetermined sequence, including large areas of ~35 Kb and ~147 Kb on Chr4.003.105 and Chr4.003.108 respectively.

Each TRBV gene is composed of i) a short leader (L) exon, generally of ~50 bp, ii) a single intron of between ~80 and ~500 bp and iii) a variable (V) exon of ~300 bp, immediately flanked at the 3' end with a 23-RS. Comparison of the nucleotide sequence of each of the bovine TRBV genes with human TRBV gene sequences, revealed maximum levels of similarity between the species ranging from 71.8% to 83.15% for all except one of the bovine TRBV genes. On the basis of these results, bovine TRBV genes were considered orthologues of their most similar human

counterpart and were assigned to subgroups named according to the orthologous human subgroup (Table 1). The single bovine TRBV gene that lacked significant homology to any of the human TRBV genes displayed 76.6% identity with the murine TRBV1 gene (which lacks a human orthologue) and was placed in subgroup TRBVX. The subgroups thus established generally adhered to the definition of members within a subgroup exhibiting > 75% nucleotide sequence identity. However, the single member of the TRBV10 subgroup displayed > 75% identity to all of the TRBV6 genes and the identity between members of the TRBV9 and TRBV5 subgroups was often > 75% (data not shown). Conversely a single member of the TRBV19 subgroup (TRBV19f) showed only 63.0–64.8%

**Table 1: TRBV gene repertoires.**

Subgroup	Number of genes identified in human TRB locus		Number of genes identified in bovine TRB locus		Number of bovine genes identified from cDNA analyses
	Total	Functional	Total	Functional	
1	1				
2	1	1			
3	1–2*	1	1	1	1
4	2–3*	2–3	2	2	2
5	8	5	4	1	1
6	8–9*	6–7 (+1)†	40	20	20
7	9	5 (+2)†	2	2	3
8	2				
9	1	1	35	23	13
10	3	2 (+1)†	1	0	2
11	3	3	1	0	
12	5	3	2	2	2
13	1	1	1	0	
14	1	1	1	1	1
15	1	1	1	1	1
16	1	(+1)†	1	0	1
17	1				
18	1	1	4	0	
19	1	1	6	3	3
20	1	1	5	5	9
21	1		16	9	11
22	1				
23	1				
24	1	1	1	1	2
25	1	1	1	1	1
26	1		1	1	1
27	1	1			1
28	1	1	1	1	1
29	1	1	5	3	8
30	1	(+1)†	1	1	1
X			1	1	1
Total	64–67	40–42 (+6)	134	79	86

The total and functional repertoire of TRBV genes in the human TRB locus and in the third assembly of the bovine genome as well as the bovine TRBV gene repertoire identified from cDNA analysis. Details of the human repertoire are taken from the IMGT database <http://imgt.cines.fr>.

\*denotes variation in germline repertoire due to insertion/deletion polymorphisms and † denotes TRBV genes which have alternative functional and non-functional alleles (TRBV6-2, TRBV7-3, TRBV7-4, TRBV10-1, TRBV16, TRBV30).

nucleotide identity with the other members of this subgroup.

Of the 24 bovine subgroups present in the genome assembly, 11 have multiple members. Subgroups TRBV6, 9 and 21 have all undergone substantial expansion, having 40, 35 and 16 members respectively – together representing 68% of the total Btau\_3.1 TRBV gene repertoire. Southern blot analysis corroborates the presence of large numbers of TRBV6 and 9 genes in the genome (Figure 1).

A prominent feature of the genomic organisation of TRBV genes (Figure 2) is that members of expanded subgroups are generally intercalated with members of other expanded subgroups in a recurrent pattern. Thus, a 165 KB region of Chr4.003.105 and virtually all of scaffold ChrUn.003.1717 are composed of alternating TRBV6 and 9 genes (reflected in the similarity in the patterns of larger bands (> 4.3 Kb) obtained in southern blots of genomic DNA when hybridised with TRBV9- and TRBV6-specific probes in Figure 1), whilst the 3' end of Chr4.003.105 and the 5' end of Chr4.003.108 contain repeated units comprising TRBV18, 19, 20 and 21 genes. Dot-plot analyses indicate that this organisation has arisen through a series of complex tandem duplication events within the regions in which TRBV9 and 6 genes and TRBV18, 19, 20 and 21

genes are located (Figure 3). Six homology units, ranging in size from ~7 Kb to ~31 Kb and encompassing from 1 to 11 TRBV genes were identified. Three of these homology units (represented by the orange, dark blue and black bars in Figure 2) have undergone multiple (2–3) duplications: variation in the length of the different copies of these homology units (represented by broken lines in Figure 2), suggests that either i) distinct iterations of a duplication event have involved different components of the homology unit or ii) the different copies have been subject to different post-duplication deletions.

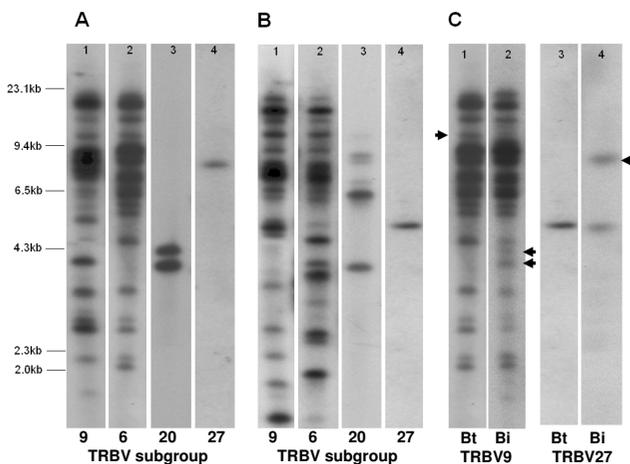
The levels of nucleotide identity between TRBV genes in corresponding positions in homology units is frequently high: 12 pairs of TRBV6 genes, 11 pairs of TRBV9 and 1 pair each of TRBV19 and TRBV20 have identical coding sequences whilst 1 pair of TRBV4 genes and 3 pairs of TRBV21 as well as 4 triplets of TRBV6 and 4 triplets of TRBV9 genes have > 97% sequence identity in the coding region.

#### Duplication has expanded the repertoire of TRBD, TRBJ and TRBC genes in the bovine genome

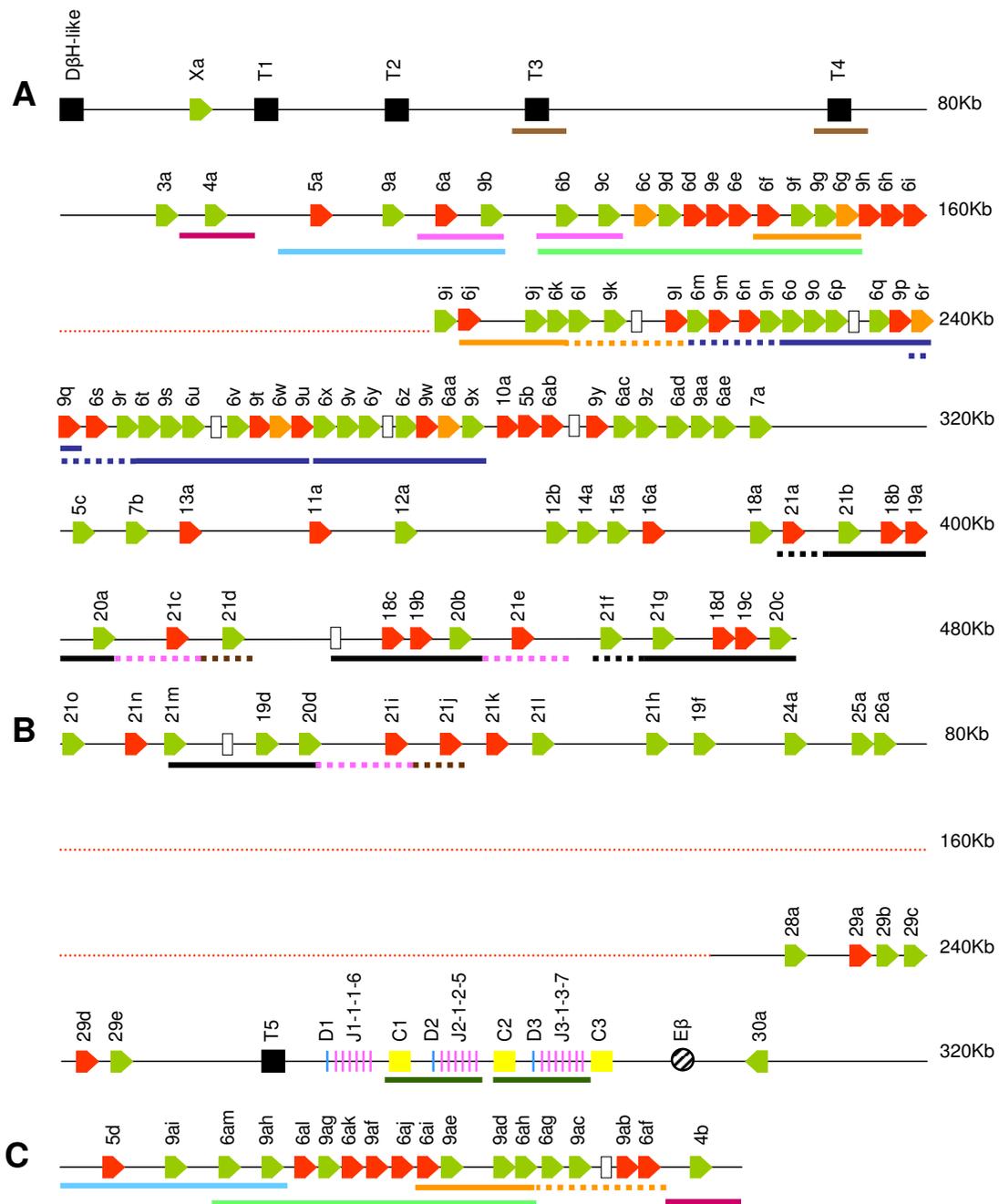
A total of 3 TRBD, 18 TRBJ and 3 TRBC genes were identified in the assembly (Additional File 1). These genes were all located within a ~26 Kb region of scaffold Chr4.003.108 and organised into 3 tandemly arranged clusters, each of ~7 Kb length and composed of a single TRBD gene, 5–7 TRBJ genes and a single TRBC gene (Figure 2). Dot-plot analysis reveals the presence of a third DJC cluster is attributable to duplication of a ~7 Kb region, one copy of which incorporates TRBC1, TRBD2 and the TRBJ2 cluster whilst the other copy incorporates TRBC2, TRBD3 and the TRBJ3 cluster (Figure 4). Numerous interruptions in the line representing the duplicated region indicate that there has been significant post-duplication deletion/insertion related modification of the duplicated region.

The nucleotide and deduced amino acid sequence of the 3 TRBD and 18 TRBJ genes as well as the flanking RS are shown in Figure 5a and 5b respectively. The 13 bp (TRBD1) or 16 bp (TRBD2 and 3) TRBD genes are G-rich and encode at least one glycine in all 3 potential reading frames with the exception of the 3<sup>rd</sup> reading frame of TRBD1. The TRBJ genes range in size from 43 bp to 59 bp in length and all encode the canonical FGXG amino acid motif that defines TRBJ genes.

As with all mammalian TRBC genes so far characterised, bovine TRBC1 and TRBC3 genes are composed of 4 exons, 3 introns and a 3'UTR region. The structure of the TRBC2 gene is anticipated to be the same but due to a region of undetermined sequence between exons 1 and 3 we were unable to identify exon 2. The exon nucleotide sequences of TRBC1 and 3 are very similar (97%), resulting in the

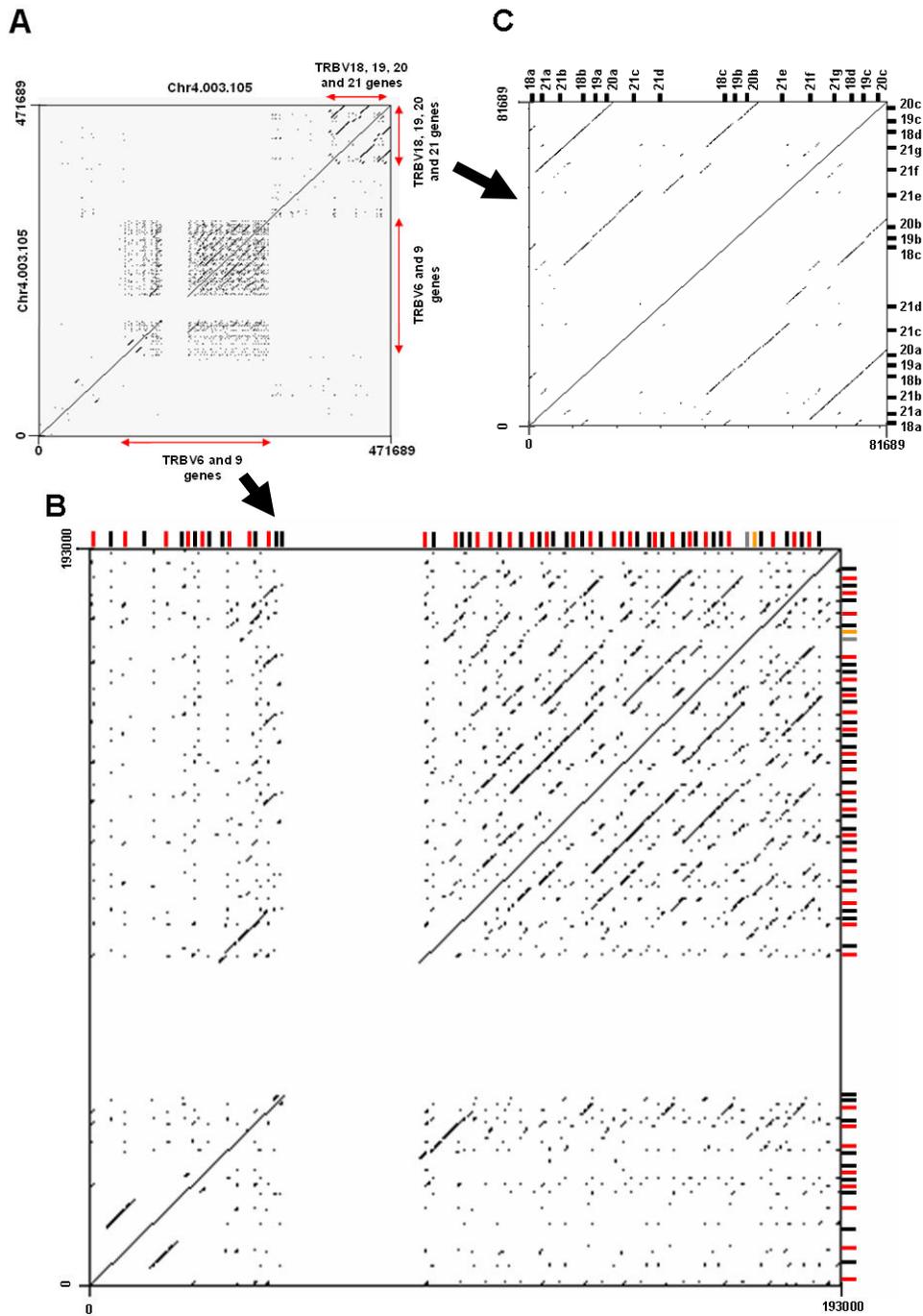


**Figure 1**  
**Southern blot analysis of bovine genomic DNA.** Genomic DNA from a *Bos taurus* animal digested with (A) HindIII or (B) SspI was hybridised with probes specific for TRBV9 (lane 1), TRBV6 (lane 2), TRBV20 (lane 3) and TRBV27 (lane 4). (C) Comparison of the banding patterns obtained from genomic DNA of a *Bos taurus* (Bt) and a *Bos indicus* (Bi) animal hybridised with a probe specific for TRBV9 (lanes 1 and 2) after digestion with HindIII and a probe specific for TRBV27 after digestion with SspI (lanes 3 and 4). Arrows indicate bands that are evident in *Bos taurus* but not *Bos indicus* DNA or vice versa.



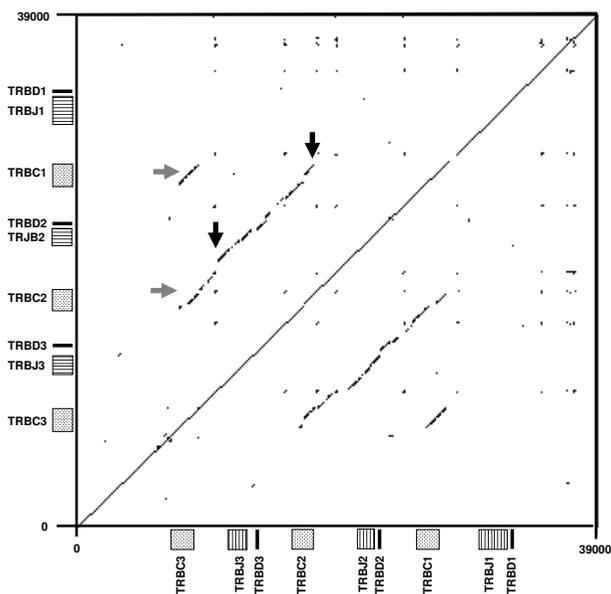
**Figure 2**

**Genomic organisation of the genes in the bovine TRB locus as described in Btau\_3.1.** The order and location of TRB genes on (A) Chr4.003.105 (B) Chr4.003.108\_RC and (C) ChrUn.003.1717. Red dotted lines represent large regions of undetermined sequence within the scaffolds. TRBV genes are classified as functional (green), open-reading frame non-functional (orange) or pseudogenes (red), and their transcriptional orientation indicated by their direction; TRBV gene 'relics' are shown as open boxes. TRBD (blue vertical lines), TRBJ (pink vertical lines) and TRBC (yellow boxes) genes are arranged into 3 DJC clusters, with a putative bovine TRB enhancer (Eβ) located 3' to the TRBC3 gene (black diagonal shading). The sizes of non TRB genes (black boxes) – dopamine-β-hydroxylase-like gene (DβH-like) and trypsinogen genes (T) are not shown to scale. Regions of duplicated DNA are indicated by the colour-coordinated boxes located beneath the scheme of gene location. Broken lines indicate regions of DNA that are not present in all copies of the duplicated region.



**Figure 3**

**Dot-plot analyses of Chr4.003.105.** (A) The TRB locus region of Chr4.003.105. The multiplicity of diagonal lines parallel to the main diagonal present in the regions containing i) the TRBV 6 and 9 genes and ii) the TRBV 18, 19, 20 and 21 genes shows that these regions have been subjected to numerous duplication events. The clear cruciform area in the TRBV 6 and 9 region (also in (B)) reflects an 35 Kb area of undetermined sequence. (B) The TRBV 6 and 9 region of Chr4.003.105. Various duplicated regions of ~7 Kb to ~31 Kb and including multiple TRBV6 (black) and TRBV9 (red) genes are evident. (C) The TRBV18, 19, 20 and 21 region of Chr4.003.105. The pattern of parallel lines in this dot-plot analysis indicates a region of DNA that includes a TRBV21, 18, 19 and 20 genes that has been duplicated twice, giving rise to 3 homology units.



**Figure 4**  
**Dot-plot analysis of the bovine DJC region on Chr4.003.108.** Duplication of a ~7 Kb region (diagonal line between black arrows) has generated a third DJC cluster. One of the homology units incorporates TRBC1, TRBD2 and the TRBJ2 whilst the other incorporates TRBC2, TRBD3 and TRBJ3. Smaller lines parallel to the main diagonal reflect the similarity in sequence of TRBC3 with TRBC1 and 2 (grey arrows).

encoded 178 amino acid products differing by only 5 residues – 3 in the extra-cellular domain and 2 in the cytoplasmic domain (Figure 6a). The incomplete sequence for TRBC2 is predicted to encode a product identical to that of TRBC1. In contrast to the high levels of pairwise identity between the exonic nucleotide sequences of all 3 TRBC genes, the nucleotide sequences of the 3<sup>rd</sup> intron and the 3'UTR regions of TRBC3 show low identity with TRBC1 and 2, whereas the latter two genes show a high level of identity (Figure 6b). The similarity in the lengths of TRBD2 and 3, the phylogenetic clustering of TRBJ2 and TRBJ3 genes in corresponding genomic positions (Figure 7) and the similarity in the sequences of the 3<sup>rd</sup> introns and 3'UTRs of TRBC1 and 2 all reflect the duplication history of the DJC region as described in Figure 4.

**The repertoire of functional TRBV, TRBD and TRBJ genes available for somatic recombination is large and phylogenetically diverse**

Computational analysis was used to predict the functional competency of the TRBV, TRBD and TRBJ genes present in the genome assembly. Fifty-five (41%) of the TRBV genes identified are predicted to encode pseudogenes (Additional File 2), whilst TRBJ1-2 (which has a 1 bp deletion

that results in the canonical FGXG motif being lost in the ORF) and TRBJ1-3 (which lacks a RS that is compatible with somatic recombination) are also predicted to be non-functional (Figure 5). Thus, the functional repertoire comprises 79 (59%) TRBV genes (comprising 66 unique coding TRBV sequences) belonging to 19 different subgroups, 3 TRBD genes and 16 TRBJ genes. This provides a potential 3168 ( $66 \times 3 \times 16$ ) unique VDJ permutations that can be used during somatic recombination of TRB chains.

Phylogenetic analysis demonstrates that the repertoire of functional TRBV genes is diverse (Figure 8), with representatives in each of the 6 phylogenetic groups (A-F) described previously in humans and mice [13,39]. The phylogenetic groupings were supported by high (99%), bootstrap values ( $P_B$ ), with the exception of group A ( $P_B = 76\%$ ). Maximum likelihood analysis using a variety of nucleotide models provides essentially similar phylogenetic clustering (data not shown) indicating the reliability of the tree presented in Figure 8. The extensive intermingling of murine, human and bovine TRBV subgroups is consistent with the establishment of distinct subgroups having occurred prior to mammalian radiation. Conversely, the formation of distinct clades of TRBV genes of orthologous subgroups from different species (e.g. TRBV6 genes from human and bovine form distinct clades) indicates that duplication within subgroups has predominantly occurred post-speciation. Despite this and the substantial disparity in the number of functional TRBV genes present in the 3 species, the distribution amongst the different phylogenetic groups is similar (Figure 8b). Phylogenetic groups C and F form a minor component of the functional TRBV repertoire, whilst the contributions from groups B and D are also fairly modest. In contrast, group E and to an even greater extent group A, are over-represented, together representing between 61.9% (in the mouse) and 81.6% (in humans) of the total functional repertoire.

Phylogenetic analysis resolves the functional TRBJ genes in human, mice and Btau\_3.1 into 12 groups (Figure 7). With the exception of group 8, each group is supported by high  $P_B$  values and is composed of orthologues that share a conserved order in the genome; consistent with the duplication history of the DJC region, TRBJ genes from both the 2<sup>nd</sup> and 3<sup>rd</sup> bovine DJC clusters group together with the respective genes from the 2<sup>nd</sup> murine and human DJC clusters. Group 8, which contains TRBJ2-3, human and murine TRBJ2-4 and bovine TRBJ3-3 and 3-4 genes is only supported by a  $P_B$  value of 57%. The diversity of the functional TRBJ repertoire across the 3 species is comparable, with humans having functional genes in each of the 12 phylogenetic groups whilst in both mice and Btau\_3.1 only 2 groups lack functional members: groups 3 (TRBJ1-

**A**

	5'flanking 12-RS				3'flanking 23-RS		
TRBD	Nonamer	Spacer	Heptamer	TRBD gene	Heptamer	Spacer	Nonamer
TRBD1	CGTTTTGT	ATAAAGCTGTAA	CGTTGTG	GGG ACA GCA GGG C G T A G G Q Q G D S R	CACGGTG	ATTCAACCCATGGGAATCCTTT	ACAAAAACC
TRBD2	CATTTTTGT	ATCACAGTATAA	CATTGTG	GGG ACT GGG GGG TGG G G T G G W G L G G G D W G V	CACAATG	ATTCAAGTAGAGGAAGTACTTTT	ACAAAAAGC
TRBD3	CATTTTTGT	ATCACAGTATAA	CATTGTG	GGA CTT CGG GGG GGG C G L R G G D F G G G T S G G	CACGATG	ATTCAAGTAGAGGAAGTACTTTT	ACAAAAAGC

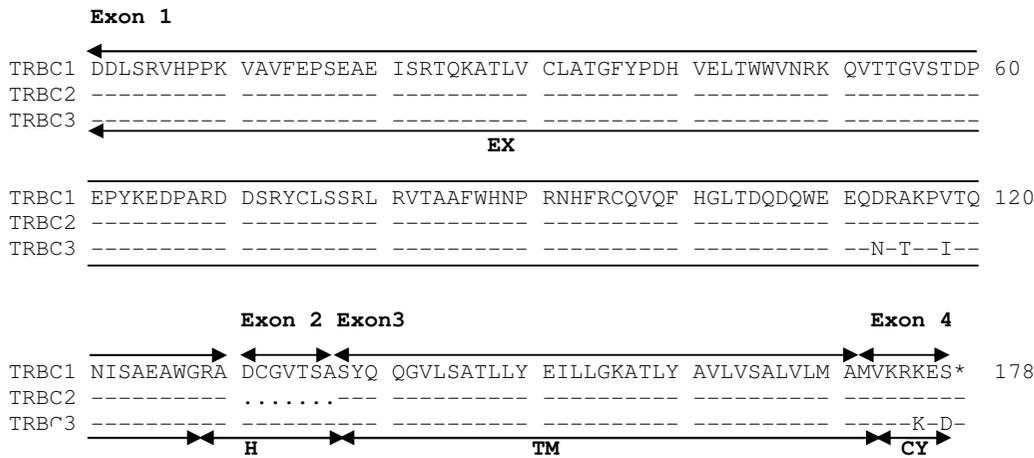
**B**

TRBJ	RS Nonamer	Spacer	Heptamer	TRBJ gene
TRBJ1-1	GCTTTCCTC	CTTGCTCCGCTT	CACTGTG	TG AAC ACT GAG GTT TTC TTT GGA AAA GGC ACC AGG CTC ACG GTT GTA G N T E V F F G K G T R L T V V
TRBJ1-2	CGTTGTAGG	GTGGCTGTGTTT	TGATGTG	T TAT GAC TAT CAC TTC GGC CCA GGG ACC AAG CTG ACG GT(T)† GTA G Y D Y H F G P G T K L T V V
TRBJ1-3	GTTTTGAAG	TGAATCTGGAG	<u><b>TCTTGGC</b></u>	T TTG GAG ACT TCT ACT TTT TTT GGG GTG GGG GGC GGG GAG GGA AGA CGG CTC GCT GCG G L E T S T F F G V G G G E G R R L A A
TRBJ1-4	AGTTTCCTC	ACTAGTCTTTAG	GGTTGTG	TG ACT AAT GAT AGA CTC TAT TTC GGC AAC GGG ACG AAG CTT TCC GTC TTG G T N D R L Y F G N G T K L S V L
TRBJ1-5	GGATTTGCC	ACACCCGCATTG	TGCTGTG	G AGC AAC CAG GCA CAG AAC TTT GGA CAC GGG ACC TGG CTC GCC GTC CTG G S N Q A Q N F G H G T W L A V L
TRBJ1-6	GGTTTTGCC	ACAGCCACCTGC	AGCTGTG	G GCC TAT AAT TCG CCC CTC CAC TTT GGA ATC GGC ACC AGG CTC ATC GTG ACA G A Y N S P L H F G I G T R L I V T
TRBJ2-1	GAATCTCTG	CTAGCCCCTTCC	CACTGTG	C TCC TAT GAG GAG CAG CAC TTC GGG CCA GGG ACC AGG CTC ATT GTG CTA G S Y E E Q H F G P G T R L I V L
TRBJ2-2	AGTTTGCGC	CTGGGTCCCCAG	GGCTGTG	TG AAC GCC GTG CAG CTG TAC TTT GGA GCT GGT TCC AAG CTG ACT GTG CTG G N A V Q L Y F G A G S K L T V L
TRBJ2-3	GGTTTTTGT	CCTGGGTCCCCG	GGCTGTG	AGC CAA AGC ACT CAG TAC TTC GGC GCG GGC ACC CGG CTG TCG GTG CTA G S Q S T Q Y F G A G T R L S V L
TRBJ2-4	GGTTTTTGC	GCGGGGCTGGGG	GGCCGTG	AC TCA GAG ACG CTG TAC TTC GGG CCG GGC ACG CGG CTC CAG GTG CTA G S E T L Y F G P G T R L Q V L
TRBJ2-5	GGTTTGCTT	GCGGGACTGGGA	CTCTGTG	A GCC AGT GAG CCG TAT TTT GGC GCC GGC ACC AGG CTC ACG GTC ACA G A S E R Y F G A G T R L T V T
TRBJ3-1	GAATCTCTG	CTAGCCCCTTCC	CACTGTG	C TCC TAT GGG GAG CTG CAC TTC GGG CCA GGG ACC AGG CTC TCT GTG CTA G S Y G E L H F G P G T R L S V L
TRBJ3-2	AGTTTGCGC	CTGGGTCCCCAG	GGCTGTG	TG AAC ACC CAG CCC CTG TAC TTT GGA GCT GGT TCC AAG CTG ACT GTG CTG G N T Q P L Y F G A G S K L T V L
TRBJ3-3	AGTTTTTGT	CCTGAGCCTCCG	GGCTGTG	AGC AAC AAC CCT CTG TAT TTT GGA GGG GGC ACT CCG CTG TTG GTA CTA G S N N P L Y F G G G T R L L V L
TRBJ3-4	TATTTTTGT	GCTGAGCCCTGA	GGCTGTG	AGC ACA GAC ACG CAG TAC TTC GGC GCG GGC ACC CGG CTG ACG GTG CTA G S T D T Q Y F G A G T R L T V L
TRBJ3-5	GGTTTTTGC	GCGAGGCTGGGG	GGCCGTG	ACT CAG ATC CAG TAC TTC GGG CCG GGC ACG CGG CTC CTG GTG CTA G T Q I Q Y F G P G T R L L V L
TRBJ3-6	GATTTTTGT	GGGGCGCCCCGA	GGCTGTG	C TTT GCA GGC GCC GCC CTG ACC TTC GGG GCC GGC AGC TGG CTG ACG GTG GTG G F A G A A L T F G A G S W L T V V
TRBJ3-7	GGTTTGAGT	GCGGGGCTCTTC	TTCCGTG	T TCC TAT GAG CAG TAT TTC GGC CCA GGC ACC AAG CTC ACG GTC TTA G S Y E Q Y F G P G T K L T V L

**Figure 5**

**The genomic sequence of the (A) 3 TRBD genes and (B) 18 TRBJ genes.** The nucleotide and predicted amino acid sequences of (A) The TRBD genes. TRBD genes have the potential to be read in all 3 reading frames, and with the exception of the 3<sup>rd</sup> reading frame of TRBD1 encode at least 1 glycine residue. (B) The TRBJ genes. TRBJ1-3 is predicted to be non-functional due to loss of consensual RS heptamer sequence (bold and underlined). (†) In the genome TRBJ1-2 has a frameshift due a single base pair deletion in the TRBJ region and would therefore be predicted to be a pseudogene, but based on sequences correlating with this TRBJ gene derived from cDNA analyses we have introduced a thymidine (shown in parentheses).

**A**



**B**

	Exon 1 390bp	Intron 1	Exon 2 18bp	Intron 2	Exon 3 109bp	Intron 3	Exon 4 21bp	3'UTR
TRBC1 vs TRBC2	99.7%	-	-	-	99.1%	96%	100%	98.5%
TRBC1 vs TRBC3	97.2%	96%	100%	79.2%	97.2%	<40%	80.9%	<50%
TRBC2 vs TRBC3	96.9%	-	-	-	96.3%	<40%	80.9%	<50%

**Figure 6**

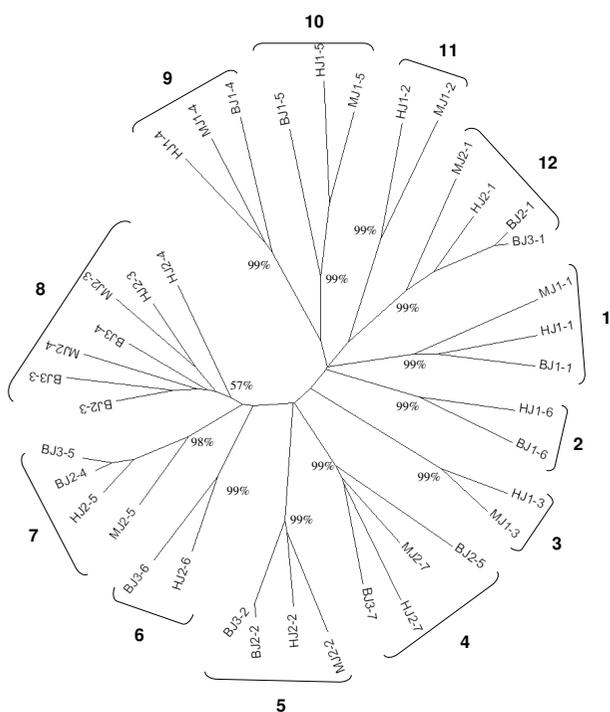
**The bovine TRBC genes.** (A) Predicted amino acid sequences of products of the TRBC1, 2 and 3 genes. The regions encoded by the 4 exons and the boundaries between the predicted extra-cellular (EX), hinge (H), trans-membrane (TM) and cytoplasmic (CY) domains have been marked. The sequence for a 7 amino acid section of TRBC2 can not be predicted due to absence of nucleotide sequence for exon 2 (represented by dots). (B) Pairwise percentage identity of nucleotide sequences between the exons, introns and 3'UTR of the 3 TRBC genes. Some pairwise comparisons have been omitted due to a region of undetermined sequence spanning the 1<sup>st</sup> intron, 2<sup>nd</sup> exon and 2<sup>nd</sup> intron of TRBC2.

3) and 11 (TRBJ1-2) in Btau\_3.1 and groups 2 (TRBJ1-6) and 6 (TRBJ2-6) in mice.

**Comparison with cDNA data identifies additional TRBV gene sequences missing from the genome assembly**

Using a variety of RT-PCR based methods, our group has isolated and sequenced over 1000 partial TRB chain cDNAs [31-33,40]. With a few exceptions, these cDNA sequences incorporated > 230 bp of the TRBV gene (i.e. over 80% of the sequence encoding the mature peptide)

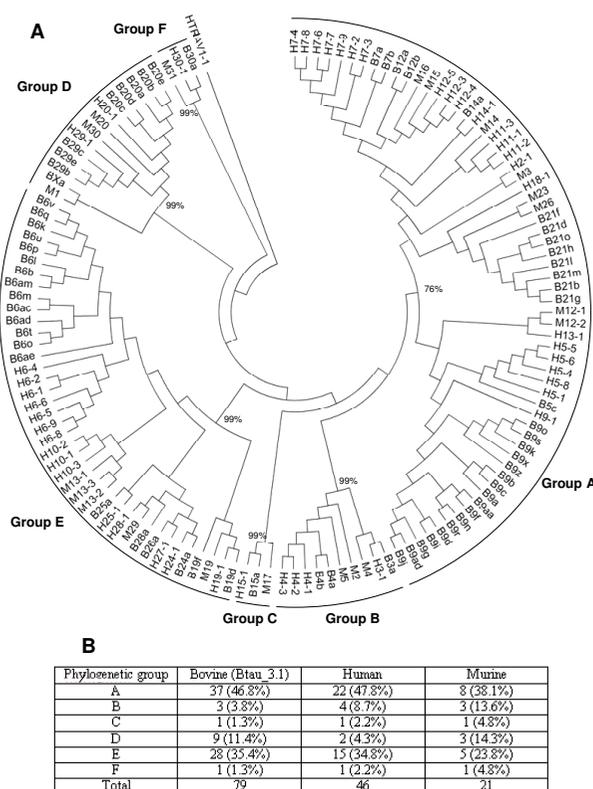
and in some cases the full length of the TRBV gene. Based on the assumption that sequences sharing ≤ 97% nucleotide identity represent distinct genes, as applied in studies of human and murine TRBV genes [41,42], our analysis identified 86 putative unique TRBV genes belonging to 22 subgroups (Table 1). Analysis of the sequence data available for each cDNA sequence indicated that only one of these genes is predicted to be non-functional (TRBV6-6 – due to a loss of a conserved cysteine encoding codon at position 104 according to the IMGT numbering



**Figure 7**  
**Neighbor-joining phylogenetic tree of the functional genomic repertoire of murine, human and bovine TRBJ genes.** Analysis was completed on the coding and RS nucleotide sequence of functional TRBJ genes following complete deletion to remove gaps in the alignment. The final dataset included 59 positions. The 12 phylogenetic groups (1–12) have been indicated and the percentage bootstrap interior branch test value ( $P_B$ ) based on a 1000 replications is shown for each of the groups. Generally each group is composed of genes from the 3 species that share a conserved order in the genome; group 8 is unique in containing the orthologues of two adjacent genes human and murine TRBJ2-3 and TRBJ2-4 (and in the bovine TRBJ3-3 and TRBJ3-4 as well as TRBJ2-3).

system [43]), consistent with evidence that mRNA expression of non-functional TRB chains is down-regulated and therefore limited [10,44,45].

All of the TRBV subgroups identified in Btau\_3.1 were also identified in the cDNA sequences. In addition a single member of subgroup TRBV27, which is not represented in Btau\_3.1, was identified. Although the repertoire of cDNA and functional genomic TRBV genes is broadly similar in both size and distribution across the subgroups (Table 1), detailed comparison shows that for the large subgroups there is substantial disparity between the genes present in the assembled genome and the cDNA repertoire. Thus, only 23 (35.9%) of the 64 TRBV genes in subgroups TRBV6, 9, 19, 20, 21 and 29 identified from cDNA analysis had genomic sequences showing 100% sequence iden-



**Figure 8**  
**Neighbour-joining phylogenetic tree of the functional genomic repertoire of murine, human and bovine TRBV genes.** (A) Analysis of the coding region nucleotide sequences of functional TRBV genes following complete deletion to remove gaps in the alignment. The final dataset included 281 positions. The sequence of HTRAVI-1 [93] has been included as an outgroup. The six phylogenetic groups (A-F) have been indicated and the percentage bootstrap interior branch test value ( $P_B$ ) based on a 1000 replications is shown for each of the groups. To reduce the size of the tree for presentation, where 2 TRBV genes have identical sequence only 1 has been included in the analysis. Therefore, bovine 6x (identical to 6t), 6an (6p), 6u (6y), 6ag (6l), 6ak (6h), 6v (6z), 9ae (9f), 9ak (9b), 9ac(9k), 9s (9v), 9aj (9d), 19d (19e) and 21m (21p) have been excluded, as has human TRBV6-3 which is identical to TRBV6-2. H = human, M = murine, B = bovine. (B) Summary of the distribution of functional TRBV genes amongst the 6 phylogenetic groups in humans, mice and Btau\_3.1.

ty, whilst 26 (40.6%) exhibited  $\leq 97\%$  identity to any genomic sequence. The remaining 15 (23.4%) sequences displayed nucleotide identities of 98–99% with genomic gene sequences. Given the presence in the genome of TRBV genes exhibiting  $> 97\%$  nucleotide identity, it is not possible to conclude whether these cDNAs represent allelic variants of already identified genes or products of additional genes absent from the current assembly. That

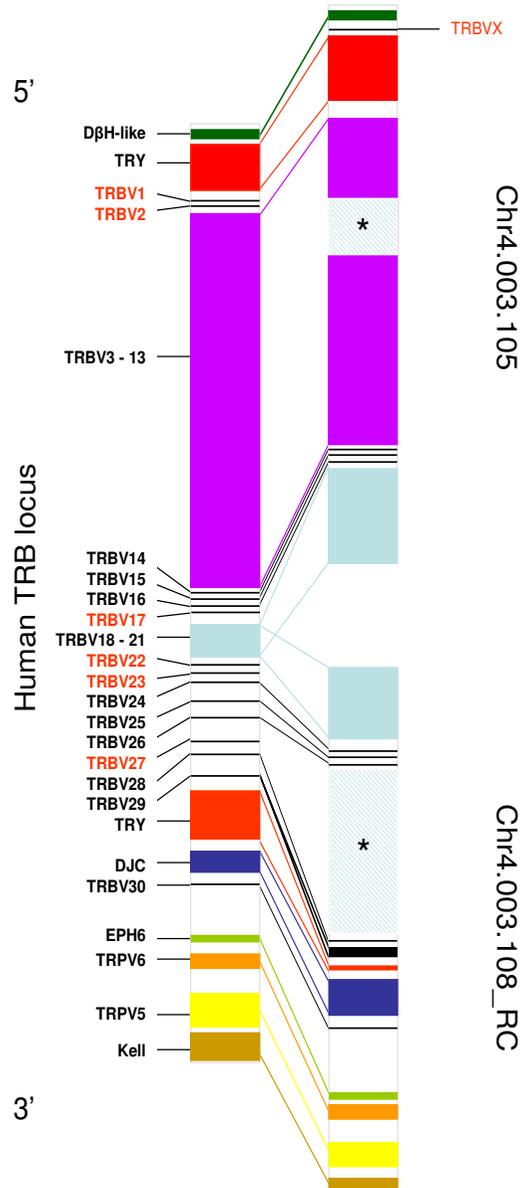
at least some of the cDNAs fall into the latter category, is supported by the identification of sequences exhibiting 100% identity with 4 of these cDNA sequences, in the genome project's WGS trace archive (data not shown). Conversely, 40 (63.5%) of the 63 predicted functional genes identified in these subgroups within the genome did not have cDNA sequences displaying 100% nucleotide identity. Twenty-two of these (34.9%) showed 98–99% identity with cDNA sequences, whilst the remaining 18 (28.6%) exhibited < 97% identity to any of the cDNA sequences. In contrast to the findings with multi-member subgroups, cDNAs corresponding to 9 subgroups with single members identified in the genome showed 100% identity with their respective genome sequence. Thus, comparison with cDNA evidence suggests that substantial numbers of genes, predominantly from the large subgroups, are still absent from Btau\_3.1.

In contrast to the TRBV situation, all TRBD and TRBC genes and the 16 functional TRBJ genes identified in Btau\_3.1 have been found expressed in cDNA. In addition, a functional allele of the TRBJ1-2 gene, which compared to the genomic sequence has a 1 bp insertion that restores the ORF encoding the FGXG motif (Figure 5), has been identified. No evidence for further TRBD, TRBJ or TRBC genes was found from cDNA analysis, suggesting the repertoire of these genes in Btau\_3.1 is complete.

**Conserved synteny between the human TRB locus and scaffolds Chr4.003.105 and Chr4.003.108**

The organisation of genes within Chr4.003.105 and Chr4.003.108 shows marked conservation in order with that of genes at the 5' and 3'ends of the human TRB locus respectively (Figure 9). Genes belonging to orthologous TRBV subgroups show a similar order although in some areas substantial tandem duplication has obscured synteny at the level of individual genes (e.g. the TRBV3-13 regions in the human TRB locus and on Chr4.003.105). TRBVX, the only bovine TRBV gene that has no human orthologue, is located in a position (between the dopamine-β-hydroxylase-like (DβH-like) gene and trypsinogen genes) syntenic with its murine orthologue (mTRBV1). As mentioned previously, synteny is also shown in the organisation of TRBJ genes, with human and bovine orthologues occupying conserved positions in their relative clusters.

Synteny also extends to non-TRB genes located within and adjacent to the human TRB locus. The 5 trypsinogen genes located on Chr4.003.105 and Chr4.003.108 are syntenic to those located towards the 5'end and 3'end of the human TRB locus respectively, and the DβH-like gene flanking the 5' end of the human TRB locus and the ephrin type-b receptor 6 precursor (EPH6), transient receptor potential cation channel subfamily V (TRPV), TRPV6 and Kell (Kell blood group glycoprotein).



**Figure 9**  
**Comparison of the genomic organisation of genes on Chr4.003.105 and Chr4.003.108 with the human TRB locus.** The relative position of genes or groups of genes in the human TRB locus and the orthologues on Chr4.003.105 and Chr4.003.108 are shown. Human TRBV genes without bovine orthologues are shown in red script, as is bovine TRBVX which lacks a human orthologue. The hatch areas marked with an asterisk in Chr4.003.105 and Chr4.003.108 indicate large areas of undetermined sequence. DβH-like (dopamine-β-hydroxylase like gene), TRY (trypsinogen genes), EPH-6 (ephrin type-b receptor 6 precursor), TRPV5 (transient receptor potential cation channel subfamily V member 5), TRPV6 and Kell (Kell blood group glycoprotein).

member 5, TRPV6 and kell blood group glycoprotein (kell) loci flanking the 3' end of the human TRB locus all have bovine orthologues in syntenic positions on the 2 scaffolds.

Although fluorescent *in situ* hybridisation studies have previously shown that the position of the TRB locus with respect to the blue cone pigment (BCP) and chloride channel protein 1 (CLCN1) genes are conserved between ruminants and humans [46], this analysis shows for the first time the high levels of synteny between human and bovine orthologues both within and adjacent to the TRB locus. Extrapolation of this synteny predicts that Chr4.003.105 and Chr4.003.108 (in reverse complement) should be juxtaposed on chromosome 4, whilst Chr4.003.106, which contains bovine orthologues to numerous genes that in humans are telomeric to the TRB locus (including CLCN1) should be located 3' to Chr4.003.108 and Chr4.003.107, which contains a bovine orthologue to the acylglycerol kinase (AGK) gene that in humans lies centromeric to the TRB locus, should be positioned 5' to Chr4.003.105. This location of Chr4.003.106 has also been predicted by clone paired-end analysis (data not shown).

#### **RS and regulatory elements sequences are conserved in the bovine TRB locus**

The RS sequences of the bovine TRBV, TRBJ and TRBD genes show a high degree of similarity to canonical RS sequences defined for the corresponding human and murine genes (Figure 10). In the bovine TRBV 23-RS sequences the CACAG of the heptamer and a poly-A stretch in the centre of the nonamer show a high degree of intra- and inter-species conservation. Although conservation of the spacer is less marked, the CTGCA sequence proximal to the heptamer is reasonably well conserved and similar to that of humans. Despite more limited conservation, the 8 bp proximal to the nonamer also displays a degree of cross-species similarity. Similarly, the bovine TRBJ RS exhibits intra- and inter-species conservation of the first 3 bp (CAC) of the heptamer sequence and a poly-A stretch in the nonamer. Conservation in the spacer is limited but overrepresentation of G at the position 6 bp from the heptamer and C 4 bp from the nonamer is seen in both the bovine and human.

We identified a 187 bp sequence ~8.7 Kb 3' to the TRBC3 gene that displays high nucleotide similarity with the sequences of the enhancers (E $\beta$ ) identified in the murine (76.2%) and human (78.3%) TRB loci [47-49]. Sequences of the protein binding sites described in the E $\beta$ s of humans (T $\beta$ 2-4) and mice ( $\beta$ E1-6) are well conserved in the aligned bovine sequence (Figure 11a); several of the transcription binding sites shown to be functionally important in the regulation in E $\beta$  function [47-50], such

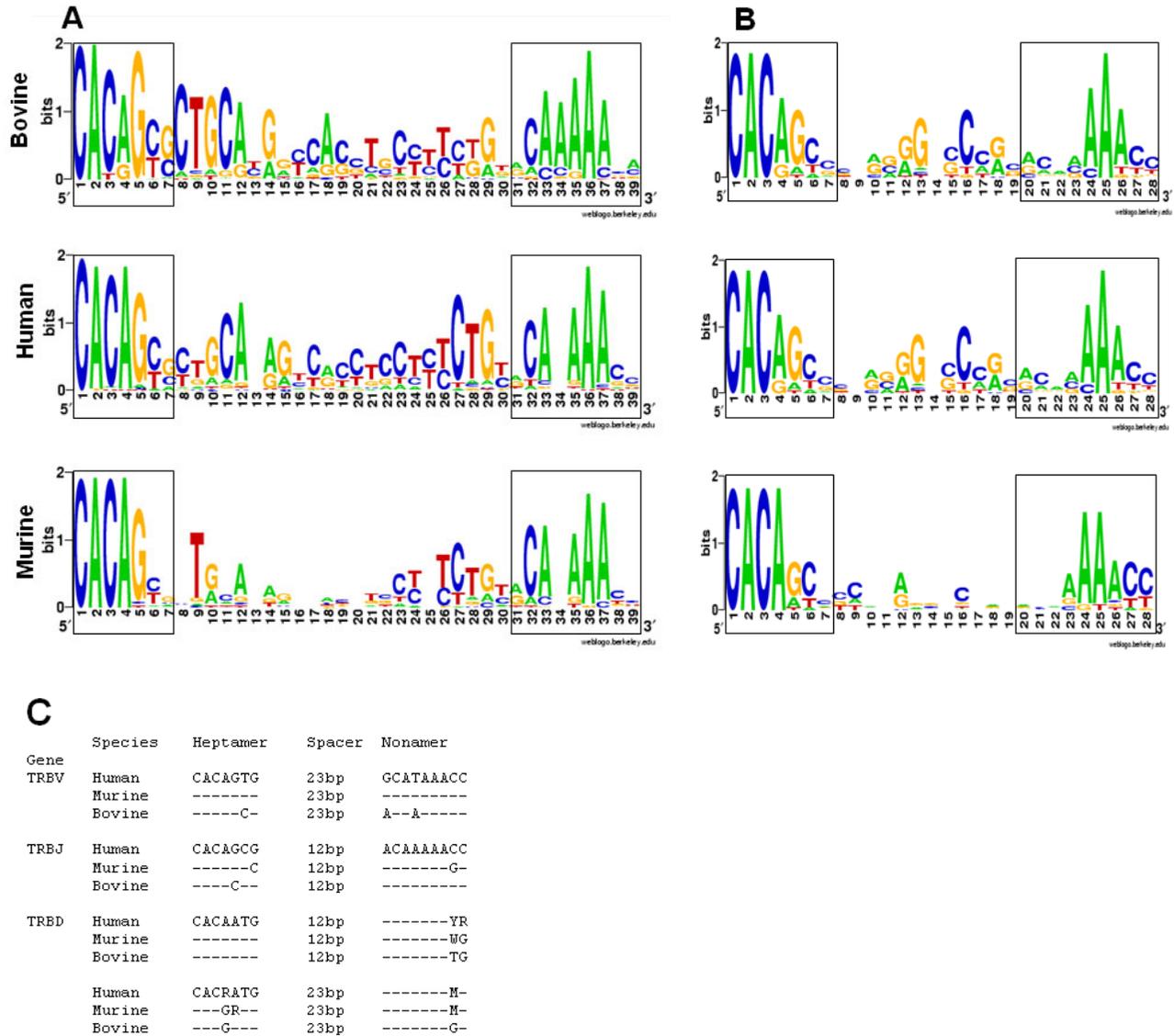
as the GATA-binding site in  $\beta$ E1/T $\beta$ 2 and the  $\kappa$ E2-binding motif in  $\beta$ E3 are absolutely conserved, whilst others (such as the CRE in  $\beta$ E2/T $\beta$ 2) show minimal sequence divergence. In contrast, the sequence of the TRBD1 promoter (PD $\beta$ 1), which includes the ~300 bp directly upstream of the TRBD1 gene and has been well defined in the mouse [51,52], displays a more limited nucleotide identity (59.2%) with the bovine sequence. As shown in Figure 11b, some transcription factor binding sites demonstrated to be important for PD $\beta$ 1 function (SP-1 and GATA) in mice and/or humans are absent from the bovine sequence, whereas others (TATA box, AP-1 and Ikaros/Lyf-1) have been well conserved [51-53].

We were also able to identify a conserved cAMP responsive element (CRE) motif (AGTG $\Delta$ xxTGA) in the ~80-120 bp upstream sequence of 57 (42.6%) of the bovine TRBV genes (Figure 11c). This motif is found within conserved decamer sequences in the promoter regions of some murine and human TRBV genes [54] and has been shown to specifically bind a splice variant of a CRE binding protein preferentially expressed in the thymus [55]. In general, the CRE motif has been found in bovine genes that are members of subgroups that are orthologous to the human TRBV subgroups in which the CRE motif is also found [10].

#### **Discussion**

Sequencing of the human and murine TRB loci has defined the repertoire of TRB genes in these species as well as provided insights into the organisation, evolution and regulation of this immunologically important locus [9,10]. Although the bovine TRB locus sequence in the third bovine genome assembly is incomplete, the analysis conducted in the present study has provided insight into the nature of the bovine TRB gene repertoire and its genomic organisation and evolution.

The most striking result from the study was the large number of TRBV genes identified (134) which is over twice the number found in humans and four times that in mice [11,12]. Although 11 of the 24 bovine subgroups identified in the genome contain multiple genes, the majority of the TRBV repertoire expansion is attributable to the extensive membership of just 3 subgroups, TRBV6 (40 members), 9 (35 members) and 21 (16 members). By comparison, the largest subgroups in humans are TRBV6 and TRBV7, with 9 members each, whilst in mice the only multi-membered subgroups are TRBV12 and 13 with 3 members each. As in humans the expansion of the TRBV repertoire has predominantly occurred through the tandem duplication of DNA blocks containing genes from more than 1 subgroup [9,10]. Dot-plot analyses shows that this duplication in the bovine is complex, leading to the generation of 6 homology units ranging in size from

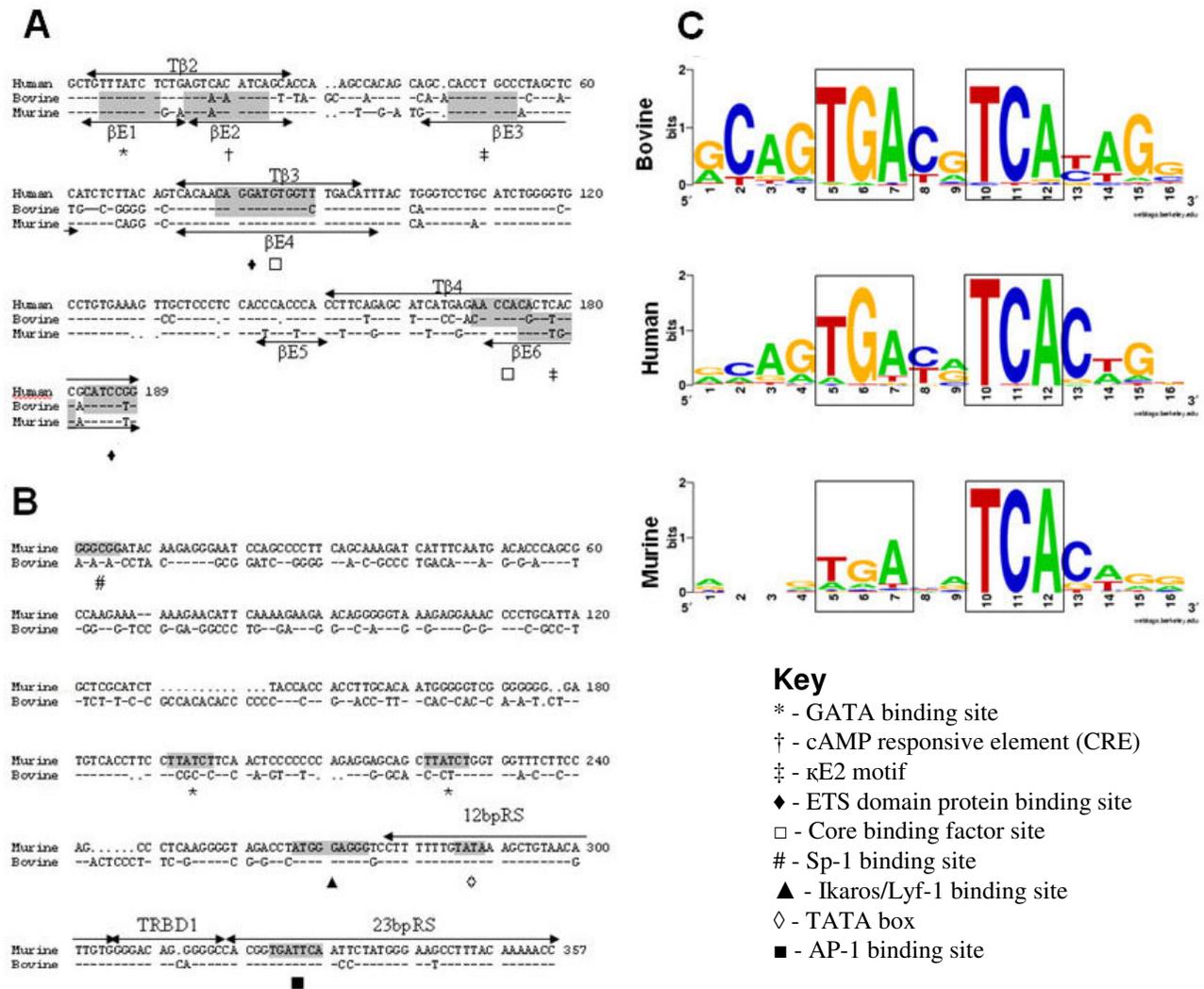


**Figure 10**  
**Comparison of recombination signal sequences of human, murine and bovine TRB genes.** Sequence logos for the RSs of (A) TRBV and (B) TRBJ genes of bovine, humans and mice. The heptamer and nonamer sequences are enclosed in boxes. The height of the letters is correlated to their conservation at each location. (C) Consensus sequences of RSs from the different groups of bovine, human and murine TRB genes. Dashes represent nucleotide identity. M = A/C, R = A/G, W = A/T, and Y = C/T. Non-functional RSs functional have been excluded from the analysis.

~7 Kb to ~31 Kb and encompassing between 1 and 11 TRBV genes. Unequal cross-over (non-homologous meiotic recombination) between genome-wide repeats (e.g. SINEs, LINEs and LTRs) has been proposed to act as the substrate for such duplication events in TR loci [9]. Although genome-wide repeats are found in the DNA surrounding the bovine TRBV genes (Additional file 3), as in the human TRB locus they are only rarely found at the

boundaries of duplicated homology units (data not shown), suggesting their contribution to mediating duplication is minimal [10].

Although gene conversion of TRBV genes has been documented [56], as with other multi-gene families in the immune system, TRBV genes predominantly follow a 'birth-and-death' model of evolution [13,57,58], by



**Figure 1**  
**Sequence comparison of regulatory elements in the bovine, human and murine TRB loci.** (A) Alignment of the human, bovine and murine minimal enhancer sequences. Protein binding sequences in the human (Tβ) and murine (βE) minimal Eβ region are indicated by arrows. (B) Alignment of the murine PDβ1 with the bovine sequence immediately upstream of the TRBD1 gene. The position of the TRBD1 gene and the flanking RSs are indicated. Shaded areas indicate the location of defined transcription regulation elements as defined in the key. Identity is shown by dashes and gaps by dots. (C) Sequence logos for the conserved TGAXxTCA CRE motif and adjacent bases found in the promoter regions (at ~80–120 bp upstream) of 57 bovine, 52 human and 23 murine TRBV genes. The location of conserved TGA and TCA are enclosed in boxes. The height of the letters is correlated to their conservation at each location.

which new genes are created by repeated gene duplication, some of which are maintained in the genome whilst others are deleted or become non-functional due to mutation. Genes maintained following duplication are subject to progressive divergence, providing the opportunity for diversification of the gene repertoire. Gene duplication within the TR loci has occurred sporadically over hundreds of millions of years with ancient duplications accounting for the generation of different subgroups and

more recent duplications giving rise to the different members within subgroups [9,59]. The continuous nature of duplication and divergence of bovine TRBV genes is evident in the multi-membered subgroups where nucleotide identity between members ranges between 75.5% and 100%. The complete identity observed between some TRBV genes suggests that some of the duplication events have occurred very recently. Similar features have been described for the murine TRA and human IGκ loci, within

which recent duplications, < 8 million years ago (MYA), have created pairs of V genes exhibiting ~97% nucleotide identity [9,60,61]. Southern blot data showing differences in RFLP banding patterns of TRBV9 and 27 genes in DNA from *Bos indicus* and *Bos taurus* animals (Figure 1c), which only diverged between 0.25 – 2 million years ago [62-64], provides further evidence of recent evolutionary development of the TRBV repertoire in cattle.

The distribution of TRBV genes over 5 scaffolds and the presence of > 180 Kb of undetermined sequence within two of the scaffolds indicate that characterisation of the genomic TRBV repertoire remains incomplete. Comparison with cDNA sequences data indicates that the number of undefined genes is substantial – only 36/86 (42%) of TRBV genes identified from cDNA analysis have corresponding identical sequences in Btau\_3.1. Most of the identified TRBV genes missing from the assembly are members of the large subgroups TRBV6, 9, 19, 20, 21 and 29, further enhancing their numerical dominance. Although it is anticipated that completion of the TRB locus sequence will incorporate significant numbers of additional TRBV genes, the possible existence of insertion-deletion related polymorphisms (IDRPs), which can lead to intra-species variation in genomic TRBV gene repertoires as described in human and murine TRB loci [65-68], may result in some of the genes identified in cDNA being genuinely absent from the sequenced bovine genome

The proportion of TRBV pseudogenes in Btau\_3.1 is 41%, comparable to that seen in both humans (29%) and mice (40%), suggesting that the 'death rate' in TRBV gene evolution is generally high [58]. Pseudogene formation has occurred sporadically throughout the evolution of TRBV genes, with genes that have lost function tending to subsequently accumulate further lesions [9]. The majority of bovine TRBV pseudogenes (57%) contain a single lesion and thus appear to have arisen recently; the remaining 43% have multiple lesions of varying severity and complexity (Additional file 2). In addition to pseudogenes we also identified 7 sequences showing limited local similarity to TRBV genes in Btau\_3.1 (Figure 2 – open boxes). Such severely mutated TRBV 'relics', 22 of which have been identified in the human TRB locus [10], are considered to represent the remnants of ancient pseudogene formation.

In contradiction to a previous report [39], the repertoire of functional TRBV genes in Btau\_3.1 exhibits a level of phylogenetic diversity similar to that of humans and mice. Phylogenetic groups A and E are over-represented in all 3 species, which in humans and cattle is largely attributable to expansion of subgroups TRBV5, 6, 7 and 10 and TRBV6, 9 and 21 respectively; in mice the expansion of subgroups

TRBV12 and 13 make a more modest contribution to this over-representation. Much of the expansion of human subgroups TRBV5, 6 and 7 occurred 24–32 MYA [13] and similarly, as described above, in bovines much of the expansion of subgroups TRBV6, 9 and 21 subgroups appears to be very recent. As these expansions have occurred subsequent to primate/artiodactyl divergence (~100MYA) [69], over-representation of phylogenetic groups A and E must have occurred as parallel but independent events in these lineages, raising interesting questions about the evolutionary pressures that shape the functional TRBV repertoire.

In contrast to the wide variation in the organisation of TRBD, TRBJ and TRBC genes in the TRB locus seen in non-mammalian vertebrates [70-74], in mammals the arrangement of tandemly located DJC clusters is well conserved [10,35,36,75,76]. Although most placental species studied have 2, variation in the number of DJC clusters has been observed, with unequal cross-over events between TRBC genes usually invoked as the most likely explanation for this variation [36,77,78]. The results from this study provide the first description of the entire bovine DJC region and confirm that like sheep, cattle have 3 complete DJC clusters [36,37]. Dot-plot and sequence analyses indicate that unequal crossover between the ancestral TRBC1 and TRBC3 genes led to duplication of a region incorporating TRBC1, TRBD3 and TRBJ3 genes, generating the DJC2 cluster. The similarity with the structure of the ovine DJC region suggests that this duplication event occurred prior to ovine/bovine divergence 35.7 MYA [69]. As with duplication of TRBV genes, expansion of TRBD and TRBJ gene numbers has increased the number of genes available to partake in somatic recombination – the 3168 different VDJ permutations possible from the functional genes present in Btau\_3.1 is considerably more than that for either humans ( $42 \times 2 \times 13 = 1092$ ) or mice ( $21 \times 2 \times 11 = 462$ ). Interestingly, the sequence of bovine TRBD1 gene is the first TRBD gene described that doesn't encode a glycine residue (considered integral to the structure of the CDR3 $\beta$ ) in all 3 reading frames [79]. However, analysis of cDNA reveals evidence of expression by functional TRB chains of TRBD1 in the reading frame that doesn't encode a glycine but have generated a glycine codon by nucleotide editing at the VJ junction (data not shown).

In contrast to TRBV, TRBD and TRBJ genes which encode products that bind to a diverse array of peptide-MHC ligands, TRBC gene products interact with components of the CD3 complex which are non-polymorphic. Consequently, due to structural restrictions TRBC genes are subject to concerted evolutionary pressures with intra-species homogenization through gene conversion evident in both humans and mice [9,80]. Similarly, the bovine TRBC

genes were found to encode near identical products, most likely as a result of gene conversion, although in the case of TRBC1 and TRBC2 genes this more probably reflects minimal divergence following duplication.

Comparison with the human and murine sequences shows that non-coding elements that regulate TRB expression, such as the E $\beta$ , promoters and RSs are highly conserved in the bovine. This is consistent with work demonstrating that the critical role of RSs has enforced a high level of evolutionary conservation [70,73,74,81] and that E $\beta$  and PD $\beta$ 1 sequences are well conserved in eutherian species [36,52]. Although transcriptional factor binding sites are less well conserved in the putative PD $\beta$ 1 than the E $\beta$  sequence, the Ikaros/Lyf-1 and Ap-1 binding sites of the PD $\beta$ 1, which are vital in enforcing stage-specific (i.e. D $\beta$ -J $\beta$  prior to V $\beta$ -D $\beta$ J $\beta$  recombination) are conserved [53,82]. Our analysis of putative TRBV promoter elements was restricted to the well described CRE motif [9,10,54]. However, TRBV promoters are complex and expression of TRBV genes whose promoters lack the CRE motif is maintained through the function of other transcriptional factor binding sites [83]. A more detailed analysis of the bovine TRBV promoters would be interesting given the potential influence this may have on shaping the expressed TRBV repertoire [25], but is beyond the scope of the current study.

The portion of the bovine TRB locus described in Btau\_3.1 encompasses > 730 Kb of sequence (excluding the regions of undetermined sequence in Chr4.003.105 and Chr4.003.108). Thus, although incomplete, the bovine TRB locus is larger than that of either humans (620 Kb) or mice (700 Kb), mainly as a consequence of the duplications leading to the dramatic expansion of the V genes. In contrast to V genes, duplication of trypsinogen genes within the TRB locus is more limited in the bovine (Figure 2), where only 5 trypsinogen genes were identified, compared to the human and murine where more extensive duplication has led to the presence of 8 and 20 trypsinogen genes respectively. Despite the differences in duplication events, the organisation of both TR and non-TR genes within and adjacent to the TRB locus exhibits a striking conserved synteny between cattle, humans and mice [9,84]. Indeed, the organisation of genes within the TRB locus and its position relative to adjacent loci is ancient, with marked conserved synteny also demonstrated between eutherian and marsupial mammalian species and, to a large extent, chickens [9,75]. Given the evidence for conserved synteny of TRBV gene organisation despite dissimilar duplication/deletion events between mice, humans and cattle, the results of the analysis completed in this study suggest that several subgroups including TRBV1, 2, 17, 22 and 23, which were not identified in the

genome assembly or from cDNA sequences, may have been deleted from the bovine genome (Figure 9). Conservation of synteny would predict that the genomic location of the TRBV27 gene identified from cDNA analysis will be within the region of undetermined sequence in Chr4.003.108 between the TRBV26 and 28 genes (Figure 9).

## Conclusion

The primary purpose of this study was to analyse the sequence data made available from the third bovine genome assembly to gain a better understanding of the bovine TRB gene repertoire and the organization and evolution of the bovine TRB locus. The results of this analysis have shown that: (1) the bovine TRBV genomic repertoire has been dramatically expanded through a complex series of duplication events and although incomplete, is the largest described to date. These duplication events have led to massive expansion in the membership of certain TRBV subgroups, particularly TRBV6, 9 and 21; (2) duplication has generated 3 DJC clusters compared to 2 in humans and mice; (3) the elements that regulate TRB expression and the organisation of genes within and adjacent to the TRB loci exhibit high levels of conservation between humans, mice and cattle. (4) cDNA evidence indicates that a substantial number of TRBV genes, predominantly from the larger subgroups are absent from the current assembly.

Notwithstanding the incomplete assembly of the TRB locus, the results of these analyses clearly demonstrate that cattle possess a phylogenetically diverse repertoire of functional TRB genes that is substantially larger than that described for other species. These findings, together with emerging evidence of similar expansions of gene repertoires for other TR chains in ruminants [85,86] suggest that strong evolutionary pressures have driven a generic enlargement of TR gene numbers, and thus greater potential TR diversity, in the ruminant lineage. Further studies are required to define the full extent of these expansions and to understand their evolutionary basis.

## Methods

### Analysis of the genome

The third bovine genome assembly (Btau\_3.1) was retrieved directly from the sequencing centre involved in the Bovine Genome Project [87]. Sequences of bovine TRB genes identified from cDNA analysis [31-34,88] and human and murine TRB genes derived from GenBank sequences [GenBank:U66059-U66061] and [GenBank:AE000663-AE000665] respectively were compared to Btau\_3.1 using the BLASTN algorithm tool on the Ensembl website [89]. The locations of the TRB genes identified in Btau\_3.1 are provided in Additional file 1.

### Sequence analysis

Basic sequence analysis such as CLUSTALW alignments [90] and translations were conducted using the DNAsis Max v2.7 programme (MiraiBio, Alameda, CA, USA). Comparison of human and genomic bovine TRBV sequences was completed using the IMGT/V-QUEST programme [91] available through the IMGT homepage, <http://imgt.cines.fr> [92,93]. Dot-plot analyses were completed using dotter programme [94] and dottup programme from EMBOSS [95]. Genome wide repeats were analysed using the RepeatMasker programme [96]. Sequence logos in Figures 10 and 11 were generated using the Weblogo programme [97].

To be considered functional TRBV gene segment sequences were required to maintain i) splice sites appropriate for RNA editing, ii) open reading frames, which include codons for the conserved cysteine, tryptophan and cysteine residues at positions 23, 41 and 104 (IMGT unique numbering system [43]) respectively and iii) a 23-RS compatible with somatic recombination [98,99].

### Nomenclature

As the sequence of the TRB locus was incomplete, it was not possible to fully implement the IMGT nomenclature system which requires knowledge of the genomic order of genes from the 5' to 3' end of the locus [100]. Genomic bovine TRBV gene subgroups have been named according to the orthologous subgroups in humans and members of subgroups have been given an alphabetic rather than numeric description to avoid confusion with previously published cDNA data [32]. The DJC region of the locus appears complete and so the TRBD, TRBJ and TRBC have been named according to their 5' to 3' order in the genome.

### Phylogenetic analysis

Phylogenetic analysis was performed on the nucleotide sequences of functional TRBV genes (coding sequences) and TRBJ genes (coding sequence + RS) of humans, mice and bovine as identified in Btau\_3.1. Neighbour-joining method [101] analysis was performed with the MEGA4 software [102,103], using the uncorrected nucleotide differences (p-distance), which is known to give better results when a large number of sequences which contain a relatively small number of nucleotides are examined [104]. Maximum likelihood analysis was performed under a variety of substitution models (Jukes-Cantor, Kimura 2-parameter, Felsenstein 81, Felsenstein 84, Tamura-Nei 93 and General Time Reversible) as implemented by the PHYML programme [105,106], using the phylogenetic tree produced by NJ analysis as the primary tree. In each case the reliability of the resulting trees was estimated by

the approximate Likelihood Ratio Test (aLRT) method [107].

### Southern blot and cDNA analysis

Southern blots were performed as described in Houston *et al.* [32]. Analysis of bovine TRB cDNA expression included use of the methods described in previous studies [31-33,88]. cDNA sequences for bovine TRBV genes were derived from sequences submitted to public databases: [D90130, [AJ006569](#), [AJ006570](#), [AJ006572](#), [AJ006573](#), [AJ006574](#), [AJ006575](#), [AJ006576](#), [D90121](#), [D90123](#), [AJ006583](#), [D90122](#), [D90127](#), [D90124](#), [D90133](#), [L18951](#), [AJ006580](#), [D90131](#), [AJ006579](#), [AJ235264](#), [AJ235265](#), [AJ006578](#), [AJ235266](#), [AJ235267](#), [AJ006577](#), [D90128](#), [D90129](#), [AJ235268](#), [D90125](#), [AJ006347](#), [AJ006346](#) and [D90132](#)] and additional un-submitted data that are available on request from the corresponding author.

### Abbreviations

IG: immunoglobulin; IMGT: IMGT<sup>®</sup>, the international ImMunoGeneTics information system<sup>®</sup>; RFLP: restriction fragment length polymorphism; TRB: T cell receptor beta chain; TR: T cell receptor; TRBV:  $\beta$  variable gene; TRBJ:  $\beta$  joining gene; TRBD:  $\beta$  diversity gene; TRBC:  $\beta$  constant gene.

### Authors' contributions

TC, JA, AL and WM conceived the study. TC performed the genome analysis, the cDNA and southern blot work, and with JA completed the sequence and phylogenetic analysis. All authors read and approved the manuscript.

### Additional material

#### Additional file 1

*Table S1 – Location of bovine TRB genes in Btau\_3.1. For each gene the start and stop positions on the scaffold as well as the orientation are shown.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-192-S1.pdf>]

#### Additional file 2

*Table S2 – Summary of the analysis of non-functional TRBV gene segments identified in Btau\_3.1. To be considered functional TRBV gene segment sequences were required to fulfil several conditions as described in the Methods. The table above summarises the lesions identified in the genomic sequences of TRBV genes that are considered to render them non-functional and therefore pseudogenes. TRBV6c, 6g, 6r, 6w and 6aa are open-reading frame (ORF) pseudogenes as their reading frame is maintained, the lesions causing their predicted loss of function resulting either in the loss of a conserved residue or loss of a consensus 23-RS.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-192-S2.doc>]

### Additional file 3

Figure S1 – Analysis of the repeat content in (A) Chr4.003.105, (B) Chr4.003.108\_RC and (C) ChrUn.003.1717. The repeat content is calculated as the percentage (y-axis) of the sequence (x-axis) composed of repeat elements in 10 Kb windows using a rolling 1 Kb step. The positions of TRB genes and interposed/adjacent non-TRB genes, as well as regions of undetermined sequence are shown according to the legend. Abbreviations for gene names used in the figure have been described in Figure 9. Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-192-S3.pdf>]

### Acknowledgements

We would like to thank Dr. Zen Lu from the Roslin Institute for help with dot-plot data. We would also like to acknowledge funding from the Veterinary Training and Research Initiative award from the Department of Food and Rural Affairs (DEFRA) and the Scottish Higher Education Funding Council (SHEFC). Some of the work on cDNA analysis was also supported by Wellcome Trust Grant No. 075820.

### References

- Nikolich-Zugich J, Slifka MK, Messaoudi I: **The many important facets of T-cell repertoire diversity.** *Nat Rev Immunol* 2004, **4(2)**:123-132.
- Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P: **A direct estimate of the human alpha beta T cell receptor diversity.** *Science* 1999, **286(5441)**:958-961.
- Casrouge A, Beaudoin E, Dalle S, Pannetier C, Kanellopoulos J, Kourilsky P: **Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes.** *J Immunol* 2000, **164(11)**:5782-5787.
- Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC: **Structure of the complex between human T-cell receptor, viral peptide and HLA-A2.** *Nature* 1996, **384(6605)**:134-141.
- Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, Teyton L, Wilson IA: **An alpha beta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex.** *Science* 1996, **274(5285)**:209-219.
- Barker PE, Ruddle FH, Royer HD, Acuto O, Reinherz EL: **Chromosomal location of human T-cell receptor gene Tc beta.** *Science* 1984, **226(4672)**:348-349.
- Caccia N, Kronenberg M, Saxe D, Haars R, Bruns GA, Goverman J, Malissen M, Willard H, Yoshikai Y, Simon M, Hood L, Mak TW: **The T cell receptor beta chain genes are located on chromosome 6 in mice and chromosome 7 in humans.** *Cell* 1984, **37(3)**:1091-1099.
- Isobe M, Erikson J, Emanuel BS, Nowell PC, Croce CM: **Location of gene for beta subunit of human T-cell receptor at band 7q35, a region prone to rearrangements in T cells.** *Science* 1985, **228(4699)**:580-582.
- Glusman G, Rowen L, Lee I, Boysen C, Roach JC, Smit AF, Wang K, Koop BF, Hood L: **Comparative genomics of the human and mouse T cell receptor loci.** *Immunity* 2001, **15(3)**:337-349.
- Rowen L, Koop BF, Hood L: **The complete 685-kilobase DNA sequence of the human beta T cell receptor locus.** *Science* 1996, **272(5269)**:1755-1762.
- Bosc N, Lefranc MP: **The mouse (*Mus musculus*) T cell receptor beta variable (TRBV), diversity (TRBD) and joining (TRBJ) genes.** *Exp Clin Immunogenet* 2000, **17(4)**:216-228.
- Folch G, Lefranc MP: **The human T cell receptor beta variable (TRBV) genes.** *Exp Clin Immunogenet* 2000, **17(1)**:42-54.
- Su C, Nei M: **Evolutionary dynamics of the T-cell receptor VB gene family as inferred from the human and mouse genomic sequences.** *Mol Biol Evol* 2001, **18(4)**:503-513.
- McBlane JF, van Gent DC, Ramsden DA, Romeo C, Cuomo CA, Gellert M, Oettinger MA: **Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps.** *Cell* 1995, **83(3)**:387-395.
- Bassing CH, Alt FW, Hughes MM, D'Auteuil M, Wehrly TD, Woodman BB, Gartner F, White JM, Davidson L, Sleckman BP: **Recombination signal sequences restrict chromosomal V(D)J recombination beyond the 12/23 rule.** *Nature* 2000, **405(6786)**:583-586.
- Sleckman BP, Bassing CH, Hughes MM, Okada A, D'Auteuil M, Wehrly TD, Woodman BB, Davidson L, Chen J, Alt FW: **Mechanisms that direct ordered assembly of T cell receptor beta locus V, D, and J gene segments.** *Proc Natl Acad Sci USA* 2000, **97(14)**:7975-7980.
- Tonegawa S: **Somatic generation of antibody diversity.** *Nature* 1983, **302(5909)**:575-581.
- Krangel MS: **Gene segment selection in V(D)J recombination: accessibility and beyond.** *Nat Immunol* 2003, **4(7)**:624-630.
- Mostoslavsky R, Alt FW, Bassing CH: **Chromatin dynamics and locus accessibility in the immune system.** *Nat Immunol* 2003, **4(7)**:603-606.
- Schlissel MS: **Regulating antigen-receptor gene assembly.** *Nat Rev Immunol* 2003, **3(11)**:890-899.
- Bories JC, Demengeot J, Davidson L, Alt FW: **Gene-targeted deletion and replacement mutations of the T-cell receptor beta-chain enhancer: the role of enhancer elements in controlling V(D)J recombination accessibility.** *Proc Natl Acad Sci USA* 1996, **93(15)**:7871-7876.
- Bouvier G, Watrin F, Naspetti M, Verthuy C, Naquet P, Ferrier P: **Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development.** *Proc Natl Acad Sci USA* 1996, **93(15)**:7877-7881.
- Mathieu N, Hempel WM, Spicuglia S, Verthuy C, Ferrier P: **Chromatin remodeling by the T cell receptor (TCR)-beta gene enhancer during early T cell development: Implications for the control of TCR-beta locus recombination.** *J Exp Med* 2000, **192(5)**:625-636.
- Oestreich KJ, Cobb RM, Pierce S, Chen J, Ferrier P, Oltz EM: **Regulation of TCRbeta gene assembly by a promoter/enhancer holocomplex.** *Immunity* 2006, **24(4)**:381-391.
- Ryu CJ, Haines BB, Lee HR, Kang YH, Draganov DD, Lee M, Whitehurst CE, Hong HJ, Chen J: **The T-cell receptor beta variable gene promoter is required for efficient V beta rearrangement but not allelic exclusion.** *Mol Cell Biol* 2004, **24(16)**:7015-7023.
- Sikes ML, Suarez CC, Oltz EM: **Regulation of V(D)J recombination by transcriptional promoters.** *Mol Cell Biol* 1999, **19(4)**:2773-2781.
- Whitehurst CE, Chattopadhyay S, Chen J: **Control of V(D)J recombinational accessibility of the D beta 1 gene segment at the TCR beta locus by a germline promoter.** *Immunity* 1999, **10(3)**:313-322.
- Baron C, Sachs DH, LeGuern C: **A particular TCR beta variable region used by T cells infiltrating kidney transplants.** *J Immunol* 2001, **166(4)**:2589-2596.
- Butler JE, Wertz N, Sun J, Sacco RE: **Comparison of the expressed porcine Vbeta and Jbeta repertoire of thymocytes and peripheral T cells.** *Immunology* 2005, **114(2)**:184-193.
- Halsey WA Jr, Palmer BE, DeMartini JC, Howell MD: **Analysis of sheep T-cell receptor beta-chain heterogeneity.** *Immunogenetics* 1999, **49(3)**:206-214.
- Connelley T, MacHugh ND, Burrells A, Morrison WI: **Dissection of the clonal composition of bovine alphabeta T cell responses using T cell receptor Vbeta subfamily-specific PCR and heteroduplex analysis.** *J Immunol Methods* 2008, **335(1-2)**:28-40.
- Houston EF, Connelley T, Parsons K, MacHugh ND, Morrison WI: **Analysis of T-cell receptor BV gene sequences in cattle reveals extensive duplication within the BV9 and BV20 subgroups.** *Immunogenetics* 2005, **57(9)**:674-681.
- Houston EF, Morrison WI: **Identification of seven new TCRBV subfamilies in cattle (*Bos taurus*).** *Eur J Immunogenet* 1999, **26(5)**:349-353.
- Tanaka A, Ishiguro N, Shinagawa M: **Sequence and diversity of bovine T-cell receptor beta-chain genes.** *Immunogenetics* 1990, **32(4)**:263-271.
- Watanabe M, Iwasaki Y, Mita Y, Ota S, Yamada S, Shimizu M, Takagaki Y: **Porcine T-cell receptor beta-chain: a genomic sequence covering Dbeta1.1 to Cbeta2 gene segments and the diversity of cDNA expressed in piglets including novel alternative splicing products.** *Mol Immunol* 2007, **44(9)**:2332-2343.

36. Antonacci R, Di Tommaso S, Lanave C, Cribeu EP, Ciccarese S, Mas-sari S: **Organization, structure and evolution of 41 kb of genomic DNA spanning the D-J-C region of the sheep TRB locus.** *Mol Immunol* 2008, **45(2)**:493-509.
37. Conrad ML, Pettman R, Whitehead J, McKinnel L, Davis SK, Koop BF: **Genomic sequencing of the bovine T cell receptor beta locus.** *Vet Immunol Immunopathol* 2002, **87(3-4)**:439-441.
38. Antonacci R, Massari S, De Iaco R, Ciccarese S: **Assignment of the TRB@ locus encoding the T-cell receptor beta chain to sheep, cattle, goat and river buffalo chromosomes by in situ hybridization.** *Cytogenet Cell Genet* 2001, **94(1-2)**:82-83.
39. Su C, Jakobsen I, Gu X, Nei M: **Diversity and evolution of T-cell receptor variable region genes in mammals and birds.** *Immunogenetics* 1999, **50(5-6)**:301-308.
40. Connelley T, Burrells A, Machugh ND, Morrison WI: **Use of a Pan-Vbeta primer permits the amplification and sequencing of TCRbeta chains expressed by bovine T-cell clones following a single semi-nested PCR reaction.** *Vet Immunol Immunopathol* 2008, **126(1-2)**:156-62.
41. Arden B, Clark SP, Kabelitz D, Mak TW: **Mouse T-cell receptor variable gene segment families.** *Immunogenetics* 1995, **42(6)**:501-530.
42. Arden B, Clark SP, Kabelitz D, Mak TW: **Human T-cell receptor variable gene segment families.** *Immunogenetics* 1995, **42(6)**:455-500.
43. Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G: **IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains.** *Dev Comp Immunol* 2003, **27(1)**:55-77.
44. Gudikote JP, Wilkinson MF: **T-cell receptor sequences that elicit strong down-regulation of premature termination codon-bearing transcripts.** *EMBO J* 2002, **21(1-2)**:125-134.
45. Li S, Wilkinson MF: **Nonsense surveillance in lymphocytes?** *Immunity* 1998, **8(2)**:135-141.
46. Di Meo GP, Perucatti A, Schibler L, Incarnato D, Ferrara L, Cribeu EP, Iannuzzi L: **Thirteen type I loci from HSA4q, HSA6p, HSA7q and HSA12q were comparatively FISH-mapped in four river buffalo and sheep chromosomes.** *Cytogenet Cell Genet* 2000, **90(1-2)**:102-105.
47. Carvajal IM, Sen R: **Functional analysis of the murine TCR beta-chain gene enhancer.** *J Immunol* 2000, **164(12)**:6332-6339.
48. Gottschalk LR, Leiden JM: **Identification and functional characterization of the human T-cell receptor beta gene transcriptional enhancer: common nuclear proteins interact with the transcriptional regulatory elements of the T-cell receptor alpha and beta genes.** *Mol Cell Biol* 1990, **10(10)**:5486-5495.
49. Takeda J, Cheng A, Mauxion F, Nelson CA, Newberry RD, Sha WC, Sen R, Loh DY: **Functional analysis of the murine T-cell receptor beta enhancer and characteristics of its DNA-binding proteins.** *Mol Cell Biol* 1990, **10(10)**:5027-5035.
50. Henderson AJ, McDougall S, Leiden J, Calame KL: **GATA elements are necessary for the activity and tissue specificity of the T-cell receptor beta-chain transcriptional enhancer.** *Mol Cell Biol* 1994, **14(6)**:4286-4294.
51. Doty RT, Xia D, Nguyen SP, Hathaway TR, Willerford DM: **Promoter element for transcription of unrearranged T-cell receptor beta-chain gene in pro-T cells.** *Blood* 1999, **93(9)**:3017-3025.
52. Sikes ML, Gomez RJ, Song J, Oltz EM: **A developmental stage-specific promoter directs germline transcription of D beta J beta gene segments in precursor T lymphocytes.** *J Immunol* 1998, **161(3)**:1399-1405.
53. Wang X, Xiao G, Zhang Y, Wen X, Gao X, Okada S, Liu X: **Regulation of Tcrb recombination ordering by c-Fos-dependent RAG deposition.** *Nat Immunol* 2008, **9(7)**:794-801.
54. Anderson SJ, Chou HS, Loh DY: **A conserved sequence in the T-cell receptor beta-chain promoter region.** *Proc Natl Acad Sci USA* 1988, **85(10)**:3551-3554.
55. Yang L, Lanier ER, Kraig E: **Identification of a novel, spliced variant of CREB that is preferentially expressed in the thymus.** *J Immunol* 1997, **158(6)**:2522-2525.
56. Funkhouser W, Koop BF, Charmley P, Martindale D, Slightom J, Hood L: **Evolution and selection of primate T cell antigen receptor BV8 gene subfamily.** *Mol Phylogenet Evol* 1997, **8(1)**:51-64.
57. Nei M, Gu X, Sitnikova T: **Evolution by the birth-and-death process in multigene families of the vertebrate immune system.** *Proc Natl Acad Sci USA* 1997, **94(15)**:7799-7806.
58. Nei M, Rooney AP: **Concerted and birth-and-death evolution of multigene families.** *Annu Rev Genet* 2005, **39**:121-152.
59. Clark SP, Arden B, Kabelitz D, Mak TW: **Comparison of human and mouse T-cell receptor variable gene segment sub-families.** *Immunogenetics* 1995, **42(6)**:531-540.
60. Bosc N, Lefranc MP: **The mouse (Mus musculus) T cell receptor alpha (TRA) and delta (TRD) variable genes.** *Dev Comp Immunol* 2003, **27(6-7)**:465-497.
61. Sitnikova T, Su C: **Coevolution of immunoglobulin heavy- and light-chain variable-region gene families.** *Mol Biol Evol* 1998, **15(6)**:617-625.
62. Bradley DG, MacHugh DE, Cunningham P, Loftus RT: **Mitochondrial diversity and the origins of African and European cattle.** *Proc Natl Acad Sci USA* 1996, **93(10)**:5131-5135.
63. Hiendleder S, Lewalski H, Janke A: **Complete mitochondrial genomes of Bos taurus and Bos indicus provide new insights into intra-species variation, taxonomy and domestication.** *Cytogenet Genome Res* 2008, **120(1-2)**:150-156.
64. MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG: **Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (Bos taurus and Bos indicus).** *Genetics* 1997, **146(3)**:1071-1086.
65. Behlke MA, Chou HS, Huppi K, Loh DY: **Murine T-cell receptor mutants with deletions of beta-chain variable region genes.** *Proc Natl Acad Sci USA* 1986, **83(3)**:767-771.
66. Jouvin-Marche E, Trede NS, Bandeira A, Tomas A, Loh DY, Cazenave PA: **Different large deletions of T cell receptor V beta genes in natural populations of mice.** *Eur J Immunol* 1989, **19(10)**:1921-1926.
67. Noonan DJ, Kofler R, Singer PA, Cardenas G, Dixon FJ, Theofilopoulos AN: **Delineation of a defect in T cell receptor beta genes of NZW mice predisposed to autoimmunity.** *J Exp Med* 1986, **163(3)**:644-653.
68. Seboun E, Robinson MA, Kindt TJ, Hauser SL: **Insertion/deletion-related polymorphisms in the human T cell receptor beta gene complex.** *J Exp Med* 1989, **170(4)**:1263-1270.
69. Hedges SB, Dudley J, Kumar S: **TimeTree: a public knowledge-base of divergence times among organisms.** *Bioinformatics* 2006, **22(23)**:2971-2972.
70. Chretien I, Marcuz A, Fellah J, Charlemagne J, Du Pasquier L: **The T cell receptor beta genes of Xenopus.** *Eur J Immunol* 1997, **27(3)**:763-771.
71. Fellah JS, Durand C, Kerfour F, Charlemagne J: **Complexity of the T cell receptor Cbeta isotypes in the Mexican axolotl: structure and diversity of the VDJCbeta3 and VDJCbeta4 chains.** *Eur J Immunol* 2001, **31(2)**:403-411.
72. Nam BH, Hirono I, Aoki T: **The four TCR genes of teleost fish: the cDNA and genomic DNA analysis of Japanese flounder (Paralichthys olivaceus) TCR alpha-, beta-, gamma-, and delta-chains.** *J Immunol* 2003, **170(6)**:3081-3090.
73. Shigeta A, Sato M, Kawashima T, Horiuchi H, Matsuda H, Furusawa S: **Genomic organization of the chicken T-cell receptor beta chain D-J-C region.** *J Vet Med Sci* 2004, **66(12)**:1509-1515.
74. Zhou H, Bengten E, Miller NVW, Clem LW, Wilson M: **The T cell receptor beta locus of the channel catfish, Ictalurus punctatus, reveals unique features.** *J Immunol* 2003, **170(5)**:2573-2581.
75. Parra ZE, Baker ML, Hathaway J, Lopez AM, Trujillo J, Sharp A, Miller RD: **Comparative genomic analysis and evolution of the T cell receptor loci in the opossum Monodelphis domestica.** *BMC Genomics* 2008, **9**:111.
76. Williams CB, Blankenhorn EP, Byrd KE, Levinson G, Gutman GA: **Organization and nucleotide sequence of the rat T cell receptor beta-chain complex.** *J Immunol* 1991, **146(12)**:4406-4413.
77. Komatsu M, Lamoyi E, Mage RG: **Genomic DNA encoding rabbit T cell receptor beta-chains: isotypes and allotypes of C beta.** *J Immunol* 1987, **138(5)**:1621-1626.
78. Kotzin BL, Barr VL, Palmer E: **A large deletion within the T-cell receptor beta-chain gene complex in New Zealand white mice.** *Science* 1985, **229(4709)**:167-171.
79. McCormack WT, Tjoelker LW, Stella G, Postema CE, Thompson CB: **Chicken T-cell receptor beta-chain diversity: an evolutionarily conserved D beta-encoded glycine turn within the hyper-**

- variable CDR3 domain. *Proc Natl Acad Sci USA* 1991, **88(17)**:7699-7703.
80. Rudikoff S, Fitch WM, Heller M: **Exon-specific gene correction (conversion) during short evolutionary periods: homogenization in a two-gene family encoding the beta-chain constant region of the T-lymphocyte antigen receptor.** *Mol Biol Evol* 1992, **9(1)**:14-26.
  81. De Guerra A, Charlemagne J: **Genomic organization of the TcR beta-chain diversity (Dbeta) and joining (Jbeta) segments in the rainbow trout: presence of many repeated sequences.** *Mol Immunol* 1997, **34(8-9)**:653-662.
  82. Yang XO, Doty RT, Hicks JS, Willerford DM: **Regulation of T-cell receptor D beta 1 promoter by KLF5 through reiterated GC-rich motifs.** *Blood* 2003, **101(11)**:4492-4499.
  83. Halle JP, Haus-Seuffert P, Woltering C, Stelzer G, Meisterernst M: **A conserved tissue-specific structure at a human T-cell receptor beta-chain core promoter.** *Mol Cell Biol* 1997, **17(8)**:4220-4229.
  84. Lai E, Concannon P, Hood L: **Conserved organization of the human and murine T-cell receptor beta-gene families.** *Nature* 1988, **331(6156)**:543-546.
  85. Antonacci R, Lanave C, Del Faro L, Vaccarelli G, Ciccarese S, Massari S: **Artiodactyl emergence is accompanied by the birth of an extensive pool of diverse germline TRDVI genes.** *Immunogenetics* 2005, **57(3-4)**:254-266.
  86. Miccoli MC, Antonacci R, Vaccarelli G, Lanave C, Massari S, Cribiu EP, Ciccarese S: **Evolution of TRG clusters in cattle and sheep genomes as drawn from the structural analysis of the ovine TRG2@ locus.** *J Mol Evol* 2003, **57(1)**:52-62.
  87. The Bovine Genome Sequencing and Analysis Consortium, Elsik CG, Tellam RL, Worley KC: **The genome sequence of taurine cattle: a window to ruminant biology and evolution.** *Science* 2009, **324**:522-528.
  88. Connelley T: **Immunodominance, clonal composition and TCR'beta' repertoire of the bovine CD8+ T-cell response to Theileria parva.** Edinburgh: University of Edinburgh; 2007. [<http://www.ensembl.org>].
  90. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22(22)**:4673-4680.
  91. Brochet X, Lefranc MP, Giudicelli V: **IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis.** *Nucleic Acids Res* 2008:W503-508.
  92. Lefranc MP, Giudicelli V, Kaas Q, Duprat E, Jabado-Michaloud J, Scaviner D, Ginestoux C, Clement O, Chaume D, Lefranc G: **IMGT, the international ImMunoGeneTics information system.** *Nucleic Acids Res* 2005:D593-597. [<http://imgt.cines.fr/>].
  94. Sonnhammer EL, Durbin R: **A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis.** *Gene* 1995, **167(1-2)**:GCI-10. [<http://emboss.bioinformatics.nl>].
  96. RepeatMasker [<http://www.repeatmasker.org>]
  97. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14(6)**:1188-1190.
  98. Akamatsu Y, Tsurushita N, Nagawa F, Matsuoka M, Okazaki K, Imai M, Sakano H: **Essential residues in V(D)J recombination signals.** *J Immunol* 1994, **153(10)**:4520-4529.
  99. Hesse JE, Lieber MR, Mizuuchi K, Gellert M: **V(D)J recombination: a functional definition of the joining signals.** *Genes Dev* 1989, **3(7)**:1053-1061.
  100. Lefranc MP: **Nomenclature of the human T cell receptor genes.** *Curr Protoc Immunol* 2001, **Appendix I**:Appendix 10.
  101. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4(4)**:406-425.
  102. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24(8)**:1596-1599. [<http://www.megasoftware.net>].
  104. Nei M, Kumar S: **Molecular evolution and phylogenetics.** Oxford, UK: Oxford University Press; 2000.
  105. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52(5)**:696-704.
  106. PHYML [<http://atgc.lirmm.fr/phym1>]
  107. Anisimova M, Gascuel O: **Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative.** *Syst Biol* 2006, **55(4)**:539-552.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

