

Research article

Open Access

## Bos taurus genome sequence reveals the assortment of immunoglobulin and surrogate light chain genes in domestic cattle

Anna Ekman<sup>1</sup>, Mikael Niku<sup>2</sup>, Jenni Liljavirta<sup>1</sup> and Antti Iivanainen\*<sup>1</sup>

Address: <sup>1</sup>Department of Basic Veterinary Sciences, University of Helsinki, Helsinki, Finland and <sup>2</sup>Department of Applied Chemistry and Microbiology, University of Helsinki, Helsinki, Finland

Email: Anna Ekman - [anna.ekman@helsinki.fi](mailto:anna.ekman@helsinki.fi); Mikael Niku - [mikael.niku@helsinki.fi](mailto:mikael.niku@helsinki.fi); Jenni Liljavirta - [jenni.liljavirta@helsinki.fi](mailto:jenni.liljavirta@helsinki.fi); Antti Iivanainen\* - [antti.iivanainen@helsinki.fi](mailto:antti.iivanainen@helsinki.fi)

\* Corresponding author

Published: 30 April 2009

Received: 11 December 2008

BMC Immunology 2009, 10:22 doi:10.1186/1471-2172-10-22

Accepted: 30 April 2009

This article is available from: <http://www.biomedcentral.com/1471-2172/10/22>

© 2009 Ekman 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:** The assortment of cattle immunoglobulin and surrogate light chain genes has been extracted from the version 3.1 of *Bos taurus* genome sequence as a part of an international effort to sequence and annotate the bovine genome.

**Results:** 63 variable lambda chain and 22 variable kappa chain genes were identified and phylogenetically assigned to 8 and 4 subgroups, respectively. The specified phylogenetic relationships are compatible with the established ruminant light chain variable gene families or subgroups. Because of gaps and uncertainties in the assembled genome sequence, the number of genes might change in the future versions of the genome sequence. In addition, three bovine surrogate light chain genes were identified. The corresponding cDNAs were cloned and the expression of the surrogate light chain genes was demonstrated from fetal material.

**Conclusion:** The bovine kappa gene locus is compact and simple which may reflect the preferential use of the lambda chain in cattle. The relative orientation of variable and joining genes in both loci are consistent with a deletion mechanism in VJ joining. The orientation of some variable genes cannot be determined from the data available. The number of functional variable genes is moderate when compared to man or mouse. Thus, post-recombinatorial mechanisms might contribute to the generation of the bovine pre-immune antibody repertoire. The heavy chains probably contribute more to recombinational immunoglobulin repertoire diversity than the light chains but the heavy chain locus could not be annotated from the version 3.1 of *Bos taurus* genome.

### Background

Immunoglobulins are the molecular mediators of the adaptive humoral immune response in jawed vertebrates. Somatic recombination during B lymphoid differentiation is required for immunoglobulin expression [1]. In the germline state, the genes encoding for the variable (V), diversity (D) and joining (J) segments are dispersed across a wide genomic stretch. A process called V(D)J joining

brings together the specific genes for each segment type and thereby creates the second exon of a transcriptionally competent immunoglobulin gene. The recombination machinery consists of two recombination activating gene products RAG1 and RAG2 as well as various other proteins, reviewed in [2]. The *cis*-acting recognition signal sequences (RSSs) target the recombination machinery to the correct genomic site. Each RSS consists of heptamer

and nonamer motifs flanking a 12 or 23 bp long central spacer. In the rearranging locus, two variably separated double strand DNA breaks are introduced next to one 12 bp and one 23 bp RSS. The nascent non-homologous DNA ends are joined into a coding joint in the middle of the recombined gene. The DNA fragment between the breaks is either deleted or inverted depending on the relative orientation of the recombining genes.

The immunoglobulin heavy chain and light chain rearrangements in many species are temporally separated during B cell development. In mouse and man but not in chicken, a population of cells can be demonstrated that has undergone rearrangement only in the immunoglobulin heavy chain locus [3,4]. A surrogate light chain (SLC) is temporarily expressed at this stage of the B cell development [5]. SLC is composed of two polypeptides VPRED and IGLL1 that are homologous to the variable and the constant domain of the immunoglobulin light chain, respectively [6]. In mice, three VPRED paralogues VPRED1, VPRED2 and VPRED3 have been described [7,8]. The IGLV-like VPRED2 is missing from the human genome. Gene targeting studies demonstrate the role of SLC genes in the production of B cells [9].

The genome sequence of *Bos taurus* permits for the first time a direct estimate of the size of the immunoglobulin light chain gene pool in domestic cattle, one of the most important farm animal species. We have characterized the structure and composition of bovine immunoglobulin and surrogate light chain gene loci as a part of a community effort to annotate the version 3.1 assembly of *Bos taurus* genome sequence [10].

## Results

The bovine immunoglobulin lambda ( $\lambda$ ) chain locus is located on chromosome 17. In version 3.1 of the genome sequence (Btau\_3.1), 63 variable, 3 joining and 5 constant genes could be identified in 10 scaffolds. 25  $\lambda$  variable genes (ca. 41%) fulfilled the criteria for classification as potentially functional (see Methods and Additional file 1).

Based on the phylogenetic analyses and nucleotide sequence identities in a gene region corresponding to FR1–FR3, the  $\lambda$  variable genes can be grouped into 8 phylogenetic subgroups (figure 1, Additional files 1 and 2). The  $\lambda$  variable gene subgroups in the present work accommodate all the characterized bovine IGLV genes from [11] and most of the ovine IGLV genes [12-15]. Interspecies comparison revealed that four of the six described ovine gene families or subgroups [12-15] are represented in the bovine collection (figure 1 and Additional file 2) and contain 43 (ca. 68%) of the bovine genes. As can be seen from Additional file 1, subgroup 1 is the largest and contains 16

(ca 64%) of the potentially functional  $\lambda$  variable genes. This subgroup seems to be ruminant specific as no human or mouse genes co-segregate with its members. Subgroups 2 and 6 are each represented in the genome by a single subgroup-specific gene cluster. The 13 bovine genes of subgroup 5 are all pseudogenes as are the ovine genes in this subgroup. With the exception of one gene, the bovine (but not the ovine) genes in this subgroup share an in-frame stop codon in framework 3 (not shown). 20 genes (ca. 32%) of which 3 are potentially functional do not co-segregate with any members of the established ovine  $\lambda$  variable gene subgroups. However,  $\lambda$  variable genes in the bovine subgroups 7, 8 and 9 are similar to genes in human specific subgroups 5, 8 and 9 respectively (i.e., 80% nucleotide sequence identity in a gene region corresponding to FR1–FR3; not shown). No ovine, human or mouse genes closely related to IGLV41, IGLV47 or IGLV53 could be identified. Further, no bovine genes from the current assembly could be mapped to the established ovine families III or IV [12-15].

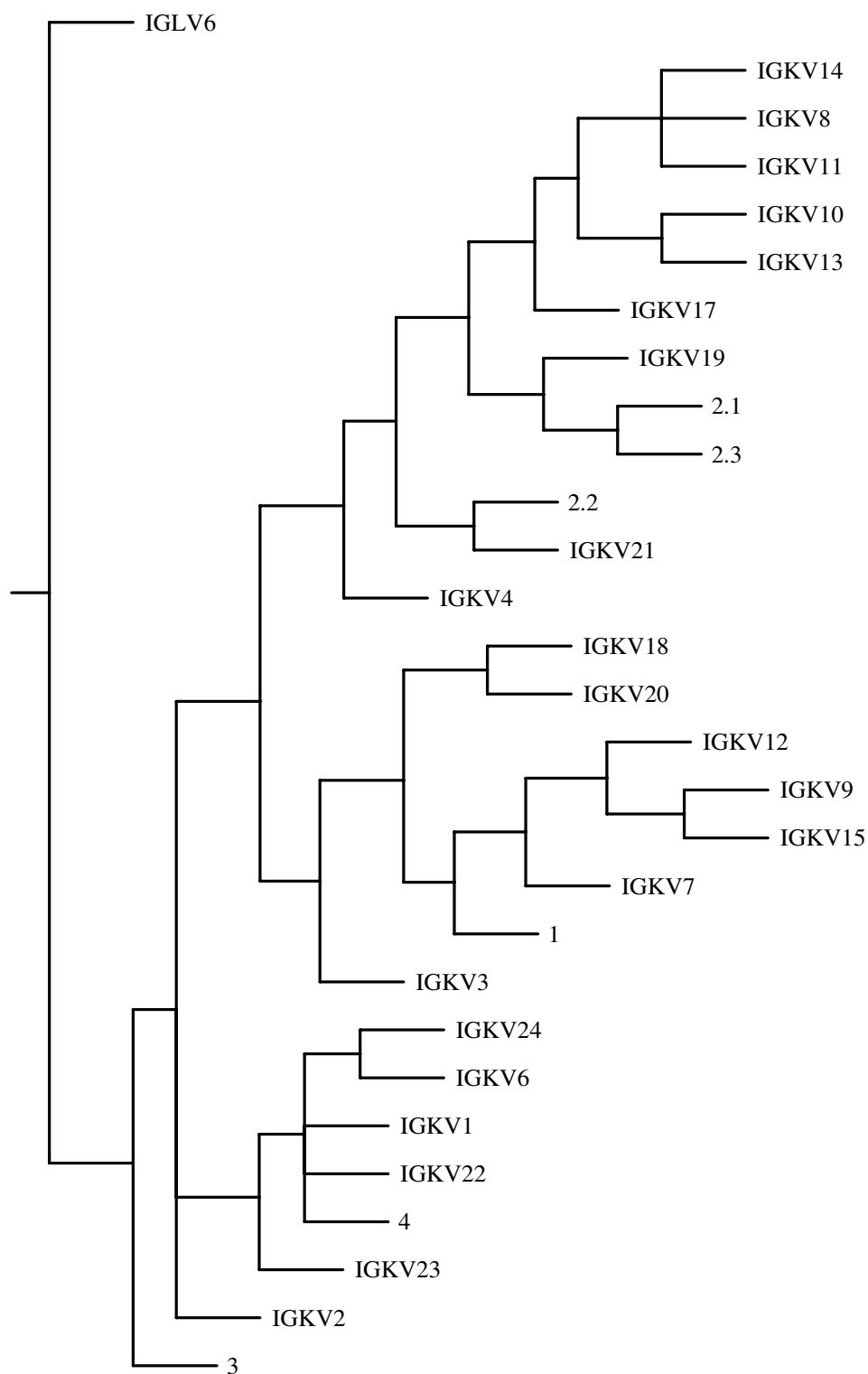
Three immunoglobulin lambda joining and five immunoglobulin lambda constant genes were identified (Additional file 1). Two of the J-C gene pairs form apparently functional units. IGLC1 and IGLC2 have identical coding sequence but differ at 3'UTR. Chen *et al.* [16] described four IGLC genes which correspond to IGLC2-IGLC5 in this paper.

The bovine immunoglobulin  $\kappa$  locus is located in chromosome 11. A blast search against Btau\_3.1 revealed matches only at a single location in scaffold Chr11.003.53. 22 variable, 3 joining and one constant immunoglobulin  $\kappa$  gene were identified. 8 variable genes (ca. 36%) were classified as potentially functional (Additional file 3). The variable  $\kappa$  genes can be grouped in 4 phylogenetic subgroups preserving the established ovine specific gene families (figure 2 and Additional file 4). 21 (ca. 95%) of the bovine genes co-segregate with ovine  $\kappa$  variable gene families I, II and IV [15]. Subgroup 2 is the largest and contains 7 (ca. 88%) of the potentially functional genes. All the seven members of subgroup 1 are pseudogenes.

Surrogate light chain genes VPRED1, IGLL1 and VPRED3 were identified in the assembly at or close to the  $\lambda$  chain locus. The VPRED1 and IGLL1 lie next to each other as in the mouse genome [8]. VPRED3 is located close to the immunoglobulin  $\lambda$  constant genes IGLC1-IGLC4 but in the opposite transcriptional orientation. None of the surrogate light chain genes is flanked by an RSS. The exon-intron boundaries of the surrogate light chain genes are conserved between cow, mouse and man (not shown). The VPRED1 and VPRED3 gene structures resemble those of the immunoglobulin  $\lambda$  variable genes with a leader and



**Figure 1**  
**Phenogram of ruminant immunoglobulin  $\lambda$  variable genes.** Sequences corresponding to the V region but excluding CRD3 were aligned and phylogenetically analyzed as described in the methods. *IGKV11* was used as an outgroup. Ovine genomic sequences 1.2, 1.3, 2.1, 2.2, 3, 3.1, 4.1, 4.2, 5.1, 5.2, 5.3, 5.4, 6.1, 6.2, 8.0, 9, 10, 12.1, 12.2, 16.1, 16.2, 17, 18, 26.1, 26.2, 26.3 (AF040900–AF040924, M60441) are from Reynaud *et al.* [13] Ovine sequences 6a, 6b, 6c and 6d (AF038145–AF038148) are derived from cDNA [14].



**Figure 2**  
**Phenogram of ruminant immunoglobulin  $\kappa$  variable genes.** Sequences corresponding to the V region but excluding CRD3 were aligned and phylogenetically analyzed as described in the methods. *IGLV6* was used as an outgroup. The ovine sequences 1, 2.1, 2.2, 2.3, 3 and 4 (AF038133–AF038138) are derived from cDNA [14].

main exon. Successful cloning of the cDNAs using primers that span the exon/intron boundaries and extend far into the 3'UTR of the germline genes confirms that the mRNA expression of surrogate light chain genes does not depend on recombination. The functionality of the surrogate light chain genes was additionally supported by demonstrating the expression of VPREB1, VPREB3 and IGLL1 mRNA in fetal liver, spleen, bone marrow, lymph node and thymus (figure 3).

The cDNA and genomic DNA sequence analysis of the surrogate light chain genes revealed several single nucleotide differences in comparison with the reference genomic sequence (Additional file 5). Therefore, it seems that the bovine surrogate light chain genes are polymorphic.

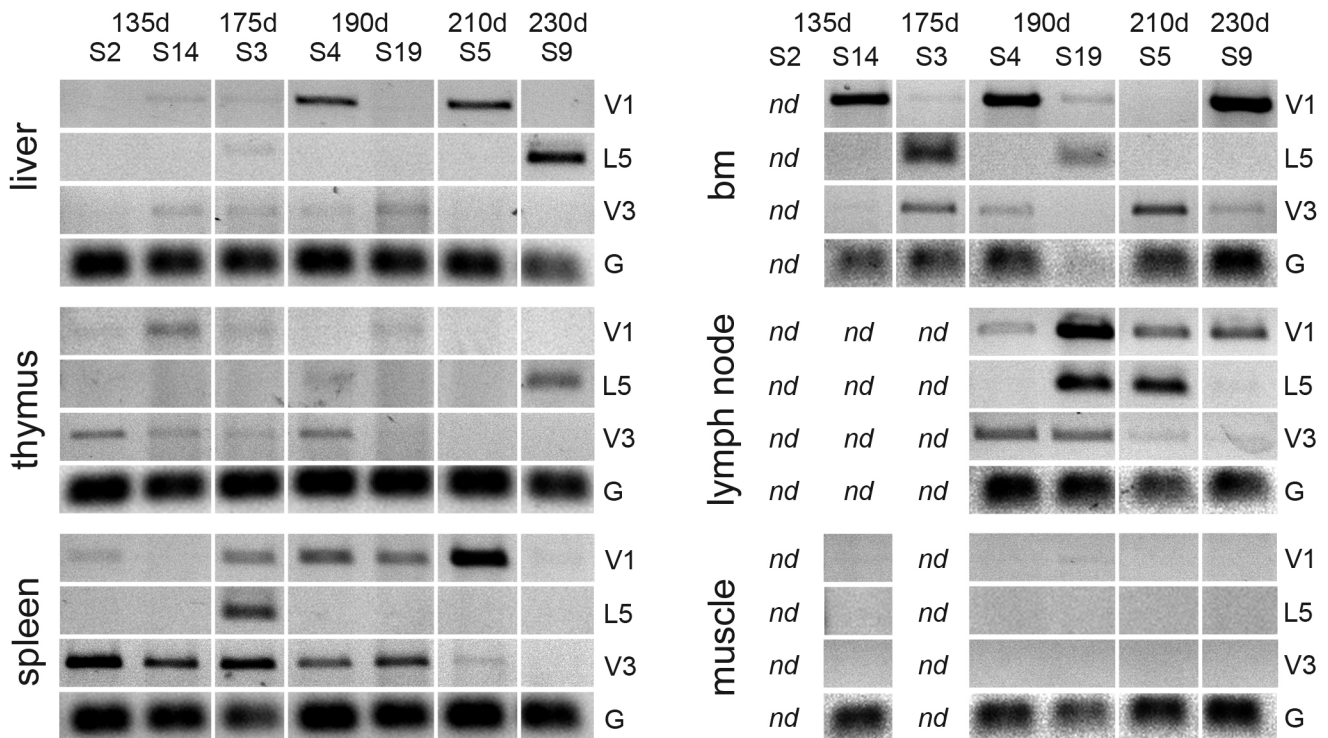
**Discussion**

In this paper, we have presented the analysis of the immunoglobulin and surrogate light chain gene assortment extracted from the *Bos taurus* genome sequence Btau\_3.1 [10]. Btau\_3.1 is nearly completely based on a whole genome shotgun sequence from a single animal (L1 Dominette 01449) with a 30% inbreeding coefficient [10,17].

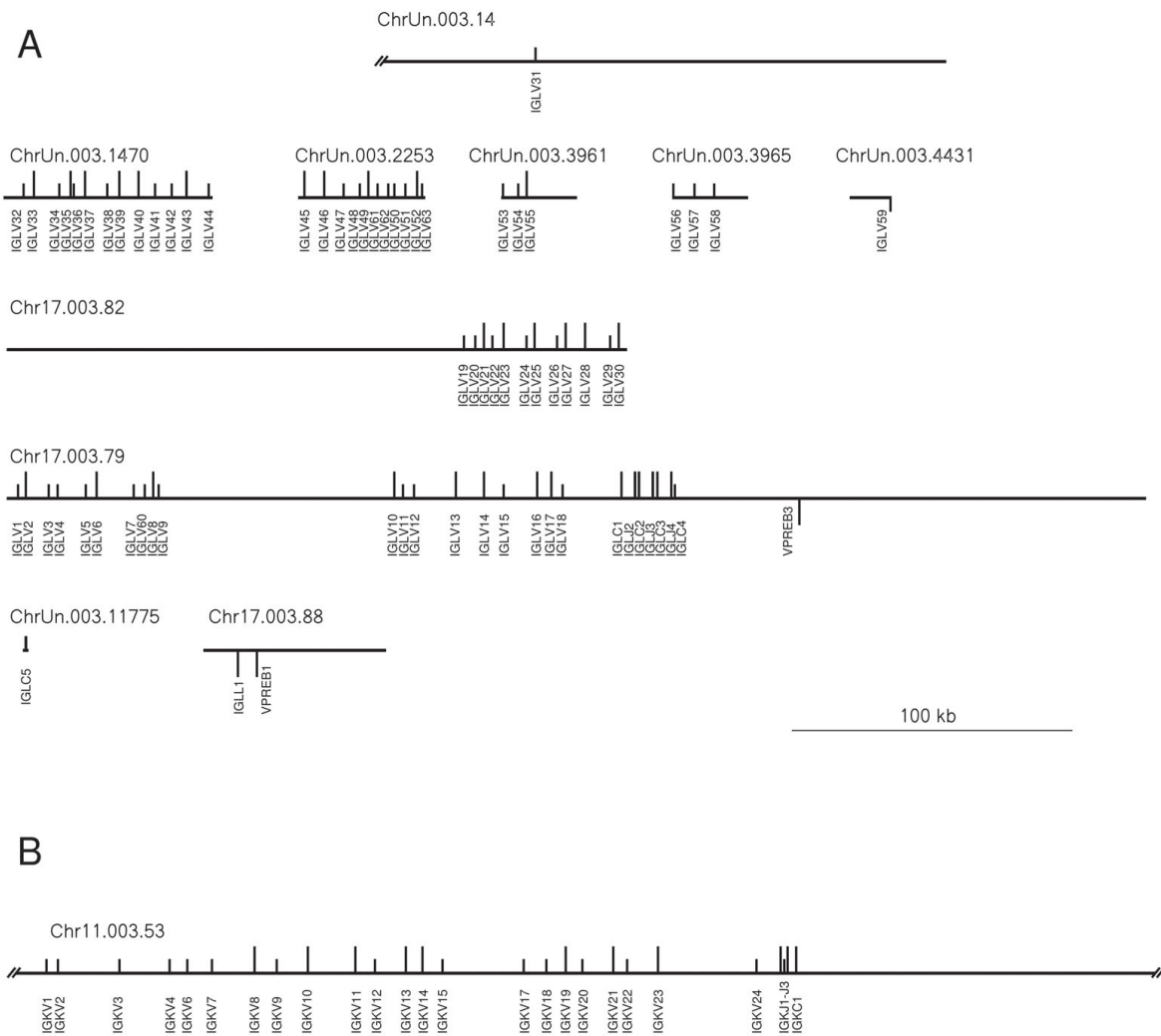
This facilitates the analysis of immunoglobulin genes, which is in mixed databases greatly complicated by gene polymorphism and targeted somatic mutations [18]. Most of the functional light chain genes have probably been included in our gene set although the exact number of genes is likely to change in the future genome versions. 32 λ variable genes were in genomic contigs not assigned to a specific chromosomal location and might include orphans.

An interspecies comparison suggests ruminant specific adaptations:

- (1.) The bovine κ locus appears small and uncomplicated when compared with the λ locus (see figure 4). This might reflect the preferential use of the λ light chain in cattle [19].
- (2.) In cattle, the number of functional immunoglobulin light chain genes is markedly lower than in mice and in man, i.e. 33 vs. 105 and 77 [20]. Further, the variation in CDR1 length is more restricted and the number of unique CDR1-CDR2 combinations is lower than in mice and man (see table 1). Even though some genes might be



**Figure 3**  
**Expression of VPREB1, VPREB3, and IGLL1 in bovine fetuses.** Gene expression was analyzed by RT-PCR as described in methods. The individual animals and their estimated fetal ages are indicated on top. The amplicon lengths were: 558 bp (VPREB1), 491 bp (VPREB3), 475 bp (IGLL1) and 150 bp (GAPDH). V1 = VPREB1, V3 = VPREB3, L5 = IGLL1, G = GAPDH, bm = bone marrow, nd = not determined.



**Figure 4**  
**Schematic organization of the kappa, lambda and surrogate light chain genes in Btau\_3.1 scaffolds.** The genes on the plus strand are projected above and those on minus strand below the scaffold line. Long projection stands for an apparently functional gene and a short projection for a nonfunctional gene. Scale bar: 100 kb. A: Lambda and surrogate light chain genes. B: Kappa light chain genes.

missing from Btau\_3.1, the number of potentially functional bovine light chain genes probably overestimates the *bona fide* functional genes for which protein evidence is required. We are not aware of any other approximations on the number of functional immunoglobulin genes in ruminants. By extrapolation, the total number of  $\lambda$  variable gene segments in the sheep genome has been estimated from 60–90 [14] to 150 [18]. The latter estimate is, however, based on cDNA data.

(3.) The phylogenetic analysis suggests that most of the potentially functional  $\lambda$  genes belong to a single subgroup

(subgroup 1, see additional file 6) that is not apparent in the human or in the mouse genomes but is present in sheep genome. This subgroup comprises 21 variable genes of which 16 are potentially functional. The CDR1 [21] is either 8 or 9 amino acids long with a characteristic hydrophobic residue at position 30. Based on similarities on primary sequences, the CDR 1 structures among the members of subgroup 1 correspond most closely to the canonical loop 1 structures 1 and 2 found on  $\lambda$  chain variable regions [22]. CDR2 is 3 amino acids long and probably adopts a hair pin structure commonly found on CDR2 of  $\lambda$  and  $\kappa$  light chains [22]. It remains to be seen whether or

**Table 1: Characteristics of CDR1 and CDR2 in the variable regions of bovine, mouse and human light chains**

	cattle	mouse	man
<b>lambda chain variable region</b>			
CDR1 length (amino acids)	6,8,9	7,8,9	6,7,8,9
CDR2 length (amino acids)	3,7	3,7	3,7
unique IGLV CDR1/CDR2 pairs	19	8	37
<b>kappa chain variable region</b>			
CDR1 length (amino acids)	6,10,11	5,6,7,10,11,12	6,7,10,11,12
CDR2 length (amino acids)	3	3	3
unique IGKV CDR1/CDR2 pairs	5	94	27

The data on human and murine immunoglobulins are from IMGT database [20].

not the CDRs adopt any of the established canonical immunoglobulin structures in reality. No high resolution structures are currently available for bovine immunoglobulins in the PDB archives [23].

(4.) The apparent expansion of the pseudogene subgroup 5 is intriguing although the reasons behind this are currently elusive. 12 subgroup members out of 13 share an identical stop codon in framework 3.

The data on the overall organization of the bovine λ chain locus is still quite fragmentary (figure 4). It could resemble the human locus, which displays a 900 kb long upstream region of 73 to 74 variable genes followed by 7 to 11 pairs

of joining and constant genes all in one transcriptional orientation [24]. However, recombination using inversion cannot be ruled out in the bovine λ chain locus at present. In contrast to what is found in man and cattle, the murine λ chain locus is much reduced in size (only about 240 kb) and contains two small clusters of different immunoglobulin lambda chain genes [reviewed in [25]].

The κ chain locus is much less complicated in cattle than in man or mouse. All identified κ genes were localized to a ca. 280 kb genomic segment within Chr11.003.53 (Additional file 3 and figure 4). In comparison, the κ locus spans ca. 1.8 Mb in man [26] and ca. 3.2 Mb in mouse [27,28]. The relative orientation of the bovine

Query id	Subject id	% identity	alignment length	mismatches	gap openings	q. star	q. end	s. start	s. end	e-value	bit score
gi 33241 emb X03957.1	11	66,98	642	142	15	169	774	1601468	1602075	1,00E-046	183
gi 33241 emb X03957.1	11	75,18	137	30	4	9	144	1601146	1601279	1,00E-015	80,6
gi 33241 emb X03957.1	11	71,08	166	45	3	876	1040	1602192	1602355	2,00E-013	73,4

**Alignment:**

```

Query: 476      acacgcagcctgggtgggactccaggagccagctcttaccctagagtttctgcacgggca 535
              ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 1601749 acactcagcctgggtaagag-ccaggggcccagctcttaccctagagtttctgcacgggca 1601807

Query: 536      gcaggttggcagcgcacactgtgggagccctagtggcagcccagggcgactcct 589
              ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 1601808 gcagattggcagcactcactgtgggagccctagttagcagcccagtgaggctcct 1601861
    
```

**Figure 5**  
**Pair wise alignment of the human kappa deleting element [gi = 33241] and Chr11.003.59.** Upper part: tabular output from bl2seq. The following parameters were used: blastn matrix: 2, -3; gap open: 5, gap extension: 2. Lower part: local alignment about 28.5 kb downstream from IGKC1. Heptanucleotide and nonanucleotide boxes are emphasized.

**Table 2: Gene specific primers used in this study**

Primer		Sequence
<b>VPREB1</b>		
VPREB1-f2	C, E	5'-catgtcctctggccctcgt-3'
VPREB1-r1	C, E	5'-gccagcctcctgtccac-3'
<b>VPREB3</b>		
VPREB3-fw1	C, E	5'-tgtgtggaggctcccgaag-3'
VPREB3-fw2	C	5'-cgcagaacacgggactcct-3'
VPREB3-rev1	C, E	5'-aggtcaggagtagaagtgg-3'
<b>IGLL1</b>		
L5-f1b	C	5'-ccagcgcgtctgcccaag-3'
L5-f2c	E	5'-tgctggctgggctctgg-3'
L5-r3a	C, E	5'-agaagggacgtaggggaccat-3'
<b>GAPDH</b>		
QGAPDfw	E	5'-ctgacctccgctggag-3'
QGAPDrev	E	5'-aagagtgagtgtcgtttgaag-3'

The use of each primer is indicated: C (cloning of the full length cDNA), E (RT-PCR expression analysis).

genes allows recombination exclusively by deletion. In mice, the relative orientation of more than 75 of the 140 murine  $\kappa$  variable genes supports inversion [29,30]. In man, a large duplication has produced nearly identical copies of 34 variable  $\kappa$  genes ca. 800 kb upstream in an opposite orientation. In addition, 2 variable genes most proximal to the J-C gene region support inversion [26]. A kappa deleting element (kde) homologous to the murine recombining sequence (RS) is located about 24 kb downstream of the human  $\kappa$  locus [31]. A permanent disruption of one or both  $\kappa$  loci by a recombination involving kde (RS) is a frequent finding in human and murine B cells that produce the  $\lambda$  light chain [32,33]. Interestingly, a highly similar sequence to kde is located 28.5 kb downstream of the bovine  $\kappa$  locus (see figure 5).

The heavy chain locus could not be annotated as most of it is missing from Btau\_3.1. The available data on the light chain loci suggests that a moderate number of potentially functional light chain genes exist in the bovine genome. Although the heavy chains add more to the recombinatorial diversity of immunoglobulins than the light chains,

post-recombinatorial mechanisms might also contribute to a fully blown bovine preimmune repertoire. The relative importance of V(D)J recombination for the generation of the preimmune repertoire in ruminants is currently controversial [14,18]. In late fetal and neonatal sheep, however, the repertoire is expanded by somatic hypermutation in the ileal Peyer's patch [12,13].

Surrogate light chain (SLC) is needed to expand the H+L-cell population in species in which heavy and light chain genes are sequentially arranged. This assures that sufficient number of cells productively rearrange both loci [34]. The expression of SLC genes in the bovine fetal tissues (figure 3) confirms their functionality. The data presented in this paper does not permit further conclusions on the role of SLC genes in cattle. Nevertheless, analyses of serial sections by immunohistochemistry have revealed specific sites in the bovine fetus where there are no light chain positive cells but which still contain heavy chain positive cells (Ekman and Iivanainen, unpublished).

## Conclusion

This study describes the bovine assortment of immunoglobulin and surrogate light chain genes based on Btau\_3.1. A large fraction of the potentially functional variable genes belong to subgroups that are shared between cattle and sheep but not found in man or in mouse. The number of functional light chain variable genes in Btau\_3.1 is moderate in comparison with the corresponding number in the human or mouse genomes. The new data on the immunoglobulin light chain genes provides novel insight on the humoral immune system of ruminants and should facilitate the development of vaccines and other therapeutic tools against cattle specific infectious diseases.

## Methods

### Gene identification and annotation

An iterative blast search against the bovine genomic sequence database was performed via Ensembl genome browser [35]. The initial query sequences were bovine light chain variable gene encoded cDNAs with frequent matches in the dbEST database at the National Center for Biotechnology Information [36]. Genome-wide annotation evidence based on Swiss-Prot, TrEMBL and various other databases at GenBank, EMBL and DDBJ were provided by The Wellcome Trust Sanger Institute [37] and by the Bovine Genome Database [38]. Annotation of the genomic sequence and its comparison against the various evidence entries was carried out using Apollo [39], Otterlace [40] and blast [41].

### Functional and phylogenetic analyses of genes

Sequence extractions were done in the European Molecular Biology Open Software Suite [42]. The extracted genes



were further analyzed using the following criteria: (a) an uninterrupted open reading frame, (b) consensus splice sites at exon/intron boundaries, (c) the presence of four conserved framework residues C23, W41, L89 and C104 for the variable and constant genes, and F/W-G-X-G motif for the joining genes [21], and (d) a likely functional recombination signal sequence. In functional recombination assays, the spacer length and three outmost nucleotides of the heptamer have been shown to be the most critical parameters for efficient recombination [43].

Multiple alignments of genomic sequences corresponding to regions spanning from FR1 up to but excluding CDR3 [21] were performed using a global alignment strategy in the MAFFT package, version 6.603b [44]. Evolutionary distances were computed and phylogenetic trees constructed in PHYLIP, version 3.67 [45] using the F84 model for nucleotide substitution and neighbor joining algorithm, respectively. The reliability of the tree topologies were evaluated using the bootstrap test ( $n = 1000$ ) in PHYLIP. The consensus tree was calculated using majority rule in the Consense consensus tree program in PHYLIP.

Since the complete gene pool is not available, *ad hoc* gene names are used in this paper. The variable gene families or subgroups identified in cattle [11] and in sheep [12-15] are used where the phylogenetic analyses indicate a close relationship. Furthermore, nucleotide sequence identity matrix for the gene region corresponding to FR1-FR3 (e.g., amino acids 1 to 104 in the IMGT numbering system [21]) was calculated from globally aligned sequences using the BioEdit Sequence Alignment Editor v. 7.0.9 [46]. Truncated or incomplete genes *IGVL59*, *IGLV61*, *IGLV62* and *IGLV63* were excluded from the initial alignment. They were subsequently assigned to the respective subgroups by phylogenetic analysis in PHYLIP, based on alignments using the local alignment strategy in the MAFFT package (Additional file 1).

#### **Cloning and expression analysis of the surrogate light chain genes**

Bovine fetal material was obtained from a local slaughterhouse. The use of animal tissues was approved by the local animal welfare authorities. Total RNA was isolated from muscle, thymus, liver, spleen, lymph node and bone marrow of fetuses at 135, 175, 190, 210 and 230 days of gestational age [47]. 50 – 400 mg of frozen tissue was crushed with a mortar, suspended in Eurozol RNA extraction reagent (Euroclone) and homogenized using Polytron PT1200 homogenizer (Kinematica AB) with a 5 mm cutter. The extraction procedure was carried out according to manufacturer's instructions. RNA was further purified by precipitating with 2.5 M LiCl (Sigma) and dissolved in

water. Prior to reverse transcription RNA was treated with RQ1 DNase (Promega) to remove possible genomic contamination. In the reverse transcription reaction 20 pmol of oligo(dT) primer was added to 1 µg of total RNA, and RevertAid M-MuLV reverse transcriptase (Fermentas) was used according to manufacturer's instructions. RiboLock ribonuclease inhibitor (Fermentas) was added to the reaction.

For the amplification and cloning of the full length cDNAs, the following primers were used: VPRED1-f2 and VPRED1-r1, VPRED3-fw1, VPRED3-fw2 and VPRED3-rev1, L5-f1b and L5-r3a (table 2). Purified PCR fragments were ligated to a pSTBlue-1 vector (Novagen). For each cDNA, several clones were sequenced on ABI3130 XL 16-capillary sequencer at the DNA-sequencing core facility at the University of Helsinki using fluorescently labeled BigDye™ dideoxynucleotides. To confirm suspected polymorphisms in the *VPRED1*, *VPRED3* and *IGLL1* genes, a selection of cDNA clones from lymph node and bone marrow, and PCR products from genomic DNA were sequenced.

The expression of *VPRED1*, *VPRED3* and *IGLL1* surrogate light chain genes was confirmed by RT-PCR using the following RNA preparations (age in gestational days): bone marrow (135d, 175d, 190d, 210d, 230d), liver (135d, 175d, 190d, 210d, 230d), lymph node (190d, 210d, 230d), muscle (135d, 190d, 210d, 230d), spleen (135d, 175d, 190d, 210d, 230d), and thymus (135d, 175d, 190d, 210d, 230d). Expression of the housekeeping gene *GAPDH* was used to monitor the variation in RNA quality and quantity. *GAPDH* specific control RT-PCRs without reverse transcriptase did not yield any products (not shown). For primers, see table 2.

#### **Abbreviations**

H: heavy chain; L: light chain; FR: framework region; CDR: complementarity determining region; RSS: recombination signal sequence; IGLV: immunoglobulin lambda variable; IGLJ: immunoglobulin lambda joining; IGLC: immunoglobulin lambda constant; IGKV: immunoglobulin kappa variable; IGKJ: immunoglobulin kappa joining; IGKC: immunoglobulin kappa constant; VPRED: pre-B lymphocyte gene; IGLL: immunoglobulin lambda-like polypeptide; SLC: surrogate light chain; RAG: recombination activating gene; GAPDH: glyceraldehyde phosphate dehydrogenase; kde: kappa deleting element; RS: recombination sequence.

#### **Authors' contributions**

AI conceived the study, analyzed and annotated the sequences, participated in the wet lab work and wrote the paper. AE and JL participated in the wet lab work. MN

participated in the phylogenetic analyses. All authors contributed to and approved the final manuscript.

## Additional material

### Additional file 1

**Table S1 – Immunoglobulin  $\lambda$  genes in Btau\_3.1.** F = functional, FL = fragmented locus, ORF = open reading frame, P = pseudogene, T = truncated pseudogene. Variations in the RSS heptamer motif and spacer length that likely prevent recombination are in bold and underlined. Scaffold coordinates are used.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2172-10-22-S1.xls>]

### Additional file 2

**Table S2- Sequence identity matrix of ruminant immunoglobulin  $\lambda$  variable genes.** Sequences corresponding to the V region but excluding CRD3 were aligned and a sequence identity matrix was calculated as described in the methods. Ovine genomic sequences 1.2, 1.3, 2.1, 2.2, 3, 3.1, 4.1, 4.2, 5.1, 5.2, 5.3, 5.4, 6.1, 6.2, 8.0, 9, 10, 12.1, 12.2, 16.1, 16.2, 17, 18, 26.1, 26.2, 26.3 (AF040900–AF040924, M60441) are from Reynaud et al. [13]. Ovine sequences 6a, 6b, 6c and 6d (AF038145–AF038148) are derived from cDNA [14]. Identities  $\geq 80\%$  are in bold. Subgroups are bounded and colored light green. Diagonal cells are marked as ID.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2172-10-22-S2.xls>]

### Additional file 3

**Table S3 – Immunoglobulin  $\kappa$  genes in Btau\_3.1.** F = functional, FL = fragmented locus, ORF = open reading frame, P = pseudogene. Variations in the RSS heptamer motif that likely prevent recombination are in bold and underlined. Scaffold coordinates are used.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2172-10-22-S3.xls>]

### Additional file 4

**Table S4 – Sequence identity matrix of ruminant immunoglobulin  $\kappa$  variable genes.** Sequences corresponding to the V region but excluding CRD3 were aligned and a sequence identity matrix was calculated as described in the methods. The ovine sequences 1, 2.1, 2.2, 2.3, 3 and 4 (AF038133–AF038138) are derived from cDNA [14]. Identities  $\geq 80\%$  are in bold. Subgroups are bounded and colored light green. Diagonal cells are marked as ID.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2172-10-22-S4.xls>]

### Additional file 5

**Table S5 – Single nucleotide differences in VPRED1, VPRED3 and IGLL1.** cDNA clones or PCR products from genomic DNA were sequenced. The sequences were compared against the Btau\_3.1 minus strand of Chr17.003.88 scaffold reference sequence. Both alleles are indicated whenever possible.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2172-10-22-S5.xls>]

### Additional file 6

**Table S6 – V and J region amino acid sequences of the apparently functional bovine immunoglobulin light chains.** Dots indicate gaps that were introduced during the alignment. White spaces separate consecutive FRs and CDRs.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2172-10-22-S6.pdf>]

## Acknowledgements

This study was supported by grants from The Academy of Finland (122540/2007 to AI), The Research Funds of The University of Helsinki (914/51/2006 to AI) and Finnish Ministry of Agriculture and Forestry (828/312/2009 to AI). The authors thank Kirsti Sihto, DVM for help in collecting the fetal material, and Kirsi Lahti and Tuire Pankasalo for expert technical assistance.

## References

1. Tonegawa S: **Somatic generation of antibody diversity.** *Nature* 1983, **302**:575-581.
2. Sekiguchi J, Alt F, Oettinger M: **The mechanism of V(D)J recombination.** In *Molecular biology of B cells* Edited by: Honjo T, Alt F, Neuberger M. Elsevier; 2003:61-82.
3. Kearney JF, Won WJ, Benedict C, Moratz C, Zimmer P, Oliver A, Martin F, Shu F: **B cell development in mice.** *Int Rev Immunol* 1997, **15**:207-41.
4. Weill JC, Reynaud CA: **Galt versus bone marrow models of B cell ontogeny.** *Dev Comp Immunol* 1998, **22**:379-385.
5. Ogawa M, ten Boekel E, Melchers F: **Identification of CD19(-)B220(+)-c-Kit(+)-Flt3/Flk-2(+) cells as early B lymphoid precursors before pre-B-I cells in juvenile mouse bone marrow.** *Int Immunol* 2000, **12**:313-324.
6. Kerr WG, Cooper MD, Feng L, Burrows PD, Hendershot LM: **Mu heavy chains can associate with a pseudo-light chain complex (psi L) in human pre-B cell lines.** *Int Immunol* 1989, **1**:355-61.
7. Kudo A, Melchers F: **A second gene, VpreB in the lambda 5 locus of the mouse, which appears to be selectively expressed in pre-B lymphocytes.** *EMBO J* 1987, **6**:2267-72.
8. Shirasawa T, Ohnishi K, Hagiwara S, Shigemoto K, Takebe Y, Rajewsky K, Takemori T: **A novel gene product associated with mu chains in immature B cells.** *EMBO J* 1993, **12**:1827-1834.
9. Shimizu T, Mundt C, Licence S, Melchers F, Mårtensson IL: **VpreB1/VpreB2/lambda 5 triple-deficient mice show impaired B cell development but functional allelic exclusion of the IgH locus.** *J Immunol* 2002, **168**:6286-6293.
10. The Bovine Genome Sequencing and Analysis Consortium: **The Genome Sequence of Taurine Cattle: A window to ruminant biology and evolution.** *Science* 2009, **324**:522.
11. Sinclair MC, Gilchrist J, Aitken R: **Molecular characterization of bovine V lambda regions.** *J Immunol* 1995, **155**:3068-3078.
12. Reynaud CA, Mackay CR, Müller RG, Weill JC: **Somatic generation of diversity in a mammalian primary lymphoid organ: the sheep ileal Peyer's patches.** *Cell* 1991, **64**:995-1005.
13. Reynaud CA, Garcia C, Hein WR, Weill JC: **Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process.** *Cell* 1995, **80**:115-125.
14. Reynaud CA, Dufour V, Weill JC: **Generation of diversity in mammalian gut-associated lymphoid tissues: restricted V gene usage does not preclude complex V gene organization.** *J Immunol* 1997, **159**:3093-3095.
15. Hein W, Dudler L: **Diversity of Ig light chain variable region gene expression in fetal lambs.** *Int Immunol* 1998, **10**:1251-1259.
16. Chen L, Li M, Li Q, Yang X, An X, Chen Y: **Characterization of the bovine immunoglobulin lambda light chain constant IGLC genes.** *Vet Immunol Immunopathol* 2008, **124**:284-294.
17. **Bovine Genome Project** [<http://www.hgsc.bcm.tmc.edu/projects/bovine/>]

18. Jenne CN, Kennedy LJ, McCullagh P, Reynolds JD: **A new model of sheep Ig diversification: shifting the emphasis toward combinatorial mechanisms and away from hypermutation.** *J Immunol* 2003, **170**:3739-3750.
19. Butler JE: **Immunoglobulin gene organization and the mechanism of repertoire development.** *Scand J Immunol* 1997, **45**:455-462.
20. **The international ImMunoGeneTics information system®** [<http://imgt.cines.fr>]
21. Lefranc MP, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G: **IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V like domains.** *Dev Comp Immunol* 2003, **27**:55-77.
22. Al-Lazikani B, Lesk AM, Chothia C: **Standard conformations for the canonical structures of immunoglobulins.** *J Mol Biol* 1997, **273**:927-948.
23. **RCSB Protein databank** [<http://www.rcsb.org>]
24. Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Schmeits JL, Wang J, Shimizu N: **One-megabase sequence analysis of the human immunoglobulin lambda gene locus.** *Genome Res* 1997, **7**:250-261.
25. Lefranc M-P, Lefranc G: **Immunoglobulin Lambda (IGL) Genes of Human and Mouse.** In *Molecular Biology of B Cells* Edited by: Honjo T, Alt FW, Neuberger M. London: Elsevier Academic Press; 2004:37-59.
26. Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Asakawa S, Sasaki T, Klobeck HG, Combriato G, Zachau HG, Shimizu N: **Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the V kappa genes.** *Eur J Immunol* 2001, **31**:1017-1028.
27. Zachau HG: **The immunoglobulin kappa genes and the kappa locus of the mouse.** [<http://biochemie.web.med.uni-muenchen.de/zachau/kappa>].
28. Zachau HG: **Immunoglobulin kappa genes of Human and Mouse.** In *Molecular Biology of B Cells* Edited by: Honjo T, Alt FW, Neuberger M. London: Elsevier Academic Press; 2004:27-36.
29. Rösenthaller F, Kirschbaum T, Heim V, Kirschbaum V, Schäble KF, Schwendinger J, Zocher I, Zachau HG: **The 5' part of the mouse immunoglobulin kappa locus.** *Eur J Immunol* 1999, **29**:2065-2071.
30. Thiebe R, Schäble KF, Bensch A, Brensing-Küppers J, Heim V, Kirschbaum T, Mitlöchner H, Ohnrich M, Pourrajabi S, Rösenthaller F, Schwendinger J, Wichelhaus D, Zocher I, Zachau HG: **The variable genes and gene families of the mouse immunoglobulin kappa locus.** *Eur J Immunol* 1999, **29**:2072-2081.
31. Klobeck HG, Zachau HG: **The human CK gene segment and the kappa deleting element are closely linked.** *Nucleic Acids Res* 1986, **14**:4591-4603.
32. Moore MW, Durdik J, Persiani DM, Selsing E: **Deletions of kappa chain constant region genes in mouse lambda chain-producing B cells involve intrachromosomal DNA recombinations similar to V-J joining.** *Proc Natl Acad Sci USA* 1985, **82**:6211-6215.
33. Siminovitch KA, Bakhshi A, Goldman P, Korsmeyer SJ: **A uniform deleting element mediates the loss of kappa genes in human B cells.** *Nature* 1985, **316**:260-262.
34. Mårtensson IL, Keenan RA, Licence S: **The pre-B-cell receptor.** *Curr Opin Immunol* 2007, **19**:137-142.
35. **The Ensembl project** [<http://www.ensembl.org>]
36. **The National Center for Biotechnology Information** [<http://www.ncbi.nlm.nih.gov>]
37. **The Wellcome Trust Sanger Institute** [<http://www.sanger.ac.uk>]
38. **The Bovine Genome Database** [<http://www.bovinegenome.org>]
39. Lewis SE, Searle SM, Harris N, Gibson M, Lyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby MA, Kaminker JS, Matthews BB, Prochnik SE, Smithy CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME: **Apollo: a sequence annotation editor.** *Genome Biol* 2002, **3**:RESEARCH0082.
40. Searle SM, Gilbert J, Lyer V, Clamp M: **The otter annotation system.** *Genome Res* 2004, **14**:963-970.
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
42. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**:276-277.
43. 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**:4520-4529.
44. Katoh K, Misawa K, Kuma K, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Res* 2002, **30**:3059-3066.
45. Felsenstein J: **PHYLIP – Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989, **5**:164-166.
46. Hall T: **BioEdit Sequence Alignment Editor.** [<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>].
47. Rüsse I: **Rind.** In *Lehrbuch der Embryologie der Haustiere* Edited by: Rüsse I. Berlin: Parey; 1991:159-168.

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)

