Immunogenetics (1999) 50:301–308

ORIGINAL PAPER

Chen Su · Ingrid Jakobsen · Xun Gu · Masatoshi Nei

Diversity and evolution of T-cell receptor variable region genes in mammals and birds

Received: 20 May 1999 / Revised: 13 July 1999

Abstract The receptor of a T lymphocyte (TCR) recognizes nonself antigens in the company of major histocompatibility complex (MHC) molecules presented to it by the antigen-presenting cell. The variable region of TCR is encoded by either a concatenation of variable region (*TCR-V*), diversity region (*TCR-D*), and joining region (*TCR-J*) genes, or a concatenation of *TCR-V* and *TCR-J* genes. The *TCR-V* genes exist as a multigene family in vertebrate species. Here we study the evolutionary relationships of *TCR-V* genes from humans, sheep, cattle, rabbits, mice, and chicken. These six species can be classified into two groups according to the frequency of $\gamma\delta$ T-cells in their peripheral T-cell populations. The "$\gamma\delta$ low" group of species includes humans and mice, in which $\gamma\delta$ T-cells constitute very limited portion of the T-cell population. The "$\gamma\delta$ high" group includes sheep, cattle, rabbits, and chicken, in which $\gamma\delta$ T-cells comprise up to 60% of the T-cell population. Here, we compiled *TCR-V* sequences from the six species and conducted a phylogenetic analysis. We identified various *TCR-V* gene subgroups based on the analysis. We found that humans and mice have representatives from nearly all of the subgroups identified, while other species have lost subgroups to different extent. Therefore, the $\gamma\delta$ low species have a high degree of diversity of *TCR-V* genes, while $\gamma\delta$ high species all have limited diversity of *TCR-V* genes. This pattern is similar to that found for immunoglobulin variable region (*IGV*) genes.

C. Su (✉) · I. Jakobsen · X. Gu · M. Nei
Institute of Molecular Evolutionary Genetics and Department of Biology, The Pennsylvania State University, 208 Mueller Lab, University Park, PA 16802, USA
e-mail: cxs513@psu.edu
Tel.: +1-814-8657030
Fax: +1-814-8637336

Present address:
X. Gu, Iowa Computational Molecular Biology Lab and Department of Zoology and Genetics, Science II Building, Iowa State University, Ames, IA 50011, USA

Key words T-cell receptors · Variable region genes · Evolution · Phylogeny · Diversity

Introduction

T lymphocytes have the ability to mount a specific cell-mediated immune response to foreign antigens. Each T lymphocyte expresses a unique TCR heterodimer that reacts with a specific antigen peptide mostly bound to a cell-associated MHC molecule. T lymphocytes can be divided into two subsets according to the specificity of their heterodimer antigen receptors: $\alpha\beta$ T lymphocytes and $\gamma\delta$ T lymphocytes. The former lymphocytes express receptors composed of α and β polypeptide chains, and the latter express receptors composed of γ and δ chains. Each of the polypeptides is encoded by a combination of separate genes. For example, α and γ polypeptides are encoded by *TCR-V* genes, *TCR-J* genes, and *TCR-C* genes, whereas β and δ polypeptides are encoded by these three genes plus diversity (*TCR-D*) genes. Furthermore, the *TCR-V* regions can be divided into complementarity-determining regions (CDRs) and framework regions (FRs), where the CDRs congregate on the surface of the TCR molecules which comes in contact with the antigen-MHC assemblage, while the FRs provide the framework structure of three-dimensional conformations. The CDRs have been shown to contribute most of the specificity of MHC recognition and antigen binding by the TCRs (e.g., Chothia et al. 1988; Davis and Bjorkman 1988; Hong et al. 1992).

The $\gamma\delta$ T-cells show different frequencies and physiological distributions in different animal species. The $\gamma\delta$ high species, which include sheep, cattle, rabbits, and chicken, have a much higher proportion (up to 60%) of $\gamma\delta$ T lymphocytes in the peripheral T-cell pool than the $\gamma\delta$ low species (e.g., humans and mice), in which the proportion of $\gamma\delta$ T-cells is around 5% (Sawasdikosol et al. 1993; see Hein and Dudler 1993 and references therein).

To investigate the difference between the $\gamma\delta$ high and $\gamma\delta$ low species, we considered whether the degree of diversity of the *TCR-V* genes is correlated with the frequency of $\gamma\delta$ T-cells. Humans and mice have been shown to have a seemly high degree of diversity of the *TCR-V* genes (e.g., Arden et al. 1995a, b; Clark et al. 1995; Rowen et al. 1996), and this may be a general feature of $\gamma\delta$ low species. However, the level of diversity of *TCR-V* genes for the $\gamma\delta$ high species has not been studied.

The purpose of this paper is to address this question using the sequence information available in the public database. The analysis of compiled sequence data now enables us to reconstruct the evolutionary history of *TCR-V* genes and examine the degree of gene diversity in different species. This study also provides insight into the mechanism of evolution and function of TCR molecules.

Materials and methods

TCR-V sequences used in the study

In humans, the *TCRA* (encoding α chain) and *TCRD* (encoding δ chain) loci are mapped to the same position, near the centromere of Chromosome (Chr) 14. In fact, the *TCRD* locus is located between *TCRAV* and *TCRAJ* clusters (Klein and Hořejši 1997) so that the *TCRAV* and *TCRDV* genes are closely linked. The *TCRB* (encoding β chain) and *TCRG* (encoding γ chain) loci occupy distinct regions on different arms of Chr 7. The *TCR-V* genes from the four loci can be grouped into families (based upon 75% sequence similarity criterion), in which members of a given family are more similar to each other than they are to members of any other families.

In this analysis, we used representative *TCR-V* sequences for each gene family found in humans, sheep, cattle, rabbits, mice, and chicken (see Table 1 for number of gene families found in each of these species and footnotes for references), excluding pseudogenes. We chose these six species because their *TCR-V* gene repertoires were relatively well characterized, and a sufficient number of *TCRAV*, *TCRBV*, *TCRGV*, and *TCRDV* sequences is available at the time of this study. The DNA sequences of the *TCR-V* genes from these species were retrieved from GenBank, along with officially adopted individual gene names (WHO-IUIS nomenclature sub-committee on TCR designation, 1995) wherever applicable. We used cDNA sequences as well as germline sequences. The mixed usage of cDNA and germ line sequences should not affect the results of our phylogenetic analysis, since somatic mutations rarely occur in the *TCR-V* genes (Ikuta et al. 1985).

Here we list the names of the sequences used, together with their GenBank accession numbers. To simplify gene notation, we used the letter "*A*" followed by a unique number (the number of the gene family to which the gene belongs wherever possible) to designate the *TCRAV* genes. Similarly, we used the letter "*B*" for *TCRBV* genes, "*D*" for *TCRDV* genes, and "*G*" for *TCRGV* genes. In humans and mice, there are a small number of *TCRV* genes that can be rearranged to *TCRAJ* or *TCRDJ* genes to form either an a chain or a d chain, and they are designated *AD* genes.

TCRAV sequences used in the analysis: (1) human (*Homo sapiens*): *A1* (D13077), *A2* (L11159), *A3* (M13726), *A4* (L06886), *A5* (D13069), *AD6* (Z14996), *A7* (L11161), *A8* (L11162), *A9* (M13737), *A10* (D13075), *A11* (M13742), *A12* (X70310), *A13* (M27374), *AD14* (M95394), *A15* (Z22965), *A16* (M17651), *AD17* (D13071), *A18* (M17661), *A19* (M17662), *A20* (M17663), *AD21* (M17664), *A22* (D13072), *A23* (X58736), *A24* (X58737), *A25* (X58738), *A26* (X58739), *A27* (M23431), *AD28* (X61070), *A29* (X58768), *A30* (X70305), *A31* (X70306), *A32* (D13073); (2) mouse (*Mus musculus*): *A1* (M31647), *AD2* (X06771), *A3* (M33586), *AD4* (M34198), *A5* (X02967), *AD6* (M37279), *AD7* (M37597), *A8* (M38680), *A9* (X60319), *AD10* (M38102), *AD11* (M73263), *A12* (X03668), *A13* (M38681), *A14* (D90229), *A15* (X57397), *AD17* (M16118), *A19* (M22604); (3) sheep (*Ovis aries*): *A622* (M55622), *A035* (U78035); (4) cattle (*Bos taurus*): *A011* (D90011), *A012* (D90012), *A013* (D90013), *A014* (D90014), *A015* (D90015), *A016* (D90016), *A017* (D90017); (5) rabbit (*Oryctolagus cuniculus*): *A885* (M12885); (6) chicken (*Gallus gallus*): *A611* (U04611), *A612* (U04612), *A613* (U04613).

TCRBV sequences used in the analysis: (1) human (*Homo sapiens*): *B1* (M27904), *B2* (L05149), *B3* (Z22967), *B4* (Z29580), *B5* (X04927), *B6* (M14262), *B7* (M13855), *B8* (X07192), *B9* (X57614), *B10* (M16309), *B11* (X74845), *B12* (L36092), *B13* (X61446), *B14* (M14267), *B15* (M11951), *B16* (X04933), *B17* (M27388), *B18* (M27189), *B19* (M27390), *B20* (M13554), *B21* (M33233), *B22* (L36092), *B23* (U03115), *B24* (U03115), *B25* (U03115), *B26* (U03115); (2) mouse (*Mus musculus*): *B1* (M29878), *B2* (M21203), *B3* (M12415), *B4* (M13674), *B5* (M15613), *B6* (M10093), *B7* (X00696), *B8* (M15616), *B9* (M13677), *B10* (X56702), *B11* (N00046), *B13* (M31648), *B14* (M11858), *B15* (X04047), *B16* (X03671), *B17* (M61184), *B18* (X16695), *B19* (X16691), *B20* (X59150), *B21* (X16689), *B22* (X16690), *B23* (X16692), *B24* (X16693), *B25* (X16694); (3) sheep (*Ovis aries*): *B1S5* (AF030011), *B2S1* (AF030012), *B3S1* (AF030013), *B4S1* (AF030016), *B6S1* (AF030017), *B7S1* (AF030018), *B8S1* (AF030019), *B10S1* (AF030020), *B12S1* (AF030021), *B13S1* (AF030023), *B15S1* (AF030024), *B17S1* (AF030025), *B22S1* (AF030026); (4) cattle (*Bos taurus*): *B122* (D90122), *B123* (D90123), *B124* (D90124), *B125* (D90125), *B126* (D90126), *B129* (D90129); (5) rabbit (*Oryctolagus cuniculus*): *B2*

Table 1 Number of *TCR-V* gene families and genomic genes identified from five mammalian and one avian species

	TCRAV	*TCRBV*	*TCRGV*	*TCRDV*
Human[b]	32 (45)[a]	34 (75)	6 (14)	3 (4)[c]
Mouse[d]	21 (100)	20 (25)	5 (7)	4 (10)[c]
Rabbit	2 (??)[e]	9 (>11)[f]	2 (>6)[g]	5 (>5)[h]
Sheep	?? (>4)[i]	17 (>20)[j]	6 (>13)[k]	7 (>28)[k]
Cattle	7 (20)[l]	9 (>22)[m]	7 (15–20)[n]	1 (40–50)[n]
Chicken	2 (20–25)[o]	2 (a few)[p]	3 (>26)[q]	2 (??)[r]

[a] Number of gene families (based upon 75% sequence similarity criterion) with the number of genomic genes shown in parentheses.
[b] Arden et al. (1995a); Klein and Hořejši (1997)
[c] A small number of *TCRAV* genes were found to be used as *TCRDV* genes in $\gamma\delta$ T-cells in humans and mice, and they are not included in the *TCRDV* gene family
[d] Wang et al. (1994); Arden et al. (1995b); Klein and Hořejši (1997)
[e] Marche and Kindt (1986)
[f] Isono et al. (1994)
[g] Isono et al. (1995)
[h] Kim et al. (1995)
[i] Hein et al. (1991); Massari et al. (1997)
[j] Grossberger et al. (1993)
[k] Hein and Dudler (1993); Hein (1994)
[l] Ishiguro et al. (1990)
[m] Tanaka et al. (1990)
[n] Hein and Dudler (1997)
[o] Gobel et al. (1994)
[p] Tjoelker et al. (1990)
[q] Six et al. (1996)
[r] Chen et al. (1996)

(D17416), *B5* (D17417), *B6* (D17418), *B7S1* (D17419), *B8* (D17423), *B9* (D17424), *B10* (D17425), *B11* (D17426), *B1* (M14576); (6) chicken (*Gallus gallus*): *B1S1* (M37798), *B2S2* (M37806).

TCRGV sequences used in the analysis: (1) human (*Homo sapiens*): *G1* (M13429), *G2* (M27335), *G3* (S60779), *G4* (S69780); (2) mouse (*Mus musculus*): *G1* (M13337), *G2* (M13338), *G3* (M13336), *G4* (Z49051), *G5* (Z22847); (3) sheep (*Ovis aries*): *G1S1* (Z12998), *G2S1* (Z12999), *G2S2* (Z13000), *G2S3* (Z13001), *G2S4* (Z13002), *G3S1* (Z13003), *G4S1* (Z13004), *G5S1* (Z13005), *G5S2* (Z13006), *G6S1* (Z13007); (4) cattle (*Bos taurus*): *G119* (D16119), *G126* (D16126), *G129* (D16129), *G130* (D16130), *G131* (D16131), *G133* (D16133), *G186* (U73186), *G187* (U73187), *G188* (U73188); (5) rabbit (*Oryctolagus cuniculus*): *G1S1* (D38135), *G1S2* (D38137), *G1S3* (D38138), *G1S4* (D38139), *G2S1* (D38142); (6) chicken (*Gallus gallus*): *G1S3* (U78210), *G1S4* (U78212), *G1S5* (U78213), *G1S8* (U78216), *G2S7* (U78225), *G2S8* (U78226), *G2S9* (U78227), *G3S3* (U78230), *G3S4* (U78231), *G3S8* (U78235).

TCRDV sequences ssed in the analysis: (1) human (*Homo sapiens*): *D101* (X14545), *D102* (X72501), *D103* (M23326); (2) mouse (*Mus musculus*): *D101* (X13314), *D104* (M37280), *D105* (M37282); (3) sheep (*Ovis aries*): *D1S1* (AJ005903), *D2* (AJ005904), *D3S2* (Z12996), *D4* (AJ005906), *D5* (AJ005907), *D6* (AJ005908), *D7* (AJ005909); (4) cattle (*Bos taurus*): *D113* (D16113), *D116* (D16116); (5) rabbit (*Oryctolagus cuniculus*): *D1* (D26555), *D4* (D38120), *D5* (D38121).

Phylogenetic analysis

The four gene sets (*TCRAV*, *TCRBV*, *TCRGV*, and *TCRDV*) were aligned separately using the CLUSTAL W computer program (Thompson et al. 1994) with minor visual adjustments. The CDRs are highly variable and difficult to align, so they were omitted. The analysis was therefore based on an alignment of the FRs only. Sequences with relatively long gaps (>6 base pairs) were excluded.

Phylogenetic analysis was done using the MEGA computer program (Kumar et al. 1993). First, we estimated distance between amino acid sequences using *p* distance (the percentage dissimilarity). Sites with gaps were ignored for each pair of sequences (pairwise deletion option in the MEGA program). This option was appropriate in this case because only a small number of gaps were present in the final alignment and most of the gaps were shared by a group of closely related sequences. We used the neighbor-joining (NJ; Saitou and Nei 1987) method to reconstruct the phylogenetic trees. The reliability of trees was examined by the bootstrap test (Felsenstein 1985) and the interior-branch test (Rzhetsky and Nei 1992; Sitnikova 1996), which produced the bootstrap probability (P_B) and confidence probability (CP) values, respectively, for each interior branch in the tree. $P_B \geq 95\%$ and $CP \geq 95\%$ are commonly considered statistically significant. However, $P_B > 80\%$ was also interpreted as high statistical support for interior branches in the tree, since the bootstrap test is known to be conservative (Hillis and Bull 1993; Sitnikova et al. 1995; Zharkikh and Li 1992).

The maximum parsimony (MP) method as implemented in PAUP*4.0 (Swofford 1998) was also used to examine the reliability of topologies generated by the NJ method. Two types of bootstrapping were conducted for the MP trees. First, the standard stepwise addition + TBR (full heuristic) search was done for 50 bootstrap replicates. We could not conduct more bootstrap replicates because this method is very time consuming. Second, the fast heuristic (random addition of taxa and no branch swapping once the first tree is created) was used for 10 000 bootstrap replicates. Although the second method is much faster, it was not very useful for our data since it had consistently lower bootstrap values and often showed topologies different from those generated by the full heuristic method and the NJ method. Therefore, we present results only from the full heuristic search for MP method below.

Results

TCRAV and TCRDV gene tree

In humans, sheep, cattle, and mice, *TCRAV* and *TCRDV* sequences have higher sequence similarity to each other than to other *TCR-V* genes, and the *TCRDV* loci are physically mapped between the *TCRAV* and the *TCRAJ* loci (Arden et al. 1995a, b; Caccia et al. 1985; Hein and Mackay 1991; Solinas-Toldo et al. 1995). Therefore, the *TCRAV* and *TCRDV* sequences from various species were combined in the same alignment and are shown in a joint phylogenetic tree (NJ) in Figure 1. A tree generated by the MP method shows essentially the same topology as that by the NJ method except that the MP tree had very low resolution for deep branches. The bootstrap values found by MP method are presented in the figure, along with P_Bs by the NJ method and CPs by the interior branch test, for the branches that are important for identification of the gene subgroups (see below).

Our phylogenetic tree shows that the *TCRAV* genes from various vertebrate species form clusters separate from the *TCRDV* genes, although the bootstrap value is not very high. It is likely that the *TCRAV* and the *TCRDV* genes diverged before the divergence of mammals and birds, because chicken *TCRAV* genes do not appear to be at the basal branch of the tree. This is consistent with the observation that both $\alpha\beta$ and $\gamma\delta$ T-cell lineages are also found in sharks and skates (Litman and Rast 1996; Rast et al. 1995, 1997), indicating that *TCRAV* and *TCRDV* lineages were established before the divergence of cartilaginous fish and bony vertebrates. Interestingly, human and mouse *AD* genes belong to both *TCRAV* and *TCRDV* clusters, with no dominance of either type, suggesting that these *AD* genes originated from both *TCRAV* and *TCRDV* genes. These *AD* genes are dispersed along the whole *TCRA* and *TCRD* chromosome region and are not located in any specific location.

We found that a small number of genes cluster with those from the other locus. For example, the mouse *D102* gene has sequence similarity to human *A4* and *A20* and mouse *A12* genes and belong to a cluster of *TCRAV* genes, while the human *A12* gene belongs to a cluster of *TCRDV* genes and has highest sequence similarity to the human *AD6* and mouse *AD17* genes. These two genes may have been subject to gene conversion from the other genes or unequal crossing-over.

We classified the *TCRAV* and *TCRDV* genes into subgroups to demonstrate the evolutionary relationship of these genes. This classification was based on P_B values ranging from 55% to 100% on the deepest nodes wherever applicable. For example, the *TCRAV* genes can be divided into four clusters that we have named subgroup A, B, C, and D (see Fig. 1). These subgroups are supported by P_B values ranging from 64% to 100%

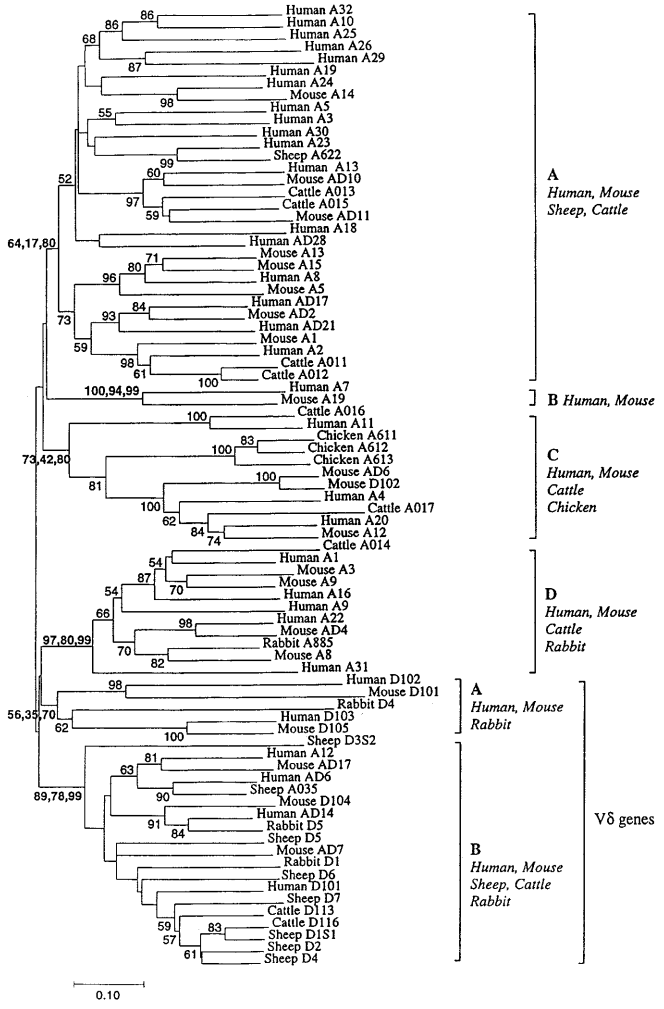

Fig. 1 Phylogenetic (NJ) tree of 48 *TCRAV*, 19 *TCRDV*, and 12 *TCRADV* sequences from five mammalian and one avian species. Four *TCRAV* subgroups and two *TCRDV* subgroups are indicated by *brackets*. A total length of 75 amino acids per sequence was used. The P_B values generated by the NJ method are shown on the branches wherever P_B values >50%. However, for those branches used to identify the gene subgroups, the P_B values from the MP method and the interior branch test are also shown, in order, following the NJ P_B values. Those P_B values are indicated by *bold* fonts. Although not every part of the topology of this NJ tree could be reproduced by the MP method, the gene subgroups identified are supported by both NJ and MP methods

and CP values ranging from 80% to 99% (Fig. 1). Each of the *TCRAV* subgroups is not necessarily represented in the genomes of all species. Human and mouse *TCRAV* genes are found in four subgroups, whereas cattle *TCRAV* genes are found in three subgroups, and sheep, rabbit, and chicken *TCRAV* genes are found in only one subgroup each. This suggests that the common ancestor of amniotes must have possessed *TCRAV* genes from all four subgroups and that three different subgroups have been lost in the bird and the rabbit lineage, respectively, and one subgroup has been lost in the artiodactyl lineage with subsequent subgroup loss in sheep.

The evolutionary pattern of the *TCRDV* genes is somewhat different. The *TCRDV* genes can be divided into two subgroups, A and B, based on the P_B values of 56% and 89%, and CPs of 70% and 99%, respectively (see Fig. 1). Human, mouse, and rabbit *TCRDV* genes appear in both of the two subgroups, whereas sheep and cattle *TCRDV* genes are seen in only one subgroup. This suggests that the earliest divergence of *TCRDV* genes occurred before mammalian radiation and subsequently one subgroup of *TCRDV* genes has been lost in the artiodactyl lineage. This scenario could be further investigated if chicken or other avian *TCRDV* sequences become available.

TCRBV gene tree

In the phylogenetic tree for *TCRBV* sequences shown in Fig. 2, we identified six *TCRBV* subgroups that are statistically supported by P_B values ranging from 70% to 100% (Fig. 2). We excluded a cattle *TCRBV* gene

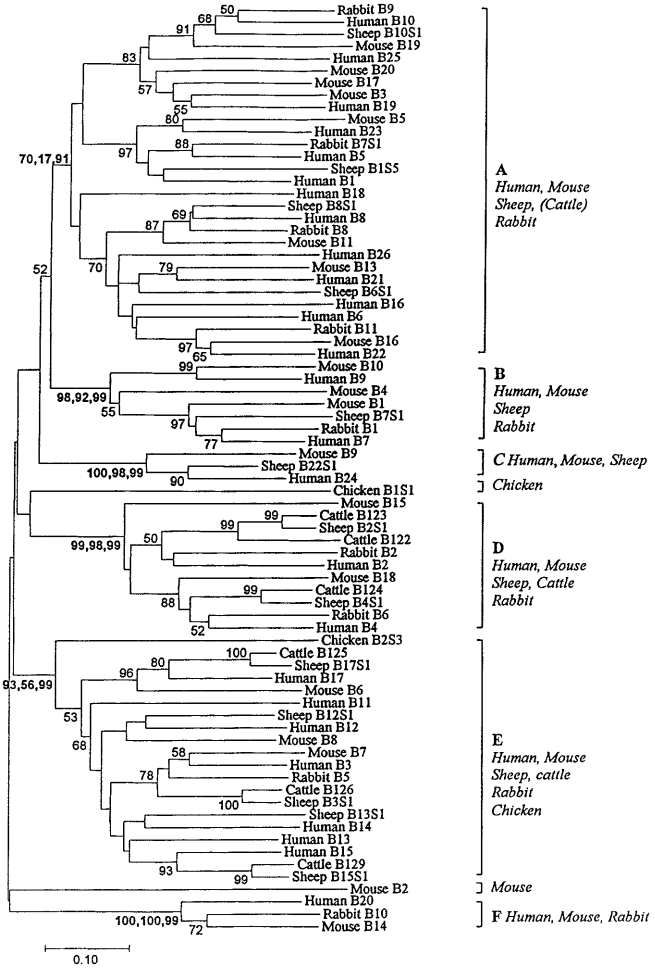

Fig. 2 Phylogenetic (NJ) tree of 75 *TCRBV* sequences from six species. Six *TCRBV* subgroups and two single *TCRBV* genes are indicated by *brackets*. A total length of 74 amino acids per sequence was used. The P_B values are indicated in the same way as for Fig. 1. All subgroups are also supported by the MP method

from our data set, because only partial sequence of this gene was available. However, when it was included, it belongs to the subgroup A, and we indicate this by marking "cattle" beside subgroup A in a pair of brackets. One chicken sequence (chicken *B1S1*) and one mouse sequence (mouse *B2*) did not reliably cluster with any other sequences and are therefore excluded from the six subgroups (see Fig. 2). The MP tree identified the same six subgroups, with the same exclusion of chicken *B1S1* and mouse *B2* genes, although the P_B value supporting the subgroup A by the MP method is very low ($P_B = 17\%$; see Fig. 2). However, all CP values supporting these six subgroups by the interior-branch test are higher than 90%.

According to our tree, human and mouse *TCRBV* genes are found in all six subgroups, while sheep and rabbit *TCRBV* genes belong to five of the subgroups, *TCRBV* genes from cattle belong to three subgroups, and chicken has *TCRBV* genes from only one subgroup. We therefore propose that the divergence of the six subgroups predated the separation between mammals and birds and that artiodactyls and rabbits each lost one subgroup while chicken lost five of them. Furthermore, cattle lost two more subgroups that are still present in sheep.

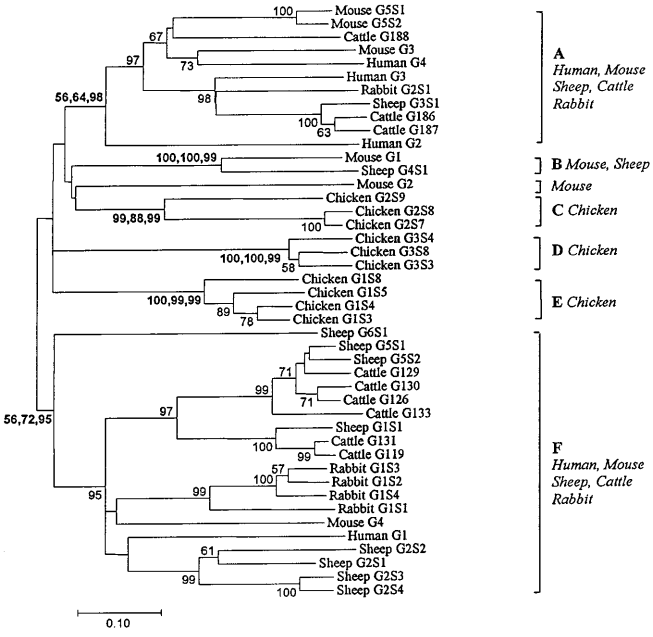

Fig. 3 Phylogenetic tree of 44 *TCRGV* sequences from six species. Six *TCRGV* subgroups and one single *TCRGV* gene are indicated by brackets. A total length of 74 amino acids per sequence was used. The P_B values are indicated in the same way as for Fig. 1. The subgroups are also supported by the MP method except for group A (see results)

TCRGV gene tree

The phylogenetic tree of *TCRGV* sequences from various species shows that *TCRGV* genes form six subgroups, A, B, C, D, E, and F, with the exception of one mouse sequence (mouse *G2*; see Fig. 3). A tree generated by the MP method generally supports the existence of these six subgroups, with the same exclusion of the mouse *G2* gene. However, in the MP tree the human *G2* gene did not necessarily cluster with subgroup A, and the MP bootstrap value indicated for the branch leading to subgroup A is for the cluster without the human *G2* gene. Four of the subgroups are supported by nearly 100% P_B and CP values, indicating high reliability of these clusters. One of the other two subgroups (subgroup F) is composed of one highly supported cluster ($P_B = 95\%$) plus one sheep sequences (sheep *G6S1*), and this subgroup is supported by 72% P_B value from MP method and 95% CP value from the interior-branch method. Among these six subgroups, three subgroups, namely, C, D, and E, consist of only chicken *TCRGV* sequences and are highly supported by P_B values (>99%), suggesting that all present chicken *TCRGV* genes are duplicated from three ancestral *TCRGV* sequences. Moreover, the sequence similarity is relatively high within each chicken subgroup, indicating that these gene duplications happened only recently.

Human, rabbit, and cattle *TCRGV* sequences belong to two subgroups, A and F, whereas sheep and mouse *TCRGV* sequences are found in three of the subgroups, A, B, and F (see Fig. 3). Therefore, the ancestor of mammals would have had three subgroups, A, B, and F, while humans, rabbits, and cattle all lost subgroups B. Chicken *TCRGV* genes are divergent from the *TCRGV* genes of mammals and generally do not intersperse with other species in the phylogenetic tree. This pattern, however, was not seen in the *TCRAV/DV* and *TCRBV* gene trees presented above.

Discussion

Evolution of TCR-V genes in humans and mice: γδ low species

The constitution of the human *TCR-V* gene repertoire is similar to that of the mouse repertoire. Humans and mice both have representatives from all subgroups identified except for *TCRGV* genes. In *TCRGV* genes, there are three subgroups found exclusively in chicken, and another small group containing only one mouse and one sheep *TCRGV* sequences (subgroup B in Fig. 3). In the other trees, human and mouse *TCR-V* genes disperse all along the tree, while genes from other species tend to be confined to a few clusters. Interestingly, humans and mice are generally considered to be more distantly related than are humans and artiodactyls or humans and rabbits (e.g., Easteal 1988; Li et al. 1990), although it is still controversial (e.g., Novacek 1992). Therefore, given that humans have *TCR-V* gene repertoires distinct from sheep, cattle, and rabbits, it is

surprising that humans and mice have similar *TCR-V* repertoires.

Sitnikova and Su (1998) studied the*IGV* genes from humans, mice, sheep, cattle, rabbits, and chicken and found that humans and mice have similar *IGV* gene repertoires and the highest diversity of *IGV* gene repertoire among these six species. When combined with these results, it seems that the immune system genes from humans and mice have more similarity to each other than to those from the other species, in contrast to the distant evolutionary relationship between them. However, it is not clear whether there are functional reasons for this.

Evolution of TCR-V genes in sheep, cattle, rabbits, and chicken: γδ high species

With respect to the *TCR-V* genes, sheep and cattle are the most well-described organisms except humans and mice. Although sheep and cattle are very closely related species, they show different degrees of diversity of *TCRAV* and *TCRBV* genes. Sheep retain *TCRBV* genes from five subgroups, two of which are not found in cattle (see Fig. 2), whereas cattle possess *TCRAV* genes from four subgroups, two of which are absent in sheep (see Fig. 1). The degree of diversity for the *TCRGV* genes is also different in sheep and cattle (see Fig. 3). Considering that sheep and cattle diverged only around 20 million years ago (Kumar and Hedges 1998), the loss of *TCR-V* subgroups seems to occur frequently in the ruminant lineages. It is worth pointing out that although sheep and cattle contain similar repertoires of *IGV* genes, pigs, a species closely related to ruminants, have a different *IGV* gene repertoire (Sitnikova and Su 1998). Furthermore, camels, a species also closely related to ruminants, have a form of Ig molecules with no light chains. This suggests that the repertoire of *TCR-V* genes as well as of other immune system genes may vary greatly within the order of Artiodactyla. In this respect, it would be interesting to study more wild species of the Artiodactyla, such as deer and hippopotamus, which occupy ecological niches different from those of domesticated sheep, cattle, pigs, and camels.

The genomic structure and function of rabbit and chicken *TCR-V* genes are not well characterized. However, the chicken *TCR-V* genes seem to have the lowest diversity of all the species we examined. For example, chicken has *TCRAV* genes from only one subgroup and *TCRBV* genes from only two subgroups. Chicken also has a limited *TCRGV* gene repertoire, which shows no particular sequence similarity to those from the other species. Interestingly, chicken also has the most restricted *IGV* gene repertoire yet found in vertebrates (Ota and Nei 1995; Reynaud et al. 1989). It is not clear how and why these apparently limited immune system gene repertoires in chicken came into being.

The rabbit *IGV* gene repertoire is probably as limited as that of chicken (see Knight 1992 and references therein), and rabbits practically use only one V_H (variable gene encoding the heavy chain) gene in *VDJ* recombination to produce antibodies. Compared with chicken, however, rabbits seem to have a higher degree of diversity for *TCR-V* genes. Rabbits have one subgroup of *TCRAV* genes, two subgroups of *TCRDV* genes, five subgroups of *TCRBV* genes, and two subgroups of *TCRGV* genes. However, there are preferences in *TCRGV* and *TCRDV* gene usage in different rabbit tissues. For example, the rabbit *D1* gene is predominantly expressed in adult peripheral blood lymphocytes (PBLs) as part of the γδ TCR heterodimer. In this respect, the usage preference of rabbit *TCR-V* genes is similar to the rabbit *IGV* genes.

Coevolutionary relationship between TCR-V and IGV gene families

A distinct feature of the phylogenetic relationship of *TCR-V* genes is that, in general, human and mouse sequences disperse all along the phylogenetic tree, while genes from sheep, cattle, rabbits, and chicken are restricted to a few clusters. This result is the same as that obtained from the *IG* V_H and V_L (variable gene encoding light chains) gene trees, where humans and mice contain diverse gene repertoires, and genes from the other species have more restricted diversity (Sitnikova and Su 1998; see Table 2).

The Ig molecules can be either anchored in the cell membrane of B cells or secreted into biological fluids, while the TCR molecules are functional only on the surface of T cells and only recognize short antigen peptides bound to MHC molecules except for some γδ T cells. Both B cells and T cells can recognize a huge spectrum of antigens. The correlation between the diversity of *TCR-V* genes and *IGV* genes in a species may be due to mechanisms for producing diversified B cells and T cells. In humans, for example, virtually no somatic hypermutation or somatic gene conversion is involved in generation of the primary antibody repertoire (Klein and Hořejši 1997). Therefore, humans maintain a diverse *IGV* gene repertoire to cope with a large spectrum of antigens. However, the lack of somatic gene diversification does not seem to be the reason for the higher level of diversity of *TCR-V* gene repertoires

Table 2 Percentage of γδ T cells in the circulating T-cell populations and diversity of *TCR-V* and *IGV* genes in six species

	Percentage of γδ T cells	Diversity of *TCR-V* genes	Diversity of *IGV* genes
Human	Low (~5%)	High	High
Mouse	Low (~5%)	High	High
Rabbit	High (~20%)	Low	Low
Sheep	High (~30%)	Low	Low
Cattle	High (~30%)	Low	Low
Chicken	High (~20%)	Low	Low

in humans and mice, since there are no somatic gene conversion/hypermutation events in *TCR-V* genes in any of the six species we studied here.

The reason humans and mice have a low frequency of $\gamma\delta$ T cells in the peripheral T-cell pool while other species have a much higher frequency of $\gamma\delta$ T cells is unclear, partly because the function of $\gamma\delta$ T cells is not fully understood. Recent studies suggest that $\gamma\delta$ T cells contribute to immune competence in a way that is distinct from $\alpha\beta$ T cells. In three cases, investigators found that neither peptides bound to foreign agents nor peptides derived from them are recognized by $\gamma\delta$ T-cell clones. Instead, protein antigens are recognized by $\gamma\delta$ T cells directly without any requirement for antigen processing (Schild et al. 1994; Sciammas et al. 1994; Weintraub et al. 1994). The $\gamma\delta$ T cells can also recognize a variety of nonpeptide antigens in addition to peptide antigens (Poccia et al. 1998). Furthermore, a comparison with antibody and $\alpha\beta$ TCR V domains reveals that the three-dimensional framework structure of Vd more closely resembles that of V_H than of Va, Vb, or V_L (Li et al. 1998). These results suggest that $\gamma\delta$ T-cell receptors may be more like immunoglobulins in their recognition properties (Chien et al. 1996; Marchalonis et al. 1997). Therefore, the low frequency of $\gamma\delta$ T cells in humans and mice may be due to the fact that the Ig molecules in humans and mice are highly diverged, and may have undertaken part of the role played by $\gamma\delta$ T cells in the other species.

In conclusion, the phylogenetic analysis of *TCR-V* genes from the six species reveals different degrees of diversity of gene repertoire, similar to those seen for *IGV* genes. This pattern of TCR diversity must be related to the diversity of other immune system genes and prompts further investigation into different mammalian and avian immune systems.

Acknowledgments The authors thank Igor Rogozin for his helpful comments. This work has been supported by grants from NIH and NSF to M. N.

References

Arden B, Clark SP, Kabelitz D, Mak TW (1995a) Human T-cell receptor variable gene segment families. Immunogenetics 42:455–500

Arden B, Clark SP, Kabelitz D, Mak TW (1995b) Mouse T-cell receptor variable gene segment families. Immunogenetics 42:501–530

Caccia N, Bruns GA, Kirsch IR, Hollis GF, Bertness V, Mak TW (1985) T cell receptor α chain genes are located on chromosome 14 at 14q11–14q12 in humans. J Exp Med 161:1255–1260

Chen CH, Six A, Kubota T, Tsuji S, Kong FK, Gobel TW, Cooper MD (1996) T cell receptors and T cell development. Curr Top Microbiol Immunol 212:37–53

Chien YH, Jores R, Crowley MP (1996) Recognition by $\gamma\delta$ T cells. Annu Rev Immunol 14:511–532

Chothia C, Boswell DR, Lesk AM (1988) The outline structure of the T-cell $\alpha\beta$ receptor. Embo J 7:3745–3755

Clark SP, Arden B, Kabelitz D, Mak TW (1995) Comparison of human and mouse T-cell receptor variable gene segment subfamilies. Immunogenetics 42:531–540

Davis MM, Bjorkman PJ (1988) T-cell antigen receptor genes and T-cell recognition. Nature 334:395–402

Easteal S (1988) Rate constancy of globin gene evolution in placental mammals. Proc Natl Acad Sci U S A 85:7622–7626

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Gobel TW, Chen CL, Lahti J, Kubota T, Kuo CL, Aebersold R, Hood L, Cooper MD (1994) Identification of T-cell receptor a-chain genes in the chicken. Proc Natl Acad Sci USA 91:1094–1098

Grossberger D, Marcuz A, Fichtel A, Dudler L, Hein WR (1993) Sequence analysis of sheep T-cell receptor β chains. Immunogenetics 37:222–226

Hein WR (1994) Structural and functional evolution of the extracellular regions of T cell receptors. Semin Immunol 6:361–372

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

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

Hein WR, Mackay CR (1991) Prominence of $\gamma\delta$ T cells in the ruminant immune system. Immunol Today 12:30–34

Hein WR, Marcuz A, Fichtel A, Dudler L, Grossberger D (1991) Primary structure of the sheep T-cell receptor a chain. Immunogenetics 34:39–41

Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol 42:182–192

Hong SC, Chelouche A, Lin RH, Shaywitz D, Braunstein NS, Glimcher L, Janeway CA, Jr. (1992) An MHC interaction site maps to the amino-terminal half of the T cell receptor a chain variable domain. Cell 69:999–1009

Ikuta K, Ogura T, Shimizu A, Honjo T (1985) Low frequency of somatic mutation in b-chain variable region genes of human T-cell receptors. Proc Natl Acad Sci U S A 82:7701–7705

Ishiguro N, Tanaka A, Shinagawa M (1990) Sequence analysis of bovine T-cell receptor a chain. Immunogenetics 31:57–60

Isono T, Isegawa Y, Seto A (1994) Sequence and diversity of variable gene segments coding for rabbit T-cell receptor β chains. Immunogenetics 39:243–248

Isono T, Kim CJ, Seto A (1995) Sequence and diversity of rabbit T-cell receptor g chain genes. Immunogenetics 41:295–300

Kim CJ, Isono T, Tomoyoshi T, Seto A (1995) Variable-region sequences for T-cell receptor-g and -d chains of rabbit killer cell lines against Shope carcinoma cells. Cancer Lett 89:37–44

Klein J, Hořejši V (1997) Immunology. Blackwell Science Ltd., Tokyo, Japan

Knight KL (1992) Restricted V_H gene usage and generation of antibody diversity in rabbit. Annu Rev Immunol 10:593–616

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis. The Pennsylvania State University, University Park, PA, USA

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

Li WH, Gouy M, Sharp PM, O'HUigin C, Yang YW (1990) Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. Proc Natl Acad Sci U S A 87:6703–6707

Litman GW, Rast JP (1996) The organization and structure of immunoglobulin and T-cell receptor genes in the most phylogenetically distant jawed vertebrates: evolutionary implications. Res Immunol 147:226–233

Marchalonis JJ, Schluter SF, Edmundson AB (1997) The T-cell receptor as immunoglobulin: paradigm regained. Proc Soc Exp Biol Med 216:303–318

Marche PN, Kindt TJ (1986) Two distinct T-cell receptor a-chain transcripts in a rabbit T-cell line: implications for allelic exclusion in T cells. Proc Natl Acad Sci USA 83:2190–2194

Massari S, Antonacci R, De Caro F, Lipsi MR, Ciccarese S (1997) Assignment of the TCRA/TCRD locus to sheep chromosome bands 7q1.4– >q2.2 by fluorescence in situ hybridization. Cytogenet Cell Genet 79:193–195

Novacek MJ (1992) Mammalian phylogeny: shaking the tree. Nature 356:121–125

Ota T, Nei M (1995) Evolution of immunoglobulin V_H pseudogenes in chickens. Mol Biol Evol 12:94–102

Poccia F, Gougeon ML, Bonneville M, Lopez-Botet M, Moretta A, Battistini L, Wallace M, Colizzi V, Malkovsky M (1998) Innate T-cell immunity to nonpeptidic antigens. Immunol Today 19:253–256

Rast JP, Haire RN, Litman RT, Pross S, Litman GW (1995) Identification and characterization of T-cell antigen receptor-related genes in phylogenetically diverse vertebrate species. Immunogenetics 42:204–212

Rast JP, Anderson MK, Strong SJ, Luer C, Litman RT, Litman GW (1997) α, β, γ, and δ T cell antigen receptor genes arose early in vertebrate phylogeny. Immunity 6:1–11

Reynaud CA, Dahan A, Anquez V, Weill JC (1989) Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell 59:171–183

Rowen L, Koop BF, Hood L (1996) The complete 685-kilobase DNA sequence of the human β T cell receptor locus. Science 272:1755–1762

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

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

Sawasdikosol S, Hague BF, Zhao TM, Bowers FS, Simpson RM, Robinson M, Kindt TJ (1993) Selection of rabbit CD4- CD8- T cell receptor-$\gamma\delta$ cells by in vitro transformation with human T lymphotropic virus-I. J Exp Med 178:1337–1345

Schild H, Mavaddat N, Litzenberger C, Ehrich EW, Davis MM, Bluestone JA, Matis L, Draper RK, Chien YH (1994) The nature of major histocompatibility complex recognition by $\gamma\delta$ T cells. Cell 76:29–37

Sciammas R, Johnson RM, Sperling AI, Brady W, Linsley PS, Spear PG, Fitch FW, Bluestone JA (1994) Unique antigen recognition by a herpesvirus-specific TCR-$\gamma\delta$ cell. J Immunol 152:5392–5397

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

Sitnikova T, Su C (1998) Coevolution of immunoglobulin heavy and light chain variable region gene families. Mol Biol Evol 15:617–625

Sitnikova T, Rzhetsky A, Nei M (1995) Interior-branch and bootstrap tests of phylogenetic trees. Mol Biol Evol 12:319–333

Six A, Rast JP, McCormack WT, Dunon D, Courtois D, Li Y, Chen CH, Cooper MD (1996) Characterization of avian T-cell receptor gamma genes. Proc Natl Acad Sci USA 93:15329–15334

Solinas-Toldo S, Lengauer C, Fries R (1995) Comparative genome map of human and cattle. Genomics 27:489–496

Swofford DL (1998) PAUP. Phylogeneetic analysis using parsimony. Sinauer Associates, Sunderland, Massachusetts

Tanaka A, Ishiguro N, Shinagawa M (1990) Sequence and diversity of bovine T-cell receptor b-chain genes. Immunogenetics 32:263–271

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

Tjoelker LW, Carlson LM, Lee K, Lahti J, McCormack WT, Leiden JM, Chen CL, Cooper MD, Thompson CB (1990) Evolutionary conservation of antigen recognition: the chicken T-cell receptor β chain. Proc Natl Acad Sci USA 87:7856–7860

Wang K, Klotz JL, Kiser G, Bristol G, Hays E, Lai E, Gese E, Kronenberg M, Hood L (1994) Organization of the V gene segments in mouse T-cell antigen receptor a/d locus. Genomics 20:419–428

Weintraub BC, Jackson MR, Hedrick SM (1994) $\gamma\delta$ T cells can recognize nonclassical MHC in the absence of conventional antigenic peptides. J Immunol 153:3051–3058

WHO-IUIS Nomenclature Sub-Committee on TCR Designation (1995) Nomenclature for T-cell receptor (TCR) gene segments of the immune system. Immunogenetics 42:451–453

Zharkikh A, Li W-H (1992) Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences: I. Four taxa with a molecular clock. Mol Biol Evol 9:1119–1147