J Mol Evol (1993) 37:260–272

Journal of Molecular Evolution

The Paradox of *MHC-DRB* Exon/Intron Evolution: α-helix and β-sheet Encoding Regions Diverge While Hypervariable Intronic Simple Repeats Coevolve with β-sheet Codons

Franz-Werner Schwaiger,[1] Eva Weyers,[1] Cornelia Epplen,[1] Jörg Brün,[1] Gabi Ruff,[2] Allan Crawford,[3] Jörg T. Epplen[1]

[1] Department of Molecular Human Genetics, Ruhr-Universität Bochum, Postfach 10 21 48, W 4630 Bochum 1, Germany
[2] Department of Animal Genetics, Bremgarten Str. 109a, University of Berne, 3012 Berne, Switzerland
[3] Department of Biochemistry, University of Otago, P.O. Box 56, Dunedin, New Zealand

Received: 6 February 1993/Accepted: 1 March 1993

Abstract. Twenty-one different caprine and 13 ovine *MHC-DRB* exon 2 sequences were determined including part of the adjacent introns containing simple repetitive $(gt)_n(ga)_m$ elements. The positions for highly polymorphic *DRB* amino acids vary slightly among ungulates and other mammals. From man and mouse to ungulates the basic $(gt)_n(ga)_m$ structure is fixed in evolution for 7×10^7 years whereas ample variations exist in the tandem $(gt)_n$ and $(ga)_m$ dinucleotides and especially their "degenerated" derivatives. Phylogenetic trees for the α-helices and β-pleated sheets of the ungulate *DRB* sequences suggest different evolutionary histories. In hoofed animals as well as in humans *DRB* β-sheet encoding sequences and adjacent intronic repeats can be assembled into virtually identical groups suggesting coevolution of noncoding as well as coding DNA. In contrast α-helices and C-terminal parts of the first *DRB* domain evolve distinctly. In the absence of a defined mechanism causing specific, site-directed mutations, double-recombination or gene-conversion-like events would readily explain this fact. The role of the intronic simple $(gt)_n(ga)_m$ repeat is discussed with respect to these genetic exchange mechanisms during evolution.

Key words: MHC class II genes — Polymorphism — Gene conversion — Intron preservation — Artiodactyls

Correspondence to: J.T. Epplen

*M*ajor *h*istocompatibility *c*omplex (*MHC*) class I and class II genes encode heterodimeric cell-surface molecules which bind foreign antigen and present it to T lymphocytes. The efficiency of immune responses depends critically on the appropriate presentation of antigen (Klein 1987). Recently characteristic sequence features were identified in peptides eluted from *MHC* binding sites (Falk et al. 1991), suggesting different affinities for various antigens (Madden et al. 1991). Many *MHC* class I and class II molecules conserve an extremely high degree of polymorphism (Kappes and Strominger 1988, Bodmer et al. 1990, Marsh and Bodmer 1993). For example, nearly 80 sequences have been identified from one to nine gene loci for the *HLA-DRB*-encoded cell-surface proteins in man (Marsh and Bodmer 1993). Almost all of the variable amino acid (aa) residues of the heterodimeric *DRA/DRB* molecule are concentrated in the first domain of the β-chain encoded by the second *DRB* exon. The degree of heterozygosity at the *DRB1* locus is higher than expected and very few alleles represent rare variants in the population (Kappes and Strominger 1988). Studies on the generation and maintenance of sequence differences have led to various models concerning *MHC* evolution in primates (Gyllenstein et al. 1991a,b; Kasahara et al. 1990) and rodents (Figueroa et al. 1990; Wakeland et al. 1990). Klein (1980, 1987) formulated the *transspecies theory* assuming that many newly evolving species commence with a whole group of *MHC* alleles, hence,

conserving polymorphism of their progenitors. The persistence of high heterozygosity rates at *MHC* loci is explained by "overdominant selection" mechanisms (Hughes and Nei 1989). In addition, the exchange of nucleotide sequences appears to generate new combinations of "old polymorphisms" in *HLA* class I (Belich et al. 1992, Watkins et al. 1992) and class II genes (Kuhner and Peterson 1992).

For several theoretical and practical reasons also the *MHC* class II genes of some artiodactyl species have been studied recently (Anderson et al. 1991; Scott et al. 1991, Ammer et al. 1992). In addition to exonic polymorphism we have previously shown that a simple repetitive element of the basic structure $(gt)_n(ga)_m$ displays hypervariability in the second intron of bovine *DRB* genes. There were also hints that the simple repeats coevolve with the exonic sequences (Ammer et al. 1992). Such an effect had been proposed even earlier for the complete human *DRB* exons 2 and adjacent intronic repeats (Rieß et al. 1990). Here these studies are expanded to cover goats, sheep, and other selected ungulate species in order to reveal the extent of coevolution in *MHC* exon and intron sequences. The mixed $(gt)_n(ga)_m$ simple repeat has been preserved for more than 7×10^7 years from primates to artiodactyls. Thus the question arises as to whether the expressed exonic sequences exert any influence on the adjacent simple repeats or vice versa.

Materials and Methods

DNAs and Amplification. Genomic DNA was obtained from artiodactyl peripheral blood leukocytes using a modified salting out procedure according to Miller et al. (1988). All DNAs were characterized by oligonucleotide fingerprinting to ensure individuality and genetic relationships (Schwaiger et al. unpublished data). The 5′-primer for PCR has been described previously (Ammer et al. 1992); the 3′-primer "GIo" was constructed on the basis of bovine and caprine intronic sequences (5′-CGTACCCAGA$^{T}/_{G}$-TGAGTGAAGTATC-3′). Amplification was performed with approximately 1 μg of DNA for 30 cycles with 30 s denaturation at 94°C, 1-min primer annealing at 59°C, and 1-min extension at 72°C. For amplification of cDNA the primers "Z32h-o" (5′-AGATACTTCCATAATGGAGAA-3′) and "HDR2o" (5′-CCATTCCACTGTGAGAGG-3′) have been used to provide biased amplification of distinct *DRB* loci. Thus, parts of the exon 2 and 3 (corresponding to aa 36–176) were synthesized during PCR. As a consequence, DNA contaminations would have been identified immediately due to the presence of intron 2. Amplification products were separated on 2% agarose gels. The gels were dried and hybridized with the ^{32}P-labeled internal oligonucleotide specific for *DRB* exon 2 region "HDBo" as described before (Ammer et al. 1992). The PCR products of all investigated animals have been separated on sequencing gels, electroblotted, and hybridized with various site-specific oligonucleotides according to Rieß et al. (1990).

RNA Isolation, cDNA Synthesis, and Molecular Cloning. RNA isolation was performed following the protocol of Sambrook et al. (1989). cDNA was synthesized with 30 μg of total RNA using the Pharmacia cDNA synthesis kit (Freiburg, Germany). All PCR products were made blunt-ended and cloned into a pUC19 or pBlueskript vector. Both strands have been sequenced for each clone. For all other molecular biological techniques previously described protocols have been followed (Ammer et al. 1992).

Sequences and Phylogenetic Analysis Methods. Human *DRB* exon and intron sequences (*DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0404, DRB1*0405, DRB1*0407, DRB1*0801, DRB1*1001, DRB1*1103, DRB1*1301, DRB1*1401, DRB1*1501, DRB4*0101, DRB5*0101*) have been compiled by Rieß et al. (1990). Mouse *I-Eβ* sequences of various haplotypes [d (Saito et al. 1983), u (Ayane et al. 1986), w17 (Vu et al. 1988), 1 (Cam et al. 1990), k (Gorski et al. 1990), non (Acha-Orbea and Scarpellino 1991), b and s (Mengle-Gaw et al. 1984, 1985), z (Ogawa et al. 1990), f and q (Begovich et al. 1990)] and a rat *RT1* sequence (*RT1-DU*, Chao et al. 1989) were included for analysis. Nucleic acid and protein sequences of bovine *Bota-DRB01-22* were taken from Ammer et al. (1992). In addition 11 *Bota-DRB* protein sequences (Anderson et al. 1991) were studied for protein relationships.

For maximum parsimony analyses and construction of phylogenetic trees several program packages were applied: PHYLIP (DNAPARS and PROTPARS; version 3.4; Felsenstein 1988), VOSTORG (Zharkikh et al. 1991), and SPLIT (Bandelt and Dress 1989). Protein and nucleic acid sequences were analyzed by the neighbor-joining, furthest-neighbor, unique substitution, maximum parsimony method of the DNAPARS, PROTPARS, or VOSTORG program. Analysis of relative silent and relative replacement substitutions was performed by the NAG program (version 2; Nei and Gobori 1986). Exonic sequences are identical in length and thus aligned. Bootstraps were performed for SPLIT and PHYLIP in order to evaluate the trees. The program VOSTORG does not include a bootstrap function. Hence, the input order of the sequences was changed 10 times to exclude artifacts. Homology was computed with the program BESTFIT of the program package HUSAR (German Cancer Research Center, Heidelberg).

Most of the phylogenetic analysis methods used here are based on distance matrices. The distance of taxons (here DNA or protein sequences) is calculated as percent base difference. Those two taxons (a, b) with the smallest distance are joined to form a new taxon (c). a and b are then replaced within the distance matrix by the new artificial taxon c. The new distances for c to all other taxons are calculated subsequently. Many methods differ only in the way of computing the new distances. (For detailed information see Sneath and Sokal 1973.)

The furthest-neighbor method follows the principles described above. The new distance for c to another taxon i is calculated by substituting the distances of the original taxons (a, b) with the distance of the most remote taxon—the "furthest" taxon from i. Using the neighbor-joining method the branch lengths for every pair of taxons are calculated on the basis of a distance matrix. Pairs of taxons are identified that minimize the sum of the calculated branch lengths at each stage of the taxons' clustering (Saitou and Nei 1987). The DNAPARS/PROTPARS programs create unrooted trees on the basis of maximum parsimony on a set of taxons. The numbers of nucleic acid changes on a given tree are counted. Trees which have the smallest number of changes (mutation) are supposed to represent the best model for evolution. The unique substitution method is also based on the principle of maximum parsimony. In a preliminary sequence alignment analysis all sites are removed that represent one of the

following situations: (1) No changes between all sequences; (2) only one change in one sequence; (3) two different changes of the same nucleotide (i.e., G → T and G → A) in two different sequences; (3) three different nucleotide exchanges (i.e., G → T, G → A and G → C) in three taxons. Using the resulting alignment a distance matrix is calculated. A second matrix is computed based on the quantities of common (supposed to be "ancestral") nucleotides. This matrix is used to identify the optimal pair of sequences to be joined. The tree is then built by comparing both matrices.

In the SPLIT analysis (Bandelt and Dress 1983) of a given similarity matrix of a set of taxa, a subset is regarded as a cluster when two objects from this subset share higher similarity in comparison to third objects from outside. These clusters can then be described in a hypergraph without triangles. In this graph one and the same taxon may be found in more than one cluster, indicating uncertain relationships. Such a situation can be interpreted as gene-conversion-like events or parallel mutations. All the analysis methods applied are based on the assumption that individual residues evolve independently.

Results and Discussion

Ungulate MHC-DRB Polymorphisms Encoded by Exon 2

A panel of 25 goats of six different breeds (*Capra aegagrus hircus*), 15 sheep of 4 different breeds (*Ovis aries), 1 mountain goat (Oreamnos americanus*), 1 gazelle (*Gazella dorcas*), and 1 giraffe (*Giraffa giraffa*) have been investigated for *MHC-DRB* exon 2 plus intron sequences by PCR amplification. In total, 37 novel different *DRB* exons 2 plus introns have been sequenced (21 caprine, 13 ovine, 1 mountain goat, 1 gazelle, and 1 giraffe, Figs. 1, 2).

All the deduced goat and sheep *DRB* exon 2 protein sequences were analyzed on the basis of the predicted class II antigen-binding groove structure (Brown et al. 1988) and compared to those of cattle, man, and mouse. Highly polymorphic residues are found mainly at the same aa positions as in cattle (11, 37, 57, 71, and 86). In positions 28, 30, and 74, which are also highly variable in cattle, man, and mouse, only two different aa were observed in goats. Positions 13, 28, 30, and 70 were less variable in sheep. In principle this apparently lower degree of polymorphism in some aa positions could be due to a biased amplification efficiency of certain *DRB* loci. Yet the extensive polymorphism at some *Ovar-DRB* positions (11, 71, 86) contradicts such an objection. In position 78, which is expected to point toward the antigen or even make contact to the peptide, four different aa residues were found in goats (2 each in sheep and cattle). Thus the additional variation of two aa residues in position 78 seems to have emerged after the separation into the precursors of present-day sheep and goats. Whereas only Gly and Val are present in position 86 in man, four different aa have been detected in cattle and sheep and five in goat.

Detailed analyses of the cattle, sheep, and goat *DRB* sequences showed the codon usage to be equivalent to the general codon usage described for these species (Wada et al. 1991). Polymorphic codons show an increased number of pyrimidines in the third position in all three species (Table 1). T nucleotides are found more frequently in the third position while G's are reduced in comparison to the codon usage of the complete exon 2. In addition, we analyzed the codon usage under the two following assumptions (Table 2): (1) In a given nucleotide position individual mutations occur only rarely and they are transmitted to other alleles or loci via recombination or gene-conversion-like event. With this hypothesis in mind the observed codons have been compared as singular units (appearing only once in evolution) and were consequently counted only once regardless of how often this codon was found at a given aa position. (2) All mutations happen independently in each sequence. Hence the total number of base exchanges in relation to the consensus sequence has been counted. Probably both extreme views are not completely justified. At highly polymorphic sites certain codons are accumulated in which the third position is predominantly occupied by one particular base: e.g., at position 11, the codon NNT is present in almost all codons throughout five different species. In four of the five codons at the highly variable aa position 86 or caprine and ovine *DRB*, the rarely observed NNT codons (Wada et al. 1991) are realized exclusively. For some polymorphic positions rather rare codons remained stable in the species of Bovidae. In contrast, identical codons are virtually missing in the identical aa residues in position 13 or 86 when one compares the orders of artiodactyls, rodents, and primates. Some codons are only present in the Bovidae or even restricted to the Caprinae (aa positions 11, 13, 21, 30, 32, 59, 66, 73, 86, 89) and could hence be interpreted as polymorphism which originated after the separation of artiodactyle or the Bovidae, respectively. Furthermore, not only single codons but complete patterns of nucleotide substitutions were conserved. Several of these substitutions suggest the group-specific organization of exonic sequences. This fact and the persistence of very similar *DRB* exon 2 sequences (*Caae-DRB12/13* and *Bota-DRB21/22;* Ammer et al. 1992) can be taken as evidence that polymorphisms survive speciation events for several million years (ancient polymorphisms).

Diversifying Selection

If *MHC* alleles are selectively neutral (Klein 1987) the relative number of synonymous substitutions

```
Amino acid #  11         21         31         41         51         61         71         81         91   #ind./breed
HLA-DRB ConsEQ SKSECHFFNG TERVRFLDRY FYNQEEYVRF DSDVGEYRAV TELGRPDAEY WNSQKDLLEQ RRAAVDTYCR HNYGVGESFT VQRR

Caae-DRB01  -y y-r------- ---------- -h-g--f--- ---w--f--- ---------- ------f--s --t------- y--------- ----   2 / z,a
Caae-DRB02  -y y-r------- ---------- yt-g--tl-- ---w--f--- ---------- ------f--s --t------- y--------- ----   3 / z,e
Caae-DRB04  -y c-r----s-- ----gl---- -h-g--i--- ---w--f--- ---------- -----ei--s --t------- -----v---- ----   2 / s,w

Caae-DRB05  -y h--------- ----wy---- ---g------ -n-w------ a-------k- -----ei--- --t----v-- ---------- ----   2 / s

Caae-DRB06  -y h--------- ---------- ---g------ -n-w--f--- a----qt-k- -----e---r --te---f-- -----f---a ----   1 / a
Caae-DRB07  -y h--------- ---------- ---g------ -n-w--f--- a-----s-k- -----e---r --te---f-- -----f---a ----   2 / s
Caae-DRB08  -y y-g------- ---------- ---g------ -n-w--f--- a-----s-k- -----e---r --te---f-- -----f---- ----   2 / s

Caae-DRB09  -y y-g------- ---------- ---g------ -n-w--f--- a----qt-k- ------f--d s------f-- -----i---a ----   4 / s
Caae-DRB10  -y y-g------- ---------- ---g------ -n-w--f--- a-------k- ------f--d s------f-- -----i---- ----   1 / e

Caae-DRB11  -y --k----s-- -----y---- ---g-----y ---r--f--- a-----e--- -----ei--- k--ea--v-- -----i---s ----   4 / s
Caae-DRB12  -y --k----s-- ---------- -h-g--f--- ---w------ a-------k- -----ei--r k--------- y--------- ----   1 / s
Caae-DRB13  -y t-k----s-- ---------- -h-g--f-l- ---w------ a-------k- -----ei--r k--------- ---------- ----   1 / s
Caae-DRB14  -y t-k----s-- ---------- ---g-----y ---w--f--- a--------- -----ei--d s--------- ---------- ----   1 / s
Caae-DRB15  -y t-k--r-s-- ---------- ---g-----y ---w--f--- a--------- -----ei--d s--------- ---------- ----   1 / s
Caae-DRB16  -y --k--r-s-- ---------- ---g-----y ---w--f--- a--------- -----ei--d s--------- ---------- ----   1 / s
Caae-DRB17  -y t-k--r-s-- ---------- ---g-----y ---w--f--- a--------- -----ei--- k--e---c-- -----i---s ----   1 / s
Caae-DRB18  -y -t-----s-- ---------- ---g-----y ---w--f--- a-----s-k- -----ei--d s--------- -----i---s ----   1 / e
Caae-DRB19  -y -t-----s-- -----l---- ---g--tl-y ---w------ a--------- -----e---- --te------ ---------- ----   1 / s
Caae-DRB20  -y -t-----s-- ---------- ---g-----y ---w------ a--------- -----e---- --te------ ---------- ----   1 / s
Caae-DRB21  -y -t-----s-- ---------- ---g-----y ---w------ a-----s-k- -----e---- --te------ -----i---s ----   1 / w

Caae-DRB22  -y y-g----s-- -g---l-h-f yt-g--nl-- ---w--f--- --q-qe---- ------f--- k--e---v-- -----l---- ----   4 / s

HLA-DRB ConsEQ SKSECHFFNG TERVRFLDRY FYNQEEYVRF DSDVGEYRAV TELGRPDAEY WNSQKDLLEQ RRAAVDTYCR HNYGVGESFT VQRR

Ovar-DRB01  -y t-k--r-s-- ---------- -h-g--tl-- ---w------ a-------k- ------f--r a--------- -----i---- ----   8 / P, R
Ovar-DRB02  -y t-k--r-s-- ---------- -h-g--tl-- ---w------ a-------k- ------f--r a--------- ---------s ----   2 / P, T
Ovar-DRB03  -y trk--r-s-- ---------- -h-g--tl-- ---w------ a-------k- ------f--r a--------- ---e-i-g-- ----   1 / P*

Ovar-DRBDH  -y t-k--r-s-- ---------- ---g------ ---w------ a-------k- -----ei--r --te------ -----i---s ----   1 / S **

Ovar-DRBFM  -y t-k--r-s-- ---------- ---g---a-- ---w------ a-------k- -----ei--r --te------ ---------- ----   1 /  ****

Ovar-DRB04  -y -t-------- ---------- ---g--tl-- ---w------ a-------k- ------f--- t-t------- -----i---s ----   1 / R

Ovar-DRB05  -y a----r---- --------e-- ---g------ ---w------ a----rs--- ------f--r k--n------ ---------- ----   1 / P
Ovar-DRB06  -y a----r---- --------e-- ---g--tl-- ---w------ a-------k- -----e---r k--n------ -----d---- ----   1 / P
Ovar-DRB07  -y a----r---- --------e-- ---g------ ---w------ a-------k- ------f--r k--n------ -----f---s ----   1 / M

Ovar-DRB08  -y h--------- ---------- -h-g--f--- ---w--f--- a----rs--- -----e---r --te------ -----f---s ----   3 / P, R
Ovar-DRB09  -y h--------- ---------- ---g------ ---w--f--- a--------h -----e---r ---e------ -----i---s ----   1 / P
Ovar-DRB10  -y h--------- ---------- ---g------ ---w--f--- a----qs--h -----e---r ---e------ -----i---s ----   2 / M, T
Ovar-DRB11  -y h----r-s-- -----y---- ---g------ -n-w------ a-------k- -----ei--r k--------- -----i---s ----   1 / M

Ovar-DRB12  -y r----r-s-- -----y---- ---g------ ---w------ a-------k- -----e---r k--n------ -----i---s ----   1 / M

Ovar-DRB13  -y h----s---- ---------- ---g------ ---w--f--- a-----e-k- -----ei--r k--------- -----i---s ----   1 / R


Oram-DRB03  -y y--------- -------g-- yt-g--l--- ---w--f--- a--------h ------f--q d---a-af-- ---r-v---- ----   1 / -


Gigi-DRB01  -y atg----s-- -q-----e-- --sg---a-- ---w--f--- -------t-- ---------r k--n------ -----i---- ----   1 / -


Gado-DRB01  -y y-g------- -------e-- -h-g--f--- ---w------ h-----a-kq -----ki--d s------f-- ---------- ----   1 / -


cDNA-clones

Exon II
Bota-A1     -y ---------- ---------- yt-g--t--- ---w--f--- ----pq-r-- ------f--e k--e--rv-- ----gm---- ----   ***

Caae-DRB10  -y y-g------- ---------- ---g------ -n-w--f--- a-------k- ------f--d s------f-- -----i---- ----   1 / e

Exon III             101        111        121        131        141        151        161        171
HLA-DRB1         VHPKVT VYPSKTQPLQ HHNLLVCSVS GFYPGSIEVR WFRNGQEEKT GVVSTGLIHN GDWTFQTLVM LETVPRSGEV YTCQVE

Bota-DRB(A1)     -e-t-- ---a------ ----------n -----h---- ----ah--ea --i-----q- -------m-- -----q---- ------   ***

Caae-DRB10       -a-t-- ---a------ ----------n -----h---- -----h--ea --i-----q- -------m-- -----q---- ------   1 / e

Ovar-DRBDH       -e-i-- ---a------ ----------n -----h---- -----h--ea --i-----q- -------m-- -----q---- -----d
Ovar-DRBFM       -e-i-- ---a------ ----------n -----h---- -----h--ea --i-----q- -------m-- -----q---- -----d   1 / ****

*Only one clone investigated per individual
Abbreviations (goat breeds) s: "Saanen"; a: "Angora"; b: "Buren"; e: "Bunte Edel"; w: "Weiße Edel"; z: "African dwarf"
            (sheep breeds) P: "Perendale"; R: "Romney"; M: "Merino sheep"; S: "Suffolk cross"; T: "Texel sheep"
** Sequence from EMBL databank [accession number EMNEW:OAMHCDRB2; Ballingal et al.]
*** Sequence from Groenen et al. (1989)
**** Sequence from EMBL databank [accession number: M73984; Fabb et al., unpublished.]
In one goat (e: "Bunte Edel") a preliminary study revealed so far that the Caae-DRB10 allele is expressed in cDNA
obtained from total cellular mRNA.
```

Fig. 1. Translated exon 2 sequences of ungulate *DRB* sequences compared to consensus *HLA-DRB*. After PCR at least two independent clones have been sequenced on both strands for each individual. The PCR amplification was controlled by sequence gel electrophoresis, electroblotting, and hybridization using polymorphism-specific oligonucleotide probes (Schwaiger et al. 1993) in order to prove the identity of the sequences.

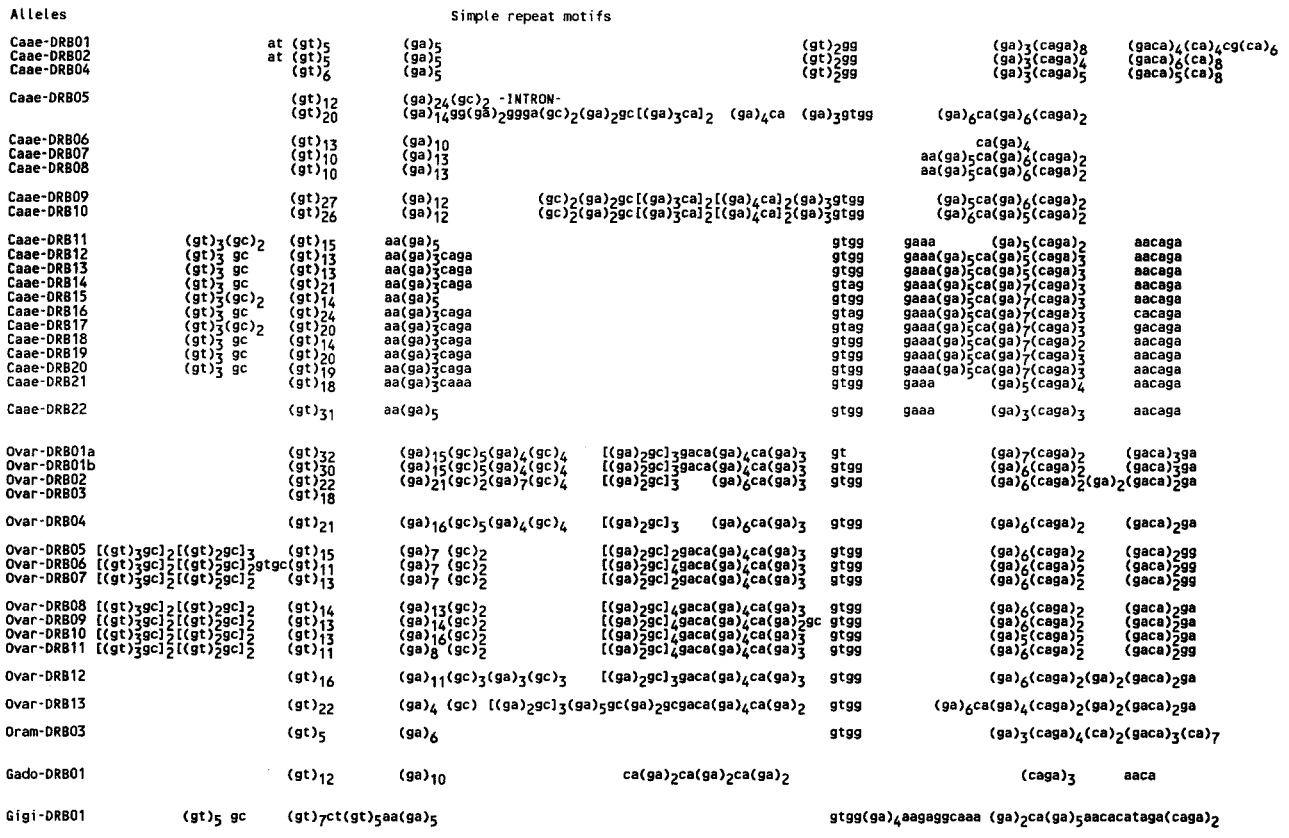


Fig. 2. Intronic simple repeat sequences in introns 2 of ungulate *DRB* genes.

Table 1. Frequencies of pyrimidine (Y) and purine (R) residues in the third position of polymorphic codons[a]

	Goat	Cattle	Man
Y (%)	75.0/86.7	69.0/80.0	75.9/87.5
R (%)	25.0/13.3	31.0/20.0	24.1/12.5

[a] Polymorphic codons of amino acids which are encoded by two triplet bases/amino acids which are encoded by three or four triplets

Table 2. Frequencies of codons harboring the nucleotides T, A, C, or G in the third position of DRB exon 2 sequences[a]

	Goat	Cattle	Man
Total exon 2			
Number of codons	155/1,445	148/1,870	151/1,190
T (%)	17.4/10.7	16.2/11.7	17.2/10.3
A (%)	4.5/6.5	5.4/7.8	6.0/3.9
C (%)	39.4/38.8	40.5/38.2	37.1/34.7
G (%)	38.7/44.0	37.8/42.9	39.7/51.2
Polymorphic residues only			
Number of codons	41/153	50/198	37/154
T (%)	**34.1/23.5**	24.0/17.7	**37.8/23.3**
A (%)	4.8/3.9	4.0/2.0	2.7/1.3
C (%)	46.3/54.2	46.0/52.5	40.5/42.9
G (%)	**14.6/18.3**	**26.0/27.8**	**18.9/32.5**

[a] The first decimal figure (%) was computed with each of the different observed codon versions counted only once for all DRB exons, whereas the second percentage value represents all observed codons for each position evaluated separately in the different DRB exons. Bold numbers represent significant changes in third codons positions of polymorphic residues.

should be equal to that of nonsynonymous substitutions. Maintenance of polymorphism in the antigen binding region, however, is influenced by positive selection (Potts and Wakeland 1990). Hughes and Nei (1989) demonstrated diversifying selection for aa substitutions in the antigen binding site of class II molecules in mice and man. In the antigen binding sites of the first domain in artiodactyl *DRB* sequences the quotient of relative silent to relative replacement substitutions is significantly smaller than 1 ($P < 0.001$). Surprisingly, this is also true for the complete *DRB* exons 2 of goat and sheep ($P < 0.001$), but not for those of cattle.

The overdominant selection hypothesis assumes that there is an unusually high degree of polymorphism and an unusually long persistence of polymorphic alleles in the population. In more than 85% of our animals two different sequences were amplified. Extensive family studies indicate that the two amplified *DRB* sequences were inherited independently according to Mendelian principles (data not shown). Thus the polymorphism seems to have originated in the progenitors of Bovidae (Ammer et al. 1992) or Caprinae (Fig. 1, Table 3), and it is

Table 3. Putative double-recombination or gene-conversion-like events as documented in the resulting exon 2 sequences of goat and sheep DRB genes[a]

Position aa		86				90			93			Simple repeat type
Caae-DRB04	–	G– –	– – –	– – –	– – –	– – –	– – –	– – –	– –C	– – –	–	$(gt)_n(ga)_m1$
Caae-DRB06	–	T– –	– – –	– – –	– – –	G– –	– – –	– – –	– – –	– – –	–	$(gt)_n(ga)_m2$
Caae-DRB07	–	T– –	– – –	– – –	– – –	G– –	– – –	– – –	– –C	– – –	–	$(gt)_n(ga)_m2$
Caae-DRB09	–	– – –	– – –	– – –	– – –	G– –	– – –	– – –	– –C	– – –	–	$(gt)_n(ga)_m3$
Caae-DRB10	–	– – –	– – –	– – –	– – –	– – –	– – –	– – –	– –C	– – –	–	$(gt)_n(ga)_m3$
Caae-DRB15	–	GG–	– – –	– – –	– – –	– – –	– – –	– – –	– –C	– – –	–	$(gt)_n(ga)_m4$
Caae-DRB16	–	GG–	– – –	– – –	– – –	– – –	– – –	– – –	– – –	– – –	–	$(gt)_n(ga)_m4$
Caae-DRB20	–	– – –	– – –	– – –	– – –	–G–	– – –	– – –	– – –	– – –	–	
Ovar-DRB01	–	– – –	– – –	– – –	– – –	– – –	– – –	– – –	– – –	– – –	–	$(gt)_n(ga)_ma$
Ovar-DRB08	–	– – –	– – –	– – –	– – –	–G–	– – –	– – –	– – –	– – –	–	$(gt)_n(ga)_mc$
Ovar-DRB02	–	GG–	– – –	– – –	– – –	–G–	– – –	– – –	– – –	– – –	–	$(gt)_n(ga)_ma$
Ovar-DRB04	–	GG–	– – –	– – –	– – –	– – –	– – –	– – –	– – –	– – –	–	$(gt)_n(ga)_mb$
Ovar-DRB06	–	T– –	– – –	– – –	– – –	–GT	– – –	– – –	– – –	– – –	–	$(gt)_n(ga)_mb$
Caae-DRB08	–	T– –	– – –	– – –	– – –	– – –	– – –	– – –	– – –	– – –	–	

Position aa	β_1-region 11	12	13	/	18	19	20	21	/	32	/	β_2-region 37	38	39	40
Caae-DRB01	–AC	– – –	– –A	/	– – –	–T	– – –	– – –	/	CA–	/	TT–	– – –	– – –	– – –
Caae-DRB02	–AC	– – –	– –A	/	– – –	–T	– – –	– – –	/	– – –	/	– – –	– – –	– – –	– – –
Caae-DRB15	A– –	– – –	–AA	/	–C–	– – –	– – –	–G	/	CA–	/	TT–	– – –	– – –	– – –
Caae-DRB20	AA–	– – –	–AA	/	–C–	– – –	– – –	–G	/	TA–	/	TA–	– – –	– – –	–A–
Caae-DRB12	– – –	–C–	– – –	/	–C–	– – –	– – –	–G	/	TA–	/	– – –	– – –	– – –	–A–
Caae-DRB13	– – –	–C–	– – –	/	–C–	– – –	– – –	–G	/	TA–	/	TA–	– – –	– – –	–A–
Caae-DRB06	CA–	– – –	– – –	/	– – –	– – –	– – –	– – –	/	TA–	/	TA–	– – –	– – –	– – –
Ovar-DRB06	CA–	– – –	– – –	/	– – –	– – –	– – –	–G	/	CA–	/	TT–	– – –	– – –	– – –

The nucleotides G [in the third codon position (aa #21) which is only found in goat and sheep] and the C [in aa position 93 which is restricted to goat] represent silent mutations. The specific type of simple repeat is characterized by numbers or letters

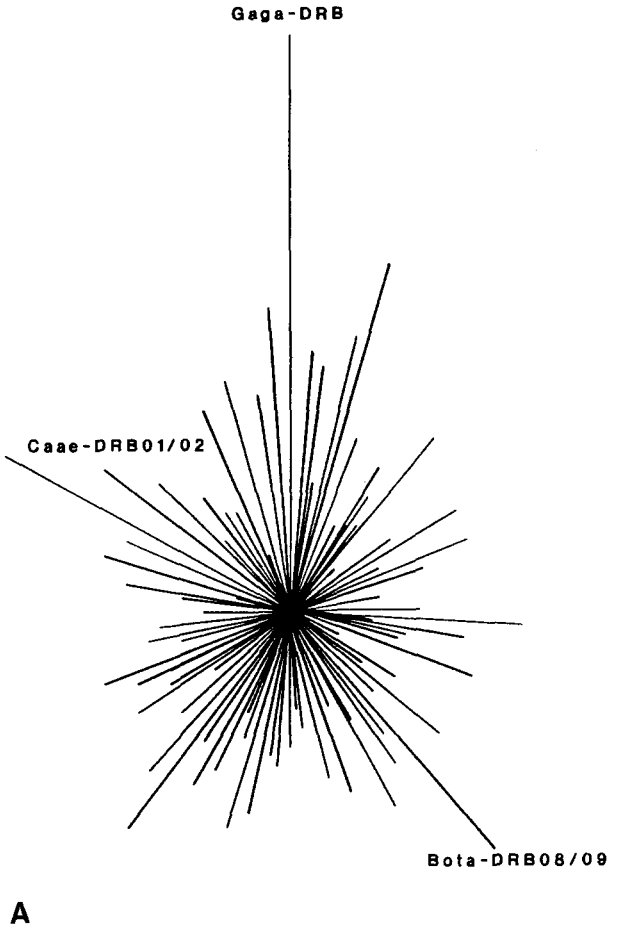
maintained in these species as an ancient polymorphism. Hence, overdominant selection seems to exert its influence also on cattle, goat, and sheep *DRB* sequences.

Phylogenetic Trees of MHC-DRB Sequences and Their Subregions

Recently it has been noticed that subregions of the *MHC* class II β1-domain subregions (β-pleated sheets and α-helices and C-terminal end) develop differently in the evolution of primates (Gyllenstein et al. 1991a,b), mouse (Wakeland et al. 1990), and cattle (Sigurdardottir et al. 1992). Here phylogenetic trees have been constructed by various methods including the three species of the Bovidae as well as mouse and human sequences—as representatives for two other mammalian orders.

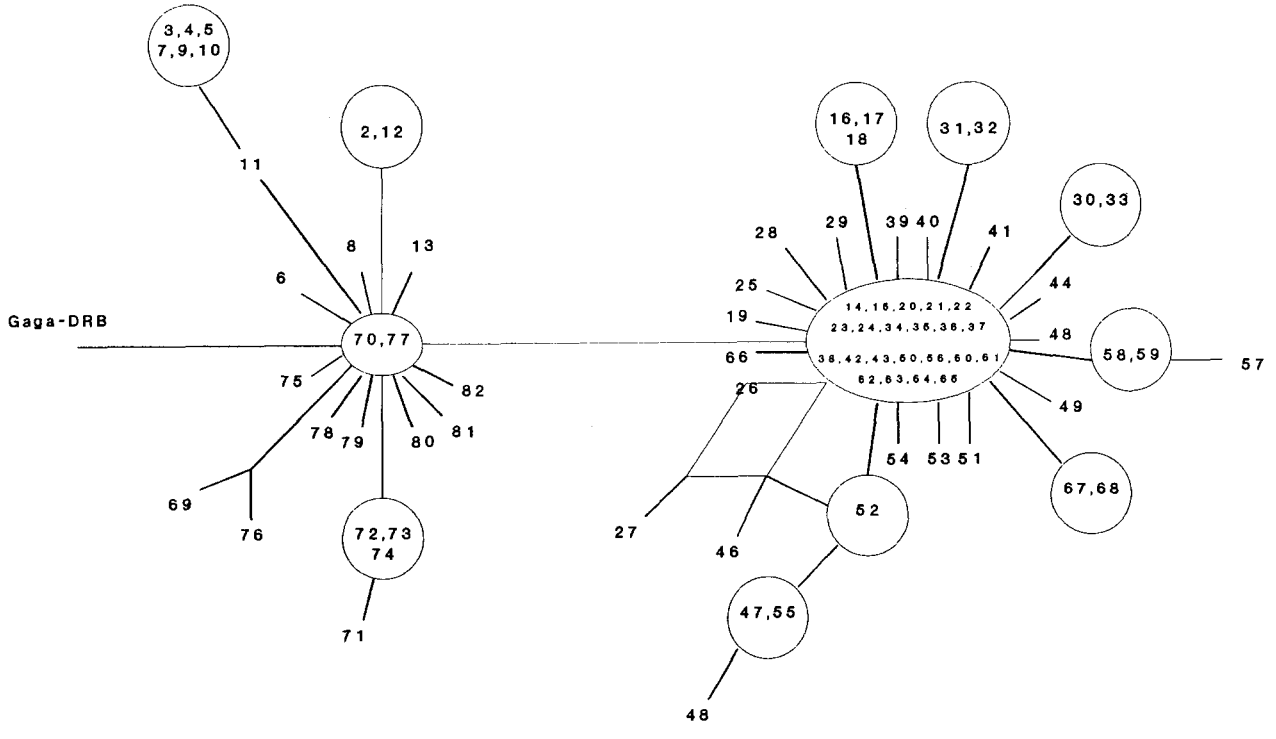
Using the SPLIT approach, the complete exon 2 sequences of the three different animal orders are separated clearly from each other, indicating monophyletic relationships. Ungulate sequences are grouped together. Interestingly, on the peptide level the mouse *DRB* orthologue represents an outgroup and the human counterparts diverge together with the three artiodactyl species. In addition, this analysis revealed that α-helices are randomized extraordinarily for both nucleotide (Fig. 3A) and protein sequences. This means that a high number of parallel mutations and recombination or gene-conversion-like events may have happened in this region of *DRB* exon 2. β-sheet regions are randomized only at the protein level, suggesting restricted variability of the polymorphic protein positions (data not shown). The DNA sequence trees of the β-sheet regions diverge into the three orders but the individual *DRB* sequences do not form a tree defining phylogenetic relationships (Fig. 3B). Bootstrapping confirmed the validity if the graphs obtained (data not shown).

To confirm these results with conventional methods, phylogenetic trees were constructed with the neighbor-joining, furthest-neighbor, unique substitution method of VOSTORG and the maximum parsimony method of DNAPARS and PROTPARS from PHYLIP. In the evolution of *DRB* exon 2 complex mechanisms may have included parallel



A

Fig. 3. Phylogenetic relationships of *DRB* nucleotide sequences. The definition of the α-helix (positions 160–234) and β-sheet encoding portions (13–157) follows Gyllenstein and Erlich (1991); the β-sheet subregions (β1: 13–47; β2: 48–89; β3: 90–157) were divided according to Wakeland et al. (1990). The C-terminal exonic region included positions 235–268. Bootstrapping with DNAPARS confirmed the appearance of the clusters in the phylogenetic trees of α-helices, β-sheets, and C-terminal regions. **A, B** Graphs representing the SPLIT phylogenetic analysis including α-helices **(A)** and β-sheets **(B)** from members of three mammalian orders. Distances are not indicated by length of the branches. All α-helix branches **(A)** originate from one center with various distances as a split of a single sequence from the rest. Splits of two sequences vs the remaining ones are mentioned explicitly. In the β-sheet graph the artiodactyls *DRB* sequences are separated from the other two orders. The relationship of individual *DRB* sequences within the orders remains unclear. The *Gaga-DRB* revealed the highest isolation index for both analyses. 2: *Mumu-DRB* U, 3: *Mumu-DRB* w17, 4: *Mumu-DRB* B, 5: *Mumu-DRB* F, 6: *Mumu-DRB* I, 7: *Mumu-DRB* K, 8: *Mumu-DRB* NON, 9: *Mumu-DRB* M12, 10: *Mumu-DRB* Q, 11: *Mumu-DRB* S, 12: *Mumu-DRB* Z, 13: *Rano-DRB* RT1.DU; 14/15: *Ovar-DRB01/02*, 16–25: *Ovar-DRB04-13;* 26/27: *Caae-DRB01/02*, 28–46: *Caae-DRB01-22;* 47–68: *Bota-DRB01-22;* 69: HLA-DRB1*0101, 70: HLA-DRB1*0301, 71: *HLA-DRB1*0401*, 72: *HLA-DRB1*0404*, 73: *HLA-DRB1*0405*, 74: *HLA-DRB1*0407*, 75: *HLA-DRB1*0801*, 76: *HLA-DRB1*1001*, 77: *HLA-DRB1*1103*, 78: *HLA-DRB1*1301*, 79: *HLA-DRB1*1401*, 80: *HLA-DRB1*1501*, 81: *HLA-DRB4*0101*, 82: *HLA-DRB5*0101*. Continued on next page.



B

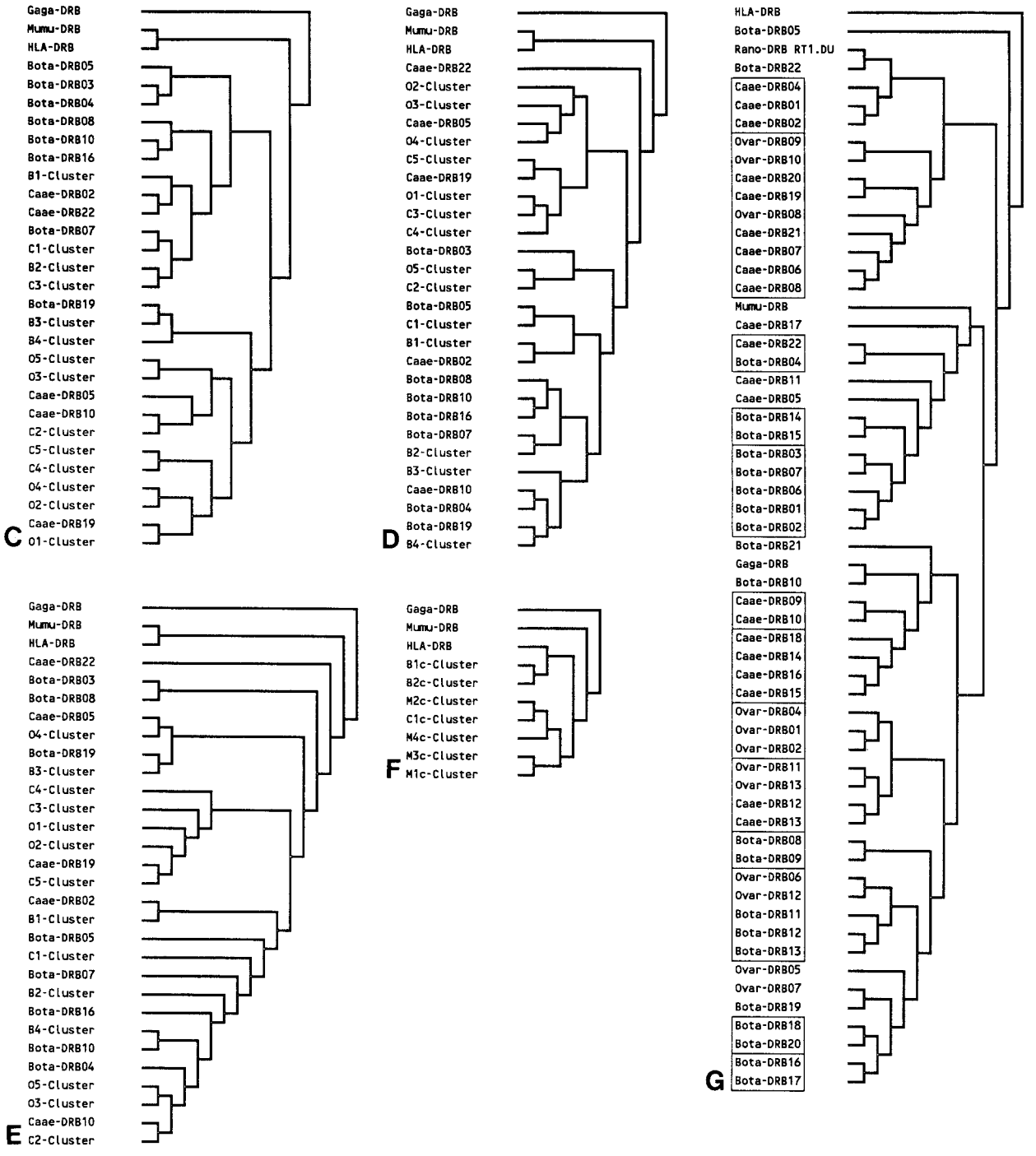


Fig. 3. Continued. **C–E** Example for unrooted phylogenetic trees of *DRB* β-sheet nucleotide sequences as obtained with the neighbor-joining **(C)**, furthest-neighbor **(D)**, and unique substitution method **(E)** of VOSTORG. Distances are not drawn to scale of evolutionary time spans. Several clusters defining relationship were generated with all three methods. C1-cluster: *Caae-DRB01/04;* C2 cluster: *Caae-DRB07-09;* C3 cluster: *Caae-DRB12/13;* C4 cluster: *Caae-DRB11/14–17;* C5 cluster: *Caae-DRB18/20/21;* O1 cluster: *Ovar-DRB01/02;* O2 cluster: *Ovar-DRB05–07;* O3 cluster: *Ovar-DRB09/13;* O4 cluster: *Ovar-DRB11/12;* O5 cluster *Ovar-DRB08/10;* B1 cluster: *Bota-DRB01/02/06/09;* B2 cluster: *Bota-DRB21/22;* B3 cluster: *Bota-DRB11-13;* B4 cluster: *Bota-DRB14/15/17/18/20.* **F** Unrooted phylogenetic tree of the *DRB* exon 2 C-terminal nucleotide sequence deduced from trees obtained with the neighbor-joining, furthest-neighbor, and unique substitution methods of VOSTORG. For some *DRB* sequences no clear relationship with one of the clusters was found when comparing the three methods (*Bota-DRB13, Ovar-DRB01/02, Caae-DRB04/22*), all other sequences appeared always in the same clusters. M1 cluster: *Bota-DRB06/10/14–17, Ovar-DRB05, Caae-DRB01/02/05/13/14/15/16/19/20*); M2 cluster: *Ovar-DRB04/07–13, Caae-DRB11/17/18/21;* M3 cluster: *Bota-DRB21, Ovar-DRB06, Caae-DRB04/12;* M4 cluster: *Bota-DRB11, Caae-DRB08;* C1 cluster: *Caae-DRB06/07/09/10*); B1 cluster: *Bota-DRB01-03/08/09/12;* B2 cluster *Bota-DRB04/05/07/18–20.* **G** One of four most parsimonious unrooted trees for the α-helical region computed by the neighbor-joining method of VOSTORG. Some clusters suggest closer relationships among different animal orders than within the same order. Boxes indicating clusters of sequences found in trees computed by different methods.

mutations and repeated recombination and/or gene-conversion-like events (Gorski and Mach 1986). Hence trees obtained via such analyses may not reveal real phylogenetic relationships—a fact that is underscored by bootstrapping—for example, with the DNAPARS program (data not shown). As expected, the various methods resulted in trees with different topology for the exon 2 and for the complete β-sheet encoding sequences, especially for comparing taxons within the same genus. Particular clusters of closely related sequences (for example, Fig. 3C–E) emerged with all analytical methods, suggesting a monophyletic relationship. The other *DRB* sequences vary in their relationships in trees even with the same method of analysis. This fact reflects the uncertainty in defining any direction relations for most of the *DRB* exon 2 sequences. Combining the results of all analysis programs, the clusters of the *DRB* α-helices were mixed (on the nucleotide and the protein level); i.e., they consisted of sequences from different animal orders (for one example see Fig. 3G). On the other hand, the three mammalian orders formed separate clusters in all trees of the C-termini, β-pleated sheets, and the three subregions of the β-sheets (Fig. 3C,F). But with respect to the β1-, β2-, or β3-regions, the various methods produced several equally probable trees, differing only at their terminal twigs. The latter trees appear more meaningful in comparison to the trees representing the complete exon 2 or the total β-sheets. Trees of these regions and those of the α-helical or C-terminal portions revealed relationships of different *DRB*s on the protein and the nucleotide level, even within one species of the Bovidae (Fig. 4). Similar data exist for the mouse class II $I\text{-}A^{\beta}$ locus (Wakeland et al. 1990), which is the most variable class II surface protein, but not for the *DRB* orthologue. This fact may be taken as evidence for evolutionary exchanges of motifs between the *DRB* sequences.

In the case of concerted evolution of *DRB* sequences a clear phylogenetic relationship can *not* be established if a high number of recombination and/or gene-conversion-like events or parallel mutations act on this locus during evolution. Hence we searched for traces of such exchange events and their distribution throughout the ungulate *DRB* sequences (Table 3). Short nucleic acid motifs (even those with silent mutations or unusual species-specific codons) are evolutionarily stable and are distributed throughout many *DRB* sequences of goats, sheep, and cattle. Such patterns exhibit almost the same codon usage even in the interspecies comparison. In addition they are often flanked by sequences varying in different *DRB* sequences on both sides. This is true also for the intraspecies comparison. For example, a specific codon representing a silent mutation at position 93 (Arg: CGG → CGC; invariably found in five *Caae-DRB* sequences) is combined with four different aa in position 86 and also with four distinguishable simple repeat structures. As no selective pressure can be assumed for this silent exchange, parallel mutations in five different *DRB* sequences appear as an insufficient explanation for this fact. Ser and Ala residues in position 90 are also found in variable combinations with 5′ and 3′ adjacent nucleotide patterns. Hence cassettes of once-established motifs are found with variable upstream and downstream sequence environments (Wakeland et al. 1990). This phenomenon is not restricted to particular exonic regions. Instead "motif sharing" appears to be demonstrable all over exon 2 with no detectable, defined breakpoint for putative recombinations or conversions. Exchanges between the sequences are somewhat accumulated toward the 3′-situated simple repeat.

Our findings regarding the evolution of *DRB* first domain α-helices and β-pleated sheets are consistent with those of other reports (Sigurdardottir et al. 1992; Erlich and Gyllenstein 1991; Wakeland et al. 1990): Defined β-sheets are combined with various α-helices within the species of cattle, goat, and sheep. The phylogenetic analysis using various methods revealed different tree topologies for the five subregions of the putative antigen binding domain, suggesting sequence exchanges. The topology of trees for such short sequences can easily be disturbed by homoplasy (e.g., convergent evolution). As a consequence the real relationships may be substantially different. Therefore we compared selectively neutral nucleic acid positions and codon usage in the characteristic sequence motifs (Table 3). Silent mutations or such mutations that do not change the physicochemical character of the protein in a usually nonpolymorphic position are found in several different sequences of obviously polyphyletic origin. The probability is extremely low that silent mutations like those at aa positions 21 and 93 would have happened independently in three or five different alleles, respectively. Thus recombinational and/or gene conversion events are quite likely. Recombination would not occur at defined positions (e.g., between β-sheet and α-helix; Wakeland et al. 1990; Gyllenstein et al. 1991a,b) as there is also evidence for exchanges of the 3′-end of exon 2 with respect to the α-helices and the adjacent intronic repeat, as well as for exchanges within the β-pleated sheet (Table 3).

DRB loci encode a high number of alleles and most individuals are heterozygous. Certain polymorphic alleles are transmitted during the speciation process and could then be mixed via recombination and/or gene-conversion-like events.

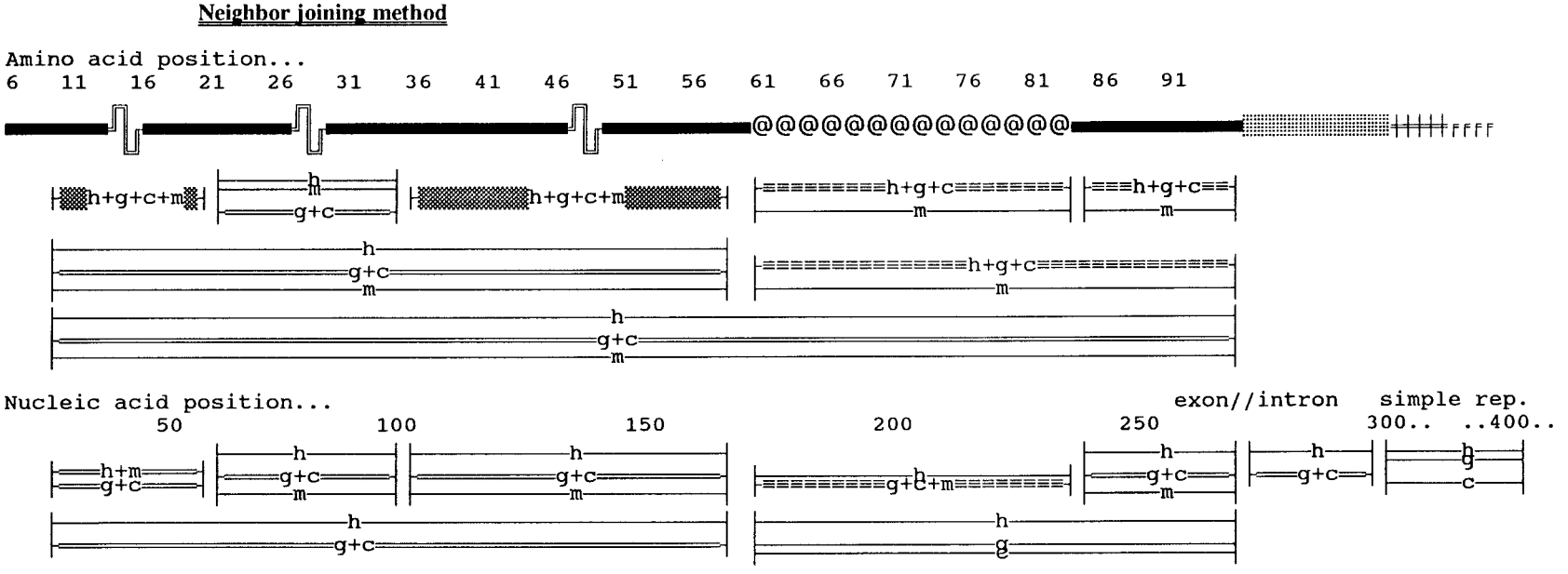


Fig. 4. Evaluation of phylogenetic trees obtained from *MHC-DRB* exon/intron sequences from ungulates (*g* = goat, *c* = cattle), mouse (*m*), and man (*h*) by the neighbor-joining method. ~~~ indicates complete mixing of the sequences of *h*, *m*, *g*, and *c*. = = = = and ═══ mean that the sequences of three or two species are not separable. The more distant the lines for two species are depicted, the fewer overlaps in the trees exist. The symbols used are explained at the bottom of the figure.

Recombination would exert a strong diversifying influence on such a locus. Taking into account the high degree of *DRB* heterozygosity nearly every recombination event should create a new allele. This effect is multiplied when the break points vary. New species-specific mutations can accumulate during the course of evolution. Different modes of positive selection (Wakeland et al. 1990) could then fix the new sequence combination or/and mutations in the population. In addition to the examples where a conversion is the most likely explanation (Mengle-Gaw et al. 1984; Gorski and Mach 1986) we identified one likely case of an unequal crossing over or deletion/insertion event in the intronic sequences (*Caae-DRB05*, Fig. 2).

Analysis of Simple Repeat Elements and Their Preservation During Evolution

In all second *DRB* introns, a simple repeat element is situated about 35–50 bases downstream of the exon/intron boundary in artiodactyls (Ammer et al. 1992), man (Rieß et al. 1990), and mouse (Braunstein and Germain 1986), respectively. The repeat length ranges from a mean of 74 base pairs (bp) in man (±12 SD among 32 examples; Rieß et al. 1990), to 97 (±13 SD) in cattle, 102 (±27 SD) in goats, and up to 161 (±19 SD) in sheep. So far we have not found stretches shorter than 44 bp in 113 *DRB* sequences of 16 species out of 4 families in 3 mammalian orders (Schwaiger, unpublished data).

The simple repeat sequences are divided into two main groups: (1) the perfect simple $(gt)_n$ repeat stretch and its derivatives and (2) the perfect simple $(ga)_m$ repeat and derivatives. The $(gt)_n$ stretch tends to evolve more conservatively (mainly transition mutations), comprising 12 (±3.9; cattle) to 20 (±3.7; man) copies of the dinucleotide motif. On the other hand, $(ga)_m$ gives rise to many derivatives and a much greater variability of the perfect repeats. In the $(ga)_m$ stretches the bulk of point mutations are transversions. Length variations of the entire repeat are mainly caused by the derivatives of the $(ga)_m$ dinucleotides (Fig. 2 and data not shown).

Usually simple repeats are hypervariable in length and quite unstable on an evolutionary time scale (Epplen et al. 1991). In general they seem to vary in the numbers of established patterns, in this case dinucleotides or derived quadruplets. Depending on the copy number of perfect repeat motifs the rate of length changes is several orders of magnitude higher than the usual mutation rate. This is reflected in the *DRB* intronic repeat since in the interspecies comparison of bovines and caprines no two identical repeats exist. In general the simple repeats show clearly distinguishable derivatives in an interspecies comparison of cattle, goat, and sheep. Surprisingly simple repeats of very similar structure surfaced within the subfamily Bovinae and even in both subfamilies of Caprinae (*C. aegagrus*) and Rupicaprinae (*O. americanus*). The closely related *Bos taurus* and *B. indicus* (2×10^6 years of evolutionary distance; Legel 1989) exhibited comparable repeats and these were associated with similar β-sheet encoding regions. The exonic sequence of *Bison bison* is not similar to *Bota-DRB04* or *Bota-DRB05* (about 5×10^6 years difference; Legel 1989) although the repeats are of the same basic structure. But the *Oram-DRB03* exon 2 shares an almost identical repeat with *Caae-DRB01*, *-DRB02*, or *-DRB04*. *O. americanus* and *C. aegagrus* are supposed to have separated more than 10^7 years ago (Thenius 1979). Here the exonic sequences of goat and mountain goat have diverged completely (Fig. 1). Simple repeats can thus be stable in their basic structure for a time period of more than 10^7 years, but they are neighboring different exonic sequences in another species.

Group-Specific Organization and the Paradox of Exon/Intron Evolution

As shown in Fig. 2, a *group-specific organization* of the simple repeats in goat and sheep is obvious, as it is for man and cattle. This group-specific organization reflects the categorization of exonic sequences, in particular the β-sheet classification (β1-region, Fig. 1; Ammer et al. 1992). Exonic sequences within a group of similar repeat structure reveal usually higher similarity to each other than to any member of all other groups within a species (BESTFIT analysis; data not shown). The constant simple repeat structure in a distinct group of *DRB* sequences within a species is still surprising. How can the occurrence of similar repeats adjacent to *DRB* exonic sequences be correlated most closely with the β1-region? As we have shown, recombination or gene-conversion-like events are obviously occurring in the second exon of *DRB* sequences. If we assume single recombination events and no selective pressure for certain repeat/exon combinations then there should be no linkage of repeats to particular β-sheets. Evidence for broken linkage within one species was found only as an exception (*Caae-DRB08*). Various explanations for coevolution of β1-encoding regions and simple repeats are possible: (1) High mutation rates in both exonic β-sheet and/or intronic sequences or parasite-driven selection result in convergent evolution. Yet there is no evidence for higher mutation rates in *DRB* exons 2 (for a review see Kasahara et al. 1990) or selection based on the presence of certain pathogens (Potts and Wakeland 1990). The specific structure of the simple repeat in intron 2 appears to depend on multiple point mutations and local amplification and reamplification events. Hence the

independent creation of closely related derivative structures in different alleles appears inconceivable. (2) Exchanges of nucleotide stretches occur between the 3′-end of the β-sheet encoding region and the intronic simple repeat. Consequently exchanges of sequences from allele to allele or locus to locus should be double-crossover or gene-conversion-like events whereby any subregion of exon 2 can theoretically be transmitted. These events are apparently more frequent at the 3′-end of exon 2 excluding the adjacent intronic simple repeat. In another instance polypurine and $(gt)_n$ elements have also been supposed to be hot spots for initiating or terminating conversion events in the evolution of γ-globin genes (Fitch et al. 1990).

Reduced Exon and Simple Repeat Polymorphism in Sheep

In each of the species cattle, goat, and sheep similar β1-regions show group-specific characteristics. The repeat structures coevolving with the β1-regions in these clusters are not closely related compared to those of different species (Fig. 1—for example, *Ovar-DRB08/09* and *Caae-DRB06/07*). It is very difficult to envision how the exchange of the basic simple repeat structure occurred consistently in all members of a group after the separation of the species. For goats and cattle at least four distinct classes of repeat groups were defined (Fig. 2, Ammer et al. 1992). Interestingly the repeats of the sheep *DRB* sequences appear to be more closely related to each other than to any of those of goats or cattle. Hence they may have descended from a single simple repeat structure that appears closely related to that of *Caae-DRB09/Caae-DRB10*. In addition sheep *DRB* exons 2 are less polymorphic in aa positions 13 and 70 in comparison to those of other ungulates. Nevertheless phylogenetic analysis revealed that some of the sheep exon 2 sequences are more homologous to goat than to sheep (see Fig. 1—e.g., *Ovar08/09/10* vs *Caae06/07*), supporting the transspecies theory. The extent of polymorphism could have been reduced in both exonic sequences and intronic simple repeats due to several bottlenecks. A founder population may have started with a few $(gt)_n(ga)_m$ repeats. Genetic drift and recombination spread this repeat in the pool of sheep *DRB* sequence. On the other hand, if there is directionality of recombination or gene-conversion-like events, new variations of ancient polymorphism could accumulate at a repeat locus and then be maintained by diversification and overdominant selection.

Conclusions and Hypothesis

By scrutinizing *MHC-DRB* introns in addition to expressed sequences new insights are gained for a deeper understanding of the evolutionary process. Also the development and nature of some simple repetitive genome components can be clarified. According to the data reported herein the simple repeat situated in all known *DRB* introns may have some biological meaning since it has been preserved in evolution. The *DRB* polymorphism appears to be generated in part by double-recombination or/and gene-conversion-like events resulting in new combinations of ancient polymorphism maintained by overdominant or diversifying selection. No specific recombination hot spot was identified in the second *DRB* exon. These events occur mainly 3′ to the β1-pleated sheet encoding region and in almost all cases they are restricted to the 5′-part of the composite simple $(gt)_n(ga)_m$ repeat. In this context it is noteworthy that certain nuclear proteins bind specifically to the intronic $(gt_n(ga)_m$ repeats (Mäueler et al. 1992). To clarify their influence on genetic exchanges requires diligent experimentation.

Acknowledgments. This work was supported by the DFG (Ep 7/6-2) and the Studienstiftung des Deutschen Volkes (F-WS). We thank E. Curio for correcting the English, H. Zoller for providing mainframe facilities, R. Wojcieszynski for advice in the Rechenzentrum (RUB), as well as A.W.M. Dress and R. Wetzel (Faculty of Mathematics, University of Bielefeld) for introducing us to the SPLIT program.

References

Acha-Orbea H, Scarpellino L (1991) Nonobese diabetic and nonobese nondiabetic mice have unique MHC class II haplotypes. Immunogenetics 34:57–59

Ammer H, Schwaiger F-W, Kammerbauer C, Gomolka M, Arriens A, Lazary S, Epplen JT (1992) Exonic polymorphism versus intronic simple repeat hypervariability in *MHC-DRB*-genes. Immunogenetics 32:330–337

Andersson L, Sigurdardottir S, Borsch C, Gustafsson K (1991) Evolution of *MHC* polymorphism: extensive sharing of polymorphic sequence motifs between human and bovine *DRB* alleles. Immunogenetics 33:188–193

Ayane M, Mengle-Gaw L, McDevitt HO, Benoist C, Mathis D (1986) E-alpha-u and E-beta-u chain association: Where lies the anomaly? J Immunol 137:948–951

Bandelt H-J, Dress AWM (1989) Weak hierarchies associated with similarity measures—an additive clustering technique. Bull Math Biol 51:133–166

Begovich AB, Vu TH, Jones PP (1990) Characterization of the molecular defects in the mouse E-beta-f and E-beta-q genes: Implications for the origin of MHC polymorphism. J Immunol 144:1957–1964

Belich MP, Madrigal JA, Hildebrand WH, Zemmour J, Williams RC, Luz R, Peztl-Erler ML, Parham P (1992) Unusual *HLA-B* alleles in two tribes of Brazilian Indians. Nature 357: 326–329

Bodmer JG, Marsh SGE, Albert E (1990) Nomenclature for factors of the HLA system. Immunol Today 11:3–10

Braunstein NS, Germain RN (1986) The mouse E-beta-2 gene: a class II *MHC* beta gene with limited intraspecies polymorphism and an unusual pattern of transcription. EMBO J 5: 2469–2476

Brown JH, Jardetzky T, Saper MA, Samraoui B, Bjorkman PJ, Wiley DC (1988) A hypothetical model of the foreign antigen

binding site of class II histocompatibility molecules. Nature 332:845–850

Cam P, Jouvin-Marche E, LeGuern C, Marche PN (1990) Structure of class II genes in wild mouse Mus saxicola: functional and evolutionary implications. Eur J Immunol 20:1337–1343

Chao NJ, Timmerman L, McDevitt HO, Jacob CO (1989) Molecular characterization of *MHC* class II antigens (beta-1 domains) in the bb diabetes-prone and -resistant rat. Immunogenetics 29:231–234

Epplen JT, Ammer H, Epplen C, Kammerbauer C, Mitreiter R, Roewer L, Schwaiger W, Steimle V, Zischler H, Albert E, Andreas A, Beyermann B, Meyer W, Buitkamp J, Nanda I, Schmid M, Nürnberg P, Pena SDJ, Pöche H, Sprecher W, Schartl M, Weising K, Yassouridis A (1991) DNA fingerprinting: Approaches and applications. Birkhäuser-Verlag, Basel

Erlich HA, Gyllenstein UB (1991) Shared epitopes among *HLA* class II alleles: gene conversion, common ancestry and balancing selection. Immunol Today 12:411–414

Falk K, Rötzschke O, Stevanovic S, Jung G, Rammensee H (1991) Allele specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351:290–296

Felsenstein J (1988) Phylogeny from molecular sequences: inference and reliability. Ann Rev Genet 22:521–565

Figueroa F, Gutknecht J, Tichy H, Klein J (1990) Class II *Mhc* genes in rodent evolution. Immunol Rev 113:207–226

Fitch DHA, Mainone C, Goodman M, Slightom JL (1990) Molecular history of gene conversion in the primate fetal τ-globin genes. J Biol Chem 265:781–793

Groenen MAM, van der Poel JJ, Dijkhof RJM, Giphart RJ (1990) The nucleotide sequence of bovine MHC class II *DQB* and *DRB* genes. Immunogenetics 31:37–44

Gorski J, Mach B (1986) Polymorphism of human Ia antigens: gene conversion between two DR β loci results in a new *HLA-D/DR* specifity. Nature 322:67–70

Gorski J, Hayes CE (1990) The I-J-disparate mouse strains B10.A(3R) and B10.A(5R) have identical I-E beta sequences, Immunogenetics 39:127–129

Gyllenstein U, Sundvall M, Ezcurra I, Erlich HA (1991a) Genetic diversity at class II *DRB* Loci of the primate *MHC*. J Immunol 146:4368–4376

Gyllenstein U, Sundvall M, Erlich HA (1991b). Allelic diversity is generated by intraexon sequence exchange at the *DRB1* locus of primates. Proc Natl Acad Sci USA 88:3686–3690

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962

Kasahara M, Klein D, Weimin F, Gutknecht J (1990) Evolution of the class II major histocompatibility complex alleles in higher primates. Immunol Rev 113:65–82.

Kappes D, Strominger JL (1988) Human class II major histocompatibility complex genes and proteins. Annu Rev Biochem 57:991–1028

Klein J (1980) Generation of diversity at *MHC* loci: implication for T cell receptor repertoires. In: Fougereau M, Dausset D (eds) Immunology 80. Academic Press, London, pp 239–253

Klein J (1987) Origin of the major histocompatibility complex polymorphism: the transspecies hypothesis. Hum Immunol 19:155–162

Kuhner MK, Peterson MJ (1992) Genetic exchange in the evolution of the human *MHC* class II loci. Tissue Antigens 39: 209–215

Legel S (1989) Nutztiere der Tropen und Subtropen (Legel S, (ed) Hirzel-Verlag, Leipzig

Madden DR, Gorga JC, Strominger JL, Wiley DC (1991) The structure of *HLA-B27* reveals nonamer self-peptides bound in an extended conformation. Nature 353:321–325

Marsh SGE, Bodmer J (1993) *HLA* class II nucleotide sequences, 1992. Immunogenetics 37:79–94

Mäueler W, Muller M, Köhne AC, Epplen JT (1992) A gel retardation assay system for studying protein binding to simple repetitive DNA sequences. Electrophoresis 13:7–10

Mengle-Gaw L, Conner S, McDevitt HO, Fathman CG (1984) Gene conversion between murine class II major histocompatibility complex loci. J Exp Med 160:1184–1194

Mengle-Gaw L, McDevitt HO (1985) Predicted protein sequence of the murine I-E-S-beta polypeptide chain from cDNA and genomic clones. Proc Natl Acad Sci USA 82:2910–2914

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

Nei M, Gobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Ogawa S, Nishimura H, Awaji M, Nozawa S, Hirose S, Shirai T (1990) Nucleotide sequence analysis of MHC Class II genes in autoimmune disease-prone (NZB X NZW)F1 mice. Immunogenetics 32:63–67

Potts WK, Wakeland EK (1990) Evolution of diversity at the major histocompatibility complex. Tree 5(6):181–186

Rieß O, Kammerbauer C, Roewer L, Steimle V, Andreas A, Albert E, Nagai T, Epplen JT (1990) Hypervariability of intronic simple $(gt)_n(ga)_m$ repeats in *HLA-DRB* genes. Immunogenetics 32:110–116

Saito H, Maki RA, Clayton LK, Tonegawa S (1983) Complete primary structures of the e-beta chain and gene of the mouse major histocompatibility complex. Proc Natl Acad Sci USA 80:5520–5524

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

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schwaiger FW, Buitkamp J, Weyers E, Epplen JT (1993) Typing of *MHC-DRB* genes with the help of intronic simple repeated DNA sequences. Mol Ecol 2:55–59

Scott PC, Maddox JF, Gogolin-Ewens KJ, Brandon MR (1991) The nucleotide sequence and evolution of ovine *MHC* class II *B* Genes: *DQB* and *DRB*. Immunogenetics 34:80–87

Sigurdardottir S, Borsch C, Gustafsson K, Andersson L (1992) Exon encoding the antigen-binding site of *MHC* class II beta-chains is divided into two subregions with different evolutionary histories. J Immunol 148:968–973

Sneath PHA, Sokal RR (1973) Numerical taxonomy. The principles and practice of numerical classification. Freeman and Co, San Francisco

Thenius E (1979) Die Evolution der Säugetiere. Gustav Fischer Verlag, Stuttgart, New York

Vu TH, Tacchini-Cottier FM, Day CE, Begovich AB, Jones PP (1988) Molecular basis for the defective expression of the mouse E-w17-Beta gene. J Immunol 141:3654–3661

Wada K, Wada Y, Doi H, Ishibashi F, Gojobori T, Ikemura T (1991) Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acid Res 19 Suppl:1981–1985

Wakeland EK, Boehme S, She JX (1990) The generation and maintenance of MHC class II gene polymorphism in rodents. Immunol Rev 113:39–46

Watkins DI, McAdam SN, Liu X, Strang CR, Milford EL, Levine CG, Garber TL, Dogon AL, Lord CI, Ghim SH, Troup GM, Hughes AL, Letvin NL (1992) New recombinant *HLA-B* alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. Nature 357:329–333

Zharkikh AA, Rzhetsky A, Morosov PS, Sitnikova TL, Krushkal JS (1991) VOSTORG: a package of microcomputer programs for sequence analysis and construction of phylogenetic trees. Gene 101:251–254