Polymorphism and selection in the major histocompatibility complex *DRA* and *DQA* genes in the family Equidae

Eva Janova[1], Jan Matiasovic[1], Jiri Vahala[2], Roman Vodicka[3], Enette Van Dyk[4] and Petr Horin[1]

(1) Institute of Animal Genetics, Faculty of Veterinary Medicine, University of Veterinary and Pharmaceutical Sciences Brno, Palackého 1/3, 612 42 Brno, Czech Republic

(2) Zoo Dvur Kralove n. L., Stefanikova 1029, 544 01 Dvur Kralove, Czech Republic

(3) Zoo Prague, U Trojskeho zamku 3/120, 171 00 Praha 7, Czech Republic

(4) Department of Companion Animal Clinical Studies, Faculty of Veterinary Science, Onderstepoort, University of Pretoria, Pretoria, Gauteng, South Africa

Abstract *The major histocompatibility complex genes coding for antigen binding and presenting molecules are the most polymorphic genes in the vertebrate genome. We studied the DRA and DQA gene polymorphism of the family Equidae. In addition to 11 previously reported DRA and 24 DQA alleles, six new DRA sequences and 13 new DQA alleles were identified in the genus Equus. Phylogenetic analysis of both DRA and DQA sequences provided evidence for trans-species polymorphism in the family Equidae. The phylogenetic trees differed from species relationships defined by standard taxonomy of Equidae and from trees based on mitochondrial or neutral gene sequence data. Analysis of selection showed differences between the less variable DRA and more variable DQA genes. DRA alleles were more often shared by more species. The DQA sequences analysed showed strong amongst-species positive selection; the selected amino acid positions mostly corresponded to selected positions in rodent and human DQA genes.*

Keywords Equus - MHC polymorphism - *DRA* and *DQA* genes - Positive selection

Introduction

Immune loci are suitable candidates for analysing effects of selection on the origins and mechanisms of maintenance of genetic diversity (Trowsdale and Parham *2004*). The major histocompatibility complex (MHC) consists of a cluster of closely linked genes playing a crucial role in the presentation of antigenic peptides to T lymphocytes (Klein *1986*). The genetic organisation of the MHC region can differ by species, suggesting rapid evolution of MHC genes (Kelley et al. *2005*). Various species differ in the numbers of specific MHC loci. The MHC genes are the most polymorphic genes in the vertebrate genome. Their high polymorphism seems to be maintained by balancing selection for millions of years, predating speciation events and reflecting the co-evolution of host with their pathogens and parasites (Bernatchez and Landry *2003*; Hedrick *1999*; Jeffery and Bangham *2000*).

Loss of genetic diversity in the MHC genes is associated with reduced ability to cope with new and changed diseases (Frankham *2003*). Analyses of within-species and interspecific MHC variation proved to be useful to detect selection on these genes in various mammalian groups (Cutrera and Lacey *2007*; Kennedy et al. *2007*; O'Brien and Yuhki *1999*; Seddon and Ellegren *2002*, *2004*). Trans-species transmission of MHC alleles was reported in cichlid bony fish (Ottova et al. *2005*), free-living rodents (Cutrera and Lacey *2007*; Seddon and

Baverstock *2000*), carnivores (Hedrick et al. *2000*; Seddon and Ellegren *2002*), ungulates (Hedrick and Kim *2000*) and primates (Otting et al. *2002*).

The family Equidae is a suitable object of genetic diversity and comparative genomic studies for several reasons. It is a rapidly evolving and variable group, at least at the karyotype level, with a relatively well-documented history of evolution (George and Ryder *1986*). The family is now composed of a single genus, *Equus* (Nowak *1999*), including domestic as well as free-ranging species living in different habitats. The process of domestication of horses was recently studied at the molecular level (Lindgren et al. *2004*). Over 6,000 years, the horse has differentiated into many often old and specific domesticated breeds and populations (Bowling and Rusinsky *2000*). The full genome sequence of the domestic horse has been published recently (http://0-www.ncbi.nlm.nih.gov.innopac.up.ac.za/sites/entrez?db=genomeprj&cmd=Retrieve&dopt=Overview&list_uids=11760).

The horse major histocompatibility complex (*ELA* or *Eqca)* is located on the horse chromosome (ECA) 20 (Bailey et al. *2000*). The equine and human MHCs that diverged less than 100 Mya (Kelley et al. *2005*) have a similar genomic organisation with contiguous class I, II and III regions (Gustafson et al. *2003*). *ELA* (*Eqca*) class II genes have been studied in the domestic horse as well as in some other members of the family. Whereas exon 2 *DRA* alleles generally exhibit if ever only small variation in mammalian species (e.g. Pimtanothai et al. *2001*; Russell et al. *1997*; Takada et al. *1998*; Yuhki et al. *2003*), extensive polymorphism of *DRA* genes has been reported in equids. In the domestic horse, four alleles have been found and a total of 11 alleles have been reported for the Equidae so far (Albright-Fraser et al. *1996*; Brown et al. *2004*; Luís et al. *2005*). High level of polymorphism of equid *DRB* and *DQB* genes was found (Fraser and Bailey *1996*; Horin and Matiasovic *2002*; Mashima *2003*; Szalai et al. *1994*). Two *DQA* genes within the *ELA/Eqca* region were originally identified (Gustafson et al. *2003*), and recently, the existence of three expressed *DQA* genes was reported (Miller and Antczak *2008*). A total number of 24 MHC *DQA* exon 2 alleles have been identified in the family Equidae (Fraser and Bailey *1998*). Whilst single nucleotide substitutions are mainly observed in *DRA* alleles, changes in consecutive nucleotides often occurred in the *DQA* genes. In both cases, these nucleotide variations result into amino acid changes in the antigen-binding site (ABS) of the molecule (Albright-Fraser et al. *1996*).

The purpose of this study was to analyse within-species and interspecies variation in exon 2 sequences of *DRA* and *DQA* genes related to their evolution and selection in captive and free-ranging members of the family Equidae.

Materials and methods

Animals

A total of 140 zebra specimens, i.e. peripheral blood samples from 13 damara zebra (*Equus burchelii antiquorum*, Smith 1841), 32 Boehm's zebras (*Equus burchelii boehmi*, Matschie 1892), 15 Chapman's zebras (*Equus burchelii chapmanni*, Layard 1865), 14 *Equus burchelii cunninghami*, Heller 1914, 39 Grevy's zebras (*Equus grevyi*, Oustalet 1882) and 27 Hartmann's mountain zebras (*Equus zebra hartmannae*, Matschie 1898), along with five asses *Equus asinus somaliensis*, Noack 1884, 12 kiangs (*Equus kiang*, Moorcroft 1841) and

six kulans (*Equus hemionus kulan*, Pallas 1775) originating from zoos in the Czech Republic were used for this study. In addition, 31 samples of Hartmann's mountain zebras and 17 Burchell's zebras (*Equus burchelii burchellii*, Gray 1824) originating from natural populations in the Republic of South Africa were available. The taxonomy commonly used for the family Equidae (Nowak *1999*) was respected. As herdbooks of captive zebras belonging to a specific subgroup called *E. b. cunninghami* are kept separately by zoological gardens, we also considered this group separate from *E. burchellii* subspecies.

Molecular analysis

Within-species and interspecies sequence variation at the *DRA* and *DQA* loci was examined. Exon 2 sequences were selected because they are known to be most subject to positive selection due to their involvement in the presentation of antigenic peptides (Militiadou et al. *2003*). Based on the assumption that MHC polymorphism can be expressed not only by true locus allelism, in all analyses of *DQA* genes, we considered all sequences as if they were alleles of one locus, similarly to Bryja et al. (*2006*).

DNA extraction, PCR and SSCP

Genomic DNA was extracted from peripheral blood using the NucleoSpin blood kit (Macherey-Nagel, Düren, Germany).

Exon 2 DRA amplification Primers Be3 and Be4 (Albright-Fraser et al. *1996*) were used for amplifying exon 2 of the horse *DRA* gene.
Be3 (5′-GCT TCT CAT CCT AGT TCC CTT)
Be4 (5′-GCC TAG GAG TGC AGC AGA)
The length of the PCR product was 307 bp. The polymerase chain reaction (PCR) mixture contained 4.15 mM $MgCl_2$, 90 μM dNTPs, 0.245 μM of each primer, 0.333 U HotStar Taq DNA polymerase (Quiagen, Valencia, CA, USA) and 0.5 μl genomic DNA in 15.5 μl total PCR volume. The amplification protocol consisted of 15 min at 95°C followed by 34 cycles of 30 s at 95°C, 45 s at 59°C and 30 s at 72°C. The final extension was done at 72°C (1 min).

Exon 2 DQA amplification Primers *DQA* 2e and *DQA* 2f (Fraser and Bailey *1998*) were used for amplifying exon 2 ELA *DQA* sequences.
DQA 2e (5′-CTG AIC ACI TTG CCT CCT ATG)
DQA 2f (5′-TGG TAG CAG CAG IAG IGT TG)
The PCR product was 247 bp long. The PCR mixture contained 3 M $MgCl_2$, 90 μM dNTPs, 0.55 μM of each primer, 0.5 U HotStar Taq DNA polymerase (Quiagen) and 0.5 μl genomic DNA in 25 μl total volume. The PCR reaction started at 95°C 15 min and included 39 cycles of 30 s at 95°C, 45 s at 52°C, 30 s at 72°C. The final extension was run at 72°C for 1 min. Both PCRs were run in the PTC-200DNA Engine Cycler (BIO-RAD).
The variation in *DRA* and *DQA* exon 2 sequences was pre-screened by single-strand conformation polymorphism (SSCP) analysis (Albright-Fraser et al. *1996*). Briefly, *DRA* and/or *DQA* PCR products were analysed on 13% and 10% acryl-amide gel, respectively, containing bis-acryl-amide (37.5:1), 0.5× TBE, ammonium persulfate and TEMED. Samples plus loading dye were heated at 95°C for 10 min then placed on ice until loaded. Gels were run in 0.5× TBE at 10 W for 18 h (*DRA*) or 20 h (*DQA*) at 10°C. SSCP patterns were visualised by standard silver staining. For maximising the numbers of alleles detected, all different SSCP patterns were sequenced. For this purpose, primarily homozygous (two-band pattern) samples were analysed. The PCR products were purified by the High Pure PCR

product purification kit (Roche, Basel, Switzerland) and sequenced. Patterns indicating genotypes heterozygous for a so far unidentified allele were sequenced as PCR products and the unknown allele sequence could be identified by manual editing and by using the Haplofinder software (Militiadou et al. *2003*). All allelic sequences used for further analysis were identified and confirmed based on at least two independent PCR amplifications.

Sequence analysis

All unequivocally identified sequences were aligned with the so far known sequences using the Bioedit sequence alignment editor (Hall *1999*) with the exception of a new *E. asinus* DRA allele (accession number FJ487912), submitted to GenBank only recently after we had completed the analysis. The domestic horse *Equus caballus DRA**0201 (Albright et al. *1991*), *DRA*(2)-1/1, ELA-*DRA*(2)-3/3 (Albright-Fraser et al. *1996*) and *DRA**JBH11 (Brown et al. *2004*) were included as well as the donkey *E. asinus* ELA-*DRA*(2)-1/5 (Albright-Fraser et al. *1996*), *DRA**0601 (accession number AF541938), ELA-*DRA**JBD3, *DRA**JBD17, ELA-*DRA**JBH45 (Brown et al. *2004*), ELA-*DRA*(2)-4/4 from *Equus hemionus* (Albright-Fraser et al. *1996*) and *DRA**JBZ185 allele reported from *E. burchellii boehmi* (Brown et al. *2004*). Similarly, the domestic horse ELA-*DQA**0101 (Szalai et al. *1994*), ELA-*DQA**0201 to ELA-*DQA**1501 (Fraser and Bailey *1998*), *DQA**1901 to *DQA**2401, accession number AF115324-AF115329 alleles were analysed along with ELA-*DQA**1601 found in *E. hemionus* and ELA-*DQA**1701 and ELA-*DQA**1801 both from *E. asinus* (Fraser and Bailey *1998*).

A revised nomenclature and numbering of the known and novel allelic sequences based on general recommendations (Klein et al. *1990*; Ellis et al. *2006*) and the IPD-MHC Database (http://0-www.ebi.ac.uk.innopac.up.ac.za/ipd/mhc) have been established. Based on personal consultations (D. Miller, Cornell University), the designation of the alleles respects the format anticipated for the horse part of the IPD-MHC database, which is in preparation. The sequences found in *E. caballus* were thus designated *ELA-Eqca*, *Eqbu* for *E. burchelii*, *Eqgr* for *E. grevyi*, *Eqze* for *E. zebra, Eqhe* for *E. hemionus,* and *Eqki* for *E. kiang*. The numbering of the *DRA* alleles respected their order of identification in the respective species, i.e. the first alleles reported were assigned the lowest numbers. The alleles were thus renamed to *ELA-Eqca-DRA*0101* (originally *DRA(2)-1/1*), *ELA-Eqca-DRA*0201* (*ELA-DRA(2)-2/2*), *Eqca-DRA*0301* (*ELA-DRA(2)-3/3*), *Eqca-DRA*0401* (*DRA*JBH11*), *Eqas-DRA*0101* (*ELA-DRA(2)-1/5*), *Eqas-DRA*0201* (*DRA*0601*), *Eqas-DRA*0301* (*ELA-DRA*JBD3*), *Eqas-DRA*0401* (*DRA*JBD17*), *Eqas-DRA*0501* (*ELA-DRA*JBH45*), *Eqhe-DRA*0101* (*ELA-DRA(2)-4/4*) and *Eqbu-DRA*0101* (*DRA*JBZ185*).

For the *DQA* sequences of the domestic horse, only the species designations (*Eqca*) were added from *ELA-Eqca-DQA*0101* (former *ELA-DQA*0101*) to *ELA-Eqca-DQA*1501* (former *ELA-DQA*1501*), followed by *ELA-Eqca-DQA*1601* through *ELA-Eqca-DQA*2101* (*DQA*1901-DQA*24*). Ass and donkey alleles were named *Eqhe-DQA*0101* (corresponding to *ELA-DQA*1601*) and *Eqas-DQA*0101* and *Eqas-DQA*0201* (*Ela-DQA*1701* and *ELA_DQA*1801*), respectively. Alleles shared between species in both loci were assigned species-specific symbols and a reference to their existence in other species was made (Table 1).

Table 1 *Numbers of alleles detected in this study/total numbers of alleles identified so far in two MHC class II (DRA and DQA) loci in equids (numbers of alleles shared across species in brackets)*

Species	Locus	
	DRA	***DQA***
E. caballus	0/4 (0)	0/21 (1)
E. asinus	0/6 (3)[a]	0/2 (0)
E. hemionus	1/2 (1)	1/2 (0)
E. kiang	2/3 (2)	3/3 (0)
E. zebra	5/5 (4)	4/4 (3)
E. burchelii	5/6 (4)	8/8 (3)
E. grevyi	2/2 (1)	2/2 (1)

[a]Including the new allele *ELA-DRA*Blk165*, not involved in other analyses

Statistical analysis of sequences

Nucleotide and amino acid sequences were analysed using MEGA, version 3.1 (Kumar et al. *2004*).

Within-species variation

The numbers of synonymous (dS) and non-synonymous distances (dN) were computed. For both genes, the CG/AT proportion was determined. The overall transition/transversion bias R was computed, and due to unequal proportion of transitional and transversional substitution, the analyses of selection were provided by modified Nei–Gojobori (Nei and Gojobori *1986*) method with Jukes–Cantor correction (Jukes and Cantor *1969*) for multiple substitution.

Molecular evolutionary analyses

The molecular tree was designed using the neighbour-joining method with Kimura-2 parameters correction. All exon 2 *DRA* and *DQA* alleles known so far in Equidae and reported above were used for constructing the *DRA* phylogenetic tree. The bovine and pig BoLA-*DRA**01011 (DQ821713), SLA-*DRA**002- (AY754888), and BoLA-*DQA*1*08 (U80874), BoLA-*DQA*2*06 (U80868) and SLA-*DQA**0101 (DQ159886) allelic sequences were used as outgroups.

Analysis of positive selection

Analysis of selection was performed using the maximum-likelihood-based random-sites model analysis in the programme CodeML within the PAML 3.14 software (Yang et al. *2000*). We used maximal likelihood (ML) models accounting for the heterogeneity amongst site partitions by using different ω (=dN/dS) parameters for the partitions. The likelihood ratio model was used to test whether ω was significantly different from 1. The random-sites models implemented in this study were M0, M1a, M2a, M3, M7 and M8. The models M0 (one ratio) with M3 (discrete), M1 (neutral) with M2 (selection) and M7 (beta) with M8 (beta and ω), respectively, were compared (see Yang et al. *2000* for details). When the alternative models (M2a, M3 and M8) suggest the presence of sites with $\omega > 1$, all three tests can be

considered as tests of positive selection (Yang et al. *2000*). Posterior probabilities for site classes have been calculated by naive Bayes and Bayes empirical Bayes analysis (Yang and Swanson *2002*) and empirical Bayes method (Yang et al. *2005*) for models M2a and M8. If the posterior means of ω for some site classes were >1, those sites were interpreted to be likely to be under positive selection (Yang et al. *2005*). The models were applied to the exon 2 reading frame matched 243 bp *DRA* and 201 bp *DQA* sequences encoding amino acids 5–84 (numbering according to Brown et al. *2004*) and 11–77 (numbering according to Fraser and Bailey *1998*), respectively.

Results

Within-species variation: allelic diversity of *DRA* and *DQA* genes

Total numbers of alleles and numbers of alleles shared between species are summarised in Table 1. The numbers of variable positions, over mean distances and numbers of synonymous and non-synonymous distances are in Table 2.

Table 2 *Comparison of diversity of DRA and DQA alleles expressed as numbers of sequences analysed (N), all variable nucleotide positions (VNP) and parsimony informative positions (PIP), over mean distance OMD (the number of nucleotide substitution) with its standard error s_e (bootstrap 1,000 replicates), average percentage proportion of CG/AT nucleotides in sequences, percentage of mean divergence of nucleotide substitutions, synonymous (dS) and non-synonymous distances (dN) with s_e (bootstrap 1,000 replicates), and overall transition/transversion bias R*

	Number	VNP/PIP	OMD ± s_e	GC:AT	dS ± s_e	dN ± s_e	dN/dS	*R*
DRA	17	12/7	3.118 ± 0.943 (1.3%)	46.4:53.6	0.036 ± 0.017	0.008 ± 0.003	0.222	4.475
DQA	37	77/61	18.952 ± 2.265 (9.4%)	45:55	0.128 ± 0.028	0.098 ± 0.018	1.306	1.14

Altogether, 11 *DRA* including six novel exon 2 allelic sequences were found in the group analysed (GenBank accession numbers EU930116–EU930129 and FJ657514, the same sequence having different accession numbers in different species); five sequences were shared by more than one species (Table 1). Novel alleles *Eqhe-DRA*0201* in *E. hemionus*, *Eqki-DRA*0101* and *Eqki-DRA*0201* in *E. kiang*, *Eqze-DRA*0101* through *Eqze-DRA*0501* in *E. zebra*, *Eqgr-DRA*0101* and *Eqgr-DRA*0201* in *E. grevyi*, *Eqbu-DRA*0201* through *Eqbu-DRA*0601* in *E. burchellii* were identified. Alleles *Eqze-DRA*0401*, *Eqbu-DRA*0501*, *Eqze-DRA*0501*, *Eqbu-DRA*0401* were identified by the Haplofinder software. Alignment of all *DRA* sequences used also for evolutionary and selection analyses is in Fig. 1.

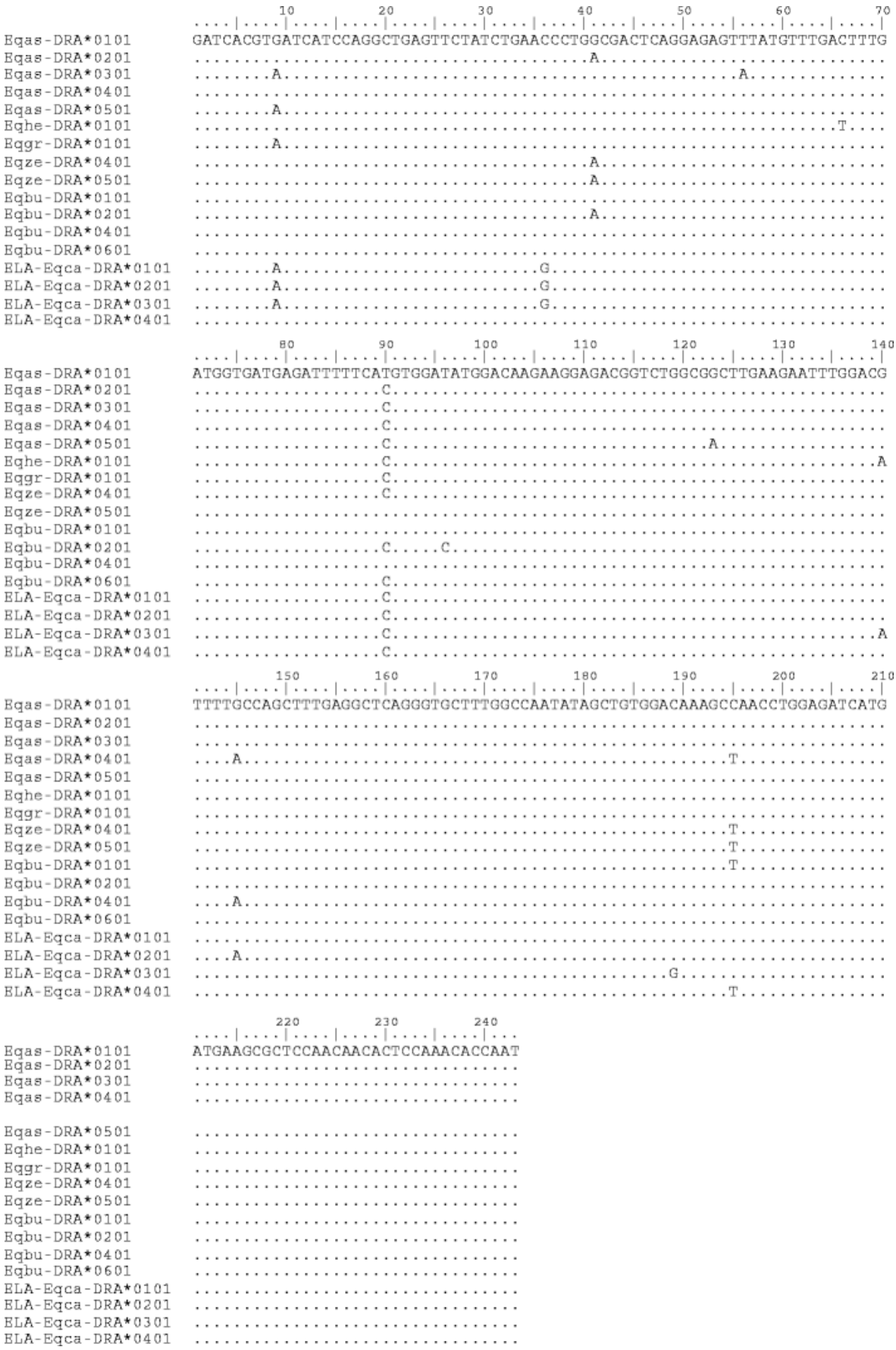
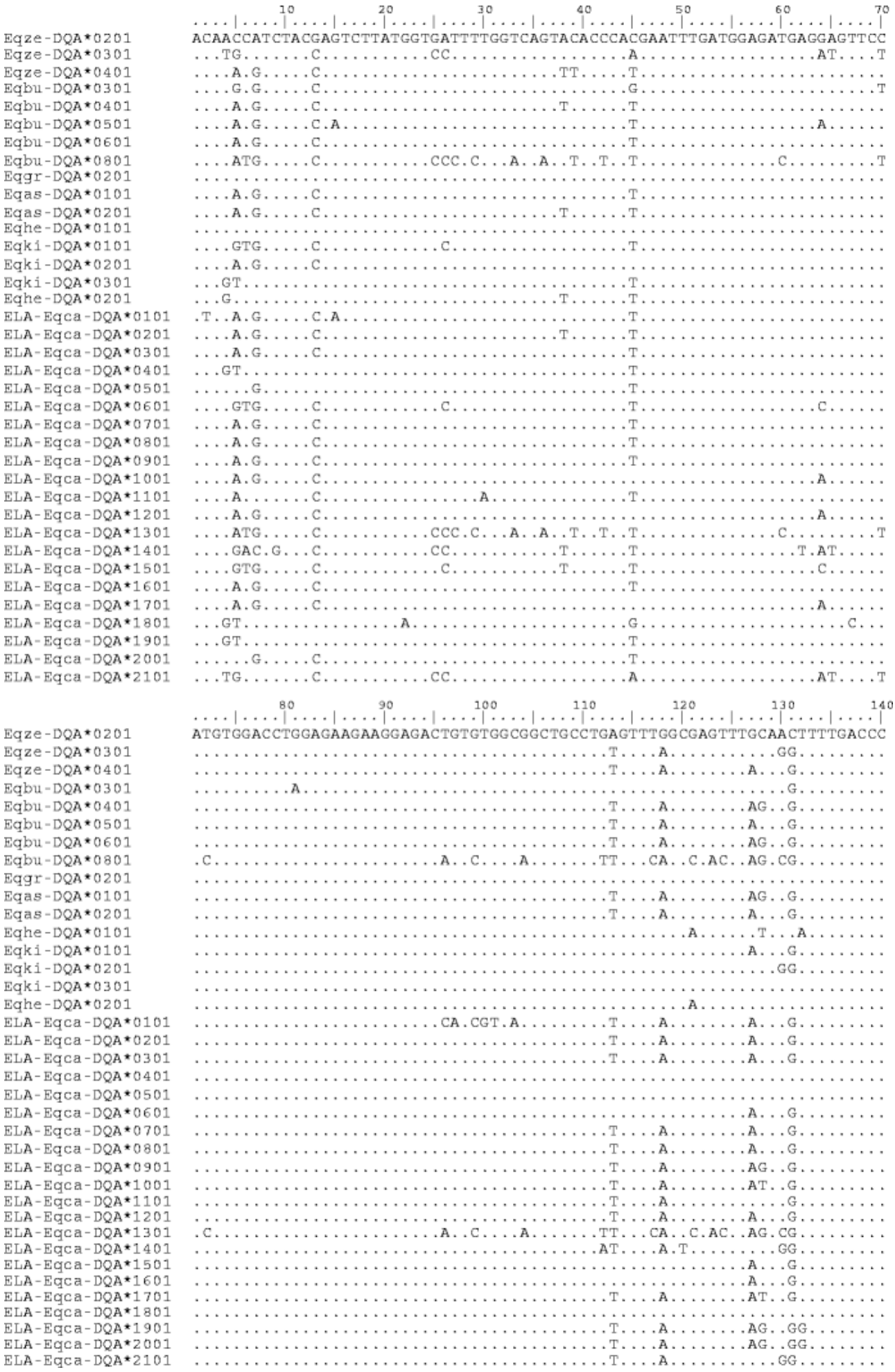


Fig. 1 *Alignment of nucleotide sequences of the second exon of the horse DRA gene used for phylogenetic and selection analysis*

Fifteen exon 2 *DQA* were found in our group of Equidae (GenBank accession numbers EU930130–EU930136 and EU935828–EU935838). Out of them, 13 novel sequences, *Eqgr-DQA*0101* and *Eqgr-DQA*0201* in *E. grevyi*, *Eqbu-DQA*0101* through *Eqbu-DQA*0801* in *E. burchelii*, *Eqze-DQA*0101* through *Eqze-DQA*0401* in *E. zebra*, *Eqki-DQA*0101* through *Eqki-DQA*0301* in *E. kiang* and *Eqhe-DQA*0201* in *E. hemionus* were found. Some *DQA* exon 2 nucleotide sequences were found to be shared by more species (Table 1). All exon 2 *DQA* sequences available used for phylogenetic and selection analyses are aligned in Fig. 2.


10 20 30 40 50 60 70
Eqze-DQA*0201 ACAACCATCTACGAGTCTTATGGTGATTTTGGTCAGTACACCCACGAATTTGATGGAGATGAGGAGTTCC
Eqze-DQA*0301 ...TG.....C.....CC.....A.....AT....T
Eqze-DQA*0401 ....A.G.....C.....TT.....T.....
Eqbu-DQA*0301 ....G.G.....C.....G.....T
Eqbu-DQA*0401 ....A.G.....C.....T.....T.....
Eqbu-DQA*0501 ....A.G.....C.A.....T.....A.....
Eqbu-DQA*0601 ....A.G.....C.....T.....
Eqbu-DQA*0801 ....ATG.....C.....CCC.C...A..A..T..T..T.....C.....T
Eqgr-DQA*0201 .....
Eqas-DQA*0101 ....A.G.....C.....T.....
Eqas-DQA*0201 ....A.G.....C.....T.....T.....
Eqhe-DQA*0101 .....
Eqki-DQA*0101 ....GTG.....C.....C.....T.....
Eqki-DQA*0201 ....A.G.....C.....
Eqki-DQA*0301 ...GT.....T.....
Eqhe-DQA*0201 ...G.....T.....T.....
ELA-Eqca-DQA*0101 .T..A.G.....C.A.....T.....
ELA-Eqca-DQA*0201 ....A.G.....C.....T.....T.....
ELA-Eqca-DQA*0301 ....A.G.....C.....T.....
ELA-Eqca-DQA*0401 ...GT.....T.....
ELA-Eqca-DQA*0501 ......G.....T.....
ELA-Eqca-DQA*0601 ....GTG.....C.....C.....T.....C.....
ELA-Eqca-DQA*0701 ....A.G.....C.....T.....
ELA-Eqca-DQA*0801 ....A.G.....C.....T.....
ELA-Eqca-DQA*0901 ....A.G.....C.....T.....
ELA-Eqca-DQA*1001 ....A.G.....C.....A.....
ELA-Eqca-DQA*1101 ....A.....C.....A.....T.....
ELA-Eqca-DQA*1201 ....A.G.....C.....A.....
ELA-Eqca-DQA*1301 ....ATG.....C.....CCC.C...A..A..T..T..T.....C.....T
ELA-Eqca-DQA*1401 ....GAC.G...C.....CC.....T.....T.....T.AT.....
ELA-Eqca-DQA*1501 ....GTG.....C.....C.....T.....T.....C.....
ELA-Eqca-DQA*1601 ....A.G.....C.....T.....
ELA-Eqca-DQA*1701 ....A.G.....C.....A.....
ELA-Eqca-DQA*1801 ...GT.....A.....G.....C...
ELA-Eqca-DQA*1901 ...GT.....T.....
ELA-Eqca-DQA*2001 ......G.....C.....T.....
ELA-Eqca-DQA*2101 ...TG.....C.....CC.....A.....AT....T
80 90 100 110 120 130 140
Eqze-DQA*0201 ATGTGGACCTGGAGAAGAAGGAGACTGTGTGGCGGCTGCCTGAGTTTGGCGAGTTTGCAACTTTTGACCC
Eqze-DQA*0301 .....T....A.....GG.....
Eqze-DQA*0401 .....T....A.....A...G.....
Eqbu-DQA*0301 .....A.....G.....
Eqbu-DQA*0401 .....T....A.....AG..G.....
Eqbu-DQA*0501 .....T....A.....A...G.....
Eqbu-DQA*0601 .....T....A.....AG..G.....
Eqbu-DQA*0801 .C.....A..C....A.....TT...CA..C.AC..AG.CG.....
Eqgr-DQA*0201 .....
Eqas-DQA*0101 .....T....A.....AG..G.....
Eqas-DQA*0201 .....T....A.....A...G.....
Eqhe-DQA*0101 .....A.....T...A.....
Eqki-DQA*0101 .....A...G.....
Eqki-DQA*0201 .....GG.....
Eqki-DQA*0301 .....
Eqhe-DQA*0201 .....A.....
ELA-Eqca-DQA*0101 .....CA.CGT.A.....T....A.....A...G.....
ELA-Eqca-DQA*0201 .....T....A.....A...G.....
ELA-Eqca-DQA*0301 .....T....A.....A...G.....
ELA-Eqca-DQA*0401 .....
ELA-Eqca-DQA*0501 .....
ELA-Eqca-DQA*0601 .....A...G.....
ELA-Eqca-DQA*0701 .....T....A.....A...G.....
ELA-Eqca-DQA*0801 .....T....A.....A...G.....
ELA-Eqca-DQA*0901 .....T....A.....AG..G.....
ELA-Eqca-DQA*1001 .....T....A.....AT..G.....
ELA-Eqca-DQA*1101 .....T....A.....G.....
ELA-Eqca-DQA*1201 .....T....A.....A...G.....
ELA-Eqca-DQA*1301 .C.....A..C....A.....TT...CA..C.AC..AG.CG.....
ELA-Eqca-DQA*1401 .....AT....A.T.....GG.....
ELA-Eqca-DQA*1501 .....A...G.....
ELA-Eqca-DQA*1601 .....A...G.....
ELA-Eqca-DQA*1701 .....T....A.....AT..G.....
ELA-Eqca-DQA*1801 .....
ELA-Eqca-DQA*1901 .....T....A.....AG..GG.....
ELA-Eqca-DQA*2001 .....T....A.....AG..GG.....
ELA-Eqca-DQA*2101 .....T....A.....GG.....

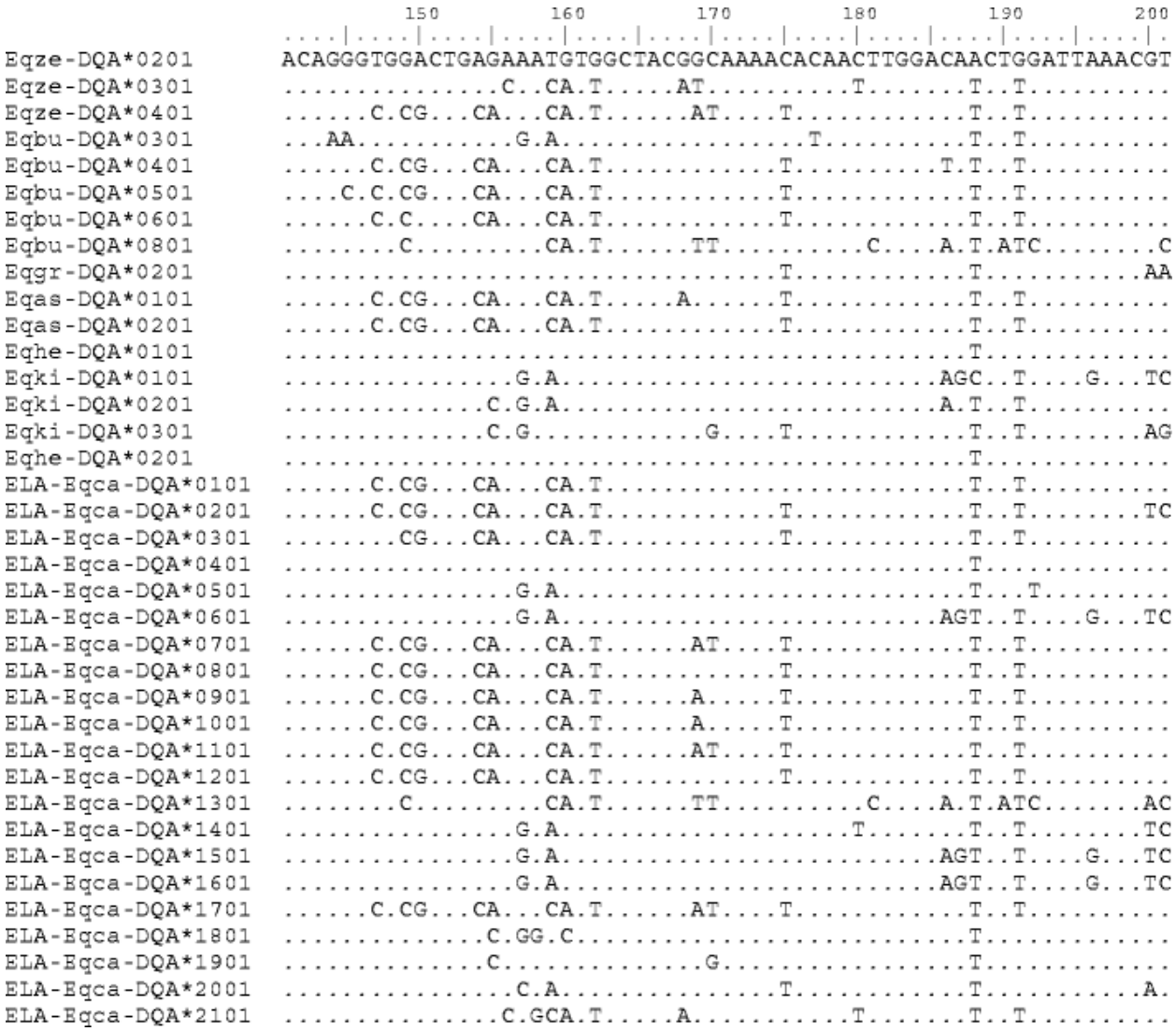


Fig. 2 *Alignment of nucleotide sequences of the second exon of the horse DQA gene(s) used for phylogenetic and selection analysis*

Evolutionary analysis

Evolutionary relationships amongst exon 2 allelic *DRA* and *DQA* sequences as well as positions of alleles shared between species are apparent from the trees constructed (Figs. 3 and 4, respectively). The *DRA* sequences of Equidae formed a clade well distinct from the outgroups used. In *DQA*, the zebra *Eqbu-DQA*08* and horse *ELA-DQA*1301* were different from other equid alleles forming a clade with the pig *SLA-DQA*0101*. A globin gene-based tree (Oakenfull and Clegg *1998*) provided a comparison with neutral loci (Fig. 5).

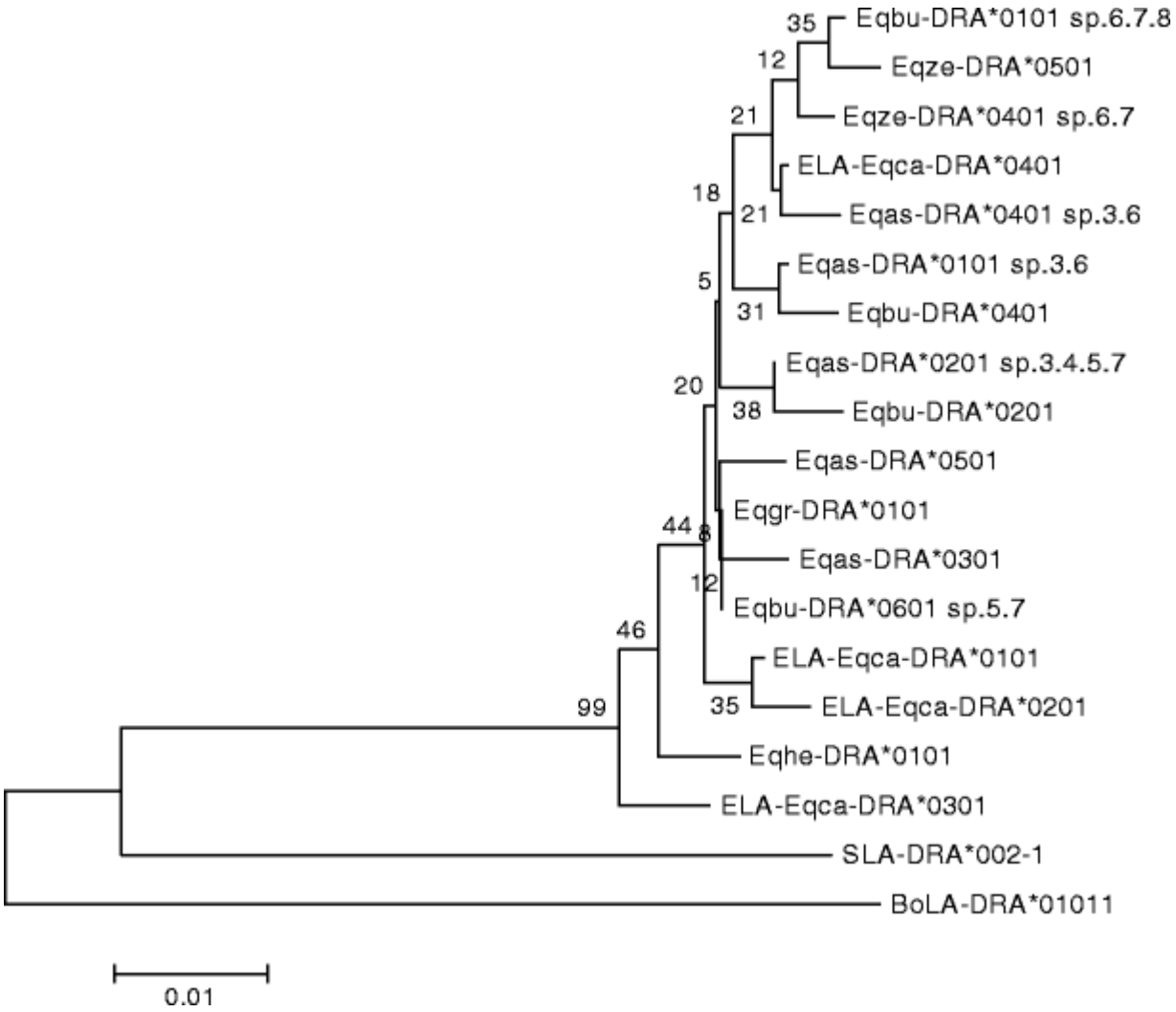


Fig. 3 *Phylogenetic relationships of the MHC exon 2 DRA gene sequences in the genus Equus expressed as Kimura's two-parameter distances. Boostrap of 1,000 replicates are shown. Trans-species sharing is indicated by numbers assigned to species: E. caballus and E. przewalski sp.1,2, E. asinus sp.3, E. hemionus sp.4, E. kiang sp.5, E. zebra sp.6, E. burchellii sp.7, E.grevyi sp.8*

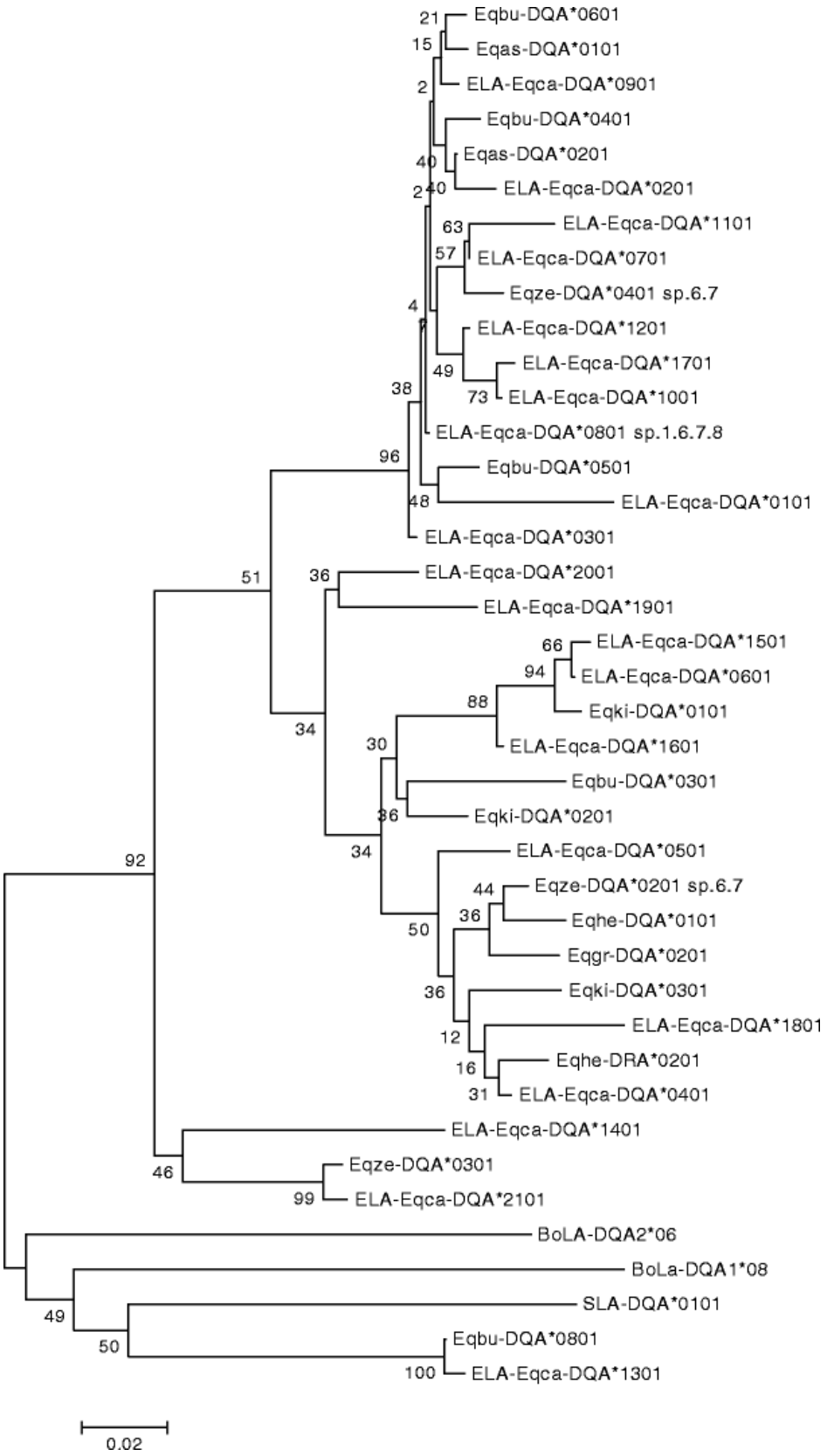


Fig. 4 *Phylogenetic relationships of the MHC exon 2 DQA gene sequences in the genus Equus expressed as Kimura's two-parameter distances. Boostrap of 1,000 replicates are shown. Trans-species sharing is indicated by numbers assigned to species: E. caballus and E. przewalski sp.1,2, E. asinus sp.3, E. hemionus sp.4, E. kiang sp.5, E. zebra sp.6, E. burchelii sp.7, E.grevyi sp.8*

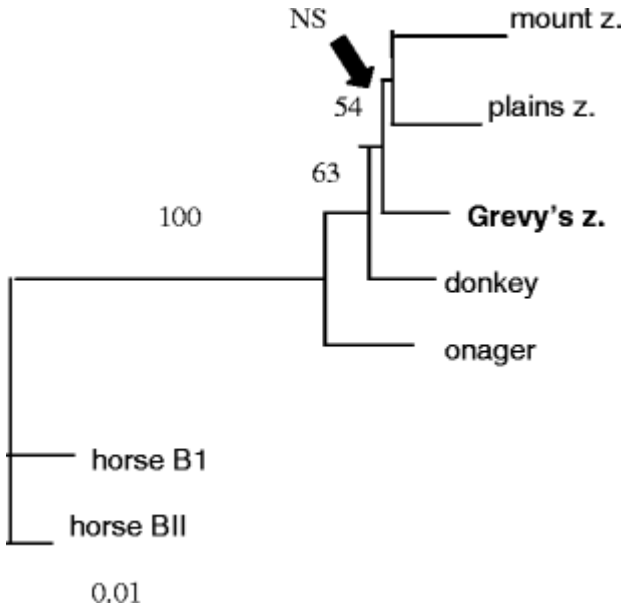


Fig. 5 *Phylogenetic tree of equids produced from the α1 and α2 globin gene DNA sequence. NS indicates a bootstrap value less than 50 (Oakenfull and Clegg 1998)*

Analysis of selection

The selection models revealed different levels of selection in both genes (Table 3). The *DRA* exon 2 sequences showed signs of limited selection with a single strongly selected amino acid site 49 and weakly selected sites 14, 19, 47 and 63. The *DQA* exon 2 seems to be under stronger selection pressure with seven positively selected amino acid sites identified (2, 43, 52, 53, 57, 67) and probably also 64 (Table 4). The posterior means of ω in both genes are in Figs. 6 and 7, respectively. Although the omega values for the *DQA* sequences sites 3 and 13 seemed to be high in the graphs, observed p values were not significant (BEB p for site 3 = 0.892, for site 13 $p = 0.747$).

Table 3 *Summary of test statistics for the likelihood-ratio test of exon 2 of the DRA and DQA genes*

	df	**Test statistic**	**Significance (*p* value)**
DRA models compared			
M1a vs M2a	2	5.37	0.068
M3 vs M0	4	13.098	0.011
M8 vs M7	2	5.634	0.06
DQA models compared			
M1a vs M2a	2	78.408	<0.001
M3 vs M0	4	170.704	<0.001
M8 vs M7	2	80.308	<0.001

The test statistics were computed as 2(Lb − La), where La and Lb are log-likelihood values for each of the nested models compared

Table 4 *Positively selected sites DRA and DQA sequences identified in models M2a and M8 by the naive empirical Bayes and Bayes empirical Bayes procedure (Yang et al. 2005)*

***DRA* model code**	***P***	**Log-likelihood**	**Estimates of parameters**	**Positively selected sites (NEB analysis)**	**Positively selected sites (BEB analysis)**
M0: one ratio	1	−436.846	$\omega = 0.513$	None	None
M1: neutral	1	−432.982	$p_0 = 0.744$, $\omega_0 = 0$ $p_1 = 0.256$, $\omega_1 = 1$	Not allowed	Not allowed
M2a: positive selection	3	−430.297	$p_0 = 0.895$, $\omega_0 = 0.019$ $p_1 = 0.000$, $\omega_1 = 1$ $p_2 = 0.105$, $\omega_2 = 5.18$	14, **47**, **49**	49
M3: discrete	5	−430.297	$p_0 = 0.051$, $\omega_0 = 0.019$ $p_1 = 0.844$, $\omega_1 = 0.019$ $p_2 = 0.105$, $\omega_2 = 5.018$	Not analysed	Not analysed
M7: beta	2	−433.115	$p = 0.005$, $q = 0.020$	Not allowed	Not allowed
M8: beta and omega *DQA* model code	4	−430.298	$p_0 = 0.893$, $p_1 = 0.107$, $p = 0.005$, $q = 0.114$, $\omega = 4.944$	**14**, 19, **47**, **49**, 63	**49**
M0: one ratio	1	−1,266.800	$\omega = 1.180$	None	None
M1: neutral	1	−1,221.653	$p_0 = 0.557$, $\omega_0 = 0.040$ $p_1 = 0.443$, $\omega_1 = 1$	Not allowed	Not allowed
M2a: positive selection	3	−1,182.449	$p_0 = 0.438$, $\omega_0 = 0.012$ $p_1 = 0.429$, $\omega_1 = 1$ $p_2 = 0.133$, $\omega_2 = 7.180$	**2**, **43**, **52**, **53**, **57**, 64, **67**	**2**, **43**, **52**, **53**, **57**, 64, **67**
M3: discrete	5	−1,181.448	$p_0 = 0.510$, $\omega_0 = 0.057$ $p_1 = 0.370$, $\omega_1 = 1.156$ $p_2 = 0.121$, $\omega_2 = 8.802$	Not analysed	Not analysed
M7: beta	2	−1,222.614	$p = 0.083$, $q = 0.100$	Not allowed	Not allowed
M8: beta and omega	4	−1,182.460	$p_0 = 0.867$, $p_1 = 0.133$, $p = 0.0117$, $q = .016$, $\omega = 7.213$	**2**, **43**, **52**, **53**, **57**, 64, **67**	**2**, **43**, **52**, **53**, **57**, 64, **67**

Sites inferred under selection at 90% level are listed in standard font and at the 99% level in bold. Test statistic was computed as 2(Lb − La), where La and Lb are log-likelihood values for each of the nested models compared

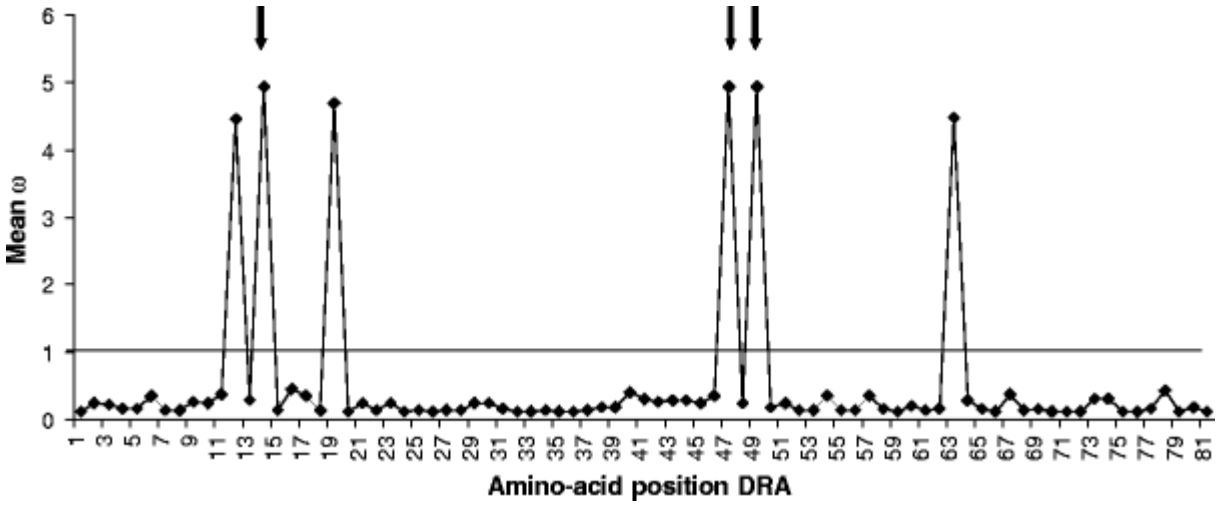


Fig. 6 *Selected amino acid positions in the DRA exon 2 antigen-binding site. Posterior means of ω, calculated as the average of ω over the 11 site classes, weighted by the posterior probabilities under the random-sites model M8 (β and ω) computed by using the CodeML procedure implemented in the PAML3.14 software package. Arrows indicate sites with significant positive selection identified by the naive empirical Bayes and Bayes empirical Bayes procedure*

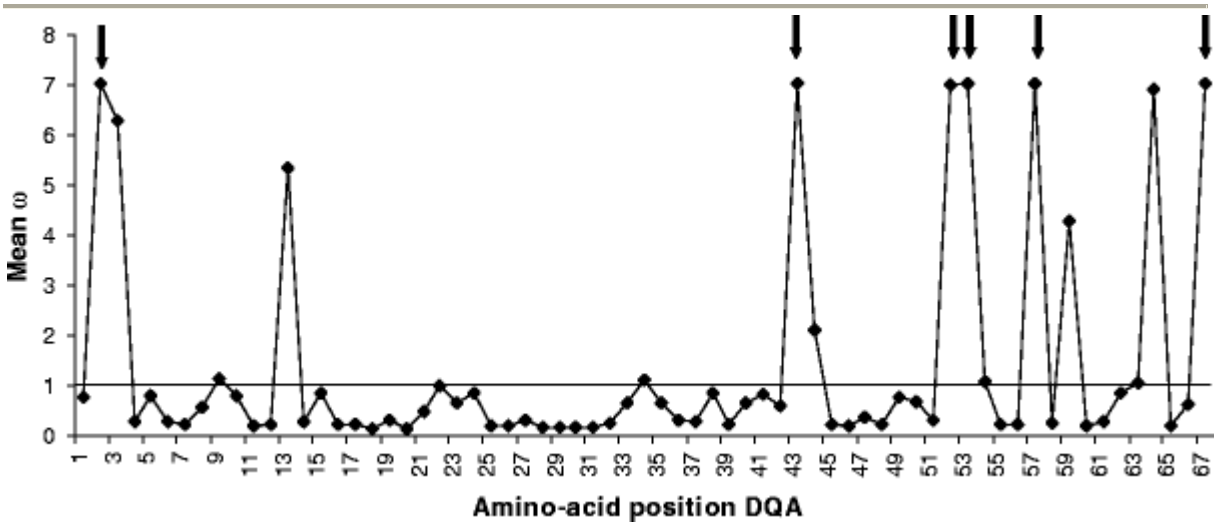


Fig. 7 *Selected amino acid positions in the DQA exon 2 ABS. Posterior means of ω, calculated as the average of ω over the 11 site classes, weighted by the posterior probabilities under the random-sites model M8 (β and ω) computed by using the CodeML procedure implemented in the PAML3.14 software package. Arrows indicate sites with significant positive selection identified by the naive empirical Bayes and Bayes empirical Bayes procedure*

Based on alignment to sequences from humans and rodents in which the amino acid positions associated with antigen binding are known, the amino acid sites 52, 53 and 54 of the horse *DQA* exon 2 sequences were selected as candidate positions for the ABS and analysed in more details. Eleven different three-triplet nucleotide motifs at these positions were found in the 37 *DQA* alleles analysed. The motif CAA AAC ATT corresponding to amino acids QNI was identified in 16 alleles, whilst AGA GAA GTG (REV) was found in seven, AGA AAT GTG (RNV) in five, and AGA AAC ATT (RNI) in two alleles. The remaining nine-nucleotide combinations occurred in only one allele.

Information on selection acting on vole *DQA* sequences (Bryja et al. *2006*) allowed us a comparison with a group of free-ranging mammals naturally exposed to pathogens. Alignment of vole and equid sequences showed that specific *DQA* sites under selection were common for voles and horses. Strongly selected amino acid positions in voles 11, 52, 62, 66 and 76 corresponded to our strongly selected sites 2, 43, 53, 57 and 67, whilst another selected six sites in voles and two sites in *Equidae* are selected in only one of the two groups. Alignment of human (Paliakasis et al. *1996*), vole and horse *DQA* sequences showed that three out of 11 ABS under selection in humans are also sites under selection in Equidae and voles (amino acids 2, 53 and 67 in equids), two other ABS corresponded to positively selected sites in voles only (positions 31 and 72 in the vole sequence), and one ABS corresponds to a selected site in Equidae (*Equid* position 64; Table 5).

Table 5 *Numbers and sharing of selected amino acid positions (SAAP) in the DQA locus in equids, voles (Bryja et al. 2006) and humans (Paliakasis et al. 1996)*

	Equids			Voles		All species
	Number of SAAP	**Shared with vole**	**Shared with human**	**Number of SAAP**	**Shared with human**	
Numbering of shared SAAP	7	**2, 43, 53, 57, 67** (11, 52, 62, 66, 76)	**2, 53, 64, 67**	11	11, 31, 62, 72, 76	**2, 53, 67** (11, 62, 76)

Numbering of the sites is based on equid sequences (in bold), vole numbering is in standard font

Discussion

The data provide evidence for within-species variation in the numbers of alleles in the species analysed as well as for interspecies and inter-locus differences (Table 1). Six novel *DRA* and 13 novel *DQA* sequences were unequivocally identified in the whole family *Equidae*. The new nomenclature suggested should make their list comparable to other species. Although the traditional designation ELA will always be used in general terms, the “systemic” species-derived designation could be useful for interspecies comparisons. Numbering of the originally reported alleles (Albright-Fraser et al. *1996*; Fraser and Bailey *1998*) did not change. Different approaches, not always based on species assignment, were used for the designation and numbering of alleles described more recently. Therefore, new numbers for these alleles reflecting their occurrence in the particular species have been introduced.

It is not clear to what extent the results are influenced by the size and non-random composition of the captive groups available. A within-species comparison of captive and free-living *E. b. burchelli* and *E. z. hartmannae* studied here revealed no differences in terms of the presence of *DRA* or *DQA* alleles. As the composition of the groups available was not suitable for population analyses, calculations of allelic frequencies and heterozygosity parameters would not be informative and therefore were not done. Further cloning and sequence analysis would provide more information about the so far undetected variation.

The numbers of MHC alleles identified in each species ranged from two to six (mean 3.7) for the *DRA* and from 2 to 21 (mean 6.0) for the *DQA* exon 2 sequences. When excluding exceptionally high numbers of *DQA* alleles found in the domestic horse, the range is two to

eight, with a mean of 3.5. Despite relatively high numbers of kiangs ($n = 12$) and *E. grevyi* ($n = 39$) analysed in this study, low total numbers of *DRA* and *DQA* alleles, respectively, have been found in these species so far. These data, comparable to the values of diversity of MHC class II loci found in other mammals like voles (Bryja et al. *2007*; Cutrera and Lacey *2007*), swine (Chardon et al. *1999*; Chen et al. *2005*) and/or humans (Bondinas et al. *2007*), demonstrate similar diversity in the respective MHC loci in the equid species analysed and support the view that most of the existing diversity has been identified. The high *DQA* diversity found in the domestic horse as compared to other equids remains unexplained. Analyses of larger numbers of animals are needed to confirm the differences observed in this study. Sequence analysis did not allow us to distinguish alleles of particular *DQA* loci; the *DQA* sequences did not form gene-specific clusters. It is possible that our primers amplified preferentially one of the loci. Although three expressed *DQA* loci were recently identified in the domestic horse (Miller and Antczak *2008*), we always retrieved only two sequences maximum from a single animal. Two DQB loci were found by a BAC contig analysis (Gustafson et al. *2003*), but it remains unclear how many functional DQ heterodimers are expressed in an individual horse. It also remains to be established whether interspecies and/or individual variation in the number of *DQA* loci does exist in equids, like in cattle (Ballingall et al. *1997*).

As expected, the diversity of the *DQA* exon 2 was higher than in the *DRA* gene whose polymorphism in equids is a rather extraordinary feature. The results confirmed high variation of the equid *DRA* exon 2 reported previously by others (Albright-Fraser et al. *1996*; Brown et al. *2004*; Luís et al. *2005*). In contrast to the domestic horse where recent analysis of large numbers of horses by pyrosequencing did not reveal existence of new alleles (Diaz et al. *2008*), we identified novel exon 2 sequences in other equid species. The reasons for large interspecies differences in the *DRA* polymorphism are unknown, although such differences in the extent of the MHC polymorphism were observed in other groups of mammals. It seems that they are not necessarily influenced by ecology of the species (Klein *1987*). Differences in selection acting on different loci are one of the possible explanations. In this context, the role of positive selection on polymorphic *DRA* genes has not yet been analysed. Balancing selection can be detected by analysing evolutionary trees with trans-species allele sharing, by determining the rate of non-synonymous vs. synonymous substitutions (dN/dS ratio) and/or by using specific software for model selection analysis at the molecular level.

Analysis of evolutionary trees showed that the domestic horse *DRA* alleles were separated from the zebra group, the horse alleles were closer to the root of the tree with a single exception (*ELA-Eqca-DRA*0401*), and they did not overlap with zebra alleles. On the other hand, the ass/donkey *DRA* sequences clustered with all other species. Clustering of *DQA* exon 2 alleles followed much less the patterns of the current taxonomic classification. The sequences of zebras and asses/donkeys clustered together with various horse sequences. Furthermore, a separate group of two alleles clustering with an SLA sequence could be identified. Comparison of the MHC *DQA* and *DRA* phylogenetic trees with trees based on other molecular markers showed patterns of clustering similar to MHC class I sequences from various equid species (Holmes and Ellis *1999*), whilst trees based on equid mitochondrial DNA (mtDNA; George and Ryder *1986*), neutral globin genes (Oakenfull and Clegg *1998*) and microsatellite trees (Krüger et al. *2005*) showed clear demarcation amongst equid species and even subspecies (Fig. 5). Discrepancies between allelic and taxonomic trees may be explained by several factors. Due to their role in antigen presentation and in host pathogen interactions, the MHC genes are subject to balancing selection (Hedrick *1999*;

Hughes and Yager *1998*; Sommer *2005*). According to the trans-species hypothesis, selected alleles cluster into allelic lineages regardless of species origin (Klein and O'Huigin *1994*).

In this study, *DRA* alleles were shared between species more often than the *DQA* sequences. Out of 17 *DRA* sequences analysed, both asses/donkeys and zebras shared three, and two *DRA* exon 2 sequences were found in more than one zebra species. The *DRA* alleles were shared even amongst the most distant species in these trees, i.e. zebras and asses, whilst the *DQA* alleles were shared either between the domestic horse and zebras or amongst various zebras (Table 1). Besides one sequence common for the domestic horse and zebra and two sequences shared by two zebra species, the remaining 34 *DQA* alleles were found in only one species. Small sample sizes and their non-random genetic composition, especially in captive animals, could, however, influence the extent of allele sharing observed. Nevertheless, the data indicate that like in other taxonomic groups, e.g. wolves and dogs (Seddon and Ellegren *2002*; Hedrick et al. *2000*), the phylogeny of MHC genes in equids is different from neutral loci and that other factors than phylogenetic diversification contributed to the allelic variation observed. The divergence time between horses and zebras based on microsatellite trees has been estimated to be 0.86–2.3 Ma and between true horses and true assess to be 0.88–2.3 Ma (Krüger et al. *2005*). Based on mtDNA analysis, the common ancestor of all extant forms could exist about 3.9 Mya, and speciation leading to zebras, asses and horses could take place within the next 0.5 My (George and Ryder *1986*). These data and our results thus show that some *DRA* and *DQA* allelic lineages can be evolutionarily old.

The PAML procedure used is considered to be more sensitive than other methods for assessing selection at the molecular level (Anisimova et al. *2001*). Here, selection on both *DQA* and *DRA* genes was demonstrated. The results thus showed that like more polymorphic MHC class II *DQA*, *DQB* and/or *DRB* loci, *DRA* polymorphism is also subject to positive selection and might be of functional importance. The intensity of selection of MHC genes can be different between genes. In the natural populations of voles, the *DQA1* locus was proposed to be exposed to stronger selection than closely linked *DRB* genes (Bryja et al. *2007*). Here, we observed strong selection acting on the *DQA* loci, whilst less intensive selection has been postulated for the *DRA* locus. The nucleotides in the ABS sites were highly variable, and the *DQA* gene was shown to be under stronger selection pressure with more amino acid sites subject to selection. In the ABS coding exon 2 sequences of MHC genes, a dN/dS ratio higher than in other domains has been well documented and explained by balancing selection (Bryja et al. *2006*; Hedrick et al. *2000*; Seddon and Ellegren *2002*). The ABS coding sequences show a high level of variation not only in terms of the numbers of alleles but also in the extent of sequence variation between alleles (Hughes and Yager *1998*), with positions of ABS variable between different alleles of the same locus (Latek et al. *2000*). These data are in agreement with the assumption that higher sequence variability of *DQA* than *DRA* could be correlated with higher selection pressure and numbers of selected sites. The evidence for positive selection at the *DRA* amino acid level is in agreement with the high degree of trans-species sharing observed for the *DRA* exon 2 sequences. The dN/dS ratio is low due probably to the presence of conserved sites.

Altogether, the ratio of non-synonymous vs. synonymous nucleotide substitutions, interspecific allele sharing and model selection analysis provided evidence for positive selection in the MHC *DRA* and *DQA* class II genes in the family Equidae. The *DQA* locus is thus comparable to other mammalian groups, whilst the *DRA* locus differs in this aspect from typical mammalian *DRA* polymorphism patterns.

Acknowledgment This study was supported by the project of the Czech National Grant Agency, GA CR 523/06/1402. We wish to thank Dr. Josef Bryja for his valuable help with the selection analysis.

References

Albright D, Bailey E, Woodward JG (1991) Nucleotide sequence of a cDNA clone of the horse (*Equus caballus*) DRA gene. Immunogenetics 34:136–138

Albright-Fraser DG, Reid R, Gerber V, Bailey E (1996) Polymorphism of DRA among equids. Immunogenetics 43:315–317 (erratum Immunogenetics 1996; 44:487)

Anisimova M, Bielawski JP, Yang ZH (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Ecol 18(8):1585–1592

Bailey E, Marti E, Fraser DG, Antzac DF, Lazary S (2000) Immunogenetics of the horse. In: Bowling AT, Rusinsky A (eds) The genetics of the horse. CABI, Wallingford (UK), pp 123–155

Ballingall KT, Luyai A, McKeever DJ (1997) Analysis of genetic diversity at the DQA loci in African cattle: evidence for a BoLA-DQA3 locus. Immunogenetics 46(3):237–244

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol 16:363–377

Bondinas GP, Moustakas AK, Papadopoulos GK (2007) The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. Immunogenetics 59(7):539–553

Bowling AT, Rusinsky A (2000) Genetic aspects of domestication, breeds and their origins. In: Bowling AT, Rusinsky A (eds) The genetics of the horse. CABI, Wallingford (UK), pp 25–52

Brown JJ, Thomson W, Clegg P, Eyre S, Kennedy LJ, Matthews J, Carter S, Ollier WER (2004) Polymorphism of the equine major histocompatibility complex class II DRA locus. Tissue Antigens 64:173–179

Bryja J, Galan M, Charbonnel N, Cosson JF (2006) Duplication, balancing selection, and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). Immunogenetics 58:191–202

Bryja J, Charbonell N, Berthier K, Galan M, Cosson F (2007) Density-related changes in selection pattern for major histocompatibility genes in fluctuating populations of voles. Mol Ecol 16(23):5084–5097

Chardon P, Renard C, Vaiman M (1999) The major histocompatibility complex in swine. Immunol Rev 167:179–192

Chen FX, Xie J, Li NL, Zhou Y, Xin L, Chou KY (2005) Novel SLA-DQ alleles and their recombinant molecules in xenogeneic stimulation of human T cells. Transpl Immunol 14(2):83–89

Cutrera AP, Lacey EA (2007) Trans-species polymorphism and evidence of selection on class II MHC loci in tuco-tucos (Rodentia: Ctenomyidae). Immunogenetics 59:937–948

Diaz S, Echeverria MG, It V, Posik DM, Rogberg-Munoz A, Pena NL, Peral-Garcia P, Vega-Pla JL, Giovambattista G (2008) Development of an ELA_DRA gene typing method based on pyrosequencing technology. Tissue Antigens 72(5):464–468

Ellis SA, Bontrop RE, Antczak DF, Ballingall K, Davies CJ, Kaufman J, Kennedy LJ, Robinson J, Smith DM, Stear MJ, Stet RJM, Waller MJ, Walter L, Marsh SGE (2006) ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. Immunogenetics 57(12):953–958

Frankham R (2003) Genetics and conservation biology. Cr Biol 326(Suppl 1):S22–S29

Fraser DG, Bailey E (1996) Demonstration of three DRB loci in a domestic horse family. Immunogenetics 44(6):441–445

Fraser DF, Bailey E (1998) Polymorphism and multiple loci for the horse *DQA* gene. Immunogenetics 47:487–490

George M, Ryder OA (1986) Mitochondrial DNA evolution in the genus *Equus*. Mol Biol Evol 3(6):535–546

Gustafson AL, Tallmadge RL, Pamlachan N, Miller D, Bird H, Antczak DF, Raudsepp T, Chowdhary BP, Skow LC (2003) An order BAC contig map of the equine major histocompatibility complex. Cytogenet Genome Res 102:189–195

Hall TA (1999) BioEdit. A user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. Evolution 53:313–318

Hedrick PW, Kim TJ (2000) Genetics of complex polymorphism: parasites and maintenance of the major histocompatibility complex variation. In: Singh RS, Krimbas CB (eds) Evolutionary genetics: from molecules to morphology. Cambridge University Press, Cambridge, pp 204–234

Hedrick PW, Lee RN, Parker KM (2000) Major histocompatibility complex (MHC) variation in the endangered Mexican wolf and related canids. Heredity 85:617–624

Holmes EC, Ellis SA (1999) Evolutionary history of MHC class I genes in the mammalian order Perissodactyla. J Mol Evol 49(3):316–324

Horin P, Matiasovic J (2002) A second locus and new alleles in the major histocompatibility complex class II (ELA-DQB) region in the horse. Anim Genet 33(3):196–200

Hughes AL, Yager M (1998) Natural selection and the evolutionary history of major histocompatibility complex loci. Front Biosci 3:509–516

Jeffery KJ, Bangham CR (2000) Do infectious diseases drive MHC diversity? Microbes Infect 2:1335–1341

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic, New York, pp 21–132

Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. Immunogenetics 56:683–695

Kennedy LJ, Angles JM, Barnes A, Carmichael LE, Radford AD, Kolier WER, Happ GM (2007) DLA-DRBI, DQAI, and DQBI alleles and haplotypes in North American gray wolves. J Hered 98(5):491–499 (Special issue)

Klein J (1986) Natural history of the major histocompatibility complex. Wiley, New York

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

Klein J, O'Huigin C (1994) MHC polymorphism and parasites. Philos Trans R Soc Lond B 346:351–358

Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllenstein UB, Heise ER, Jones PP, Parham P, Wakeland EK, Watkins DI (1990) Nomenclature for the major histocompatibility complexes of different species—a proposal. Immunogenetics 31(4):217–219

Krüger K, Gaillard C, Stranzinger G, Rieder S (2005) Phylogenetic analysis and species allocation of individual equids usány microsatellite data. J Anim Breed Genet 122(Suppl.1):78–86

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–153

Latek RR, Suri A, Petzold SJ, Nelson CA, Kanagawa O, Unanue ER, Fremont DH (2000) Structural basis of peptide binding and presentation by the type I diabetes-associated MHC class II molecule in NOD mice. Immunity 12:699–710

Lindgren G, Backström N, Swinburne J, Hellborg L, Einarsson A, Sandberg K, Cothran G, Vilà C, Binns M, Ellegren H (2004) Limited number of patrilines in horse domestication. Nat Genet 36:335–336

Luís C, Cothran EG, Oom MM, Bailey E (2005) Major histocompatibility complex locus DRA polymorphism in the endangered Sorraia horse and related breeds. J Anim Breed Genet 122:69–72

Mashima S (2003) Comparative sequence analysis of equine and human MHC class II DOB genes. Cytogenet Genome Res 102(1–4):196–200

Militiadou D, Law AS, Russell GC (2003) Establishment of a sequence-based typing system for BoLA-DRB3 exon 2. Tissue Antigens 62(1):55–65

Miller D, Antczak DF (2008) Polymorphism and genomic organization of the Equine MHC class II region. XXXI Conference of the International Society for Animal Genetics, 20–24 July 2008, RAI Conference Center, Amsterdam, The Netherlands, Book of Abstracts, poster 2125

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

Nowak RM (1999) Walker's mammals of the world, vol II. The Johns Hopkins University Press, Baltimore

Oakenfull EA, Clegg JB (1998) Phylogenetic relationship within the genus *Equus* and the evolution of alpha and theta globin genes. J Mol Evol 47(6):772–783

Otting N, de Groot NG, Doxiadis GGM, Bontrop RE (2002) Extensive Mhc-DQB variation in humans and non-human primate species. Immunogenetics 54:230–239

Ottova E, Simkova A, Martin JF, Gouy de Bellocq J, Gelnar M, Allienne JF, Morand S (2005) Evolution and trans-species polymorphism of MHC class IIβ genes in cyprinid fish. Fish Shellfish Immunol 18:199–222

O'Brien SJ, Yuhki N (1999) Comparative genome organization of the major histocompatibility complex: lessons from the Felidae. Immunol Rev 167:133–144

Paliakasis K, Routsias J, Petratos K, Kokkinidis M, Papadopoulos GK (1996) Novel structural features of the human histocompatibility molecules HLA-DQ as revealed by modeling based on the published structure of the related molecule HLA-DR. J Struct Biol 117:145–163

Pimtanothai N, Hurley CK, Leke R, Klitz W, Johnson AH (2001) HLA-DR and-DQ polymorphism in Cameroon. Tissue Antigens 58:1–8

Russell GC, Gallagher A, Craigmile S, Glass E (1997) Characterization of cattle cDNA sequences from two DQA loci. Immunogenetics 45(6):455–458

Seddon JM, Baverstock PR (2000) Evolutionary lineages of RT1.Ba in the Australian rat. Mol Ecol 17:768–772

Seddon JM, Ellegren H (2002) MHC class II genes in European wolves: a comparison with dogs. Immunogenetics 54:490–500

Seddon JM, Ellegren H (2004) A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent with neutral evolution. Proc R Soc Lond B 271:2283–2291

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16

Szalai G, Antczak DF, Gerber H, Lazary S (1994) Molecular-cloning and characterization of horse DQB CDNA. Immunogenetics 40(6):457

Takada T, Kikkawa Y, Yonekawa H, Amano T (1998) Analysis of goat MHC class II *DRA* and *DRB* genes: identification of the expressed gene and new DRB alleles. Immunogenetics 48(6):408–412

Trowsdale J, Parham P (2004) Defense strategies and immunity-related genes. Eur J Immunol 34(1):7–17

Yang Z, Swanson WJ (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol 19:49–57

Yang Z, Swanson WJ, Vacquier VD (2000) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysine reveals variable selective pressures among lineages and sites. Mol Biol Evol 17(10):1446–1455

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Yuhki N, Beck T, Stephens RM, Nishigaki Y, Newmann K, O'Brien SJ (2003) Comparative genome organization of human, murine, and feline MHC class II region. Genome Res 13(6):1169–1179