Genetic variation and selection in major histocompatibility complex Class II gene in Guizhou pony

Chang Liu[1,2], Hongmei Lei[1], Xueqin Ran[1]*, JiafuWang[1,3,]*

[1]College of Animal Sciences, Guizhou University, Guiyang, China;

[2]College of Pharmacy, Guizhou University of Traditional Chinese Medicine, Guiyang, China;

[3]Tongren University, Tongren, China.

Corresponding author Email: xqran@gzu.edu.cn (X. Ran), jfwang@gzu.edu.cn (J. Wang)

ABSTRACT

The Guizhou pony (GZP) is an indigenous species of equid??, and raised in mountain district of Guizhou province, Southwest of China. To assess the diversity of the major histocompatibility complex (MHC) class II gene in GZP, four regions of Equine leukocyte antigen (ELA) involving in *DQA*, *DRA*, *DQB* and *DRB* were chosen to investigate by direct sequencing technology. Compared with other three loci, *DRA* harbored the lowest d_N/d_S ratio (0.560), which indicated that the *DRA* was conserved and might be under a purifying selection. Many codons in four *ELA* regions were screened out to undergo the significant positive selection occurring at the antigen binding sites (ABS), including nine *DQA*, five *DQB*, nine *DRA* and seven *DRB* codons. The selected residues in ABS would play a significant role in the innate immune system of GZP. Furthermore, two alleles of GZP were shared with Przewalski horse, and six older haplotypes of GZP were much close among GZP, European horse, Przewalski horse and wild ass by one or two mutational steps, providing another clue for GZP as a natural ancient variety. The specific diversity of ABS sites and the numbers of unique haplotypes implied to significant evolution process for the species in a better fitness and adaptation to the indigenous hard environment for the pony inhabited.

Keywords

Guizhou pony, MHC, antigen binding sites, evolution process, adaptation

INTRODUCTION

Commented [HX1]: Not clear what this means

Commented [HX2]: Are predicted to?

Deleted: function

Commented [HX3]: Not clear what this means.

Commented [HX4]: Variety of equid?

Commented [HX5]: No need to say “sites” after “antigen binding sites”

Commented [HX6]: This sentence doesn’t make sense. How do we define what constitutes a ‘significant’ evolutionary process, or ‘better’ fitness? What’s a ‘hard environment’?

The major histocompatibility complex (MHC) genes play a major role in vertebrate immune systems and have a high degree of genetic diversity, which is associated with adaptation and evolution (Reed et al., 2003). The MHC system could be clustered into two classes, including class I and class II, which are key parts of the immune system (Hughes et al.,1988). The MHC class II are highly polymorphic involved in the immune response by presenting extracellular antigens to T lymphocyte cells. These molecules are heterodimers constituted by α and β chains encoded by A and B genes. Most previous studies showed that exon 2 of MHC class II genes exhibited the greatest polymorphism, and encoded the α and β domains that are principally responsible for peptide binding (O'Connor et al., 2007). Frequently, the polymorphism of MHC loci has been found to associated with differences in the susceptibility to infectious diseases (Hill et al., 2001), especially in sheep (Paterson et al., 1998), mouse (Meyer-Lucht et al., 2005), voles (Kloch et al., 2010) and lemur (Schad et al., 2005). The equine MHC class II loci may also help to interpret associations between the polymorphism of equine MHC class II gene region and the encountered pathogens of the horse (Miller et al., 2017).

MHC variants play key roles in mate preference, kin recognition and maternal-fetal interactions (Edwards et al., 1998; Bernatchez et al., 2003; Piertney et al., 2006). These diverse functions and characteristics of MHC molecules reflected evolutionary relevant and adaptive processes within and between populations (Sommer et al., 2005). It has been extensively studied that the mechanisms of negative frequency-dependent and over-dominant selection in MHC genes. The correlative and experimental support for the negative frequency-dependent selection at MHC genes has been provided in humans (Trachtenberg et al. 2003), reed warbler population (Westerdahl et al. 2004), mice (Kubinak et al. 2012), stickleback (Eizaguirre et al. 2012; Bolnick and Stutz 2017) and guppies (Phillips et al. 2018). Examples of asymmetric overdominant selection have been shown in a number of natural and laboratory populations (Landry et al. 2001; Richman et al. 2001; Lenz et al. 2009; Schwensow et al. 2010; Lenz et al. 2013), and supported by several computer-based binding prediction studies (Lenz 2011; Lau et al. 2015; Buhler et al. 2016; Pierini and Lenz 2018). At present, there are three primary sources of evidence to support balancing selection (i) elevated levels of polymorphism, (ii) the rates of nonsynonymous (d_N) to synonymous (d_S) of nucleotide substitutions (Hughes et al., 1988&Hughes et al., 1989), and (iii) trans-species polymorphisms with alleles among species (Klein et al., 1993). The d_N/d_S ratio has been widely used to measure selective pressure on genes (Yang et al., 2000), and substantially differing

Commented [HX7]: Adaptation to what?

Deleted: of animal

Commented [HX8]: Is clustered?

Deleted: is

Deleted: only the

Deleted: was

Deleted:

Deleted: n

Commented [HX9]: "may also be important for determining the host response to encountered pathogens in horses"

Commented [HX10]: Define these terms if they are important to the study

Deleted: ed

rates of evolution across the MHC genes (Hughes et al., 1995). Furthermore, MHC variation within species and interspecies is proved to be useful to determine selection effect in various mammals (Cutrera et al., 2007).

In the family Equidae, the horse MHC class II gene, i.e. Equine leukocyte antigen (ELA) class II, is located on the short arm of chromosome 20q14-q22 (Mäkinen et al., 1989; Ansari et al., 1988), and contains the *DQA*, *DQB*, *DRA* and *DRB*. The *DQA* and *DRA* genes encode for the α-chain of ELA class II molecules, and the polymorphism of *DQA* and *DRA* genes have been determined in European equids (Luís et al., 2005; Janova et al., 2009; Kamath et al., 2011). The *DQB* and *DRB* genes encode the β-chain of ELA class II complex, and high level of *DRB* and *DQB* polymorphism have been reported for Arabian and European horses (Fraser et al., 1996; Mashima et al., 2003). Site-specific methods have found elevated d_N/d_S ratios at these antigen binding sites (ABS), suggesting substantially differing rates of evolution across the MHC. The Exon2 code for a part of the pocket of the MHC molecules, within this sequence some of the amino acid residues are called antigen binding site (ABS), which are much diverse. Previous reports indicated that the exon2 of ELA class II gene is rich in genetic variation among horse populations (Kamath et al., 2011).

The Guizhou pony (GZP) is an indigenous species raised in Guizhou province started from the Warring States Period (475-221 B.C.), ancient China. It is one of the five Chinese pony species, and the body height is only 1.1 meters (10-11 hands). Plenty of nucleotide polymorphism have been determined from several pony populations derived from Irish, Canadian and China native breeds by using mtDNA/SSR markers (McGahern et al. 2006; Prystupa et al. 2012). The purpose of present study is to analyze the variation in MHC II exon 2 of *DQA*, *DRA*, *DQB*, and *DRB* regions, and the relationship with selection and evolution of GZP.

MATERIAL AND METHODS

Animal collection and DNA isolation

A total of 50 blood samples of 4 to 8 years old Guizhou pony (GZP) were collected from Ziyun county, Anshun, Guizhou Province, China. All animal procedures were approved by the Institutional Animal Care and Use Committee of Guizhou University (Approval number EAE-GZU-2018-P007). All sampled ponies were randomly selected and well-developed in good health. The heights of GZP were from 102 to 118 cm in height with weight between 210 to 265 kg. Blood samples of pony were collected from jugular vein and kept in

Commented [HX11]: among species

Commented [HX12]: Historical patterns of selection

Deleted: ing

Commented [HX13]: acronym already defined

Commented [HX14]: What does this mean for the hosts?

Commented [HX15]: During?

Commented [HX16]: This is quite vague. How much is 'plenty'?

Commented [HX17]:

Commented [HX18]: Acronym already defined

Commented [HX19]: How were they caught?

anticoagulants EDTA.Na2. All samples were stored at -20°C until DNA extraction. Genomic DNA was extracted from blood samples using SQ Blood DNA Kit (OMEGA, USA). Nucleic acid concentration of the extracted genomic DNA from GZP was calculated by determination of OD260/OD280, and detected by 0.7% agarose gel electrophoresis.

PCR amplification, cloning and sequencing

The exon 2 regions of the *ELA-DQA*, *DQB*, *DRA* and *DRB* genes were amplified from genomic DNA by PCR method with specific primers. We amplified 246 bp of the *DRA* using equid-specific primers *DRA*-F and *DRA*-R(AlbrightFraser, et al., 1996), 246 bp of the *DQA* using the primer *DQA*-F and *DQA*-R (Fraser et al., 1998), 276 bp of the *DRB* using the primer *DRB*-F and *DRB*-R(Fraser et al., 1996), 230 bp of the *DQB* using the primer *DQB*-F and *DQB*-R (Mashima et al., 2003), and synthesized by the Bio-Engineering (Shanghai) company (**Table 1**). The PCR volume was 20 μL, which contained 10 μL of 2 × PCR Mixture (0.1 U Taq Plus Polymerase/μL, 500 μM dNTP each, 20 mM Tris-HCl (pH8.3), 100 mM KCl, 3 mM $MgCl_2$), 0.4 μL of upstream/downstream primers (10 μmol/L) and 1 μL templates. The PCR amplification was carried out with initial denaturation at 95 °C for 5 min, followed by 30 cycles (95 °C for 30 s, 58 °C for 30 s, and 72 °C for 30 s), final extension at 72°C for 10 min. PCR products were extracted and purified with the Gel Extraction Kit (OMEGA, USA), and ligated into pGEM®-T vectors and transformed into *E. coli* competent cells. Twenty positive clones per sample were picked out with a sterile toothpick and detected by Sanger sequencing method (Invitrogen, China). Each allele was confirmed for at least two observations, meaning that it could be accepted more than one homozygous individual or two heterozygous individuals to be included in the following analyses.

DNA sequence polymorphism analysis

The base composition of the *DRA*, *DRB*, *DQA* and *DQB* genes was counted by MEGA7 software (Kumar et al., 2016). Standard descriptive diversity indices for each locus within the GZP were calculated using MEGA7 software, including the variable sites (V), parsim-info sites (P), singleton sites (S), transition/transversion bias ratio (R). In considering the nature of the variable region, it is of importance to ascertain whether the variability is uniformly distributed or is confined to small segments of the variable regions. Thus, the variation of amino acids calculated by the mutation rate (Variability = The number of different amino acids at a certain

Commented [HX20]: Clones?

Deleted: twice

Commented [HX21]: This isn't clear.

Commented [HX22]: Analyzed in?

Deleted: were

position/Frequency of the most common amino acids at this position) (Wu et al., 1970). Selection was estimated using MEGA7 software in terms of the relative rates of nonsynonymous (d_N) and synonymous (d_S) mutations, according to the Nei and Gojobori's method with the Jukes and Cantor (JC) correction (Nei et al., 1986). Z-Test ($P < 0.05$) of selection were performed over all sites, under the null hypothesis of neutrality ($d_N = d_S$) and the alternative hypotheses of non-neutrality ($d_N \neq d_S$), positive selection ($d_N > d_S$), and purifying selection ($d_N < d_S$).

Site-specific selection analyses and protein 3D structure analysis

The nonsynonymous and synonymous substitution in the overall domain as well as antigen binding site (ABS) and non-ABS were estimated for *DQA*, *DQB*, *DRA* and *DRB* allele. Positive selection was assessed by CodeML subroutine in PAML program (Yang, 2007). The PAML procedure was considered to more sensitive than other methods for assessing selection at the molecular level (Anisimova et al. 2001). The PAML program used maximum likelihood estimation to examine heterogeneity in rates of $\omega = d_N/d_S$ among codons (Bielawski et al., 2003). Compared to other methods, the PAML program sensitively detected molecular evidence of selection (Anisimova et al., 2003). In this analysis, we assessed heterogeneity in ω ($\omega < 1$: purifying selection, $\omega = 1$: neutral evolution, $\omega > 1$: positive selection) across the four alleles (*DQA*, *DQB*, *DRA* and *DRB*) to identify codons under positive selection. Observed ω value followed six models in PAML: M0 (one ratio, average ω across all sites), M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta) and M8 (beta and omega) (Yang et al., 2000). Predictions of the *DQA*, *DQB*, *DRA* and *DRB* protein structure were performed online using the SWISS-MODEL program (Biasini et al., 2014; Waterhouse et al., 2018) (https://swissmodel.expasy.org/interactive).

Phylogenetic allele networks

We inferred phylogenetic relationships among sequence haplotypes by constructing median-joining haplotype network (Bandelt et al., 1999) using maximum parsimony by Network 4.6.1 (http://www.fluxus-engineering.com/sharenet.htm). Allele frequency information and population proportion were incorporated into the visualization of the network. Sequences from the horse (*E. callabus*, *E. przewalskii*, *E. burchellii*, *E. asinus* and so on) were incorporated to evaluate the distance from Guizhou pony haplotypes (**Table 2**).

RESULTS

Analysis of nucleotide diversity

A total of 184 alleles were identified from 1000 sequencing data, and the effective numbers of alleles was 118. It contained 18 novel *DQA* alleles (GenBank accession number: MT304744 - MT304761), 38 new *DQB* alleles (MT304705 - MT304743), 22 new *DRA* alleles (MT304762 - MT304783) and 28 new *DRB* alleles (MT304784 - MT304811) (**Table S1**). The alignment result listed in Table S1 for all the effective alleles from *DQA*, *DQB, DRA*, *DRB* and sequences of JQ254059, AF034122, AJ575295, and AF144564. Based on *DQA*, *DQB*, *DRB* alignment results, it revealed considerable sequence diversity within the same genus. The diversity of nucleotides in *DRA* is much lower than that in *DQA*/*B* and *DRB* in GZP, which, however, is comparable with the level of nucleotide diversity in *DRA* from other species within Equus genus (Kamath et al., 2011). Within GZP, genetic diversity was much higher in *DQA*, *DQB* and *DRB* than that in *DRA*, the ratio (Variable sites/Length) was the lowest at *DRA* locus (15.04%), and the highest at DQB locus (46.08%).

Commented [HX23]: clones

Deleted: were

Analysis of Nucleotide compositions

The GC content in *DQB* and *DRB* were higher than *DQA* and *DRA* (**Table S2**). The content of G+C (48.10 %) was slightly less than A+T (51.90 %) at *DQA*, and the content of G+C (48.10 %) was litter lower than A+T (51.90 %) at *DRA*, which revealed that *DQA* and *DRA* possessed lower GC percentages. On the contrary, the base composition of G+C (64.20 %) was higher than A+T (35.80 %) in *DQB* alleles, and the base composition of G+C (61.90 %) was much more than A+T (38.10 %) in DRB alleles, which showed that *DQB* and *DRB* contained higher GC content area. The R (Transitions/Transversions) was 1.357 and 2.241 at *DQA* and *DRA* allele, respectively. In contrast, it was opposed with the R of 0.778 and 0.573 at *DQB* and *DRB* allele, respectively. The results revealed that *DRA* locus was much more conservative compared with the other loci.

The amino acid composition analyses

It was deduced that the exon 2 of *DQA*, *DQB*, *DRA* and *DRB* nucleotide sequence encoded 82, 76, 81, 79 amino acid sequence, respectively. The underlined residues in **Fig.1** indicated putative ABS based on the HLA equivalents (Brown et al. 1988; Brown et al. 1993), which were thought to be the position to contact with antigens peptide (**Fig** 1.A-D). Among of the predicted amino acid sites, 38 sites (45.12 %), 51 sites (67.10 %), 16 sites (19.75 %), and 49 sites (62.02 %) were variable at the *DQA*, *DQB*, *DRA* and *DRB* of GZP populations,

respectively. For the ABS sites, 17 of 21 sites (80.95 %), 12 of 18 (66.67 %) sites, 5 of 20 (25.00 %) sites, and 13 of 14 (92.85 %) sites were diverse at *DQA*, *DQA*, *DQB*, *DRA* and *DRB* loci. At ABS, the amino acid composition of *DQA*, *DQB*, *DRA* and *DRB* were calculated by MEGA 7 software (**Fig.** 2). The polar R-amino acids at *DQA* locus occupied the highest percentage (46.30 %), which consisted of Gly, Cys, Ser, Tyr, Thr, Asn, Gln (**Fig.** 2). The non-polar R-amino acids at *DRA* locus occupied the highest percentage (58.86 %), which consisted of Ala, Leu, Val, Trp, Ile, Phe, Pro, Met (**Fig.** 2). The largest proportion (30.70 %) of charged R-amino acids was located at *DQB*, composed of Arg, Lys, His, Glu, Asp (**Fig.** 2).

Global selection analyses

Not all the variable amino acids were selected based on Wu-Kabat variability index (Wu et al., 1970). The Wu–Kabat variability index with respect to amino acids at peptide-binding pockets was calculated using the formula given by Wu and Kabat (Wu et al., 1970). At *DQA* locus, total of fifteen amino acids were strongly selected at 10, 17, 18, 21, 23, 30, 46, 51, 52, 58, 60, 61, 62, 65 and 72, with residue 60 the highest variability (**Fig**.3). At *DQB* locus, many polymorphic sites were observed, with eight high mutation loci at 16, 27, 38, 46, 47, 57, 61 and 65 residues (**Fig**.3). However, 12 residues were found to be polymorphic at DRA locus, including residues 12, 14, 15, 19, 29, 39, 47, 49, 63, 64, 67 and 69 (**Fig**.3). At *DRB* locus, amino acid residues at six different positions had high values (more than 30) of Wu-Kabat variability index, the strongly selected amino acid regions could be found at residues 1, 2, 4, 5, 6, 7, 8, 12, 19, 28, 36, 47, 48, 50, 56, 58, 61, 62, 65 and 69 (**Fig**.3). Compare of the d_N/d_S ratio averaged across the whole coding region suggested that positive selection occurred at loci *DQB* (d_N/d_S = 1.127, p=0.322) and *DRB* (d_N/d_S =1.228, p=0.202), and purifying selection appeared at *DQA* (d_N/d_S = 0.779, p=0.143) and *DRA* (d_N/d_S =0.560, p=0.069) (**Table 3**). All codon sites were not statistically significant by Z-tests (p > 0.05, **Table 3**). The estimates of d_N/d_S suggested that both of *DQA* and *DRA* was unlikely affected by the positive selection at the level of the entire gene.

Site-specific selection analyses

It is unlikely for selection to act uniformly across MHC? genes over evolutionary time, but more probable for it to occur at specific sites based on their functional role. For the ABS sites, the rate of nonsynonymous substitutions (d_N =0.594±0.132) exceeded four times than the number of synonymous substitutions (d_S=0.128±0.080) at the *DRB* (**Table 3**). These results are agreement with that observed in the Argentine Creole

Commented [HX24]: repeats 'sites' from AB[Sites]

Commented [HX25]: move to methods

Commented [HX26]: codons?

Commented [HX27]: repeats

horse, which exhibited rate of nonsynonymous substitutions more than four times the number of synonymous substitutions at the exon 2 of *ELA-DRB* (Díaz et al., 2001). For the *DQA* and *DQB*, the ABS sites rate of synonymous substitutions and nonsynonymous substitutions was much close to each other (d_N=0.330±0.088, d_S=0.287±0.110; d_N=0.206±0.079, d_S=0.133±0.076, respectively) (**Table 3**). At the *DRA*, ABS sites exhibited less nonsynonymous substitutions (d_N=0.017±0.008) than synonymous substitutions (d_S=0.028±0.022) with the d_N/ d_S ratio of 0.607(**Table 3**). For the *DRB*, *Z*-tests performed on ABS separately were significant (p = 0.001), providing evidence for positive selection at these sites. However, we could not reject the null hypothesis of neutral evolution at the non-ABS (**Table 3**). At the *DQA*, *DQB* and *DRA*, *Z*-tests by site type also could not reject the null hypothesis of neutrality (p > 0.05). In contrast, non-ABS sites showed more synonymous substitutions than nonsynonymous substitutions with the d_N/ d_S ratio of 0.581, 0.895, 0.520 and 0.678 at the *DQA*, *DQB*, *DRA* and *DRB,* respectively (**Table 3**).

However, for four loci, results from the selection analyses in PAML revealed different levels of selection (**Table 4**). The variable evolutionary rates across codon sites (M3) fit the data better than the M0 model, and models (M2a and M8) had higher log-likelihoods than positive selection (M1a and M7). At the *DQA*, both M2a and M8 models implied that approximately 2 % of sites were possibly under positive selection (ω = 8.583, ω = 8.425) (**Table 4**). Posterior means of ω estimated across *DQA* codons under positive selection models predicted fourteen sites (positions 10, 17, 18, 30, 46, 51, 52, 60, 61, 65, 68, 69, 70, 72) that may be under selection (ω > 1), nine (10, 30, 46, 51, 52, 61, 65, 68, and 72) of which are also putative ABS based on the HLA equivalents (**Fig**. 4). However, the discrete model (M3: 3 discrete evolutionary rate classes) had the highest log-likelihood and estimated that only 6.6 % of codon sites had ω values greater than one (ω = 10.031) with the remaining 93.4 % of sites being assigned ω values close to 0 (**Table 4**).

However, posterior means of ω across *DQB* codon sites estimated by models M2a (ω=6.240) and M8 (ω=6.373), and predicted that only five codons (positions 16, 27, 47, 57, 61) were under significant positive selection. These five codons were also known as putative ABS based on the HLA equivalents (**Fig.** 4). At the *DQB*, M3 model estimated that approximately 8 % of codon sites had ω values greater than one (7.3 % with ω = 1.283; 0.6 % with ω = 6.935) with the remaining 92 % of sites being assigned ω values close to 0 (ω = 0.054) (**Table 4**).

For the *DRA*, M3 model estimated that only 6.6 % of codon sites had ω values greater than one (ω = 10.031) with the remaining 93.4 % of sites being assigned ω values close to 0 (**Table 4**). Moreover, posterior means of ω across *DRA* codon sites estimated by models M2a (ω=10.286) and M8 (ω=10.323), predicted that nine codons (positions 12, 14, 15, 16, 18, 19, 49, 64, 68) were under significant positive selection. But, only two sites (19 and 49 sites) were also known as putative ABS based on the HLA equivalents (**Fig**. 4).

Furthermore, we found that M3 model estimated that 14.6 % of codon sites had ω values greater than one (ω1=1.351, ω2=6.823) at the *DRB* (**Table 4**), and higher than other MHC calss II *DQA*, *DQB* and *DRA*. The posterior means of ω across *DRB* codon sites estimated by models M2a (ω=5.972) and M8 (ω=5.961), and predicted that twelve codons (positions 1, 2, 23, 28, 47, 48, 58, 61, 62, 65, 69, 77) were under significant positive selection, seven (2, 28, 48, 58, 61, 69, and 77) of which are also putative ABS based on the HLA equivalents (**Fig**. 4).

Evolutionary analysis

The haplotype median network of *DQA*, *DQB*, *DRA* and *DRB* between GZP and known horse species (*Eqca*, *E.callabus*; *Eqpr, E.przewalski*; *Eqki*, *E.kiang*; *Eqgr*, *E.grevyi*; *Eqas*, *E.asinus*; *Eqbu*, *E.burchelli*; *Eqze, E.zebra*; *Eqhe*, *E.hemionus*.) in GenBank were plotted by software NetWork 4.6. Owing to the presence of loops in the network, we could not determine the genealogy of the *DQA*, *DQB*, *DRA* and *DRB* (**Fig** 5.A-D). Therefore, some alleles seemed more likely to be ancestor based on their internal position in the network and on their greater numbers of mutational connections. At the *DQA* locus, these alleles seemed more likely to be ancestor, including *DQA1*, *DQA3*, *Eqca17* and *Eqca18*. Among these, allele *DQA1* appears to be ancestral for most alleles, namely *DQA12*, *DQA13*, *DQA14*, *DQA15*, *DQA9*, *Eqca10*, *Eqbu6*, *Eqca20*, *Eqas1*, *Eqas2*, *Eqbu5*, *Eqca19*, *Eqbu4*, *Eqbu12*. Three haplotypes of Przewalski wild horse *Eqpr 3*, *Eqpr4* and *Eqpr2*, were separated from *DQA3* by two mutational steps and are most closely related to GZP haplotypes. Meanwhile, the *DQA1* allele was shared among four species (*Eqca15*, *Eqgr1*, *Eqbu2* and *Eqze1*), *DQA2* was shared with *Eqca14*, and *DQA3* were shared between two species (*Eqca16* and *Eqbu20*). At *DQB* locus, allele *DQB1* appears to be ancestral for most alleles, including *DQB31*, *DQB23*, *DQB32*, *DQB10*, *DQB30*, *DQB33*, *DQB29*, *DQB3*, *DQB24*, *DQB34*, *DQB40*, *Eqas7* and *Eqas4*. We found that haplotype *DQB5* and *DQB13* were shared between *Eqca1* and *Eqca7*, respectively. For *DRA* locus, allele *DRA5* was shared between *Eqbu7* and *Eqca5*, and *DRA1*

Deleted: the

Commented [HX28]: move to methods

Commented [HX29]: ancestral?

Commented [HX30]: ancestral

Deleted: as

Deleted: as

Deleted: were

was shared with *DRA3*. Interestingly, allele *DRA1* seemed more likely to be ancestral for vast alleles, containing twenty alleles of *DRA21*, *DRA15*, *DRA13*, *DRA8*, *DRA18*, *DRA7*, *DRA14*, *DRA6*, *DRA20*, *DRA19*, *DRA17*, *DRA16*, *DRA2*, *DRA11*, *DRA10*, *DRA12*, *Eqca2*, *Eqca6*, *Eqca7*, and *Eqca8*. Haplotypes *Eqhe* and *DRA9* were separated from *DRA1* by as two mutations step as are most closely related GZP haplotypes. Allele *DRA5* seemed more likely to be ancestral, *Eqbu*, *Eqze*, *Eqgr, Egas* were separated from *DRA5* by as one or two mutations step and are most closely related to GZP haplotypes. Most of *DRB* alleles were found to disperse in the whole network, and much more genetic relationship presented between GZP and the other horse species. Wild ass haplotypes *Eqas3*, *Eqas4* and *Eqas6* were separated from *DRB28* by as one mutational step as are most closely related GZP haplotypes. Furthermore, the haplotypes *DRB2* (*Eqpr1*) and *DRB3* (*Eqpr2*) were shared by GZP and Przewalski's horse. The haplotypes *DRB1* (*Eqca5*), *DRB2* (*Eqca12*), *DRB4* (*Eqca7*), *DRB5* (*Eqca1*), *DRB15* (*Eqca2*) and *DRB23* (*Eqca8*), were shared between GZP and European horse.

DISCUSSION

This study found elevated levels of diversity at four ELA class II gene regions, *DQA*, *DQB*, *DRA* and *DRB*, with the contribution of many novel alleles identified from GZP. The data provided evidence for within-species variation in the numbers of alleles, in which total of 21 *DQA* alleles, 45 *DQB* alleles, 22 *DRA* alleles and 31 *DRB* alleles were unequivocally identified from the GZP.

In four GZP loci, the *DRA* locus was relatively conservative compared with the other three loci. The alignments of *DQA*, *DQB*, and *DRB* genes revealed considerable sequence diversity, in contrast, *DRA* contained much lower diversity of nucleotide. It is consistent with the level of nucleotide diversity at genus Equus *ELA* genes reported by Kamath *et al* (2011). The polymorphism of *DQB* was the highest, the ratio of polymorphic sites was 46.08 %, and followed by *DRB* and *DQA*. The polymorphism of *DRA* was the lowest level, only 15.04 %, which was consistent with the results of *DRA* locus in dogs (Wagner et al., 1999), cats (Yuhki et al., 1995), goats (Takada et al., 1998) and pigs (Chardon et al., 1999). It is reported that the ELA gene diversity was important for immune functions involving the resistance and susceptibility to pathogens (Trowsdale et al., 2004), with a probable mechanism of gene selection in the evolution process of pony (Penn et al., 1999).

We detected the balancing selection events by determining the rate of non-synonymous/synonymous

Commented [HX31]: not clear what this means

Deleted: as

Deleted: Kamath et al.,

substitutions (d_N/d_S ratio) of nucleotides, as well as d_N/d_S test. It revealed a high genetic variability at *DQA*, *DQB*, *DRA*, and *DRB* loci. Similar to the low polymorphism detected by sequence alignment, d_N/d_S ratio (d_N/d_S =0.560) at the *DRA* locus was the lowest. It is well known that one reason for the number of synonymous substitutions greater than that of non-synonymous is due to strong functional and structural constraints on the protein (Kamath et al., 2011). The low level of polymorphism observed from locus *DRA* might be attributed to the high selective pressure during the horse survival (Albright-Fraser et al., 1996).

The selection effective of four *ELA* loci were further analyzed by using PAML program. We found that nine *DQA* codons, five *DQB* codons, nine *DRA* codons and seven *DRB* codons under significant positive selection, and most of them were predicted to be the ABS of *ELA*. Based on swiss-model prediction (**Fig.** 4), these amino acids under site-specific selection located on the protein surface, which consisted of the inner surface of MHC cleft to bind peptide in the antigen presentation (Madden et al.,1995). Several reports indicated that the diversity and nonsynonymous mutations at the ABS site could increase the host ability to recognize pathogens (Hughes et al.,1988; Hughes et al., 1989). Obviously, the different rates of non-synonymous and synonymous substitutions in *DQA*, *DQB*, *DRA* and *DRB* had a close connection with the ABS changes in the GZP. In particular, the d_N/d_S ratio in the ABS site was greater than that in the non-ABS region at the *DQA*, *DQB* and *DRB* loci, and this feature is common in Argentine Creole horse (Díaz et al., 2001). The d_N/d_S ratio of ABS, higher than the other regions, has been explained by balancing selection (Albright-Fraser et al., 1996), and the positive selection results in MHC polymorphism (Yang et al., 2000).

The diversity of *ELA* genes might harbor the evolutionary relationship of GZP with the other horse. The haplotype median network of *DQA*, *DQB*, *DRA* and *DRB* between GZP and the reported horses (*Eqca*, *E. callabus*; *Eqpr*, *E. przewalski*; *Eqki*, *E. kiang*; *Eqgr*, *E. grevyi*; *Eqas*, *E. asinus*; *Eqbu*, *E. burchelli*; *Eqze*, *E. zebra*; *Eqhe*, *E. hemionus*) were built by NetWork program. Among these, several wild ass haplotypes were separated from *DQA1*, *DQB1*, and *DRB28* by as one or two mutational steps and are most closely related to GZP haplotypes. The divergence time between horse and ass has been estimated to be 0.88–2.3 Ma (Krüger et al. 2005). One *E. hemionus* haplotype (*Eqhe2*) was separated from *DRA* by as two mutational steps and is most closely related to GZP haplotypes. It suggested that *DQA1*, *DQA3*, *DQB1*, *DRA1*, *DRA5*, and *DRB28* might be the oldest alleles. The haplotype *DRB2* and *DRB3* were shared between GZP and Przewalski horse at the

Commented [HX32]: probably not necessary

Deleted: the rates of

Commented [HX33]: add some more detail here. What selective pressure? Disease outbreaks?

Commented [HX34]: Methodological detail probably not needed

Commented [HX35]: repeat

Commented [HX36]: Is it obvious? Perhaps 'these data suggest that'

Commented [HX37]: repeat

Deleted: ly

Commented [HX38]: not clear what this means

Commented [HX39]: methodological detail not needed

Deleted: as

Formatted: Font: Italic

Deleted: as are

Deleted: in

DRB locus. Przewalski haplotypes *Eqpr3*, *Eqpr4* and *Eqpr2* were separated from *DQA3* by two mutational steps. The Przewalski horses were discovered in the Asian steppes in the 1870s, which is the only survived species of wild horse in the world up to now (Wakefield et al., 2002). It is supposed that the Przewalski horse and domesticated horse populations separated from each other about 45,000 years ago and maintained a certain degree of gene-flow for a long time (Der Sarkissian et al., 2015). Meanwhile, some haplotypes were shared between the GZP and European horses, including *DQA1*, *DQA3*, *DQB5*, *DQB13*, *DRB1*, *DRB2*, *DRB15*, *DRB23*, *DRB4*, and *DRB5*. Moreover, allele *DQA1* appears to be the ancestor for three alleles contained *Eqca10*, *Eqca20*, and *Eqca19*. Allele *DRA1* seemed more likely to be ancestral for four alleles included *Eqca2*, *Eqca6*, *Eqca7, Eqca8*. The genes of domesticated Asian horse might also have dispersed into European populations because of the gene flow (Bjørnstade al., 2003). Interestingly, haplotype *DQA1*, *DQA3*, *DRA5* were shared between the GZP and *E. burchelli, E. grevyi* and zebra. The divergence time between horses and zebras based on microsatellite trees is estimated in 0.86–2.3 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 zebra, ass and horses might have occurred within the following 0.5 My (George et al., 1986). These data and our results thus indicate that the GZP is an ancient variety of equid after long-time of adaptive evolution in GZP. Furthermore, some unique haplotypes of GZP appear gradually in the evolutionary process, and correspond to specific adaptation under the local environmental pressures such as the unique sets of pathogenic microorganisms present in the mountainous and humid districts in Guizhou province, China.

CONCLUSION

Plenty of nucleotide diversity was detected from the exon 2 of *ELA-DQA*, *DQB*, *DRA*, *DRB* genes in the GZP by direct sequencing technology. Of those four loci, the *DRA* locus was relatively conservative. Many codons in the antigen binding sites (ABS) underwent positive selection, including nine *DQA* codons, five *DQB* codons, nine *DRA* codons and seven *DRB* codons. Those amino acids coded by selected codons would stand on the inner surface of cleft of *ELA* complex to bind to antigen peptides. The selected sites might be related to the GZP ability to defend against foreign pathogens from the surrounding environment. Furthermore, many ancient alleles were detected at the *DQA*, *DQB*, *DRA* and *DQB* gene regions of GZP. Two older haplotypes at *DRB* (*DRB2* and *DRB3*) were shared by the GZP and Przewalski horse, two older haplotypes at *DRA* (*DRA1* and

Deleted: as

Deleted: still have kee

Deleted: p

Deleted: extent

Deleted: And, a

Deleted: exchange

Formatted: Font: Italic

Deleted: take place

Deleted: next

Commented [HX40]: not sure what this means, it's a little vague

Deleted: d

Deleted: It's

Commented [HX41]: need clarification.

Commented [HX42]: Do your data allow you to conclude this?

Commented [HX43]: Vage term

Deleted: were

Commented [HX44]: "possessed the lowest diversity"?

Deleted: defense

Deleted: at

Deleted: of habitat

Deleted: from

DRA5) seemed separate from *Eqbu*, *Eqze*, *Eqgr*, *Egas* by one or two mutations step, and four older haplotypes of GZP (*DQA1*, *DQA3*, *DQB1*, and *DRB28*) were much close to wild ass, Przewalski horse by only one or two mutational steps. It indicated that the indigenous breed, GZP, retained ancient haplotyes in *ELA* genes. Meanwhile, a fairly large number of unique haplotypes dispersed in GZP crowns might have been acquired in the long process of ELA molecule evolution. For a long time, the unique genetic characteristics of GZP are unclear, and thus it is undervalued and frequently confused with other ponies. The genetic uniqueness revealed in this study is helpful to understand its genetic conservation of the ancient variety.

Deleted: as

Acknowledgments

The Talents of Guizhou Science and Technology Cooperation Platform (QKHPTRC[2019]-5615), The Guizhou Province Hundred Innovative Talents Project (QKHRC[2016]-4012], and the Guizhou Agriculture Research program (QKHNYZ[2012]3009, QKHNYZ[2009]3068).

Competing interests

The authors declare that they have no competing interests.

Author Contributions

Chang Liu performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables and wrote drafts of the paper. Hongmei Lei performed the experiments. Xueqin Ran and JiafuWang conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Supplementary Materials

Table S1. The nucleotide of the *DQA*, *DQB*, *DRA*, and *DRB* locus (XLSX)

Table S2. Nucleotide composition of *DQA*, *DQB*, *DRA*, and *DRB* in Guizhou pony (XLSX)

References

Albright-Fraser DG, Reid R, Gerber V, Bailey E. 1996. Polymorphism of DRA among equids. Immunogenetics 43(5):315-7.

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting

adaptive molecular evolution. Mol Biol Evol. (8):1585-92.

Anisimova M, Nielsen R, Yang Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics (3):1229-36.

Ansari HA, Hediger R, Fries R, Stranzinger G.1988. Chromosomal localization of the Major Histocompatibility Complex of the horse (ELA) by insitu hybridization. Immunogenetics 28(5):362-364.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies.MolBiolEvol 16(1):37-48 DOI: 10.1093/oxfordjournals.molbev.a026036.

Bernatchez L, Landry C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol 16(3):363-77.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014;42:W252-8.

Bielawski JP, Yang Z. 2003. Maximum likelihood methods for detecting adaptive evolution after gene duplication. J Struct Funct Genomics 3(1-4):201-12.

Bjørnstad G, Nilsen NØ, Røed KH. 2003. Genetic relationship between Mongolian and Norwegian horses? Anim Genet 34(1):55-8.

Bolnick DI, Stutz WE. 2017. Frequency dependence limits divergent evolution by favouring rare immigrants over residents. Nature. 2017;546(7657):285-288.

Bontrop RE, Otting N, de Groot NG, Doxiadis GG. 1999. Major histocompatibility complex class II polymorphisms in primates. Immunol Rev 167:339-50.

Brown JJ, Thomson W, Clegg P, Eyre S, Kennedy LJ, Matthews J, Carter S, Ollier WE. 2004. Polymorphisms of the equine major histocompatibility complex class II DRA locus. Tissue Antigens 64(2):173-9 DOI: 10.1111/j.1399-0039.2004.00269.x.

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 333(6175):786 DOI: 10.1038/332845a0.

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC.1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature. 364(6432):33-9. DOI: 10.1038/364033a0

Buhler S, Nunes JM, Sanchez-Mazas A. 2016. HLA class I molecular variation and peptide-binding properties suggest a model of joint divergent asymmetric selection. Immunogenetics 68(6-7):401-416.

Chardon P, Renard C, Vaiman M. **1999.** The major histocompatibility complex in swine. Immunol Rev 167:179-92.

Cutrera AP, Lacey EA .2007. Trans-species polymorphism and evidence of selection on class II MHC loci in tuco-tucos (Rodentia: Ctenomyidae). Immunogenetics 59(12):937-48 DOI: 10.1007/s00251-007-0261-3.

Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, Fumagalli M, Jónsson H, Bar-Gal GK, Albrechtsen A, Vieira FG, Petersen B, Ginolhac A, Seguin-Orlando A, Magnussen K, Fages A, Gamba C, Lorente-Galdos B, Polani S, Steiner C, Neuditschko M, Jagannathan V, Feh C, Greenblatt CL, Ludwig A, Abramson NI, Zimmermann W, Schafberg R, Tikhonov A, Sicheritz-Ponten T, Willerslev E, Marques-Bonet T, Ryder OA, McCue M, Rieder S, Leeb T, Slatkin M, Orlando L. 2015. Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. Curr Biol 25(19):2577-83 DOI: 10.1016/j.cub.2015.08.032.

Díaz S, Giovambattista G, Dulout FN, Peral-García P. 2001. Genetic variation of the second exon of ELA-DRB genes in Argentine Creole horses. Anim Genet 32(5):257-63.

Edwards SV, Hedrick PW.1988. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol Evol 13(8):305-11.

Eizaguirre C, Lenz TL, Kalbe M, Milinski M.2012. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nat Commun 3:621.

Figueroa F, Günther E, Klein J. 1988. MHC polymorphism pre-dating speciation. Nature 335(6187):265-7 DOI: 10.1038/335265a0.

Fraser DG, Bailey E.1996. Demonstration of three DRB loci in a domestic horse family. Immuno genetics 44(6):441-5 DOI: 10.1007/bf02602805.

Fraser DG, Bailey E.1998. Polymorphism and multiple loci for the horse DQA gene. Immunogenetics 1998,

47(6):487-490.

George M Jr, Ryder OA. 1986. Mitochondrial DNA evolution in the genus Equus. Mol Biol Evol 3(6):535-46 DOI: 10.1093/oxfordjournals.molbev.a040414.

Hill AV. 2001. The genomics and genetics of human infectious disease susceptibility. Annu Rev Genomics Hum Genet 2:373-400 DOI: 10.1146/annurev.genom.2.1.373.

Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335(6186):167-70 DOI: 10.1038/335167a0.

Hughes AL, Nei M.1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl AcadSci U S A 86(3):958-62 DOI: 10.1073/pnas.86.3.958.

Hughes AL, Hughes MK. 1995. Natural selection on the peptide-binding regions of major histocompatibility complex molecules. Immuno genetics 42(4):233-43 DOI: 10.1007/bf00176440.

Janova E, Matiasovic J, Vahala J, Vodicka R, Van Dyk E, Horin P. 2009. Polymorphism and selection in the major histocompatibility complex DRA and DQA genes in the family Equidae. Immunogenetics 61(7):513-27 DOI: 10.1007/s00251-009-0380-0.DOI: 10.1007/s00251-009-0380-0.

Kamath PL, Getz WM. 2011. Adaptive molecular evolution of the Major Histocompatibility Complex genes, DRA and DQA, in the genus Equus. BMC Evol Biol 11:128. DOI: 10.1186/1471-2148-11-128.

Klein J, Satta Y, Takahata N, O'hUigin C.1993. Trans-specific Mhc polymorphism and the origin of species in primates. J Med Primatol 22(1):57-64.

Kloch A, Babik W, Bajer A, Siński E, Radwan J. 2010. Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole Myodesglareolus. Mol Ecol 19 Suppl 1:255-65.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 33(7):1870-4. DOI: 10.1093/molbev/msw054.

Krüger K, Gaillard C, Stranzinger G, Rieder S. 2005. Phylogenetic analysis and species allocation of individual equids using microsatellite data. 122 Suppl 1:78-86.

Kubinak JL, Ruff JS, Hyzer CW, Slev PR, Potts WK. 2012. Experimental Viral Evolution to Specific Host MHC Genotypes Reveals Fitness and Virulence Trade-Offs in Alternative MHC Types. Proc Natl Acad Sci U S A 109(9):3422-7.

Landry C, Bernatchez L. 2001. Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (Salmo salar). Mol Ecol 10(10):2525-39.

Lau Q, Chow N, Gray R, Gongora J, Higgins DP. 2015. Diversity of MHC DQB and DRB Genes in the Endangered Australian Sea Lion (Neophoca cinerea). J Hered 106(4):395-402.

Lenz TL, Wells K, Pfeiffer M, Sommer S. 2009. Diverse MHC IIB Allele Repertoire Increases Parasite Resistance and Body Condition in the Long-tailed Giant Rat (Leopoldamys Sabanus). BMC Evol Biol 9:269.

Lenz TL, Eizaguirre C, Kalbe M, Milinski M. 2013. Evaluating patterns of convergent evolution and trans-species polymorphism at MHC immunogenes in two sympatric stickleback species. Evolution 67(8):2400-12.

Lenz TL. 2011. Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. Evolution. 2011 ;65(8):2380-90.

Lewin HA, Russell GC, Glass EJ. 1999. Comparative organization and function of the major histocompatibility complex of domesticated cattle. Immunol Rev. 167:145-58.

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(1):69-72.

Mäkinen A, Chowdhary B, Mahdy E, Andersson L, Gustavsson I. **1989.**Localizationof the equine Major Histocompatibility Complex (ELA) tochromosome-20 by insitu hybridization. Hereditas 1989, 110(1):93-96.

Madden KS, Felten SY, Felten DL, Bellinger DL.1995. Sympathetic nervous system--immune system interactions in young and old Fischer 344 rats. Ann N Y Acad Sci 771:523-34.

Mashima S. 2003. Comparative sequence analysis of equine and human MHC class II DQB genes. Cytogenet Genome Res 102(1-4):196-200 DOI: 10.1159/000075748.

McGahern AM, Edwards CJ, Bower MA, Heffernan A, Park SD, Brophy PO, Bradley DG, MacHugh DE, Hill EW. 2006. Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. Anim Genet 37(5):498-502DOI: 10.1111/j.1365-2052.2006.01506.x.

Meyer-Lucht Y, Sommer S. 2005. MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*).Mol Ecol 14(7):2233-43 DOI: 10.1111/j.1365-294X.2005.02557.x.

Miller D, Tallmadge RL, Binns M, Zhu B, Mohamoud YA, Ahmed A, Brooks SA, Antczak DF. 2017. Polymorphism at expressed DQ and DR loci in five common equine MHC haplotypes. Immunogenetics 69(3):145-15 DOI: 10.1007/s00251-016-0964-4.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3(5):418-26 DOI: 10.1093/oxfordjournals.molbev.a040410.

O'Connor SL, Blasky AJ, Pendley CJ, Becker EA, Wiseman RW, Karl JA, Hughes AL, O'Connor DH. 2007. Comprehensive characterization of MHC class II haplotypes in Mauritian cynomolgus macaques. Immunogenetics 59(6):449-62.

Paterson S, Wilson K, Pemberton JM. 1998. Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population. Proc Natl AcadSci U S A 95(7):3714-9 DOI: 10.1073/pnas.95.7.3714.

Piertney SB, Oliver MK. 2006. The evolutionary ecology of the major histocompatibility complex.Heredity (Edinb) 96(1):7-21 DOI: 10.1038/sj.hdy.6800724.

Penn DJ, Potts WK. 1999. The Evolution of Mating Preferences and Major Histocompatibility Complex Genes. Am Nat 153(2):145-164 DOI: 10.1086/303166.

Phillips KP, Cable J, Mohammed RS, Herdegen-Radwan M, Raubic J, Przesmycka KJ, van Oosterhout C, Radwan J. 2018. Immunogenetic novelty confers a selective advantage in host-pathogen coevolution. Proc Natl Acad Sci U S A 115(7):1552-1557.

Pierini F, Lenz TL. 2018. Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection. Mol Biol Evol 35(9):2145-2158.

Prystupa JM, Hind P, Cothran EG, Plante Y. 2012. Maternal lineages in native Canadian equine populations and their relationship to the Nordic and Mountain and Moorland pony breeds. J Hered 103(3):380-90 DOI: 10.1093/jhered/ess003 DOI: 10.1093/jhered/ess003.

Reed DH, Frankham R. 2003. Correlation between fitness and genetic diversity. Conservation Biology 17: 230–237.

Richman AD, Herrera LG, Nash D. 2001. MHC class II beta sequence diversity in the deer mouse (Peromyscus maniculatus): implications for models of balancing selection. Mol Ecol 10(12):2765-73.

Schad J, Ganzhorn JU, Sommer S. 2005. Parasite burden and constitution of major histocompatibility complex in the Malagasy mouse lemur, *Microcebus murinus*. Evolution 59(2):439-50.

Schwensow N, Dausmann K, Eberle M, Fietz J, Sommer S.2010. Functional associations of similar MHC alleles and shared parasite species in two sympatric lemurs. Infect Genet Evol 10(5):662-8.

Sommer S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16 DOI: 10.1186/1742-9994-2-16.

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. Immuno genetics. 48(6):408-12 DOI: 10.1007/s002510050452.

Takahata N, Nei M. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124(4):967-78.

Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E, Funkhouser R, Fugate M, Theiler J, Hsu YS, Kunstman K, Wu S, Phair J, Erlich H, Wolinsky S. 2003. Advantage of rare HLA supertype in HIV disease progression. Nat Med 9(7):928-35.

Trowsdale J, Parham P. 2004. Mini-review: defense strategies and immunity-related genes. Eur J Immunol 34(1):7-17 DOI: 10.1002/eji.200324693.

Singh VK, Mangalam AK, Dwivedi S, Naik S. 1998. Primer premier: program for design of degenerate primers from a protein sequence. Biotechniques 24(2):318-9.

Wagner JL, Burnett RC, Storb R. 1999. Organization of the canine major histocompatibility complex: current perspectives. J Hered 90(1):35-8 DOI: 10.1093/jhered/90.1.35.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T.2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46(W1):W296-W303.

Wakefield S, Knowles J, Zimmermann W, van Dierendonck M. 2002. Chapter 7: Status and Action Plan for the Przewalski's Horse (Equus ferus przewalskii) In: Moehlman PD, editor. Equids: Zebras, Asses,

and Horses: Status Survey and Conservation Action Plan. IUCN/SCC Equid Specialist Group, IUCN (The World Conservation Union); Gland Switzerland and Cambridge.

Westerdahl H, Wittzell H, von Schantz T, Bensch S. 2004. MHC Class I Typing in a Songbird With Numerous Loci and High Polymorphism Using Motif-Specific PCR and DGGE. Heredity (Edinb) 92(6):534-42.

Wu TT, Kabat EA. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J Exp Med 132(2):211-50 DOI: 10.1084/jem.132.2.211.

Yuhki N, O'Brien SJ. 1997. Nature and origin of polymorphism in feline MHC class II DRA and DRB genes. J Immunol 158(6):2822-33.

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15(12):496-503.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24(8):1586-91 DOI: 10.1093/molbev/msm088.

Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155(1):431-49.