Asian Herpetological Research 2014, 5(2): 91–103
DOI: 10.3724/SP.J.1245.2014.00091

# MHC Class I Exon 4 in the Multiocellated Racerunners (*Eremias multiocellata*): Polymorphism, Duplication and Selection

Xiuyun YUAN[1,2], Xiaomao ZENG[1] and Xianguang GUO[1*]

[1] *Chengdu Institute of Biology*, *Chinese Academy of Sciences*, *Chengdu* 610041, *Sichuan*, *China*
[2] *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*

**Abstract** The major histocompatibility complex (MHC) is a dynamic genetic region with an essential role in the adaptive immunity of jawed vertebrates. The MHC polymorphism is affected by many processes such as birth-and-death evolution, gene conversion, and concerted evolution. Studies investigating the evolution of MHC class I genes have been biased toward a few particular taxa and model species. However, the investigation of this region in non-avian reptiles is still in its infancy. We present the first characterization of MHC class I genes in a species from the family Lacertidae. We assessed genetic diversity and a role of selection in shaping the diversity of MHC class I exon 4 among 37 individuals of *Eremias multiocellata* from a population in Lanzhou, China. We generated 67 distinct DNA sequences using cloning and sequencing methods, and identified 36 putative functional variants as well as two putative pseudogene-variants. We found the number of variants within an individual varying between two and seven, indicating that there are at least four MHC class I loci in this species. Gene duplication plays a role in increasing copy numbers of MHC genes and allelic diversity in this species. The class I exon 4 sequences are characteristic of low nucleotide diversity. No signal of recombination is detected, but purifying selection is detected in $\beta$2-microglobulin interaction sites and some other silent sites outside of the function-constraint regions. Certain identical alleles are shared by *Eremias multiocellata* and *E. przewalskii* and *E. brenchleyi*, suggesting trans-species polymorphism. The data are compatible with a birth-and-death model of evolution.

**Keywords** MHC class I, *Eremias multiocellata*, trans-species polymorphism, balancing selection, purifying selection, birth-and-death evolution

## 1. Introduction

The genetic region that scientists today call the major histocompatibility complex (MHC) was discovered by Gorer (1936) in his pioneering studies of antigenic responses to transplanted sera by inbred mouse strains. The MHC was genetically defined more precisely by Snell (1948), who first used the term. It has been recognized as a large multigene family present in all jawed vertebrates (Danchin *et al.*, 2004; Kelley *et al.*, 2005). During the last two decades, MHC genes, especially class I and class II genes, have been the subjects of the majority of research in the fields of ecology and evolution (reviewed in Sommer, 2005; Milinski, 2006; Piertney and Oliver, 2006; Ujvari and Belov, 2011). This is because MHC includes the most polymorphic genes in vertebrate populations. Given the extraordinary richness and diversity of continuously evolving pathogens in the environment, it is not surprising that the MHC harbors the most polymorphic genes described thus far, with some loci, such as the human HLA-B locus, possessing more than 2000 alleles (de Bakker and Raychaudhuri, 2012). The number of genes in the MHC varies enormously between species, even between individuals of the same species (Malaga-Trillo *et al.*, 1998; Figueroa *et al.*, 2001). It has been noted that balancing selection is an evolutionary force that can maintain many haplotypes over long periods of evolutionary time. Genes that

* Corresponding author: Dr. Xianguang GUO, from Chengdu Institute of Biology, Chinese Academy of Sciences, with his research focusing on molecular phylogenetics, biogeography, and evolutionary history of lizards.
E-mail: guoxg@cib.ac.cn
Received: 1 March 2014 Accepted: 6 June 2014

are evolving under balancing selection are distinct from neutral loci in that they show high levels of heterozygosity and a large number of alleles at similar frequencies (Maruyama and Nei, 1981). Some alleles can be maintained over great lengths of time, possibly exceeding speciation events. Phylogenetic reconstructions therefore reveal trans-species polymorphism (Klein, 1987; Klein *et al.*, 2007), where alleles between species are more closely related (or even identical) than alleles within species. Phylogenetic reconstructions can also reveal orthologous relationships, by which sequences group by gene rather than by species, forming orthologous gene clusters (Nei *et al.*, 1997). In orthologous clusters, the divergence pattern of genes reflects species phylogeny (when trans-species evolution does not also occur) because genes have diverged from a common ancestor due to a speciation event (Fitch, 2000). The evolution of MHC is also affected by gene conversion (Spurgin *et al*., 2011), concerted evolution, and birth-and-death evolution (Nei and Rooney, 2005). These processes can explain the considerable amount of copy number variation that exists among species (Mehta *et al.,* 2009), within species (Bonhomme *et al.*, 2008), and within populations (Eimes *et al.*, 2011).

The MHC class I (MHC-I) genes encode a single polypeptide α chain and consist of seven exons. Among them, exons 2 to 4 encode three extracellular domains (α1, α2, and α3, respectively). The α1 and α2 domains are highly polymorphic and are responsible for binding to the antigenic peptide (Bjorkman *et al.*, 1987). The α3 domain is the least polymorphic domain, with functional constraint of binding *β*2-microglobulin non-covalently and interacting with CD8 molecules. So far, data from the α3 domain are frequently used to recover evolutionary relationships among class I sequences across species because they are relatively conserved in comparison to the two peptide-binding domains (α1, α2), which are often under balancing selection. MHC class I has been well-studied in mammals (Kelley *et al.*, 2005), birds (Shiina *et al.*, 2004), and fishes (Shand and Dixon, 2001; Kulski *et al*., 2002). However, there is still limited knowledge of the MHC in non-avian reptiles, particularly in the key group of squamates. This is likely attributed to the fact that the isolation and characterization of MHC genes still remains a challenging and time-consuming task in non-model species (Babik, 2010).

MHC-I research has not been evenly distributed across vertebrate taxa, but studies from many underrepresented groups such as amphibians and non-avian reptiles are slowly emerging. In a pioneering paper, Grossberger and Parham (1992) isolated the MHC-I sequences from the Giant Ameiva (*Ameiva ameiva*) and Northern Water Snake (*Nerodia sipedon*), and revealed the conserved elements in reptilian class I structure. Recently, MHC-I evolution has been studied at the population level in some non-avian reptiles, such as sand lizard (*Lacerta agilis*) and Northern Viper (*Vipera berus*) (Madsen *et al.*, 2000), tuatara (Miller *et al.*, 2007), saltwater crocodiles (*Crocodylus porosus*) (Jaratlerdsiri *et al.*, 2012), and loggerhead sea turtle (*Caretta caretta*) (Stiebens *et al.*, 2013). Few studies of MHC-I have been performed among closely related species of non-avian reptiles and have been limited to six species of geckos (Radtkey *et al.*, 1996) and three species of iguanas (Glaberman and Caccone, 2008). Geckos have MHC-I genes differentiated by point mutations (changes at the nucleotide level) (Radtkey *et al.*, 1996), while each of the iguana species has at least two to three MHC-I genes showing species-specific evolution in exon 4 (Glaberman and Caccone, 2008). These complex patterns and mechanisms in the evolution of squamates MHC-I genes suggest that more species should be assessed to enrich our understanding of the evolution of squamates MHC-I genes. Consequently, studies examining MHC evolution of a greater number of lizard species may be pivotal in better understanding the evolutionary process of adaptive immunity in jawed vertebrates.

Racerunner lizards of the genus *Eremias* (Family Lacertidae) are the dominant reptiles in Central Asia deserts and steppes (Szczerbak, 2003). Here, we selected a widespread viviparous species, the Multiocellated Racerunner (*E. multiocellata*), to study the structure, variation and evolutionary mechanisms acting on MHC-I exon 4. Our primary objectives were to (i) characterize the genetic diversity of alleles and infer putative numbers of loci for *E. multiocellata*; (ii) screen for recombination in the exon 4 sequences; (iii) determine if diversifying selection or purifying selection has acted on racerunner lizards MHC-I exon 4 (including what codons may have experienced diversifying selection or purifying selection). As exon 4 displays strong functional constraints, we predict there is a signal of purifying selection to be detected. Additionally, we have combined some data from two closely related congeneric species to further investigate if there is species-specific evolution in MHC-I exon 4 in *E. multiocellata*. If exon 4 appears to be evolving in a species-specific manner in *Eremias*, we would expect different MHC-I exon 4 variants are grouped in the phylogenetic tree by species.

## 2. Materials and Methods

**2.1 Samples collection and DNA extraction** In total, 37 individuals of *E. multiocellata* were sampled from Lanzhou, Gansu Province, China (N 36º07'14.7'', E 103º47'53.9''). All voucher specimens were deposited in the herpetological collections of Chengdu Institute of Biology, Chinese Academy of Sciences. The liver tissue was taken and stored in 95% ethanol at –20°C until DNA extraction. Genomic DNA was extracted using the EasyPure® Genomic DNA Kit (TransGen Biotech, Beijing, China).

**2.2 PCR amplification, cloning, SSCP and sequencing** A 183 bp fragment of MHC-I loci exon 4 was amplified using the primers MHCIF (5'-GCCGAGTTCACGGCTTCTACCCC-3') and MHCIR (5'-CATGCTCCACGTGGCACTGGTA -3') from Murphy *et al*. (2009). Each of our 50 μL PCR reactions contained 25 μL of 2 × EasyTaq SuperMix (TransGen Biotech, Beijing, China), 0.4 μM of each primer, and 1–2 μL genomic DNA. The PCR protocol involved initial denaturation at 94°C for 180 s followed by 30 cycles of 94°C for 30 s, 60°C for 30 s, and elongation at 72°C for 45 s; and final extension at 72°C for 8 min. Negative controls were run for all amplifications. Amplified products were visualized on 2% agarose gel stained by ethidium bromide and purified using the EasyPure Quick Gel Extraction Kit (TianGen Biotech, Beijing, China).

Purified products were cloned using *Escherichia coli* DH5α (TransGen Biotech, Beijing, China) competent cells, using pEASY-T1 vector (TransGen Biotech, Beijing, China) according to the manufacturer's instructions. Clones were picked from plates and dipped directly into 25 μL PCR reactions containing 10 pmole of M13 primers, 12.5μL of 2 × EasyTaq SuperMix (TransGen Biotech, Beijing, China), and the products were resolved on agarose gel. Inserts were then amplified from positive clones using the MHC primers, so that orientation of the insert did not affect subsequent applications. These products were a suitable size for allele identification using Single Strand Conformation Polymorphism (SSCP) method (Orita *et al.*, 1989), and ten clones per individual were randomly screened, which allowed us to sample the genomes of a relatively large number of lizards in the population. The SSCP method is most suitable for fragments of 100 bp–300 bp, in which it is able to detect the overwhelming majority of single-base substitutions (Sunnucks *et al.*, 2000). We performed the SSCP analysis following the protocol of Xu *et al.* (2007). Different conformations were verified as different alleles by sequencing PCR products from the initial colony screen using standard M13 primers with ABI Big Dye Terminator chemistry on an ABI 3730 automated sequencer. By comparing the SSCP profiles from different individuals, we can avoid multiple repeated sequencing of the same alleles.

**2.3 Sequence alignment and identification of true variants** Nucleotide sequences were assembled and their quality was checked using SEQMAN in DNASTAR (Burland, 2000) by generating a single sequence from the overlap of forward and reverse sequences. Sequences were screened for similarities to reptile MHC-I exon 4 using the basic local alignment search tool (http://www.ncbi.nlm.nih.gov). Alignments of nucleotide and deduced amino acid (aa) sequences were performed using MEGA 5.0 (Tamura *et al.*, 2011). Nucleotide sequences were translated into amino acids using a standard genetic code in MEGA 5.0. We also performed data screening to minimize the inclusion of PCR and cloning artifacts (spurious errors) in our data set. MHC sequences were confirmed when they were observed in at least two independent PCRs (Lukas and Vigilant, 2005; Cummings *et al.*, 2010). To avoid analysis of PCR artefacts, we only considered variants as 'true' alleles if they occurred in at least two individuals. Our analysis method is thus very conservative with regard to PCR errors. This stringency provides confidence about the polymorphism detected with this method.

**2.4 Sequence analyses and molecular diversity indices** Molecular diversity indices, including a pairwise nucleotide/aa difference between sequences and the number of synonymous and nonsynonymous substitution sites, were calculated in order to compare degrees of polymorphism between all MHC-I exon 4 variants generated in the current study. The pairwise differences between sequences and overall sequences were assessed using MEGA 5.0, while nucleotide diversity ($\pi$), number of segregating sites ($S$), and number of synonymous and nonsynonymous sites was counted using DnaSP v5 (Librado and Rozas, 2009). The frequency of individuals carrying a particular MHC variant amongst the 37 individuals was manually inspected. Putative pseudogenes were identified by the presence of premature stop codons and/or deletions, while putative functional genes were assessed by the presence of complete open reading fragments corresponding to MHC-I exon 4. Since the MHC-I exon 4 contained residues that interact with the $\beta$2-microglobulin domain and CD8 molecules, both of which facilitate binding between MHC molecule and

T-cells (Otten *et al.*, 1992; Apasov and Sitkovsky, 1993), we detected some functional amino acid sites according to the structure of HLA-A2 (Saper *et al.*, 1991) and MHC class I gene of some other species (e.g., Siddle *et al.*, 2006; Murphy *et al.*, 2009).

**2.5 Selection test** Given that recombination can mislead selection analyses if it is not accounted for, we run a recombination screen prior to performing selection analyses. The single-break point (SBP) method (Kosakovsky Pond *et al.*, 2006) in the HyPhy package (available at http://www.datamonkey.org) was chosen. Then we used several approaches to test if selection has acted on the MHC-I exon 4 evolution in *E. multiocellata*. First, non-synonymous($d_N$) and synonymous ($d_S$) substitution rates were calculated for the whole amino acid positions following the method of Nei and Gojobori using the Jukes-Cantor correction for multiple substitutions (Nei and Gojobori, 1986). These results were then used for the codon-based *Z*-test of purifying selection in MEGA, where the null hypothesis was that $d_N = d_S$ across all pairwise comparisons and the alternative hypothesis was $d_N < d_S$ (Tamura *et al.*, 2011). Value of *P* less than 0.05 is considered to be of sufficient significance to allow the null hypothesis to be rejected. Second, we tested whether historical selection had acted on each codon site using four likelihood tests: single likelihood ancestral counting (SLAC), fixed effects likelihood (FEL), random effects likelihood (REL), and maximum likelihood models of codon evolution in CODEML (implemented in PAML; Yang, 2007). SLAC, FEL, and REL are included in the HyPhy software package (hosted at the Datamonkey server; Delport *et al.*, 2010). SLAC is the most conservative, whereas REL is the most powerful (Kosakovsky Pond and Frost, 2005). In CODEML, we compared models allowing for positive selection (M2a and M8) with those assuming nearly neutral (M1a and M7) (Yang *et al.*, 2005). The tested models differ in the number and type of included parameters that are based on ω ratio, i.e. $d_N$ versus $d_S$. Comparison of nested models (M1a *vs.* M2a, and M7 *vs.* M8, respectively) was obtained using likelihood ratio test statistics. The Bayes empirical Bayes method was used to calculate posterior probabilities for site classes in models M2a and M8. If the posterior probabilities for some sites are significant ($\omega > 1$), those sites are inferred to be under positive selection.

To reconstruct the input tree used for all four likelihood methods, we first tested the alignment for the best-fit model of nucleotide evolution using the Datamonkey server model selection tool. This analysis compared the fit of over 200 nucleotide substitution models to the observed data using the Akaike information criterion. The tree was reconstructed by maximum likelihood approach using the PhyML 3.0 online web server (Guindon *et al.*, 2010). The BIONJ distance-based tree was used as the starting tree and HKY+G+I as the substitution model.

**2.6 Phylogenetic inference** In order to identify the number of major evolutionary lineages and/or subgroups that MHC-I variants may represent in *E. multiocellata*, phylogenetic analyses of all novel variants generated were conducted using Bayesian inference in MrBayes 3.2 (Ronquist and Huelsenbeck, 2012). Phylogenetic analyses were conducted using the novel alleles along with several outgroup sequences, including three alleles of Gobi Racerunner (*Eremias przewalskii*) and seven alleles of Ordos Racerunner (*Eremias brenchleyi*) (Yuan *et al.*, unpublished data) as well as the following vertebrate class I sequences from GenBank: Tuatara (*Sphenodon punctatus*), *Sppu-U*01* DQ145788, *Sppu-U*02* DQ145789; Galápagos marine iguana (*Amblyrhynchus cristatus*), *Amcr-UA* EU839664, *Amcr-UB*01* EU604308, *Amcr-UB*02* EU604309, *Amcr-UB*03* EU604310, *Amcr-UB*0401* EU604311, *Amcr-UB*0402* EU604312; Galápagos land iguana (*Conolophus subcristatus*), *Cosu-UB*0101* EU604313, *Cosu-UB*0102* EU604314, *Cosu-UB*02* EU604315, *Cosu-UB*03* EU604316; Green iguana (*Iguana iguana*), *Igig-UB*0101* EU604317, *Igig-UB*0102* EU604318, *Igig-UB*02* EU604319; Ameiva lizard (*Ameiva ameiva*), *LC13* M81096, *LC5* M81095; Green anole (*Anolis carolinensis*) XM_003227748; Northern Water Snake (*Nerodia sipedon*), SC1 M81099; Chinese soft-shelled turtle (*Pelodiscus sinensis*), AB185243; Chinese alligator (*Alligator sinensis*), *Alsi01* HQ158339; American crocodile (*Crocodylus acutus*), *Crac04* HQ158328; Saltwater crocodile (*Crocodylus porosus*), *Crpo01* HQ158304; Black caiman (*Melanosuchus niger*), *Meni02* HQ158358; Chicken (*Gallus gallus*), B-F10 X12780; Mallard (*Anas platyrhynchos*), *Du2* AB115242; Mouse (*Mus musculus*), *H2K* L36312.

The best-fit model, GTR+G+I, was selected by using the AIC criterion in jModelTest2 (Darriba *et al.*, 2012). We ran two Markov chains for 10 million generations, retaining every 1000th sample from the posterior distribution and starting values for each chain chosen randomly. MCMC convergence was explored by examining the potential scale reduction factor (PSRF; Gelman and Rubin, 1992) convergence diagnostics for all parameters in the model and graphically using Tracer v1.5 (Rambaut and Drummond, 2009). The first three million generations, before this chain reached apparent

stationarity, were discarded, and the remaining samples from the independent runs were pooled to obtain the final approximation of the posterior distribution of trees. To yield a single hypothesis of phylogeny, the posterior distribution was summarized as a 50% majority-rule consensus.

## 3. Results

**3.1 MHC class I polymorphism of *E. multiocellata*** In total, 67 distinct sequence variants were identified using SSCP and sequencing (Table 1). Because two variants with base deletions, *Ermu-4*66* and *Ermu-4*67*, were

**Table 1** List of specimens and variants information for MHC class I gene exon 4 of *Eremias multiocellata.*

| Specimen voucher | "True" alleles | Private alleles | Putative pseudogenes |
|---|---|---|---|
| GUO1932 | *Ermu-4*01*, *Ermu-4*02*, *Ermu-4*03*, *Ermu-4*04*, *Ermu-4*05*, *Ermu-4*06*, *Ermu-4*07* | | *Ermu-4*64* |
| GUO1933 | *Ermu-4*01*, *Ermu-4*03*, *Ermu-4*04*, *Ermu-4*05*, *Ermu-4*08*, *Ermu-4*09* | | *Ermu-4*64* |
| GUO1935 | *Ermu-4*06*, *Ermu-4*07*, *Ermu-4*09*, *Ermu-4*10*, *Ermu-4*11*, *Ermu-4*12* | | |
| GUO1936 | *Ermu-4*07*, *Ermu-4*10*, *Ermu-4*11*, *Ermu-4*12*, *Ermu-4*13* | *Ermu-4*37* | |
| GUO1937 | *Ermu-4*05*, *Ermu-4*07*, *Ermu-4*09* | *Ermu-4*38* | |
| GUO1938 | *Ermu-4*07*, *Ermu-4*14* | *Ermu-4*39* | |
| GUO1939 | *Ermu-4*05*, *Ermu-4*07*, *Ermu-4*14* | | |
| GUO1940 | *Ermu-4*05*, *Ermu-4*07*, *Ermu-4*15*, *Ermu-4*16* | | |
| GUO1941 | *Ermu-4*15*, *Ermu-4*16*, *Ermu-4*17* | | |
| GUO1942 | *Ermu-4*07*, *Ermu-4*16*, *Ermu-4*17* | *Ermu-4*40* | |
| GUO1943 | *Ermu-4*07*, *Ermu-4*08*, *Ermu-4*18*, *Ermu-4*19* | | |
| GUO1944 | *Ermu-4*07*, *Ermu-4*09*, *Ermu-4*19* | *Ermu-4*41* | |
| GUO1945 | *Ermu-4*09*, *Ermu-4*20*, *Ermu-4*21* | | |
| GUO2058 | *Ermu-4*01*, *Ermu-4*07*, *Ermu-4*19*, *Ermu-4*21*, | | *Ermu-4*65* |
| GUO2059 | *Ermu-4*05*, *Ermu-4*07*, *Ermu-4*10*, *Ermu-4*11*, *Ermu-4*22* | | *Ermu-4*65* |
| GUO2061 | *Ermu-4*05*, *Ermu-4*23* | *Ermu-4*42* | *Ermu-4*65* |
| GUO2062 | *Ermu-4*08*, *Ermu-4*09*, *Ermu-4*23* | *Ermu-4*43*, *Ermu-4*44*, *Ermu-4*45* | |
| GUO2063 | *Ermu-4*05*, *Ermu-4*24*, *Ermu-4*25* | *Ermu-4*46*, *Ermu-4*47* | |
| GUO2064 | *Ermu-4*05*, *Ermu-4*19*, *Ermu-4*24*, *Ermu-4*25* | | |
| GUO2065 | *Ermu-4*07*, *Ermu-4*13*, *Ermu-4*16*, *Ermu-4*22*, *Ermu-4*26* | | |
| GUO2066 | *Ermu-4*15*, *Ermu-4*26* | *Ermu-4*48*, *Ermu-4*49* | |
| GUO2067 | *Ermu-4*20*, *Ermu-4*27*, *Ermu-4*28* | *Ermu-4*50* | |
| GUO2664 | *Ermu-4*09*, *Ermu-4*24*, *Ermu-4*28*, *Ermu-4*29* | | |
| GUO2665 | *Ermu-4*07*, *Ermu-4*09*, *Ermu-4*29*, *Ermu-4*30* | | |
| GUO2666 | *Ermu-4*05*, *Ermu-4*20*, *Ermu-4*27* | *Ermu-4*51*, *Ermu-4*52*, | *Ermu-4*65* |
| GUO2667 | *Ermu-4*07*, *Ermu-4*09*, *Ermu-4*11* | *Ermu-4*53*, *Ermu-4*54*, | |
| GUO2668 | *Ermu-4*09*, *Ermu-4*31* | *Ermu-4*55*, | *Ermu-4*67* |
| GUO2669 | *Ermu-4*22*, *Ermu-4*24*, *Ermu-4*31*, *Ermu-4*32*, *Ermu-4*56* | *Ermu-4*56*, | |
| GUO2670 | *Ermu-4*07*, *Ermu-4*08*, *Ermu-4*32* | *Ermu-4*57* | |
| GUO2671 | *Ermu-4*18*, *Ermu-4*32* | *Ermu-4*58* | |
| GUO2672 | *Ermu-4*07*, *Ermu-4*18* | *Ermu-4*59* | *Ermu-4*66* |
| GUO2673 | *Ermu-4*02*, *Ermu-4*19* | *Ermu-4*60*, *Ermu-4*61* | *Ermu-4*65* |
| GUO2674 | *Ermu-4*18*, *Ermu-4*19*, *Ermu-4*33*, *Ermu-4*34* | | |
| GUO2675 | *Ermu-4*12*, *Ermu-4*16*, *Ermu-4*33*, *Ermu-4*34*, *Ermu-4*35* | | |
| GUO2676 | *Ermu-4*16*, *Ermu-4*24*, *Ermu-4*30*, *Ermu-4*35*, *Ermu-4*36* | | |
| GUO2677 | *Ermu-4*05*, *Ermu-4*07*, *Ermu-4*09*, *Ermu-4*28* | | |
| GUO2678 | *Ermu-4*07*, *Ermu-4*23*, *Ermu-4*36* | *Ermu-4*62*, *Ermu-4*63* | |

Note: *Ermu-4*64*, *65*, *66*, *67* contained premature stop codons or deletion bases.

detected in only one individual respectively, we did not consider them as valid pseudogenes. Two other variants, *Ermu-4*64* and *Ermu-4*65*, contained premature stop codons, suggesting they may represent putative pseudogenes (sequences are deposited in GenBank, under accession numbers: KJ143608 and KJ143609). No premature stop codons or frameshift mutations were detected in the remaining 63 variant sequences. Among the 63 variants, we found 27 private variants, which were not considered as 'true' alleles (Lukas and Vigilant, 2005; Cummings *et al.*, 2010). Accordingly, 36 sequence variants were validated and included in the subsequent analyses (sequences are deposited in GenBank, under accession numbers KJ143572−KJ143607). The assignment of sequences to specific loci based on phylogenetic analysis was not possible and, for simplicity, we thus call MHC-I loci exon 4 variants as 'alleles'. In accordance with the proposed nomenclature for MHC in nonhuman species (Klein *et al.,* 1990), we referred to the exon4 alleles *Ermu-4** for *E. multiocellata* with serial numbers attached.

In total, 35 variable nucleotide positions were identified in the sequence alignment of the 36 alleles, representing 19.44% of the sequences analyzed. Average pairwise differences between alleles showed 3.52 substitutions accounting for 0.023 of low nucleotide diversity ($\pi$). The number of pairwise nucleotide differences between alleles ranged from one (e.g., *Ermu-4*02 vs. Ermu-4*03*) to five (*Ermu-4*01 vs. Ermu-4*04*).The most common allele, *Ermu-4*07*, was present in 19 out of 37 individuals (frequency = 0.514), followed by *Ermu-4*05* and *Ermu-4*09* (frequency = 0.297), *Ermu-4*16* and *Ermu-4*19* (frequency = 0.162), and *Ermu-4*24* (frequency = 0.135). In contrast, the majority of the alleles (18 alleles) were less frequent and occurred only in two individuals (frequency = 0.054).

**3.2 Putative amino acid sequences** Amino acid (aa) translation of these alleles generated 36 deduced sequences, each 60 aa in length; 20 of which were polymorphic among sequences. Amino acid replacements between alleles ranged from 0 (e.g., *Ermu-4*02 vs. Ermu-4*03*) to 5 (*Ermu-4*01 vs. Ermu-4*04*). Positions 1−9 and 53−60 were conserved and sequence identity levels were 100%. The 36 confirmed sequences were aligned with classical and non-classical sequences from other vertebrates to analyze residues that interact with $\beta$2-microglobulin domain and CD8 molecules. At most of these interaction sites, each racerunner sequence displays a residue also found at the corresponding site of functional class I molecules in other vertebrates, suggesting that these *E. multiocellata* sequences represent functional class I genes. Sites at which *E. multiocellata* display a unique residue appear to be particularly variable across vertebrates. Cysteine residue was completely consistent among the whole sequences in our analysis, suggesting its physicochemical property was quite conservative. Two sequences deduced from *Ermu-4*64* and *65* each contained premature stop codon and were therefore considered as pseudogenes.

**3.3 Allelic copy number variation** The number of MHC-I variants isolated from each individual varied between two and seven (mean, four variants per individual). This indicates that at least four gene loci, but no more than seven occur in *E. multiocellata* as a heterozygote at one MHC locus can contain a maximum of two variants. The large number of MHC-I loci within each individual could explain the large number of novel variants (36 'true' alleles and two putative pseudogene variants) observed in the current study. Twenty-six out of 37 individuals examined (frequency = 0.73) had optimal numbers of MHC 'true' alleles per individual with 13 (frequency = 0.351) containing three variants per individual, eight (frequency = 0.216) containing four variants per individual, and six (frequency = 0.167) containing five variants per individual. Despite lack of specific explanation for the optimal gene copy numbers in *E. multiocellata*, a possible scenario, which needs to be tested, is that individuals with optimal levels of diversity might suffer least from parasitization, as has been proposed for many vertebrates such as sticklebacks (*Gasterosteus aculeatus*) (Wegner *et al.*, 2003), sand lizard (*Lacerta agilis*) (Olsson *et al.*, 2003), house sparrow (*Passer domesticus*) (Griggio *et al.,* 2011), and saltwater crocodile (*Crocodylus porosus*) (Jaratlerdsiri *et al.*, 2012), which their mate choices favored offspring with an optimum number of MHC variants.

**3.4 Purifying selection on MHC class I of *E. multiocellata*** For the whole sequences, the number of nonsynonymous substitutions per site ($d_N$ = 0.022 ) was lower than the number of synonymous substitutions per site ($d_S$ = 0.026 ) with a $P$ value of 0.211 through $Z$-test for purifying selection using MEGA5.0, indicating nearly neutral evolution of the exon 4. Maximum likelihood models also show weak evidence for positive selection acting on the MHC-I exon 4 in the Multiocellated Racerunners (Table 2). Both alternative models which take into account the positive selection did not fit our data significantly better than the basic models of neutral evolution (M2a *vs.* M1a: $df$ = 2, $\chi^2$ = 3.948, $P$ > 0.05; M8

*vs.* M7: $df = 2, \chi^2 = 4.008, P > 0.05$). Models M2a and M8 both identified only one positively selected site (27) with > 95% probability. The SLAC, REL, and FEL yielded similar results to CODEML (see Table 3). Analyses using SLAC, FEL, and REL all suggested that *E. multiocellata* MHC-I exon 4 alleles underwent purifying selection, although the negative selected sites identified by different methods varied (Figure 1 and Table 3). REL predicted the most sites under selection, while SLAC predicted the fewest. Unexpectedly, only approximately one fifth of the sites identified as under purifying selection were $\beta$2-microglobulin interaction sites. The CD8 interaction sites (positions 20−27) showed no signal of purifying selection.

**3.5 Phylogenetic analyses** The 50% majority consensus tree was illustrated in Figure 2. There is strong support for the monophyly of all squamate class I sequences included in the analysis. This suggested that these genes derive from a common ancestral locus whose descendant lineages have not been identified in any other vertebrate group including tuatara, which is also a member of the Lepidosauria superorder. Within squamates, there is also strong node support for the monophyly with all *Eremias* exon 4 sequences included in the analysis. However, the different exon 4 variants of *Eremias* are not grouped in the phylogenetic tree by loci or by species. There is no clear pattern of orthology of within or between species. Mixing of alleles from different species could be observed, although the tree exhibited a star-like topology, likely due to the lack of informative mutations. Alleles from the same species did not always display more similarity than those from different taxa. Specifically, some alleles of Multiocellated Racerunner were identical to those of Gobi Racerunner and Ordos Racerunner, such as *Ermu-4*07*, *Erbr-4*01* and *Erpr-4*01*. In conclusion, exon 4 of *Eremias* MHC-I genes appears to be not evolving in a species-specific manner, but possess trans-species polymorphism.

## 4. Discussion

**4.1 MHC class I exon 4 polymorphism** Based on the observation that as many as seven alleles were present in individuals of *E. multiocellata* and the knowledge that this species is diploid (Dai *et al.*, 2004), we conclude that there are at least four loci in this *Eremias* species. Similar

**Table 2** CODEML results of maximum-likelihood models for exon 4 of the MHC-I loci gene in *Eremias multiocellata.*

| Model code | Parameters | Log-likelihood | Parameter estimates | Positively selected sites (posterior probability) |
|---|---|---|---|---|
| M1a | 2 | –533.026 | $p_0 = 0.722$, $p_1 = 0.278$, | Not allowed |
| M2a | 4 | –531.052 | $p_0 = 0.711$, $p_1 = 0.289$, $p_2 = 0.000$, $\omega_2 = 32.883$ | 10 I(0.781) , **27 R(0.957)**, 39 Y(0.775), 52 E(0.591) |
| M7 | 2 | –533.064 | p = 0.005, q = 0.012 | Not allowed |
| M8 | 4 | –531.060 | $p_0 = 0.999$, p = 0.005, q = 0.012, $\omega$ = 33.667 | 10 I(0.869), 26 S(0.622), **27 R(0.976)**, 30 V(0.505), 39 Y(0.865), 52 E(0.780), |

Note: the maximum likelihood models were performed in CODEML integrated in PAML package v4 (Yang, 2007). *P* indicates a number of parameters considered by the model, $\omega$ is a parameter based on $d_N/d_S$ ratio (selection parameter), $p_n$ is a proportion of sites in a specific $\omega$ site class, p and q are the indicators of shape of $\beta$ distribution; models M2a and M8 estimated the positively selected sites of exon 4 by Bayes Empirical Bayes procedure (Yang *et al.*, 2005). Site with probability > 0.95 is in bold. Amino acids refer to the sequence of *Ermu-4*22*.

**Table 3** Summary statistics for codon sites undergoing purifying or positive selection identified by different methods in HyPhy package.

| Method | Negatively selected sites | | | | | | | | | | Positively selected sites |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 12 | 13 | 17 | 19 | **33** | 34 | 35 | 37 | **41** | 48 | 27 |
| SLAC | | | | | ■ | | | | | | |
| FEL | | ■ | | ■ | ■ | | | ■ | | | + |
| REL | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |

Note: Numbers correspond to the alignment shown in Figure 1. Bold sites represent the predicted function sites. The '■' or '+' indicates that the posterior probability is >95% with a Bayes factor over 50 for REL, or the significance level is at 0.05 for SLAC and FEL.

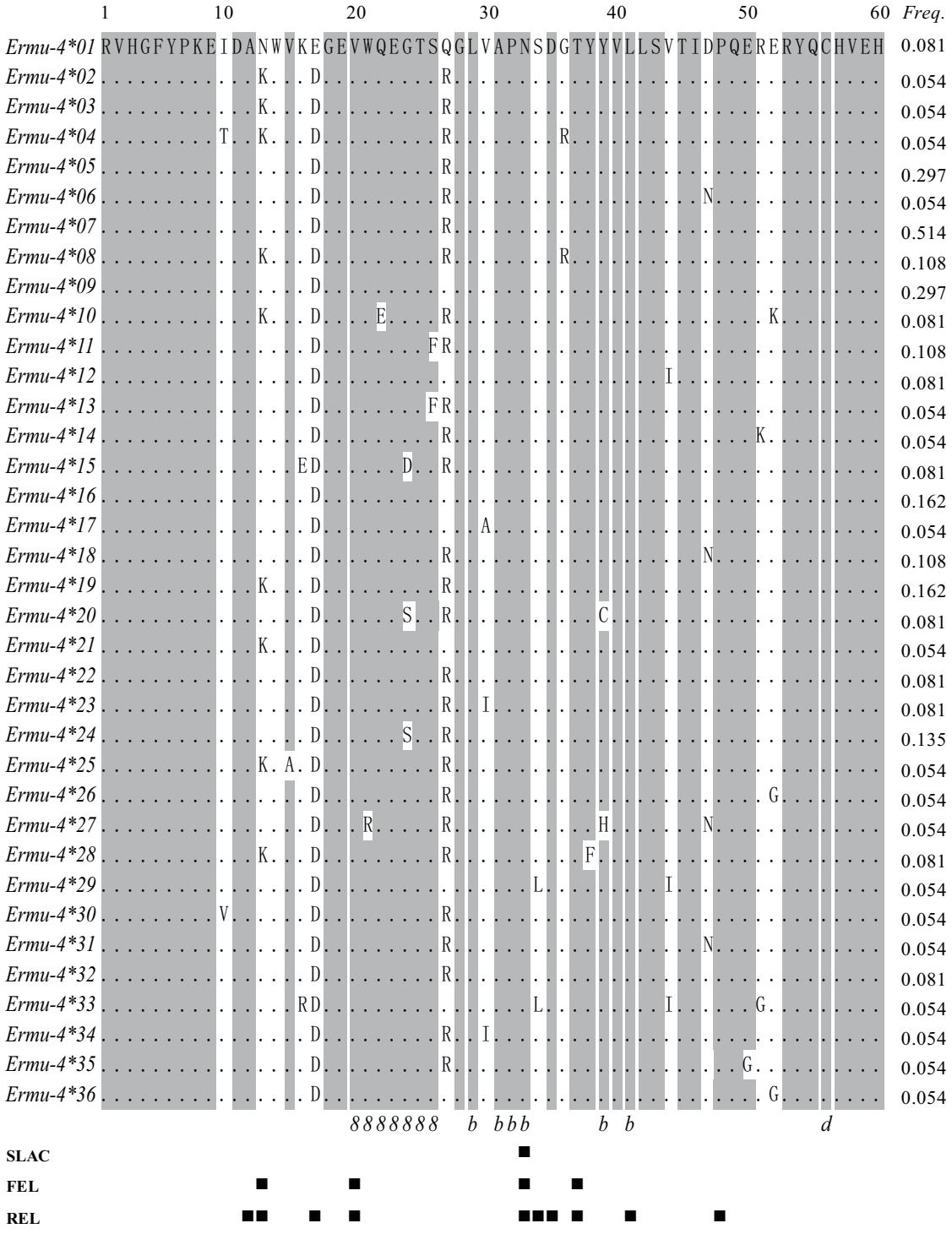


**Figure 1** Amino acid alignment of 36 distinct MHC class I exon 4 alleles and their frequencies in 37 individuals of *Eremias multiocellata*. Dots indicate identity to *Ermu-4*01*. Variable positions are relative to sequence variants on the top. Residues of interest are indicated below the sequences following Siddle *et al.* (2006). *8*, the predicted CD8 interaction sites; *b*, the predicted *β*2-microglobulin interaction sites; *d*, cysteine residues. Signs "■" refer to negative selection sites identified by SLAC, FEL, and REL.

pattern also exits in *E. brenchleyi* (Yuan *et al.*, unpublished data). Our finding of at least four gene loci of MHC-I in *E. multiocellata* could suggest that gene duplication plays a role in increasing copy numbers of MHC genes and, therefore, diversity. Noticeably, the actual number of loci is probably larger than our prediction because we identified the alleles using a very conservative strategy that a sequence was determined as a novel allele only if it

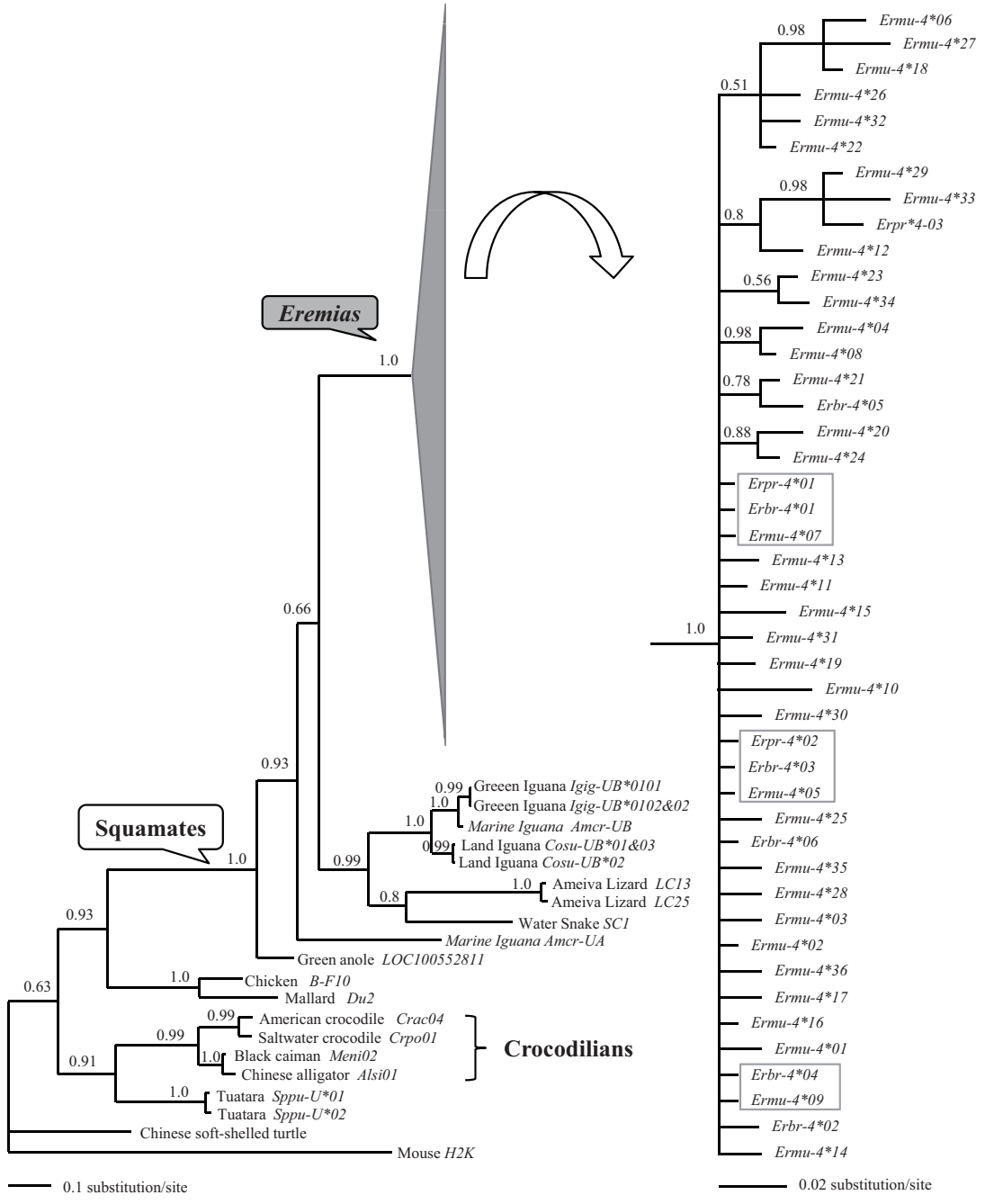


**Figure 2** Bayesian phylogenetic reconstruction using exon 4 (α3) sequence data from the major reptile groups, with particular attention to *Eremias multiocellata*. The number at each node is the Bayesian posterior probability supporting the phylogram displayed. *Erpr-4*01–03* represent three MHC class I exon 4 alleles isolated from an individual of Gobi Racerunner (*Eremias przewalskii*) from Minqin, Gansu, China (Voucher No. GUO2055). *Erbr-4*01–06* represent six MHC class I exon 4 alleles isolated from an individual of Ordos Racerunner (*Eremias brenchleyi*) from Taishan, Shandong, China (Voucher No. GUO1913). Identical alleles among the racerunners were marked with frame.

appeared in more than one individual. Gene duplication among MHC-I genes has also been observed in many other representative vertebrates such as Atlantic salmon (Miller *et al.*, 2006), frogs (Kiemnec-Tyburczy *et al.*, 2012), birds (Shiina *et al.*, 2004), iguanas (Glaberman and Caccone, 2008), and saltwater crocodile (Jaratlerdsiri *et al.*, 2012; 2014). Meanwhile, we found two putative pseudogenes, which contain premature stop codons. Thus our present study showed one additional reptile species with multiple MHC-I loci, supported results of the recent study on reptile MHC-I genes, and presented the ubiquity of multiple MHC-I loci in vertebrates. Our knowledge on

reptile MHC-I genes will be updated and broadened by studying more representative species.

Eighteen MHC variants, representing 50 % of the total MHC alleles, were found only twice among the samples studied here. Together with the large number of novel variants detected ($V$ = 36), this could indicate that there is a rapidly evolving locus within this species consistent with MHC-I gene loci in mammals that are found to have high duplication rates (Piontkivska and Nei, 2003). Given the high frequency (over 50 %) of the MHC variant *Ermu-4*07* among the individuals of *E. multiocellata* studied, it is possible to consider that this variant may correspond to ancestral sequence. This variant may be subject to particular selective pressures that allow the variant to persist longer than expected under neutrality (Klein, 1987). To sum, the data seem to be compatible with a birth-and-death model of evolution (Nei *et al.*, 1997).

**4.2 Effects of recombination and purifying selection** As a ubiquitous evolutionary force, recombination has been shown to contribute substantially to MHC diversity, creating new alleles in isolated populations (Consuegra *et al.*, 2005). However, in the present study, we have not detected signal of recombination in the MHC-I exon 4 sequences. The detection of recombination is based on detecting discordant phylogenetic signal in alignments of DNA or protein sequences, which provides estimates of the number and location of break points (Kosakovsky Pond *et al.*, 2006). Based on the result of SBP method, there shows no signal of recombination in the exon 4 data set. Thus, it is likely that the sequences polymorphism could be explained by point mutations at the polymorphic sites. Nevertheless, gene conversion events among exon 4 sequences may not have been detected because of insufficient nucleotide polymorphism weakens the statistical power of this analysis (Kosakovsky Pond *et al.*, 2006). Indeed, gene conversion tends to decrease the nucleotide diversity among sequences but increase the number of haplotypes (Nei and Rooney, 2005). It is therefore possible that in *E. multiocellata* MHC class I, gene conversion may have reduced the nucleotide variation in exon 4, thereby reducing the statistical power to detect it whilst increasing the number of distinct sequences in exon 4. At present, we are unable to test this hypothesis.

Several lines of evidence suggest that MHC-I exon 4 sequences have undergone purifying selection. First, the low nucleotide diversity and the relative lack of amino acid replacement substitutions observed in exon 4 is a signature of purifying selection (Wright and Gaut, 2005), and this has been observed in many other species (Kiemnec-Tyburczy *et al.*, 2012). Second, the nucleotide sequence divergence of these alleles was increased compared to the amino acid sequence divergence, consistent with the purifying selection signal showing that synonymous were more frequent than nonsynonymous variations. Third, several codon-based methods have detected negative selection signals as well as the sites probably under purifying selection in the 36 MHC-I exon 4 alleles. The sites that were identified by more than one method were considered robust. Purifying selection is detected at $\beta$2-microglobulin interaction sites, and sites outside of but close to portions which are involved in CD8 or $\beta$2-microglobulin binding (Figure 1). Given that the functional constraint imposed on the α3 domain (Kaufman *et al.*, 1992), exon 4 is expected to have lower $d_N/d_S$. Our results are consistent with this prediction; very few sites were under positive selection in exon 4. Interestingly, among the ten sites detected under purifying selection (Figure 1; Table 3), six are silent sites, i.e. synonymous sites, including residues 12, 33, 35, 37, 41, 48. However, as noted by van Oosterhout (2009), purifying selection is inefficient with low recombination rates. It is reasonable to observe the low nucleotide diversity and high number of variants because drift may overcome 'weak' purifying selection. Thus, in *E. multiocellata*, MHC-I exon 4 variability is predominantly caused by the accumulation of point mutations over millions of years, with gene duplication generating additional allelic diversity.

**4.3 Implications for trans-species polymorphism** It was interestingly noted that certain identical alleles were shared by different racerunner species. Of all the exon 4 alleles identified in this study and those from Gobi Racerunner and Ordos Racerunner (Yuan *et al.*, unpublished data), at least three groups (Figure 2) were identical between the three species. Each of these identical alleles was identified from at least two individuals. The shared polymorphism observed of different racerunner species would have to be the product of ancestral polymorphism at the original locus where multiple allelic lineages were maintained after species divergence. In general, such cases of trans-species polymorphism are one of the hallmarks of MHC genes and are an expected consequence of strong balancing selection (Hughes and Nei, 1988; Hughes and Yeager, 1998; Piertney and Oliver, 2006). Under the scenario of concerted evolution, it is expected to generate a high degree of sequence identity among multiple MHC gene families within species. This can mask gene orthology, leading to species-specific clusters of sequences. The

finding from the phylogenetic tree in this study can preclude the possibility that concerted evolution caused the trans-species polymorphism.

In trans-species polymorphism, mixing of alleles from different species results from the long-time persistence of MHC genes which originated before births of the corresponding species (Klein *et al.*, 1998). So we can estimate the time when MHC genes come into being according to divergence time of related species. In the present study, the racerunner MHC-I lineage was likely to originate at least before divergence of the three species (subgenus *Pareremias*, Guo *et al.*, 2011). As the divergence event date back to about 6.3 million years ago (with the 95% credible interval ranging from 5.3 to 8.5 Ma) (Guo *et al.*, 2011), the oldest MHC-I allelic lineage for the subgenus *Pareremias* was estimated to originate about 6 Ma. This time to main MHC-I alleles for racerunners is consistent with that for some other vertebrates. For example, human HLA class I lineages are recognized in great apes and thus have been maintained for 6 Ma (Vogel *et al.*, 1999).

**Acknowledgements** We thank Jinlong LIU, Dajiang LI, Bo CAI, and Tianhe ZHOU for assistance in collecting specimens or tissue samples. This research was financially supported by the Science and Technology Project for Outstanding Youths in Life Science (KSCX2-EW-Q-6) from the Chinese Academy of Sciences, and National Natural Science Foundation of China (31272281).

## References

**Apasov S., Sitkovsky M.** 1993. Highly lytic $CD8^+$, $\alpha\beta$ T-cell receptor cytotoxic T cells with major histocompatibility complex (MHC) class I antigen-directed cytotoxicity in $\beta$2-microglobulin, MHC class I-deficient mice. Proc Natl Acad Sci U S A, 90: 2837–2841

**Babik W.** 2010. Methods for MHC genotyping in non-model vertebrates. Mol Ecol Resour, 10: 237–251

**Bjorkman P. J., Saper M. A., Samraoui B., Bennett W. S., Stromiger J. L., Wiley D. C.** 1987. Structure of the human class I histocompatibility antigen, HLA-A2. Nature, 329: 506–512

**Bonhomme M., Doxiadis G. G. M., Heijmans C. M. C., Vervoort V., Otting N., Bontrop R. E., Crouau-Roy B.** 2008. Genomic plasticity of the immune-related Mhc class I B region in macaque species. BMC Genomics, 9: 514–524

**Burland T. G.** 2000. DNASTAR's Lasergene sequence analysis software. Methods Mol Biol, 132: 71–91

**Consuegra S., Megens H. J., Schaschl H., Leon K., Stet R. J., Jordan W. C.** 2005. Rapid evolution of the MH class I locus results in different allelic compositions in recently diverged populations of Atlantic salmon. Mol Biol Evol, 22: 1095–1106

**Cummings S. M., McMullan M., Joyce D. A., van Oosterhout C.** 2010. Solutions for PCR, cloning and sequencing errors in population genetic analysis. Conserv Genet, 11: 1095–1097

**Dai X., Zeng X., Chen B., Wang Y.** 2004. The research on the karyotypes of six species in the genus *Eremias* from China. Hereditas, 26: 669–675(In Chinese with English abstract)

**Danchin E., Vitiello V., Vienne A., Richard O., Gouret P., McDermott M. F., Pontarotti P.** 2004. The major histocompatibility complex origin. Immunol Rev, 198: 216–232

**Darriba D., Taboada G. L., Doallo R., Posada D.** 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods, 9: 772–784

**Delport W., Poon A. F., Frost S. D., Kosakovsky Pond S. L.** 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics, 26: 2455–2457

**de Bakker P. I., Raychaudhuri S.** 2012. Interrogating the major histocompatibility complex with high-throughput genomics. Hum Mol Genet, 21: 29–36

**Eimes J. A., Bollmer J. L., Whittingham L. A., Johnson J. A., van Oosterhout C., Dunn P. O.** 2011. Rapid loss of MHC class II variation in a bottlenecked population is explained by drift and copy number variation. J Evol Biol, 24: 1847–1856

**Figueroa F., Mayer W. E., Sato A., Zaleska-Rutczynska Z., Hess B., Tichy H., Klein J.** 2001. MHC class I genes of swordtail fishes, *Xiphophorus*: variation in the number of loci and existence of ancient gene families. Immunogenetics, 53: 695–708

**Fitch W. M.** 2000. Homology: a personal view on some of the problems. Trends Genet, 16: 227–231

**Gelman A., Rubin D. B.** 1992. Inference from iterative simulation using multiple sequences. Stat Sci, 7: 457–511

**Glaberman S., Caccone A.** 2008. Species-specific evolution of class I MHC genes in iguanas (Order: Squamata; Subfamily: Iguaninae). Immunogenetics, 60: 371–382

**Gorer P. A.** 1936. The detection of antigenic differences in mouse erythrocytes by the employment of immune sera. Int J Clin Exp Pathol, 17: 42–50

**Griggio M., Biard C., Penn D. J., Hoi H.** 2011. Female house sparrows "count on" male genes: experimental evidence for MHC-dependent mate preference in birds. BMC Evol Biol, 11: 44–50

**Grossberger D., Parham P.** 1992. Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. Immunogenetics, 36: 166–174

**Guindon S., Dufayard J. F., Lefort V., Anisimova M., Hordijk W., Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol, 59: 307–321

**Guo X., Dai X., Chen D., Papenfuss T. J., Ananjeva N. B., Melnikov D. A., Wang Y.** 2011. Phylogeny and divergence times of some racerunner lizards (Lacertidae: *Eremias*) inferred from mitochondrial 16S rRNA gene segments. Mol Phylogenet Evol, 61: 400–412

**Haddrill P. R., Halligan D. L., Tomaras D., Charlesworth B.** 2007. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. Genome Biol, 8: 181–189

**Hedrick P. W.** 1998. Balancing selection and MHC. Genetica, 104: 207–214

**Hughes A. L., Nei M.** 1988. Pattern of nucleotide substitution

at major histocompatibility complex class I loci reveals overdominant selection. Nature, 335: 167–170

**Hughes A. L., Yeager M.** 1998. Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet, 32: 415–435

**Jaratlerdsiri W., Isberg S. R., Higgins D. P., Gongora J.** 2012. MHC class I of saltwater crocodiles (*Crocodylus porosus*): polymorphism and balancing selection. Immunogenetics, 64: 825–838

**Jaratlerdsiri W., Isberg S.R., Higgins D.P., Ho S.Y., Salomonsen J., Skjodt K., Miles L.G., Gongora J.** 2014. Evolution of MHC class I in the Order Crocodylia. Immunogenetics, 66: 53–65

**Kaufman J., Andersen R., Avila D., Engberg J., Lambris J., Salomonsen J., Welinder K., Skjødt K.** 1992. Different features of the MHC class I heterodimer have evolved at different rates. Chicken B-F and $\beta$2-microglobulin sequences reveal invariant surface residues. J Immunol, 148: 1532–1546

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

**Kiemnec-Tyburczy K. M., Richmond J. Q., Savage A. E., Lips K. R., Zamudio K. R.** 2012. Genetic diversity of MHC class I loci in six non-model frogs is shaped by positive selection and gene duplication. Heredity, 109: 146–155

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

**Klein J., Sato A., Nagl S., ÓhUigín C.** 1998. Molecular trans-species polymorphism. Annu Rev Ecol Evol Syst, 29: 1–21

**Klein J., Sato A., Nikolaidis N.** 2007. MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. Annu Rev Genet, 41: 281–304

**Kosakovsky Pond S. L., Frost S. D.** 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol, 22: 1208–1222

**Kosakovsky Pond S. L., Posada D., Gravenor M. B., Woelk C. H., Frost S. D.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol, 23:1891–1901

**Kulski J. K., Shiina T., Anzai T., Kohara S., Inoko H**. 2002. Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol Rev, 190: 95–122

**Librado P., Rozas J.** 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25: 1451–1452

**Lukas D., Vigilant L.** 2005. Reply: facts, faeces and setting standards for the study of MHC genes using noninvasive samples. Mol Ecol, 14: 1601–1602

**Madsen T., Olsson M., Wittzell H., Stille B., Gullberg A., Shine R., Andersson S., Tegelström, H.** 2000. Population size and genetic diversity in sand lizards (*Lacerta agilis*) and adders (*Vipera berus*). Biol Conserv, 94: 257–262

**Malaga-Trillo E., Zaleska-Rutczynska Z., McAndrew B., Vincek V., Figueroa F., Sultmann H., Klein J.**, 1998. Linkage relationships and haplotype polymorphism among cichlid MHC class II B loci. Genetics, 149: 1527–1537

**Maruyama T., Nei M.** 1981. Genetic variability maintained by mutation and over-dominant selection in finite populations. Genetics, 98: 441–459

**Mehta R. B., Nonaka M. I., Nonaka M.** 2009. Comparative genomic analysis of the major histocompatibility complex class I region in the teleost genus *Oryzias*. Immunogenetics, 61: 385–399

**Miller H. C., Andrews-Cookson M., Daugherty C. H.** 2007. Two patterns of variation among MHC class I loci in tuatara (*Sphenodon punctatus*). J Hered, 98: 666–677

**Miller H. C., Belov K., Daugherty C. H.** 2006. MHC class I genes in the tuatara (*Sphenodon* spp.): evolution of the MHC in an ancient reptilian order. Mol Biol Evol, 23: 949–956

**Millinski M.** 2006. The major histocompatibility complex, sexual selection, and mate choice. Annu Rev Ecol Evol Syst, 37: 159–186

**Murphy B. F., Thompson M. B., Belov K.** 2009. Evolution of viviparity and the maternal immune system: major histocompatibility complex (MHC) class I genes in skinks. Orbit, 1: 1–17

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

**Nei M., Rooney A. P.** 2005. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet, 39: 121–152

**Nei, M., Gu X., Sitnikova T.** 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci U S A, 94: 7799–7806

**Nonaka M. I., Aizawa K., Mitani H., Bannai H. P., Nonaka M.** 2011. Retained orthologous relationships of the MHC class I genes during euteleost evolution. Mol Biol Evol, 28: 3099–3112

**Olsson M., Madsen T., Nordby J., Wapstra E., Ujvari B., Wittsell H.** 2003. Major histocompatibility complex and mate choice in sand lizards. Proc Biol Sci, 270: 254–256

**Orita M., Iwahana H., Kanazawa H., Haysdhi K., Sekiya T.** 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci U S A, 86: 2766–2770

**Otten G. R., Bikoff E., Ribaudo R. K., Kozlowski S., Margulies D. H., Germain R. N.** 1992. Peptide and $\beta$2-microglobulin regulation of cell surface MHC class I conformation and expression. J Immunol, 148: 3723–3732

**Piertney S. B., Oliver M. K.** 2006. The evolutionary ecology of the major histocompatibility complex. Heredity, 96: 7–21

**Piontkivska H., Nei M.** 2003. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. Mol Biol Evol, 20: 601–609

**Radtkey R. R., Becker B., Miller R. D., Riblet R., Case T. J.** 1996. Variation and evolution of class I MHC in sexual and parthenogenetic geckos. Proc Biol Sci, 263: 1023–1032

**Rambaut A., Drummond A. J.** 2009. Tracer v1.5. Retrieved from http://beast.bio.ed.ac.uk/Tracer

**Ronquist F., Teslenko M., van der Mark P., Ayres D. L., Darling A., Höhna S., Larget B., Liu L., Suchard M. A., Huelsenbeck J. P.** 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol, 61: 539–542

**Saper M., Bjorkman P., Wiley D.** 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. J

Mol Biol, 219: 277–319

**Shand R., Dixon B.** 2001. Teleost major histocompatibility genes: diverse but not complex. Mod Asp Immunobiol, 2: 66–72

**Shiina T., Shimuzu S., Hosomichi K., Kohara S., Watanabe S., Hanzawa K., Beck S., Kulski J.K., Inoko H.** 2004. Comparative genomic analysis of two avian (quail and chicken) MHC regions. J Immunol, 172: 6751–6763

**Siddle H. V., Deakin J. E., Baker M. L., Miller R. D., Belov K.** 2006. Isolation of major histocompatibility complex class I genes from the tammar wallaby (*Macropus eugenii*). Immunogenetics, 58: 487–493

**Snell G. D.** 1948. Methods for the study of histocompatibility genes. J Genet, 49: 87–108

**Sommer S.** 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool, 2: 1–18

**Spurgin L. G., van Oosterhout C., Illera J. C., Bridgett S., Gharbi K., Emerson B. C., Richardson D. S.** 2011. Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. Mol Ecol, 20: 5213–5225

**Stiebens V. A., Merino S. E., Chain F. J., Eizaguirre C.** 2013. Evolution of MHC class I genes in the endangered loggerhead sea turtle (*Caretta caretta*) revealed by 454 amplicon sequencing. BMC Evol Biol, 13: 1–11

**Sunnucks P., Wilson A. C., Beheregaray L. B., Zenger K., French J., Taylor A. C.** 2000. SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. Mol Ecol, 9: 1699–1710

**Szczerbak N. N.** 2003. Guide to the Reptiles of the Eastern Palearctic. Malabar, Florida: Krieger Publishing Company, 260 pp

**Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S.** 2011. MEGA5: molecular Eevolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol, 28: 2731–2739

**Ujvari B., Belov K.** 2011. Major histocompatibility complex (MHC) markers in conservation biology. Int J Mol Sci, 12: 5168–5186

**van Oosterhout C.** 2009. A new theory of MHC evolution: beyond selection on the immune genes. Proc Biol Sci, 276: 657–665

**Vogel T. U., Evans D. T., Urvater J. A., O'Connor D. H., Hughes A. L., Watkins D. I**. 1999. Major histocompatibility complex class I genes in primates: co-evolution with pathogens. Immunol Rev, 167: 327–337

**Wegner K. M., Reusch T. B. H., Kalbe M.** 2003. Multiple infections drive major histocompatibility complex polymorphism in the wild. J Evol Biol, 16: 224–232

**Xu S., Sun P., Zhou K., Yang G.** 2007. Sequence variability at three MHC loci of finless porpoises (*Neophocaena phocaenoides*). Immunogenetics, 59: 581–592

**Yang Z.** 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol, 24: 1586–1591

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