Diversity and selection of MHC class IIb gene exon3 in Chinese alligator

Chuanpeng Nie · Juan Zhao · Yanyan Li · Xiaobing Wu

Abstract Our study used MHC class IIb gene exon3 complete sequence as markers to investigate genetic variability, selection and population differentiation in Chinese alligator. In this study, 282 bp MHC IIb exon3 complete sequence was got, none of the sequences contained insertions/deletions or stop codons, suggesting that all sequences might come from functional molecules in the genome. The neighbor-joining (NJ) tree revealed that Xuangzhou and Changxing populations were genetically close related, while Wild population showed the most diverse from the other. Gene flow (Nm) was very higher than one, suggesting that inter-group gene flow may have occurred. Furthermore, the results showed that MHC IIb gene might be a good molecular marker, we think that this technology could be used for Chinese alligator breeding and releasing in future.

Keywords MHC class IIb gene exon3 · Diversity · Selection · Chinese alligator · Endangered

Introduction

Chinese alligator (*Alligator sinensis*), one of the 23 critically endangered crocodile species in the world, has been closed to extinct during the past decades due to habitat loss and serious illegal hunting [1]. The investigation indicated that the number of Chinese alligators has decreased from 500 individuals in the 1980s to currently less than 120 or 150 individuals in the wild [1, 2].

Genetic variability in this relict species is obviously essential for the genetic management of the captive alligator. Until now, the genetic diversity studies about Chinese alligator have mostly focused on neutral DNA markers, such as RAPD [3], AFLP [4], mtDNA D-loop sequencing [5, 6] and microsatellite [7–10]. However, because of selective processes involving the environment or the capacity for future adaptive change, the genetic structure may differ from that of neutral genetic regions [11–14]. Therefore, adaptive non-neutral markers have become valuable especially [15–17], such as the genes of the major histocompatibility complex (MHC), which are found in all jawed vertebrates and play a critical role in an organism's immune response [18, 19]. Shi et al. [20] analyzed the fragments of MHC class IIb exon2 in three Chinese alligators from ARCCAR, and Liu et al. [21] researched the exon3 partial sequences in 14 Chinese alligators. They found there was higher polymorphism of MHC class IIb genes in Chinese alligators.

In this study, we would define MHC IIb gene exon3 complete sequences and expand on these earlier studies by examining genetic variation across populations based on these complete sequences and investigate selection at the sequence level by examining patterns of sequence substitutions. Our specific goals were (1) to define MHC class IIb gene exon3 complete sequences and analyze selection, (2) to provide more detailed genetic information for conservation and management strategies for this endangered population [22].

C. Nie · J. Zhao · X. Wu (✉)
College of Life Sciences, Institute of Traditional Chinese Medicine Resources, Anhui Normal University, Wuhu 241000, People's Republic of China
e-mail: wuxb@mail.ahnu.edu.cn

C. Nie · Y. Li
College of Life Sciences, Fuyang Teachers College, Fuyang 236041, People's Republic of China

Materials and methods

Samples and DNA extraction

Sample of individuals was given in Table 1. Sample collection, transportation and storage were same to the previous study [3, 5]. DNA extraction followed a conventional phenol/chloroform procedure [23], and genomic DNA dissolved with ddH2O. The extracted DNA was examined on 1 % agarose gels stained with 10 mg/ml ethidium bromide, stored at −20 °C for further use.

PCR procedure

The primers amplifing the MHC designed by the sequence of *Caiman crocodilus* (AF256651, AF256652 and AF277661) through Primer premier 5.0 and Oligo 6.0 [24]. Amplification reactions (30 μl) containing 30 ng genome DNA, 3 μl 10× PCR buffer (Sangon in Shanghai), 2 μl 25 mM MgCl$_2$ (Sangon in Shanghai), 1 μl 25 mM dNTP (Sangon in Shanghai), 2 μl 10 mM primer (Synthesized by Genscript in Nanjing), 1 U *Taq* DNA polymerase (Sangon in Shanghai). The PCR was performed in a Mastercycler gradient (Eppendorf, Germany). Initial denaturation of 95 °C for 5 min was followed by 35 cycles of 94 °C for 30 s, 47 °C for 30 s (primer annealing), and 72 °C for 1 min (primer extension). A final extension of 72 °C for 10 min was incorporated, followed by cooling to 4 °C until recovery of the samples. PCR products were separated on a 1.5 % agarose gel following staining with 10 mg/ml ethidium bromide and visualized using UV light.

Cloning and sequencing

PCR purified products were cloned into pMD18-T vectors (Takara, Dalian, China) and transformed into DH5a competent cells following the manufacturer's recommendation. After incubation at 37 °C on LB agar-ampicillin plates overnight, at least 15 clones per individual were checked for an insert by PCR. Between 8 and 11 insert-positive clones per individual were sequenced with the M13 forward and reverse primer by Genscript in Nanjing.

Table 1 Samples of Chinese alligator used for this study

Populations	Sources	Numbers	Type
Wild population (Wild)	Xuanzhou, Anhui	7	Blood
Captive population in Xuanzhou (Xuanzhou)	ARCCAR	7	Blood
Captive population in Changxing (Changxing)	Changxing, Zhejiang	7	Blood

Sequences were obtained from at least two clones in the same individual were used in the subsequent data analysis; the sequences which had different length in the same individual were kept. This strategy should allow for the removal of potential artificial polymorphisms, because of the use of *Taq* DNA polymerase by which mis-incorporation could occasionally occur.

Statistic analysis

After splicing with the ContigExpress software and correcting with peak Chromas [25], all sequences were compared (using the algorithm BLASTn) with those available via the NCBI database (http://www.ncbi.nlm.nih.gov) [26]. Nucleotide multiple alignments were performed with ClustalX [27].

Nucleotide polymorphism and diversity were calculated using DnaSP v5 [28]. Analysis of molecular variance (AMOVA) was performed using Arlequin v3.1 [29] to assess genotypic variations across all the populations studied. Population's relationships were estimated using the NJ method.

Statistical analysis of nucleotide and amino acid sequences was performed with MEGA v4 [30]. The non-synonymous (dN) and synonymous (dS) substitution rates between alleles were calculated for the entire sequence using the Modified Nei–Gojorbori method applying the Jukes–Cantor correction [31]. In addition, we used MEGA v4 to perform Z tests of positive, neutral and purifying selection on the entire exon3.

Pairwise analyses of transitions and transversions, and computation of DNA and amino acid entropy were performed with the program DAMBE [32, 33]. To see how the alleles were related, we reconstructed a nucleotide phylogeny by the NJ method with Kimura's two-parameter model, using MEGA v4 [30]. Bootstrap values were calculated by 1,000 replicates, and branches corresponding to partitions reproduced in less than 50 % of bootstrap replicates were collapsed.

Results

Definition of MHC IIb gene exon3 complete sequences

By sequencing, sequences of three length types were obtained in Chinese alligators, namely 913, 928, 943 bp. Similarity search of the sequences at the NCBI database showed 86–94 % similarity with *C. crocodilus* (AF256651, AF256652 and AF277661). According to the intron–exon boundary GT–AG rules, and by aligning with the sequence of *C. crocodilus* (AF256651, AF256652 and AF277661), we had found the sequences included 282 bp MHC IIb exon3 complete sequence. The sequence did not show any gaps, the nucleotide sequences were corresponded with

putative 93 amino acids peptide, none of the sequences contained insertions/deletions or stop codons, suggesting that all sequences might come from functional molecules in the genome. The sequences from 928 and 913 bp length types could be cut to the same exon sequence.

Nucleotide and amino acid variation

For simplicity, different sequences are tentatively referred to as alleles even though they may be from different loci [34]. In total, 206 clones from 21 Chinese alligators were sequenced; an average of ten clones (range 8–11) per individual. By sequencing, we indentified 43 alleles from 21 individuals. These sequences have been deposited to GenBank (accession nos. JQ048623–JQ048665). Moreover, 21 alleles indentified in exon3 DNA complete sequence and 16 alleles indentified in exon3 amino acid

complete sequence. 11.7 % (33 out of 282) of the nucleotide sites was variable and 16.1 % (15 out of 93) of the amino acid sites was variable. Entropy measures the variation index at per site were given (Fig. 1). Nucleotide diversity in Wild population was higher than that in two captive populations, Xuangzhou population higher than Changxing population. About haplotype diversity, Wild population was highest, followed by Changxing and Xuangzhou populations (Table 2).

Selection on MHC IIb exon3 complete sequence

Z tests allowed us to exclude the possibility that the exon3 experiences balancing selection ($Z = -3.167$, $P = 1.000$); Z tests indicated that the dN/dS ratio does not differ from neutrality ($Z = -2.990$, $P = 0.003$), and a test of purifying selection was significant ($Z = 3.134$, $P = 0.001$) (Table 3).

Fig. 1 a Entropy measures the variation index at per base site in exon3 sequences, **b** entropy measures the variation index at per amino acid site Hi = −sum [Pi·log 2 (Pi)]; Hi: 0–2; Pi: the substitution frequencies of different nucleotides per site within the allelic sequences sampled
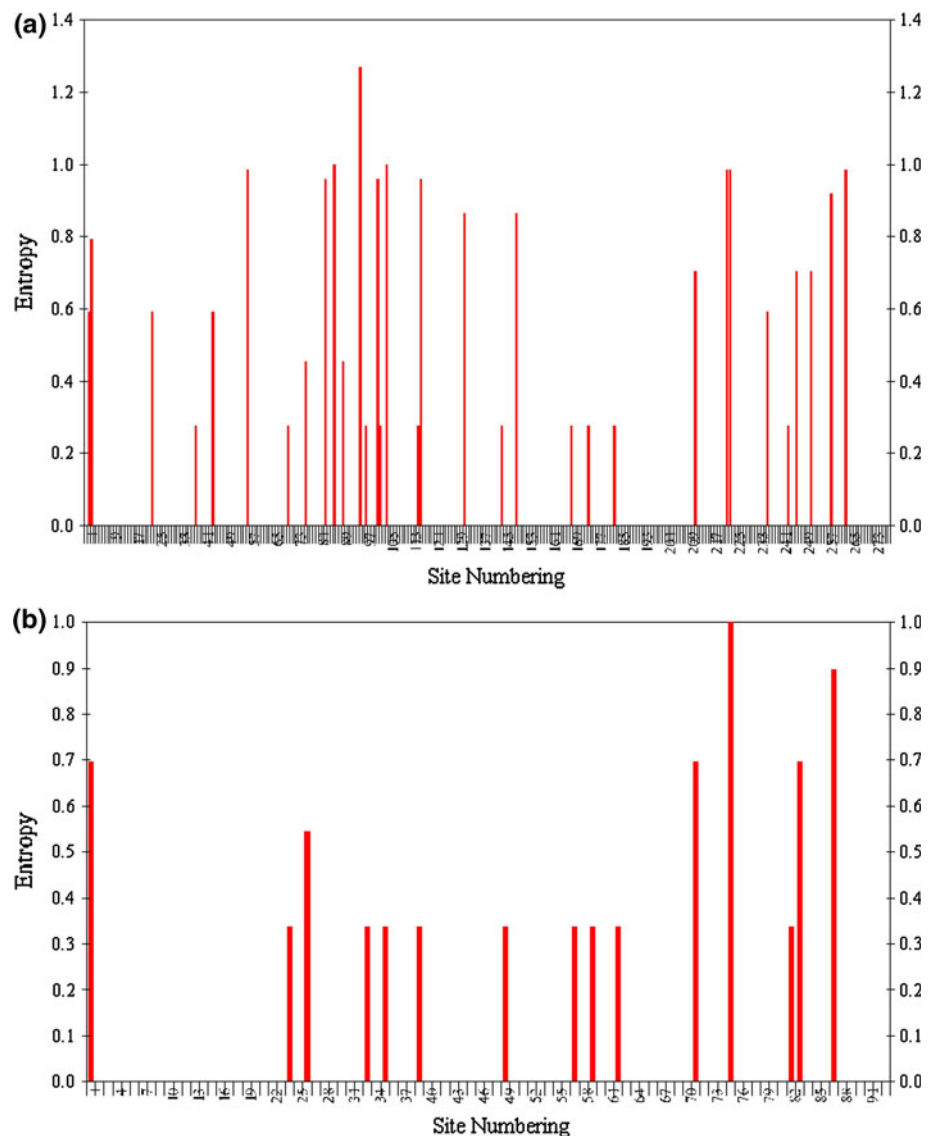
Table 2 Analysis of genetic diversity of different populations

Population	Number of specimen	Average number of nucleotide differences (k)	Haplotype diversity (Hd)	Nucleotide diversity (π)	Number of haplotype (h)
Xuanzhou	7	9.543	0.84762	0.03420	7
Wild	7	10.04737	0.94211	0.03601	13
Changxing	7	6.93464	0.88889	0.02486	10
Total	21	8.876	0.894	0.03181	21

Table 3 Summary of nucleotide substitution rates in the MHC IIb gene exon3 of China alligator

	dS	dN	Purifying selection	
			Z	P
Xuanzhou	0.093 ± 0.024	0.014 ± 0.007	3.393	0.001
Wild	0.091 ± 0.025	0.017 ± 0.006	3.054	0.001
Changxing	0.067 ± 0.018	0.010 ± 0.005	3.081	0.001
Total	0.083 ± 0.021	0.014 ± 0.006	3.134	0.001
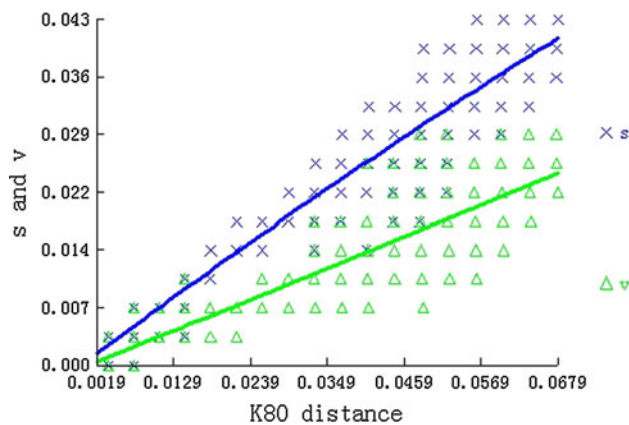
Fig. 2 Substitution saturation analysis transitions-type and transversions-type nucleotide differences plotted against the evolutionary distance of Kimura's two-parameter method

Fig. 3 The cluster of three populations in Chinese alligator based on NJ method

Substitution saturation analysis

The analysis of substitution saturation was conducted and is shown in Fig. 2. When transitions occur much more frequently than transversions, no saturation of substitution is recognized. On the other hand, when transversions gradually outnumber transitions, substitution saturation is suspected because multiple substitutions may have occurred at each site. Therefore, we conducted the comparison of two regression slopes. The slope of transitions was significantly steeper than that for transversions against evolutionary distances, so the substitution was considered not to be saturated.

Population structure

The genetic distances between the Wild population and the two captive populations were higher than that between the two captive populations. The NJ tree (Fig. 3) revealed that Xuangzhou population and Changxing population were genetically close related, while Wild population showed the most diverse from the others.

The difference among three populations was not significant ($P > 0.05$) in AMOVA. The degree of differentiation within a population accounted for 99.53 %, while only 0.47 % among populations (Table 4). All those indicated that the genetic differences mainly occurred within populations.

Alleles phylogeny

In this study, the sequences included exon3 complete sequences. Phylogenetic NJ tree showed alleles group coming from 928 bp sequences were relatively concentrated (Fig. 4). We also found trans-species polymorphism

Table 4 Hierarchical analysis of molecular variance (AMOVA) within/among populations

Source of variation	df	Sum of squares	Variance components	Percentage of variation	Fst	P value
Among populations	2	9.579	0.02084 Va	0.47	–	–
Within populations	50	221.194	4.42389 Vb	99.53	–	–
Total	52	230.774	4.44473	–	0.00469	0.35582 ± 0.01200

df degrees of freedom, Va variance components among populations, Vb variance components within populations, Fst fixation index

Fig. 4 **a** NJ phylogeny of 21 alleles based on exon3 DNA complete sequence, **b** NJ phylogeny of 16 alleles based on exon3 amino acids sequence. A NJ phylogeny based on genetic distance of the Kimura two-parameter model and bootstrap re-sampling for 1,000 times. The tree also includes three alleles from the *Caiman crocodilus* (AF256651, AF256652 and AF277661) (Bootstrap values >50 % were indicated above branches)
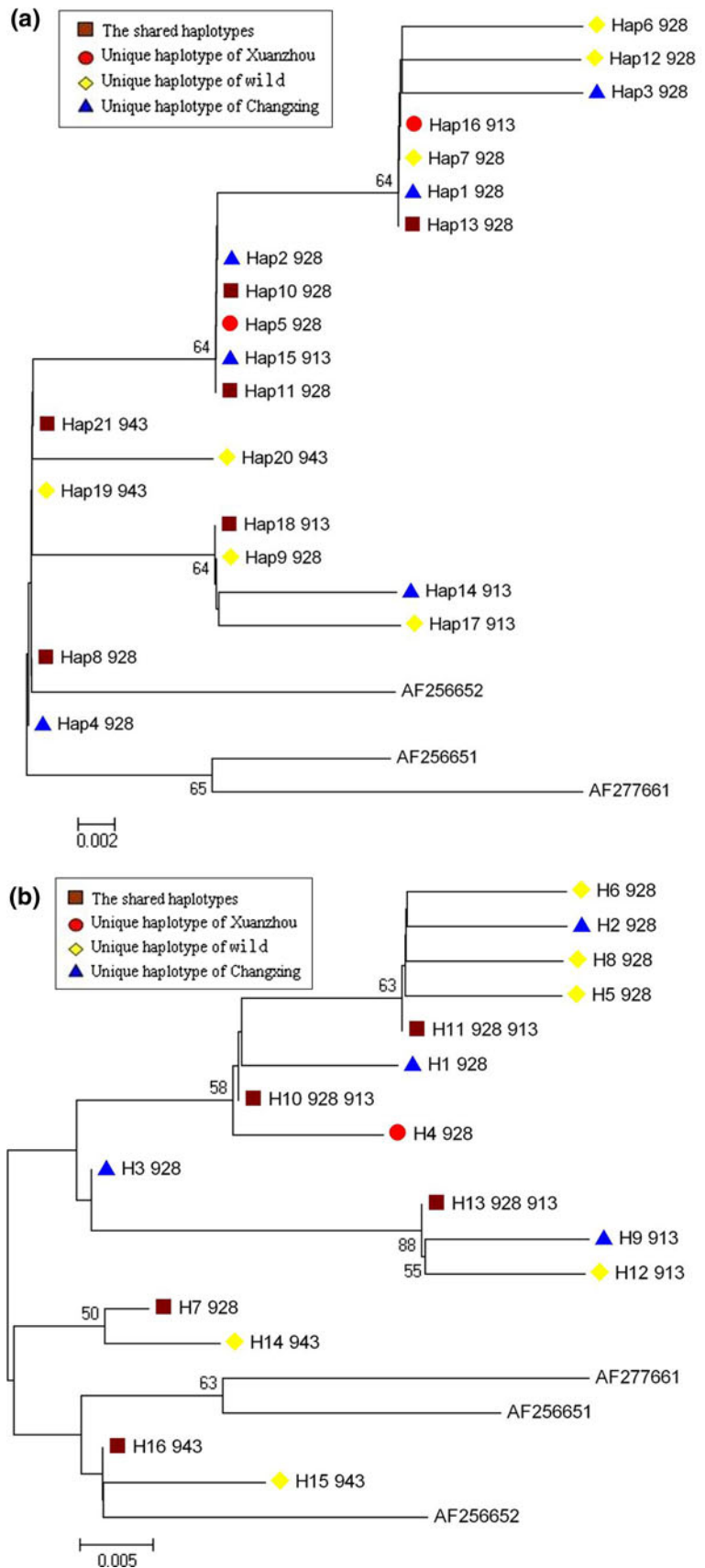
within the exon3 regions, three alleles of *C. crocodilus* were scattered in the alleles of Chinese alligator, suggesting that exon3 had high variation in or between species (Fig. 4).

Discussion

Genetic variation of MHC IIb exon3 complete sequence

In previous studies, 38 variable sites among ten nucleotide sequences and 23 variable sites among amino acid sequences were detected in MHC IIb gene exon2 partial sequences, which were 166 bp long (except one 160 bp) [20]; 34 sequence haplotypes of exon3 were detected in the sampled Chinese alligators, and 83 polymorphic (variable) sites were found within MHC IIb gene exon3 partial sequences (260 bp) [21]. In this study, 11.7 % (33 out of 282) of the nucleotide sites was variable and 16.1 % (15 out of 93) of the amino acid sites was variable. The research based on MHC IIb gene was a more powerful method for genetic diversity of Chinese alligator, this technology could be used for Chinese alligator breeding and releasing in future.

In contrast to the three populations, Wild population had higher levels of MHC IIb diversity. This difference in diversity is interesting, because we would expect uniform distribution of alleles throughout the range. Different patterns of diversity might be explained by differences in population size, as larger populations have more genetic diversity. Changxing population, however, is smaller than Xuangzhou population.

Population fixation index (Fst) represents the genetic differentiation among populations; the larger Fst values are, the higher the degree of differentiation among populations is [35]. In this study, the result that Fst values was very low (Table 4), indicated a lack of isolation among three populations and there were no differentiation. The result was the same as the researches by Wu et al. [3] and Liu et al. [21]. Usually, gene flow values (Nm) less than one indicates a limited group of gene flow, Nm values higher than one may represent large levels of gene flow and genetic exchanging now or in the past [36]. In this study, gene flow (Nm) was 5.77, suggesting that inter-group gene flow may have occurred.

Evidence for purifying selection

An intriguing alternative explanation is that high diversity in MHC genes is driven by balancing selection against diverse pathogens [11, 37]. However, patterns of diversity differ somewhat across exons in the same gene; for instance, the exon3 of class II loci encodes an extracellular domain close to the transmembrane region, which may experience purifying selection [38].

Such selection leaves a characteristic mutational signature in coding regions; in particular, codons that have experienced balancing selection should have more non-synonymous changes than synonymous changes, while codons that experienced purifying selection will have very few non-synonymous changes relative to the number of synonymous changes.

In this study, Z tests indicated that the dN/dS ratio does not differ from neutrality (dS > dN), and a test of purifying selection was significant ($P = 0.001$) (Table 3). This result showed that the exon3 was expected to be under purifying selection.

Conclusion

In this study, we had defined the MHC IIb exon3 complete sequence. The sequence included 282 bp nucleotide and did not show any gaps, the nucleotide sequences were corresponded with putative 93 amino acids peptide. Z tests indicated that the dN/dS ratio does not differ from neutrality (dS > dN), which showed that the exon3 was expected to be under purifying selection. At the same time, we found that the research based on MHC IIb gene was a more powerful method for genetic diversity of Chinese alligator, this technology could be used for Chinese alligator breeding and releasing in future.

References

1. Thorbjarnarson J, Wang XM, Ming S (2002) Wild populations of the Chinese alligator approach extinction. Biol Conserv 103:93–102
2. Wu XB, Gu CM, Zhu JC (2008) Comprehensive study of Anhui Chinese alligator national nature reserve. HeFei University of Technology Press, Anhui
3. Wu XB, Wang YQ, Zhou KY, Zhu WQ, Tong ZZ, Nie JS, Wang CL, Xie WS (2002) Genetic variation in captive population of Chinese alligator, *Alligator sinensis*, revealed by random amplified polymorphic DNA (RAPD). Biol Conserv 106:435–441
4. Wang YQ, Zhu WQ, Huang L, Zhou KY, Wang RP (2006) Genetic diversity of Chinese alligator (*Alligator sinensis*) revealed by AFLP analysis: an implication on the management of captive conservation. Biodivers Conserv 15:2945–2955
5. Wang YQ, Zhu WQ, Wang CL (2003) D-loop sequence variation of mitochondrial DNA in captive Chinese alligator. Acta Genet Sinica 30:425–430

6. Wu XB, Liu H, Xue H, Amato G, Thorbjarnarson J (2007) Low genetic variation with Mitochondrial DNA control region sequence in Chinese Alligator (Alligator sinensis) and implification for its conservation. J Anhui Normal Univ (Nat Sci) 30(3):349–353

7. Huang L, Wang YQ (2004) SSR polymorphism of Alligator sinensis and conservation strategy of genetic diversity. Acta Genet Sinica 31(2):143–150

8. Jing W, Wang XL, Lan H, Fang SG (2009) Eleven novel microsatellite markers for the Chinese alligator (Alligator sinensis). Conserv Genet 10:543–546

9. Xu QH, Fang SG, Wang ZP, Wang ZW (2005) Microsatellite analysis of genetic diversity in the Chinese alligator (Alligator sinensis) Changxing captive population. Conserv Genet 6:941–951

10. Zhu HT, Wu XB, Xue H, Wei L, Hu YL (2009) Isolation of polymorphic microsatellite loci from the Chinese alligator, Alligator sinensis. Mol Ecol Res 9(3):892–894

11. Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proc Natl Acad Sci USA 101:3490–3494

12. Meyers LA, Bull JJ (2002) Fighting change with change: adaptive variation in an uncertain world. Trends Ecol Evol 17:551–557

13. Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16–34

14. van Tienderen PH, de Haan AA, van der Linden CG, Vosman B (2002) Biodiversity assessment using markers for ecologically important traits. Trends Ecol Evol 17:577–582

15. Alcaide M, Edwards S, Negro JJ, Serrano D, Tella J (2008) Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (Falco naumanni). Mol Ecol 17:2652–2665

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

17. Koskinen MT, Haugen TO, Primmer CR (2002) Contemporary fisherian life-history evolution in small salmonid populations. Nature 419:826–830

18. Klein J, Gutknecht J, Fischer N (1990) The major histocompatibility complex and human evolution. Trends Genet 6(1):7–11

19. Karaiskou N, Moran P, Georgitsakis G, Abatzopoulos TJ, Triantafyllidis A (2010) High allelic variation of MHC class II alpha antigen and the role of selection in wild and cultured Sparus aurata populations. Hydrobiologia 638:11–20

20. Shi Y, Wu XB, Yan P, Chen BH (2004) Cloning and sequences analysis of the second exon of MHC class IIB genes in Chinese alligator (Alligator sinensis). Zool Res 25(5):415–421

21. Liu H, Wu XB, Yan P, Jiang ZG (2007) Polymorphism of exon 3 of MHC class IIB gene in Chinese alligator (Alligator sinensis). J Genet Genomics 34(10):918–929

22. Axtner J, Sommer S (2007) Gene duplication, allelic diversity, selection processes and adaptive value of MHC class II DRB genes of the bank vole, clethrionomys glareolus. Immunogenetics 59:417–426

23. Sambrook J, Russel DW (2001) Molecular cloning: a laboratory manual, vol 2 edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

24. Kamel A, Abd-Elsalam (2003) Bioinformatic tools and guideline for PCR primer design. Afr J Biotechnol 2(5):91–95

25. Lu GQ (2004) Software review Vector NTI, a balanced all-in-one sequence analysis suite. Brief Bioinform 5(4):378–388

26. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402

27. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins GD (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

28. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11):1451–1452

29. Exeoffier L, Laval G, Sehneider S (2005) Arlequin ver 3.0: an integrated software package for population genetics data analysis. Evol Bioinform 1:47–50

30. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis, MEGA. software version 4.0. Mol Biol Evol 24:1596–1599

31. Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. Oxford University Press, New York

32. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92:371–373

33. Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. Mol Phylogen Evol 26:1–7

34. Miller HC, Lambert DM (2004) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). Mol Ecol 13:3709–3721

35. López-Fanjul C, Almudena F (2007) Miguel A. The effect of dominance on the use of the QST-FST contrast to detect natural selection on quantitative traits. Genetics 176(1):725–727

36. Slatkin M (1985) Gene flow in natural populations. Ann Rev Ecol Syst 16:393–430

37. Yang G, Yan J, Zhou K, Wei F (2005) Sequence variation and gene duplication at MHC DQB loci of baiji (Lipotes vexillifer), a Chinese river dolphin. J Hered 96:310–317

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