



## MHC class I loci of the Bar-Headed goose (*Anser indicus*)

Qinglong Liang<sup>1,2</sup>, Lei Wei<sup>1</sup>, Xinwei Wang<sup>1</sup> and Hongxuan He<sup>1</sup>

<sup>1</sup>Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, National Research Center For Wildlife-Borne Diseases, Chinese Academy of Sciences, Beijing, China.

<sup>2</sup>Graduate University, Chinese Academy of Sciences, Beijing, China.

### Abstract

MHC class I proteins mediate functions in anti-pathogen defense. MHC diversity has already been investigated by many studies in model avian species, but here we chose the bar-headed goose, a worldwide migrant bird, as a non-model avian species. Sequences from exons encoding the peptide-binding region (PBR) of MHC class I molecules were isolated from liver genomic DNA, to investigate variation in these genes. These are the first MHC class I partial sequences of the bar-headed goose to be reported. A preliminary analysis suggests the presence of at least four MHC class I genes, which share great similarity with those of the goose and duck. A phylogenetic analysis of bar-headed goose, goose and duck MHC class I sequences using the NJ method supports the idea that they all cluster within the anseriforms clade.

**Key words:** MHC I, *Anser indicus*, polymorphism, phylogenetic analysis.

Received: August 8, 2009; Accepted: March 17, 2010.

The major histocompatibility complex (MHC) is a gene complex encoding the receptors which bind endogenous or exogenous antigenic peptides for presentation to cytotoxic T-lymphocytes, an important step for the initiation of the majority of immune responses in vertebrates. Much of the polymorphism in the MHC class I occurs in the regions encoded by exons 2 and 3, which bind the peptides to create variation in binding specificity. This polymorphism is considered important, both for handling a wide range of evolving pathogens and for the health and conservation of species. The bar-headed goose (*Anser indicus*) breeds in Central Asia and migrates southward over the Himalayas to India or northern Burma for the winter.

We chose to study the MHC class I of this species because it may provide relevant information regarding viral disease sensitivity in this group of birds. Successful cloning of an MHC class I gene would make it the first nuclear protein-coding gene studied at the DNA level in this species. Because of the large body of data accumulated on different vertebrates (Jarvi *et al.*, 1999; Kaufman *et al.*, 1999; Westerdahl *et al.*, 1999; Richardson and Westerdahl, 2003; Mesa *et al.*, 2004; Moon *et al.*, 2005; Xia *et al.*, 2005; Alcaide *et al.*, 2008, 2009), several studies have emphasized the potential of MHC genes as valuable molecular markers for assessing the evolutionary and adaptive potential of endangered populations and species in relation to the

menace of changed and emerging diseases (*e.g.*, Yuhki and O'Brien, 1990; Hedrick and Parker, 1998; Garrigan and Hedrick, 2001; Wan *et al.*, 2006; Bollmer *et al.*, 2007). It is important to point out that nearly all the above cited studies investigating MHC class I diversity in non-model avian species have done so at the cDNA level. In the present study, we cloned and sequenced the  $\alpha$  1 domain encoded by exon 2 and the  $\alpha$  2 domain encoded by exon 3 of the bar-headed goose MHC class I heavy chain. We identified at least four MHC class I genes in one individual, which share great similarity with the MHC class I of the goose and duck. These studies will ultimately improve our understanding of the immune response to pathogens in the bar-headed goose and of avian MHC evolution.

Liver samples from six dead bar-headed geese were collected randomly at the same site of Lake Qinghaihu, Qinghai province, China. The samples were stored at -80 °C at the National Research Center for Wildlife-Borne Diseases, Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences. Genomic DNA (gDNA) was extracted from a piece of liver tissue using the TIANamp Genomic DNA Kit. The first pair of conserved regions of the MHC class I primers designed for the amplification was based on the MHC class I sequence of chicken, duck, goose, turkey and quail. BHGMF 5'-GGTTGTGTTACAGGGTCTCA-3' and BHGMR 5'-TAGCCCTTCTCCTTCTCCCT-3' were designed to amplify partial MHC class I heavy-chain sequences, which included the expected size of the fragment amplified by using BHGMF and BHGMR primers.

Another set of primers was designed based on the sequences obtained from the first primers and corresponding sequences from goose and duck. GMF 5'-AGACGGGTGGGGTCTGGA-3' and AIMR 5'-ACTTCCTCTTGGTGATTTGTGCC-3' were designed to amplify partial upstream MHC class I sequences. We used MHC class I sequences from duck (Moon *et al.*, 2005) and goose (Xia *et al.*, 2005) to identify the exons 2 and 3 of the bar-headed goose MHC class I partial sequences which were amplified in our experiments. These two pairs of primers were designed by software Primer Premier 5.0 (PREMIER Biosoft International, Palo Alto, CA)

PCR was performed in a total volume of 30  $\mu$ L, containing 20-50 ng of gDNA, 0.5  $\mu$ M of each primer, and 15  $\mu$ L of 2PCR Taqmix. Cycling conditions were: 94 °C for 5 min, then 30 cycles at 94 °C for 50 s, 58 °C for 50 s, and 72 °C for 1 min; and a final step at 72 °C for 7 min. The PCR products were analyzed on 1% agarose gel stained with ethidium bromide. Purified PCR products were ligated to the pMD18-T vector and then transformed into *E. coli* DH5 $\alpha$ -competent cells. Positive clones were identified by insert release after *Bam*HI and *Hind*III restriction digestion as well as colony PCR and restriction analysis of the PCR products and then sequenced using primer M13. We amplified and cloned DNA from each individual at least three times, and we sequenced at least six positive clones from each individual. The sequences of the six bar-headed geese were submitted to BLAST queries, to ensure that the proper sequences had been amplified. Because Taq polymerase has a significant error rate, nucleotide sequences that varied by only a single base were assumed to represent PCR artifacts and were not identified as unique alleles, unless they were confirmed in a certain number of animals or in separate PCR assays. The nucleotide sequences reported in this article were submitted to GenBank (accession numbers FJ606105 to FJ606113) (Table 1).

Alignments of the bar-headed goose MHC class I gene sequences were obtained by the CLUSTAL X program (1.8) (Thompson *et al.*, 1997) and refined by visual inspection. Manual adjustments of protein-coding nucleotide sequence alignments were facilitated through translation to amino acid sequences. The MEGA 3.1 software (Kumar *et al.*, 2004) was used to estimate the rate of nonsynonymous ( $d_n$ ) and synonymous ( $d_s$ ) substitutions, as

described by Nei and Gojobori (1986). *Anin-A1* and *Anin-A2* PBR amino acid sequences were imported into the NCBI Molecular Modeling Database to compare the tertiary structure with the structure of human MHC class I (*HLA-A2*). A phylogenetic tree was constructed with MEGA 3.1 from the predicted amino acids plus previously published vertebrate sequences, using the Neighbor-Joining (NJ) method.

In the six bar-headed geese studied, we confirmed the presence of five sequences of exon 2: *Anin-A1\*01*, *Anin-A1\*02*, *Anin-A1\*03*, *Anin-A1\*04*, and *Anin-A1\*05*. By this nomenclature, "A1" refers to the MHC class I  $\alpha$  1 chain gene, while the first two letters of *Anin* are derived from the genus name *Anser* and the last two from the species name *indicus* (Klein *et al.*, 1990; Ellis *et al.*, 2006). Two of these sequences, *Anin-A1\*01* and *Anin-A1\*03*, were found in all six animals, while *Anin-A1\*02*, *Anin-A1\*04* and *Anin-A1\*05* were present in only one of them (Table 1). The amino acid translations deduced from these five sequences produced five unique protein sequences, aligned in Figure 1a. A comparison to the GenBank database revealed that the bar-headed goose MHC class I exon 2 sequences shared great nucleotide similarity with goose and duck MHC class I (79.1% to 91.6%), while their amino acid similarity ranged from 65.9% to 86.3% with goose MHC class I and 65.9% to 73.8% with duck MHC class I, respectively.

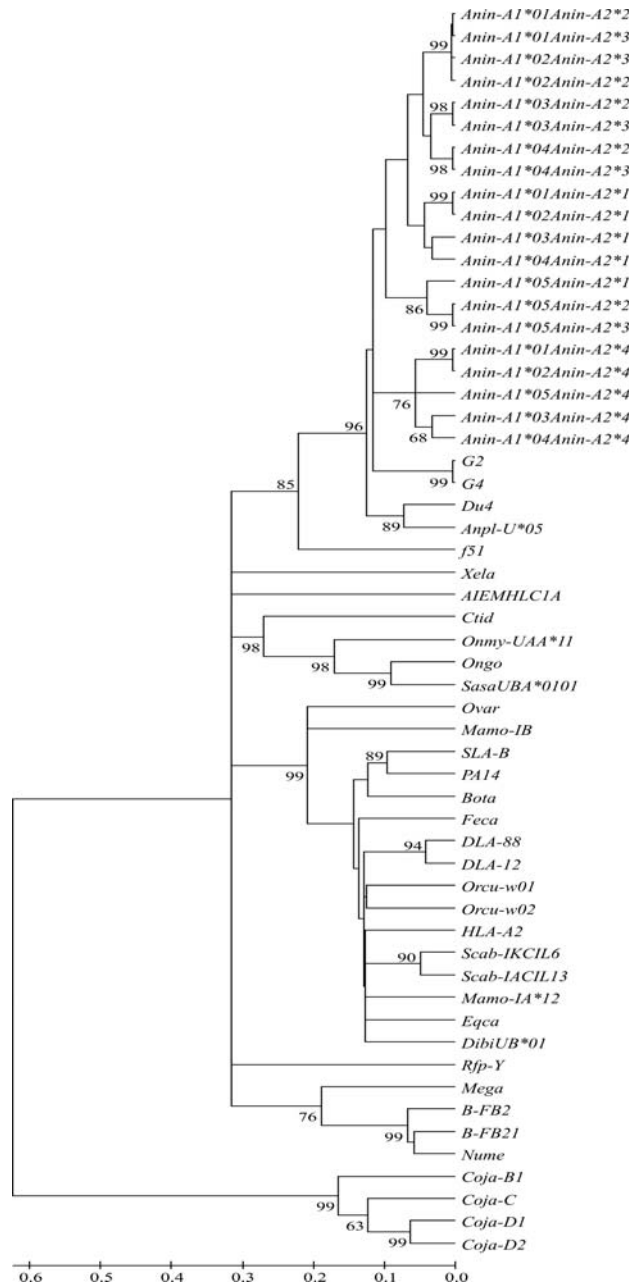
Positive clones sequenced from the six bar-headed geese studied yielded four distinct nucleotide sequences. An alignment of the amino acid translation of the four unique bar-headed goose sequences *Anin-A2\*01*, *Anin-A2\*02*, *Anin-A2\*03* and *Anin-A2\*04* is shown in Figure 1b. "A2" refers to the MHC class I  $\alpha$  2 chain gene. A comparison to the GenBank database revealed high nucleotide similarity of bar-headed goose exon 3 sequences to goose and duck MHC-I (85.8% to 90.5%) and a high amino acid similarity between them and the MHC class I of goose and duck, which ranged from 86.9% (goose) to 82.6% (duck).

The amino acid sequences deduced for the bar-headed goose MHC class I exons 2 and 3 correspond, respectively, to amino acid positions 26-113 in the  $\alpha$  1 domain and 114-205 in the  $\alpha$  2 domain of the duck MHC class I (Mesa *et al.*, 2004). To determine which genes could be involved in antigen presentation, we examined the amino acid sequences for hallmarks of classical MHC class I

**Table 1** - The sequences of exon 2 and exon 3 in the six bar-headed geese studied.

| Exon 2            | GenBank accession number | Animal ID                    | Exon 3            | GenBank accession number | Animal ID               |
|-------------------|--------------------------|------------------------------|-------------------|--------------------------|-------------------------|
| <i>Anin-A1*01</i> | FJ606105                 | A01, A02, A03, A04, A05, A06 | <i>Anin-A2*01</i> | FJ606110                 | A01, A02, A03, A05, A06 |
| <i>Anin-A1*02</i> | FJ606106                 | A06                          | <i>Anin-A2*02</i> | FJ606111                 | A01, A02, A04, A05, A06 |
| <i>Anin-A1*03</i> | FJ606107                 | A01, A02, A03, A04, A05, A06 | <i>Anin-A2*03</i> | FJ606112                 | A03, A04                |
| <i>Anin-A1*04</i> | FJ606108                 | A01                          | <i>Anin-A2*04</i> | FJ606113                 | A01                     |
| <i>Anin-A1*05</i> | FJ606109                 | A06                          |                   |                          |                         |





**Figure 2** - Phylogenetic tree created by the NJ method based on the amino acid sequences of bar-headed goose MHC class I  $\alpha 1$  and  $\alpha 2$  domains and other vertebrates. Only bootstrap levels exceeding 60% are shown (as a percentage) and are based on 1,000 replications. The sequence sources are as follow: goose (*G2*, AY387649; *G4*, AY387651), duck (*Du4*, AB115244; *Anpl-U\*05*, AY294419), human (*HLA-A2*, K02883), rainbow trout (*Onmy-UAA\*11*, AB012064), chicken (*B-FB2*, AF013492; *Rfp-Y*, AF218783; *B-FB21*, AF013493), turkey (*Mega*, DQ993255), helmeted guineafowl (*Nume*, EF643463), grass carp (*Ctld*, AB109779), African clawed frog (*Xela*, L20733), dog (*DLA-88*, U55028; *DLA-12*, U55026), pig (*SLA-B*, AY135599; *PA14*, AF014001), *Ameiva ameiva* (*AIEMHLC1A*, M81094), pink salmon (*Ongo*, D58386), Atlantic salmon (*SasaUBA\*0101*, AF504019), horse (*Eqca*, X71809), domestic cat (*Feca*, U07674), *Diceros bicornis minor* (*DibiUB\*01*, AF055347), sheep (*Ovar*, M34676), cattle (*Bota*, Y09208), rabbit (*Orcu-w01*, K02442; *Orcu-w02*, K02441), woodchuck (*Mamo-IA\*12*, AF146092; *Mamo-IB*, AF201912), tassel-eared squirrel (*Scab-ICKIL6*, M97618; *Scab-IACIL13*, M97617), Japanese quail (*Coja-C*, BAA83671; *Coja-D1*, BAC82515; *Coja-D2*, BAC82516; *Coja-B1*, BAC82519), Florida sandhill crane (*f51*, AF033106).

synonymous and nonsynonymous substitutions were calculated at probable pocket residues and at the codons that are nonpocket residues for the species studied (Table 2). Nonsynonymous rates were higher at the pocket residues than at the nonpockets in the  $\alpha 1$  domain. At the pocket residues of the  $\alpha 1$  domain,  $d_N$  was higher than  $d_S$  for any species ( $d_N/d_S > 1$ ), while at the nonpocket residues,  $d_N$  was not higher than  $d_S$  in any of the three species.

A phylogenetic analysis of bar-headed goose MHC class I sequences was conducted using the NJ method. Bootstrap values supported bar-headed goose and goose MHC class I exon 2 and exon 3 sequences clustering within the anseriform clade at the order level, and with anseriforms (goose and duck) at the family level (Figure 2). Bar-headed goose exons 2 and 3 clustered with strong support with anseriforms MHC class I sequences. The position of bar-headed goose MHC class I exon 2 and exon 3 sequences in the same clade with goose and duck, along with long branch lengths (Figure. 2), suggests an ancient history of these MHC class I sequences.

To our best knowledge, this is the first report on the isolation of polymorphism patterns in classical MHC class I genes of the bar-headed goose and one of the very few studies of class I gene structure in non-model avian species. Although we did not perform gene expression analyses in this study, other studies have generally observed a correlation between signatures for balancing selection and level of expression of MHC genes (Zoorob *et al.*, 1990; Jacob *et al.*, 2000).

In conclusion, we present here the first MHC class I sequences of the bar-headed goose, a worldwide migrant bird. The molecular methods and sequence data collected in this paper should contribute to a better understanding of the evolutionary significance and conservation implications of the MHC in this species. Moreover, the group-specific primers designed for this study targeted highly conserved regions across several kinds of bird MHC class I genes, and therefore similar fragments of other avian groups are likely to be cross-amplified successfully. Given that MHC genes can decisively determine virus and parasite resistance (Hedrick, 2001), this study may also aid in the preservation of the genetic diversity of the bar-headed goose.

## Acknowledgments

The study was supported by grants from the Chinese Academy of Sciences (CAS) Innovation Program (KSCX2-YW-N-063), National Key Basic Research and Development Program of China (9732007BC109103), United States Department of Agriculture (USDA) and International Development Research Centre (IDRC).

## References

Alcaide M, Edwards SV, Negro JJ, Serrano D and Tella JL (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.
- Alcaide M, Edwards SV, Cadahia SV and Negro JJ (2009) MHC class I genes of birds of prey: Isolation, polymorphism and diversifying selection. *Conserv Genet* 10:1349-1355.
- Bollmer JL, Vargas FH and Parker PG (2007) Low MHC variation in the endangered Galapagos penguin (*Spheniscus mendiculus*). *Immunogenetics* 59:593-602.
- Ellis SA, Bontrop RE, Antczak DF, Ballingall K, Davies CJ, Kaufman J, Kennedy LJ, Robinson J, Smith DM, Stear MJ *et al.* (2006) ISAG/IUIS-VIC comparative MHC nomenclature committee report, 2005. *Immunogenetics* 57:953-958.
- Garrigan D and Hedrick PW (2001) Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. *Immunogenetics* 53:483-489.
- Hedrick PW (2001) Conservation genetics: Where are we now? *Trends Ecol Evol* 16:629-636.
- Hedrick PW and Parker KM (1998) MHC variation in the endangered *Gila topminnow*. *Evolution* 52:194-199.
- Jacob JP, Milne S, Beck S and Kaufman J (2000) The major and a minor class II beta-chain (B-LB) gene flank the Tapasin gene in the B-F /B-L region of the chicken major histocompatibility complex. *Immunogenetics* 51:138-147.
- Jarvi SI, Goto RM, Gee GF, Briles WE and Miller MM (1999) Identification, inheritance, and linkage of B-G-like and MHC class I genes in cranes. *J Hered* 90:152-159.
- Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R and Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401:923-925.
- Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER, Jones PP, Parham P, Wakeland EK and Watkins DI (1990) Nomenclature for the major histocompatibility complexes of different species: A proposal. *Immunogenetics* 31:217-219.
- Kumar S, Tamura K and Nei M (2004) MEGA 3.1: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150-163.
- Mesa CM, Thulien KJ, Moon DA, Veniamin SM and Magor KE (2004) The dominant MHC class I gene is adjacent to the polymorphic TAP2 gene in the duck, *Anas platyrhynchos*. *Immunogenetics* 56:192-203.
- Moon DA, Veniamin SM, Parks-Dely JA and Magor KE (2005) The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed class I genes. *J Immunol* 175:6702-6712.
- Nei M and Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418-426.
- Richardson DS and Westerdahl H (2003) MHC diversity in two *Acrocephalus* species: The outbred Great reed warbler and the inbred Seychelles warbler. *Mol Ecol* 12:3523-3529.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997) The clustal-X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876-4882.
- Wan QH, Zhu L, Wu H and Fang SG (2006) Major histocompatibility complex class II variation in the giant panda (*Ailuropoda melanoleuca*). *Mol Ecol* 15:2441-2450.
- Westerdahl H, Wittzell H and von Schantz T (1999) Polymorphism and transcription of MHC class I genes in a passerine bird, the great reed warbler. *Immunogenetics* 49:158-170.
- Xia C, Lin CY, Xu GX, Hu TJ and Yang TY (2004) cDNA cloning and genomic structure of the duck (*Anas platyrhynchos*) MHC class I gene. *Immunogenetics* 56:304-309.
- Xia C, Hu T, Yang T, Wang L, Xu G and Lin C (2005) cDNA cloning, genomic structure and expression analysis of the goose (*Anser cygnoides*) MHC class I gene. *Vet Immunol Immunopathol* 107:291-302.
- Yuhki N and O'Brien SJ (1990) DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. *Proc Natl Acad Sci USA* 87:836-840.
- Zoorob R, Behar G, Kroemer G and Auffray C (1990) Organization of a functional chicken class II B gene. *Immunogenetics* 31:179-187.

Associate Editor: Fábio de Melo Sene

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.