Greater prairie chickens have a compact MHC-B with a single class IA locus

J. A. Eimes · K. M. Reed · K. M. Mendoza · J. L. Bollmer · L. A. Whittingham · Z. W. Bateson · P. O. Dunn

Received: 18 July 2012 / Accepted: 22 October 2012
© Springer-Verlag Berlin Heidelberg 2012

Abstract The major histocompatibility complex (MHC) plays a central role in innate and adaptive immunity, but relatively little is known about the evolution of the number and arrangement of MHC genes in birds. Insights into the evolution of the MHC in birds can be gained by comparing the genetic architecture of the MHC between closely related species. We used a fosmid DNA library to sequence a 60.9-kb region of the MHC of the greater prairie chicken (Tympanuchus cupido), one of five species of Galliformes with a physically mapped MHC. Greater prairie chickens have the smallest core MHC yet observed in any bird species, and major changes are observed in the number and arrangement of MHC loci. In particular, the greater prairie chicken differs from other Galliformes in the deletion of an important class I antigen binding gene. Analysis of the remaining class IA gene in a population of greater prairie chickens in Wisconsin, USA revealed little evidence for selection at the region responsible for antigen binding.

Keywords MHC · Fosmidlibrary · Galliformes · Comparative genomics · Tympanuchus cupido

Introduction

The major histocompatibility complex (MHC) is a family of genes found in all jawed vertebrates (Trowsdale 1988) that facilitates the adaptive immune response to foreign parasites and pathogens (Janeway et al. 2005). Comparisons across distantly related taxa have revealed major differences in the size, number, and arrangement (i.e., architecture) of MHC loci (Kelley et al. 2005). For example, the MHC in galliform birds appears to be dramatically more compact than that of mammals. The core MHC-B of the domestic chicken (Gallus gallus) is 1/20th the size of the MHC in humans (HLA) and contains just 19 coding genes compared with over 60 in humans (Guillemot et al. 1988; Kelley et al. 2005).

The compact nature of the MHC in domestic chickens and strong associations between individual haplotypes and disease resistance led to the description of the chicken MHC as a "minimal essential MHC" (Kaufman et al. 1995). Kaufman (1999) hypothesized that in birds, fewer antigen presenting molecules will recognize fewer pathogens, but with lessened risk of autoimmune disorders, while in mammals, the larger suite of expressed molecules protects against more pathogens but at the expense of greater risk of autoimmunity. For example, in humans, there are up to six expressed MHC class I genes, and several associations with diseases and specific loci have been described (reviewed by Shiina et al. 2004a). Domestic chickens, in contrast, have only two class I genes, BF1 and BF2, and BF2 is dominantly expressed, and it has been hypothesized that BF2 plays a role in resistance to Rous sarcoma virus and Marek's disease (Kaufman and Venugopal 1998; Wallny et al. 2006).

In addition to the domestic chicken, the core MHC-B has been sequenced in four additional species within the Galliformes: domestic turkey (Meleagris gallopavo), golden pheasant (Chrysolophus pictus), the black grouse (Tetrao...
tetrix, Genbank accession number AC:JQ028669), and Japanese quail (Coturnix japonica), and the MHC has many fewer genes, less intergenic space and smaller introns than the mammalian MHC (Shina et al. 2004a, b; Chaves et al. 2009; Ye et al. 2012). Assembled sequences of the Galliform MHC-B suggest that even among closely related taxa the gene number, order and orientation are variable. For example, the MHC-B of the turkey and Japanese quail contain several BG genes between Blec2 and BTN2, while the domestic chicken has just one (Shina et al. 2004a, b; Chaves et al. 2009). The domestic chicken has two MHC-B class IIB genes (Kaufman et al. 1999), whereas the turkey and golden pheasant have at least three (Chaves et al. 2009; Ye et al. 2012), and in the Japanese quail, block duplications have generated up to eight MHC failed, suggesting a different and our efforts to amplify class I and other regions of the core gested variation in class II gene number between individuals, Previous work on this species (Eimes et al. 2011) had sug-

To further examine variation in the MHC of the locally threatened species, the greater prairie chicken (Tympanuchus cupido), we produced a physical map of the MHC-B region. Previous work on this species (Eimes et al. 2011) had suggested variation in class II gene number between individuals, and our efforts to amplify class I and other regions of the core MHC failed, suggesting a different MHC-B genetic architecture than other Galliformes. The MHC-B assembly presented here covers the region extending 60.9 kb from the BTN2 gene to the centromere protein gene CenpA. Our results show that the genetic architecture of the MHC-B, while generally conserved in Galliformes, exhibits some important variations, which is in part due to major inversions and deletions that have occurred between closely related species. In addition to a physical map of the MHC in the greater prairie chicken, we investigated genetic variation and natural selection on the previously undescribed (in this species) class I peptide binding gene.

Materials and methods

Fosmid library construction

gDNA insert preparation

The library was constructed from DNA extracted from whole blood from an adult female captured in Polk County, Minnesota in 2008. High molecular weight genomic DNA was extracted using the Masterpure Complete DNA Purification Kit (Epicentre Biotechnologies, USA) and resuspended in TE buffer at a concentration of 0.5 μg/μl. To generate approximately 40 kb DNA fragments for insertion into the fosmid vector, 40 μg of purified gDNA was randomly sheared by passing the sample through a 200-μl pipette tip 50 times. To confirm correct size of the insert, 1.0 μl of the resulting DNA was run on a 20-cm long, 0.8 % agarose gel overnight at 50 V using the Fosmid Control DNA (Epicentre) (40 kb) as a size marker. The gDNA was additionally sheared by vortexing for 40-s intervals until an intact, 40 kb fragment was generated. We increased the efficiency of the insert ligation (on which all subsequent cloning steps are dependent) by omitting further size selection of the gDNA. The MaxPlax Lambda Packaging Extracts, EPI300-T1 Plating Strain (pCC1FOS, Epicentre) preferentially accepts inserts of 30–40 kb and, thus, size selects the fragment during the packaging step.

Cloning

Single-copy gDNA fragments (~40 kb) were cloned using the CopyControl™ Fosmid Library Production Kit with pCC1FOS Vector and Phage T-1 Resistant EPI300™-T1R Escherichia coli Plating Strain (Epicentre). We end-repaired 15.0 μg of the DNA fragments using the Epicentre Fosmid kit according to the manufacturer's instructions, and 0.25 μg of that DNA was ligated into 0.5 μg pCC1FOS vector DNA. Transformed fosmids were packaged into MaxPlax Lambda Packaging Extract (Epicentre) and processed according to the manufacturer's instructions followed by infection of EPI300-T1R E. coli cells (Epicentre). Infected bacteria were spread on 50 LB plates (22 cm) supplemented with 12.5 mg/mL chloramphenicol and incubated at 37 °C overnight.

Colony screening

Colonies were harvested from each plate by flooding with 1.5 ml LB and gently scraping 700 μl of the mixture into 1 ml microcentrifuge tubes supplemented with 20 % glycerol for immediate storage at –80 °C. The remaining mixture from each plate was placed into a separate microcentrifuge tube for DNA extraction using the Epicentre FosmidMax Purification kit. Each extracted fosmid sub-library was then screened for MHC genes using the PCR primers Blex2F (Eimes et al. 2010) and E3R (Chaves et al. 2010), which together amplify a 279-bp fragment of the MHC class II gene encoding the β chain, and the primers pcTAP2F1 and pcTAP2R1 which amplify a 380-bp fragment of the TAP2 antigen peptide transporter gene in the MHC class I (Table 1). The following PCR conditions were used: 50.0–100.0 ng fosmid DNA, 0.5 U of HotStar Taq polymerase (Qiagen, USA), 4.0 μl of Q solution (Qiagen) and final concentrations of 0.5 μM of each primer, 200 μM each dNTP and 1× of HotStar Taq PCR Buffer (Qiagen) and water to a total volume of 25.0 μl. Cycling conditions were: 15 min at 95 °C initial activation followed by 30 cycles of 98 ° C for 20 s, 60 °C for 30 s, and 72 °C for 30 s.

Positive sub-libraries were identified by sequencing the PCR amplicons and confirming homology to MHC genes in the domestic turkey and chicken using BLAST. The corresponding glycerol stock (200 μl) from each positive
A sub-library (one TAP1 and one class II) was diluted with LB and plated at a density of 200–300 colonies per plate for secondary screening. Colonies were harvested and DNA was extracted and screened by PCR as described above. A positive sub-library for both class I and class II was identified and diluted to 200–300 colonies on a single plate. Whole colony PCR (using conditions described above) was performed until single positive colonies for each MHC class were identified. Single positive colonies were cultured in 7 ml LB+chloramphenicol and 10.5 μl induction solution in a 14-ml culture tube and shaken overnight at 37 °C at 215 rpm. The fosmid DNA was extracted and prepared for sequencing using the Epicentre FosmidMax Purification Kit.

### Sequencing and contig construction

The MHC class II-positive fosmid was sequenced at Research and Testing Laboratory, LLC (Lubbock, TX) using a Roche 454 FLX Genome Sequencer. Because of the relatively small size of the fosmid sequence, only a very small fraction of a full 454 plate was used for sequencing. The 33,559 raw 454 sequence reads (average length of 513 bp) were initially assembled with Sequencher software (Gene Codes Inc, Ann Arbor, MI.) and contigs compared to the NCBI sequence databases by BLAST. After removal of E. coli sequences and quality trimming, the remaining reads (4,357, average length 273 bp) were de novo-assembled and aligned to the turkey MHC as a reference. The resulting seven contigs were manually edited to resolve homopolymer ambiguities resulting in a single sequence of 27.3 kb, representing approximately 40× coverage (maximum depth 119). The TAP1 positive fragment was sequenced using an Illumina HiSeq 2000 at the Biomedical Genomics Center, University of Minnesota. Raw Illumina data (49,976,444 reads representing a subset of a single flow cell) were trimmed for vector and Illumina adaptors and process control sequences and double filtered to remove fosmid vector and E. coli sequences using BLAST and a custom Perl scripts. The resulting sequences (35,522,364 reads with an average length of 76 bp) were then assembled with VELVET (Zerbino and Birney 2008) using standard defaults and two different k-mer lengths (57 and 63). The assembly from the 63 k-mer resulted in 1,019 sequence contigs ranging in size from 66 to 5,733 bp. These were aligned to the turkey whole genome sequence using BLAST to assign contigs to position. The VELVET contigs were assembled using Sequencher and positioned relative to each other based on alignment to the turkey and chicken MHC assemblies resulting in a final assembly of 30 kb with nine gaps averaging 33 bp.

### Table 1  PCR primers and conditions used to amplify regions of the prairie chicken MHC-B

<table>
<thead>
<tr>
<th>Locus/prime</th>
<th>Primer sequence</th>
<th>Product size (bp)</th>
<th>Annealing temp (°C)</th>
<th>Extension time</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTN2 F2</td>
<td>cagggctgtgggtcttg</td>
<td>2,850</td>
<td>58</td>
<td>4 m</td>
</tr>
<tr>
<td>BLEC1 R3</td>
<td>caaaggctgtgggtcttg</td>
<td>2,850</td>
<td>58</td>
<td>4 m</td>
</tr>
<tr>
<td>TAPBP F2</td>
<td>tggctgtgggtcttggtc</td>
<td>2,850</td>
<td>58</td>
<td>4 m</td>
</tr>
<tr>
<td>BRD2 R2</td>
<td>cagggctgtgggtcttg</td>
<td>2,850</td>
<td>58</td>
<td>4 m</td>
</tr>
<tr>
<td>TAPBP 18 F</td>
<td>cagggctgtgggtcttg</td>
<td>2,850</td>
<td>58</td>
<td>4 m</td>
</tr>
<tr>
<td>TAPBP 18 R</td>
<td>cagggctgtgggtcttg</td>
<td>2,850</td>
<td>58</td>
<td>4 m</td>
</tr>
<tr>
<td>IIB-2 15 F</td>
<td>cagggctgtgggtcttg</td>
<td>2,850</td>
<td>58</td>
<td>4 m</td>
</tr>
<tr>
<td>IIB-2 15 R</td>
<td>cagggctgtgggtcttg</td>
<td>2,850</td>
<td>58</td>
<td>4 m</td>
</tr>
</tbody>
</table>

**Immunogenetics**

Author’s personal copy

---

*E. coli* sequences and quality trimming, the remaining reads (4,357, average length 273 bp) were de novo-assembled and aligned to the turkey MHC as a reference. The resulting seven contigs were manually edited to resolve homopolymer ambiguities resulting in a single sequence of 27.3 kb, representing approximately 40× coverage (maximum depth 119). The TAP1 positive fragment was sequenced using an Illumina HiSeq 2000 at the Biomedical Genomics Center, University of Minnesota. Raw Illumina data (49,976,444 reads representing a subset of a single flow cell) were trimmed for vector and Illumina adaptors and process control sequences and double filtered to remove fosmid vector and *E. coli* sequences using BLAST and a custom Perl scripts. The resulting sequences (35,522,364 reads with an average length of 76 bp) were then assembled with VELVET (Zerbino and Birney 2008) using standard defaults and two different k-mer lengths (57 and 63). The assembly from the 63 k-mer resulted in 1,019 sequence contigs ranging in size from 66 to 5,733 bp. These were aligned to the turkey whole genome sequence using BLAST to assign contigs to position. The VELVET contigs were assembled using Sequencher and positioned relative to each other based on alignment to the turkey and chicken MHC assemblies resulting in a final assembly of 30 kb with nine gaps averaging 33 bp.
Gaps and ambiguities in the assemblies were resolved by PCR using the primers and conditions in Table 1. Additional long-range PCR was performed to confirm gene presence/absence and orientation for MHC regions in the prairie chicken that did not align with the turkey or chicken (Table 1). gDNA from the same individual used to construct the fosmid library was used to amplify the approximately 4.5 kb region separating the two fosmid fragments (DMB1 to DMB2) and generate one contiguous MHC-B fragment of class I, II, and III genes. PCR products that exceeded 1.5 kb were sequenced by primer walking. All PCR products were sequenced on Applied Biosystems (Carlsbad, CA, USA) 3730XL and 3130 automated DNA sequencers (University of Chicago Cancer Research Center).

Gene identification, annotation, and sequence analyses

Final sequences were aligned with the fully annotated sequence of the turkey and chicken MHC for comparisons of predicted gene sequences, gene order and orientation and expressed sequence tags (ESTs) using Sequencher and Geneious version 5.4 (Biomatters, Auckland, NZ) (Figs. 1 and 2). Gene annotations, CG content analysis, taxonomic comparisons of gene nucleotide similarity (Table 2) and sequence similarity dot matrix plots (Fig. 3) were made using Geneious software. MHC nomenclature is consistent with that used for the turkey and chicken (Chaves et al. 2009; Shiina et al. 2004a, b).

In order to estimate polymorphism and test for selection at the single class I locus, IA, we sampled 30 unrelated individuals from a locally threatened population (N=~1,200) in Portage County, Wisconsin (provenance data available in Johnson et al. (2003)). Individuals were genotyped over the entire exon 3, which, along with exon 2, codes for the peptide binding region (PBR) in domestic chickens (Bjorkman et al. 1987; Wallny et al. 2006). A 674-bp fragment extending from exon 2 to exon 4 was amplified using the PCR and Sanger sequencing conditions described above and the primers pcIA ex3F and pcIA ex3R (Table 1). We aligned the 273-bp exon 3 sequences using Geneious and identified heterozygous sites from chromatograms. For four individuals, some polymorphic sites could not be unambiguously assigned, and these individuals were cloned, and haplotypes were identified by sequencing 12 colonies from each PCR insert. We cross-checked genotypes from cloned sequences with those from direct sequencing to eliminate false polymorphisms often associated with cloning (Alcaide et al. 2011). We then used the program PHASE (Stephens et al. 2001) in the DnaSP v.5.0 software package (Librado and Rozas 2009) to identify specific haplotypes for each individual. PHASE employs Bayesian algorithms for inferring haplotypes from genotype data in a population sample. PHASE assigned haplotypes using information based on heterozygous genotypes, the cloned haplotypes and the 13 homozygous individuals that were entered into the program. \( ds/dN \) ratios (Codon Z test) were calculated using the Nei and Gojobori (Jukes-Cantor correction) method of overall means in Mega 5.10 (Nei and Gojobori 1986; Tamura et al. 2011). Estimates of nucleotide diversity were also calculated in Mega 5.10. In addition, an a priori test for selection on individual codons was performed using the program CODEML, found in the PAML 4.0 statistical software package (Yang 2007). The CODEML models used in this study were M7 and M8 which have been shown to provide robust estimates for selection at highly polymorphic loci (Nielsen and Yang 1998; Bos and Waldman 2006; Alcaide et al. 2008).

In order to compare nucleotide diversity and diversifying selection of the class I PBR of prairie chickens with other avian species, single locus class I exon 3 sequences were obtained from NCBI for the domestic chicken (Acc No. AM282692-AM282700), the Eurasian kestrel (Falco tinnunculus; Acc No. EU120698–EU120722), and the lesser kestrel (Falco naumanni; Acc No. JF831087-JF831120). We chose these species because they are the only taxa for which an adequate number of confirmed, single locus class I sequences were available. In this study, the PBR is study-defined as those codons identified from analysis of structural models of the peptide binding grooves in class I proteins of the domestic chicken (Bjorkman et al. 1987; Wallny et al. 2006; Koch et al. 2007; Promerova et al. 2009).

cDNA

Expression of class I and II genes was confirmed using RNA extracted from the spleens of two adult female prairie chickens from Minnesota found recently deceased in winter of 2009. Spleen tissue (5 mg) was preserved in RNAlater (Qiagen, USA) and RNA was extracted and DNase treated using the RNeasyMini Kit (Qiagen), and cDNA was synthesized using the QuantiTect Reverse Transcription Kit (Qiagen) using the manufacturer’s protocol. The IA gene was amplified using the primers IAex2F and IAex3R (Table 1), and the class II B genes (IIB1 and IIB2) were amplified using primers RNA Fl (Strand et al. 2007) and E3R (Chaves et al. 2010). The primers RNA Fl and E3R amplified cDNA to produce a 167-bp fragment that extended from the 108th bp of exon 2 to the 85th bp of exon 3 (minus intron 2). Class II B PCR products were cloned using Clonesmart blunt-end high-count cloning vectors.
kanamycin cloning kits with chemically competent cells (Lucigen, Middleton, WI, USA). Sequences and annotated map (Fig. 1) are deposited online at Genbank (accession numbers pending).

Fig. 1 An annotated map of the B-F/BL region of the MHC of the greater prairie chicken (JX971120). Black = entire genes; Gray = predicted coding DNA sequences; White = fosmid sequences; Black triangles = primer positions.

Fig. 2 The comparative genomics of the core MHC-B in five species of Galliformes. The phylogeny is from Kriegs et al. (2007). (Tycu = Tymananchus cupido; Chpi = Chrysolophus pictus; Mega = Meleagris gallopavo; Gaga = Gallus gallus; Coja = Coturnix japonica). Arrows indicate major gene inversions relative to the domestic chicken. Although not to scale, the maps reflect the relative sizes and densities of the MHC-B in each species. Note the absence of a second class I peptide binding gene in the Tycu MHC. Maps (other than Tycu) are from Shiina et al. 2004a, b (Coja), Chaves et al. 2009 (Mega), Ye et al. 2012 (Chpi), and Jacob et al. 2000 (Gaga).
Results

Gene number and homology

The two fosmid clones and intervening sequence spanned 60.9 kb from exon 7 of the butyrophilin-like gene (BTN2) to exon 4 of the centromere protein A gene (CenpA) (Genbank Accession number: JX971120), a region that is homologous to both the domestic chicken (AP011531) and turkey (DQ993255) (Fig. 1). The MHC B-F/B-L region of the prairie chicken is more compact than the domestic chicken or turkey. The region spanning the end of BTN2 to the end of C4 was 59,800 compared to 76,190 bp in chicken and 90,270 bp in turkey (Fig. 2). The sequence contained just 14 genes (compared to 17 and 20 in chicken and turkey, respectively). The 14 predicted genes included a portion (1,487 bp) of BTN2, one C-type lectin-like gene (Blec), two classical MHC class IIB (β chain) loci, the TAP associated glycoprotein (TAPBP), the bromodomain-containing protein 2 gene (BRD2), three non-classical class II domains: α1 (DMA), β1 (DMB1) and IgG1 β2 (DMB2), two class 1 transporter proteins (TAP1 and TAP2), one classical MHC class I locus (IA; α1 and α2 chains), a complement protein (C4) and a portion (863 bp) of CenpA.

Rearrangement of BG and Blec2 genes

All other Galliformes sequenced to date have at least one BG gene located between BTN2 and Blec1 (Shiina et al. 2004a, b; Chaves et al. 2009; Ye et al. 2012), but this gene was not present in this region of the prairie chicken MHC (Fig. 2). Also absent from this region of the MHC-B of the prairie chicken was the putative NK cell receptor Blec2 (NK2 in the golden pheasant), which is conserved in the core MHC-B of other Galliformes (Shiina et al. 2004a, b; Chaves et al. 2009; Ye et al. 2012). We tested for the presence of these genes in other regions of the prairie chicken genome by designing PCR primers from conserved regions of the turkey and domestic chicken BG1 as well as Blec2 genes (Table 1). We successfully amplified both BG1 and Blec2-like fragments in the greater prairie chicken outside the B-F/B-L MHC region. The Blec2 fragment was most similar to exon 3 and intron 3 of Blec2 in the turkey (92 % pairwise sequence identity). Given the duplicated nature of the BG genes in other Galliformes, homology of this amplified fragment is uncertain. The amplified fragment aligned to exon 1 and intron 1 of several Galliformes BG genes but was most similar to BG1 of the domestic chicken (81.0 % pairwise sequence identity). When the comparison was restricted to the turkey, the fragment was most similar to BG3 (78 % pairwise sequence identity).

Loss of class I locus and inversion of TAP genes

Relative to the domestic chicken and turkey, the TAP1 and TAP2 gene block was inverted in the greater prairie chicken, and the second IA gene was absent (Fig. 2). Long-range PCR between DMB2 and TAP2 confirmed the absence of the second IA gene in the prairie chicken core MHC-B, and it was also absent in the gDNA from 12 additional, unrelated birds from Wisconsin.

A single class IA gene, IA, was found between TAP1 and C4 (Fig. 2). Based on its location and expression, this gene is probably homologous to BF2 in domestic chickens (IA2 in turkeys). Thirty additional Wisconsin birds were tested for the presence of class IA genes outside the MHC-B region by amplifying the polymorphic region of IA (Wallny et al.

<table>
<thead>
<tr>
<th>Gene</th>
<th>CDS size (bp)</th>
<th>Predicted protein size (aa)</th>
<th>No. exons</th>
<th>Percent nucleotide identity</th>
<th>Percent AA sequence similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gallus</td>
<td>Turkey</td>
</tr>
<tr>
<td>Blec 1</td>
<td>567</td>
<td>189</td>
<td>5</td>
<td>91.7</td>
<td>94.2</td>
</tr>
<tr>
<td>Class II IIB1</td>
<td>786</td>
<td>262</td>
<td>6</td>
<td>91.0</td>
<td>92.9</td>
</tr>
<tr>
<td>TAPBP</td>
<td>1,297</td>
<td>431</td>
<td>7</td>
<td>91.9</td>
<td>95.1</td>
</tr>
<tr>
<td>Class II IIB2</td>
<td>792</td>
<td>264</td>
<td>6</td>
<td>92.9</td>
<td>93.2</td>
</tr>
<tr>
<td>BRD2</td>
<td>2,343</td>
<td>781</td>
<td>11</td>
<td>91.5</td>
<td>97.6</td>
</tr>
<tr>
<td>DMA</td>
<td>792</td>
<td>264</td>
<td>4</td>
<td>90.0</td>
<td>93.1</td>
</tr>
<tr>
<td>DMB1</td>
<td>978</td>
<td>326</td>
<td>5</td>
<td>90.0</td>
<td>95.0</td>
</tr>
<tr>
<td>DMB2</td>
<td>777</td>
<td>259</td>
<td>6</td>
<td>90.6</td>
<td>95.6</td>
</tr>
<tr>
<td>TAP2</td>
<td>2,094</td>
<td>697</td>
<td>10</td>
<td>91.9</td>
<td>94.2</td>
</tr>
<tr>
<td>TAP1</td>
<td>1,856</td>
<td>586</td>
<td>11</td>
<td>92.7</td>
<td>95.1</td>
</tr>
<tr>
<td>Class I IA</td>
<td>1,080</td>
<td>360</td>
<td>8</td>
<td>85.5</td>
<td>92.7</td>
</tr>
<tr>
<td>C4</td>
<td>5,538</td>
<td>1,846</td>
<td>40</td>
<td>85.5</td>
<td>95.6</td>
</tr>
</tbody>
</table>

Table 2 Comparison of MHC-B genes (coding DNA sequence, CDS) between prairie chicken and domestic chicken (Gallus) and turkey.
2006; Koch et al. 2007) with universal primers. None of these individuals had more than two distinct sequences, which indicated only a single class IA locus in greater prairie chickens. Expression of a single IA locus was confirmed by PCR amplification of cDNA from spleen tissue (primers IAex2F and IAex3R).

Polymorphism and selection at the IA locus

Of 30 individuals genotyped at exon 3 of the IA locus, 13 were homozygotes, which, combined with the haplotypes from the four cloned individuals, provided sufficient power for PHASE to identify haplotypes with near 100% probability (Alcaide et al. 2011). A total of 14 unique sequences were found with an average nucleotide diversity of 0.042. Translation of the nucleotide sequences produced 14 unique amino acid sequences (Fig. 4). At both the exon level and at peptide binding codons only, \(d_{\text{S}}/d_{\text{G}}\) ratios were not significantly different than one, indicating strong selection was not acting on the IA locus in prairie chickens. Among bird species for which single IA locus data are available, the prairie chicken was the only species without a significant
level of selection on this region based on $d_S/d_S$ ratios (Table 3). In the CODEML a priori test for selection, a maximum likelihood ratio test showed a better fit for model M8, than the neutral model, M7 (P<0.001). The M8 model did identify positive selection at the IA locus; however, only three codons were so identified: 4, 22, and 61 (>95% posterior probability), all of which correspond to putative peptide binding codons identified in domestic chickens (Fig. 4).

Class IIB

Two class IIB loci were present in the prairie chicken core MHC-B, each flanking the TAPBP gene, an arrangement identical to the domestic chicken (Figs. 1 and 2). TAPBP was oriented in the same direction as in the domestic chicken (the gene is in reverse orientation in the turkey) (Fig. 2). Based on the similarity of gene order between the prairie chicken and the domestic chicken, the IIB1 gene is homologous to the BLB1 and IIB2 is homologous to BLB2. Sequencing of the PCR amplicons (exon 2 to 3: 167 bp; see Table 1) of cDNA from spleen samples from two individuals revealed two unique sequences per individual indicating that at least one locus was expressed. The IIB3 gene of the turkey is expressed while the IIB2 gene is not (Chaves et al. 2009), suggesting that the IIB3 of turkeys and BLB2 of domestic chicken are homologous. These results suggest that the prairie chicken IIB1 is homologous to the turkey IIB1, and the IIB2 gene of the prairie chicken is putatively homologous to the IIB3 gene of the turkey.

Gene similarity between species

Nucleotide and amino acid sequences for the fully sequenced genes were compared between the three species of Galliformes (Table 2). In all cases, the coding sequences between the prairie chicken and turkey were more similar (mean=94.2, range=87.7–97.6) than those between prairie chicken and domestic chicken (mean=90.4, range=85.5–92.7), consistent with the accepted phylogenetic relationships of the taxa (Pereira and Baker 2006). Amino acid residues were also more similar between the prairie chicken and turkey (mean=90.9, range=77.1–99.0) compared to the domestic chicken (mean=86.0, range=74.5–98.9). Comparative identity matrix plots also indicated that the prairie chicken was more similar to the turkey than the domestic chicken (Fig. 3).

Discussion

A highly compact core MHC

We assembled a 60.9-kb region of the core MHC-B of the greater prairie chicken and found 14 genes that were homologous to those in the domestic chicken and turkey. Genes of the prairie chicken MHC-B were more similar to the turkey than the domestic chicken at both the nucleotide and protein level, which reflects the closer phylogenetic relationship of the taxa (Pereira and Baker 2006). The MHC of the prairie chicken appears to be smaller than the domestic chicken and turkey, in both the number of coding DNA sequences and overall size. The core MHC-B has now been sequenced in five species of Galliformes, and a surprising amount of variation in gene presence, order, number, and orientation has been demonstrated. However, within prairie chickens, we found the smallest core MHC-B of any bird yet sequenced.

Several loci appeared to be absent from the core MHC-B of the greater prairie chicken. One of these, Blec2, is a putative NK cell receptor found in the MHC of other Galliformes and has been implicated in resistance to Marek’s disease in...
domestic chickens (Kelley and Trowsdale 2005; Rogers and Kaufman 2008). We found a Blec2-like gene in the gDNA of the individual from which the fosmid library was constructed, but not within the core MHC-B, indicating this Blec2-like gene is located elsewhere in the MHC or genome.

Unexpectedly, there was no BG gene between the BTN2 and Blec1 genes in the prairie chicken MHC. The domestic chicken and golden pheasant have at least one BG gene located between BTN2 and Blec1, the turkey and Japanese quail have three in that location (Fig. 2) (Shiina et al. 2004a, b; Bauer and Reed 2011; Ye et al. 2012). Our PCR results from gDNA indicate that the prairie chicken has at least one BG gene, but its location remains unknown. A possible explanation for the absence of the BG and Blec2 genes from the core MHC-B in the prairie chicken is a block inversion of the BTN2-BG-Blec2 gene complex which would agree with our PCR and sequencing results placing BTN2 ~2,000 bp from Blec1. However, this scenario would require an additional inversion of BTN2 since our PCR results require a reverse orientation of BTN2 that a single block inversion would have generated.

The results of this study show that even in closely related species of birds, the genetic architecture of the core MHC can vary substantially. Although gene rearrangements, deletions, and duplications are observed in the MHC of Galliformes, the general size of the core MHC is relatively stable within the order. Copy number variation is also observed, but it is limited to a few paralogs in both class I and II loci. This contrasts considerably with the picture that is emerging in the Order Passeriformes. The only extensive map of a passerine to date, that of the zebra finch (Taeniopygia guttata), suggests that the MHC in this order is dramatically expanded relative to Galliformes, with potentially dozens of duplications of class I and II genes. Furthermore, in contrast to Galliformes, the MHC in passerines may be scattered on several chromosomes (Balakrishnan et al. 2010, but see Ekblom et al. 2011). The extensive duplication of class I genes in passerines also occurs in mammals and may be caused by the disruption of the tightly linked TAP-class I gene complex that occurs in the domestic chicken (Kaufman 1999; Balakrishnan et al. 2010).

**Table 3** Polymorphism and tests for selection at exon 3 and the peptide binding codons (PBC) of the IA gene in four bird species

<table>
<thead>
<tr>
<th>Species</th>
<th>N</th>
<th>Alleles</th>
<th>π Ex 3</th>
<th>S Ex 3</th>
<th>π PBC</th>
<th>S PBC</th>
<th>dN/dS PBC</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tymanuchus cupido</td>
<td>30</td>
<td>14</td>
<td>0.044</td>
<td>27</td>
<td>0.121</td>
<td>10</td>
<td>1.02</td>
<td>0.410</td>
</tr>
<tr>
<td>Gallus gallus</td>
<td>7</td>
<td>9</td>
<td>0.050</td>
<td>30</td>
<td>0.229</td>
<td>17</td>
<td>3.55</td>
<td>0.006</td>
</tr>
<tr>
<td>Falco tinnunculus</td>
<td>25</td>
<td>23</td>
<td>0.041</td>
<td>33</td>
<td>0.146</td>
<td>13</td>
<td>4.35</td>
<td>0.023</td>
</tr>
<tr>
<td>Falco naumanni</td>
<td>36</td>
<td>18</td>
<td>0.031</td>
<td>33</td>
<td>0.100</td>
<td>11</td>
<td>5.88</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Additional information for Gallus is in Worley et al. (2008) and for Falco is in Alcaide et al. (2009)

N=individuals; π=nucleotide diversity; S=number of segregating sites; dN/dS=ratio of non-synonymous and synonymous substitutions; P=level of significance for Codon Z test of dN/dS ratio

**Minor selection on a single class I locus**

The most notable difference between the MHC-B of prairie chickens and other Galliformes was the absence of a second class I A locus in prairie chickens. The prairie chicken class I A aligns with the BF1 and BF2 genes in the domestic chicken and IA1 and IA2 genes in the turkey and golden pheasant. However, in the prairie chicken, there is no IA locus between DMB2 and the TAP genes and only a single IA gene is located between TAP1 and C4, which is the location of the class IA2 gene in the other Galliformes (Fig. 2). These differences in the prairie chicken may be associated with the inversion of the TAP gene block accompanied by a deletion of the second IA locus (Fig. 2). The TAP gene block is also inverted in the golden pheasant; however, unlike the prairie chicken, the pheasant has retained the IA2 gene (Ye et al. 2012). Both class I A genes are expressed in the turkey (Reed et al. 2011); however, in domestic chickens, only the BF2 locus is highly expressed, and it appears to bind to a wider spectrum of antigenic peptides than any single class I gene in mammals (Wallny et al. 2006). Studies of other bird species have revealed a similar pattern of expression of a single class I locus among duplicated loci (Moon et al. 2005).

Given the presence of a single IA locus, another interesting result of this study was the low level of detectable selection at the IA locus. Several studies have tested for selection on class I loci in birds. With the exception of the domestic chicken (Worley et al. 2008) and some species in the genus Falco (Alcaide et al. 2009, 2010; Gangoso et al. 2012), these relied on sequences from several paralogs (duplicated loci within the same species). However, dN/dS ratios derived from such mixed locus samples may not be accurate, because differences in synonymous substitutions between loci may be larger than within loci, resulting in an underestimate of the true dN/dS ratio of a specific gene (Hughes and Nei 1989). Furthermore, it is common for only a single class I locus to be expressed in vertebrate taxa, regardless of the number of paralogs and therefore only one locus is presumably under selection (Kaufman 1999; Flajnik and Kasahara 2001). Comparisons of dN/dS ratios between the minimally expressed chicken BF1 and the dominantly expressed BF2 loci support this idea.
It is not surprising, then, that most studies testing for balancing selection at the MHC class I in birds have produced \( d_S/d_S \) greater than one (when combining data from several loci), although none of them have shown selection at a level of significance. However, using data from species for which sequences were available for an individual class I locus, we found a strong, significant signal of selection at the putative PBR for the domestic chicken and two species of Falco (Table 3). Thus, the apparent absence of selection at the IA locus in the greater prairie chicken was unexpected and contrasts with our previous study of class II loci in prairie chickens (Eimes et al. 2011), where we found that the \( d_S/d_S \) ratio in the PBR (2.02) was significantly larger than one (codon-based \( Z \) test, \( Z=2.01; P=0.047 \)).

Among vertebrates, conclusive evidence for a species with a single, classical MHC class I locus has only been found in the amphibian genus Xenopus (Shum et al. 1993). Similar to our study, the \( d_S/d_S \) within the PBR of exon 3 (\( x^2 \)) in Xenopus was not significantly greater than one; however, \( d_N \) was higher in Xenopus (36±7) than in prairie chickens (26±1.0) (Flajnik et al. 1999; Bos and Waldman 2006). Flajnik et al. (1999) argued that selection was acting on the class I locus in Xenopus despite the low \( d_N/d_S \) because the \( d_S \) across the entire gene had reached saturation due to the age of the allelic lineages and the higher mutation rate of frogs relative to most other vertebrates. These factors do not explain our results because MHC class I PBR allelic lineages in birds have not been shown to persist for very long periods, (Hess and Edwards 2002), mutation rates in birds are not unusually high and the \( d_S \) in exon 3 in our sample was relatively low (9.13±1.1).

An alternative explanation for the lack of polymorphism and detectable balancing selection at the IA locus is that strong directional selection has acted on this locus and reduced variation at the PBR. Recently, Gangoso et al. (2012) reported relatively low levels of class I polymorphism among several species of Falco and argued that strong directional selection could account for the lower levels of class I variation found in some species relative to others within the same genus. They argued that strong selection from particular parasites might reduce variation. Nonetheless, Gangoso et al. (2012) found significant evidence of diversifying selection at the putative PBR of a single class I locus (as defined by Alcaide et al. 2009, 2010), in contrast to our study. One interesting finding in our study is that all of the 14 nucleotide sequences obtained from 30 individuals coded for unique amino acid sequences, which may suggest selection for a wide variety of amino acid sequences. This scenario would seem to contradict the hypothesis that class I motifs in birds are typically less variable than mammals because they bind a wider spectrum of peptides (Wallny et al. 2006; Koch et al. 2007) and suggests that, at least in greater prairie chickens, a wider diversity of different sequences increases immunocompetence.

Another possible explanation for our results is that drift in this fragmented population has removed most evidence of selection at this locus. The Wisconsin study population experienced a bottleneck, resulting in a decline from approximately 50,000 to 1,200 individuals over the last 75 years (Johnson et al. 2003). We previously demonstrated a significant loss of MHC class II variation in this population (Eimes et al. 2011). Tests for positive selection at class II loci were significant, but only in historic samples (1950; class I was not studied). Earlier studies of this population have found a similar loss of genetic variation at neutral markers (Bellinger et al. 2003; Johnson et al. 2003); however, the loss of MHC class II variation exceeded that of neutral genetic variation, possibly because of drift acting on copy number variation (Eimes et al. 2011). Low genetic variation at both MHC and neutral genes combined with a lack of detectable selection at either class I or II (possibly due to drift rather than purifying selection) indicates that the “genetic health” of this population may have been reduced due to the population bottleneck. Studies of larger populations of prairie chickens that have not experienced as severe a bottleneck are needed to assess whether drift has erased traces of diversifying selection on the IA locus in this population or if strong directional selection may have acted on this locus throughout the range of the species.

Acknowledgments Funding was provided by grants from the Research Growth Initiative, University of Wisconsin-Milwaukee Graduate School, and National Science Foundation (DEB-0948695) to POD, JLB and LAW, American Ornithologists’ Union, The American Museum of Natural History, and Ruth Walker Research Award from the University of Wisconsin-Milwaukee to JAE, and USDA-CSREES NRI Grants (#2005-1326 and 2009-35205-05302) to KMR. We thank J. Toepfer and R. Bellinger for tissue collection and processing, C. Wimpee for advice and assistance in the laboratory, and R. Settlage, University of Wisconsin-Milwaukee to JAE, and USDA-CSREES NRI Grants (#2005-1326 and 2009-35205-05302) to KMR. We thank J. Toepfer and R. Bellinger for tissue collection and processing, C. Wimpee for advice and assistance in the laboratory, and R. Settlage, for assistance with bioinformatics.

References


Bauer MM, Reed KM (2011) Extended sequence of the turkey MHC B-locus and sequence variation in the highly polymorphic B-G loci. Immunogenetics 63:209–221


Gangoso L, Alcaide M, Grande JM, Munoz J, Talbot SL, Dufourt JCM, Burt DW (2010) The major histocompatibility complex (Mhc) class IIB region has greater genomic structural flexibility and diversity in the quail (Coturnix coturnix japonica) than the chicken. BMC Genomics 7:322


Rogers SL, Kaufman J (2008) High allelic polymorphism, moderate sequence diversity and diversifying selection for B-NK but not B-Lc, the pair of lectin-like receptor genes in the chicken MHC. Immunogenetics 60:461–475


numbers of B and Y genes with variable diversity and expression. Immunogenetics 59:725–734


Trowsdale J (1988) Molecular genetics of the MHC. Immunology 41:1–17


