

## ORIGINAL ARTICLE

A locus-wide approach to assessing variation in the avian MHC: the *B*-locus of the wild turkeyLD Chaves<sup>1,3</sup> GM Faile<sup>1</sup>, JA Hendrickson<sup>1</sup>, KE Mock<sup>2</sup> and KM Reed<sup>1</sup><sup>1</sup>Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, St Paul, MN, USA and <sup>2</sup>Wildland Resources Department, Utah State University, Logan, UT, USA

Studies of major histocompatibility complex (MHC) diversity in non-model vertebrates typically focus on structure and sequence variation in the antigen-presenting loci: the highly variable and polymorphic class I and class IIB genes. Although these studies provide estimates of the number of genes and alleles/locus, they often overlook variation in functionally related and co-inherited genes important in the immune response. This study utilizes the sequence of the MHC *B*-locus derived from a commercial turkey to investigate MHC variation in wild birds. Sequences were obtained for nine interspersed MHC amplicons (non-class I/II) from each of 40 birds representing 3 subspecies of wild turkey (*Meleagris gallopavo*). Analysis of aligned sequences identified 238 single-nucleotide variants approximately one-third of which had minor allele frequencies >0.2 in the sampled birds. PHASE analysis identified 70 prospective

MHC haplotypes in the wild turkeys, whereas a combined analysis with commercial birds identified almost 100 haplotypes in the species. Denaturing gradient gel electrophoresis (DGGE) of the class IIB loci was used to test the efficacy of single-nucleotide polymorphism (SNP) haplotyping to capture locus-wide variation. Diversity in SNP haplotypes and haplotype sharing among individuals was directly reflected in the DGGE patterns. Utilization of a reference haplotype to sequence interspersed regions of the MHC has significant advantages over other methods of surveying diversity while identifying high-frequency SNPs for genotyping. SNP haplotyping provides a means to identify both divergent haplotypes and homozygous individuals for assessment of immunological variation in wild and domestic populations. *Heredity* (2011) **107**, 40–49; doi:10.1038/hdy.2010.153; published online 22 December 2010

**Keywords:** MHC; B-locus; SNP; haplotype; DGGE; turkey

## Introduction

The major histocompatibility complex (MHC) is the most polymorphic and gene-dense region in the vertebrate genome (Kelley *et al.*, 2005). The locus contains numerous genes involved in immune responses including genes for cytokines, complement and antigen-processing proteins, and the highly polymorphic glycoproteins responsible for presenting endogenous and exogenous antigens (MHC class I and class II, respectively) to T cells. In addition to class I and class II genes, clusters of other related genes can be found within the MHC, such as tripartite motif loci (*TRIM*), the butyrophilin loci (*BTN*) and in birds, the B-G loci (Kaufman *et al.*, 1999a).

Variation in MHC genes is vital for generation of a diverse array of antigenic peptides for T-cell presentation. The rapidly evolving pathogen repertoire requires multiple MHC loci with highly polymorphic molecules necessary for presentation of pathogen-derived peptides to effector T cells. For example, single loci in the human MHC have 100s of alleles and allelic combinations may

generate 10 000s of potential haplotypes (Marsh *et al.*, 2005). In contrast, domestic poultry show more constrained diversity in their MHC (Hosomichi *et al.*, 2008; Chaves *et al.*, 2010) although; a recent study of domestic chickens identified nearly 100 *MHC-B* haplotypes (Fulton *et al.*, 2006). MHC haplotypes have been found to have strong influences on immune response and disease resistance in chickens. For example, studies have shown that MHC haplotypes influence immune response to Rous sarcoma virus (Aeed *et al.*, 1993; White *et al.*, 1994; Taylor 2004).

In the context of the MHC, the best-studied bird species are those of agricultural significance; MHC regions have been examined in chicken, quail, duck and turkey (Kaufman *et al.*, 1999b; Moon *et al.*, 2005; Shiina *et al.*, 2004, 2007; Chaves *et al.*, 2009b, 2011). Sequencing the MHC locus of the chicken, quail and turkey (*MHC-B*) identified the classical MHC class I and II genes within an ~50 kb distance (Shiina *et al.*, 2004, 2007; Chaves *et al.*, 2009b). Commercial selection has clearly reduced allelic diversity in chicken lines (Muir *et al.*, 2008). A sampling of genetic variation within commercial turkeys shows that some MHC haplotypes occur at high frequency within breeder lines (Chaves *et al.*, 2010).

Population level studies of MHC variation in wild birds have focused mainly on two technologies to survey variation at class I and class II loci. Southern hybridization methods have been used to estimate the number of genes and alleles/locus for the class I and II loci thereby providing a rough estimate of overall genetic variation at

Correspondence: Dr KM Reed, Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, 295 ASVM, 1988 Fitch Avenue, St Paul, MN 55108, USA.  
E-mail: reedx054@umn.edu

<sup>3</sup>Current Address: National Jewish Health, Division of Allergy and Clinical Immunology, Department of Medicine, 1400 Jackson Street, Denver, CO 80206, USA.

Received 1 July 2010; revised 11 October 2010; accepted 4 November 2010; published online 22 December 2010

the MHC (Emara *et al.*, 1992; Westerdahl *et al.*, 2000; Freeman-Gallant *et al.*, 2002; Richardson and Westerdahl, 2003; Alcaide *et al.*, 2008). PCR-based techniques such as reference strand-mediated conformational analysis, single strand conformation polymorphism and denaturing gradient gel electrophoresis (DGGE) are also effective in estimating the number of alleles, especially when locus-specific primers are available (Ramon *et al.*, 1998; Goto *et al.*, 2002; Westerdahl *et al.*, 2004; Knapp, 2005; Alcaide *et al.*, 2010). Studies have generally identified 2–4 class IIB loci in birds and provide evidence for increased MHC diversity in outbreeding populations. These methods, however, can lead to overestimation in the number of loci due mainly to problems of distinguishing loci and alleles (Jacob *et al.*, 2000; Miller *et al.*, 2004). Recently, Babik *et al.* (2009) tested the use of high-throughput DNA sequencing technology (Roche 454 sequencing) as an alternative for MHC genotyping. This approach used degenerate primers based on previously sequenced loci/alleles to simultaneously amplify multiple loci by PCR. Primer tags allowed for segregation of sequencing reads to allow for digital genotyping of multiple individuals in a single sequencing pool. Sequencing and PCR errors remain problematic, but bioinformatic methods can be used to help identify low-frequency artifact alleles.

Amplifying, cloning and sequencing the highly variable MHC domains of class I and class II loci is labor intensive, but can provide the best estimate of the number of loci present in the genome. Sequence surveys typically focus on the peptide-binding region (PBR) encoded in exons 2 and 3 of class I and exon 2 of the class IIB genes. The PBR is the region that determines which antigens will be presented to T lymphocytes and sequence differences are indicative of functionally different alleles. When combined with mRNA sequencing, expressed alleles can be matched to their genomic counterpart, improving allele designations. Studies that characterize only the PBR, however, risk concluding that variation observed at the DNA level directly translates into significant differences in disease resistance/susceptibility.

Studies that focus only on the class I and class II loci also overlook the potential importance of variation at other genetically linked MHC genes. As comparative genome sequences become increasingly available, it is possible to expand MHC surveys beyond the multilocus class I and class II genes to include tightly linked adjacent loci, potentially avoiding the difficulties associated with genotyping these loci. The goals of this project were twofold: first to demonstrate the ability of locus-wide single-nucleotide polymorphism (SNP) genotyping to identify MHC haplotypes and second, to quantify variation within the *MHC-B* locus of wild turkeys as compared with commercial birds. The turkey (*Meleagris gallopavo*) with six recognized subspecies, is the only domesticated agricultural species native to

North America (Dickson, 1992). Commercial breeds are believed to be derived primarily from the south Mexican subspecies (*M. g. gallopavo*) via reintroduction to the New World from Europe (Reed, 2009; Speller *et al.*, 2010). Results of this study show the benefit of an assembled MHC haplotype in investigating variation and confirm that MHC diversity in wild birds is significantly greater than in commercial lines.

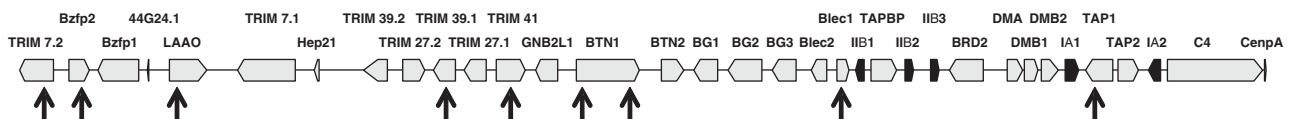
## Materials and methods

### Genomic sequencing

A panel of 40 individuals from three subspecies (Eastern, Merriam's and Rio Grande) of turkey was examined in this study (Supplementary Table S1). The eastern subspecies (*M. g. silvestris*), has the largest geographical distribution occurring from the east coast through the Great Plains; Merriam's (*M. g. merriami*), is located in parts of Colorado, New Mexico and Arizona; and the Rio Grande subspecies (*M. g. intermedia*), is present in most of Texas and parts of northern Mexico. DNA samples were from 11 different locations, representing a subset of samples examined by Mock *et al.* (2002), Latch *et al.* (2002) and a hunter harvested bird from Winona, MN, USA were used in this study. Samples from the populations studied by Mock *et al.* (2002) were relict, native populations, with presumably higher sequence diversity. Individuals were sequenced at nine interspersed locations across the *MHC-B* region (Figure 1, Supplementary Table S2) as previously described (Chaves *et al.*, 2010). PCR products were purified using a MinElute PCR Purification Kit (Qiagen, Inc., Valencia, CA, USA) and sequenced with an automated ABI Sequencer (Applied Biosystems, Carlsbad, CA, USA). Sequence data were manually analyzed using Sequencher software (Gene Codes, Corp., Ann Arbor, MI, USA) and single-nucleotide variants (SNVs, SNPs and deletion/insertion polymorphisms (DIPs), respectively) were recorded.

### Haplotype identification and phylogenetic analysis

Polymorphisms were analyzed using Arlequin, PHASE and Haploview software (Schneider *et al.*, 2000; Stephens and Donnelly, 2003; Barrett *et al.*, 2005) to quantify haplotype frequency and diversity within the 40 sequenced birds. Analysis was performed using 45 SNVs (42 SNPs and 3 DIPs) with minor allele frequencies (MAFs) of 0.2 or greater that were successfully genotyped on all 40 individuals. An additional haplotype inference analysis included samples from a previous study of commercial turkeys (Chaves *et al.*, 2010) using SNVs of high MAF common among the two groups. Analysis of haplotype relationships was performed using ClustalW (Larkin *et al.*, 2007) and SplitsTree4 (Huson and Bryant, 2006).



**Figure 1** The turkey *MHC-B* locus and position of amplicons (arrows) sequenced for haplotype analysis and genetic mapping. Positions of primers within the reference sequence are given in Supplementary Table S2.

### Class IIB DGGE and sequencing

To compare variation in one of the *MHC-B* multigene families among individuals with different SNP haplotypes, the class IIB genes of the *B*-locus were examined by DGGE. For DGGE, primers (cIIB\_DGGE\_F—5'-AAAA CCCGAGCCCTTGT-3' and cIIB\_DGGE\_R—5'-ACTCA CCGCTCCTCTGCAC-3') were designed from aligned sequences (Chaves *et al.*, 2010) to simultaneously amplify the highly variable PBR (exon 2) from the three class IIB genes of the *B*-locus. The designed primers are specific for the *MHC-B* loci and do not amplify the other known class IIB-like genes that reside outside of the *B*-locus. Polymerase chain reactions for DGGE analysis were optimized with genomic DNA template using *Taq* Mastermix (Promega, Corp., Madison, WI, USA) supplemented with 1 × Q solution (Qiagen). Amplifications were performed for 35 cycles with 58 °C annealing temperature and 30 s extension times. Control birds and clone constructs with known class IIB haplotypes and were used to help quantify the differences observed in the wild turkeys.

DGGE was performed using the Dcode Universal Mutation Detection System (Bio-Rad, Hercules, CA, USA). Optimal conditions for examining the class IIB genes (25–65% urea/formamide gradient in 6% acrylamide, 1 × Tris/acetic acid/EDTA buffer, at 130 V and 60 °C for 4 h) were determined using perpendicular DGGE and a time-series analysis. These conditions were used in parallel denaturing gels to compare class IIB amplicons between individuals. PCR products were denatured at 95 °C for 5 min, incubated at 65 °C for 1 h and then allowed to slowly cool to room temperature (RT) before addition of loading dye. Gels were visualized by staining with ethidium bromide.

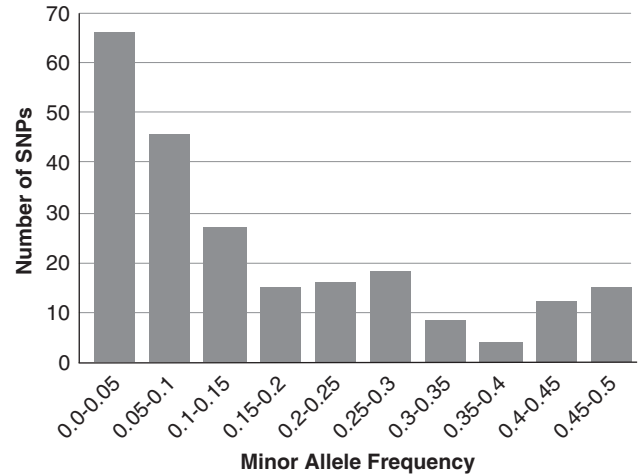
On the basis of the DGGE results, four birds were selected for sequencing of the class IIB DGGE PCR products to verify the amplification of multiple loci. For consistency, the same PCR products used for DGGE were cloned using a pDrive Cloning Kit (Qiagen), transformed into DH5 $\alpha$  cells (Invitrogen, Carlsbad, CA, USA), and over 20 purified plasmids were sequenced per individual. In addition, locus-specific primers (Chaves *et al.*, 2009b) were used to individually sequence portions of the three known *MHC-B* class IIB genes from the single wild bird collected from Minnesota. Sequences of exon 2 were translated to identify putative PBR alleles.

## Results

### MHC-B polymorphisms

Over 9 kb of the *MHC-B* region was sequenced on 40 wild turkeys from across North America. A total of 238 SNVs (Appendix) were identified with MAF ranging between 0.01 and 0.5 (average 0.15) (Figure 2). In all, 37% of the loci had a MAF  $\geq$  0.2. The frequency of SNVs in this region (all polymorphisms), 1/40 bp, is higher than the 1/200 bp found in other regions of the turkey genome (Reed *et al.*, 2006) and the 1/70 bp identified within the *MHC-B* in commercial breeder lines (Chaves *et al.*, 2010). Most SNPs were transition substitutions and located in introns (Table 1). Of the 45 SNPs located in coding sequence, 18 represent non-synonymous substitutions.

In addition to SNPs, nine DIPs (deletion–insertion) were present in the studied amplicons. Included in these is a 57 bp biallelic DIP in the first intron of *Bzfp2* not



**Figure 2** MAFs of SNPs and DIPs identified in the 40 wild turkeys sequenced across the *MHC-B* locus.

**Table 1** Characterization of SNPs identified within wild turkey subspecies

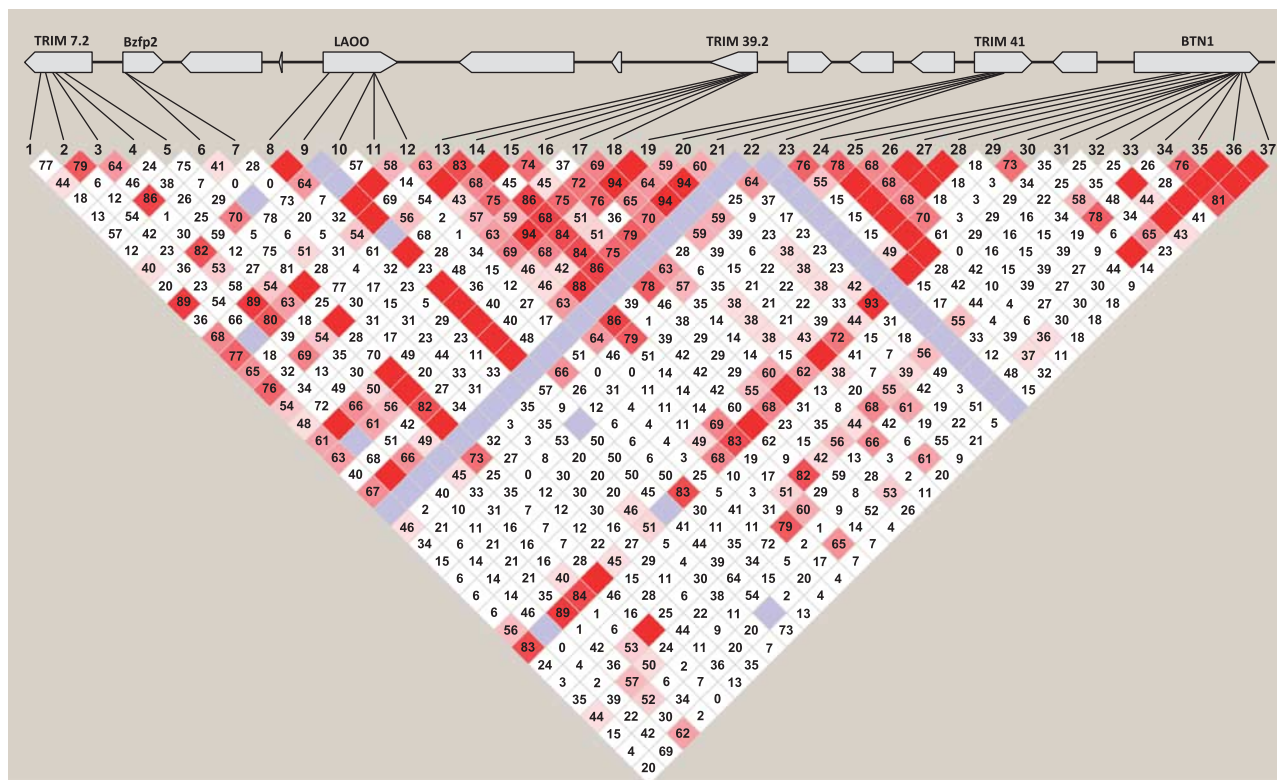
	Observed	Frequency	Newly identified	Frequency
Non-coding	38	0.165	12	0.112
Intron	154	0.667	76	0.71
Exon	45	0.195	22	0.206
Syn	27	0.117	11	0.103
Non-syn	18	0.078	11	0.103
Transistion	173	0.749	88	0.822
Transversion	54	0.234	19	0.178
Triallelic	2	0.009	2	0.019
Indel	8	NA	3	NA
MAF $\geq$ 0.2	84	0.367	5	0.047
Total SNP	229		107	

Abbreviations: MAF, minor allele frequency; SNP, single-nucleotide polymorphism; Syn, synonymous.

previously identified in commercial populations and a large polymorphic repeat element present within intron 6 of *LAO*. Examination of this locus by electrophoresis of the PCR amplicons on 3% agarose and 5% denaturing acrylamide gels found five alleles at this locus (one more than identified in commercial breeder lines, Chaves *et al.*, 2010) ranging in size from 180 to ~350 bp.

Of the 238 SNVs, 112 (109 SNPs and 3 DIPs) were not previously identified in commercial turkeys (Chaves *et al.*, 2010). The novel SNPs had MAF ranging from 0.01 to 0.49 with an average of 0.07. Sixteen SNPs previously found in commercial birds were not present in the wild turkey data set. However, some of these may have been missed through technical failures. Two tri-allelic SNPs were identified at positions 56210 (H) and 93867 (H). Allele frequencies at these loci ranged from 0.038 to 0.875 and averaged 0.25. SNPs at these loci were excluded from further analysis. The *TRIM39.1* amplicon showed poor amplification and sequencing success. This is perhaps due to null alleles caused by nucleotide variations within the primer-binding site(s) within wild birds. This locus was not included in haplotype inference. Finally, none of the polymorphisms in the *TAP1* amplicon had a MAF  $\geq$  0.1.





**Figure 3** The limited extent of LD across the *MHC-B* region identified in wild turkeys.  $D'$  values ( $\times 100$ ) given for all pairwise comparisons  $< 1.0$ . Values not given for blocks in complete LD ( $D' = 1.0$ ). Block colors correspond to  $D'/\text{LOD}$ , where  $D' < 1$  and  $\text{LOD} < 2$  is white,  $D' < 1$  and  $\text{LOD} \geq 2$  is shades of pink/red,  $D' = 1$  and  $\text{LOD} < 2$  are grey, and  $D' = 1$  and  $\text{LOD} \geq 2$  are bright red.

### Haplotype reconstruction

Haplotype inference was conducted using 45 polymorphic loci that displayed varying allele representation among the subspecies (Figure 3) and that had  $\text{MAF} \geq 0.2$  (Appendix). Initial analysis combining all individuals identified four homozygotes (MCS6, MCS7, ECL10 and MINN). Of the 40 individuals examined, 70 potential haplotypes were inferred by PHASE analysis with the majority of individuals possessing at least one unique haplotype.

A second haplotype analysis of 37 markers with  $\text{MAF} \geq 0.2$  incorporated genotypes from a panel of 52 commercial turkeys (Chaves *et al.*, 2010). This analysis identified a total of 99 haplotypes, of which the wild turkeys possessed 66 (Supplementary Table S3). Posterior probabilities for haplotype allele assignment at a given SNV position ranged from 0.5 (no support) to 1.0 (complete linkage disequilibrium (LD)) within the sampled individuals. The two rounds of PHASE inference had similar posterior probabilities when averaged per locus. Wild turkey haplotypes differed from each other at 1–26 SNPs with an average of 15 differences (s.d. = 5.5). Twelve cases of haplotype sharing were observed among the wild birds (Supplementary Table S3). Only a single haplotype was shared between the commercial and wild birds; EOM6 had a haplotype in common with Nici (NTWF06) the only fully sequenced turkey haplotype (Chaves *et al.*, 2009a).

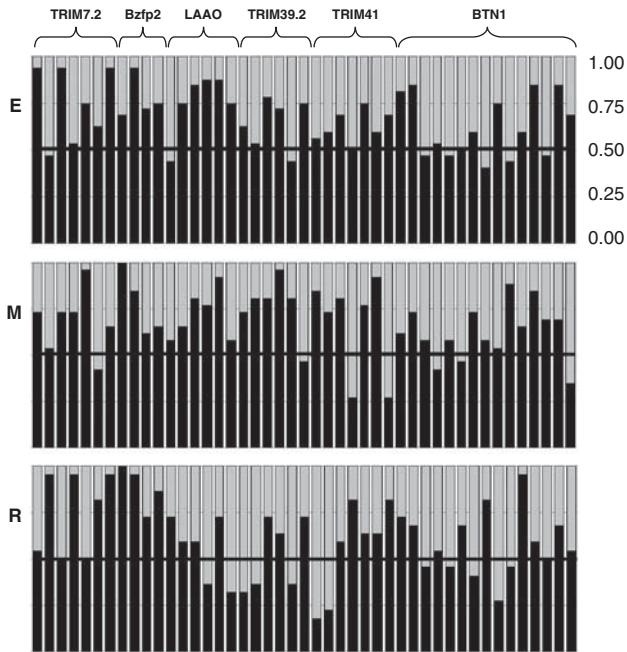
### LD and subspecies variation

Patterns of LD among 37 SNPs are shown in Figure 3. LD within the region is limited with the highest values

centered in *TRIM39.2*. The low level of LD is indicative of recombination within the *MHC-B*. Analysis of sequence variation in commercial turkeys suggests that recombination in the *B*-locus occurs at a higher frequency (18 cM per Mbp, Chaves *et al.*, 2010) than the whole genome average (2.5 cM per Mbp, Reed *et al.*, 2007). Within the UMN/NTBF mapping families, the genetic distance between *TRIM7.2* and *BTN1* (physical distance  $< 100$  kb) is just over 1.1 cM.

Rare alleles within individuals were more frequent in the EWW group and the Eastern subspecies, respectively (Figure 4, Table 2). As expected given the physical and genetic distance between populations, loci in Hardy–Weinberg equilibrium occurred more frequently in the samples from distinct locations rather than pooling individuals within a specific subspecies (Table 2). Pairwise  $F_{ST}$  analysis found significant variation between the three subspecies. The Eastern vs Merriam's subspecies comparison had the lowest  $F_{ST}$  value (0.08) whereas the value for Eastern vs Rio Grande was 0.12. The greatest pairwise  $F_{ST}$  value occurred between Rio Grande vs Merriam's (0.18). Genetic distances (corrected average pairwise differences ( $p_i$ )) between subspecies were similar, with values for Eastern vs Merriam's, Eastern vs Rio Grande, and Merriam's vs Rio Grande of 1.51, 2.52 and 3.75, respectively.

Relationships of haplotypes derived from 37 shared SNVs were examined to compare the origin and variation of wild turkey *MHC* haplotypes within subspecies and to compare wild turkeys to commercial breeder lines (Figure 5). The unrooted neighbor joining (NJ) tree shows



**Figure 4** Frequency of major alleles at 45 SNP loci used in haplotype inference within three subspecies of wild turkey (E-Eastern, M-Merriam's and R-Rio Grande). Within each subspecies, the frequency of the major allele as determined in the total data set is indicated for each SNP by the black bar. Genes containing the SNPs are indicated above.

**Table 2** Locus comparison for sample locations with four or more individuals

Group	N	Exclusive alleles	H-W loci	Homozygous loci	Haplotypes identified	Homozygous individuals
ECL (Eastern)	6	2	36	8	11	1
EWV (Eastern)	4	8	38	6	8	0
MCS (Merriam's)	5	2	31	11	6	2
MSL (Merriam's)	5	4	39	6	9	0
REN (Rio Grande)	4	1	35	4	8	0
RKW (Rio Grande)	6	5	37	6	12	0
Eastern	16	32	25	0	29	2
Merriam's	13	9	19	1	19	2
Rio Grande	11	16	27	1	22	0
Total	40	57		0	70	4

Abbreviations: H-W, Hardy-Weinberg; SNP, single-nucleotide polymorphism.

For all polymorphisms identified, alleles exclusive to populations are noted. For each sample location and subspecies, the SNP loci present in H-W equilibrium ( $P \geq 0.05$ ), SNP loci found to be homozygous, and the number of SNP haplotypes calculated for the 45 SNPs and DIPs used for initial haplotype inference are indicated. Distribution of homozygous individuals by population and subspecies are shown.

an interleaving of commercial and wild haplotypes, and a separate cluster of partially divergent wild haplotypes (Figure 5a). Haplotypes of wild turkeys did not assort by subspecies. Most of the commercial haplotypes cluster

near each other but support for major branches, as evidenced by bootstrap values, is weak. The haplotype of Nici (NTWF06) is found in a distant node and is the only commercial haplotype completely shared with the wild birds surveyed. A phylogenetic network (Figure 5b) was constructed to better model haplotype relationships which are assumed to be influenced by reticulate events. The haplotype network retains the overall NJ tree topology, but suggests historical recombination with the *B*-locus. This is especially evident for haplotypes such as RKW6b that are positioned between major clusters.

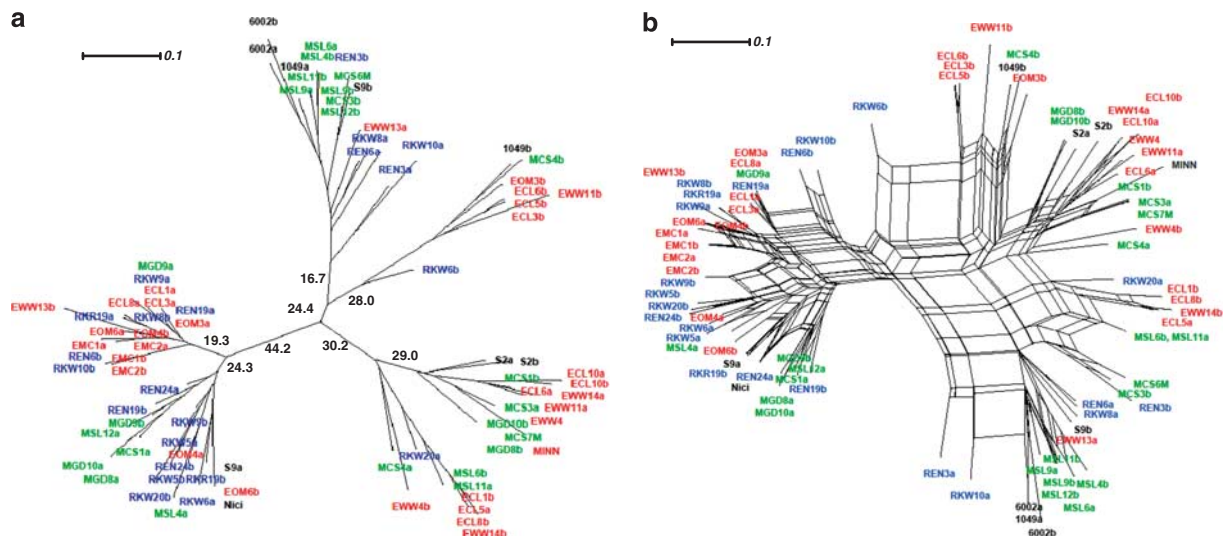
#### Class IIB variation

DGGE was used as a means to visualize variation at the class IIB genes to determine the effectiveness of locus-wide SNP genotyping. Observed DGGE patterns for control birds and 37 of the wild turkeys are presented in Figure 6. Included for comparison were birds homozygous at all three class IIB loci (6002, S-02, Nici and MN Wild; two or three total alleles), a triple heterozygote (1049, five total alleles) and an individual heterozygous at one locus, but with unique alleles at the other two loci (S-09, four total alleles). The artificial single locus heterozygote (B1\*1/1\*3) was created using cloned DNAs with a total of 42 bp differences between the cloned alleles.

As expected, individuals homozygous at the class *IIB1*, class *IIB2* and class *IIB3* loci tend to display fewer bands and more simplified DGGE banding patterns. For example, contrast in Figure 6 S-02 (two alleles; allele B1\*2 and B3\*2 are identical), Nici (three alleles), S-09 (four alleles) and 1049 (five alleles; amplified portions of alleles B1\*2 and B3\*2 are identical). In this multi-locus analysis, DGGE pattern is determined not only by the number of alleles, but also sequence similarities among alleles. For example, the average number of sequence differences between alleles was lower in the homozygous commercial bird 6002 (45.3) than in MN Wild (52.2) (Figure 6 inset). Both of these birds have three known alleles but very different DGGE patterns. In contrast, the DGGE patterns for birds 6002 and 1049 are complex, but very similar in appearance. Individual 1049 has five known class IIB alleles that differ by an average of 47.9 nucleotides.

Unique banding patterns were observed for most of the wild turkeys, although shared bands were observed within sample location. For example, banding patterns are highly similar for ECL5 and 6, MGD8 and 10; EWW11 and 14; and MSL 4, 9 and 12. This is consistent with the SNP data (Supplementary Table S3) in that ELC5 and 6 share a common haplotype (ECL5b and 6b) and MGD8 and 10 have two shared haplotypes (MGD8a/10a and 8b/10b). Haplotype similarities are also reflected by their positions in the cladograms in Figure 5. Thus, haplotype sharing and sequence diversity between SNP haplotypes is directly reflected in the DGGE patterns.

To verify amplification of multiple loci and determine underlying sequence differences, the same PCR products used for DGGE were cloned and sequenced for four individuals. Birds EOM4 and EOM6 were selected because they each displayed several well-defined DGGE bands. MSL9 and MSL12 were selected because they both displayed a DGGE pattern characterized by a single bright band with many indistinct minor bands and



**Figure 5** Relationships of inferred haplotypes from wild and selected commercial turkeys. (a) Neighbor-joining tree constructed with Uncorrected P distances. Bootstrap values (1000 replicates) are given for the basal branches. (b) Neighbor-net network constructed using ordinary least squares variance. Haplotypes originating in Eastern, Merriam's and Rio Grande subspecies are indicated in red, green and blue, respectively. Selected commercial haplotypes are indicated in black.

shared SNP haplotypes. Of the clones examined, 18 unique DNA sequences were obtained for EOM4 (22 total clones), 14 for EOM6 (21 clones), 8 for MSL9 (23 clones) and 7 for MSL12 (21 clones). In each case, the number of unique sequences exceeds the number of expected class IIB alleles (assuming three loci). As the clones were obtained from PCR amplicons generated with non-proof reading *Taq* polymerase, many of the unique sequences may be artifact in that they differ at as few as a single nucleotide. Also possible is the production of sequence hybrids during the simultaneous amplification of multiple loci. Finally, we cannot discount the possibility of additional class IIB loci in wild turkeys versus their commercial counterparts.

Relationships among the class IIB sequences for each individual are shown in Supplementary Figure S1. Sequences of EOM4 and EOM6 each cluster into three major groups indicative of three loci, whereas MSL 9 and MSL12 show less sequence complexity. This result is consistent with the number of major bands observed in DGGE (Figure 6). Translation and alignment of the sequence region corresponding to exon 2 reduces the number of different sequences (alleles) per individual; and resulted in similar tree topologies (Supplementary Figure S2). A total of 27 unique exon 2 amino acid sequences were observed in the four birds. These sequences group into three main clusters (Supplementary Figure S3), which do not assort by locus as defined by alleles previously observed in commercial birds (Chaves et al., 2010). Two PBR sequences observed in the wild birds (W24 and 2\*5) were identical to PBR alleles seen in commercial turkeys. Aligned amino acid sequences (Supplementary Figure S4) show that the variant peptides correspond well with the position of putative peptide-binding sites defined for humans by Brown et al. (1993).

## Discussion

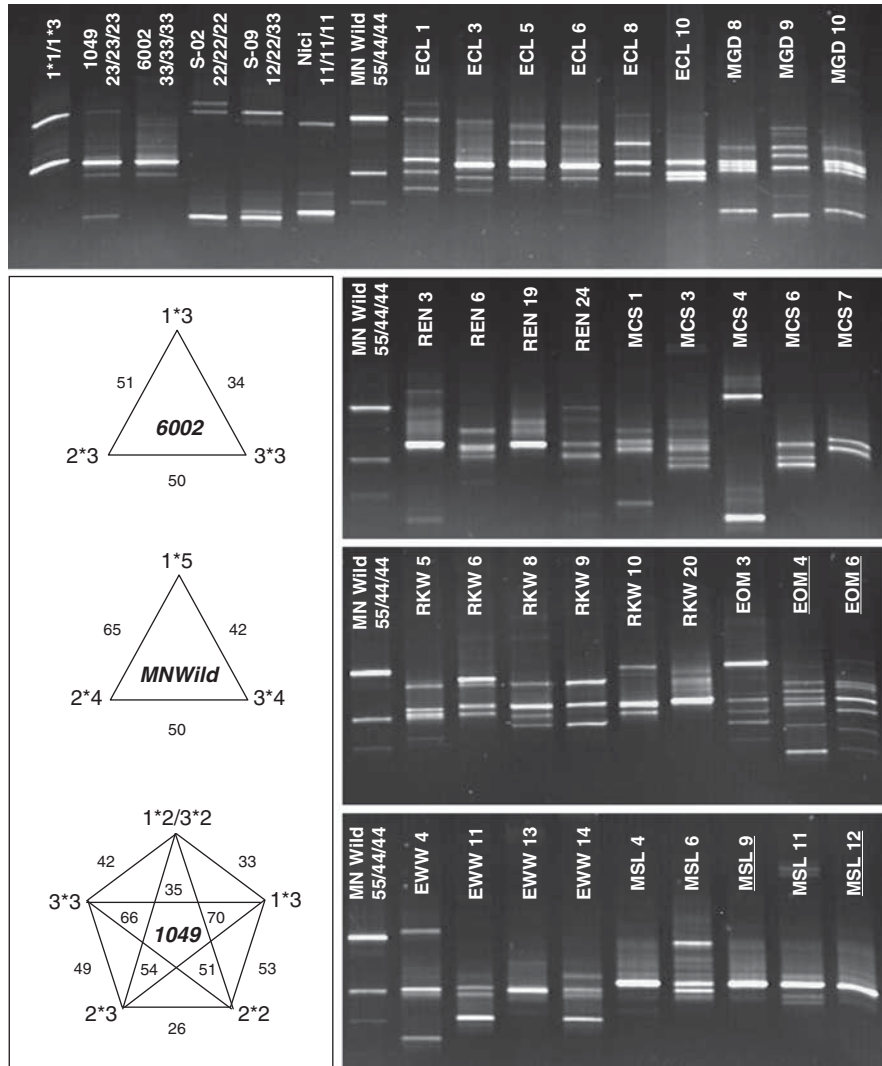
MHC diversity (allelic richness) is particularly advantageous in the face of widely varied pathogens (Wegner et al., 2004).

Studies of non-model vertebrate species have focused on polymorphism in the class I and class II MHC genes and pathogen-mediated selection has been considered a primary mechanism in maintaining MHC diversity (Jeffery and Bangham, 2000). Early population genetic studies of MHC variation in turkey relied on Southern hybridizations to elucidate gene content and variation (Zhu et al., 1995, 1996; Jones et al., 2003). Although this approach can be an easy and powerful tool to examine genomic variation within the species, Southern hybridizations (restriction fragment length polymorphism (RFLP)) rely on relatively few variable nucleotides and the technique ultimately fails to successfully capture the full extent of genetic diversity at the MHC. Similarly, sequencing of the exon(s) encoding class I and/or class IIB PBRs may not accurately identify the diversity within the species when PBRs are highly polymorphic and multiple loci occur within the genome. The goals of this project were twofold. First, to demonstrate the ability of locus-wide SNP genotyping to identify MHC haplotypes and second, to examine variation within the *MHC-B* locus of wild turkeys in contrast to that observed in commercial birds.

## MHC genotyping

Detailed genomic descriptions of avian MHC loci are limited to a few species and often focus on variation in one or two of the classical genes. MHC variation can be found at both extremes in wild bird populations, with some species having seemingly unlimited alleles and others presenting almost no polymorphism. For example, a study of class IIB gene diversity in Red-winged Blackbirds (*Agelaius phoeniceus*) found extensive variation within and between individuals (Edwards et al., 1998) while a sequence survey of the PBR of a class IIB locus from the endangered Galápagos penguin (*Spheniscus mendiculus*) found only three highly similar alleles (Bollmer et al., 2007). At the upper extreme are species like the bluethroat (*Luscinia svecica*), in which a minimum of 61 unique alleles of at least 11 functional loci were identified in 20 individuals (Anmarkrud et al., 2010). It is





**Figure 6** DGGE analysis of class IIB loci in wild turkeys. Lane 1 contains the PCR products from two clones representing alternate alleles ( $B1^*1:B1^*3$ ) at a single locus (*class IIB1*). Included for comparison (lanes 2–7) are five commercial turkeys and the wild turkey from MN all with sequenced class IIB haplotypes. Numbers of birds selected for sequencing of the DGGE products are underlined. Inset: the number of sequence differences among alleles at the class IIB loci for three of the control birds. Individual 6002 is homozygous for  $1^*3/2^*3/3^*3$  and MN Wild is homozygous for  $1^*5/2^*4/3^*4$ . Individual 1049 is a triple heterozygote ( $1^*2:1^*3/2^*2:2^*3/3^*2:3^*3$ ) with alleles  $B1^*2$  and  $B3^*2$  being identical.

important to recognize that the total number of MHC class I and class II loci is typically not known and heterozygosity and rare polymorphisms may inadvertently be misidentified as novel haplotypes or additional genes.

PCR-based techniques for surveying the polymorphic, multicopy MHC genes rely on primer sequences designed to amplify conserved regions. Of concern is the potential amplification of pseudogenes and/or non-classical loci and the possibility of missing loci and/or alleles because of null amplification (Aguilar *et al.*, 2006). A recent examination of class IIB variation in the Common Yellowthroat (*Geothlypis trichas*) demonstrates the depth of sequencing that may be required to characterize the number of class IIB alleles/loci (Bollmer *et al.*, 2010). In that study, 39 different exon 2 sequences (up to 20 class IIB loci) were identified in a single bird and expression analysis suggested at least eight transcribed loci. The total extent of variation in locus number and expression within the turkey is currently unknown.

Sequencing of the turkey *B*-locus (Chaves *et al.*, 2009b) established the presence of three classical class IIB loci with at least two loci being expressed in the spleen.

Goto *et al.* (2002) described the use of single-strand conformation polymorphism for assaying MHC polymorphism in the chicken. This approach relied on PCR primers specific to the class I and class IIB loci for amplification of the highly variable PBRs. The present study used DGGE as an alternative because of the ability to examine larger fragments (440 bp vs 100–300 for single-strand conformation polymorphism) in a relatively quick and repeatable format. As demonstrated, DGGE clearly identifies individuals with divergent class IIB alleles as predicted by the SNP genotyping and haplotype analyses. Sequencing of DGGE PCR products found the number of unique sequences to exceed the number of expected class IIB alleles. Polymerase-induced errors and mosaic sequences can lead to false identification of alleles/loci (Lenz and Becker, 2008) and clearly some of

the turkey sequence differences may be artifacts. The reality that PBR alleles can be shared between loci (Ahmed *et al.*, 2007; Chaves *et al.*, 2010, 2011) and the possibility of additional class IIB loci in the wild turkey are confounding factors. As demonstrated in this study, availability of a reference haplotype and the ability to genotype interspersed MHC regions while avoiding multigene families has advantages over other methods of surveying the MHC. The identification of a set of high-frequency SNPs will provide the basis for an efficient SNP genotype assay for the turkey MHC. SNP genotyping of MHC regions proximal to the multi-locus class I and class II loci combined with haplotype inference also provides an effective means to identify individuals with variant MHC haplotypes that can be used in closer scrutiny of the MHC through expanded sequencing.

### MHC diversity in wild turkeys

A significant loss of species-wide minor alleles in commercial populations has been well documented in the chicken (Muir *et al.*, 2008). Contraction of MHC haplotypes in commercial lines relative to wild birds has also been observed in a previous study of turkeys. Zhu *et al.* (1996) found a single, predominate haplotype (identified by Southern hybridization) present in both commercial lines as well as wild turkeys. However, as in this study, wild birds had a higher proportion of unique haplotypes. Similarly, Buchholz *et al.* (2004) surveyed 40 game farm wild-derived turkeys and identified 40 unique genotypes by Southern hybridization. Interestingly, the single turkey sampled from Minnesota was found to be homozygous at the surveyed loci. Wild turkey populations were at one point extirpated from Minnesota after 1880, as the result of habitat loss during settlement and unregulated hunting. However, the species was successfully reintroduced in the 1970s and the population in the state was estimated at 60 000 in 2006 (Minnesota Department of Natural Resources) with annual harvests over 5000 birds per year since 2000. It would be interesting to determine the degree to which MHC variation may have been restricted in the Eastern subspecies as a result of this reintroduction as it could present a concern with respect to disease resistance.

The high level of sequence diversity is reflected in the inability to establish common haplotypes among wild turkeys, even while using relatively common alleles of extremely close genomic proximity. These results can be explained by high levels of recombination and gene conversion within the region, as has been previously suggested (Chaves *et al.*, 2010). Novel mutations are not likely responsible for the inability to infer and assemble common haplotypes as the polymorphisms used were of higher allele frequencies and thus are likely ancient polymorphisms segregating within the species.

LD was virtually non-existent within the wild turkey MHC-B, although the MHC is one of the highest LD regions of the human genome (Horton *et al.*, 2004). Although LD within the chicken MHC has yet to be reported, LD has been studied on other chromosomes. In a large study of several commercial breeding lines, Andreescu *et al.* (2007) found the extent of LD in the chicken genome lower than other agricultural species. Strong LD ( $r^2 > 0.8$ ) was observed in 10% of SNPs within 0.5 cM. Further, 24% of SNP pairs (within 0.5 cM) had an

$r^2$  value  $> 0.5$ . In a study of wild chickens (Red Jungle Fowl) LD was found to be higher across a 200 kb region of chromosome 1 compared with two other commercial lines (Rao *et al.*, 2008). In the study of a second galliform, Berlin *et al.* (2008) examined nucleotide diversity at 18 autosomal loci in wild populations of the Scandinavian willow grouse (*Lagopus lagopus*). Measures of recombination and population structure at the nuclear level indicated a mature panmictic population with large effective population size. High levels of nucleotide diversity and low LD were observed suggesting LD was shaped by high recombination and purifying selection. Unfortunately, the MHC was not included in the willow grouse study.

The SNPs used to measure LD in the turkey MHC span a distance of 100 kb, which correspond to a genetic distance of 0.25 cM in commercial families (Chaves *et al.*, 2010). Within the wild birds, only four SNP pairs (0.6%) had an  $r^2$  value  $> 0.8$  and 14 (2%) had an  $r^2 > 0.5$ . These values are lower in the wild birds. Several factors could be contributing to the low LD. As stated above, high levels of recombination and/or gene conversion can lead to the loss of LD. On the basis of the results of this study, the total number of haplotypes present in wild turkeys is high. The total number of samples surveyed may not have been large enough to accurately and successfully identify alleles in disequilibrium. Further, sequencing samples in this manner may not sufficiently address the potential for null alleles, which can lead to the loss of LD.

This study focused on sequencing dispersed amplicons in the turkey B-locus to avoid the inherent complications of quantifying variation in polymorphic loci within multi-gene families. This sequencing of 40 individuals found nearly twice as many polymorphisms per base pair in wild turkey as compared with commercial birds (Chaves *et al.*, 2010) indicating a reduction in MHC alleles and potential heterozygosity loss through domestic selection. Over 100 new MHC sequence polymorphisms were identified, with many novel SNPs having lower MAF compared with those shared among wild and commercial birds. Although some of these may be indicative of relatively recent mutations, lower MAF directly reflects the MHC diversity of the sampled individuals. This study indicates that wild turkeys represent a valuable reservoir of MHC diversity. MHC alleles from wild birds could be introgressed into commercial populations if, for example, a wild MHC haplotype is shown to be particularly resistant to a disease of commercial significance.

### Conflict of interest

The authors declare no conflict of interest.

### Acknowledgements

We thank Kristelle Mendoza for laboratory assistance. This research was supported by grants from the University of Minnesota Agriculture Experiment Station and the Cooperative State Research, Education and Extension Service, US Department of Agriculture (2004-35205-14217 and 2009-35205-05302).



## References

- Aeed PA, Briles WE, Zsigray RM, Collins WM (1993). Influence of different B-complex recombinants on the outcome of Rous sarcomas in chickens. *Anim Genet* **24**: 177–181.
- Aguilar A, Edwards SV, Smith TB, Wayne RK (2006). Patterns of variation in MHC class II beta loci of the little greenbul (*Andropadus virens*) with comments on MHC evolution in birds. *J Hered* **97**: 133–142.
- Ahmed KA, Saxena VK, Saxena M, Ara A, Pramod AB, Rajaram ML *et al.* (2007). Molecular cloning and sequencing of MHC class II beta 1 domain of turkey reveals high sequence identity with chicken. *Int J Immunogenet* **34**: 97–105.
- Alcaide M, Edwards SV, Negro JJ, Serrano D, 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, López L, Tanferna A, Blas J, Sergio F, Hiraldo F (2010). Simultaneous analysis of multiple PCR amplicons enhances capillary SSCP discrimination of MHC alleles. *Electrophoresis* **31**: 1353–1356.
- Andreescu C, Avendano S, Brown SR, Hassen A, Lamont SJ, Dekkers JCM (2007). Linkage disequilibrium in related breeding lines of chickens. *Genetics* **177**: 2161–2169.
- Anmarkrud JA, Johnsen A, Bachmann L, Lifeld JT (2010). Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes. *J Evol Biol* **23**: 1206–1217.
- Babik W, Taberlet P, Ejsmond MJ, Radwan J (2009). New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Mol Ecol Res* **9**: 713–719.
- Barrett JC, Fry B, Maller J, Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**: 263–265.
- Berlin S, Quintela M, Hoglund J (2008). A multilocus assay reveals high nucleotide diversity and limited differentiation among Scandinavian willow grouse (*Lagopus lagopus*). *BMC Genet* **9**, .
- Bollmer JL, Dunn PO, Whittingham LA, Wimpee C (2010). Extensive MHC class II B gene duplication in a passerine, the Common Yellowthroat (*Geothlypis trichas*). *J Heredity* **101**: 448–460.
- Bollmer JL, Vargas FH, Parker PG (2007). Low MHC variation in the endangered Galapagos penguin (*Spheniscus mendiculus*). *Immunogenetics* **59**: 593–602.
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL *et al.* (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**: 33–39.
- Buchholz R, Jones Dukes MD, Hecht S, Findley AM (2004). Investigating the turkey's 'snood' as a morphological marker of heritable disease resistance. *J Anim Breed Genet* **121**: 176–185.
- Chaves LD, Harry DE, Reed KM (2009a). Genome-wide genetic diversity of 'Nici', the DNA source for the CHORI-260 turkey BAC library and candidate for whole genome sequencing. *Anim Genet* **40**: 348–352.
- Chaves LD, Kreuth SB, Reed KM (2009b). Defining the turkey MHC: sequence and genes of the B-locus. *J Immunol* **183**: 6530–6537.
- Chaves LD, Faile GM, Kreuth SB, Hendrickson JA, Reed KM (2010). Haplotype variation, recombination, and gene conversion within the MHC-B of the turkey. *Immunogenetics* **62**: 465–477.
- Chaves LD, Kreuth SB, Bauer MM, Reed KM (2011). Sequence of a turkey BAC clone identifies MHC class III orthologs and supports ancient origins of immunological gene clusters. *Cytogenet Genome Res* **132**: 55–63.
- Dickson JG (1992). *The Wild Turkey: Biology and Management*. Stackpole Books: Harrisburg, PA.
- Edwards SV, Gasper J, March M (1998). Genomics and polymorphism of Agph-DAB1, an MHC class II B gene in red-winged blackbirds (*Agelaius phoeniceus*). *Mol Biol Evol* **15**: 236–250.
- Emara MG, Nestor KE, Foster DN, Lamont SJ (1992). The turkey major histocompatibility complex: identification of class II genotypes by restriction fragment length polymorphism analysis of deoxyribonucleic acid. *Poult Sci* **71**: 2083–2089.
- Freeman-Gallant CR, Johnson EM, Saponara F, Stanger M (2002). Variation at the major histocompatibility complex in Savannah sparrows. *Mol Ecol* **11**: 1125–1130.
- Fulton J, Juul-Madsen H, Ashwell C, McCarron A, Arthur J, O'Sullivan N *et al.* (2006). Molecular genotype identification of the *Gallus gallus* major histocompatibility complex. *Immunogenetics* **58**: 407–421.
- Goto RM, Afanassieff M, Ha J, Iglesias GM, Ewald SJ, Briles WE *et al.* (2002). Single-strand conformation polymorphism (SSCP) assays for major histocompatibility complex B genotyping in chickens. *Poult Sci* **81**: 1832–1841.
- Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK *et al.* (2004). Gene map of the extended human MHC. *Nat Rev Genet* **5**: 889–899.
- Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK *et al.* (2008). Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. *J Immunol* **181**: 3393–3399.
- Huson DH, Bryant D (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.
- Jacob JP, Milne S, Beck S, 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.
- Jeffery JKM, Bangham CRM (2000). Do infectious diseases drive MHC diversity? *Microb Infect* **11**: 1335–1341.
- Jones D, Locklear CC, Buchholz R, Hecht SJ, Findley AM (2003). Production of an MHC class II B molecular probe in the turkey, *Meleagris gallopavo*. *J Appl Genet* **44**: 369–373.
- Kaufman J, Jacob J, Shaw I, Walker B, Milne S, Beck S *et al.* (1999a). Gene organisation determines evolution of function in the chicken MHC. *Immunol Rev* **167**: 101–117.
- Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C *et al.* (1999b). The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **401**: 923–925.
- Kelley J, Walter L, Trowsdale J (2005). Comparative genomics of major histocompatibility complexes. *Immunogenetics* **56**: 683–695.
- Knapp LA (2005). Denaturing gradient gel electrophoresis and its use in the detection of major histocompatibility complex polymorphism. *Tissue Antigens* **65**: 211–219.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.
- Latch EK, Smith EJ, Rhodes OE (2002). Isolation and characterization of microsatellite loci in wild and domestic turkeys (*Meleagris gallopavo*). *Mol Ecol Notes* **2**: 176–178.
- Lenz TL, Becker S (2008). Simple approach to reduce PCR artifact formation leads to reliable genotyping of MHC and other highly polymorphic loci—implications for evolutionary analysis. *Gene* **427**: 117–123.
- Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA *et al.* (2005). Nomenclature for factors of the HLA system, 2004. *Intl J Immunogenet* **32**: 107–159.
- Miller MM, Bacon LD, Hala K, Hunt HD, Ewald SJ, Kaufman J *et al.* (2004). 2004 nomenclature for the chicken major histocompatibility (B and Y) complex. *Immunogenetics* **56**: 261–279.
- Mock KE, Latch EK, Rhodes Jr OE (2004). Assessing losses of genetic diversity due to translocation: long-term case histories in Merriam's turkey (*Meleagris gallopavo merriami*). *Conserv Genet* **5**: 631–645.
- Mock KE, Theimer TC, Rhodes Jr OE, Greenberg DL, Keim P (2002). Genetic variation across the historical range of the wild turkey (*Meleagris gallopavo*). *Mol Ecol* **11**: 643–657.

- Moon DA, Veniamin SM, Parks-Dely JA, Magor KE (2005). The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed class I genes. *J Immunol* **175**: 6702–6712.
- Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA et al. (2008). Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *Proc Natl Acad Sci* **105**: 17312–17317.
- Ramon DS, Argüello JR, Cox ST, McWhinnie A, Little AM, Marsh SG et al. (1998). Application of RSCA for the typing of HLA-DPB1. *Hum Immunol* **59**: 734–747.
- Rao YS, Liang Y, Xia MN, Shen X, Du YJ, Luo CG et al. (2008). Extent of linkage disequilibrium in wild and domestic chicken populations. *Hereditas* **145**: 251–257.
- Reed KM (2009). Turkey. In: Cockett NE, Kole C (eds). *Genome Mapping and Genomics in Domestic Animals*. Springer: New York.
- Reed KM, Chaves LD, Mendoza KM (2007). An integrated and comparative genetic map of the turkey genome. *Cytogenet Genome Res* **119**: 113–126.
- Reed KM, Hall MK, Chaves LD, Knutson TP (2006). Single nucleotide polymorphisms for integrative mapping in the Turkey (*Meleagris gallopavo*). *Anim Biotechnol* **17**: 73–80.
- Richardson DS, Westerdahl H (2003). MHC diversity in two *Acrocephalus* species: the outbred Great reed warbler and the inbred Seychelles warbler. *Mol Ecol* **12**: 3523–3529.
- Schneider S, Roessli D, Excoffier L (2000). *Arlequin: A Software for Population Genetics Data Analysis*. Genetics and Biometry Lab, Department of Anthropology, University of Geneva, Geneva, Switzerland.
- Shiina T, Briles WE, Goto RM, Hosomichi K, Yanagiya K, Shimizu S et al. (2007). Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. *J Immunol* **178**: 7162–7172.
- Shiina T, Shimizu S, Hosomichi K, Kohara S, Watanabe S, Hanzawa K et al. (2004). Comparative genomic analysis of two avian (quail and chicken) MHC regions. *J Immunol* **172**: 6751–6763.
- Speller CF, Kemp BM, Wyatt SD, Monroe C, Lipe WD, Arndt UM et al. (2010). Ancient mitochondrial DNA analysis reveals complexity of indigenous North American turkey domestication. *Proc Natl Acad Sci USA* **107**: 2807–2812.
- Stephens M, Donnelly P (2003). A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* **73**: 1162–1169.
- Taylor Jr RL (2004). Major histocompatibility (B) complex control of responses against Rous sarcomas. *Poult Sci* **83**: 638–649.
- Wegner KM, Kalbe M, Schaschl H, Reusch TBH (2004). Parasites and individual major histocompatibility complex diversity—an optimal choice? *Microb Infect* **6**: 1110–1116.
- Westerdahl H, Wittzell H, von Schantz T (2000). MHC diversity in two passerine birds: no evidence for a minimal essential MHC. *Immunogenetics* **52**: 92–100.
- Westerdahl H, Wittzell H, von Schantz T, Bensch S (2004). MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. *Heredity* **92**: 534–542.
- White EC, Briles WE, Briles RW, Taylor Jr RL (1994). Response of six major histocompatibility (B) complex recombinant haplotypes to Rous sarcomas. *Poult Sci* **73**: 836–842.
- Zhu J, Nestor KE, Lamont SJ (1995). Survey of major histocompatibility complex class II haplotypes in four turkey lines using restriction fragment length polymorphism analysis with nonradioactive DNA detection. *Poult Sci* **74**: 1067–1073.
- Zhu J, Nestor KE, Tang Y (1996). Frequencies and genetic diversity of major histocompatibility complex class II haplotypes in commercial turkey lines. *Poult Sci* **75**: 954–958.

## Appendix

Sequence polymorphisms identified in the MHC-B of wild turkeys. Position is given with respect to the B-locus sequence (DQ993255). For each gene, I, E/N, E/S and N indicate intron, exon: non-synonymous, exon: synonymous or non-coding, respectively. Alleles use the genetic ambiguity code (R=A or G, Y=C or T, S=G or C, K=G or T, M=A or C, W=A or T). For DIPs, ‘.’ denotes a deletion.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)