CHARACTERIZATION OF MAJOR HISTOCOMPATIBILITY COMPLEX CLASS I LOCI OF THE
LARK SPARROW (*CHONDESTES GRAMMACUS*) AND INSIGHTS INTO
AVIAN MHC EVOLUTION

Amanda C. Lyons

A Thesis

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE

December 2013

Committee:

Dr. Juan L. Bouzat, Advisor

Dr. Daniel D. Wiegmann

Dr. Ron C. Woodruff

ABSTRACT

Juan L. Bouzat, Advisor

During the last several years, the Major Histocompatibility Complex (MHC) has become increasingly more important in the study of the immunocapabilities of non-model organisms due to its direct involvement in the immune response of vertebrates. The characterization of MHC class I loci in the lark sparrow (*Chondestes grammacus*) revealed multiple MHC class I loci with elevated genetic diversity at exon 3, evidence of positive selection acting upon the peptide-binding region (PBR), and the presence of pseudogenes with limited divergence. The minimum number of functional MHC class I loci is estimated at six, based upon phylogenetic analysis and the detection of multiple haplotypes per individual. Sequence analysis revealed d_N/d_S ratios greater than one at PBR sites, indicative of positive selection for variable amino acids, facilitating antigen recognition of multiple pathogens. In contrast, sequences associated with structural (non-PBR) regions of the protein revealed d_N/d_S ratios less than one, indicating negative selection. A phylogenetic analysis of previously reported avian MHC class I sequences revealed significant clustering of sequences based upon order, family, and species despite the apparent orthology between some of these genes. Results from this study suggest that adaptive selection, concerted evolution, and gene duplication have driven the evolutionary dynamics of MHC class I loci across avian species.

ACKNOWLEDGMENTS

First and foremost, I owe a great deal of gratitude and admiration to my advisor, Dr. Juan L. Bouzat, for his guidance, patience, and time over the past three years. His unwavering trust in my capabilities has been a constant encouragement throughout my graduate studies.

I would also like to thank my remaining graduate committee members, Dr. Ron C. Woodruff and Dr. Daniel Wiegmann, for generously giving me their time, important suggestions, and support in graduate school and during my undergraduate studies at BGSU. Dr. Woodruff was the first to recognize my research potential as only a sophomore undergraduate, greatly impacting my academic decisions and sparking my interest in graduate school. Dr. Wiegmann has continually provided much needed encouragement and confidence in my abilities as a graduate student.

Past and present members of the Bouzat Lab have had a special hand in my success as a graduate student. Previous lab members Dr. Jeremy Ross and Bethany Swanson completed the field work and DNA extractions used in this study. My fellow graduate students, Gabrielle Knafler, Joseph Schalk, and Ashley ‘Lani’ Estopinal, each provided me with thoughtful discussion, encouragement, and friendship during my time in the Bouzat Lab. Lastly, our undergraduate lab member, Mercedes Chumbley, has brought a great deal of laughter and encouragement to the latter part of my time in graduate school.

Lastly, I owe more gratitude to my family and friends than words can properly convey. The continual love and support from my parents, Thomas and Claudette, and older brother, Matthew, throughout my education has been the cornerstone of my persistence in higher education. Also, to all my dear friends who, while patiently tolerating my frequent absence, always seem to reassure me that our friendships will remain unchanged. A very special recognition remains for my partner, James Buie, for his immense amount of love, patience, and encouragement throughout my achievements and difficulties in graduate school.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
MATERIALS AND METHODS.....	6
Sampling	6
DNA Sequencing	6
Data Analyses	7
Database Survey of Avian MHC Class I Genes	8
Phylogenetic Analyses	8
RESULTS	10
Lark Sparrow MHC Class I Characterization	10
Avian MHC Evolution.....	12
DISCUSSION	16
Lark Sparrow MHC Class I Characterization	16
Avian MHC Evolution.....	18
TABLES AND FIGURES	21
REFERENCES	29
APPENDIX.....	37

LIST OF TABLES

Table		Page
1	Characterization of MHC class I sequences of the lark sparrow (<i>Chondestes grammacus</i>).....	21
2	Comparison of genetic diversity estimates between functional MHC class I alleles and pseudogenes sampled from four lark sparrow adults.....	22
3	Avian MHC class I sequences reported to GenBank included in this study	23

LIST OF FIGURES

Figure		Page
1	Hypothetical phylogenies based on divergent and concerted evolution of avian MHC loci	25
2	Unrooted maximum likelihood phylogeny of lark sparrow MHC class I sequences spanning exon 3 and extending into flanking introns 2 and 3 (396 bp).....	26
3	Maximum likelihood phylogeny of Aves MHC class I exon 3 DNA sequences (235 bp) retrieved from GenBank	27
4	Maximum likelihood phylogeny of passerine MHC class I exon 3 DNA sequences retrieved from GenBank	28

INTRODUCTION

The major histocompatibility complex (MHC) is a multi-gene family that codes for diverse proteins directly involved in the immune response of vertebrates. These genes are one of the most polymorphic genes studied (Parham and Ohta 1996) with a remarkable degree of gene duplication and divergence, characteristics reflective of its essential function in combating infectious diseases and parasites. Protein products of the MHC have varying functions within the immune response, but possibly the most commonly studied are the cell surface receptors produced by the MHC class I and class II loci. These molecules bind to fragments of foreign peptides, presenting them on the surface of cells to be recognized by T-cells and thus initiating the appropriate immune response (Abbas et al. 1994). Class I molecules function throughout the vertebrate body to target intracellular foreign peptides, whereas class II molecules primarily function in the liver where extracellular foreign peptides are the main target (Abbas et al. 1994). The structure of the MHC has a remarkable degree of diversity across vertebrate taxa, although the function remains conserved. This diversity of structure as well as overall high genetic polymorphism in relation to the immune response has been studied in a variety of research fields, including disease (Langefors et al. 2001; Hedrick 2002; Westerdahl 2007), mate choice (Landry et al. 2001; Paterson and Pemberton 1997; Knafler et al. 2012), genomics (Kulski et al. 2002; Kelley et al. 2005), and gene duplication (Nei et al. 1997).

Elevated genetic diversity at MHC loci is maintained by natural selection favoring sequence variation (Klein 1986; Hughes and Yeager 1998). Perhaps the most measurable evolutionary effect is positive selection, which occurs when changes in a nucleotide sequence lead to an increased fitness advantage. In MHC, positive selection is most evident at the so-called peptide-binding region (PBR) of the cell surface receptor. At this location of the protein, certain amino acids bind to foreign peptides with a relatively high degree of specificity. As a consequence, a set of amino acids at the PBR may bind to and recognize only a handful of peptides that may not be recognized by a similar, yet different, set of PBR amino acids. This characteristic creates a selective advantage for PBR diversity, a trait that would allow an individual the potential to overcome a wide range of pathogen pressures (Hughes and Yeager

1998). Positive selection at the MHC is evidenced by increased nucleotide diversity at the PBR codons which results in a relatively high rate of nonsynonymous amino acids changes within the mature protein (Hughes and Nei 1998).

Another form of selection that may act upon the MHC is balancing selection. Balancing selection is fueled by the evolutionary “arms-race” between the vertebrate immune system and pathogens (Edwards and Hedrick 1998; Takahata and Nei 1990). In the case of the MHC, balancing selection is believed to occur specifically by the processes of overdominance (heterozygote-advantage) and negative frequency dependence. Individuals that are heterozygous at MHC loci are better equipped with MHC proteins that are capable of recognizing presumably twice as many foreign peptides than their homozygous conspecifics. This leads to a selective advantage for heterozygotes within a population under a range of pathogenic pressures, thus maintaining a diverse array of MHC alleles (Hughes and Nei 1988; Penn et al. 2002). Negative frequency dependent selection, on the other hand, maintains MHC allelic diversity by the relative fitness advantage of having a rare MHC allele within a population under the pressure of a rapidly evolving pathogen (Borghans et al. 2004). As one MHC allele, for example, becomes more frequent within a population due to its respective fitness advantage in overcoming a pathogen, the pathogen population may in turn evolve the ability to overcome its detection. Individuals with a less frequent MHC allele may then have a relative fitness advantage within the population if the allele is able to initiate an effective immune response against the pathogen. This relationship between the MHC initiated immune response and pathogen pressures leads to continually coevolving populations of hosts and pathogens, which consequently results in a great diversity of MHC alleles in a vertebrate population.

Diverse MHC alleles are also maintained by disassortative mating in which individuals preferentially pair with genetically dissimilar mates, likely the result of selection favoring behaviors that avoid inbreeding (Apanius et al. 1997; Tregenza and Wedell 2000); however, this is not always the case (Paterson and Pemberton 1997; Knafler et al. 2012). In some studies, disassortative mating is driven by mate choice cues, such as plumage colorations and overall body condition (Hamilton and Zuk 1982; von Schantz et al. 1996; Dunn et al. 2013). Because of the close association between MHC diversity and

fitness and the interesting selection dynamics acting on these loci, research investigating the characteristics of MHC has become increasingly more frequent in non-model species.

Detailed studies of the MHC in humans (human leukocyte antigen, HLA) have provided the basis for investigations of different aspects of the MHC in countless non-model vertebrate species [e.g. foxes, Aguilar et al. (2004); tuataras, Miller et al. (2010); crocodiles, Jaratlerdsiri et al. (2012); salmon, Garrigan and Hedrick (2001); and guppies, van Oosterhout et al. (2006)]. These studies revealed a diverse organization, gene orientation, and gene duplication of MHC loci, generating varying hypotheses as to how diversity among these genes is maintained or lost. Avian species have been found to provide a very intriguing example of the evolutionary history of the MHC, possibly influenced by the exceptional degree of avian diversity and relatively recent species radiation of many orders. The MHC of the domestic chicken (*Gallus gallus*) has been thoroughly characterized over the past few decades (Kaufman et al. 1995; Shiina et al. 2004; Shiina et al. 2007). Chicken MHC is extremely simplified and compact compared to the MHC of other vertebrate species. Because of this, it has been characterized as the “minimal-essential” MHC (Kaufman et al. 1995). Codons that specify the PBR region in both humans and homologous regions of the chicken and other avian species have been identified (Björkman et al. 1987; Kaufman et al. 1995; Balakrishnan et al. 2010; Alcaide et al. 2009). Indeed, we now have detailed sequences of MHC loci of many avian species, including game birds (Shiina et al. 2007; Chaves et al. 2009; Shiina et al. 2004; Eimes et al. 2013; Wang et al. 2012; Ye et al. 2012; Singh et al. 2010), ratites (Miller et al. 2011), waterfowl (Moon et al. 2005; Xia et al. 2005; Liang et al. 2010), raptors (Alcaide et al. 2009; Gangoso et al. 2012), penguins (Tsuda et al. 2001; Bollmer et al. 2007; Knafler et al. 2012), and songbirds (Westerdahl et al. 2004; Richardson and Westerdahl 2003; Bollmer et al. 2012; Balakrishnan et al. 2010; Promerová et al. 2009; Spurgin et al. 2011; Borg et al. 2011; Sepil et al. 2012; Wutzler et al. 2012). A remarkable range of results concerning the structural organization of the MHC from multiple avian families has been discovered, leading to major insights on the potential mechanisms driving MHC evolution within birds (Hess and Edwards 2002).

As an extensive gene family, the MHC exhibits two main patterns of evolution when comparing

homologous loci across species: divergent and concerted evolution (Hess and Edwards 2002). Divergent evolution drives loci to diverge independently of other homologous loci. This may be caused by the evolution of a new yet similar function favored by pathogenic selection. Concerted evolution, comparatively, occurs when loci of a gene family within a species (representing paralogous genes that evolved through gene duplication) experience unequal crossing over or gene conversion and, as a result, have similar DNA sequences (Högstrand and Böhme 1999; Walsh and Stephan 2001). The phylogenetic effects of divergent evolution upon homologous loci following pre-speciation duplication events in related species are illustrated in Figure 1A. In this case, orthologous loci cluster together regardless of species. On the other hand, concerted evolution of homologous loci that duplicated prior to speciation events will cause phylogenetic clustering of paralogous genes within each species (Figure 1B). This same phylogenetic effect, though, will also occur if loci originated through gene duplication after speciation. Although these are simplified models for explaining the evolution of avian MHC, phylogenetic results based on MHC sequence data from non-model species have the potential to further our understanding of the dynamic nature of MHC evolution.

The lark sparrow (*Chondestes grammacus*) represents a passerine species with documented populations of reduced size subjected to extensive pathogen pressures (Swanson 2012). In Ohio, where the species is classified as a state endangered bird (Ohio Department of Natural Resources 2012), the breeding range of the lark sparrow has been in decline over recent decades due to decreases in suitable habitat (Martin and Parrish 2000). Denser breeding populations exist mainly in the Central Plains of North America and California (Martin and Parrish 2000; Baepler 1968). Previous studies of avian malaria within lark sparrow populations have revealed relatively high infection rates across its range (Swanson 2012), which may act as an important selection pressure on MHC diversity. Characterization of the lark sparrow MHC provides, therefore, the potential to expand our current knowledge on the evolutionary processes influencing this extensive gene family.

The goal of this study is to provide a molecular characterization of the major histocompatibility complex class I genes of the lark sparrow to estimate the minimum number of MHC class I loci,

individual alleles, and the potential presence of pseudogenes in this species. DNA sequences of the exon 3 region of MHC class I loci were targeted by PCR and gene cloning techniques, which allowed for estimates of genetic diversity and assessment of selective pressures. A phylogenetic analysis of characterized MHC gene sequences provided some insight into the evolutionary relationships between MHC loci. As commonly found in other avian species, it is hypothesized that levels of MHC class I variation in the lark sparrow are adaptive in relation to pathogen pressures. Elevated levels of genetic diversity within loci, sequence divergence between functional loci, and a ratio of nonsynonymous to synonymous base pair substitution rates significantly greater than 1 would provide evidence of positive selection in relation to the immune function of these genes.

In addition to the characterization of lark sparrow MHC genes, I performed a detailed survey of available MHC class I sequences from previously studied avian species to compare the estimated minimum number of loci and the extent of MHC class I pseudogene prevalence in avian genomes. Avian MHC phylogenies provided insights regarding the timing and rates of gene duplications as well as the potential roles of divergent and concerted evolution. Results for this study provide insights into the dynamics of the evolutionary processes driving avian MHC diversity and bring to light important avenues for further studies on avian MHC evolution.

MATERIALS AND METHODS

Sampling

The genetic characterization of MHC class I loci of the lark sparrow was based on the analysis of samples collected during the breeding seasons of 2005 and 2007 from two populations: one from the Oak Openings region in northwest Ohio (OH) and one from central Texas (TX). Individual birds were captured by mist-netting. Blood samples were drawn from the ulnar vein using a heparinized capillary tube and stored at 4°C in Longmire's lysis buffer (100 mM Tris, 100 mM EDTA, 10 mM NaCl, 0.5% SDS; Longmire et al. 1986). Standard DNA extractions were completed by proteinase-K digestion, sodium acetate protein precipitation, and subsequent ethanol precipitation of DNA (Sambrook et al. 1989). DNA was resuspended in sterile water at a concentration of 30 ng/μl for genetic analysis. Four adult lark sparrows were chosen at random for this study: two samples from the isolated Ohio population in the Oak Openings region and two samples from the large Texas population.

DNA Sequencing

Polymerase chain reaction (PCR) was conducted using primers designed for the amplification of MHC class I exon 3 of the great reed warbler (*Acrocephalus arundinaceus*). Primers HN34 (5' – CCA TGG GTG TCT GTG GGT A – 3') and HN45 (5' – CCA TGG AAT TCC CAC AGG AA – 3') are located in the flanking introns of exon 3, the coding region that constitutes half of the peptide binding site in mature MHC molecules (subunit α2) (Bonneaud et al. 2004; Westerdahl et al. 2004; Promerová et al. 2009). PCR amplifications were carried out in 20 μl volumes containing 1 μl extracted DNA, 2.0 mM MgCl₂, 0.5 μM of each primer, 150 μM of each dNTP, 1X Promega GoTaq Buffer, and 1 unit of Promega GoTaq DNA Polymerase (Promega; Madison, Wisconsin). Samples were amplified using a MJ Research Thermal Cycler (MJ Research; Waltham, Massachusetts) with a thermal profile of 25-30 cycles at 94°C for 30 sec, 63°C for 30 sec, and 72°C for 30 sec. Positive amplifications were visualized under UV light by 1% agarose gel electrophoresis with ethidium bromide staining. A later attempt to exclude the amplification of nontranslatable pseudogenes involved PCR with the lab designed forward primer HN34-2 (5' – CAT GGG TCT CTG TGG GTA CAA T – 3') and HN45 using the same PCR protocol.

PCR products were cloned into *E. coli* competent cells following manufacturers' directions using Invitrogen TOPO TA (Life Technologies; Carlsbad, California) and Promega pGEM-T Easy (Promega; Madison, Wisconsin) kits. Purified clones were verified using *Eco*R1 restriction enzyme digestion. Positive clones were sequenced by the University of Chicago Comprehensive Cancer Center's DNA Sequencing and Genotyping Facility by cycle sequencing reactions using fluorescent dye terminators and standard sequencing primers. Each clone was sequenced at least twice to confirm the validity of the resulting sequence. Several clones per individual (10 – 44) were prepared and sequenced in order to estimate the minimum number of MHC class I loci. The high number of clones sequenced per individual increased the probability of identifying alleles from multiple loci as well as related pseudogenes.

Data Analyses

Editing and analyses of MHC class I sequences were performed with several bioinformatic software programs. Initial editing and organization of sequence files were completed with BioEdit (Hall 1999). Sequences were aligned using CLUSTAL X (Larkin et al. 2007) and unique haplotypes were named according to standardized nomenclature proposed by Klein et al. (1990). Each haplotype will hereafter be referred to as *Chgr0##* (*Chondestes grammacus* [haplotype number]). After removing primer sequences, DNA sequences were included in the characterization of lark sparrow MHC if a sequence was either identified by more than one clone or differed by at least two base pairs from all sequences found in an individual. This is a conservative strategy to control for the possibility of *Taq* polymerase or DNA sequencing errors.

DNA sequences were confirmed to be the intended target by use of NCBI BLAST (Altschul et al. 1990) indicating close identity to previously published MHC class I exon 3 sequences of closely related passerine species. Exon 3 and introns 2 and 3 were assigned based upon published avian MHC class I sequences. Codons were translated using MEGA 5 (Tamura et al. 2011) and assignment of the PBR was based upon assumed passerine PBR (Balakrishnan et al. 2010; Alcaide et al. 2009) previously inferred from the HLA (Björkman et al. 1987) and chicken MHC (Kaufman et al. 1995). Sequences indicating the presence of at least one stop codon (due to changes in reading frame from indels or nonsense sequences)

were classified as pseudogenes. Based upon sequences found to be translatable, a minimum number of functional MHC class I loci was estimated with the conservative assumption that all loci are in heterozygosity.

DnaSP (Librado and Rozas 2009) was used to identify haplotypes and estimate measures of genetic diversity (nucleotide diversity, number of nucleotide differences, and number of polymorphic sites). Estimates of nonsynonymous (d_N) and synonymous (d_S) base pair substitution rates of functional alleles were calculated using MEGA 5. Ratios of d_N to d_S (ω) of the PBR and non-PBR codons comprising exon 3 were calculated separately, along with an overall estimate of ω for exon 3.

Database Survey of Avian MHC Class I Genes

In order to compare MHC class I sequences across avian orders and families, a survey of GenBank (Benson et al. 2005) was performed on February 6-7, 2013 with the criteria of class Aves, MHC class I, exon 3, and/or $\alpha 2$. Species with reported sequences were recorded and published articles were surveyed for further information with regards to estimates of the number of functional MHC class I loci and the presence of pseudogenes characterized within each species. Representative sequences for each avian species, distinct loci, and putative pseudogenes were downloaded based upon published information. A nucleotide sequence alignment of representative avian sequences and lark sparrow sequences reported in this study was performed using CLUSTAL X (Larkin et al. 2007).

Phylogenetic Analyses

Two phylogenies were created with MEGA 5 using the maximum likelihood method. Branch support was determined with bootstrapping (1000 replications). A lark sparrow MHC class I phylogeny included all pseudogenes and functional alleles reported in this study spanning the entirety of exon 3 and partial sequences of flanking introns. A published chicken (*Gallus gallus*) MHC class I sequence (BF1, AB268588) was included as the outgroup to determine accurate placement of the root, although the sequence was excluded from the final phylogeny.

Phylogenetic analysis of avian MHC class I included representative MHC class I sequences downloaded from GenBank and selected lark sparrow MHC class I sequences representative of multiple

loci and pseudogenes characterized in this study. A published saltwater crocodile (*Crocodylus porosus*) MHC class I exon 3 sequence (HQ008787) was included as the outgroup (Jaratlerdsiri et al. 2012). Due to the limited sequence lengths of many species' MHC class I exon 3, only a partial sequence of 235 base pairs was analyzed in the overall avian phylogeny.

RESULTS

Lark Sparrow MHC Class I Characterization

A total of 103 sequences (spanning MHC class I exon 3 and including partial sequences of flanking introns 2 and 3) were analyzed from four adult lark sparrows. Sequences varied between 386 and 629 base pairs in length, with the majority being 427 base pairs long. For analytical reasons, residues extending into intron 3 were removed along with forward and reverse primer sequences. The multiple sequence alignment of all lark haplotypes was 396 bp long. Initial analysis of sequences revealed a total of 46 distinct DNA sequences (*Chgr001-Chgr046*), but 25 of these sequences were excluded from further analysis because they represented unique sequences with single base pair differences. A total of 21 sequences were therefore characterized and submitted to GenBank (Table A1). Each sequence was confirmed to exhibit identity with previously reported Passerine MHC class I sequences based upon individual BLAST results. Sampling effort per individual (based on the number of clones sequenced per individual) ranged from ten to 44 and resulted in a range of two to nine MHC sequences per individual (Table 1).

Reading frames of exon 3 regions of DNA sequences were translated using MEGA 5. Of the 21 sequences, ten were found to be nontranslatable by changes in reading frame due to indels and/or the presence of stop codons. These sequences were thus considered to be pseudogenes. The number of pseudogenes identified within the four individuals ranged between two and five (Table 1) and two of the ten pseudogene sequences were found to be identical in at least two of the four individuals. The remaining 11 translatable exon 3 sequences were 273-276 bp DNA sequences that translated into 91-92 amino acids in length, respectively, with 14 of these codons representing the peptide binding region (PBR). Interestingly, one of the PBR codons (nucleotide positions 193-195; Appendix Figure A1) was missing from the majority of the lark sparrow MHC class I sequences recovered, including all pseudogene sequences. Only three functional haplotypes (*Chgr033*, *Chgr036*, and *Chgr039*) exhibited nucleotides representing this specific PBR codon. For one individual, OH043, seven translatable exon 3 sequences were characterized, indicating that there must be a minimum of four presumably functional MHC class I

loci within the lark sparrow genome. Translatable sequences of the three remaining individuals ranged between zero and two (Table 1). Each of the 11 translatable sequences was unique to an individual; that is, no two individuals were found to share identical translatable MHC class I haplotypes. However, some resulting protein sequences were found to be identical within an individual (OH043; one pair) and between two individuals (OH043 and TX01; one pair) due to the redundancy of the genetic code.

In order to identify signals of selection on lark sparrow MHC class I genes, sequences were pooled together for analysis (Table 2). Overall, genetic diversity estimates were found to be elevated (nucleotide diversity $\pi=0.126$, average number of nucleotide differences $k=48.1$; $n=21$) in considering all sequences with the exon and introns combined. As expected, estimates of genetic diversity were found to be considerably different between exon and intron regions, with elevated diversity estimates for the exon region ($\pi=0.140$ versus $\pi=0.091$). Further differences became evident when analyzing functional and pseudogene sequences ($\pi=0.115$ and $\pi=0.032$, respectively). The ratio of nonsynonymous to synonymous base pair substitution rates (d_N/d_S) of exon 3 from functional sequences was found to be 0.619, yet differences in d_N/d_S of non-PBR and PBR codons composing the exon were clearly evident. Non-PBR codons had a d_N/d_S of 0.483, characteristic of DNA sequences under negative selection, whereas PBR codons had a d_N/d_S of 1.127, signaling weak positive selection. Estimates of nucleotide diversity of each of these regions also differed significantly with π values of 0.284 and 0.085 for the PBR and non-PBR, respectively.

Analysis of the phylogenetic relationships of lark sparrow MHC class I sequences revealed limited divergence among pseudogene sequences in comparison to the divergence between functional sequences, as indicated by measures of genetic diversity within the two groups of sequences (Figure 2). Nine of the ten characterized pseudogenes formed a monophyletic group with 100% bootstrap support. The remaining pseudogene (*Chgr034*) revealed a basal clustering with a group of three functional alleles (*Chgr017*, *Chgr018*, *Chgr026*) with 58% bootstrap support. The 11 characterized MHC class I functional alleles formed six monophyletic groups (loci A-F; Figure 2) with 93-100% bootstrap support, each containing one to three haplotypes. Each group contained no more than two haplotypes found in a single

individual (Figure 2).

Avian MHC Evolution

GenBank searches of avian MHC class I exon 3 DNA sequences revealed over 1000 sequences reported from 36 species representing ten of the 29 extant avian orders (Table 3). Each order was represented by one to ten species, with Passeriformes being the most represented group. Based upon findings from the published studies reporting the MHC class I sequences, a minimum number of functional loci and pseudogene presence was determined for each species, when possible. The MHC gene family has been extensively mapped and studied within most of the species representing Galliformes (seven species reported, six mapped; Shiina et al. 2004; Shiina et al. 2007; Chaves et al. 2009; Singh et al. 2010; Wang et al. 2012; Ye et al. 2012; Eimes et al. 2013; Table 3). Two MHC class I loci (BF1 and BF2) were originally identified in the chicken (*Gallus gallus*) and similar organization of the gene family with two class I loci was found in the turkey (*Meleagris gallopavo*), the black grouse (*Tetrao tetrix*), and the golden pheasant (*Chrysolophus pictus*). Interestingly, characterization of the greater prairie chicken (*Tympanuchus cupido*) revealed only one class I locus (Eimes et al. 2013) and the Japanese quail (*Coturnix japonica*) was found to have four loci, apparently the result of gene duplications (Shiina et al. 2004). Furthermore, three pseudogene loci were also identified within the Japanese quail MHC, although none of these contained sequences homologous to exon 3 so they could not be included in the phylogenetic analysis of this study. The MHC of the helmeted guineafowl (*Numida meleagris*) has not been mapped, so estimates of the number of functional loci are limited (Singh et al. 2010).

Species representing Anseriformes have been found to have between one (domestic goose *Anser anser*; Xia et al. 2005) and two (mallard *Anas platyrhynchos*; Moon et al. 2005, and bar-headed goose *Anser indicus*; Liang et al. 2010) functional loci (Table 3), although the number of MHC class I loci of the domestic goose may be underestimated due to detection methods based upon cDNA. Three pseudogene loci (named UBA, UCA, and UEA) have been identified in the mallard MHC (see below).

Orders Gruiformes, Procellariiformes, and Charadriiformes were each represented by MHC sequences from a single species: the Florida sandhill crane (*Grus canadensis pratensis*), the blue petrel

(*Halobaena caerulea*), and the red-billed gull (*Larus scopulinus*), respectively. Studies reporting these sequences suggest the presence of at least two functional MHC class I loci in the Florida sandhill crane genome (only one locus sequenced), at least five loci in the blue petrel (with pseudogenes), and at least two loci in the red-billed gull genome (Table 3).

MHC class I sequences from raptors (orders Strigiformes, Cathartiformes, Accipitriformes, and Falconiformes) were represented by a total of 13 species (Table 3). A minimum of two MHC class I loci were identified in Cathartiformes and Accipitriformes and only one locus was estimated to be present in the genomes of Falconiformes. Conclusions regarding the MHC class I of Strigiformes, represented by only the Eurasian eagle-owl (*Bubo bubo*), were not possible due to limited data. No pseudogenes were identified in any raptor species.

Ten species representing seven of the more than 110 families of the order Passeriformes were found to have reported MHC class I DNA sequences (Table 3). Passeriformes exhibited the highest number of MHC class I loci, with a range of at least three presumably functional loci in the Seychelles warbler (*Acrocephalus sechellensis*) to a minimum of 16 in the great tit (*Parus major*). Pseudogenes were also identified in six of the ten Passeriformes species with at least three pseudogene loci estimated in both the zebra finch (*Taeniopygia guttata*) and the great tit (*Parus major*).

Combined with lark sparrow MHC class I sequences, representative sequences of reported Aves MHC class I functional genes and pseudogenes were aligned to create a maximum likelihood phylogeny (Figure 3). Six functional lark sparrow MHC sequences (*Chgr*018, 019, 036, 033, 015, and 012), representing each a functional locus (Groups A-F; Figure 2), and four pseudogenes (*Chgr*001, 005, 021, and 034) were included in the phylogeny. For simplicity, remaining lark sparrow MHC class I sequences were excluded from the final phylogeny as these were found to cluster with their assigned locus (A-F) or with other pseudogenes within a preliminary phylogeny (data not shown). The alignment of all sequences included 235 base pairs of exon 3 due to only partial sequences reported for many of the included avian species. A total of 88 avian MHC class I sequences were included in the alignment along with a reptilian outgroup (the saltwater crocodile *Crocodylus porosus*; GenBank accession number HQ008787). The

phylogeny included all species with clearly defined loci based upon extensive gene mapping, with reported locus names following the species name (Figure 3).

Phylogenetic results revealed varying patterns of relatedness among avian MHC class I sequences (Figures 3 and 4). Species consistently clustered according to their order with 71-99% bootstrap support. Passeriformes formed a separate group with 99% support whereas the remaining eight avian orders formed a cluster with 79% support. MHC class I loci of Galliformes (BF1/IA1 and BF2/IA2) tended to group together by species (*Chrysolophus pictus*, *Tetrao tetrix*, *Coturnix japonica*, *Gallus gallus*) with high levels of bootstrap support (78-100%; Figure 3). The two loci (IA1 and IA2) of the turkey (*Meleagris gallopavo*), on the other hand, did not group together. Locus IA2 clustered with both *Chrysolophus pictus* loci and the single greater prairie chicken (*Tympanuchus cupido*) locus (IA) with 66% bootstrap support. The remaining turkey locus IA1 clustered with the two black grouse sequences (*Tetrao tetrix*) with low bootstrap support (<50%).

MHC class I sequences from Anseriformes did not group by species, but may have grouped based upon loci. The mallard (*Anas platyrhynchos*) UAA locus (found to be highly expressed; Moon et al. 2005) clustered with two bar-headed goose sequences with 65% bootstrap support. In contrast, the mallard UDA (found to have low expression) and UBA (a pseudogene locus) clustered with 84% support but also grouped with the single domestic goose locus and the remaining bar-headed goose sequence with 70% support.

All Falconiformes sequences, with only one known MHC class I locus per species, grouped together by species with high internal bootstrap support. Species grouped into a kestrel species cluster (*Falco naumanni* and *Falco tinnunculus*) and a falcon species cluster (*Falco peregrines* and *Falco eleonora*) with 98% and 78% bootstrap support, respectively. Accipitriformes sequences, with two characterized MHC class I loci per species and three species included, formed a monophyletic group with 71% bootstrap support, but with low bootstrap support (<50%) for most internal clades. MHC class I sequences from orders with only one representative species (i.e., Cathartiformes, Procellariiformes, Charadriiformes, and Gruiformes) clustered consistently by species/order with the exception of the red-

billed gull (Charadriiformes; *Larus scopulinus*). This species, with two characterized loci (UAA and UBA), revealed more similarity between its UBA locus and the one available sequence from the Florida sandhill crane (*Grus canadensis pratensis*; 71% bootstrap) than to its other reported locus, UAA.

Phylogenetic results of Passeriformes are shown in Figure 4. Of the seven families represented in this phylogeny, only two have MHC class I sequences from more than one species. Representative sequences of these families, Acrocephalidae (great reed warbler *Acrocephalus arundinaceus* and Seychelles warbler *Acrocephalus sechellensis*) and Paridae (great tit *Parus major*, green-backed tit *Parus monticolus*, and blue tit *Cyanistes caeruleus*), clustered individually by family. However, with the exception of Paridae (93% bootstrap), family groups did not have bootstrap support (all <50%). The two MHC class I sequences of the green-backed tit grouped with 99% bootstrap support whereas a cluster of all great tit pseudogenes and functional genes and two blue tit pseudogenes grouped with 89% support. The two blue tit functional genes did not group significantly (<50% support) with any single sequence within this cluster. A cluster of two pseudogenes and two functional genes of the house sparrow (*Passer domesticus*) have 97% bootstrap support, yet two other house sparrow functional genes did not cluster significantly with any other sequence. Remaining sequences from reported Passeriformes MHC class I genes did not cluster with high bootstrap support. Lark sparrow MHC class I sequences reported in this study revealed high bootstrap support for some groups. Loci E and F (based upon designations assigned in Figure 2) clustered independently with 99% support. On the other hand, locus A clustered with pseudogene *Chgr034* with 71% support. The remaining pseudogenes clustered together with 100% support and indicated similarity to a common yellowthroat (*Geothlypis trichas*) pseudogene. Similarly, the three remaining functional loci (B, C, and D) did not cluster together, consistent with the idea that these sequences represent independent loci.

DISCUSSION

Lark Sparrow MHC Class I Characterization

Characterization of MHC class I loci of the lark sparrow indicated increased levels of genetic diversity, evidence of extensive gene duplication, the presence of multiple functional loci, and the presence of pseudogenes. These MHC characteristics are consistent with the majority of other Passerine MHC studies, which also reported multiple copies of MHC class I genes and pseudogenes (Westerdahl et al. 2004; Bollmer et al. 2012; Balakrishnan et al. 2010; Borg et al. 2011; Sepil et al. 2012; Wutzler et al. 2012; Alcaide et al. 2013). The minimum number of functional loci for the lark sparrow was estimated to be between four (based upon the number of unique functional sequences identified in a single individual; Table 1) and six (as inferred by phylogenetic analysis; Figure 2). The number of loci may very well be higher, as this number is likely conservative and based upon limited sampling effort per individual. Other orders of birds did not exhibit the same degree of gene duplication and extent of pseudogenes identified in Passeriformes, although only a limited number of orders were represented. One apparent trend, though, is that an increase in the number of functional MHC class I loci often coincides with the presence of pseudogenes (e.g., Japanese quail, mallard, blue petrel, and Passeriformes; see Table 3). Further MHC class I characterizations of species within other orders will certainly reveal more information about the evolution of the avian MHC class I gene family. Specifically, the organization of the MHC class I gene family in the basal avian orders Struthioniformes and Tinamiformes (ratites and tinamous, respectively) may be essential in determining the structure of the ancestral avian MHC.

Some cautionary limitations should be considered in drawing conclusions based upon the generalizations made from MHC characterizations. Studies using only DNA sequences constructed from complementary DNA (cDNA) do not allow for the identification of pseudogenes, as in the helmeted guineafowl (Singh et al. 2010) and domestic goose (Xia et al. 2005). Because of their apparent prevalence in the MHC, pseudogenes are sometimes avoided all together simply due to the goals of some studies, so limited information about their presence may be available. It should also be noted that in some studies, the estimation of the minimum number of functional loci may be underestimated due to the

unknown specificity of primers, extent of sampling effort, and the conservative assumption of heterozygosity. Likewise, some presumably functional loci based upon partial gene sequences (such as exon 3) may in fact be pseudogenes due to, for example, the presence of mutations in regulatory regions. This has been observed, for instance, for three MHC class loci in the mallard (Moon et al. 2005). There may also be differences in the degree of expression of each MHC locus (Shiina et al. 2007; Moon et al. 2005) and technical issues that may prevent DNA replication of specific alleles. Many of these issues have the potential to be resolved in the future with extensive and detailed characterization of individual species' MHC gene families based on a range of laboratory methods, including whole genome sequencing.

The characterization of lark sparrow MHC class I sequences reported in this study revealed evolutionary genetic signatures of varying selective processes. Differing levels of genetic diversity and divergence between functional and pseudogene MHC sequences in the lark sparrow are indicative of selective effects. Functional sequences revealed elevated levels of genetic diversity and significant divergence between sequences whereas pseudogenes had both limited genetic variation and low divergence (Figure 2). Selection favoring a diverse range of MHC molecules in order to cope with multiple, fast-evolving pathogens would cause functional MHC genes to exhibit increased genetic diversity (Hughes and Nei 1988; Penn et al. 2002). Furthermore, gene duplication events may further increase diversity through diversification of multiple gene loci (Walsh and Stephan 2001). Similar results have been reported for most other passerine species, including the great reed warbler (Westerdahl et al. 2004), the common yellowthroat (Bollmer et al. 2012), and the great tit (Sepil et al. 2012).

The characterization of exon 3 functional sequences also revealed differential effects of selection on structural regions associated with the PBR and non-PBR codons of the MHC class I gene. The peptide binding region revealed higher levels of nonsynonymous to synonymous substitution rates, documenting the effects of weak positive selection ($d_N/d_S > 1$), whereas the structural regions of the MHC molecule associated with non-PBR codons revealed negative selection ($d_N/d_S < 1$). These results are expected, as diverse foreign peptide recognition is advantageous whereas errors in protein structure can

render the protein nonfunctional, and limit the immunocapabilities of individuals (Hughes and Nei 1988). Similar d_N/d_S ratios have been documented in many avian MHC studies (e.g. Wutzler et al. 2012; Borg et al. 2011; Spurgin et al. 2011), although this is not always the case (Eimes et al. 2013). A recent study investigating MHC class I diversity within several passerine species found elevated d_N/d_S ratios of the PBR in comparison to the non-PBR of most species (Alcaide et al. 2013). The degree of differences between the two ratios was variable, however, revealing the complex nature of selection acting upon the passerine MHC.

Avian MHC Evolution

The limited genetic diversity and divergence of nonfunctional MHC sequences identified in this study are likely due to the neutral nature of these sequences, especially in comparison to similar measures in functional sequences. The MHC class I pseudogene characteristics identified in the lark sparrow have also been reported in other passerine species. For example, MHC class I pseudogenes of the great reed warbler (Westerdahl et al. 2004) and the great tit (Sepil et al. 2012) also have limited divergence and group into a single phylogenetic cluster. However, MHC class I pseudogenes of passerines included in this study generally seemed to cluster by species (Figure 4), indicating either post-speciation gene duplications in each of these species or concerted evolution (e.g., due to unequal crossing over or gene conversion) between pseudogene loci (Hess and Edwards 2002; Spurgin et al. 2011). Furthermore, given the extensive evidence of a high number of pseudogenes in multiple passerine species and the relatively recent and rapid divergence of the Passeriformes order (about 60 to 55 million years ago; Ericson et al. 2002), these loci seem to be the result of ancestral gene duplication events.

Phylogenetic results of representative MHC class I sequences of the lark sparrow and published avian species provided some insight into the relationships between MHC class I loci of avian species and the long term evolutionary processes influencing these genes. Within the passerine clade, relationships of lark sparrow functional MHC class I haplotypes to other available passerine MHC sequences suggest that multiple gene duplications have preceded speciation events (Figure 4). Clustering of lark sparrow loci with other passerine species may be representative of orthologous relationships of distinct MHC loci. For

example, lark sparrow locus A and all lark sparrow pseudogenes clustered with the common yellowthroat pseudogenes and functional genes. This relationship may have been caused by a pre-speciation gene duplication event that resulted in these pseudogenes. A similar relationship is also evident between lark sparrow loci C and D and a cluster of house sparrow MHC class I sequences. Interestingly, these two loci contain a PBR codon that is typically present in other MHC class I characterizations but absent in all other lark sparrow MHC class I sequences. Lark sparrow loci E and F clustered independently and locus B did not show a significant clustering with any sequence, although further study of passerine MHC surely has the potential to reveal orthology with other passerine loci.

Analysis of MHC class I sequences from species of other avian orders included in this study provide examples of differential selective pressures and evolutionary processes acting on MHC genes. Galliformes, with its detailed characterizations and minimal-essential MHC, revealed evidence of concerted evolution by the tendency of presumably homologous loci to cluster by species (Figure 3). The two common loci of galliform species (BF1/IA1 and BF2/IA2) have similar gene organization across species and are most likely the result of an ancestral gene duplication, yet within a species the two paralogous genes have evolved to be most similar (Hess and Edwards 2002). Similar phylogenetic effects could also be the result of post-speciation gene duplication within each species, but this seems unlikely due to previously established similarities in gene organizations across multiple species (Wang et al. 2012; Eimes et al. 2013). Phylogenetic results of the Anseriformes MHC class I sequences, on the other hand, may be indicative of divergent evolution of multiple loci. Clustering of the mallard UAA locus with two MHC sequences identified in a bar-headed goose appear to be orthologous. Likewise, grouping of the mallard UDA and pseudogene UBA loci with the remaining bar-headed goose sequence and the only known MHC class I locus in the domestic goose may also be orthologous.

Further evidence of divergent evolution of MHC class I loci is found with the grouping of the two red-billed gull loci with sequences from different orders. The UAA locus clustered with representatives of the Procellariiformes and Cathartiformes, although bootstrap support was weak. The UBA locus grouped with the only representative sequence of the Florida sandhill crane with 63% bootstrap support,

which suggests that the Florida sandhill crane MHC locus is orthologous to the red-billed gull UBA locus, although data in this instance is severely limited.

Results from the avian MHC phylogeny are consistent with those reported in a recent study, which characterized the MHC class I of 16 passerine species (Alcaide et al. 2013). This study found extensive evidence of positive selection acting upon the PBR codons and, furthermore, found those effects to be strongest within the passerines in comparison to other orders (Alcaide et al. 2013). The researchers also determined that, based upon nucleotide composition, gene conversion may be having a slightly greater effect on nonpasserine species. Overall, the present study, as well as that of Alcaide et al. (2013), reinforces the need for further studies of non-model avian species in order to reveal a clearer picture of avian MHC evolution.

In summary, the characterization of MHC class I loci of the lark sparrow has revealed evidence of extensive gene duplication events, differential selection (including both positive and negative selection) upon PBR and non-PBR codons, and high levels of genetic diversity. In addition, the phylogenetic analysis of lark sparrow MHC haplotypes and other published avian MHC class I sequences provided insights into the evolutionary processes driving gene duplication, divergence of MHC class I genes, and concerted evolution of MHC sequences across avian taxa. This study provides, therefore, the first characterization of MHC class I genes in the lark sparrow and a solid basis for the continued investigation of the evolutionary processes driving MHC evolution in birds.

Table 1. Characterization of MHC class I sequences of the lark sparrow (*Chondestes grammacus*). DNA sequencing results (number of clones sequenced, number of alleles, and the minimum number of functional loci estimated) and genetic diversity estimates (nucleotide diversity, average number of nucleotide differences, and number of polymorphic sites) of MHC class I exon 3 haplotypes sampled from four adult lark sparrows.

Individual	No. of Clones Sequenced	No. of Alleles	Minimum No. of Loci	Nucleotide Diversity (π)	Avg. No. of Nucleotide Diff. (k)	No. of Polymorphic Sites (s)
OH007	10	2		0.010	4.0	4.0
Functional		0	-	-	-	-
Pseudogene		2		0.010	4.0	4.0
OH043	29	9		0.120	45.8	107
Functional		7	4	0.097	37.0	88
Pseudogene		2		0.003	1.0	1
TX01	44	7		0.098	37.5	101
Functional		2	1	0.165	63.0	63
Pseudogene		5		0.009	3.4	7
TX02	20	6		0.126	48.5	105
Functional		2	1	0.119	46.0	46
Pseudogene		4		0.065	23.3	46
Total:	103	21		0.126	48.1	144
Functional		11		0.115	43.9	113
Pseudogene		10		0.032	12.4	53

Table 2. Comparison of genetic diversity estimates between functional MHC class I alleles and pseudogenes sampled from four lark sparrow adults. “Exon” and “intron” assignments for pseudogenes are included for comparison only.

	No. of Sequences	Base Pair Length	Nucleotide Diversity (π)	Avg. No. of Nucleotide Diff. (k)	No. of Polymorphic Sites (s)	d_N/d_S (ω)
Functional	11	388	0.115	43.9	113	
Exon		276	0.113	30.9	78	0.619
(PBR)		42	0.284	11.1	24	1.127
(non-PBR)		234	0.085	19.8	54	0.483
Intron		112	0.120	13.1	35	
Pseudogene	10	390	0.032	12.4	53	
Exon		278	0.035	9.6	39	
Intron		112	0.025	2.8	14	
All Sequences	21	393	0.126	48.1	144	
Exon		281	0.140	38.3	105	
Intron		112	0.091	9.9	39	

Table 3. Avian MHC class I sequences reported to GenBank included in this study. Only studies with at least partial exon 3 sequences are included. The minimum number of MHC class I loci reported is based on an estimated number of functional sequences. GenBank accession numbers of sequences included in the phylogenetic analysis are listed.

Order/Species		Family	Minimum No. of Loci	Evidence of Pseudogenes	MHC Typing Method	Ref	Accession No.
Galliformes							
Chicken	<i>Gallus gallus</i>	Phasianidae	2	-	gDNA, cloning, sequencing	1	AB268588
Turkey	<i>Meleagris gallopavo</i>	"	2	-	gDNA, cloning, sequencing	2	DQ993255
Japanese Quail	<i>Coturnix japonica</i>	"	4	Y	gDNA, cloning, sequencing	3	AB078884
Greater Prairie Chicken	<i>Tympanuchus cupido</i>	"	1	-	gDNA, cloning, sequencing	4	JX237356
Black Grouse	<i>Tetrao tetrix</i>	"	2	-	gDNA, cloning, sequencing	5	JQ028669
Golden Pheasant	<i>Chrysolophus pictus</i>	"	2	-	gDNA, cloning, sequencing	6	JQ440366
Helmeted Guineafowl	<i>Numida meleagris</i>	Numididae	≥1	N/A	cDNA, cloning, sequencing	7	EU430728
Anseriformes							
Mallard	<i>Anas platyrhynchos</i>	Anatidae	2	Y	gDNA, cloning, sequencing	8	AY885227
Domestic Goose	<i>Anser anser</i>	"	1	N/A	cDNA, cloning, sequencing	9	AY387649
Bar-headed Goose	<i>Anser indicus</i>	"	≥2	-	gDNA, cloning, sequencing	10	FJ606110-11, FJ606113
Gruiformes							
Florida Sandhill Crane	<i>Grus canadensis pratensis</i>	Gruidae	≥2	-	gDNA, cloning, sequencing	11	AF033106
Procellariiformes							
Blue Petrel	<i>Halobaena caerulea</i>	Procellariidae	≥5	Y	gDNA, cloning, sequencing	12	JF276835, JF276870
Charadriiformes							
Red-billed Gull	<i>Larus scopulinus</i>	Laridae	≥2	-	gDNA, cloning, sequencing	13	HM025950, HM025973
Strigiformes							
*Eurasian Eagle-owl	<i>Bubo bubo</i>	Strigidae	unknown		gDNA, cloning, sequencing	14	
Cathartiformes							
Andean Condor	<i>Vultur gryphus</i>	Cathartidae	≥2	-	gDNA, cloning, sequencing	14	EU120692-93
Accipitriformes							
Black-shouldered Kite	<i>Elanus caeruleus</i>	Accipitridae	≥2	-	gDNA, cloning, sequencing	14	EU120680, EU120682
*White-backed Vulture	<i>Gyps africanus</i>	"	≥2	-	gDNA, cloning, sequencing	14	
Eurasian Black Vulture	<i>Aegypius monachus</i>	"	≥2	-	gDNA, cloning, sequencing	14	EU120688, EU120691
Spanish Imperial Eagle	<i>Aquila adalberti</i>	"	≥2	-	gDNA, cloning, sequencing	14	EU120684, EU120687
*Bearded Vulture	<i>Gypaetus barbatus</i>	"	≥2	-	gDNA, cloning, sequencing	14	

Table 3 continued.

Order/Species		Family	Minimum No. of Loci	Evidence of Pseudogenes	MHC Typing Method	Ref	Accession No.
Falconiformes							
Lesser Kestrel	<i>Falco naumanni</i>	Falconidae	≥1	-	gDNA, cloning, sequencing	14	EU120665, EU120671
Common Kestrel	<i>Falco tinnunculus</i>	"	≥1	-	gDNA, cloning, sequencing	14	EU120719, EU120722
*Mauritius Kestrel	<i>Falco punctatus</i>	"	≥1	-	gDNA, cloning, sequencing	15	
Peregrine Falcon	<i>Falco peregrines</i>	"	≥1	-	gDNA, cloning, sequencing	15	JN613264, JN613271
Eleonora's Falcon	<i>Falco eleonora</i>	"	≥1	-	gDNA, cloning, sequencing	15	JN613263, JN613265
*Gyr Falcon	<i>Falco rusticolus</i>	"	≥1	-	gDNA, cloning, sequencing	15	
Passeriformes							
Great Reed Warbler	<i>Acrocephalus arundinaceus</i>	Acrocephalidae	≥6	Y	gDNA, cloning, DGGE, sequencing	16	AF449689, AF449691-93, AF449697-701
Seychelles Warbler	<i>Acrocephalus sechellensis</i>	"	≥3	-	gDNA, DGGE, sequencing	17	AJ557874, AJ557882
Common Yellowthroat	<i>Geothlypis trichas</i>	Parulidae	≥8	Y	gDNA, 454 pyrosequencing	18	JQ859962, JX213824, JX213848, JX213853
*Zebra Finch	<i>Taeniopygia guttata</i>	Estrilididae	1	Y	gDNA, BAC library scanning	19	
Common Rosefinch	<i>Carpodacus erythrinus</i>	Fringillidae	≥5	-	gDNA, SSCP, sequencing	20	FJ392763, FJ392767
Berthelot's Pipit	<i>Anthus berthelotii</i>	Motacillidae	≥4	-	gDNA, 454 pyrosequencing	21	JN799601, JN799625, JN799641
House Sparrow	<i>Passer domesticus</i>	Passeridae	≥4	Y	gDNA, cloning, sequencing	22	JN609623, JN609630, JN609637-38, JN609346-47
Lark Sparrow	<i>Chondestes grammacus</i>	Emberizidae	≥4	Y	gDNA, cloning, sequencing		This study (see Appendix)
Great Tit	<i>Parus major</i>	Paridae	≥16	Y	gDNA, 454 pyrosequencing	23	JQ035444, JQ035449, JQ035461, JQ035485
Green-backed Tit	<i>Parus monticolus</i>	"	unknown		gDNA, (unknown)	n.p.	EF446974-75
Blue Tit	<i>Cyanistes caeruleus</i>	"	10	Y	gDNA, cloning, sequencing	n.p.	AM232717
						24	HQ393911-12, HQ393950

*Excluded from phylogenies due to a limited number of reported sequences

Ref: 1. Shiina et al. 2007; 2. Chaves et al. 2009; 3. Shiina et al. 2004; 4. Eimes et al. 2013; 5. Wang et al. 2012; 6. Ye et al. 2012; 7. Singh et al. 2010; 8. Moon et al. 2005; 9. Xia et al. 2005; 10. Liang et al. 2010; 11. Jarvi et al. 1999; 12. Strandh et al. 2011; 13. Cloutier et al. 2011; 14. Alcaide et al. 2009; 15. Gangoso et al. 2012; 16. Westerdahl et al. 2004; 17. Richardson & Westerdahl 2003; 18. Bollmer et al. 2012; 19. Balakrishnan et al. 2010; 20. Promerová et al. 2009; 21. Spurgin et al. 2011; 22. Borg et al. 2011; 23. Sepil et al. 2012; 24. Wutzler et al. 2012; n.p. Not Published

Figure 1. Hypothetical phylogenies based on divergent and concerted evolution of avian MHC loci.

Phylogenies represent hypothetical gene trees of MHC sequences where shaded circles indicate three distinct species and numbers represent loci. (A) Expected phylogenetic effects of divergent evolution of MHC loci originating prior to speciation. Loci group regardless of species. (B) Expected phylogenetic effects of concerted evolution of MHC loci. Species group regardless of loci. Gene conversion has occurred between loci that originated prior to speciation. Alternatively, these phylogenetic effects may also be indicative of gene duplication events after speciation. Modified from Hess and Edwards (2002).

A.

B.

Figure 2. Unrooted maximum likelihood phylogeny of lark sparrow MHC class I sequences spanning exon 3 and extending into flanking introns 2 and 3 (396 bp). Numbers at branching nodes represent bootstrap values greater than 50% (based on 1000 replicates). Pseudogenes are designated with the Greek letter psi (Ψ). Shaded circles represent each of the four lark sparrow individuals in which sequences were found. Letters A to F represent potential functional loci.

Figure 3. Maximum likelihood phylogeny of Aves MHC class I exon 3 DNA sequences (235 bp) retrieved from GenBank. Numbers at branching nodes represent bootstrap values greater than 50% (based on 1000 replicates). The Passeriformes clade is expanded in Figure 4. A saltwater crocodile (*Crocodylus porosus*) MHC class I DNA sequence (GenBank accession number HQ008787) was included as the outgroup. Pseudogenes are designated with the Greek letter psi (Ψ). GenBank accession numbers for each sequence are listed in Table 3.

Figure 4. Maximum likelihood phylogeny of passerine MHC class I exon 3 DNA sequences retrieved from GenBank. Numbers at branching nodes represent bootstrap values greater than 50% (based on 1000 replicates). Functional lark sparrow sequences reported in this study are labeled based upon assigned loci from Figure 2. Pseudogenes are designated with the Greek letter psi (Ψ). GenBank accession numbers for each sequence are listed in Table 3.

REFERENCES

- Abbas, A.K., Lichtman, A.H., and Pober, J.S. 1994. Cellular and molecular immunology. WB Saunders, Philadelphia, P.A.
- Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D., and Wayne, R.K. 2004. High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *PNAS* **101**(10): 3490-3494.
- Alcaide, M., Edwards, S.V., Cadahía, L., and Negro, J.J. 2009. MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. *Conserv. Genet.* **10**(5): 1349-1355.
- Alcaide, M., Liu, M., and Edwards, S.V. 2013. Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. *PeerJ* **1**: e86. Available from <http://dx.doi.org/10.7717/peerj.86> [accessed 23 August 2013].
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**(3): 403-410.
- Apanius, V., Penn, D., Slev, P.R., Ruff, L.R., and Potts, W.K. 1997. The nature of selection on the major histocompatibility complex. *Crit. Rev. Immunol.* **17**(2): 179-224.
- Baepler, D.H. 1968. Lark Sparrow. *In* Life histories of North American cardinals, grosbeaks, buntings, towhees, finches, sparrows, and allies. *Edited by* A.C. Bent. U.S. National Museum Bulletin 237. pp. 886-902.
- Balakrishnan, C.N., Ekblom, R., Völker, M., Westerdahl, H., Godinez, R., Kotkiewicz, H., Burt, D.W., Graves, T., Griffin, D.K., Warren, W.C., and Edwards, S.V. 2010. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biology* **8**: 29. Available from <http://www.biomedcentral.com/1741-7007/8/29> [accessed 8 February 2013].
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. 2005. GenBank. *Nucl. Acids Res.* **33**: D34-D38.
- Björkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., and Wiley, D.C. 1987. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens.

- Nature **329**: 512-518.
- Bollmer, J.L., Vargas, F.H., and Parker, P.G. 2007. Low MHC variation in the endangered Galápagos penguin (*Spheniscus mendiculus*). Immunogenetics **59**(7): 593-602.
- Bollmer, J.L., Dunn, P.O., Freeman-Gallant, C.R., and Whittingham, L.A. 2012. Social and extra-pair mating in relation to major histocompatibility complex variation in common yellowthroats. Proc. R. Soc. B **279**(1748): 4778-4785.
- Bonneaud, C., Sorci, G., Morin, V., Westerdahl, H., Zoorob, R., and Wittzell, H. 2004. Diversity of MHC class I and IIB genes in house sparrows (*Passer domesticus*). Immunogenetics **55**(12): 855-865.
- Borg, Å.A., Pedersen, S.A., Jensen, H., and Westerdahl, H. 2011. Variation in MHC genotypes in two populations of house sparrow (*Passer domesticus*) with different population histories. Ecol. Evol. **1**(2): 145-159.
- Borghans, J.A.M., Joost, B.B., and De Boer, R.J. 2004. MHC polymorphism under host-pathogen coevolution. Immunogenetics **55**(11): 732-739.
- Chaves, L.D., Krueth, S.B., and Reed, K.M. 2009. Defining the turkey MHC: sequence and genes of the B locus. J. Immunol. **183**(10): 6530-6537.
- Cloutier, A., Mills, J.A., and Baker, A.J. 2011. Characterization and locus-specific typing of MHC class I gene in the red-billed gull (*Larus scopulinus*) provides evidence for major, minor, and nonclassical loci. Immunogenetics **63**(6): 377-394.
- Dunn, P.O., Bollmer, J.L., Freeman-Gallant, C.R., and Whittingham, L.A. 2013. MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats. Evolution **67**(3): 679-687.
- Edwards, S.V., and Hedrick, P.W. 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. TREE **13**(8): 305-311.
- Eimes, J.A., Reed, K.M., Mendoza, K.M., Bollmer, J.L., Whittingham, L.A., Bateson, Z.W., and Dunn, P.O. 2013. Greater prairie chickens have a compact *MHC-B* with a single class IA locus. Immunogenetics **65**(2): 133-144.

- Ericson, P.G.P., Christidis, L., Cooper, A., Irestedt, M., Jackson, J., Johansson, U.S., and Norman, J.A. 2002. A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. *Proc. R. Soc. Lond. B* **269**(1488): 235-241.
- Gangoso, L., Alcaide, M., Grande, J.M., Muñoz, J., Talbot, S.L., Sonsthagen, S.A., Sage, G.K., and Figuerola, J. 2012. Colonizing the world in spite of reduced MHC variation. *J. Evol. Biol.* **25**(7): 1438-1447.
- Garrigan, D., and Hedrick, P.W. 2001. Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. *Immunogenetics* **53**(6): 483-489.
- Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acid Symp. Ser.* **41**: 95-98.
- Hamilton, W.D., and Zuk, M. 1982. Heritable true fitness and bright birds: a role for parasites? *Science* **218**(4570): 384-387.
- Hedrick, P.W. 2002. Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**(10): 1902-1908.
- Hess, C.M., and Edwards, S.V. 2002. The evolution of the major histocompatibility complex in birds. *BioScience* **52**(5): 423-431.
- Högstrand, K., and Böhme, J. 1999. Gene conversion can create new MHC alleles. *Immunol. Rev.* **167**(1): 305-317.
- Hughes, A.L., and Nei, M. 1988. Pattern of nucleotide substitution at the major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167-170.
- Hughes, A.L., and Yeager, M. 1998. Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* **32**: 415-435.
- Jaratlertsiri, W., Isberg, S.R., Higgins, D.P., and Gongora, J. 2012. MHC class I of saltwater crocodiles (*Crocodylus porosus*): polymorphism and balancing selection. *Immunogenetics* **64**(11): 825-838.
- Jarvi, S.I., Goto, R.M., Gee, G.F., Briles, W.E., and Miller, M.M. 1999. Identification, inheritance, and linkage of *B-G*-like and MHC class I genes in cranes. *J. Hered.* **90**(1): 152-159.

- Kaufman, J., Völk, H., and Wallny, H.J. 1995. A “minimal essential MHC” and an “unrecognized MHC”: two extremes in selection for polymorphism. *Immunol. Rev.* **143**(1): 63-88.
- Kelley, J., Walter, L., and Trowsdale, J. 2005. Comparative genomics of major histocompatibility complexes. *Immunogenetics* **56**(10): 683-695.
- Klein, J. 1986. Natural history of the major histocompatibility complex. Wiley, New York City, N.Y.
- Klein, J., Bontrop, R.E., Dawkins, R.L., Erlich, H.A., Gyllenstein, U.B., Heise, E.R., Jones, P.P., Parham, P., Wakeland, E.K., and Watkins, D.I. 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* **31**: 217-219.
- Knafler, G.J., Clark, J.A., Boersma, P.D., and Bouzat, J.L. 2012. MHC diversity and mate choice in the Magellanic penguin *Spheniscus magellanicus*. *J. Hered.* **103**(6): 759-768.
- Kulski, J.K., Shiina, T., Anzai, T., Kohara, S., and Inoko, H. 2002. Comparative genomic analysis of the MHC: evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunol. Rev.* **190**(1): 95-112.
- Landry, C., Garant, D., Duchesne, P., and Bernatchez, L. 2001. ‘Good genes as heterozygosity’: the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proc. R. Soc. Lond. B* **268**(1473): 1279-1285.
- Langefors, Å., Lohm, J., Grahn, M., Andersen, Ø., and von Schantz, T. 2001. Association between major histocompatibility complex IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc. R. Soc. Lond. B* **268**(1466): 479-485.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21): 2947-2748.
- Liang, Q., Wei, L., Wang, X., and He, H. 2010. MHC class I loci of the bar-headed goose (*Anser indicus*). *Genet. Mol. Biol.* **33**(3): 573-577.
- Librado, P., and Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**(11): 1451-1452.

- Longmire, J.L., Lewis, A., Meincke, L.J., and Hildebrand, C.E. 1986. A new and rapid procedure for the isolation of ultra-high molecular weight eukaryotic DNA. *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **45**: 6.
- Martin, J.W., and Parrish, J.R. 2000. Lark sparrow (*Chondestes grammacus*). In *The Birds of North America*, No. 488. *Edited by* A. Poole and F. Gill. Academy of Natural Sciences, Philadelphia, P.A., and American Ornithologists' Union, Washington, D.C.
- Miller, H.C., Allendorf, F., and Daugherty, C.H. 2010. Genetic diversity and differentiation at MHC genes in island populations of tuatara (*Sphenodon spp.*). *Mol. Ecol.* **19**(18): 3894-3908.
- Miller, H.C., Bowker-Wright, G., Kharkrang, M., and Ramstad, K. 2011. Characterisation of class II B MHC genes from a ratite bird, the little spotted kiwi (*Apteryx owenii*). *Immunogenetics* **63**(4): 223-233.
- Moon, D.A., Veniamin, S.M., Parks-Dely, J.A., and Magor, K.E. 2005. The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed class I genes. *J. Immunol.* **175**(10): 6702-6712.
- Nei, M., Gu, X., and Sitnikova, T. 1997. Evolution by birth-and-death process in multigene families of the vertebrate immune system. *PNAS USA* **94**(15): 7799-7806.
- Ohio Department of Natural Resources – Division of Wildlife. 2012. Wildlife that are considered to be endangered, threatened, species of concern, special interest, extirpated, or extinct in Ohio. Publication 5356.
- Parham, P., and Ohta, T. 1996. Population biology of antigen presentation by MHC class I molecules. *Science* **272**(5258): 67-74.
- Paterson, S., and Pemberton, J.M. 1997. No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population. *Proc. R. Soc. Lond. B* **264**(1389): 1813-1819.
- Penn, D.J., Damjanovich, K., and Potts, W.K. 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. *PNAS USA* **99**(17): 11260-11264.

- Promerová, M., Albrecht, T., and Bryja, J. 2009. Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). *Immunogenetics* **61**(6): 451-461.
- Richardson, D.S., and Westerdahl, H. 2003. MHC diversity in two *Acrocephalus* species: the outbred Great reed warbler and the inbred Seychelles warbler. *Mol. Ecol.* **12**(12): 3523-3529.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Sepil, I., Moghadam, H.K., Huchard, E., and Sheldon, B.C. 2012. Characterization and 454 pyrosequencing of major histocompatibility complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evol. Biol.* **12**: 68. Available from <http://www.biomedcentral.com/1471-2148/12/6> [accessed 8 February 2013].
- Shiina, T., Shimizu, S., Hosomichi, K., Kohara, S., Watanabe, S., Hanzawa, K., Beck, S., Kulski, J.K., and Inoko, H. 2004. Comparative genomic analysis of two avian (quail and chicken) MHC regions. *J. Immunol.* **172**(11): 6751-6763.
- Shiina, T., Briles, W.E., Goto, R.M., Hosomichi, K., Yanagiya, K., Shimizu, S., Inoko, H., and Miller, M.M. 2007. Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type gene with a subregion of the chicken MHC-B affecting infectious disease. *J. Immunol.* **178**(11): 7162-7172.
- Singh, S.K., Mehra, S., Shukla, S.K., Kumar, V., Tiwari, A., Mehra, M., Goyal, G., Mathew, J., and Sharma, D. 2010. Nucleotide sequence variation in MHC class I region in guinea fowl. *Int. J. Poult. Sci.* **9**(3): 236-239.
- Spurgin, L.G., van Oosterhout, C., Illera, J.C., Bridgett, S., Gharbi, K., Emerson, B.C., and Richardson, D.S. 2011. Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Mol. Ecol.* **20**(24): 5213-5225.
- Strandh, M., Lannefors, M., Bonadonna, F., and Westerdahl, H. 2011. Characterization of MHC class I and II genes in a subantarctic seabird, the blue petrel, *Halobaena caerulea* (Procellariiformes).

- Immunogenetics **63**(10): 653-666.
- Swanson, B.L. 2012. Genetic characterization of avian malaria parasites across the breeding range of the migratory lark sparrow (*Chondestes grammacus*). M.Sc. thesis, Department of Biological Sciences, Bowling Green State University, Bowling Green, O.H.
- Takahata, N., and Nei, M. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**(4): 967-978.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**(10): 2731-2739.
- Tregenza, T., and Wedell, N. 2000. Genetic compatibility, mate choice and patterns of parentage: invited review. *Mol. Ecol.* **9**(8): 1013-1027.
- Tsuda, T.T., Tsuda, M., Naruse, T., Kawata, H., Ando, A., Shiina, T., Fukuda, M., Kurita, M., LeMaho, I., Kulski, J.K., and Inoko, H. 2001. Phylogenetic analysis of penguin (*Spheniscidae*) species based on sequence variation in MHC class II genes. *Immunogenetics* **53**(8): 712-716.
- van Oosterhout, C., Joyce, D.A., Cummings, S.M., Blais, J., Barson, N.J., Ramnarine, I.W., Mohammed, R.S., Persad, N., and Cable, J. 2006. Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution* **60**(12): 2562-2574.
- von Schantz, T., Wittzell, H., Goransson, G., Grahn, M., and Persson, K. 1996. MHC genotype and male ornamentation: genetic evidence for the Hamilton-Zuk model. *Proc. R. Soc. Lond. B* **263**(1368): 265-271.
- Walsh, J.B., and Stephan, W. 2001. Multigene Families: Evolution. *In* Encyclopedia of Life Sciences.
- Wang, B., Ekblom, R., Strand, T.M., Portela-Bens, S., and Höglund, J. 2012. Sequencing of the core MHC region of black grouse (*Tetrao tetrix*) and comparative genomics of the galliform MHC. *BMC Genomics* **13**: 553. Available from <http://www.biomedcentral.com/1471-2164/13/553> [accessed 8 February 2013].

- Westerdahl, H. 2007. Passerine MHC: genetic variation and disease resistance in the wild. *J. Ornithol.* **148**(2): S469-S477.
- Westerdahl, H., Wittzell, H., von Schantz, T., and 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.
- Wutzler, R., Foerster, K., and Kempenaers, B. 2012. MHC class I variation in a natural blue tit population (*Cyanistes caeruleus*). *Genetica* **140**(7-9): 349-364.
- Xia, C., Hu, T., Yang, T., Wang, L., Xu, G., and Lin, C. 2005. cDNA cloning, genomic structure and expression analysis of the goose (*Anser cygnoides*) MHC class I gene. *Vet. Immunol. Immunopathol.* **107**(3-4): 291-302.
- Ye, Q., He, K., Wu, S.Y., and Wan Q.H. 2012. Isolation of a 97-kb minimal essential MHC B locus from a new reverse-4D BAC library of the golden pheasant. *PLoS ONE* **7**(3): e32154.

APPENDIX

Lark sparrow MHC class I haplotypes characterized in this study

Table A1. List of the 21 MHC class I haplotypes of the lark sparrow (*Chondestes grammacus*) characterized in this study. GenBank accession numbers and descriptions are included. Haplotypes were named according to Klein et al. (1990), with the first two letters of the genus and species name and the assigned haplotype number.

Haplotype Name	GenBank Accession No.	Sequence Description
Chgr001	KF803763	Chondestes grammacus MHC class I antigen pseudogene, Chgr001 haplotype, exon 3, partial sequence
Chgr002	KF803764	Chondestes grammacus MHC class I antigen pseudogene, Chgr002 haplotype, exon 3, partial sequence
Chgr003	KF803765	Chondestes grammacus MHC class I antigen gene, Chgr003 haplotype, exon 3, partial cds
Chgr004	KF803766	Chondestes grammacus MHC class I antigen pseudogene, Chgr004 haplotype, exon 3, partial sequence
Chgr005	KF803767	Chondestes grammacus MHC class I antigen pseudogene, Chgr005 haplotype, exon 3, partial sequence
Chgr012	KF803768	Chondestes grammacus MHC class I antigen gene, Chgr012 haplotype, exon 3, partial cds
Chgr014	KF803769	Chondestes grammacus MHC class I antigen gene, Chgr014 haplotype, exon 3, partial cds
Chgr015	KF803770	Chondestes grammacus MHC class I antigen gene, Chgr015 haplotype, exon 3, partial cds
Chgr017	KF803771	Chondestes grammacus MHC class I antigen gene, Chgr017 haplotype, exon 3, partial cds
Chgr018	KF803772	Chondestes grammacus MHC class I antigen gene, Chgr018 haplotype, exon 3, partial cds
Chgr019	KF803773	Chondestes grammacus MHC class I antigen gene, Chgr019 haplotype, exon 3, partial cds
Chgr020	KF803774	Chondestes grammacus MHC class I antigen pseudogene, Chgr020 haplotype, exon 3, partial sequence
Chgr021	KF803775	Chondestes grammacus MHC class I antigen pseudogene, Chgr021 haplotype, exon 3, partial sequence
Chgr026	KF803776	Chondestes grammacus MHC class I antigen gene, Chgr026 haplotype, exon 3, partial cds
Chgr033	KF803777	Chondestes grammacus MHC class I antigen gene, Chgr033 haplotype, exon 3, partial cds
Chgr034	KF803778	Chondestes grammacus MHC class I antigen pseudogene, Chgr034 haplotype, exon 3, partial sequence
Chgr036	KF803779	Chondestes grammacus MHC class I antigen gene, Chgr036 haplotype, exon 3, partial cds
Chgr037	KF803780	Chondestes grammacus MHC class I antigen pseudogene, Chgr037 haplotype, exon 3, partial sequence
Chgr039	KF803781	Chondestes grammacus MHC class I antigen gene, Chgr039 haplotype, exon 3, partial cds
Chgr041	KF803782	Chondestes grammacus MHC class I antigen pseudogene, Chgr041 haplotype, exon 3, partial sequence
Chgr045	KF803783	Chondestes grammacus MHC class I antigen pseudogene, Chgr045 haplotype, exon 3, partial sequence

Figure A1. DNA sequence alignment of the 21 MHC class I lark sparrow (*Chondestes grammacus*) haplotypes characterized in this study.

Variable positions are relative to the reference haplotype *Chgr018* (top). Dots represent nucleotides that are identical to *Chgr018*. Shaded regions of the alignment represent the introns flanking exon 3. Functional genes are labeled A-F to represent their assigned loci from Figure 2. PBR codons (14 total) are designated with solid line rectangles and three filled triangles below the alignment. Pseudogenes are designated with the Greek letter psi (Ψ) and surrounded by the dashed line rectangle.

Figure A1. continued

[illegible]

Figure A1. continued

					350						
Chgr018	A	ATTCTGTGG	GAATTTGGGA	TTTGGGAATG	GAGGACTTGG	CAACACAGTA	ATTGCCGTGG	GAATTCCTATG	GGGTAGATAT	TACCTCCACC	TCA
Chgr017	A								C.		...
Chgr026	A								C.		...
Chgr034Ψ								T	C.		...
Chgr001Ψ			.G.....			G.....	...C.T...		...-...C.	.G..C..GT.	C..
Chgr037Ψ			.G.C....			G.....	...C.T...		...-...C.	.G..C..GT.	C..
Chgr045Ψ			.G.....		.A.....	G.....	...C.T...		...-...C.	.G..C..GT.	C..
Chgr021Ψ			.G.....			G.....	...C.T...		...-...C.	.G..C..GT.	C..
Chgr041Ψ			.G.....		C.	G.....	...C.T...		...-...C.	.G..C..GT.	C..
Chgr020Ψ			.G.....			G.....	...C.T...		...-...C.	.G..C..GT.	C..
Chgr002Ψ			.G.....			G.....	...C.T...		...-...C.	.G..C..GT.	C..
Chgr004Ψ			.G.....			G.....	...C.T...		...-...C.	.G..C..GT.	C..
Chgr005Ψ			.G.....			G.....	...C.T...		...-...C.	.G..C..GT.	C..
Chgr019	B		.G.....	T....	A.....T....	G.....G....	T.T...	C.A.	...-...C.	CGTT.....T.	C..
Chgr036	C		.G.....	T....	...A.....	G.....TG...	T....	C....	...-...C.	C..T.T..T.	C..
Chgr039	D		.G.....	...C..G...	C.A	G.....TGTC.	...C.TA...	C..T	...-...C.	C.....T.	C..
Chgr033	D		.G.....A.	...C..G...	C.A	G.....TGTC.	...C.TA...	C..T	...-...C.	C.....T.	C..
Chgr015	E								C.		...
Chgr014	E		.G.....	G...		G.....	...C.T...	C...	C.		...
Chgr003	F		.G.....	G...		G.....	...C.T...		...-...C.	.G..C..GT.	C..
Chgr012	F		.G.....	G...		G.....	...C.T...	C...	...-...C.	.G..C..GT.	C..