B-G Gene Structure, Genetic Variation and Expression in the Turkey (*Meleagris gallopavo*) Major Histocompatibility Complex

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Miranda Marie Bauer

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Kent M. Reed

February 2012

© Miranda Marie Bauer 2012

Acknowledgements

Special thanks go out to my friends and lab members I've met during my graduate school career. I am particularly grateful to my advisor, Dr. Kent Reed for giving me the opportunity to study the genetics of the turkey major histocompatibility complex with him. He provided steady belief in my abilities throughout my project as well as offered many insightful additions to my manuscripts and dissertation. Not only have my research, writing and presentation skills benefited from his guidance, but my intrapersonal skills and personal life have been enriched by his friendship as well. I also thank my committee members, Dr. Mark Rutherford, Dr. Michael Murtaugh and Dr. Doug Foster for their encouragement and support of free and critical thinking, which was the cause of much personal and scientific growth.

I thank my lab manager, mentor and friend, Kristelle Mendoza for providing necessary encouragement, advice and technical experimental assistance, especially regarding qRT-PCR. I've learned many important life lessons from this valuable woman. I also want to recognize Melissa Monson, another friend, fellow graduate student and lab member. I will remember many of our conversations, and our discussions on various scientific topics led me to many novel insights and inspirations. Also, I appreciate Cheryl Lancto and Dr. Juan Abrahante, who were instrumental with the qRT-PCR portion of the study.

I would like to recognize the USDA for providing funding for my PhD studies through the National Needs Fellowship, which forced me into an unfamiliar field of study and provided me with a foundation of knowledge for bioinformatic studies.

Dedication

I dedicate my dissertation to two of the most important women in my life. Firstly, my sister, Marissa Bauer, who is ever an inspiration to me. I wish nothing but the best for your future, and I know you will go far. You have been more of a muse for me than you know, and I hope I can enrich your life as you have mine. Secondly, Caprise Adams, who has helped me grow immensely as a person and continuously challenges me to reach past current paradigms of thought and action. You have taught me that unconventional thinking extends far past the science world and is a necessity for forward progress in all fields of study and aspects of life.

Abstract

The major histocompatibility complex (MHC) is a genomic region encoding for genes with various immune functions in vertebrates. Although this region has been studied extensively in humans, mice and agriculturally important species such as cattle, horses and chickens, the organization and functional significance of the turkey MHC is relatively unknown. One set of loci located in the avian MHC, designated as B-G genes, have been identified and sequenced within the *B*-locus of the turkey, chicken and quail, with additional B-G loci identified in the extended MHC of the latter two species. These genes are linked to the class I and class II loci of the MHC and also show regions of extreme polymorphism (Miller, 1984). The total number, function and significance of the B-G genes are yet unknown in any bird species. In turkeys, three B-G genes were previously sequenced within the *B*-locus (Chaves et al., 2009a), with evidence suggesting additional functional B-G genes located past the 5' ends of the sequenced MHC region. Evidence in the chicken shows differential expression of B-G genes in various immune tissues, which suggests potential immune function.

This research used 454 FLX Next Generation Sequencing technology (Roche) for sequencing a bacterial artificial chromosome (BAC) and identified two additional B-G genes located past the 5' end of the core *B*-locus of the turkey. These genes were annotated using *in silico* analysis and show organization similar to those in the chicken and quail. Using this information, sequence variation of the B-G genes was compared in different stocks of turkeys. Because each of the three groups, (commercial, heritage and wild), have been selected for different characteristics, variation within these loci was expected. This experiment found variation among loci within the three groups of turkeys.

Lastly, B-G gene expression was investigated with quantitative real-time PCR using liver tissue of aflatoxin challenged birds. A low level of expression was observed for three of the four BG genes investigated, with *BG5* expression invariant in all individuals across treatment groups. *BG4* expression levels fluctuated within and between groups and a higher level of expression was measured in the lactobacillus + aflatoxin group. This work extended the turkey *B*-locus sequence past the homologous region in the chicken and marks the first examination of sequence variation and gene expression of multiple B-G genes in any species. Sequence variation and expression differences among loci support a hypothesis of distinct functions for these molecules.

Table of Contents

Page

List of Tables

List of Figures

Chapter 1

Literature Review

Major histocompatibility (MHC) gene composition and evolution

The immune system contains many mechanisms to fight invasion by non-self pathogens while minimizing host cellular damage, including both innate and adaptive responses. These responses include the ability to differentiate between self and non-self, induction of inflammatory responses, antigen recognition and presentation and clearance of the pathogen. The Major Histocompatibility Complex (MHC) is of central importance in the adaptive immune system.

There are three major MHC regions that contain different types of molecules common to vertebrate species. These have been reorganized several times throughout vertebrate evolution, but most retain a common set of genes and gene families. Class I molecules are found on the surface of all nucleated cells and present antigens to CD8+ T cells, whereas class II molecules are expressed mainly on immune cells and present antigens to CD4+ T cells. These two classes of molecules are responsible for the alloreactivity of T cells involved in allograft rejection and are associated with active immunity. There are several genes (e.g. *TAP1, TAP2, Tapasin, LMP2*) involved in the pathways for intracellular (e.g. viral, parasitic), and extracellular (bacterial) peptide breakdown located within the MHC region. Class III MHC molecules encode genes involved in innate immunity, such as complement receptors (Thorsby et al., 1997) and some have non-immune functions.

Class I and Class II molecules are further subdivided into two categories, classical and non-classical molecules. Classical molecules are highly polymorphic cell surface

glycoproteins directly related to antigen peptide presentation to T cells. Non-classical molecules are structurally related to their corresponding classical molecule but responsible for functions other than peptide presentation, such as T cell activation and peptide loading onto the MHC molecule (Maenaka et al., 1999). Notably, the non-classical class I CD1 genes, which are unlinked to the MHC, are responsible for lipid antigen presentation. Non-classical class II genes include members of the DMA, DMB and DO gene families, which facilitate peptide loading onto classical class II molecules.

Other well conserved genes in the MHC include framework genes (Ex: TRIM, BTN and Blec families) and have varying functions, some of which have been correlated with immune function and disease susceptibility. Notably, TRIM genes have been associated with tumor development and antiviral properties (van der Aa et al., 2012; Hatakeyama, 2011; Gack, 2011; Uchil et al., 2008) along with innate immune responses, including modulation of pattern recognition receptor (PRR) signaling pathways through ubiquitylation events (Kawai and Akira, 2011; Ozato et al., 2008). Butyrophilin genes share structural homology to members of the B7 family, are members of the Ig superfamily and encode membrane proteins. Increasingly well characterized are the immunosuppressive effects through inhibition of T cell activation and cytokine secretion, which appear to inhibit inflammation (Smith et al., 2010; Yamashiro et al 2010). Recent studies show many micro RNAs distributed across certain mammalian MHC regions, which is presumably related to gene expression regulation (Gao et al., 2010). Their effect on MHC genes is not yet described.

The MHC loci of some avian species have been found to contain an additional set of genes not found outside of the class Aves. These genes, classified as B-G (class IV), are disulfide linked cell surface receptors originally described as erythroid markers. The function of these molecules is not yet described, but they share some structural similarities with a complex of natural killer (NK) genes, the chicken leukocyte receptor complex family (CHIR), such as an N-terminal signal peptide, an IgV-like region, a transmembrane domain, cytoplasmic tails of variable length and the occasional presence of immunoreceptor tyrosine-based inhibition motifs (ITIMS) in the exon encoding the C terminus (Goto et al., 2009; Kaufman et al., 1989, 1990; Laun et al., 2006). B-G molecules are extremely variable and immunogenic (Miller, 1984, 1991). The genes in this family are not glycosylated but are linked with other MHC Class I and II loci (Simonsen et al., 1980; Salomonsen et al., 1987; Morgan et al., 1990). Monoclonal antibodies have detected B-G homodimers and heterodimers in specific tissues and blood. Although these studies suggest an important immune function of B-G genes, very little is known about the actual function of members of this gene family.

The region encoding the MHC is the most polymorphic, gene dense region of the genome. High sequence variation (SNPs and insertions/deletions) between individuals, populations and species are well documented (Ho et al., 2009; Kumanovis et al., 2003; Trahene et al., 2006). The classical class I and class II genes show particularly high degrees of polymorphism, although high levels of allelic variation have also been documented in other MHC genes (Ballingall et al., 2010; Rogers and Kaufman, 2008; Marusina et al., 1997; Sena et al., 2011; Sironi et al., 2008; Tanaka-Matsuda et al., 2009; Walker et al., 2011). This appears to be necessary to combat the extreme diversity and

mutation rates of infectious agents. The sequence composition of the MHC is such that stretches of highly polymorphic sequence are dispersed between sequence that is highly conserved in individuals between and among species. Genes in the MHC are in strong linkage disequilibrium, which promotes co-evolution and increases the complexity of identifying single genes contributing to host immunity. Identifying specific loci associated with disease phenotypes is further complicated by the presence of gene families, some of which have highly conserved sequence motifs within and across species. Specific haplotypes in many species, including humans, are associated with differential resistance to over 100 detrimental conditions not limited to cancer, diabetes, narcolepsy, systemic lupus erythematosus, HIV and a variety of bacterial and viral infectious agents (Deshpande et al., 2008; Lombard et al., 2006; Niens et al., 2007; Sheehy, 1991; Svejgaard et al., 2001; Saaringen et al., 2002; Thorsby, 1997; Trowsdale, 2011).

Mechanisms such as convergent, concerted and birth-and-death evolution are proposed to act on the MHC leading to rapid evolution of this locus compared to other regions of the genome, with higher rates of non-synonymous (d_N) vs synonymous (d_S) mutations as well as documented gene duplications (Hughes and Nei, 1988, 1989; Nei et al., 1997). A variety of different selective pressures have been proposed to maintain both individual and population level diversity at the MHC including balancing, positive, rare-allele, and pathogen-mediated selection. These forces, along with the rapid evolutionary change, cause the MHC to be an enigmatic and challenging region of study.

Popular MHC molecules often chosen for evolutionary and population studies are the Class IIβ loci, which are members of the immunoglobulin superfamily. Mature class

II molecules are heterodimers of the class IIα/IIβ gene products and are composed of two α and two β subunits of approximately equal length encoded by multiple loci within the MHC. These molecules generally present exogenous peptides derived from extracellular pathogens to T cells. The α1 and β1 subunits comprise the peptide binding groove, which are the most polymorphic regions of the class II molecule. The variability in this region facilitates the binding of foreign peptides with different amino acid compositions.

Evidence of population history affecting the extent of MHC class IIβ diversity in groups of animals has been reported (Spurgin and Richardson, 2010). Genetic drift appears to drive and limit class IIβ gene selection on recently bottlenecked populations, with variation equivalent to or less than that measured using neutral markers (Miller and Lambert, 2004). Diversity increases in comparison to neutral markers as population expansion occurs, regardless of founder effects. Social species show a trend towards higher diversity in the MHC class IIβ region (Aguilar et al., 2004), whereas more solitary species show less diversity regardless of population expansions after bottlenecks (Miller et al., 2010).

High variation at the MHC is correlated with elaborate male ornamentation in some species and optimal allele frequencies in others, both supporting disassortative mate choice (Ekblom et al., 2004; von Schantz et al., 1996). Low intraspecific MHC diversity may increase a population's susceptibility to epidemics (Siddel et al., 2007). Studies characterizing these genes are confounded by various properties of MHC gene families, including high sequence homology between loci and variable numbers of loci (CNV) between individuals among populations (Moore et al., 2002; Nadachowska-Brzyska et al., 2011). For example, 71 MHC class IIβ alleles were found among a group of 24

individual sparrows with 1-6 alleles observed per individual (Bonneaud et al., 2004) and 299 MHC class IIβ alleles were found among groups of newts sampled throughout Europe with 2-6 alleles observed per individual (Nadachowka-Brzyska et al., 2011). In contrast, a recent study using 454 pyrosequencing in conjunction with qPCR showed the loss of class II genes in the Atlantic cod, although approximately 100 class Ia molecules were estimated, an expanded number compared to other teleosts (Star et al., 2011).

Resistance to disease (infections and epidemics) of domestic and wildlife populations has been correlated to both the number and sequence diversity of MHC loci (Alcaide et al., 2010; Jarvi et al., 2004; Spurgin and Richardson, 2010). These studies focused on assessing genetic variability of classical class I and II molecules to aid with efficient animal management and conservation. Class IIβ haplotype diversity in commercial lines, appears to be negatively correlated with selection, with lines subject to increased egg production and growth rate selection having less diversity than random bred control lines (Zhu et al., 1995). Comparative studies have identified differences in MHC haplotype diversity among wild, heritage and commercial breeds of turkey (Chaves et al., 2010), likely due to the historical selective pressures on each of these genetic backgrounds (commercial, heritage and wild). Disease resistance between wild and commercial turkey populations has been studied relating to Newcastle virus (Saif and Nestor, 2002), avian influenza, various bacterial diseases, including *Pasteurella (*Nestor and Saif, 1996) as well as phagocytic responses (Li et al., 2001). The availability of diverse genetic material allows for detailed investigations of the role of MHC diversity in population dynamics and disease challenge studies.

Avian MHC gene organization

The avian MHC comprises two genetically unlinked regions, the *B*-locus (*MHC-B*) and the *Rfpy* region (*MHC-Y*) (Briles et al., 1993; Miller et al., 1996; Reed et al., 2011; Schierman and Nordskog, 1961). These regions are co-located on a microchromosome in both the chicken (GGA16) and the turkey (MGA18). The *B*-locus of the domestic chicken (haplotype B12) was sequenced and was originally described as the 'minimal MHC," as a majority of genes identified in this region contained homologues or orthologues found in other mammalian MHCs. Of particular interest is the conspicuous lack of an extensive class III region. The 'minimal essential' chicken *MHC-B* is about 20 fold more compact than the analogous human region, contains 19 genes with small introns (about 200bp) and spans 92kb. The region is divided into two subregions, the *BF-BL*, which contains 2 classical class I and class IIβ genes, and the *B-G* which contains 1 B-G gene (Kaufman et al., 1999). In 2004, Shiina et al. published a 180kb assembly of the Japanese quail MHC region followed by an extension of the chicken MHC sequence. Both species have similar gene organization, but the Japanese quail appears to have more class I (7 vs 2), class IIβ (10 vs 2) and *B-G* genes (8 vs 3) than the chicken. Evidence in the quail suggests some of the B-G genes are pseudogenes, whereas others have functionality 243kb (Shiina et al., 2007). More recent advancements in turkey MHC research include the mapping and sequencing of the core *B*-locus (Chaves et al., 2009a). To accomplish this, a BAC was sequenced (~10x) and assembled into a single contiguous sequence comprising 197kb. This turkey haplotype contained three more genes in comparison to the homologous regions in the chicken, including two additional B-G genes and a third class IIβ locus. Subsequent screening of the BAC library identified

clones overlapping the 5' end of this *B*-locus sequence, with end sequences matching to BG-like sequences, consistent with Shiina's findings in the chicken. The reason and impacts of these findings are not yet known, but as the genetic composition of the MHC region affects disease resistance/clearance, more comparative studies are needed (Shiina et al., 2007).

The structure of B-G genes makes them difficult targets of study. The first 3 exons have highly conserved intergenic and interspecies sequence identity and comprise the extracellular and transmembrane domains. Limited pockets of SNPs are found scattered among these first three exons. These domains are followed by a cytoplasmic tail comprised of 21 base pair tandem repeats, the sequences of which are not conserved between species or even genes. Studies investigating B-G gene variation in various avian species, including the sandhill crane and ring-necked pheasant confirm the high degree of polymorphism in this gene family (Jarvi et al., 1996, 1999). Most of these population studies used techniques such as serological typing, Southern blots and single stranded conformational polymorphism (SSCP). Southern blots of DNAs collected from a population of WI wild turkeys hybridized with both class IIβ and B-G probes corresponding to *B*-locus sequences (prepared in 1993 by Marcia Miller's lab) show more polymorphism in the B-G genes compared to the class IIβ.

Limited expression studies have been reported in the turkey on MHC genes. A single MHC-*B* class IIβ gene (the major locus), shows predominant expression in the chicken and quail (Jacob et al., 2000; Hosomichi et al., 2006). RT-PCR was used to amplify class IIβ genes expressed in the spleen to determine if selective expression of these genes also occurs in the turkey. PCR experiments identified 6 class IIβ alleles from

genomic DNA, but only 4 expressed alleles corresponded to IIβ1 and IIβ3 (Chaves et al., 2009a) and reduced expression was found in class IIβ2 alleles. Considering this information, expression studies of genes comprising the turkey *MHC-B* are needed, which may correlate with variable functions.

B-G gene expression studies are few, but evidence in the chicken points to specific gene expression in erythrocytes, intestinal epithelium and liver. The B-G molecules expressed in tissue appear to be larger and show more size heterogeneity than the B-G antigens expressed on erythrocytes (Miller et al., 1990). Recently a study found a correlation between a variant of *BG1* and malignant lymphoma resistance in chickens, showing the first known association between a gene in this family and a disease condition (Goto et al., 2009). Clearly, generating an expression profile of this gene family in the turkey is imperative to identifying related functional differences between B-G genes.

The possible expression of B-G molecules in the liver is of particular interest because turkeys are now known as the most susceptible animal to aflatoxin B_1 (AFB_1), likely due to a combination of efficient microsomal activation of AFB_1 by cytochrome P450 enzymes to the toxic intermediate, *exo*-AFB_1-8,9-epoxide (AFBO) and inefficient detoxification of the intermediate mediated by glutathione-S-transferase (*GST*) enzymes (Diaz et al., 2010; Klein et al., 2000; Rawal and Coulombe, 2011). Wild poults show more resistance to aflatoxicity than their domestic counterparts and exhibit less weight loss and mortality than domestic poults in response to aflatoxin challenge (Quist et al., 2000). There have been few studies investigating host gene expression changes correlated with aflatoxin exposure and possible associations between MHC genes and

mycotoxin exposure are unknown, despite the known role of these genes in disease resistance.

Significance of disease correlations to the avian MHC

Most of the accumulated knowledge of the avian MHC stems from the chicken, which was the second vertebrate MHC to be recognized (Briles and McGibbon, 1948). Of particular interest is the relation of MHC haplotypes in the chicken to Marek's disease susceptibility (Briles et al., 1977; Hepkema et al., 1993; Dalgaard et al., 2005; Sarson et al., 2008) and vaccine efficacy (Bacon and Witter, 1993). Investigating host genetics, particularly variation in the MHC, in connection with this and other important turkey diseases such as avian influenza or turkey coronavirus is instrumental in gaining understanding of differences in vaccine response and disease resistance. Some studies suggest host response to vaccination is connected to MHC haplotypes (Saif and Nestor, 2002). This concept is of obvious interest to those wanting to control both current and future flock outbreaks. However, the tools needed to study this are simply not available yet, hence the need for further characterization of the composition and diversity of the avian MHC.

Increasing the health and well-being of humans and animals, both commercial and domestic has many benefits. According to USDA statistics, the value of turkey production approaches 4.4 billion dollars per year, with the United States being the world's largest producer, consumer and exporter of turkey products (USDA/NASS, 2010). In 2010, 7.1 billion pounds of turkey was produced and the value of exported turkey meat products totaled more than $464 million (USDA/FAS, 2011). The estimated

cost of disease to the turkey industry is difficult to estimate, because yearly mortality must be considered along with the costs of vaccinations, antibiotics, eradication programs and production losses such as feed efficiency and lower average daily weight gain. The USDA estimates disease related costs to be about 17% of total production costs per year. An outbreak of avian metapneumovirus (aMPV) in 1997 infected Minnesota turkeys. This pathogen causes an estimate $15 million in losses per year to Minnesota turkey farms (Goyal et al., 2003). Morbidity rates within infected flocks reached 50-100% with mortality of up to 30% (Jirjis et al., 2002). The virus not only hit commercial flocks, but birds in wild populations were infected as well (Bennett et al., 2004), which complicated eradication efforts. Costs associated with single disease outbreaks are significant. For example, the World Organization for Animal Health estimates the efforts to eradicate a highly pathogenic avian influenza strain between 1984 and 1985 cost nearly $65 million (USDA/APHIS, 2001; WHO, 2004). Eradication of exotic Newcastle disease in the early 1970's cost tax payers $56 million (USDA/APHIS, 2001) with more than $160 million spent on the subsequent 2002-03 outbreak (USDA/ARS, 2011).

Commercial turkey production mainly focuses on improving weight gain and feed efficiency with little to no selection for improvement of the immune system. With the rapid genetic advancements made in the field of immunology, the pathologies of many diseases are being connected to their underlying genetic components, with many involving MHC genes. Application of this knowledge could be used to create healthier flocks, decreasing morbidity and mortality rates of the birds as well as lowering costs related to vaccinations, antibiotics, and eradication programs. This helps create a healthier food product for consumers as well. Research on the relation between turkey

host genetics and disease is relatively uncharted with few molecular tools developed for this purpose, which makes investigation of the MHC all the more critical. The B-G genes prove particularly interesting, with little known about this family found exclusively in the avian MHC.

Research Aims and Goals

Given the described role of polymorphic MHC genes, such as the classical class I and II genes, to host immunity and the described polymorphic nature of the B-G gene family, further characterization of these genes is needed to elucidate their role in the avian immune response. The first aim of this research was to sequence and annotate the B-G genes located upstream of the current turkey MHC *B*-locus. The second aim was to test the hypothesis that B-G genes have differing degrees of variability within individuals of diverse genetic backgrounds. Patterns of interspecific and intraspecific sequence variation will be compared to hypothesize a model for B-G gene evolution. Finally, B-G gene expression was examined at the level of mRNA transcripts. Specifically, a B-G tissue expression profile is generated and differential B-G gene expression examined in response to aflatoxin challenge in the context of *Lactobacillus* probiotic. Since B-G genes are expressed in the liver, it is hypothesized that their expression levels vary in response to treatment groups.

Chapter 2

Extended Sequence of the Turkey MHC *B*-locus and Sequence Variation in the Highly Polymorphic B-G Loci

Miranda M. Bauer and Kent M. Reed

Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, St Paul, MN 55108

"With kind permission from Springer Science + Business Media: Immunogenetics, Extended sequence of the turkey MHC *B*-locus and sequence variation of the highly polymorphic B-G loci, 63, 2011, 209-221, Bauer MM and Reed KM."

Abstract

Genetic variation in the major histocompatibility complex (MHC) is directly correlated to differences in disease resistance. Immunity is greatly dependent on highly polymorphic genes in the MHC, such as class I, class II and class III complement genes. Preliminary studies of wild turkey populations show extreme polymorphisms in a family of genes exclusive to the avian MHC, the class IV or B-G genes. Significance of this variation is unclear as there are few and conflicting studies of the expression of these genes. Confounding understanding of B-G variation is the lack of a complete delineation of the number of loci in the turkey genome. Direct 454 sequencing of a clone from the CHORI-260 BAC library was used to extend the turkey MHC *B*-locus sequence, identifying two additional B-G genes in the species. Phylogenetic analysis shows species-specific gene evolution supporting a birth-death model of evolution for the B-G gene family. Analysis of variation within the signal peptide sequence (exon 1) found two clusters of polymorphism among the turkey B-G genes. Resequencing of exon 1 in a diverse sample including wild, heritage and commercial turkeys confirmed multiple alleles at each B-G gene. Future studies aim to correlate B-G variation with group and individual immunological differences.

Key words: MHC, *B*-locus, B-G genes, SNP haplotype

INTRODUCTION

The major histocompatibility complex (MHC) encodes genes responsible for host immune defense. As antigen diversity increases, so must the diversity of host molecules that can bind to and aid in clearance of ever evolving pathogens. MHC molecules have a variety of properties and mechanisms for this purpose. Two striking characteristics of many MHC molecules are their extreme polygenic and polymorphic nature, particularly in respect to antigen presenting molecules and those molecules with which they directly interact.

The B-G gene family of the MHC continues to be an enigmatic area of study in the avian genome, despite their discovery in the chicken (*Gallus gallus*) over 30 years ago (Pink et al., 1977; Vilhelmová et al., 1977). Early studies of the chicken originally described these molecules as erythrocyte antigens closely linked to the B-F (class I) and B-L (class II) regions of the chicken MHC (Koch et al., 1983; Longenecker et al., 1980). Each of these gene families encode highly polymorphic molecules, with class I and class II molecules responsible for antigen presentation to CD8+ and CD4+ T cells, respectively. Function of the B-G antigens has not yet been fully elucidated, although they loosely share characteristics with the chicken leukocyte receptor (CHIR) complex family (Laun et al., 2006). For example, the expressed mRNAs have an extracellular domain consisting of an N-terminal signal peptide, an IgV-like region, a transmembrane domain, individually variable cytoplasmic tails and immunoreceptor tyrosine-based inhibition motifs (ITIM) located in the exon encoding the C-terminus (Goto et al., 2009; Kaufman et al., 1989, 1990). ITIMs are conserved amino acid sequences [I/V]XYXX[L/I] located in the cytoplasmic tails of some proteins that bind to phosphatases and control the intensity of

signal received by the cell. The extreme immunogenicity of B-G antigens is well established. Exposure to B-G antigens induces an antibody response before immunization in non-avian species, such as mice and rabbits, although the genes have only been identified in birds (Goto et al., 1988; Miller et al., 1982; Salomonsen et al., 1987). Early studies in the chicken hypothesized interaction of B-G and class I antigens expressed on erythrocytes (Hala et al., 1981) and with B cells (Salomonsen et al., 1991). Studies using mAbs found conflicting results regarding tissue localization of B-G proteins in immune tissues such as the bursa and thymus (Miller et al., 1990; Salomonsen et al., 1991), but convincing evidence for antigen expression in a variety of tissues, including the caecum, small intestine and liver. A protein showing high sequence similarity to *BG2* was found to localize in chicken to the intestinal epithelial cell microvillus, where it reportedly interacts with brush-border myosin-1 protein and actin (Bikle et al., 1996).

Previous studies have focused on elucidating the polymorphic nature of the B-G genes. Of particular interest are studies of recombinant chicken lines showing differences in immunological resistance to infection, including Marek's disease, associated with certain B-G haplotypes (discussed in Goto et al., 2009). Recently, a recombinant *BG1* haplotype was shown to increase resistance of chickens to Rous sarcoma virus (Goto et al., 2009). A main challenge is delineating allelic differences caused by phenomena such as exon shuffling and alternate splicing versus actual nucleotide polymorphisms (Hosomichi ei al., 2008). In addition, studies that rely on typing with B-G alloantisera may ignore potential contributing polymorphisms in the adjacent BF-BL region.

It is currently not known how many B-G genes are present in the avian MHC, but clearly, the number of genes varies between species. Most studies have used techniques

such as hemagglutination, restriction fragment length polymorphism (RFLP), Southern hybridization and immunoprecipitation for haplotype characterization (Kaufman et al., 1990; Miller et al., 1988; Nishibori et al., 2000 ; Ruby et al., 2005). The best studied species is the chicken, where one B-G gene was originally identified in the MHC *B*-locus (Kaufman et al., 1999), and two additional B-G genes were more recently discovered in the extended *B*-locus sequence upstream of a cluster of TRIM genes (Shiina et al., 2007). In the quail, three B-G genes and five B-G-like pseudogenes have been described (Shiina et al., 2004). B-G genes have also been identified in the ring-necked pheasant (Jarvi et al., 1996, 1999), and three B-G genes are found in the MHC *B*-locus of the turkey (Chaves et al., 2009a).

This study was designed to further characterize the B-G genes in the MHC *B*-locus of the turkey (*Meleagris gallopavo*). Specifically, we describe new B-G genes in the extended *B*-locus and compare these to the previously identified B-G genes. Examination of interspecies phylogenetic relationships and intraspecies variation in the B-G genes provides new insights into the pattern of evolution and importance of these unique members of the avian MHC.

MATERIALS and METHODS

MHC BAC identification and sequencing

Two bacterial artificial chromosome (BAC) clones (56J17 and 38O12) from the CHORI-260 turkey BAC library were identified as containing B-G sequences by overgo hybridization. The library was generated from DNA of a female bird (Nici) from a partially inbred Nicholas subline and is currently believed to be homozygous at the MHC

(Chaves et al., 2009b). End sequencing of 56J17 and 38O12 determined that the inserts of these clones partially overlapped and extended the 5' end of the MHC *B*-locus sequence (GenBank DQ993255), with both BACs terminating in *BG*-related sequences (ET222701 and ET222704). The 3' end sequence of 56J17 (ET222703) was located closer to the 5' end of the current turkey *B*-locus sequence than 38O12 (ET222701). As such, this BAC was hypothesized to extend further upstream of the known *B*-locus sequence and was chosen for further sequencing. The 5' end sequence of 38O12 was later verified to be enclosed within 56J17.

For sequencing, a starter culture consisting of 5 ml of LB media supplemented with 100ug/ml of ampicillin was inoculated from a glycerol stock and grown in a shaker for 8 hours at 37°C. From this culture, 1 ml was transferred into 500 ml of LB with 100ug/ml of ampicillin added and grown for 16 hours at 37°C with gentle shaking. BAC DNA was purified using the Large Construct kit (Qiagen) following the standard protocol and with the removal of contaminating genomic DNA. Purified BAC DNA was sequenced using the Roche 454 FLX platform at the Advanced Genetic Analysis Center, University of Minnesota. Approximately 12x coverage of the BAC was obtained with 16,940 total reads. About 5.5% (940 total) of these aligned to vector or *E. coli* sequence and were discarded.

Sequence reads were assembled into contigs using gsAssembler (Roche) and Sequencher software (Gene Codes, Inc.). To aid in assembly of repetitive regions, BAC DNA was restriction digested with *Eco*RI or *Bam*HI to obtain fragments 2-3 kb in length. These were ligated into compatibly digested pBluescript II KS+ and transformed into *E. coli* DH5α competent cells. A total of 20 clones were screened for insert size and

selectively end sequenced using vector-specific primers. Primer walking and PCR sequencing were used to verify assembly and fill unresolved gaps. PCR reactions with HotStar Taq polymerase (Qiagen) used the following parameters: 95°C for 15 minutes, 35 cycles of 95°C for 30 sec, 58°C for 30 sec, and an extension time of 1min /kb at 72°C with a final extension of 72°C for 10 minutes. PCR amplicons were ligated into either pGEM or pDrive vector, transformed into electrocompetent *E. coli* DH5α cells and sequenced using vector and/or gene specific primers.

Sequence annotation

The *in silico* analysis tool, BLAST, was used to identify genes within the assembled BAC sequence. Open reading frames were predicted by aligning the assembled BAC sequence with the annotated chicken MHC gene sequences (AB268588) and where available turkey mRNA sequences. The program PIPmaker (Schwartz et al., 2000) was used to determine CG content and identify CpG islands.

To verify predicted exons, select transcripts were amplified from liver cDNA prepared from a commercial turkey of a Hybrid converter line. For PCR, RNA was isolated using Trizol reagent (Invitrogen) and reverse transcribed using the AffinyScript Multiple Temperature cDNA Synthesis Kit (Stratagene). Oligo-dT primers were combined with 500 ng of RNA template. First-strand cDNA synthesis was performed with the following cycling parameters: 42ºC for 60 min and 72ºC for 15 min. Amplicons were purified using ExoSapit (USB) and directly sequenced using the original amplification primers. Synonymous and non-synonymous mutations between chicken and turkey transcripts were manually documented and percent nucleotide and amino acid similarities

were calculated. Sequence variation in exons 1, 2 and 3 of turkey B-G genes was analyzed by performing pairwise comparisons. Identification of the signal peptide sequence was verified for all B-G genes identified in the turkey BAC using SignalP 4.0 (Petersen et al., 2011).

Interspecies sequence comparisons and phylogenetic analysis

The NCBI databases were queried with *BG* exon 1 sequences using BLAST to identify B-G genes in other avian species. Recovered sequences were aligned and compared with turkey B-G genes sequenced from Nici using Clustalx (Larkin et al., 2007) and Mega (Tamura et al., 2007). Neighbor Joining trees were visualized using TreeViewx for 1000 bootstrap replicates (Page, 2002).

Intraspecies genetic variation

A subset of five commercial, six heritage breed and four wild turkeys were selected for comparative sequencing of B-G genes. Samples were obtained from the following sources; commercial birds were all Nicholas bred animals obtained directly from Nicholas Turkey Breeding Farms, Inc (Lewisburg, WV). Two of these birds are F1s from the UMN/NTBF mapping families (Reed et al., 2003) and the remaining three are each from different breeder lines (Chaves et al., 2009b). Samples from heritage breed birds were obtained from Dr. Ed Smith (Virginia Tech University) and consisted of 2 Royal Palm, 1 Spanish Black, 1 Narragansett and 2 Bourbon Red. Genomic DNA was extracted from blood samples using Trizol (Invitrogen). The DNA from the wild birds was provided by

Dr. Marcia Miller (City of Hope Research Hospital, Duarte, CA). These animals originated from a wild population in Grant County, Wisconsin sampled in 1993. Primers were designed based on sequences from Nici and used to amplify exon 1 of the five sequenced turkey B-G genes. Exon 1 of *BG1*, *2* and *3* were amplified using a universal forward primer (BGuniv, 5'-TGAAACCCAAGTGTGTGTGG-3') in conjunction with locus-specific reverse primers (BG1R, 5'-CTTCCGCATCTCCTTCTC AG-3', BG2R, 5'-TCTTCTCAGTCTCATTCCTCTCC-3' and BG3R, 5'-CATCTTCTAC GTTTCCCTCCTC-3'). The newly identified *BG4* and *BG5* genes were amplified using locus specific forward (BG4F, 5'-CCATCCCCAATCTCTTTTCC-3', BG5F, 5'-TCCAA GTCCTCTTCCCCATC-3') and reverse (BG4R, 5'-TGAAACCCAAGTGTGTCTGG-3' and BG5R, 5'-CAAGTGTGTGTGGGTGCTG-3') primers. Sequences were aligned and variants were identified for each gene. Phylogenetic analysis was conducted using Mega (Tamura et al., 2007) and Neighbor Joining trees were generated for 1000 bootstrap replicates.

RESULTS

Assembly of BAC 56J17

Sequencing using the Roche GS FLX provided ~8000 reads from BAC 56J17 representing approximately 12x coverage. After removal of contaminating genomic sequence, average read length was 251 bp. Removal of vector (PTARBAC2.1) sequence resulted in a completed insert of 119 kb. Of the 119 kb insert, 79,437 bp directly overlapped the 5'end of the MHC *B*-locus (DQ993255) previously described by Chaves et al. (2009a), with the 3' end of the insert terminating in the middle of the TRIM41 gene. Sequence of the

overlapping portion of 56J17 was identical to the original sequence. The remaining 39959 bp of 56J17 extends the 5' boundary of the turkey MHC *B*-locus and is hereafter referred to as the extended sequence. The assembled sequence is accessioned in Genbank as HQ008883. Analysis of the extended sequence found nucleotide content to average 52% CG. Two CpG islands were predicted; the first spans positions 17714 to 18489 and the second occurs from position 30885 to 31256.

Gene annotation

Using predicted chicken mRNA sequences and turkey ESTs, the position, exon coding regions and amino acid sequences of each gene in the extended turkey MHC *B*-locus were identified. Organization of this region was similar to the orthologous region in the chicken MHC (Shiina et al., 2007). Included in the 40 kb extended sequence, are five complete genes: two *BG*, *KIFC*, *Blec3*, *Bzfp3* (Figure 2.1) and the terminal exon of *TRIM7.2*. For each gene identified, the position, predicted coding sequence and resulting number of amino acids are given in Table 2.1. Predicted coding sequences of the newly identified genes were compared to syntenic chicken orthologs to determine percent identity and to identify synonymous and non-synonymous mutations.

KIFC

KIFC (Carboxy-terminal kinesin 1) is a member of the C-terminal motor domain-type kinesin superfamily. The closest human homologue is *KIFC1*, a class III MHC gene, which is associated with the nuclear membrane and is known to affect mitotic spindle formation and microtubule function (Yang and Sperry, 2003). In the turkey, *KIFC* is 8227

bp in length with a coding region of 1944 bp and comprises 14 exons (Table 2.1). A CpG island is located 272 bp before the start codon and extends through exon 3. The turkey sequence shares 92% nucleotide identity and 92% amino acid identity with the chicken orthologue. In comparison to the chicken, there are 2 codon deletions in the turkey at position 70 and 3 codon insertions at position 229. In total there are 163 nucleotide substitutions between the species, with 45% of the substitutions being non-synonymous. The dN/dS ratio of this gene (0.45) is indicative of purifying selection, which selectively eliminates deleterious mutations (Axelsson et al., 2005).

Blec3

Blec3 (a c-type lectin receptor) is part of a multigene family that codes for C-type lectin proteins and are not found in the MHC region of mammals. Genes in this family encode signaling molecules closely related to activation-induced lymphocyte receptors, particularly *CD69*. In the turkey, *Blec3* is 3171 bp in length with a coding region of 516 bp comprising of 5 exons (Table 2.1). The turkey *Blec3* shares 89.7% nucleotide identity and 82.6% amino acid identity with that of the chicken. The turkey mRNA sequence is 9 bp shorter, with a deletion of 3 codons spanning position 494 to 504. Additionally, there are 56 nucleotide substitutions, 54% of which are non-synonymous. The dN/dS ratio for *Blec3* is >1 (1.15) suggesting diversifying selection at this locus.

Bzfp3

Bzfp3 (a zinc finger protein) is a member of a multigene family that encodes for zinc finger proteins. These genes are particularly well known for their ability to bind to DNA

and affect processes related to transcription. In the turkey, *Bzfp3* is 6166 bp long with a coding region of 1557 bp comprising 11 exons (Table 2.1). There is a CpG island 941 bp before the start codon. The turkey gene shares 93% nucleotide and amino acid identity with its chicken orthologue. Between the species, there are a total of 40 amino acid substitutions with the transcript in the turkey being 12 bp shorter as the result of a deletion at position 212, an 8 base pair gap between positions 219 and 226 and a codon deletion from positions 688-690. Based on dN/dS ratio (0.53), *Bzfp3* appears to also be under purifying selection.

B-G genes

The extended MHC sequence of the turkey contains two additional B-G genes, designated *BG4* and *BG5*. *BG4* is 6876 bp in length, comprising 37 exons and *BG5* is 5580 bp, comprising 30 exons (Table 2.1). These B-G genes show similar exon patterns as those of the chicken and the three previously described B-G genes of the turkey. Included in the predicted transcripts are a 5'UTR that is not highly conserved between B-G genes, three conserved exons (1-3) comprising the extracellular and transmembrane domains, variable length cytoplasmic tails comprised of 21 bp tandem repeats and a moderately conserved 3'UTR (Figure 2.2). The cytoplasmic tails of *BG4* and *BG5* include 34 and 27 exons, respectively, resulting in a slightly longer coding region [1251 bp (*BG4*) and 1131 bp (*BG5*), compared to 948 bp (*BG1*), 1080 bp (*BG2*) and 879 bp (*BG3*)]. There is no evidence of immunoreceptor tyrosine-based inhibition motifs (ITIMs) in the tails of the alleles represented in the turkey BAC. Each of the turkey B-G transcripts was amplified by PCR from liver cDNA confirming that *BG4* and *BG5* are functional genes. Sequencing

of PCR products confirmed the predicted exon sequences. Substitution rates are not reported between the turkey B-Gs and the syntentic chicken B-Gs, since orthology cannot be determined with certainty by sequence similarity comparisons.

Turkey B-G gene comparisons

Exons 1, 2 and 3 are more conserved between and within species than the highly variable cytoplasmic tail (Figure 2.2). Length of the tail is not only variable between genes, but previous studies in the chicken indicate variability of loci among individuals (Kaufman et al., 1990). Comparisons among turkey B-G loci found variation in length and number of exons; *BG1* and *BG3* are more similar to each other than to the other B-G genes (Figure 2.2). Nucleotide and amino acid differences for exons 1-3 of all sequenced turkey B-G genes (*BG1-5*) were compared to quantify variability within the species. Pairwise comparisons of exons 1, 2 and 3 show an average of 8.5%, 6.25% and 8.72% nucleotide difference, respectively. This establishes the signal peptide as the most variable part of the extracellular domain.

As the most variable part of the extracellular domain, the B-G signal peptide region is of particular interest. Previous reports in the chicken identify the 100 base pair exon 1 as encoding the signal peptide (Miller et al., 1991). This was confirmed for the turkey B-G genes using SignalP (Petersen et al., 2011). In a recent study, Goto et al. (2009) provided evidence of *BG1* acting as a surface receptor in the chicken. Sequence Variation in the sequence of exon 1 between the B-G loci identified in *B*-locus occurred primarily in 2 regions, one located between nucleotides 4 and 14 and the other between 36 and 45. In contrast, nucleotide differences in exons 2 and 3 were dispersed across the

exons (Figure 2.3). Exon 1 of *BG1* is most similar to that of *BG4*, having only 2 nucleotide differences and identical predicted amino acid sequences. There are 5 nucleotide differences between *BG1* and *BG3*, resulting in 3 amino acid differences. Exon 1 is least conserved between *BG4* and *BG1-3* with nucleotide differences averaging just over 13%. Exon 3 is most conserved between *BG1* and *BG3* with a nucleotide difference of 0.9%.

Alignment of amino acid sequences of exon 2 with the V motif (Hunkapiller and Hood, 1989; Miller et al., 1991) (Figure 2.3) demonstrates both the IgV-like sequence of this exon and the variability of this hypothesized ligand binding region. *BG2* and *BG5* are most similar and share 86% amino acid sequence with the IgV motif. The remaining genes, *BG1*, *3* and *5* are 84.2%, 83.3% and 85% similar to this motif. Some of these nucleotide differences would result in amino acid compositions with different functional or physical properties, possibly changing the conformation of the molecule. This clearly establishes exon 2 as the coding sequence of a variable IgV-like domain for all known turkey *B-G* molecules.

BG interspecies analysis

Nucleotide and amino acid sequences of exons 1- 3 of turkey genes *(BG1-5)* were compared with the corresponding exons of the syntenic chicken B-G genes (referred to as *cBG1-3*, Table 2.2). Whether these genes are truly orthologous is not known. Exons comprising the cytoplasmic tails were also excluded, as sequence variation precluded reliable comparative alignments.

Sequence comparisons demonstrate the reticulate nature of the B-G genes. The first 3 exons of turkey *BG4* share an average 86.5% nucleotide and 78.7% amino acid similarity with chicken *cBG2* (Table 2.2). Comparable values are seen for *cBG3* (exons 2-3) with average similarities of 90.8% and 84.6%. This comparison is somewhat biased because exon 1 of *cBG3* is not known. Similarity values for *BG5* versus *cBG2* (first 3 exons) are 86.8% nucleotide and 78.2% amino acid, respectively. Average similarities for *BG5* with *cBG3* (exons 2 and 3) are slightly higher at 90.6% and 84.0%. Although turkey *BG4* and *BG5* are syntenic in position to *cBG2* and *cBG3*, similarity values for these comparisons are not substantially different than comparisons contrasting turkey *BG1*, *2*, or *3* with the chicken BG genes (Table 2.2). Examination of intron sequences did not further resolve homologies (Figure 2.7). Sequences of introns 1 and 3 are more conserved than intron 2 where several length variants occur among the compared genes.

Interspecies evolutionary relationships of genes can provide insights into gene function. The first three exons of the B-G genes are the most phylogenetically informative because of their highly conserved nature and confidence in sequence alignments. Exons 1 and 2 are potentially the most interesting because of their hypothetical functions in signaling and ligand binding. Comparisons between published exon sequences from other galliformes[(1) (2)] revealed B-G phylogenetic relationships between species. Full gene sequences were not available for some species; for example, the sequence of *cBG3* begins in the middle of exon 2 (Shiina et al., 2007). Similarly, the sequenced B-G gene in the ring-necked pheasant extends only through exon 2. The quail MHC is unique with several

[1] GenBank AB078884, U32559.1,

[2] GenBank AB078884, U32559.1,

identified B-G pseudogenes, *qBG1, 2, 3, 5* and *8.* Predictably, some of these are missing exons.

Gene phylogenies were initially constructed separately for each of the first three exons. Sequences of exons 1 and 2 showed distinct clustering of B-G genes by species. Exon 1 (signal peptide) of turkey *BG2* was most closely related to that of *cBG1*. Clustering based only on exon 3 sequences (transmembrane domain) identified two major clusters. One cluster contained the turkey *BG1, 2, 3* and *5* genes along with quail *BG4* and pseudogenes, *qBG2* and *qBG3*. The other cluster contains *BG4, cBG2*, *cBG3*, *qBG1, qBG6*, and *qBG7*. Exon 3 may be least related to the hypothesized receptor function of the protein, making mutations between species likely to be more random in type and position. A phylogenetic tree depicting the evolutionary relationship among B-G genes based on the combined sequences of the first 3 exons is depicted in Figure 2.4. This tree supports the hypothesis that the B-G gene family has diverged through expansion/contraction after speciation and is consistent with the B-G molecules having a role as a ligand receptor of species-specific pathogens.

Intraspecies sequence variation

A diverse set of 5 commercial, 6 heritage and 4 wild turkeys was chosen for exon 1 resequencing to further characterize variation of the signal peptide sequences between B-G alleles and loci. Sequences for each individual at each of the 5 amplicons were grouped by locus and the minimum number of alleles per locus was determined as the number of unambiguous sequences (1 SNP or fewer) in the group. Among the exon 1 sequences, 22 variable sites were identified. A majority of the variable nucleotides were found in two

clusters (positions 4 to 15 and 32 to 45), consistent with the sequence variation observed between loci in the assembled *B*-locus BACs (Figure 2.5).

A minimum of 19 distinct exon 1 alleles was identified among the 15 birds (Figure 2.5). Although many nucleotide variants were shared across B-G genes, locus-specific SNPs were identified for each gene. For all loci except *BG-4*, an allele sequence identical to the corresponding locus in the assembled BAC sequences was recovered in the sample set. The number of alleles per locus ranged from 6 at *BG-1* to 2 at *BG-2*. Individual birds showed the least variation at *BG-2* where the sequences were invariant for all but one wild turkey. The sequence from this bird (allele H) was originally thought to represent a PCR artifact because of the TTTCCAA motif (positions 39-45) that is nearly universal in *BG-1* and *BG-3* loci (Figure 2.5). However, examination of flanking sequence confirmed it as a *BG-2* related sequence. Phylogenetic analysis was used to examine relationships among the 19 alleles (Figure 2.6). Alleles grouped into two major clusters with strong bootstrap support. The first cluster contains three subclusters comprised of sequences derived from the *BG1*, *BG3*, and *BG4* primer sets. The second cluster is comprised of two groups containing sequences obtained with the *BG2* and *BG5* primer sets.

DISCUSSION

Five new genes, *Bzfp3*, *Blec3*, *KIFC* and two B-G genes, were sequenced in the turkey extended *B*-locus. These genes are members of MHC multigene families which are a common feature of the MHC and often encode for quickly evolving proteins of critical immune related function. Immune function and variability of these genes is not well characterized, but studies indicate gene-dependant tissue expression. KIFC proteins are

most extensively studied in mice and are responsible for organelle transport (Noda et al., 2001; Saito et al., 1997), *Blec3* is most closely related to the human C-type lectin transmembrane receptor gene *CD69* (Natarajan et al., 2000), and *Bzfp3*, a nucleic acid binding protein, is closely related to mammalian Zinc finger protein 91 (Unoki et al., 2003).

Gene content and order of the extended turkey MHC appear identical to the syntenic region in the chicken; however, both the turkey and chicken differ considerably from the syntenic region in the quail. The higher stability of MHC organization and gene content of the turkey and chicken when compared to the quail may relate to the diversity of pathogen exposure of each species. The quail, as a migratory bird, is hypothesized to be exposed to a more diverse set of pathogens than either the turkey or chicken (Shiina et al., 2007), stimulating genetic variability in regions related to immunity, such as the MHC. Studies in the quail show increased resistance of this species to Marek's disease compared to the chicken. This phenomenon is hypothesized to relate to variances in the number of gene loci and/or alleles between the species (Mikami et al., 1975).

B-G proteins were originally isolated from chicken erythrocytes (Longenecker and Mosman, 1980), but were later found in a variety of tissues, namely the liver and intestinal epithelium. Proteins isolated from these two tissues exhibit considerable size heterogeneity. Variable cytoplasmic tail lengths and alternate splice sites have been shown to contribute to variable size B-G proteins (Figure 2.2; Kaufman et al., 1989). Although length variation may be the result of alternate splicing, it is likely that multiple B-G proteins are expressed in both the liver and intestinal epithelium, and that size heterogeneity is the result of length differences between loci. Differences in exon number

between the turkey B-G loci would clearly lead to size differences in the translated proteins. The B-G molecules specific to erythrocytes may have a more limited expression profile, with the least size variation found in this cell type (Miller et al., 1990). The extent of variation between B-G proteins in the turkey caused by these phenomena is currently being examined.

Pairwise comparisons between the first three B-G exons show most variation in exons 1 (the signal peptide) and 3 (the transmembrane domain). The IgV region is highly similar between genes, perhaps due to selection pressure favoring conservation of function. Signal peptides are best known for their role in protein translocation. However, they have also been shown to affect co- and post-translational events, along with the behavior of downstream transmembrane domains. The N-terminal region of signal peptides is responsible for the translocation of the translated protein to the endoplasmic reticulum. The hydrophobic region is critical for subsequent cleavage of the signal peptide, which is considered crucial for post-translational function. The time point and location of signal peptide cleavage after protein translation can result in cytoplasmic build up or a change in the level of protein expression. This can change the function of the molecule, leading to detrimental effects on the host cell (Hegde and Bernstein, 2006). The location of SNPs in the signal peptide between and within turkey B-G loci is intriguing in that one cluster of variants is located in the N-terminal region and the other is adjacent to the hydrophobic core. These may alter the molecule's function. Thus far, there has been no evidence of post-transcriptional or post-translational modifications of B-G proteins.

Interspecies comparisons

Phylogenetic trees examining the interspecies relationship of B-G genes show clustering by species rather than by gene. This is typical of gene families that independently expand or contract after speciation. A similar pattern is observed for other Ig-like receptors in humans and mice (Martin et al., 2002). Molecular data suggest MHC genes follow a birth-death evolution model more closely than other models such as concerted evolution, which is primarily characterized by the processes of gene conversion and unequal crossover (Wan et al., 2011). Gene conversion results in interspecies sequence homogenization and does not change the number of genes in the same family between species. The process of unequal crossover results in homogenization of genes and may result in an increase or decrease in the number of genes in a family between species as a result of random mutation after species divergence. However, a certain range is maintained so as to not alter the functionality.

The birth-death model is characterized by duplication and subsequent loss of genes and has been described in the class Ia, class Ib and class II gene families between mammalian orders (Debenham et al., 2005; Nei et al., 1997; Nei and Rooney, 2005). The B-G gene family has multiple characteristics that support this pattern such as the variation in gene number within subregions of the *B*-locus between closely related taxa. For example, the chicken has one B-G gene in the core *B*-locus sequence (*cBG1*) whereas the turkey has three (*BG1-3*) in the syntenic position. Whether this is a result of an expansion in the turkey or a contraction in the chicken is unknown. The quail is different from both species with one B-G pseudogene in this region, which suggests a regional expansion in the turkey. At least three structurally sound B-G genes are encoded in the sequenced quail

haplotype (Shiina et al., 2007), and several B-G pseudogenes are also present. The total number of B-G genes in the turkey has yet to be determined but the results of the present study confirm a minimum of six. Examination of unassigned contigs form the draft whole genome sequence (Dalloul et al., 2010) and preliminary experimental results suggest at least one more functional gene and one pseudogene (Bauer unpublished). Presence of pseudogenes is one of the characteristic features of the birth-death evolution model. Although these are seen in the extended *B*-locus sequence of the quail, further examination of the turkey and chicken are needed to determine if additional B-G genes or pseudogenes are present upstream of the B-G cluster.

Intraspecies variation

In the turkey, B-G genes exhibit high conservation of the first three exons, suggestive of duplication and subsequent recombination events that are consistent with the birth-death evolution hypothesis. Partial locus gene conversion and unequal crossover may contribute to B-G gene variation in parallel with this process. Gene conversion and unequal crossover result in genes with highly similar exon sequence due to meiotic recombination. Unequal crossover occurs between loci located in gene clusters that are not homologous in position. In order for exon pairing to occur during this recombination event, gene copies must have highly related exons and introns, which is true for the first three exons and two introns of the five turkey B-G genes (Figure 2.7). The degeneracy and varying numbers of exons in the cytoplasmic tails of B-G genes seems inconsistent with gene conversion and unequal crossover theories. However, this feature of the B-G loci may be the result of exon shuffling and selection towards structural rather than sequence conservation.

The extracellular regions of B-G molecules are more likely to be under immunological selection than exon 3, the transmembrane domain, due to their potential location on the cell surface and structural features (ie IgV motif and possibly the signal peptide). The signal peptide sequence is generally cleaved off of the translated protein in the endoplasmic reticulum. However, several post cleavage functions of this type of sequence have been identified. In humans, *HLA-E* presents a 9 amino acid signal peptide sequence on the cell surface (Braud et al., 1997). The presented sequence is a conserved signal sequence on several MHC-class I proteins (Borrego et al., 1998; Braud et al., 1998). Sequence comparisons and reconstructed phylogenies suggest similar evolutionary constraints. Loci under the influence of gene conversion should show admixture, which is observed between *BG1* and *BG3* alleles. The remaining B-G loci separate into monophyletic clades, typical of genes undergoing birth-death evolution.

MHC haplotypes have been associated with variation in disease resistance in many species. Genotyping experiments by restriction fragment length polymorphism (RFLP) of various turkey show line-dependent banding patterns (Emara et al., 1993; Sacco et al., 2001; Zhu et al., 1995), and MHC variation between wild and commercial turkey haplotypes has been established by SNP haplotyping (Chaves et al., 2010). Southern blots and RFLP experiments measuring B-G variation demonstrate striking polymorphisms within and between groups of chickens (Kaufman et al., 1990; Miller et al., 1988; Nishibori et al., 2000; Uni et al., 1995; Yonash et al., 2000). Wild and commercial birds show differences in antibody response, vaccine efficacy and disease resistance related to polymorphisms in the MHC (Bacon and Witter, 1993; Bonneaud et al., 2005, 2006; Gehad et al., 1999; Loiseau et al., 2008, 2009; Westerdahl et al., 2005). Genes encoded by the

MHC are known to have high linkage disequilibrium and in commercial turkeys are inherited as haplotypes (Chaves et al., 2010). MHC haplotypes have been correlated with immunological resistance and the level of antibody response in numerous species, with certain haplotypes correlated with increased susceptibility to pathogenic infections. Commercial turkeys are bred primarily for production traits such as growth rate and body weight and generally demonstrate reduced variability at MHC loci (Chaves et al., 2010; Nestor et al., 1996; Zhu et al., 1996). Wild birds experience no artificial selection pressure and show higher MHC sequence variability compared with commercial birds (Chaves et al., 2011). Wild turkeys in the United States have historically passed through a significant genetic bottleneck resulting from severe population elimination with subsequent restocking (Mock et al., 2002; Reed, 2009). Studies of MHC variability in heritage birds are limited, but these turkey breeds are primarily selected for color traits and show varying degrees of relatedness to commercial turkeys (Kamara et al., 2007).

Phylogenetic trees examining relationships of B-G exon 1 alleles show segregation of clusters by locus. The divergence of alleles provides evidence supporting differential selection pressures acting on each group. The significance of this divergence is unknown but may relate to differences in disease resistance as sequences differences are predicted result in amino acid differences. The exon 1 sequences of *BG2* seemingly contradict this observation as this exon was nearly invariant among the individuals examined. This is typical of loci with specialized functions undergoing purifying selection. B-G alleles of the other four genes show varying degrees of polymorphism with few locus-specific polymorphic sites.

CONCLUSION

The B-G gene family remains a complex and enigmatic area of study. Intraspecies and interspecies comparisons of the turkey B-G genes support a birth-death model of evolution acting in tandem with recombination processes such as gene conversion, unequal crossover and exon shuffling. The similarity in the position of these genes among avian species supports an initial duplication of these loci in a common ancestor. The polymorphic nature of these genes, particularly evident in the signal peptide and cytoplasmic tail, may have effects on the protein structure and function of these genes. Polymorphisms in the signal peptide may affect the efficiency of translocation or signaling properties of the molecule via interaction with downstream exons. Locus-specific segregation suggests divergence in function of these molecules. Expression of these genes in various tissues is under investigation.

The effect of B-G gene variation on vaccine response and disease resistance in the turkey is unknown. Future studies aim to investigate variation in the IgV region and splice variants in the cytoplasmic tail lengths both between individuals and among groups of turkeys. Variation in the IgV region may affect ligand binding properties and differences in cytoplasmic tail length may, as in the chicken, include in the presence of ITIMs on certain molecules. It will be necessary to examine several cell types to characterize the cellular distribution of these proteins and determine locus-specific differences in expression. Likewise, protein studies are needed to elucidate the behavior of the signal peptide and its effects on the functional properties of B-G proteins. The ligands of these receptors remain to be discovered, but there is evidence of B-G proteins interacting with other MHC molecules (Hala et al., 1981), and other Ig-like receptors have

proven essential for Ig binding and crosstalk between the innate and adaptive immune system. The effect of the highly polymorphic family of B-G receptor molecules on the innate and adaptive immune responses await further investigation, but their linkage with the class I and class II molecules makes them an interesting and unique area of study.

Acknowledgements

We thank Zheng Jin Tu of the Minnesota Supercomputing Institute for computational assistance. This research was supported by a grant from the Cooperative State Research, Education and Extension Service, U.S. Department of Agriculture (2009-35205-05302); MMB was supported in part by a USDA training grant in Microbial and Animal Health Genomics (2006-38411-17096).

				Turkey		Chicken					
Gene	**Strand**	**Position**	**Exons**	**Bp**	**AA**	**Bp**	**AA**	**Nucleotide Identity**	**dN/dS**	**AA Sub**	**Amino Acid Identity**
BG5	+	2325-8410	30	1131	377						
BG4	+	9921-17010	37	1251	417						
KIFC	+	18000-28201	14	1944	648	1944	648	92.0%	0.45	51	92.0%
Blec3	-	28806-31977	5	516	172	507	168	89.7%	1.15	30	82.6%
Bzfp3	+	33820-42172	11	1581	527	1569	522	93.0%	0.53	40	92.4%

Table 2.1. Genes in the extended turkey MHC. For each gene, the position, number of exons, length of coding region and predicted number of amino acids are given. Comparative data are presented for *KIFC*, *Blec3* and *Bzfp3*.

	Exon1				Exon 2				Exon3			
Turkey/Chicken	**Diff (bp)**	**Nucleotide Similarity**	**Diff (AA)**	**AA Similarity**	**Diff (bp)**	**Nucleotide Similarity**	**Diff (AA)**	**AA Similarity**	**Diff (bp)**	**Nucleotide Similarity**	**Diff (AA)**	**AA Similarity**
BG1/cBG1	16	0.84	9	0.73	25	0.93	2	0.98	16	0.85	10	0.72
BG1/cBG2	16	0.84	10	0.70	23	0.93	16	0.86	12	0.89	7	0.81
BG1/cBG3	NA	NA	NA	NA	22	0.94	8	0.93	17	0.84	11	0.69
BG2/cBG1	8	0.92	6	0.82	28	0.92	14	0.88	17	0.84	9	0.75
BG2/cBG2	11	0.89	8	0.76	39	0.89	14	0.88	15	0.86	8	0.78
BG2/cBG3	NA	NA	NA	NA	22	0.94	7	0.94	18	0.83	14	0.61
BG3/cBG1	17	0.83	9	0.73	27	0.92	14	0.88	17	0.84	10	0.72
BG3/cBG2	18	0.82	10	0.70	39	0.89	15	0.87	13	0.88	7	0.81
BG3/cBG3	NA	NA	NA	NA	26	0.92	11	0.90	18	0.83	11	0.69
BG4/cBG1	14	0.86	9	0.73	28	0.92	14	0.88	19	0.82	10	0.72
BG4/cBG2	14	0.86	9	0.73	43	0.87	19	0.83	17	0.84	11	0.69
BG4/ cBG3	NA	NA	NA	NA	23	0.93	9	0.92	18	0.83	14	0.61
BG5/cBG1	11	0.89	9	0.73	28	0.92	15	0.87	19	0.82	10	0.72
BG5/cBG2	15	0.85	12	0.64	41	0.88	19	0.83	17	0.84	9	0.75
BG5/cBG3	NA	NA	NA	NA	25	0.93	11	0.90	17	0.84	13	0.64

Table 2.2. Comparison of exons 1- 3 of turkey BG genes (*BG1-5*) to the syntenic chicken B-G genes (*cBG1-3*). For each gene comparison the number of nucleotide (bp) and amino acid (AA) differences and similarities are given.

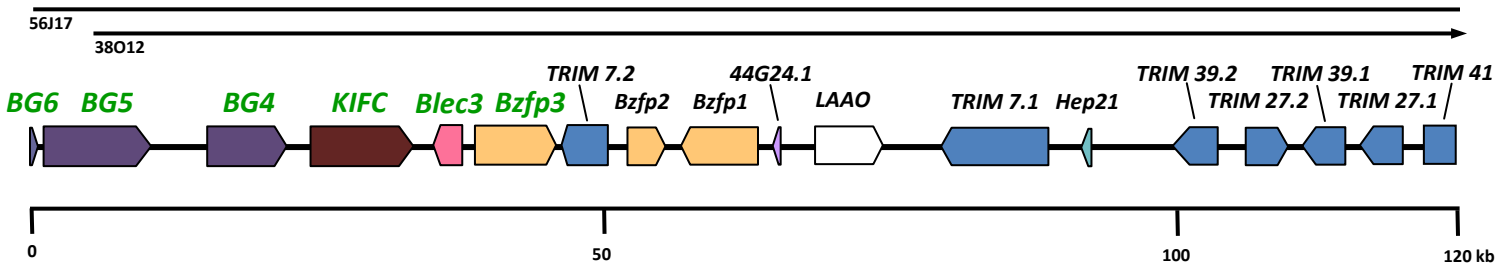


Figure 2.1. Organization of assembled BAC clone, 56J17 (Genbank HQ008883). Newly sequenced genes are labeled in green. Genes overlapping with the previously sequenced *MHC-B* locus (DQ993255) are labeled in black.

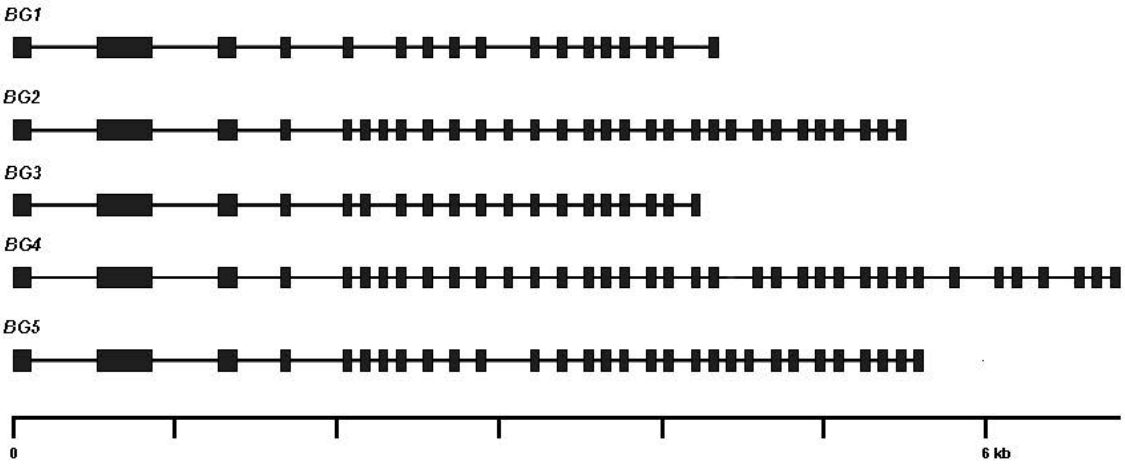


Figure 2.2. Exon structures of sequenced turkey B-G genes.

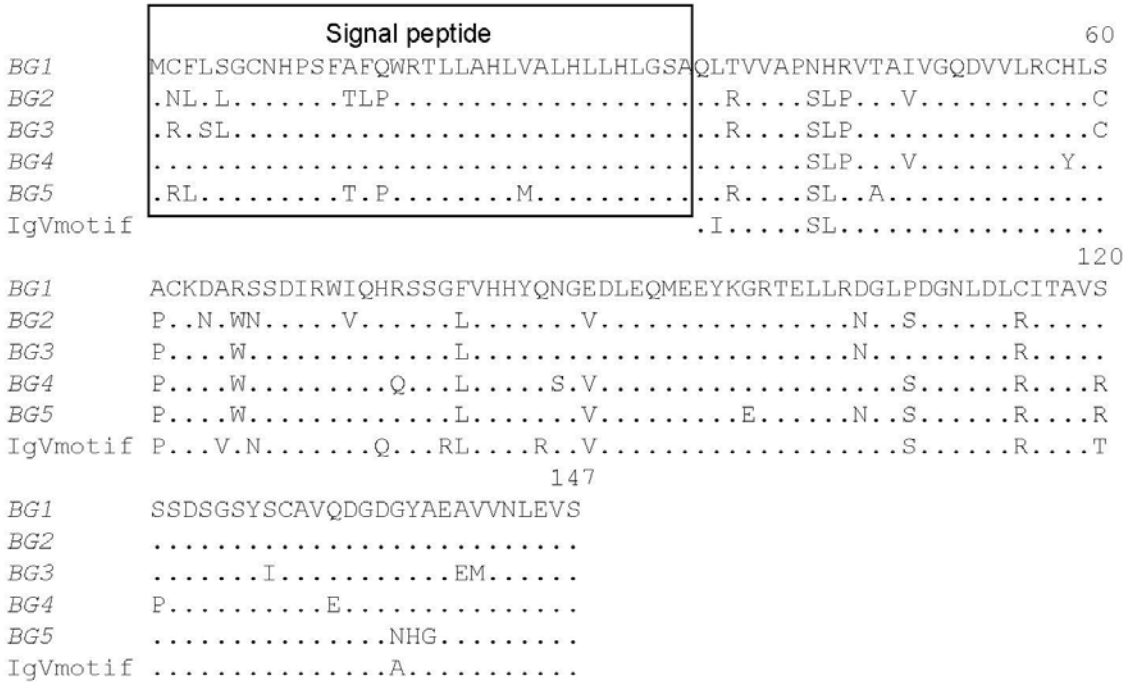


Figure 2.3. Predicted amino acid sequences of the ectodomain (exons 1 and 2) of sequenced turkey B-G genes aligned to *BG1*. The B-G signal peptide region as defined by Miller et al., 1991 and confirmed by SignalP 4.0 (Petersen et al., 2011) is indicated and the chicken IgV motif (Miller et al., 1991) is included for comparison.

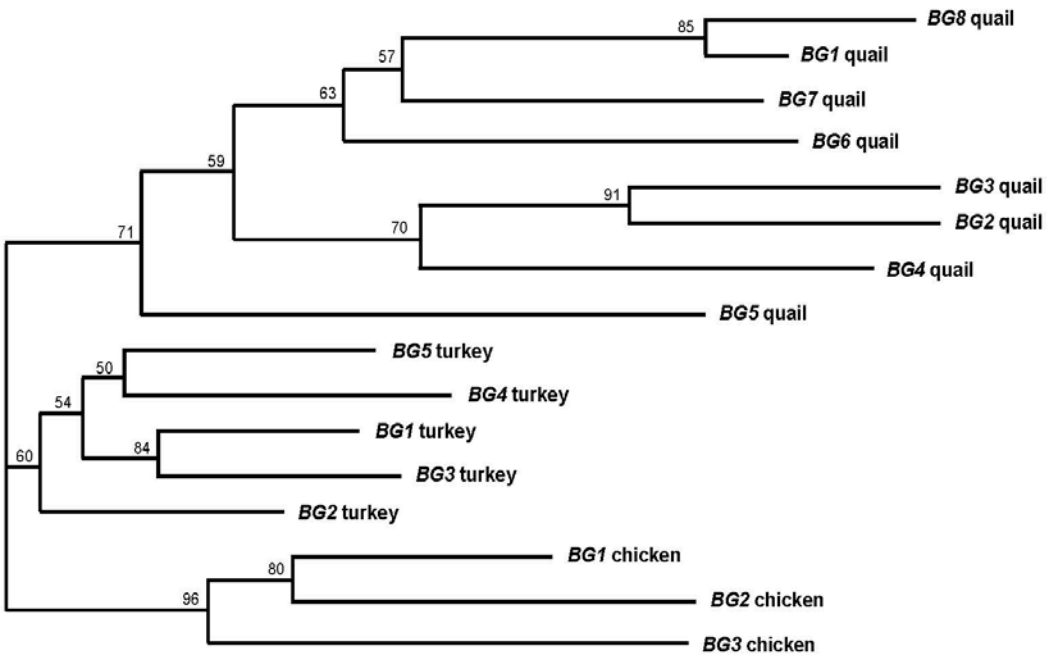


Figure 2.4. Phylogenetic analysis of sequenced B-G genes from galliform species based on nucleotide sequences of exons 1 through 3. Bootstrap percentages resulting from 1000 iterations are given at each node of the unrooted Neighbor Joining tree.

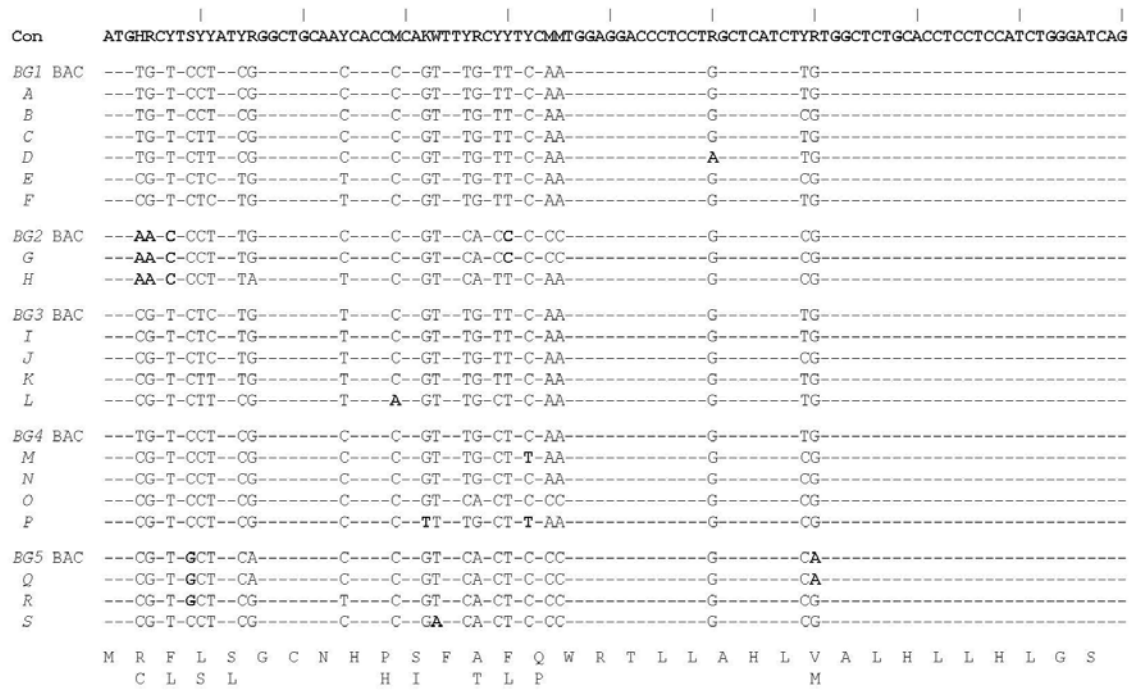


Figure 2.5. Aligned exon 1 alleles of *BG1-5* show clusters of variation spanning nucleotides 4-15 and 32-45 that cause predicted amino acid differences. Consensus nucleotide and amino acid sequences of the coding region are given above and below, respectively. For each locus, alleles observed in the sequence survey are listed below the original BAC sequence. Periods indicate identical bases and locus-specific SNPs at each gene are indicated in bold.

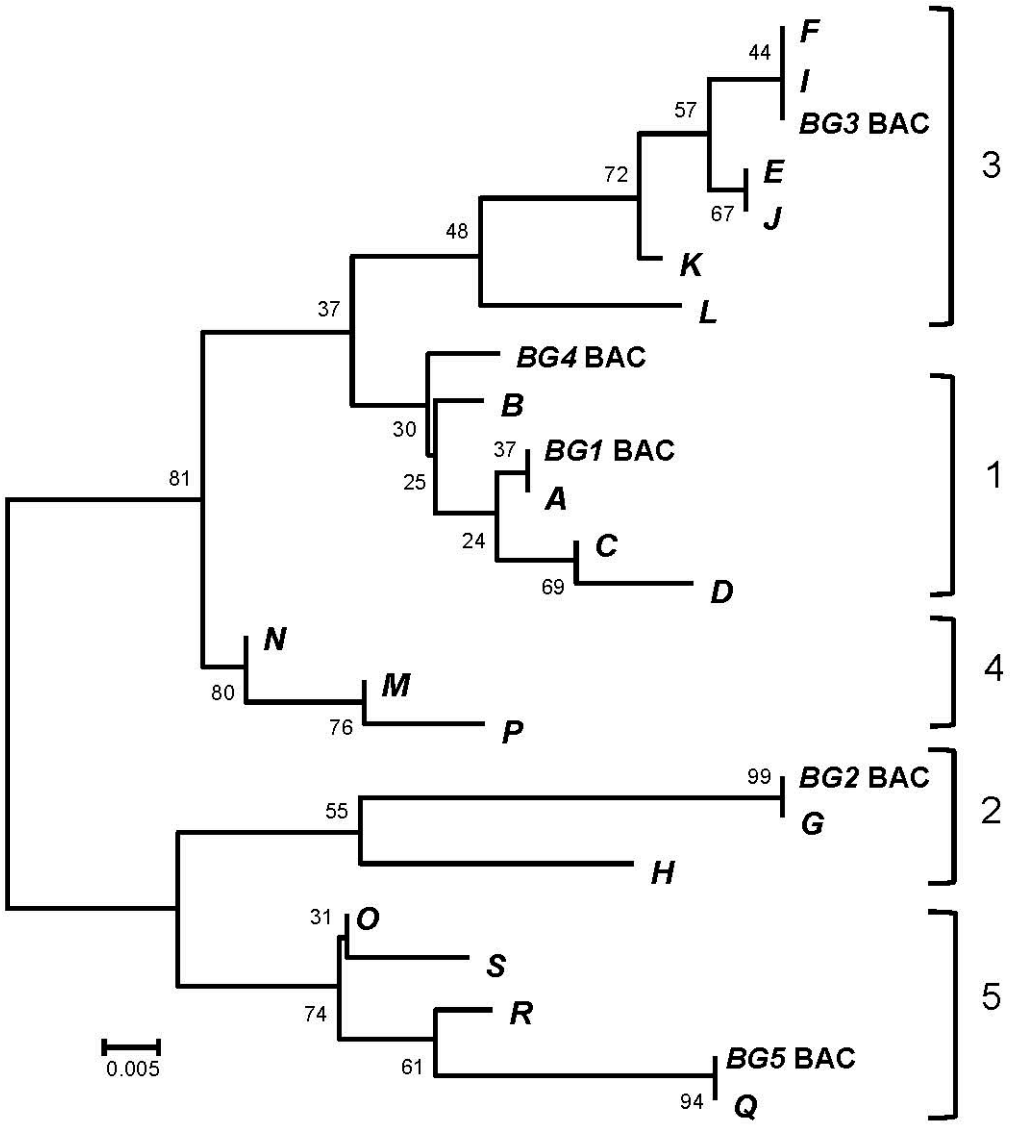
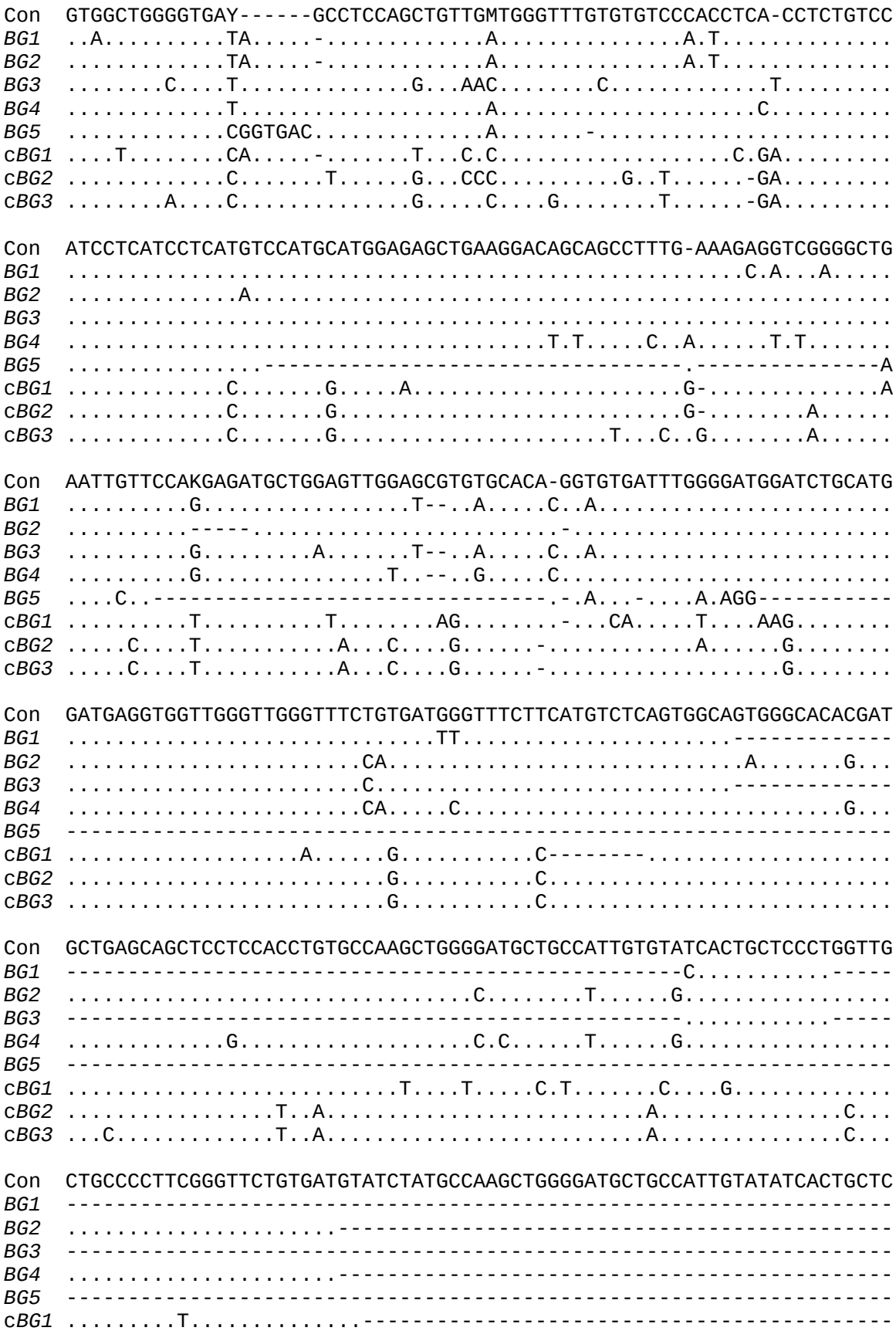
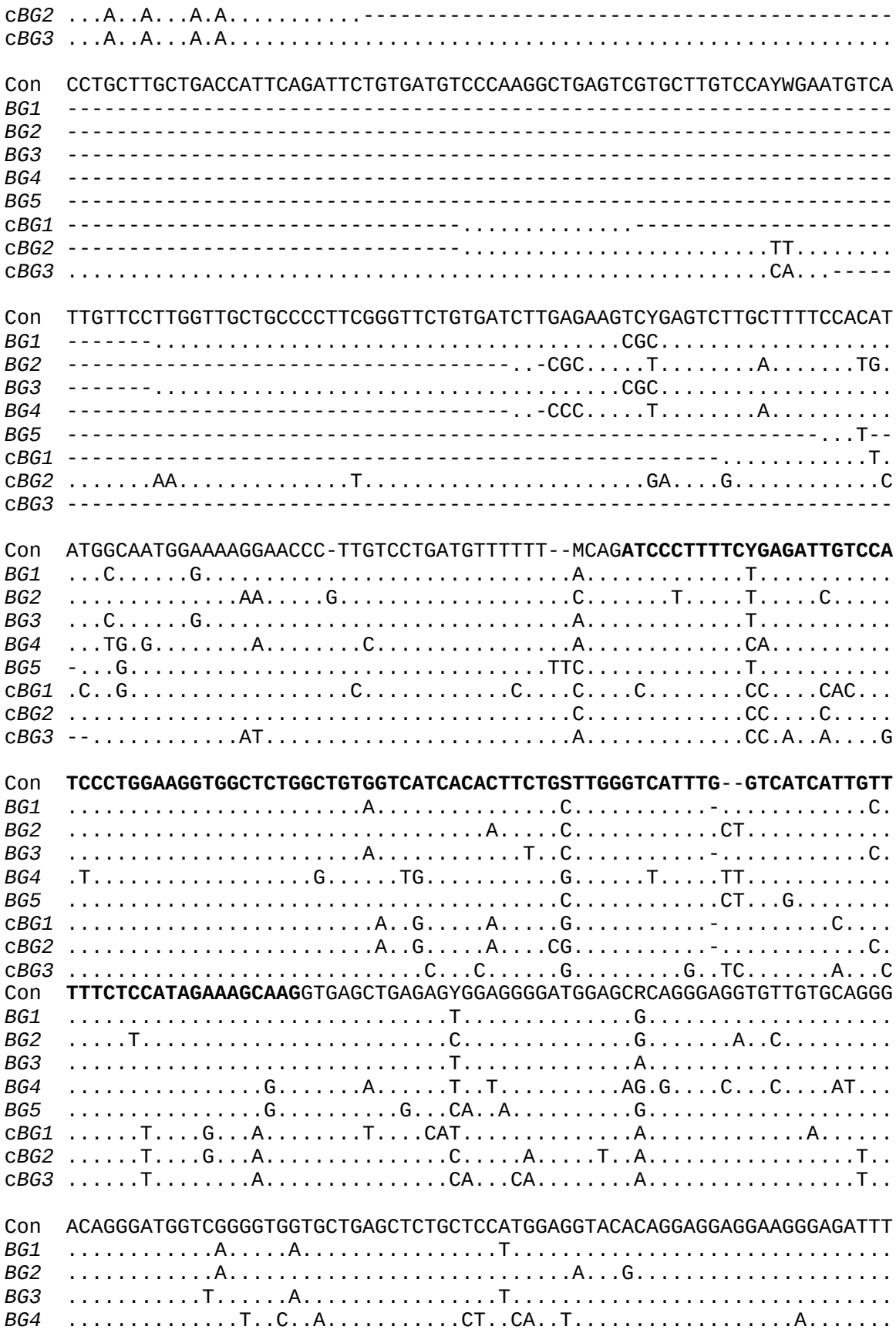


Figure 2.6. Phylogenetic analysis of B-G exon 1 alleles identified in a set of genetically diverse turkeys. The unrooted Neighbor Joining tree is based on aligned nucleotide sequence of the coding region (Fig. 5). Bootstrap percentages resulting from 1000 iterations are given at each node.

```
                                                                                                                    72
Con   ATGCGCTTCCTATTGGGCTGCAACCACCCCAGTTTCGCCTTCCCCTGGAGGACCCTCCTGGCTCATC
BG1   ...T.........C.....................T..T....AA.....................
BG2   ...AA.C.............................A..C.........................
BG3   .........TC............T..........T..T....AA.....................
BG4   ...T.........C.....................T.......AA.....................
BG5   ........G....CA.....................A.........................
cBG1  ....A...T.........................A..C.T.................C..T...
cBG2  T.T.A...................A.............C.................A..CT.T...

Con   TCGTGGCTCTGCACCTCCTCCATCTGGGATCAGGTAGGGGTCCTGTGGGGCTGCTGTGCCGGCCACA
BG1   .T.........................................G.....................
BG2   ....................................................................
BG3   .T....................................................................
BG4   .T....................................................................
BG5   ..A...................................................................
cBG1  ......................G.C.................................T.G....
cBG2  ......................G.......................CA..........T.G....

Con   GCTGTTGCTGTGGGGTGGGGGAGCGGCCATGGGGCAGGGAGGACCCAAGCCCAGCACCCACACACAC
BG1   ........................T.....................................
BG2   ....................................................................
BG3   ....................................................................
BG4   ............A.....................................................G......
BG5   ....................................................................
cBG1  .G.....................A.........G........G..............GC.CTG.
cBG2  .G.....................A......A............G..............-.CTG.


Con   TTGGGTTTCACTTTCACTTT-GGTTATAYCATGAAA-GACGCCATTYTGGGTAGAATTTCTGTCTCT
BG1   ...................G......TC.......T..-.TG...T.................T.
BG2   ...................T..A.G..G.................C...................
BG3   .......................C.G..T.........G.......C...................
BG4   ............A.......T..G....T.................A..............C....
BG5   ..............TG...CC-.C.G..T.................C........T...........
cBG1  ............................TC...........A.....T...................
cBG2  ............................TC...........A.....T...................

Con   TCTCCACCTCCACCACACGCTGTCAGTGGGCTCCCACCCACACAAGTCCTTCCCCGCTCCCTCCTGC
BG1   ....................................................................
BG2   ....................................................................
BG3   C...................................................................
BG4   ....................................................................
BG5   ....................................................................
cBG1  ...T..............G...G............T..G..G..A.........A..........A
cBG2  ...T..............G...G..............C..GC.A.........A..........A

Con   TCCCTCTCCAGTGCTGTTACATGGGATGGAGCACACACCAACTCACCCTGTGCCGCTCCATGCCCCC
BG1   ....................................................................
BG2   ....................................................................
BG3   ....................................................................
BG4   ....................................................................
BG5   ...................................................T................
cBG1  ......C...C....T.....................................................
cBG2  .....TC...C....T.....CC..............................................
```

```
Con   ACATTCACACAGCCACCATCTCACCGTCTCTCCGTGCCCTTCTCATTGCCCAGC**CCAGCTCAGGGTG**
BG1   ....................T.........T............................C....
BG2   ..T.....................................................................
BG3   ...............T.....................................................
BG4   ...........C..T.....................................................CA...
BG5   .....................................................................
cBG1  .....A...................A......T.....................................
cBG2  .....A......A...........A......T.....................................C....

Con   **GTGGCACCGAGCCTCCGTGTCACTGCCATCGTGGGACAGGACGTCGTGCTGCGCTGCCACTTGTCCC**
BG1   ..........A..A...G.....................................................G
BG2   ...............C.........G.....................................GT.
BG3   ...............C.....................................................GT.
BG4   ...............C.........G.............................T.........
BG5   ................G...G.....................................................
cBG1  ............................A............T..T.............G.......
cBG2  ............................A............T..T.............T....G..

Con   **CTTGCAAGGATGCTTGGAGCTCAGACATCAGATGGATCCAGCACCGGTCCTCTGGCCTTGTGCACCA**
BG1   ..............C.....................................................T.........
BG2   ........A.........A.................G.............T...................
BG3   .....................................................................
BG4   ...................................................A.....................
BG5   .....................................................................
cBG1  ...................................................A.......TT..........
cBG2  ........A................T.................TG...........TT..........
cBG3                                          .............C........GA..........

Con   **CTACCAAAATGGAGWGGACCTGGAGCAGATGGAGGAGTATAAAGGGAGGACAGAACTGCTCAGGRAT**
BG1   ..............A....T.......................G.....................G..
BG2   ..............T...T.......................G.....................A..
BG3   ..............A.....................................................A..
BG4   ........G...G.T...T.....................................................G..
BG5   ............G.T.............................A...................A..
cBG1  ...T..........A.....................A..........A................G..
cBG2  .....G........A................ACA..A.............................A.G
cBG3  .......G.C....T................ACA..A..........A................G..

Con   **GGTCTCTCTGATGGAAACCTGGATCTGCGCATCACTGCTGTGAGCTCCTCTGATAGTGGCTCATACA**
BG1   ......C..........T.........T.....................................
BG2   ..G.....................................................................
BG3   ......C.....................................................................T.
BG4   .............................................AC..............C.....
BG5   ..G.....................................................A................C....
cBG1  ........................T.....................A.....................
cBG2  .....T.............A...T.........................C..........TG....
cBG3  ........................T.............................C........G....

Con   **GCTGTGCTGTGCAAGA---TGGTGATGGCTATGCAGAAGCTGTGGTGAACCTGGAGGTGTCAG**GTCA
BG1   .....................................................................
BG2   .....................................................................
BG3   T...C.................................AGA..........A.............
BG4   ..........G.....................................................................
BG5   ............................AA.C...G.....................................
cBG1  ..............G.......................C.........G...................
cBG2  ................CGG..A...................T.....G.G.................
cBG3  ......T...T.....................................G.................
```


Con GTGGCTGGGGTGAY------GCCTCCAGCTGTTGMTGGGTTTGTGTGTCCCACCTCA-CCTCTGTCC
BG1 ..A.....TA.....-.....A.....A.T.....
BG2 .....TA.....-.....A.....A.T.....
BG3 .....C....T.....G...AAC.....C.....T.....
BG4 .....T.....A.....C.....
BG5 .....CGGTGAC.....A.....-.....
cBG1 ....T.....CA.....-.....T...C.C.....C.GA.....
cBG2 .....C.....T.....G...CCC.....G..T.....-GA.....
cBG3 .....A....C.....G.....C....G.....T.....-GA.....
Con ATCCTCATCCTCATGTCCATGCATGGAGAGCTGAAGGACAGCAGCCTTTG-AAAGAGGTCGGGGCTG
BG1 .....C.A...A.....
BG2 .....A.....
BG3 .....
BG4 .....T.T.....C..A.....T.T.....
BG5 .....-----.-----A
cBG1 .....C.....G.....A.....G-.....A
cBG2 .....C.....G.....G-.....A.....
cBG3 .....C.....G.....T...C..G.....A.....
Con AATTGTTCCAKGAGATGCTGGAGTTGGAGCGTGTGCACA-GGTGTGATTTGGGGATGGATCTGCATG
BG1 .....G.....T--..A.....C..A.....
BG2 .....-----.....-.....
BG3 .....G.....A.....T--..A.....C..A.....
BG4 .....G.....T..---..G.....C.....
BG5 ....C..-----.-..A...-....A.AGG-----
cBG1 .....T.....T.....AG.....-...CA.....T....AAG.....
cBG2 .....C....T.....A...C....G.....-.....A.....G.....
cBG3 .....C....T.....A...C....G.....-.....G.....
Con GATGAGGTGGTTGGGTTGGGTTTCTGTGATGGGTTTCTTCATGTCTCAGTGGCAGTGGGCACACGAT
BG1 .....TT.....-----
BG2 .....CA.....A.....G...
BG3 .....C.....-----
BG4 .....CA.....C.....G...
BG5 -----
cBG1 .....A.....G.....C-----.....
cBG2 .....G.....C.....
cBG3 .....G.....C.....
Con GCTGAGCAGCTCCTCCACCTGTGCCAAGCTGGGGATGCTGCCATTGTGTATCACTGCTCCCTGGTTG
BG1 -----C.....
BG2 .....C.....T.....G.....
BG3 -----.....
BG4 .....G.....C.C.....T.....G.....
BG5 -----
cBG1 .....T....T.....C.T.....C....G.....
cBG2 .....T..A.....A.....C...
cBG3 ...C.....T..A.....A.....C...
Con CTGCCCCTTCGGGTTCTGTGATGTATCTATGCCAAGCTGGGGATGCTGCCATTGTATATCACTGCTC
BG1 -----
BG2 .....-----
BG3 -----
BG4 .....-----
BG5 -----
cBG1 .....T.....-----


cBG2 ...A..A...A.A.....----------
cBG3 ...A..A...A.A.....
Con CCTGCTTGCTGACCATTCAGATTCTGTGATGTCCCAAGGCTGAGTCGTGCTTGTCCAYWGAATGTCA
BG1 ----------
BG2 ----------
BG3 ----------
BG4 ----------
BG5 ----------
cBG1 ----------.....----------
cBG2 ----------.....TT.....
cBG3 .....CA...-----
Con TTGTTCCTTGGTTGCTGCCCCTTCGGGTTCTGTGATCTTGAGAAGTCYGAGTCTTGCTTTTCCACAT
BG1 -------.....CGC.....
BG2 ----------..-CGC.....T.....A.....TG.
BG3 -------.....CGC.....
BG4 ----------..-CCC.....T.....A.....
BG5 ----------...T--
cBG1 ----------.....T.
cBG2 .....AA.....T.....GA....G.....C
cBG3 ----------
Con ATGGCAATGGAAAAGGAACCC-TTGTCCTGATGTTTTTT--MCAGATCCCTTTTCYGAGATTGTCCA
BG1 ...C.....G.....A.....T.....
BG2 .....AA.....G.....C.....T.....T.....C.....
BG3 ...C.....G.....A.....T.....
BG4 ...TG.G.....A.....C.....A.....CA.....
BG5 -...G.....TTC.....T.....
cBG1 .C..G.....C.....C....C....C.....CC....CAC...
cBG2 .....C.....CC....C.....
cBG3 --.....AT.....A.....CC.A..A....G
Con TCCCTGGAAGGTGGCTCTGGCTGTGGTCATCACACTTCTGSTTGGGTCATTTG--GTCATCATTGTT
BG1 .....A.....C.....-.....C.
BG2 .....A.....C.....CT.....
BG3 .....A.....T..C.....-.....C.
BG4 .T.....G.....TG.....G.....T.....TT.....
BG5 .....C.....CT...G.....
cBG1 .....A..G.....A.....G.....-.....C....
cBG2 .....A..G.....A....CG.....-.....C.
cBG3 .....C...C.....G.....G..TC.....A...C
Con TTTCTCCATAGAAAGCAAGGTGAGCTGAGAGYGGAGGGGATGGAGCRCAGGGAGGTGTTGTGCAGGG
BG1 .....T.....G.....
BG2 .....T.....C.....G.....A..C.....
BG3 .....T.....A.....
BG4 .....G.....A.....T..T.....AG.G....C...C....AT...
BG5 .....G.....G...CA..A.....G.....
cBG1 .....T....G...A.....T....CAT.....A.....A.....
cBG2 .....T....G...A.....C.....A.....T..A.....T..
cBG3 .....T.....A.....CA...CA.....A.....T..
Con ACAGGGATGGTCGGGGTGGTGCTGAGCTCTGCTCCATGGAGGTACACAGGAGGAGGAAGGGAGATTT
BG1 .....A.....A.....T.....
BG2 .....A.....A...G.....
BG3 .....T.....A.....T.....
BG4 .....T..C..A.....CT..CA..T.....A.....

```
BG5  ............A..C.A.....T.T--A..---.............................
cBG1 .T.............................G...G........TG.........---.........
cBG2 .............................G....C.........G.......T......T..C..
cBG3 ..........T..................G....C.......T.....................

Con  TTCCTGGGATTCCCA--GTGCTCATTAAATAACGTTTGCCTTTYTTTTT-GGGAATGAAGGAAGGGG
BG1  C................G.........T...T..........T........C.........T....
BG2  C............C...C...T..........T...A.....TG...T................
BG3  C..................G.......T...T..........T........C.........T....
BG4  .-A................AAATG..............TG..CC.....-..........C.-.A.
BG5  ...............TCA....TG........T..........C....-G.............-...
cBG1 ...TG......GA.......G............A......G..TGA..-.............A...
cBG2 .C.A.............................A....A....C.....................
cBG3 ......A.........................T.........C.....................

Con  AAAAARG--AAAGGGGTGTGGGTGGGAAGATAGGAAYGTGGCTGGAYYGTGGGGCAGGTGGAAAGGT
BG1  ....GG-.............A...............A.........TT..................
BG2  ....TGT.................A..........A.........TC............A......
BG3  ....GG-.............A...............T.........TT..................
BG4  -TT..AAAT.............A.............C.........GC..................
BG5  GGT..A.G.-.....-.........A..........C.........GT..................
cBG1 ----.A-........-......C..A..........T....G....CC.............G.....
cBG2 ......-.G.----------.................T....G....CT..................
cBG3 ......-.G.....T.....................C....G.C..CCA...A.............

Con  CCAGACCCTCTGGAGACGTCCCCAGAAACCAAGCTGCCCTGCTGACCACCTCTTCCTCTGCTTTGTT
BG1  .T...A--...CA...T..T..A...G.............G.......T.....G..C...C....
BG2  ........A...............C...T.....................................
BG3  .T...A--...CA......T..A...G.............G.......T.................
BG4  ........A......T........C............T........G...................
BG5  ........A......T........C...........T.....................A.........
cBG1 ..............................................G..A..T...........
cBG2 ..T........T...G.A..A.....................................CA.A.....
cBG3 ......A...CT.....A..A....T.......................C....T..C........

Con  TTCCAG
BG1  ......
BG2  ...T..
BG3  ......
BG4  ......
BG5  ......
cBG1 ......
cBG2 ......
cBG3 ......
```

Figure 2.7. Aligned contiguous sequences of turkey and chicken BG genes through the start of exon 4. Sequences are as reported in Genbank accessions DQ993255, HQ008883, and AB268588**.** Bold font denotes exons, periods indicate identical bases and dashes denote alignment gaps.

Chapter 3

Estimating Genetic Diversity at the MHC in a Population of Introduced Wild Turkey Highly Polymorphic at the B-G Region.

Miranda M. Bauer[1], Marcia M. Miller[2], W. Elwood Briles[3] and Kent M. Reed[1]

[1]Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, St Paul, MN 55108

[2]Division of Molecular Biology, City of Hope, Beckman Research Institute, Duarte CA, USA

[3]Department of Biological Sciences, Northern Illinois University, Dekalb IL, USA

Abstract

Genetic variation in the major histocompatibility complex (MHC) is known to affect disease resistance in many species**.** The antigen presenting class I and class II molecules located in this region of the genome are particularly known for their polymorphic nature, and investigation of the genetic diversity of these antigens has been performed on many species. Thus far, only limited studies of MHC haplotype variation in the turkey have been attempted, with variation assessed serologically, genetically (by restriction fragment length polymorphism, RFLP) and to a limited extent, by direct nucleotide sequencing. This study was designed to investigate MHC diversity in a collection of wild turkeys (*Meleagris gallopavo silvestris*) collected during population expansion following reintroduction in southern Wisconsin, USA. Single nucleotide polymorphisms (SNPs) were identified by sequencing select MHC class I gene regions in individuals with distinct MHC class IIβ Southern blot phenotypes. All individuals were subsequently genotyped by PCR/RFLP for haplotype analysis. To further characterize genetic diversity of the sampled birds, a portion of the mitochondrial D-loop was also sequenced. Results show that diversity predictions based only on class IIβ Southern blot analysis underestimate the number of MHC haplotypes.

INTRODUCTION

The major histocompatibility complex (MHC) plays a vital role in vertebrate immune response and disease resistance, mainly through facilitation of communication between immune cells, such as B cells, T cells and antigen presenting cells. In mammals, the MHC is typically divided into three regions, class I, class II and class III, all of which contain genes related to immune response. Class I and class II molecules are well known for their role in presenting processed peptides to T cells. The extent of variation in MHC genes within and between individuals and populations is indicative of the spectrum of pathogens to which the immune system is able to effectively respond, making it a critical area of study relative to individual fitness and the adaptive genetic variation in a population (Alcaide et al., 2010; Jarvi et al., 2004; Spurgin and Richardson, 2010).

Selection pressures affecting the evolution of MHC genes is a subject of much scrutiny, and studies purported several forces acting on MHC gene evolution, including balancing selection, pathogen-mediated selection and genetic drift in a variety of species (Alcaide et al., 2010; Bollmer et al., 2011; Jarvi et al., 2004; Miller and Lambert, 2004; Miller et al., 2010). Evidence suggests that the MHC is under different selective forces compared to neutral markers such as microsatellites and the mitochondrial D-loop, which are primarily affected by genetic drift (Aguilar et al., 2004; Sommer S., 2005; Piertney and Oliver, 2006). Population history may influence the dominant selective force on MHC alleles (Spurgin and Richardson, 2010). Balancing selection appears to be the primary force in populations having experienced genetic bottlenecks and subsequent founder effects. In these populations, MHC variation is high, in contrast to nuclear

markers and mitochondrial genomes that show reduced variation in response to the same conditions (Miller and Lambert, 2004). Population bottlenecks alone (without subsequent founder effects) tend to be primarily influenced by genetic drift, resulting in reduced MHC gene variability (Miller and Lambert, 2004).

Organization and copy number of MHC genes are variable between and within avian orders. Studies examining MHC loci in songbirds (Order: Passeriformes) and seabirds (Order: Procellariiformes) have found decreased gene density in the MHC accompanied with increased copy number of MHC genes compared to ground-feeding gamebirds of the order galliformes (Aguilar et al., 2006; Bonneaud et al., 2004; Hosomichi et al., 2006; Shiina et al., 2007; Standh et al., 2011; Westerdahl et al., 2000; Westerdahl et al., 2004). Social species show a trend towards higher diversity at the MHC (Aguilar et al., 2004), whereas more solitary species show less diversity regardless of population expansions after bottlenecks (Miller et al., 2010). Low intraspecific MHC diversity may increase a population's susceptibility to epidemics (Siddel et al., 2007).

In some species, allelic diversity at the MHC is associated with mate selection, with an apparent high frequency of disassortative mating (von Schantz et al., 1996; Aspanius et al., 1997; Freemen et al., 2003). For example, in the ring-necked pheasant, males with more diverse MHC alleles have more elaborate ornamentation, which is preferred by females (von Schantz et al., 1996). This supports an indirect correlation of mating preferences and increased allele diversity of MHC genes. In contrast, in the great snipe, where there is no difference of ornamentation between males and females, no correlation between MHC class IIβ allelic richness and female mate preference was

observed. Instead, males with one particular MHC class IIβ allele profile had greater mating success than those with other class II profiles (Ekblom et al., 2004). The correlation between parental mating preference and offspring disease resistance is unknown.

Haplotyping individuals at the MHC, a region containing genes in high linkage disequilibrium, is a method for characterizing immunological diversity. Certain haplotypes have a known association with disease resistance and antibody response in both wild and domestic birds (Alcaide et al., 2010; Bacon et al., 2001; Liu et al., 2002; Haeri et al., 2004). Previous studies have used a combination of methods for characterizing MHC haplotypes, mainly focusing on comparing variation in classical class I and class II genes (Afanassieff et al., 2001; Jarvi et al., 1999; Goto et al., 2002).

Traditional methods used to detect variations in avian MHC haplotypes, including serological and Southern blotting are confounded by the use of nonspecific antibodies and nonspecific probes directed towards class I, class II and B-G antigens. Comparing haplotypes using techniques such as RSCA (reference strand conformation analysis), SSCP (single strand conformation polymorphism) and DGGE (denaturing gradient gel electrophoresis) are more cost effective but do not adequately address problems associated with multiple loci in gene families (Agudo et al., 2011; Goto et al., 2002; Jarvi et al., 2004; Sacco et al., 2001; Zhu et al., 1995; Zhu et al., 1996). RSCA has recently gained popularity as a method for genotyping MHC loci because this method minimizes co-migration of same sized alleles. The main limitation of RSCA is the inability to

assign sequence variants to a particular locus (Brower et al., 2010; Strand and Hoglund, 2011).

The most detailed approach used to assess MHC diversity is direct sequencing, which can identify all variations between individuals of a single locus (Ewald and Livant, 2004; Hosomichi et al., 2008). A limitation of this method is the time and cost of examining long stretches of DNA, particularly for large sample sizes, although several studies advocate for next-generation sequencing approaches (Babik et al., 2009; Zagalska-Neubauer et al., 2010). This method reduces the time needed to genotype large samples but introduces other limitations such as sequencing errors, generation of PCR and sequence artifacts and chimeras. Recently Chaves et al. (2011) demonstrated a locus-wide approach to measuring MHC diversity. This approach avoids the inherent problems associated with surveying the polymorphic, multi-copy MHC genes by using a MHC-*B* locus-wide single nucleotide polymorphism (SNP) genotyping approach.

In certain regions of the United States, wild turkeys (*Meleagris gallopavo)* have passed through a significant bottleneck or even extermination. In areas of regional extirpation, reintroduction efforts involving the translocation of turkeys from areas where they were more prevalent have allowed for recent population expansions. Few studies have documented genetic diversity in wild turkeys. Mock et al. (2001) examined genetic diversity of two native subspecies (Gould's and Merriam's) using mitochondrial (mtDNA) sequences in combination with amplified fragment length polymorphisms (AFLPs) and microsatellite markers. This study identified mitotype, AFLP pattern and microsatellite marker differences between the two subspecies as well as established

AFLP analysis as an equivalent alternative to microsatellite analysis for subspecies-level investigations. The genetic distinctness of the turkey subspecies populating the United States was demonstrated using the same molecular techniques (Mock et al., 2002). The present study is the first implementation of MHC SNP genotyping in a population of wild birds from a localized area to determine the level of MHC variation in contrast to estimates based on more traditional Southern blotting and compared to neutral markers.

It is hypothesized that SNP genotyping will give a more detailed picture of MHC diversity than Southern blots of the class II region. Furthermore, substantial gene flow is expected between the groups of turkeys used in this study, given the limited geographic range of the sampled localities. This is the first study measuring MHC genetic diversity in a population of turkeys with known geographical location. Given the success of the reintroduction effort and the effect MHC diversity has on population fitness, it is hypothesized that the Wisconsin turkeys exhibit greater MHC diversity than that of neutral markers in the study groups.

MATERIALS AND METHODS:

Turkey Samples

The Eastern subspecies (*M. g. silvestris*) has the largest geographical distribution occurring from the east coast of the US through the Great Plains. The subspecies was extirpated from many parts of its historical range as the result of severe winters in combination with habitat loss during settlement and unregulated hunting. Turkeys were absent from Wisconsin from 1881 until the 1950s, when 827 Pennsylvania turkeys were

released in Monroe and Juneau counties, located in the central part of the state (Wise S., 1987). These reintroduced turkeys consisted of wild gobblers mated with game-farm hens. The reintroduction effort produced a stagnant population of turkeys in regards to both number and range until a severe winter in 1969 decreased the population to 75 birds. A different approach was implemented from 1976-1985 when 334 wild birds from Missouri were trapped and transplanted into 9 counties in the central and southeastern portions of the state. The population expanded to a current level of more than 320,000 individuals ranging throughout the state, with some ingression of individuals from the neighboring states of Minnesota and Iowa. The genetic history of these birds makes them a valuable resource for population genetics analysis. This study examines MHC haplotypes within and among 5 groups of wild turkeys collected from a limited area of Wisconsin in 1988-1990.

The turkey samples examined in this study were obtained from the Wisconsin Department of Natural Resources and collected from a region of southern Wisconsin from 1988-1990 between December and February. Blood samples were collected from birds live-trapped in five different locations in Wisconsin encompassing a 30-35 mile radius (Table 3.1; Figure 3.3). Ten birds were sampled in Iowa County, 12 from Richland County, 11 from Crawford County and 33 birds from two locations in Grant County (Grant 1 = 16 and Grant 2 = 17). The birds in Iowa County originated from a group of Missouri wild turkeys consisting of two groups of 5 toms/15 hens released at two sites. The other sampled individuals originated from turkeys trapped in Iowa county

and later transplanted to sites in the other 3 counties. Birds from different locations in Iowa County were included in each transplanted group of 5 toms/15 hens.

Southern Blotting

Genomic DNA (10ug) was extracted from blood samples of sampled wild turkeys, and MHC variation was initially examined by Southern blotting. Genomic DNA was digested in overnight incubations with *Pvu* II, size fractioned by gel electrophoresis, transferred to nitrocellulose membranes and probed with a class IIβ or B-G (bg11) probe as described in Miller et al., 1994. Briefly, membranes were hybridized overnight at 65°C, subjected to stringency washes at 65°C, and exposed to autoradiograph film. Banding patterns were visually compared between individuals. Samples with the same banding pattern were considered to have the same MHC types and were grouped together for analysis (Figure 3.1; Table 3.1).

MHC Gene Sequencing

Two individuals from the 10 most common class IIβ groups were chosen for sequencing and SNP discovery. In total, four genes flanking the class IIβ region: *TRIM41*, *BTN1*, *BTN2* and *TAP1*, were partially sequenced. PCR was performed on genomic DNA with Promega Master Mix (Promega Corp) using oligonucleotide primers (Table 3.2) designed from known turkey gene sequences (Chaves et al., 2009). The following reaction parameters were used: 94ºC for 5 min and 35 cycles of 94ºC for 30 sec, 58ºC for 30 sec and 72ºC for 1 min/kb with a final extension of 72ºC for 15 min. PCR products were

visualized by electrophoresis on a 1% agarose gel. Exo-sapit (USB Corp) was used for amplicon purification. Briefly, 1µl of ExoSAP-IT (USB Corp) was added to 4µl PCR product and incubated at 37°C for 15 min and 80°C for 15 min. Purified DNA was sent to the Biomedical Genomics Center, University of Minnesota for Sanger sequencing. Sequences were aligned and compared using Sequencher 4.9 software (Gene Codes Corp., Ann Arbor, MI) to identify SNPs.

SNPs throughout coding and noncoding regions encompassing exons 4 through 7 of *TRIM41* (1118 bp), introns 2 through 4 of *BTN1* (504 bp), an 807 bp region upstream of *BTN2* exon 1 and in regions between exon 1 through intron 2 of *TAP1* (401 bp) were identified (Table 3.2). Sequence variants were called if there were more than two heterozygous birds at one position, thereby discounting single variant individuals. SNPs occurring with the recognition sequence of restriction enzymes were identified with NEB cutter v2.0 (http://tools.neb.com/NEBcutter2/index. php).

Exon 1 through the first 221bp of exon 2 of *BG5* was amplified by PCR from genomic DNA for all 65 birds using the primer combinations 5'-GACCCCTTGTCGTGA GACTG -3' and 5'-ACCACAGCTTCTCCATGGTT-3'. Sequences were aligned and SNPs were identified throughout the region. PCR products from ten birds with suspected mixed sequences were cloned into PGEM-TEZ vector following the manufacturer's protocol to verify amplification of a single product. Briefly, PCR products were purified and ligated overnight incubation at 4ºC. Ligation reactions were transformed into DH5α competent cells. Between three and five colonies per individual were picked, grown in 5 mL LB media overnight and purified using the Qiaprep Spin Miniprep kit (Qiagen).

Purified plasmids were digested with *EcoR*I and clones positive for inserts were sequenced.

SNP genotyping

One SNP for each of the four genes flanking the class IIβ genes was chosen for genotyping on each of the 66 individuals by PCR/RFLP. PCR was performed on all individuals using the same oligonucleotide primers used for SNP discovery (Table 3.2). Reactions were performed on genomic DNA with the following reaction parameters: 94ºC for 5 min and 35 cycles of 94ºC for 30 sec, 58ºC for 30 sec and 72ºC for 1 min/kb with a final extension of 72ºC for 15 min. PCR products were visualized on a 1% agarose gel. The enzymes used for restriction digest are as follows: *Msp*I for *TRIM 41*, *Msl*I for *BTN1*, *Nco*I for *BTN2* and *Fok*I for *TAP1*. PCR products were digested in a 25µl reaction at conditions specified by the enzyme manufacturer. PCR/RFLP fragments were examined by gel electrophoresis on a 1% agarose gel, and individual genotypes were manually scored. Fixation indexes (Fst values; Nei, 1986) and mean heterozygosities were calculated and departure from Hardy-Weinberg equilibrium within each group at each locus was tested using ARLEQUIN VS. 3.5. Haplotypes were imputed using the software program, PHASE v2.1.1 (Stephens et al., 2001; Stephens and Donnelly, 2003). For each test, the dememorization number was set to 100000 and the number of iterations per batch to 1000000. Allele frequencies for each SNP were calculated using an expectation maximization (EM) algorithm.

Mitochondrial Haplotyping

To further characterize genetic diversity of the sampled birds, nucleotides 89 through 465 (380bp) in the 5' region of the mitochondrial D-loop, the most variable portion of the control region in vertebrates, was sequenced. Primers LND185 (5'-ACGGCTTGAAAA GCCATTGTTG-3') and HCR684 (5'-GGGGTTACCTCACAGGTGGC-3') were used to amplify the region of interest based on turkey sequence data (Mock et al., 2001). PCR was performed on genomic DNA of 65 birds with HotStart taq polymerase at the following parameters: 94ºC for 5 min and 35 cycles of 94ºC for 30 sec, 58ºC for 30 sec and 72ºC for 1 min/kb with a final extension of 72ºC for 15 min. The PCR products were visualized by gel electrophoresis on a 2% agarose gel. ExoSAP-IT was used for amplicon purification. Briefly, 1µl of ExoSAP-IT was added to 4µl PCR product and incubated at 37°C for 15 min and 80°C for 15 min.

Purified DNA was sent to the BMGC, University of Minnesota for Sanger sequencing. Sequences were aligned in Sequencher 4.9 (Gene Codes Corp., Ann Arbor, MI) and SNPs were identified. PHASE v2.1.1 (Stephens et al., 2001; Stephens and Scheet, 2005) was used to determine mitotype frequencies for each group of birds. DNASP vs 5.10 was used to calculate haplotype diversity, nucleotide diversity and Tajima's *D* (Librido and Rozas, 2009). Chi-square analysis was used to test for independence between Southern MHC group and mitotype as well as between SNP haplotype and mitotype. The p-value was computed by simulation with 10,000 simulations run.

RESULTS

Class IIβ variation (Southern Blotting)

Examination of the 66 sampled birds using Southern blot analyses identified thirteen distinct class IIβ groups (Groups A-M, Figure 3.1A; Figure 3.3; Table 3.1) and 16 birds with unique Southern banding patterns (Group N). Because inheritance of specific banding patterns is not known, individual banding patterns will from here on be referred to as Southern blot phenotypes. Group C was the most common phenotype in the sampled birds, with 12 of the 65 (20%) birds showing this class IIβ phenotype. Eleven of these 13 birds were sampled from the Grant 1 locality. The second most frequent phenotype observed was that of Group E (11%). This phenotype was found in birds located in the Iowa, Richland and Crawford localities but not at either of the Grant localities. Phenotypes A and B were observed in 4 of the 5 groups, with Crawford being the only locality with no birds of these phenotypes. The birds sampled at the Richland and Grant 2 localities had 3 additional phenotypes in common (H, K and L). The greatest frequency of birds with unique phenotypes, 45% and 42% respectively, were from the Crawford and Richland localities.

Eight different phenotypes were observed from the 10 birds sampled in Iowa County, 11 from Richland, 8 from Crawford and 12 from Grant 2. Grant 1 had the least number of class IIβ phenotypes with 6 detected in the 16 sampled birds. Based on these results, two individuals each from the 10 most common class IIβ phenotypes were chosen for sequencing.

SNP Discovery and Genotyping

MHC diversity as suggested by Southern blotting, was tested by examining sequence variation in four genes (*TRIM41*, *BTN1*, *BTN2* and *TAP1*) flanking the class II region. Alignment of gene sequences identified 4 SNPs with alternate alleles present in at least 2 individuals in *TAP1*, 14 in *BTN1*, 8 in *BTN2* and 11 in *TRIM41* (Table 3.2), indicating high polymorphism in this region. The SNPs identified in *TRIM41*, *BTN1* and *BTN2* were located in introns. The 4 SNPs in *TAP1* were located in exon 2. The SNP density across the MHC amplicons ranged from 1 SNP per 100 base pairs (*TAP1*, *BTN2*, *TRIM41*) to 7 SNPs per 252 base pairs (*BTN1*).

PCR/RFLP genotypes at SNPs for each of the four genes were obtained for 65 individuals. Data were not obtained for individual 1460 due to repeated PCR amplification failure at *TRIM41* and *BTN1*. SNP frequencies were compared for each population at each locus and showed locus-dependent differences between localities for the sampled individuals. For example, at the Iowa, Richland and Grant 2 localities, adenine occurred most frequently at the *TRIM41* SNP, compared to guanine for the Crawford and Grant 1 localities. At the *BTN1* SNP, thymine occurred most frequently in all localities except Grant 2 where adenine was more common. At the *BTN2* SNP, guanine occurred more frequently than cytosine in all localities, and at all localities except Crawford, guanine occurred more frequently than cytosine at the *TAP1* SNP (Table 3.3).

Genetic diversity of the sample birds was measured using a locus by locus exact test of Hardy-Weinberg equilibrium. Heterozygosity of all sampled turkeys (Table 3.4)

was significantly lower than expected for a population in H-W equilibrium at *TRIM41* ($p < 0.01$), and *BTN2* ($p < 0.05$), indicating population subdivision between localities. This possibility was further investigated by measuring deviations from H-W at each locus for each locality. None of the SNP loci showed a deviation from H-W at the Iowa locality. There were significantly fewer heterozygotes observed at *TAP1* in Crawford, at *TRIM41* in Richland ($p = 0.03$) and at *BTN2* in Grant 2 ($p = 0.01$). Heterozygosity at *BTN1* and *TAP1* was significantly higher than expected in the Grant 1 group ($p = 0.008$), suggesting selection for diversity acting on these two genes at this locality.

Pairwise Fst values were estimated using allele frequency data to investigate MHC related genetic differentiation between localities (Table 3.5). The largest Fst value, 0.13, occurred between Grant 1 and Grant 2, indicating a moderate level of differentiation between these two localities. The Richland and Grant 1 localities had a low but discernable level of differentiation ($Fst = 0.078$). Fst values between other localities ranged between 0.0 and 0.05, suggesting low levels of differentiation. Estimation of migrants per generation suggest high levels of gene flow between localities, with the lowest Nm value ($Nm = 1.12$) occurring between Grant 1 and Grant 2.

SNP haplotypes

PHASE analysis detected 16 SNP haplotypes in the 65 individuals sampled (Figure 3.4; Table 3.6). Three haplotypes (1, 2 and 5) were shared among the five localities. The locality showing the greatest SNP haplotype diversity was Iowa (16), followed by Grant 1 (15 haplotypes), Grant 2 and Richland (9 each) and Crawford (6). Haplotype 1

(GTGG) was the most common SNP haplotype in three out of the five populations (Iowa, Crawford and Grant 1). This haplotype was most frequent in the Crawford population. Haplotype 2 (AAGG) was most dominant in the remaining two populations, Richland and Grant 2. This is consistent, as their origins are from different populations of birds. Birds from Iowa County originated from Missouri wild-caught turkeys while those from Crawford, Richland and Grant counties were transplanted from localities in Wisconsin. A total of 26 distinct MHC genotypes (SNP haplotype combinations) were represented in the sample birds (Table 3.1). A total of 14 birds were homozygotes: 1 from the Iowa locality, 4 from Richland, 2 each from Crawford and Grant 1 and 5 from Grant 2.

B-G Variation

Comparison of banding patterns on the B-G Southern blots identified unique banding patterns for all 66 individuals. This is likely due to the combined presence of multiple B-G loci and alleles in this gene family, all of which show high sequence homology. One B-G gene in the extended region of the turkey MHC was chosen for partial sequencing based on its location and preliminary data comparing the sequence variability of the five known B-G genes in the turkey MHC (Chapter 3). Of those five genes, *BG5* was found to be most variable in a subset of individuals from this group of birds. A portion of *BG5* was sequenced for all 65 birds. Aligned sequences identified 28 nucleotide variants within the sequenced region. Nineteen sequence differences occurred in exon 1, concentrated between base pairs 8 and 45. Nine sequence differences were found in exon 2 located between base pairs 490-580 of the amplified product.

Examination of aligned sequences suggested that the *BG5* PCR amplicons of individuals contained more than one product. To test this, PCR products of 3 individuals were cloned and multiple clones sequenced. Aligned sequences revealed presence of multiple sequences in all three individuals (1459, 1461 and 1468). When compared to the described turkey B-G genes (Chapter 3), eight of the ten clones showed highest sequence similarity to *BG5*. One clone from 1461 had an equal number of sequence differences from *BG5* and *BG2*, while one clone from 1468 was most similar to *BG2*. These results suggest the presence of additional, undescribed B-G genes in the turkey.

Mitochondrial DNA analysis

Mitochondrial sequence variation was used as an independent estimate of genetic structure in the sampled birds. The D-loop is a neutral marker and the most variable part of the control region. As such, this region provides an informative measure of relatedness between sample groups. D-loop sequences were obtained from 65 of the 66 birds; individual 1457 was not included in the analysis due to PCR failure. Seven SNPs were identified in the 535 bp amplicon. All SNPs were transitions between cytosine and thymine, and one transition was unique to a single mitotype in a bird at the Grant 1 locality. Two mitotypes were observed in only single birds, one of these from the Crawford locality and the other from Grant 1 (Table 3.7).

As expected, all birds were homozygous at the variant positions. Eighty percent of the turkeys sampled could be grouped into one of two mitotypes. Mitotype 1 was the most common in two of the five localities, Crawford and Grant 2 (Table 3.7, Figure 3.5).

Mitotype 2 was the most common in the three remaining localities, Iowa, Richland and Grant 1. Among the sampled turkeys, Grant 1 had five mitotype groups, while Richland and Grant 2 each had four, Iowa had three and Crawford had two. Pairwise Fst values and migrants per generation (Nm) were used to investigate gene flow between the localities. Negative pairwise Fst values calculated between localities were interpreted as zero. Birds from the Crawford locality showed moderate levels of differentiation when compared to Iowa and Richland (Fst = 0.11 and 0.08 respectively) and higher levels of differentiation when compared to Grant 1 (Fst = 0.33). Grant 1 and Grant 2 showed low levels of differentiation (Fst = 0.098).

These results support the presence of family groups intermixed between localities. Conventionally, Nm levels much greater than 1 indicate gene flow levels high enough to counteract effects of genetic drift. The lowest Nm value (Nm = 0.5) was observed between Grant 1 and Grant 2, indicating little gene flow between these two localities. All other Nm values were greater than 1, suggesting high levels of female-mediated gene flow between localities and limited population substructure. Nucleotide diversity was low at all localities (π = 0.0007-0.0036).

Mitotype diversity (*h*) ranged from *h* = 0.257 (Grant1) to *h* = 0.848 (Richland) with an average of *h* = 0.663 across all samples. Tajima's *D* estimates fell within neutral expectations for all localities (Table 3.8). Chi-square tests for association between Southern phenotype and mitotype as well as SNP haplotype and mitotype found no significant association (p = 0.450 and p = 0.552, respectively).

Southern blotting method of MHC diversity compared to SNP analysis.

Comparisons between Southern phenotype and SNP analysis show several PHASED genotypes encompassed within each group. For example, there are 8 Southern blot phenotype groups that contain birds with different SNP genotypes (Table 3.1). The greatest correlation between Southern phenotype and SNP genotypes is present in phenotype group B, where 3 out 4 birds shared one SNP genotype. Phenotype group C is also interesting as it contains 5 birds sharing one SNP genotype (4/12), 3 sharing (4/13), 2 sharing (2/12) and 3 birds not sharing a SNP genotype.

The number of haplotype combinations (phased genotypes) was similar to the number of class IIβ phenotypes shown in the Southern blots (15 vs 16, Table 3.8). However, there was one more genotype at the Iowa and Grant1 localities and one less genotype at the Crawford locality compared to Southern phenotypes. Clearly, SNP haplotypes indicate greater MHC variability than that predicted by Southern blot in two localities; there are 8 additional SNP haplotypes in birds from the Iowa locality and 9 additional SNP haplotypes those from Grant 2. In contrast, the Richland and Crawford samples had 2 fewer SNP haplotypes than Southern haplotypes and 3 less were present in Grant 2. Chi square analysis found no association between Southern haplotypes and SNP haplotype ($p = 0.135$).

DISCUSSION

Comparison of methods to detect MHC variability

Investigations of MHC diversity in domestic and wild birds have focused mainly on the class IIβ peptide-binding region (PBR) because of its association with adaptive immune response. However, the genes in this family are difficult to study, due to their multiplicity and the polymorphic nature of the PBR. Another complicating factor is the high rate of gene conversion and gene duplication in avian taxa, which result in high sequence similarity among loci of even intron sequences surrounding the PBR. Sequence similarity complicates primer design for PCR and ultimately the designation of loci in many species.

Techniques such as RFLP, SSCP and DGGE can detect distinct class IIβ phenotype/haplotypes in closely related individuals, however, the variation detected within and between individuals is likely inaccurate for multiple reasons. Optimization of RFLP, SSCP and DGGE experimental conditions for each class IIβ allele is typically not possible, and genotype variability often cannot be determined due to the amplification of multiple loci in PCR reactions. This results in inflated estimates of heterozygosity. In addition, sequence homology between class IIβ genes, can cause overlap of similar alleles, which would underestimate diversity (Freeman-Gallant et al., 2002; Goto et al., 2002; Ekblom et al., 2004; Chaves et al., 2011). For example Bonneaud et al (2004) used RFLP to compare class IIβ diversity in sparrows and found variation within individuals to be extensive in that none of the individuals examined shared the same genotype.
One way to overcome the interpretation of methods such as SSCP or DGGE is by direct sequencing. For example, direct sequencing of class IIβ genes (8-15 clones per PCR) in New Zealand Robins revealed up to 28 predicted amino acid differences between class

IIβ alleles. This and other studies have shown that RFLP approaches, compared to direct PBR sequencing, underestimate class IIβ diversity likely due to non-polymorphic pseudogenes and the inability of RFLP to detect small sequence differences (Miller and Lambert, 2004).

Direct sequencing of class IIβ genes also has many drawbacks including: PCR artifacts, PCR recombination, heteroduplex formation and preferential or non-amplification of some alleles. Primer design is difficult due to the lack of extensive locus-specific sequences, and intron sequence may be highly similar between genes resulting in the amplification of multiple class IIβ genes in one PCR. Equally problematic is the possible exclusion of alleles due to the failure of PCR primers to recognize divergent sequences. Some of the problems associated with artifacts can be mitigated using next-generation sequencing approaches which effectively increase the ratio of signal to noise (Zagalska-Neubauer et al., 2010).

In the present study, MHC diversity of 66 wild turkeys in southern Wisconsin was initially determined by Southern blotting of the class IIβ and B-G regions, followed by PCR/RFLP SNP analysis of four genes flanking the class IIβ region. This implementation of PCR/RFLP focused on single-copy genes to investigate variation within and between populations by eliminating the many pitfalls (ie amplification of multiple alleles and pseudogenes) of previous studies that used class IIβ genes as a model for MHC diversity. This approach is possible because of the tight linkage between the class IIβ and flanking genes (Chaves et al., 2010). The use of single copy genes allows

for more accurate estimates of population structure, and gene flow, as allele frequencies can be used to estimate differences in heterozygosity and Fst/Nm.

The results from this study demonstrate PCR/RFLP SNP genotyping as a more informative measure of MHC diversity than class IIβ phenotyping by Southern blot analysis. The PCR/RFLP SNP method provided both genotype and haplotype information on MHC variability in these wild birds, which cannot be definitively determined from the class IIβ Southern blots. Discrepancies in the number of haplotypes is to be expected because without inheritance data, clustering by Southern blot pattern does not distinguish between homozygous and the many possible heterozygous haplotypes. This is further demonstrated in the 14 homozygous birds where eight different Southern haplotypes (ranging from 3 to 6 bands) were found, demonstrating that some SNP haplotypes are more variable than others in the class IIβ region.

Interestingly, the number of PHASED genotypes (haplotype combinations) indicated by SNP genotyping was nearly identical to the number of phenotypes observed by class IIβ Southern blotting (Table 3.9). The number of genotypes identified using PHASE showed more than twice the variability indicated by class IIβ Southern blot in the Iowa locality and 1.5x the variability in the Grant 1 locality. The lack of significant association between Southern group, SNP haplotype and mitotype is consistent with the mode of inheritance of MHC genes and mitochondrial DNA, as they are inherited independently. Twenty percent of the sampled birds had unique class IIβ Southern banding patterns so it is not surprising to find no association between this variable and SNP haplotypes.

Sequencing of *BG5*, the most variable *BG* locus based on previous studies (Chapter 3), clearly demonstrated the pitfalls of investigating gene families with high sequence homology. Sequence data supports the presence of at least one additional *BG* gene in addition to the five described in the species. The unique amplicons sequenced from three individuals were most similar to *BG5*, and may indicate a duplicated locus or copy number variation. The total number of B-G genes has yet to be determined in the turkey (domestic or wild). Thus it can not be determined if the unique alleles (or loci) observed in the present study are exclusive to wild birds. Previous studies have found differences in MHC gene copy number within and between domestic and wild species (Delarbre et al., 1992, Cutler et al., 2007). Sequence of the *B*-locus (Chapter 3) and preliminary RNA-seq data (KM Reed, unpublished) also suggest additional loci.

Genetic Relatedness of Wild Turkeys Based on Mitochondrial DNA

Our data shows that wild turkeys located in southern Wisconsin have a genetically diverse MHC and originate from several maternal lineages. The ancestral genetic background of the original birds transplanted to Wisconsin is unknown, but it seems certain from our data that these individuals were genetically diverse. Genetic diversity in the transplanted birds would mitigate a founder effect. The mitochondrial D-loop was previously shown to be most variable within the Eastern subspecies in comparison to the other subspecies inhabiting the U.S. (Mock et al., 2002), but the degree of relatedness among birds in Wisconsin was unknown. Two common mitotypes were found in 80% of the sampled birds, with two additional mitotypes at low frequency (<10%) in multiple

localities and 3 mitotypes each found in single individuals. Individuals in the Grant 1 and Crawford localities showed the least relatedness both to each other (Fst = 0.33, Nm = 1) and to the other three localities. The Grant 1 locality is closer to the Iowa state border than any other sampled locality. Interbreeding of wild birds in Wisconsin with wild birds from the state of Iowa could account for the increased genetic divergence of birds at this locality compared to those at the other localities. The Crawford locality is located on the border of Minnesota and Wisconsin, where influx of wild turkeys from Minnesota could decrease the relatedness of birds in this locality compared to the other groups. Birds located in the other three localities show no population differentiation as detected by Fst. Panmixis between these localities is not unexpected, since the sampled birds were located within a 35 mi radius, easily within the seasonal range of *M. gallopavo*. In comparison, a study investigating relatedness between two other subspecies of turkey (Merriam's and Gould's) in the southwestern U.S. found 9 mitotypes (5 unique) distributed across a 250km area, with two mitotypes containing fixed differences between the subspecies (Mock et al., 2001). A similar study of prairie chickens in central Wisconsin showed 7 mitotypes (Fst = -0.033-0.09) in 80 birds from four locations sampled between 1998-2000, a decrease from the 23 (Fst = -0.07-0.025) observed 50 years earlier (Johnson et al., 2004). Average mitotype and nucleotide diversity in the turkeys included in the present study ($h = 0.663$, $\pi = 0.0031$) was generally lower than that observed in the Eastern Wild turkeys ($h = 0.895$, $\pi = .00451$) sampled in the southern U.S. by Mock et al. (2002). However, one group (Richland) did show a similar level of diversity ($h = 0.848$, $\pi = 0.0061$). The comparatively decreased diversity observed in the present study could be

due to geographical proximity of sampled individuals, chance netting of related individuals, which tend to flock together in the winter, or the limited sample size, which is known to affect haplotype and nucleotide diversity values. In comparison Mock et al. (2002) characterized 10-20 hunter harvested birds per group sampled across 5 different states over a 4 year time period.

MHC-based genetic diversity

Selection pressures affecting MHC variation in vertebrates have been widely debated. Balancing selection caused by mechanisms such as gene duplication, negative assortative mating and pathogen-mediated selection are thought to act in tandem with genetic drift to affect MHC allele diversity. Balancing selection appears to have a prominent role in maintaining MHC diversity in populations on the long-term, even in otherwise monomorphic organisms (Aguilar et al., 2004). The magnitude of effect caused by each of these forces on MHC alleles is thought to correlate with the population's history. Genetic drift appears to have a pronounced effect in the short-term on MHC genes in isolated, small and recently bottlenecked populations, and alleles in these populations show a loss of adaptive genetic diversity (Alcaide et al., 2010; Eimes et al., 2011; Miller et al., 2010). Class IIβ genes are a common target in assessing MHC variability because of their known foreign peptide binding and presentation functions, which are related to an organism's ability to resist and clear pathogens. Higher levels of variation in the class IIβ PBR are thought to correlate with increased chance of survival (Miller and Lambert 2004; von Schantz 1996). Studies focusing on the effects of pathogen-mediated selection

on MHC diversity have found a positive association between class I and class IIβ PBR variability and pathogen diversity (Alcaide et al., 2010). Studies examining the effects of variation in other MHC genes on fitness are few (Jensen et al., 2008).

The number of SNPs located across the sequenced amplicons indicate an increased level of MHC heterozygosity compared to the rest of the turkey genome. The average SNP density across the turkey genome is reported at 1 SNP per 240 bp, which was determined using intron sequences in a commercial line (Reed et al., 2006). Chaves et al. (2010) examined commercial lines and found an average SNP density in the *B*-locus of 1 SNP per 70 bp (1 SNP/438 bp specifically in protein coding regions). The SNP frequencies within the MHC genes examined in the present study were also examined in a subset of commercial (Chaves et al., 2010) and Eastern wild (Chaves et al., 2011) turkeys. In comparison to commercial birds, MHC diversity in the WI wild birds was approximately 2x greater. This is not surprising, since commercial lines are under selection pressure for production traits, a phenomenon observed both in chickens and turkeys (Buchholz et al., 2004; Muir et al., 2008; Zhu et al., 1996)

The 15 Eastern wild turkeys previously examined by Chaves et al. (2011) came from geographical areas having no recorded reintroductions and no known extirpations or severe population declines. Populations in these areas have been used as sources for other relocation efforts because of their persistence and suspected genetic diversity (Mock et al., 2002). Those individuals exhibited between 1.2x to 3x greater SNP diversity (Chaves et al., 2011) than the wild birds sampled in this study. The reduced diversity of the WI wild birds may be due to high individual relatedness, close geographical proximity of

sampled individuals or a result of the potential bottleneck induced by population expansion from a reduced number of individuals (founder effect). This could be resolved by resampling birds in Wisconsin from a broader geographical area in the present day.

The class IIβ Southern phenotypes and SNP haplotypes suggest high levels of MHC diversity at the class IIβ and surrounding genes compared to the mtDNA D-loop in this sample of wild birds, supporting the hypothesis that high MHC diversity is maintained in these introduced populations. At the SNP haplotype level, heterozygosity was high, with the least MHC diversity seen in the Richland sample. Maintenance of heterozygosity is consistent with balancing selection. Of particular interest at the locus level is the high heterozygosity values (86.7%) observed for of *BTN1* and *TAP1* in the Grant 1 birds, indicating a selection pressure towards diversity at these two loci. Southern phenotypes indicate Grant 1 as the least diverse group of individuals, in contrast to the SNP haplotype profile, which found Grant 1 to be the most diverse group. MHC related gene flow between localities appears high based on Fst values with the highest Fst and lowest Nm occurring between Grant 1 and Grant 2 (Fst = 0.13, Nm = 3.23).

Locus by locus Hardy-Weinberg tests revealed less than expected heterozygosity between all localities for *BTN2* and *TRIM41* and no difference in expected heterozygosity for *TAP1* and *BTN1*. All localities were located within a 35 mile radius in the winter months, when family groups tend to flock together, so lower heterozygosity than expected under H-W is not surprising. The birds in the Grant 1 sample showed significantly greater heterozygosity than predicted by H-W equilibrium, which may be due to the influx of birds from the neighboring state (Iowa). With this data, it would be

useful to compare the diversity of the turkeys used in this study to turkeys currently located in the same regions to determine recent effects of gene flow on heterozygosity. Turkeys from other regions of the state, surrounding states and Missouri would give a more complete picture of how MHC diversity has evolved since the reintroduction of this bird.

CONCLUSION

We are just beginning to measure the overall diversity in the turkey MHC. Although the range of the wild turkey has been expanded through recent reintroductions, considerable MHC haplotype diversity appears to be segregating in these populations. Balancing selection may be acting to maintain MHC heterozygosity after their considerable population expansion. Although Southern blotting gives a generalized picture of MHC diversity, the method has limitations and provides a limited representation of total MHC diversity. SNP genotyping is more informative with the ability to determine the extent of heterozygosity at the locus and haplotype levels. These data provide information on selection pressures acting on a population, the number of segregating MHC chromosomes, and estimations of gene flow.

Individual	Location	Sex	Southern blot phenotype	*TRIM 41*	*BTN1*	*BTN2*	*TAP1*	SNP genotype	Mitotype
1427	Iowa	female	C	1.2	1.2	1.2	1.1	(1,12)	1
1428	Iowa	female	J	1.2	1.2	1.1	1.2	(4,12)	2
429	Iowa	female	B	1.1	1.1	1.1	1.1	(12,12)	2
1430	Iowa	female	A	2.2	1.2	1.2	1.1	(1,7)	5
1431	Iowa	female	C	1.2	1.2	1.2	1.2	(2,12)	2
1432	Iowa	female	E	2.2	1.2	2.2	1.1	(1,5)	2
1433	Iowa	female	E	2.2	2.2	1.2	2.2	(6,8)	1
1434	Iowa	female	I	2.2	2.2	1.2	1.2	(6,7)	2
1435	Iowa	female	F	1.2	1.1	1.1	1.1	(3,11)	1
1436	Iowa	female	D	1.2	1.2	1.1	1.2	(4,12)	2
1438	Richland	female	A	1.2	1.2	1.1	1.1	(3,12)	2
1439	Richland	female	H	2.2	2.2	1.2	1.1	(5,7)	2
1440	Richland	female	O	2.2	2.2	1.1	2.2	(8,8)	5
1441	Richland	female	E	2.2	2.2	1.2	2.2	(6,8)	6
1442	Richland	female	L	1.2	1.2	1.1	1.1	(3,12)	1
1443	Richland	female	B	1.1	1.1	1.1	1.1	(11,11)	1
1444	Richland	female	K	2.2	1.2	1.1	1.1	(3,7)	2
1445	Richland	female	P	1.2	2.2	1.1	1.2	(8,12)	6
1446	Richland		Q	2.2	1.1	1.1	1.1	(3,3)	5
1447	Richland	female	E	2.2	1.2	1.2	1.1	(1,7)	6
1448	Richland	female	R	1.1	2.2	1.1	1.1	(12,12)	2
1449	Richland		S	2.2	1.1	1.2	1.1	(1,3)	5
1450	Grant1	female	C	1.2	1.2	1.2	1.2	(2,12)	4
1451	Grant1	female	C	1.2	1.2	1.1	2.2	(4,13)	2

1452	Grant1	female	D	1.2	1.2	1.2	2.2	(2,13)	2
1453	Grant1	female	T	1.2	1.2	1.1	1.1	(3,12)	2
1454	Grant1	female	C	1.2	1.2	1.1	2.2	(4,13)	2
1455	Grant1		C	1.2	1.2	1.1	1.2	(4,12)	7
1456	Grant1	female	C	1.2	1.2	1.1	2.2	(4,13)	2
1457	Grant1	female	C	1.2	1.2	1.1	1.2	(4,12)	1
1458	Grant1	female	C	1.2	1.2	1.1	1.2	(4,12)	2
1459	Grant1	female	B	1.1	2.2	1.1	1.1	(12,12)	2
1460	Grant1	female	C			1.2	1.1		2
1461	Grant1	female	C	1.2	1.2	2.2	1.2	(5,10)	2
1462	Grant1		C	1.2	1.2	1.1	1.2	(4,12)	2
1463	Grant1	female	C	1.2	1.2	1.1	1.2	(4,12)	2
1464	Grant1	female	U	1.1	2.2	1.1	1.1	(12,12)	2
1465	Crawford	female	A	1.2	1.2	1.1	2.2	(4,13)	1
1466	Crawford	female	N	1.1	1.2	1.1	1.1	(11,12)	3
1467	Crawford	female	E	2.2	1.1	2.2	1.2	(1,2)	1
1468	Crawford	female	E	2.2	1.1	1.2	2.2	(2,4)	1
1469	Crawford	female	M	1.2	1.2	1.2	1.2	(2,12)	2
1470	Crawford	female	V	1.1	1.2	1.1	1.1	(11,12)	1
1471	Crawford	female	E	2.2	1.1	1.2	2.2	(2,4)	1
1472	Crawford	female	M	1.2	1.2	1.1	1.2	(4,12)	2
1473	Crawford	female	W	1.1	2.2	1.1	1.1	(12,12)	1
1474	Crawford	male	N	1.1	1.2	1.1	1.1	(11,12)	1
1475	Crawford		C	2.2	1.2	1.1	1.2	(3,8)	2
1476	Grant2	female	B	1.1	2.2	1.1	1.1	(12,12)	1
1477	Grant2	male	I	1.1	1.1	2.2	1.2	(9,10)	2
1478	Grant2	male	G	1.2	2.2	1.1	1.1	(3,12)	2
1479	Grant2	male	G	1.1	1.2	1.1	1.1	(11,12)	2
1480	Grant2	male	H	2.2	1.1	2.2	1.1	(1,1)	1

1481	Grant2	female	G	1.2	1.1	1.1	1.1	(3,11)	1
1482	Grant2	female	H	2.2	1.1	2.2	2.2	(4,4)	1
1483	Grant2	female	J	2.2	2.2	1.2	1.1	(1,7)	1
1484	Grant2	male	A	1.2	1.1	1.1	1.1	(3,11)	1
1485	Grant2		K	2.2	1.2	1.1	1.1	(3,7)	2
1486	Grant2	female	X	2.2	2.2	1.1	1.1	(7,7)	2
1487	Grant2	female	F	1.1	1.2	1.1	1.1	(11,12)	1
1488	Grant2	female	L	2.2	1.1	1.1	1.1	(3,3)	5
1489	Grant2	female	Y	2.2	1.2	1.1	1.1	(3,7)	5
1490	Grant2	female	Z	1.2	1.2	1.1	1.1	(3,12)	5
1491	Grant2	female	H	2.2	1.1	1.1	1.1	(3,3)	1
1492	Grant2	female	G	1.2	1.2	1.1	1.1	(3,12)	2
1493	Grant2	female	AA	1.2	2.2	1.1	1.2	(8,12)	6

Table 3.1. Genetic variation in wild turkeys from southern Wisconsin. For each individual, the collection location, sex, class IIβ Southern phenotype, locus genotypes, SNP genotype (haplotype combination) and mtDNA haplotype (D-loop mitotype) are given.

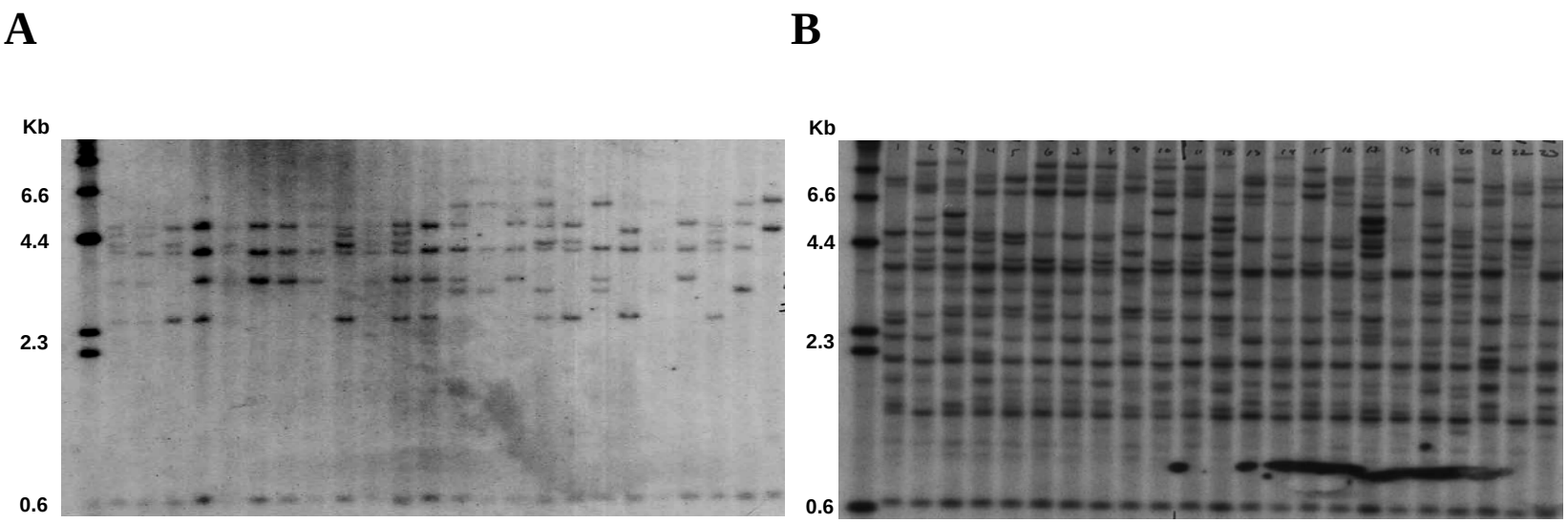


Figure 3.1A-B. Example Southern blot of wild turkey genomic DNA digested with *Pvu* II and hybridized with a (A) class IIβ (Genbank M29763) or (B) B-G (bg11) probe.

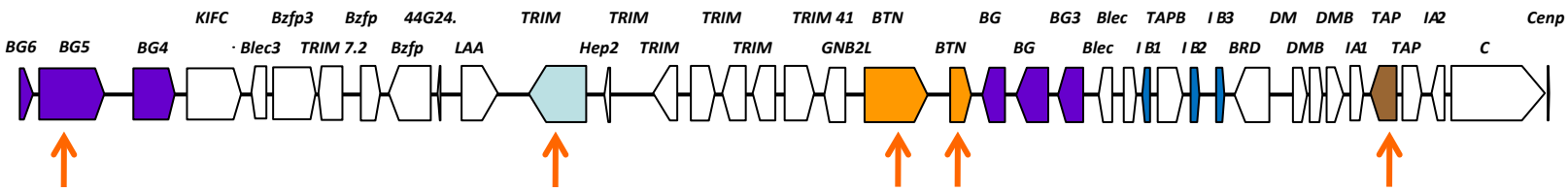


Figure 3.2. The turkey *MHC-B* locus. The genes examined by Southern blotting and SNP genotyping are colored. The known class IIβ and B-G genes represented by the Southern blot are dark blue and purple, respectively. Position of amplicons resequenced for haplotype analysis are dentified by arrows. Primer sequences are given in Table 3.2.

Gene	Primers (5'-3')	Product size (bp)	Enzyme	Digest time	Position /Ambiguity code														
					21197	21198	212041	212197											
TAP1	GCCAGATACCACAGCAGGAG	401	*Fok*I	1 hr															
	GGCCGTGCCCTACTACAC				Y	Y	Y	S											
					133534	133539	133568	133570	133642	133643	133647	133736	133739	133812	133816	133833	133902	133971	
BTN1	AGAGCGTCACACCAAAATAGG	504	*Msl*I	2 hr															
	GGGAAGTCTTGGTTGTGTTTC				R	W	R	Indel	Y	R	S	R	W	M	W	W	R	W	
					140073	140116	140219	140412	140513	140629	140700	140755							
BTN2	CCCATTGACATGAAGTGACC	800	*Nco*I	2 hr															
	AAACAGGTGGCCAAGAACAC				Y	S	K	Y	R	Y	S	R							
					117418	117426	117483	117497	117628	117675	117697	117753	117942	118205	118317				
TRIM41	CGGAGTCACAAGCATCACAG	1100	*Msp*I	4 hr															
	GGCCAGCAAGCTCTCATAAC				W	R	R	Indel	R	Y	R	Y	R	Y	R				

Table 3.2. Summary of SNPs identified in *MHC-B* genes. For each gene, the primers used in PCR, product size, genotyping enzyme (PCR/RFLP) and conditions are given. Position of SNPs within the *MHC-B* (Genbank DQ993255) are listed with the corresponding IUPAC nucleotide ambiguity code and SNPs used for genotyping are underlined.

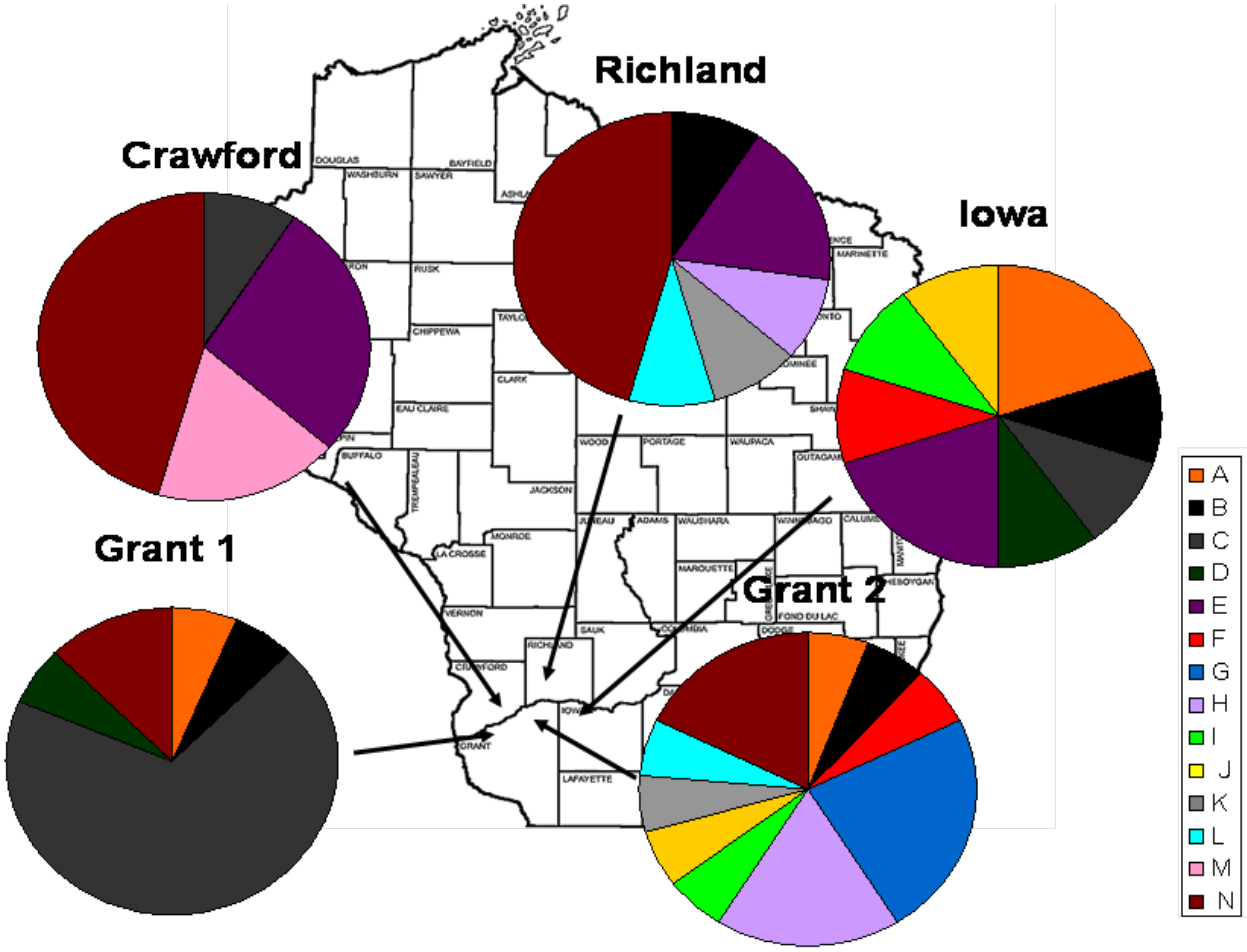


Figure 3.3. Class IIβ haplotype distributions determined by Southern blot. Group N are birds with unique Southern banding patterns.

Location	*TRIM 41*		*BTN1*		*BTN2*		*TAP1*	
	A	G	T	A	G	C	G	C
Iowa	0.65	0.35	0.60	0.40	0.65	0.35	0.70	0.30
Richland	0.71	0.29	0.58	0.42	0.83	0.17	0.79	0.21
Crawford	0.45	0.55	0.55	0.45	0.77	0.23	0.64	0.36
Grant 1	0.43	0.57	0.57	0.43	0.87	0.13	0.43	0.57
Grant 2	0.65	0.35	0.35	0.65	0.85	0.15	0.88	0.12
All Turkeys	0.58	0.42	0.50	0.50	0.82	0.18	0.69	0.31

Table 3.3. Allele frequencies by sample at SNPs in four genes flanking the class IIβ genes of the turkey. Frequencies for locus were calculated using the EM algorithm.

		TRIM41			***BTN1***			***BTN2***			***TAP1***		
Location	#	Obs. Het.	Exp. Het.	P-value	Obs. Het.	Exp. Het.	P-value	Obs. Het.	Exp. Het.	P-value	Obs. Het.	Exp. Het.	P-value
Iowa	10	0.4000	0.4421	1.0000	0.6000	0.5053	1.0000	0.5000	0.4789	1.0000	0.5000	0.4789	1.0000
Richland	12	0.0833	0.3442	0.0310	0.3333	0.5073	0.2901	0.3333	0.2899	1.0000	0.2500	0.4312	0.1938
Crawford	11	0.3636	0.4848	0.5372	0.5454	0.5195	1.0000	0.2727	0.3680	0.4385	0.1818	0.5195	0.0610
Grant 1	16	0.4667	0.5081	1.0000	0.8667	0.5080	0.0001	0.1333	0.2391	0.2031	0.8667	0.5081	0.0001
Grant 2	17	0.1176	0.2139	0.1803	0.4706	0.4706	1.0000	0.0588	0.2585	0.0145	0.3529	0.4706	0.3366
Total	65	0.2769	0.4293	0.0077	0.5692	0.5039	0.3281	0.2308	0.3131	0.0453	0.4462	0.4919	0.6093

Table 3.4. Heterozygosity of the sampled birds measured at each locus. Observed heterozygosity was compared to expected heterozygosity using a Hardy-Weinberg Exact Test. P-values ($\alpha = 0.05$) are shown for each locus.

	Iowa	Richland	Crawford	Grant1	Grant2
Iowa		0.000	0.000	0.047	0.043
Richland	0.024		0.019	0.078	0.000
Crawford	0.000	0.085		0.000	0.130
Grant 1	0.000	0.150	0.330		0.030
Grant 2	0.000	0.000	0.098	0.000	

Table 3.5. SNP haplotype (above diagonal) and mitotype (below diagonal) Fst values observed for southern WI wild turkeys.

Haplotype	Haplotype sequence	Total Freq (%)	Iowa	Richland	Crawford	Grant 1	Grant 2
1	GTGG	19.5	14.5	14.0	35.0	26.0	11.0
2	AAGG	17.0	11.0	23.0	0.0	7.0	31.0
3	GAGG	11.0	10.0	13.0	15.0	3.0	17.0
4	ATGG	10.5	10.0	17.0	5.0	3.0	19.0
5	AAGC	8.0	3.0	0.0	20.0	20.0	6.0
6	ATGC	7.5	11.0	14.5	0.0	6.0	0.0
7	AACG	6.0	9.0	6.0	5.0	1.5	8.0
8	GTGC	5.0	3.0	2.0	0.0	16.0	0.0
9	ATCG	4.0	12.0	6.5	0.0	0.5	1.0
10	AACC	3.0	1.0	0.0	20.0	4.0	0.0
11	GAGC	3.0	3.0	0.0	0.0	6.0	0.0
12	ATCC	2.5	8.0	4.0	0.0	1.0	0.0
13	GACC	1.0	0.5	0.0	0.0	1.0	3.0
14	GACG	1.0	2.0	0.0	0.0	0.0	3.0
15	GTCG	0.5	1.5	0.0	0.0	2.0	0.0
16	GTCC	0.5	0.5	0.0	0.0	3.0	0.0
% Het		80.7	90.0	66.7	90.0	92.3	72.7

Table 3.6. SNP haplotypes determined by PHASE analysis of genotyped SNPs in genes flanking the class IIβ genes of the turkey *MHC-B*. For each haplotype, the SNP combination (sequence) and frequencies by population are given.

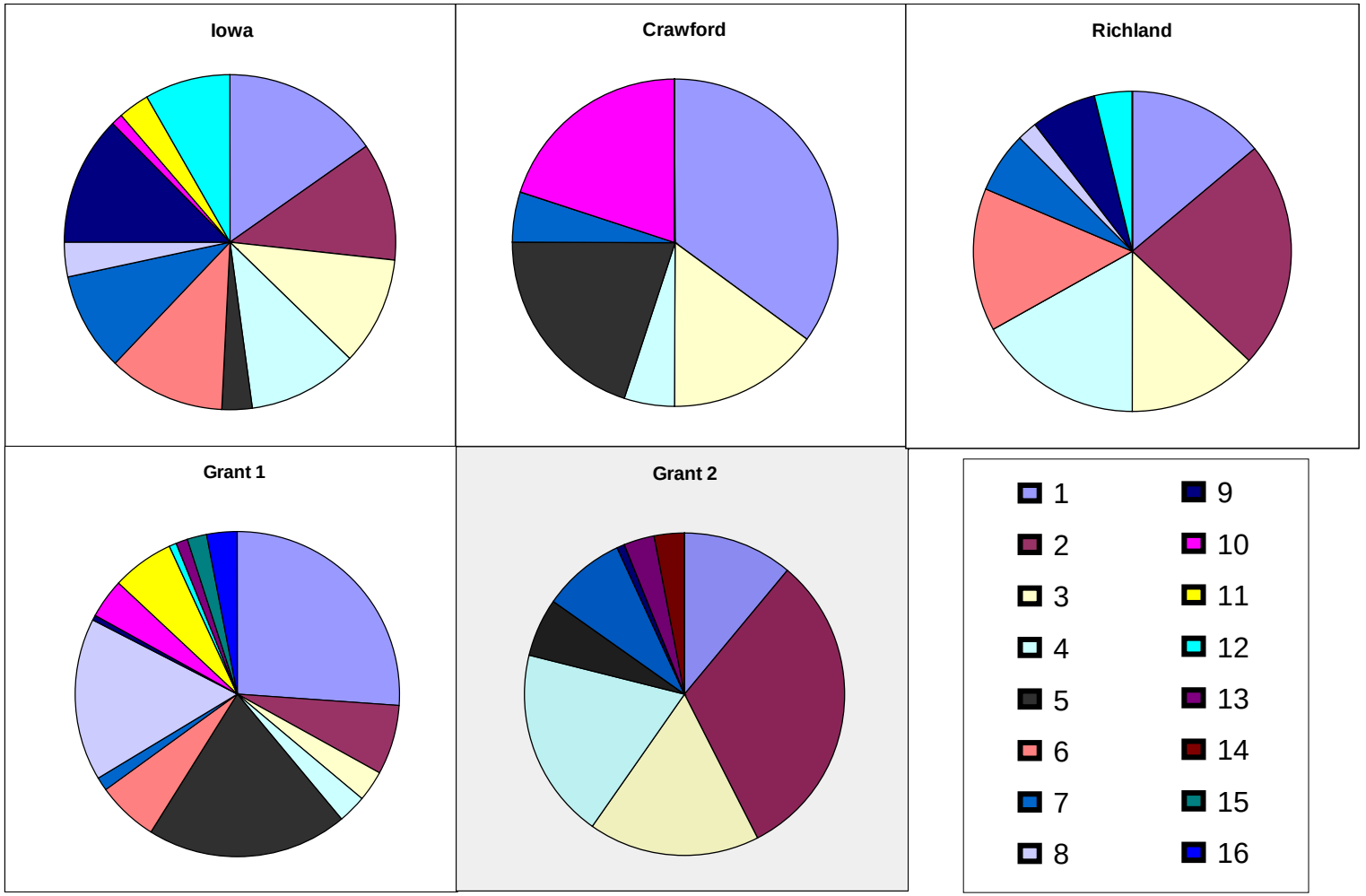


Figure 3.4. Distribution of phased MHC-SNP haplotypes in wild turkeys from southern Wisconsin.

Group	Mitotype	Total Freq (%)	Iowa	Richland	Crawford	Grant 1	Grant 2
1	TCCCTC	32.0	3 (30)	2 (17)	7 (60)	2 (13)	7 (41)
2	TCCCCC	46.0	6 (60)	4 (33)	3 (30)	13 (73)	6 (35)
3	TCCTTT	1.5	0	0	1 (10)	0	0
4	TCCTCC	1.5	0	0	0	1 (7)	0
5	TCCTCT	11.0	1 (10)	3 (25)	0	0	3 (18)
6	TTTCTC	6.0	0	2 (25)	0	0	1 (16)
7	CCCCCC	1.5	0	0	0	1 (7)	0

Table 3.7. Mitochondrial DNA haplotypes (mitotypes) observed in wild turkeys from southern Wisconsin. For each mitotype, the overall frequency and the number and frequency of each type observed (in brackets) in the five sampled localities are given.

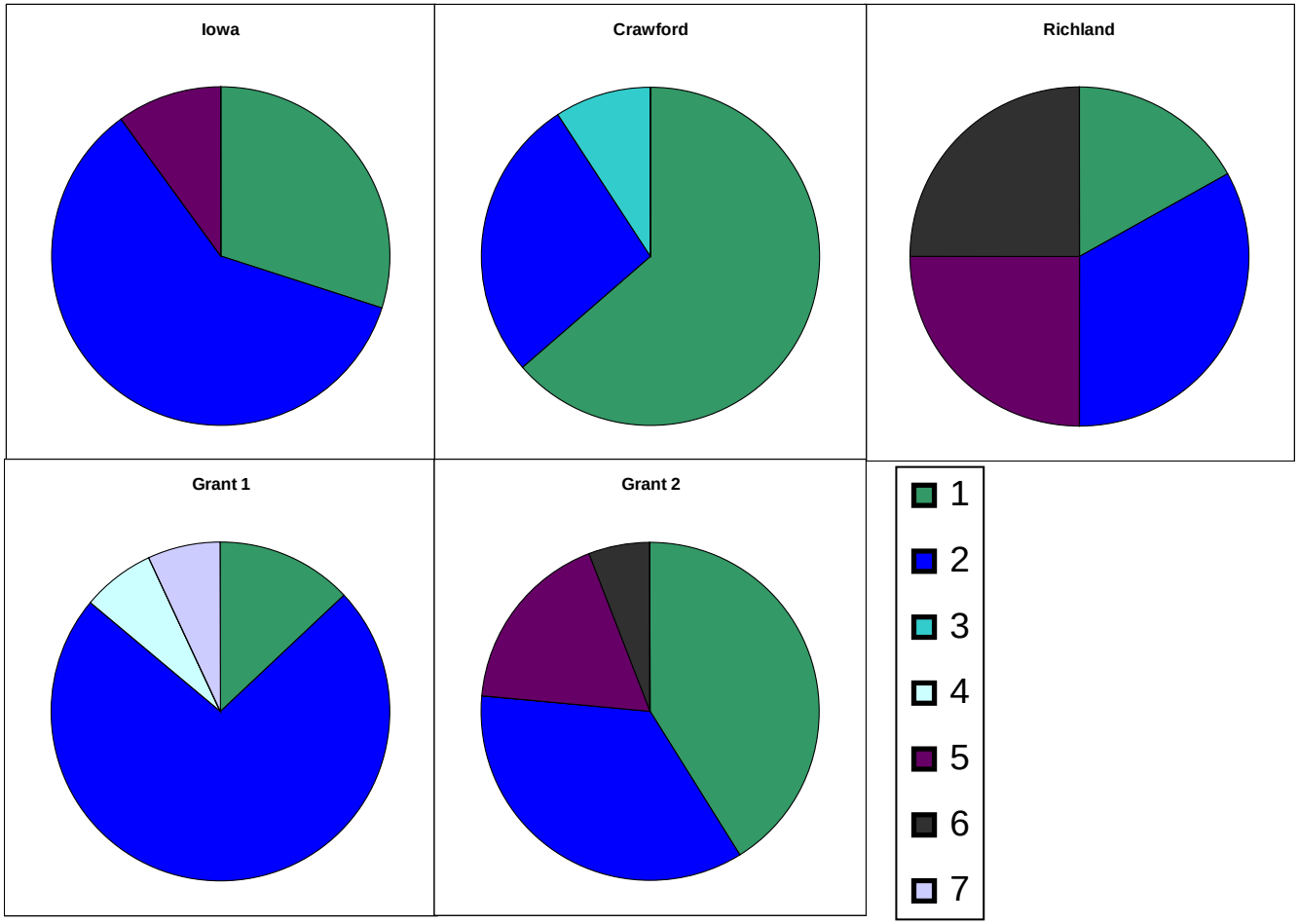


Figure 3.5. Mitotype distribution in wild turkey populations sampled in southern Wisconsin.

Location	No. of mitotypes	*h* (Haplotype diversity)	π (Nucleotide diversity)	Tajima's D
Iowa	3	0.600	0.0023	-0.6575
Richland	4	0.848	0.0061	0.6651
Crawford	3	0.436	0.0011	0.6714
Grant1	4	0.257	0.0007	-1.4905
Grant2	4	0.713	0.0036	-0.2108
All Turkeys	7	0.663	0.0031	-0.5360

Table 3.8. Summary of mtDNA diversity measures among sampled localities of wild turkeys.

Location	Class IIβ Southern phenotypes	SNP haplotypes	Phased Genotypes	Mitotypes
Iowa	8	16	9	3
Richland	11	9	11	4
Crawford	8	6	7	3
Grant 1	6	15	7	4
Grant 2	12	9	12	4
All birds	27	16	28	7

Table 3.9. Summary of MHC variation in WI wild turkeys as estimated by Southern blot, SNP haplotyping (determined by PHASE) and PHASE genotypes as compared to variation observed in the mitochondrial D-loop (mitotypes).

.

Chapter 4

Expression of MHC B-G Genes in Response to Aflatoxin Challenge in the Turkey.

Abstract

Aflatoxin B_1 (AFB_1) is a ubiquitous contaminant of animal feed, causing severe liver damage in a variety of organisms. Turkeys show extreme sensitivity to aflatoxin exposure, with low doses producing visible symptoms of aflatoxicity including severe decreases in growth rate, immunosuppression and liver necrosis. Multi-gene families are characteristic of the major histocompatibility complex (MHC), some of which exhibit high degrees of polymorphism, critical to development of an immune response. One example is the B-G gene family. The turkey B-G genes contain highly polymorphic extracellular domain sequences surrounded by tightly conserved regions and variable length cytoplasmic tails. This study was designed to generate an expression profile of the turkey B-G genes and to measure *BG* expression in birds exposed to aflatoxin in the context of *Lactobacillus* probiotic. B-G expression in liver tissue of aflatoxin challenged birds was measured using quantitative real-time PCR. A low level of expression was observed for three of the four B-G genes investigated. *BG4* expression levels fluctuated within and between groups and a higher level of expression was measured in birds challenged with aflatoxin after treatment with the probiotic *Lactobacillus*. This study provides insights into the potential differences in B-G gene functions relative to aflatoxin exposure in the context of *Lactobacillus* and the groundwork for future studies quantifying expression levels of this gene family.

INTRODUCTION

Aflatoxin B_1 (AFB_1) is a potent carcinogen commonly present in maize and other corn-based animal feeds. The potent effects of this mycotoxin on young poults were first identified in turkeys as a result of the Turkey X outbreak of 1960, which involved the deaths of more than 100,000 birds (Blount 1961; Spensley 1963). AFB_1 is produced by fungi in the *Aspergillus* group and is transmitted to animals and humans through ingestion of contaminated food. In poultry, aflatoxin induces many deleterious effects, including reduced body weight gain, decreased organ weights, immunosuppression and irreversible liver damage.

Turkeys are known as the most susceptible animal to AFB_1 with extreme sensitivity to aflatoxicosis due to a combination of efficient microsomal activation of AFB_1 to the toxic intermediate, *exo*-AFB_1-8,9-epoxide(AFBO), via cytochrome P450 enzymes and inefficient detoxification of the intermediate mediated by glutathione-S-transferase (GST) enzymes (Diaz et al., 2010; Klein et al., 2000; Rawal and Coulombe, 2011). Even at levels far below the US FDA action level of 110 ppb, AFB_1 causes poor performance, along with decreased growth rate, body weight, weight gain, egg production, reproductive performance, feed efficiency (Arafa et al., 1981; Pandey and Chauhan, 2007) and overall productivity in commercial birds (Weibking et al., 2004). Wild turkey poults show more resistance to aflatoxicity than their domestic counterparts and exhibit less weight loss and mortality than domestic poults in response to aflatoxin challenge (Quist et al., 2000).

Several studies have shown that certain species of viable or heat-killed bacteria, including *Lactobacillus* and *Propionibacterium*, are capable of binding to, absorbing and detoxifying AFB_1 in many organisms including chickens, rats, mice, humans and turkeys (Da-Silva and Oluwafemi, 2009; El-Nazami et al., 2000; Lahtinen et al., 2004; Lee et al., 2003). *Lactobacillus sp.* are of particular interest because of their widespread use in animal feed in addition to their anti-carcinogenic and anti-genotoxic effects in humans and animals (Commane et al., 2005). Studies show several species of *Lactobacillus* improve daily feed intake of aflatoxin challenged animals, returning measurements to normal levels (Hathout et al., 2011).

Host gene expression of aflatoxin-challenged animals is limited, and there are no published data regarding aflatoxin-induced changes in expression of genes in the major histocompatibility complex (MHC), despite the well-known role of these in disease resistance and antibody production. Embryonic exposure to AFB_1 results in a compromised immune response in hatched chicks through suppression of humoral and cellular immunity (Quershi et al., 2000). Immunosuppression appears to be at least partly due to upregulation of cytokines, including IL-6 (Yarru et al., 2009). Microarray gene expression profiling of broiler chickens exposed to dietary aflatoxin found down regulation of many hepatic genes related to immune function, consistent with the immunosuppressive effects of aflatoxicity (Yarru et al., 2009).

B-G genes are part of a polymorphic multi-gene family in the major histocompatibility complex (MHC) of avian genomes. There are six B-G genes described in the turkey MHC and additional genes are suspected. These are organized in

two clusters separated by ~125kb (Chapter 3; Chaves et al., 2009), which are tightly linked to the class I and class II loci. Expression data on these genes is limited, and their function has yet to be fully elucidated, although several possible functions have been hypothesized (Kaufman et al., 1991; Goto et al., 2009). Early studies describe B-G molecules as erythrocyte antigens (Koch et al., 1983; Longenecker and Mosmann, 1980). Later studies expanded on this data, presenting convincing evidence for cell surface expression on erythocytes as well as in caecal, small intestine and liver tissues (Salomonsen et al., 1987; Miller et al., 1990; Salomonsen et al., 1991). B-G polypeptides isolated from liver and caecal extracts show size heterogeneity, unlike those isolated from erythocytes (Miller et al., 1990). B-G proteins have highly conserved immunoglobulin (Ig) variable (V)-like and transmembrane domains with extensive divergence in their intracellular regions. Structurally, *BG*s share some characteristics with the chicken leukocyte receptor complex family (*CHIR*) (Laun et al, 2006). Recently, a *BG1* allele in the chicken was shown to confer significant resistance to gallid herpesvirus-2-induced lymphomas (Marek's disease) further supporting a hypothesized immune function of *BGs* (Goto et al., 2009). Preliminary RNA-seq data suggest that B-G loci in the turkey may be responding to AFB_1 exposure (KM Reed, unpublished observations). This study was designed to generate an expression profile of the turkey B-G genes and to measure *BG* expression in birds exposed to aflatoxin in the context of *Lactobacillus* probiotic.

MATERIALS AND METHODS

Initial Survey of B-G gene expression

Expression of B-G genes was initially examined by testing for PCR amplification with locus-specific primers from cDNA templates prepared from total RNA of non-challenged birds (Table 4.1). Briefly, RNA was extracted from heart, bursa, spleen and liver tissues of the RBC2 turkey line (tissues provided by Dr. Sandra Velleman, The Ohio State University) using Trizol (Invitrogen). cDNA was synthesized using the Multi-temp Affinity Script cDNA Synthesis kit (Agilent). Locus-specific primers were designed to amplify segments of *BG1-5* between exons 1 and 3. PCR was performed using HotStar Taq (Qiagen, Inc) at the following parameters: 94°C for 10 minutes followed by 35 cycles of 95°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec and a final extension of 72°C for 10 min. Expression of each locus was determined by visualization on a 2% agarose gel followed by sequencing of the PCR products. For sequencing, PCR samples of expressed loci were purified by adding 1µl of ExoSAP-IT (USB Corp) to 4µl PCR product and incubated at 37°C for 15 min and 80°C for 15 min. Treated DNA was sent to the Biomedical Genomics Center, University of Minnesota for Sanger sequencing. Sequences were aligned and compared using Sequencher 4.9 software (Gene Codes Corp.).

Aflatoxin/Lactobacillus challenge

Bacterial strains. A freeze-dried mixture of probiotics, "LGG" (Valio Ltd., Helsinki, Finland), consisting of *Lactobacillus rhamnosus* GG (2.3×10^{10} cfu/ g), *Lactobacillus*

rhamnosus LC-705 (3.0 x 10^{10} cfu/ g), prepared by directly suspending lyophilized bacteria in PBS (phosphate buffered saline, Gratz et al., 2006).

Animals and treatment. Liver tissue from 32 day old Nicholas strain turkeys was obtained as previously described (Rawal 2010). Briefly, the experimental design was as follows: after acclimatization for 10 days, birds were randomly placed in one of the four groups, AFB_1 (PA), LGG (LB), LGG + AFB_1 (LA) and PBS (control). For the first 10 days, the LB and LA groups received pretreatment with 0.5 ml of LGG (final concentrations of 5 X 10^{10} CFU/0.5 ml PBS) and the control and PA groups received pretreatment with PBS (0.5 ml), once daily, by oral gavage, after which AFB_1 treatment (for groups: PA and LA) was started at 1 ppm in the diet. Aflatoxin B_1 and LGG or PBS treatment continued for the next 11 days. Thereafter, animals were euthanized and livers were rapidly removed, frozen on dry ice and stored at − 80 °C.

Quantitative Analysis of BG Expression

RNA isolation and cDNA synthesis. Livers from three individuals were chosen from each treatment group, PA and LA, and each control group, PBS and LB, for sequencing and expression analysis. Total RNA was isolated from liver tissue using Trizol (Invitrogen), extracted using phenol/chloroform following the manufacturer's protocol and including denaturation and DNase treatment (Ambion Turbo DNA-*free*™ kit RNA, Applied Biosciences) and stored at -80°C. Quality of the extracted RNA was determined

by visualization on a 1% formaldehyde agarose gel to verify integrity of the 18S and 28S bands, and by measuring the $A_{260/280}$ ratio with an NanoDrop-1000 spectrophotometer.

Reverse transcription was performed on 1 µg of total RNA per sample using the USB First-Strand cDNA Synthesis kit for Real-Time PCR (Affymetrix/USB). The manufacturer's protocol was followed using a mix of random and poly-A tail specific primers and including the denaturation step for secondary RNA structure to produce template for quantitative real-time PCR. Samples of RNA were prepared concurrently to control for variation in PCR and reverse transcription efficiencies. Briefly, 1 µl of random and 1 µl polyA-tail specific primers were combined with 1 µg of RNA. The RNA/primer mix was heated at 75ºC for 5 minutes to minimize secondary structure. The PCR reaction was performed in 25µl reaction volumes as follows: 44ºC for 60 minutes and 92ºC for 10 minutes. The cDNA products were stored at -20ºC and a 1:10 dilution of each sample was used for the qRT-PCR assay.

Normalization. *β-actin*, *GAPDH* and *RNA polymerase 2D* (*RNA2D*) were chosen as potential reference genes based on previous studies reporting gene expression changes *in vitro* and in liver tissues of other organisms in response to aflatoxin challenge (Ratajewski et al., 2011). Each primer set was initially tested on each individual, and delta Ct values were examined with NormFinder (Andersen et al., 2004) to determine the best reference genes for this system. The gene with the lowest stability values was selected for normalization.

Real-time quantitative PCR (qRT-PCR) assay. *BG1*, *BG2*, *BG4* and *BG5* were chosen as test genes for the assay based on the initial gene expression survey of liver tissue. Primer3 v .0.4.0 software (Rozen and Skaletsky, 2000) was used to generate primers using genomic sequence of a commercial bird from a Nicholas line (Chapter 3). Where possible, primers were designed at exon-exon junctions to help identify potential DNA contamination. Each B-G primer pair was initially tested for specificity on one individual per test group with PCR using HotStar Taq (Qiagen, Inc) at the following reaction parameters: 94°C for 10 minutes followed by 35 cycles of 95°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec and a final extension of 72°C for 10 min. PCR products were bidirectionally sequenced and aligned to published locus specific sequence.

Quantitative Real-Time PCR. The expression assay included standard curves for each gene, two biological replicates and 3 individuals per treatment group and no template and genomic controls. A master mix (16µl) consisting of 4µl template, 10µl SYBR green super-mix (Bio-Rad Laboratories) and 2µl H_20 was pipetted onto a 96-well plate, and 4ul (500 µM) of the appropriate primer was added to each well. Forward and reverse primers of each gene were initially mixed and diluted to 0.5µM. The SYBR green with dissociation curve assay was run on an Agilent MX3000p (Stratagene) with cycling conditions as follows: 1 cycle at 95°C for 3 min, followed by 40 cycles of 95°C for 20 sec, 60°C for 30 sec, 72°C for 30 sec. A dissociation curve was run to ensure specificity of the amplicons using 1 cycle of 95°C for 30 sec, 60°C for 30 sec and 95°C for 30 sec. ROX was chosen as the reference dye and both no template and genomic DNA controls

were used for each primer set to detect potential contamination. Threshold cycles (Ct) were obtained in duplicate for each individual. Quantitative real-time PCR was performed using PerfeCta SYBR Green Fastmix Low ROX (Quanta Biosciences).

Standard curve. A standard curve was generated for each gene by running multiple 20µl reactions of cloned plasmid DNA containing gene-specific cDNA fragments with varying amounts of known starting template (0.005 pg- 10 ng dependent on the target). For each amplicon, clones were constructed by amplifying fragments from each gene via PCR and purified with Qiaquick PCR purification kit according to manufacturer's protocol (Qiagen). These products were ligated into pGEM-T easy T/A cloning vector according to manufacturer's protocol (Promega). Ligations were transformed into DH5α subcloning efficiency cells according to manufacturer's protocol, plated onto selective media (100 µg/ml ampicillin plus Xgal), and grown overnight at 37°C. Single colonies were isolated in 2 ml Luria Bertani (LB) broth supplemented with 100 µg/ml ampicillin then grown overnight in a 37° C shaking incubator. Plasmid DNA was purified with a Qiaprep miniprep purification kit according to manufacturer's protocol (Qiagen). Isolated plasmid DNA was digested with *Eco*RI to verify cloned insert size and gene identity was verified by DNA sequencing. Efficiency of primers was assessed via fit to standard curve (Table 4.1).

Data analysis. MxPRO QPCR software (Agilent) was used for data analysis. Initial template quantity for each sample was interpolated based on standard curve values for

BG4. Relative expression differences were calculated using the Relative Quantitation module. Automated gene expression analysis was calculated by comparing baseline-corrected normalized fluorescence on a log base 2 scale and normalizing to reference genes. Differences in reaction efficiencies were compensated for using standard curve values for each gene. Upper and lower error estimates were based on replicate variability.

Results

B-G Expression Profile

B-G transcript screening in a variety of tissues revealed a diverse expression profile, with four locus-specific transcripts (*BG1, 2, 4* and *5*) identified in the liver, three (*BG1, 2* and *5*) identified in spleen, heart, bursa and skeletal muscle (*pectoral major*) and two (*BG3* and *4*) in intestinal epithelium (Table 4.2). Analysis of whole blood showed exclusive *BG2* expression. This confirms *BG4* and *BG5* as expressed genes and shows the first documented expression of *BG3* in any tissue (Figure 4.1).

Quantitative Analysis of BG Expression

Genes with the lowest stability value are considered to have the least change in expression between samples. Of the candidate normalization genes, *RNA2D* had the lowest stability value (0.035) (Table 4.3). In addition, one of the B-G genes (*BG5)* was invariant across samples (stability value = 0.016). Standard curves were run on all potential reference and test genes to check for reaction efficiency and specificity of

primers. The *BG5* and *RNA2D* standard curves showed highest reaction efficiencies (93.8% and 93.2%). Standard curves for *BG1*, *BG2* and *BG4* had lower efficiencies (89%, 73.1% and 66.6%, respectively). Correlation coefficients were high for all standard curves (>0.991, except *BG5* = 0.94).

Preliminary tests of primers on the liver samples showed little change in Ct values for *BG1* and the stability value, as evaluated by NormFinder, was low (0.044) (Table 4.3). In addition, the dissociation curve for this gene showed two prominent peaks in the test samples. This suggests amplification of a second transcript, perhaps corresponding to an as of yet undescribed B-G family member. As a result, *BG1* was excluded from further analysis. For *BG2*, sample dissociation peaks showed a prominent shoulder. Only two samples had Ct values less than 35 with the remaining samples having late cycle Ct values (35.56-38.62), likely a result of extremely low gene expression. The Ct value of only one sample was encompassed by the *BG2* standard curve with the lowest calculated template quantity of 2.72×10^{3} transcripts (1×10^{-5}ng). No amplification was observed at lower plasmid concentrations. Dissociation curves for *BG4* and *BG5* each had one prominent peak, indicating amplification of a single product. Ct values for *BG5* ranged from 23.02- 25.92, while values for *BG4* ranged from 28.12 to 36.43, the most variable gene across the experimental samples. The *BG4* standard curve showed Ct values down to 3.47×10^{1} transcripts (1.28×10^{-7} ng), which encompassed all sample Ct values.

Hepatic B-G expression changes between treatment groups

The numbers of mRNA transcripts were calculated using gene-specific standard curves. When examined across all samples, *RNA2D* and *BG5* showed low individual expression change (Figure 4.2). Then number of transcripts of the normalizing gene *RNA2D* ranged from $2.84x10^2$ transcripts per ng RNA (T/ng) (PBS2) to $2.24x10^3$ (LA9) and averaged $7.62x10^2$ across all samples. In comparison, transcript number for *BG5* averaged $3.50x10^2$ T/ng with the narrowest range across individuals ($1.13x10^2$ to $7.76x10^2$ T/ng RNA, in samples PBS9 and PA3, respectively).

Of the two B-G genes examined, *BG4* showed both the greatest within group and among group variation. Transcript numbers ranged from $7.60x10^1$ (PBS9) to $3.00x10^4$ (PA8) with an overall average of $9.04x10^3$ T/ng. Dramatic differences in the overall expression of *BG5* and *BG4* were observed across samples (Figure 4.3). Transcript fold changes between *BG5* and *BG4* ranged from 0 (PBS10 and LB9) to 2.25 (Log10) (PA8). This difference in expression was significant across the 12 individuals [two-tailed t-test assuming unequal variances ($t(12) = 3.02$, $p = 0.01$, $\alpha = 0.05$)].

Comparison of *BG4* expression among groups was confounded by the elevated expression observed for one sample in each of the PBS, PA and LB groups and this is reflected in the group summary statistics (Table 4.4a, Figure 4.2). Average transcript number ranged from $1.39 x10^4$ T/ng (LB group) to $9.47 x10^4$ for the LA group. Most notable in comparing groups was the consistent upregulation of *BG4* in the LA group (Table 4.1A-B; Figure 4.2; Figure 4.4). Individual expression values were higher in this group than for all other individuals except PA8, which had the highest level of expression

(Figure 4.2). This corresponds to a 6.5 to 7.5 X fold increase in *BG4* when compared to the individual with the lowest *BG4* expression (PBS9) (Figure 4.4). Similar to the within individual comparisons, a t-test comparing *BG4* and *BG5* levels within each group revealed a significant difference only in the LA group ($t(4) = -7.74$, $p = 0.0015$, $\alpha = 0.05$).

Since *BG5* was identified by NormFinder as the gene with the lowest stability value (Table 4.2), average fold change in *BG4* expression was also normalized to both *RNA2D* and *BG5* for comparative purposes among groups (Figure 4.4a-b). Average *BG4* fold changes in expression are slightly higher when normalized to *RNA2D* than *BG5*. The standard error bars show the large variation in *BG4* expression within the PBS, PA and LA groups and the low variation within the LA group. The LA group shows higher average *BG4* expression compared to the other three test groups. Differences in treatment group means are not significantly different from the control group means (two-tailed T test assuming unequal variances, $p > 0.05$), likely due to the high standard deviations in the PBS, PA, and LB groups (Table 4.4a; Figure 4.2). Therefore, this study provides no evidence of either *Lactobacillus* supplementation or aflatoxin exposure significnatly affecting the expression of *BG4* or *BG5* in the liver. However, the consistent upregulation of *BG4* in the LA group suggests a co-stimulatory effect of both treatments (*Lactobacillus* and aflatoxin).

DISCUSSION

BG expression profiles

Tissue-based expression profiles can provide important clues to gene function. Early studies in the chicken indicated close linkage of the B-G genes to class I and class II genes (Simonsen, 1980) and the close physical proximity of these loci, separation of only 40 kb, was later established by Kaufman (1999). The function of the class I and class II genes in antigen presentation has been extensively studied and definitively correlated with resistance to a wide variety of pathogens (reviewed in Lamont 1998). Evidence suggests the B-G genes are more polymorphic than members of the class I and class II families, but the functional significance of this property is unknown (Kaufman et al., 1990; Li et al., 1999; Miller et al., 1988; Nishibori et al., 2000; Ruby et al., 2005). This study reports the first evidence for expression of multiple B-G genes in a panel of tissues and the first description of *BG3* expression in any tissue.

Although B-G molecules are described as cell surface receptor molecules (Goto et al., 2009), the function of B-G genes has not yet been elucidated. Several functions have been postulated including interaction between cells, the extracellular matrix or cytoskeleton, and B cell selection and interaction with class I molecules, but these hypotheses have never been directly tested (Kaufman et al., 1991). The expression of multiple B-G genes in the examined tissues is suggestive of gene-specific signaling functions. However, it is likely that B-G genes have multiple functions, some of which may be tissue specific and may not be related to the immune function. The latter seems most likely of *BG5*, the most ubiquitously expressed B-G gene.

Evidence provided by previous studies consistently hypothesize cell signaling and regulatory functions for B-G molecules. Receptors with IgV domains have been shown

to have a variety of functions including, glycoprotein binding and viral receptors (Munguia and Federspiel, 2008; Capello et al., 2008). B-G genes share Ig domain similarities with the mammalian butyrophilin immunoglobulin superfamily (*BTN*) and myelin oligodendrocyte glycoprotein (*MOG*) which contains a single IgV-like domain (Pham-Dinh et al., 1995). Members of the butyrophilin family generally have both an IgV domain and an IgC domain, whereas B-G genes, like NK receptors, have only one IgV. MOG is located within the mammalian MHC, has a single IgV domain, is able to elicit a strong immune response and is expressed only in the CNS. Its function is unknown, but it has been implicated as an autoantigen (Bernard et al., 1997). B-G genes share these characteristics with MOG, but their role as autoantigens has not been investigated.

B-G genes are structurally most similar to NK receptors. Like the NK family, B-G genes have high sequence homology in the extracellular domain, and the IgV-like and the transmembrane domains, but there is a lack of homology between genes in the exons encoding the cytoplasmic tail. Both gene families are also highly polymorphic in length. Non-synonymous mutations between B-G genes supports functional differences between genes in this family. Within gene comparisons among individuals show pockets of nucleotide variation in the signal peptide region and intermittently throughout the IgV domain, some of which are non-synonymous mutations (Chapter 3).

Evidence supporting tissue specific functions of B-G genes is the exclusive expression of *BG2* in whole blood and of *BG3* in intestinal epithelium, the only tissue in which *BG3* transcription has been observed. Typical of genes with specialized functions,

conserved sequences in the signal peptide of *BG2* have been identified among wild, heritage and commercial birds (Chapter 3). Comparisons of *BG1* and *BG3* sequences among individuals show greater similarity to each other than to other genes in the family (Chapter 3). Interestingly, there is no evidence of co-expression of these two genes in any of the examined tissues, inviting the possibility that these genes have similar but, tissue-dependent functions.

B-G genes have variable cytoplasmic tail lengths and some may have motifs in the cytoplasmic domain that affect cellular regulation. For example, a *BG1* variant in the chicken contains an immunoreceptor tyrosine-based inhibitory motifs (ITIM) located in the tail (Goto et al., 2009). ITIMs bind to phosphatases and control the intensity of the signal received by the cell. ITIMs have not been identified in turkey B-G genes.

Turkey B-G transcripts have different lengths, due to variation in the cytoplasmic tail region (Chapter 3). There is evidence in chicken for variation between individuals in the length of B-G cytoplasmic tails (Kaufman et al., 1990). Length of the cytoplasmic tails in human killer cell immunoglobulin-like receptors (KIR) and chicken immunoglobulin-like receptor (CHIR) genes, families of NK receptor molecules, is correlated with the presence or absence of motifs. Molecules with longer tails contain ITIMs in their cytoplasmic domains and associate with their ligand, MHC class I molecules. The receptor molecules with shorter tails interact with adaptor molecules containing immunoreceptor tyrosine activation motifs (ITAM) and are responsible for cellular activation (Barrow and Trowsdale, 2008; Viertlboeck et al., 2004, 2007). Given the importance of tail length in receptor function, the varying intergenic and intragenic

tail lengths in poultry and the presence of ITIMs in some B-G variants in the chicken, it would be pertinent to characterize the further cytoplasmic domains of B-G genes, which may play a role in individual expression differences.

B-G expression in response to Lactobacillus and aflatoxin treatment

Transcripts of four B-G genes were found present in turkey liver tissue, making it a complex target for analysis of B-G expression. In general, low level of overall B-G expression was observed in this tissue. Although levels of *BG1* could not be definitively determined due to the likely amplification of multiple gene products, stability values for *BG1* were consistent with limited gene expression. *BG2*, although expressed, showed consistently high Ct values (35-39) for all samples, also indicating minimal expression. Lliekwise, low and invariant expression of *BG5* was observed in all individuals across treatment groups (Table 4.1A). Given the invariant expression of this gene, it would be an ideal control gene in future studies of liver tissue. In contrast, *BG4* showed wide, intergroup variation in transcript number for three of the four treatment groups, supporting gene-specific function in this tissue. The difference in expression pattern suggests a variant function of *BG4* in the liver compared to the other B-G genes.

Individual variance of *BG4* expression may be related to outside stimulus or perhaps reflect sequence variation between individuals. Epigenetic alterations, regulatory DNA sequence elements in the promoter regions or motifs located within gene variants may cause differences in expression among individuals. Subclinical infections and other environmental stimuli may also alter the expression of immune-related genes. One study

investigating the effects of aflatoxin on wild turkey poults reported deaths of birds due to effects not related to aflatoxicosis (Quist et al., 2000). These deaths were attributed to capture stress, ingestion of foreign bodies, synovitis, *Staphylococcus* infection and unknown cause (Quist et al., 2000). Clearly, young poults are susceptible to a variety of conditions that could alter B-G expression and which may not have been detected during the duration of the present study.

Variance in B-G liver expression is particularly obvious when expression levels of *BG4* and *BG5* are compared. No statistically significant differences in B-G expression were observed in response to aflatoxin challenge or *Lactobacillus* supplementation in the exposure groups compared to controls. It is therefore, unlikely that these genes are directly involved in aflatoxicity. However, birds supplemented with *Lactobacillus* and subsequently challenged with aflatoxin (LA group) showed consistent up-regulation of *BG4*, possibly due to an effect in the liver caused by the co-presence of aflatoxin and *Lactobacillus*. The mechanism for AFB_1 binding by probiotic bacteria in the gut is not known, but AFB_1 likely binds to cell wall components of the bacteria, with published evidence implicating peptidoglycans and teichoic acids as binding sites (Hernandez-Mendoza et al., 2008; Latingen et al., 2004).

Acute aflatoxicosis causes stressful effects such as acute hepatic injury, biliary obstruction, degeneration, necrosis and hypofunction with more severe cases causing liver cancer. Ancillary results from this study (Rawal, 2010), show significant differences in biliary hyperplasia and hepatic necrosis between the control (PBS and LB) and the treatment groups (PA and LA). Liver and body weight was significantly

decreased within the PA group but was rescued to levels of the control group in the LA group. Cytokine levels showed high variation within group. IL-8 and IL-1β were upregulated while IL-8 and IL-6 were downregulated in the PA group compared to the LA group, but differences were not significant. Oxidative damage of lipids and DNA is caused by the reactive intermediate AFB_1-8,9-epoxide with lipid peroxidation specifically known to cause liver membrane damage (Abdel-Waheb et al., 2003). Hepatocarcinogenicity in humans is at least in part mediated by an AFB_1 induced base substitution within the p53 gene (Hsu et al., 1991).

The results of this study clearly demonstrate a *BG4* response to the Lactobacillus + aflatoxin treatment. Comparison with the other treatment groups suggests this response is unique to LA group, however additional studies with an increased number of individuals are needed to confirm these findings. Given the co-presence of *B-G3-5* in the small intestine and its known role as the site of *Lactobacillus* binding to AFB1, future aflatoxin challenge studies using larger sample size will help elucidate the signaling function of this molecule and may provide insights to the function of *BG3*

Gene	Primer Sequence	Amplicon Size (RNA)	PCR Efficiency
BG1	F-ACTGCTCAGGGATGGTCTCC R-ATGACAAATGACCCAAGCAG	162bp	75.4%
BG2	F-GGAACTCAGACATCAGATGGG R-GCAAATGACCCAAGCAGAAT	320bp	73.1%
BG4	F-GGATCAGCCCAGCTCACAG R-TAACCCAACCAGAAGTGTGAC	400bp	74.7%
BG5	F-GACCCTCCTGGCTCATCTCA R-CCTGAGCAGTTCTGTCCTCT	260bp	94.1%
β-actin	F-CAGACATCAGGGTGTGATGG R-GCCTTCATAGATGGGCACAG	391bp	87.2%
GAPDH	F-CATCTGAAGGGTGGTGCTAAG R-AACAGAGACATTGGGGGTTG	396bp	105.9%
RNA2D	F-AGAATGAGAGCGCGGAAG R-GCTCCCTTTATCCACTGCTC	394bp	86.9%

Table 4.1. Primers used for qRT PCR, amplicon size and reaction efficiency for each gene

Tissue	*BG1*	*BG2*	*BG3*	*BG4*	*BG5*
Spleen	+	+			+
Liver	+	+		+	+
Heart	+	+			+
Bursa	+	+			+
Intestinal Epi.			+	+	+
Muscle	+	+			+
Whole Blood		+			

Table 4.2. Expression of B-G genes as determined by PCR amplification from cDNA templates prepared from total RNA of non-challenged birds.

Rank	Gene	Stability value	Standard error
1	*BG5*	0.016	0.014
2	*RNA2D*	0.035	0.012
3	*B-actin*	0.041	0.013
4	*BG1*	0.044	0.013
5	*GAPDH*	0.063	0.016
6	*BG2*	0.082	0.019
7	*BG4*	0.121	0.026

Table 4.3. Stability values and standard error of each potential reference and test gene as determined by NormFinder (Andersen et al., 2004). Ct values were log transformed and genes with the lowest stability value were considered to have the least change in expression between samples.

A.

	BG4		*BG5*		*RNA2D*	
Group	**T/ng**	**SD**	**T/ng**	**SD**	**T/ng**	**SD**
PBS	4.11E+03	6.90E+03	1.99E+02	9.85E+01	3.25E+02	4.32E+01
PA	1.03E+04	1.70E+04	5.21E+02	2.30E+02	9.15E+02	4.39E+02
LB	2.79E+03	4.31E+03	2.67E+02	1.25E+02	6.73E+02	8.58E+01
LA	1.89E+04	9.22E+03	4.15E+02	2.11E+02	1.14E+03	9.75E+02

B.

Group	***BG4***	***BG5***	**T**	**p value**
PBS	2.96 +/- 0.13	3.44 +/- 0.68	-0.69	0.56
PA	3.39 +/- 0.11	3.97 +/- 0.61	-0/94	0.45
LB	3.10 +/- 0.11	3.65 +/- 0.47	-1.14	0.37
LA	3.27 +/- 0.15	4.93 +/- 0.15	-7.75	0.0007
All groups	3.18 +/- 0.07	4.00 +/- 0.28	-2.82	0.015

Table 4.4 A-B. Summary of B-G expression analysis by group. A. For each group the average transcript numbers per ng RNA (T/ng) and standard deviations as determined for each gene (*BG4*, *BG5* and *RNA2D*) are given. **B.** Average log fold changes in transcript number were calculated for *BG4* and *BG5* and subjected to a T-test to determine significant differences between these two genes within each group. For each gene, the mean, T value and p-values are given

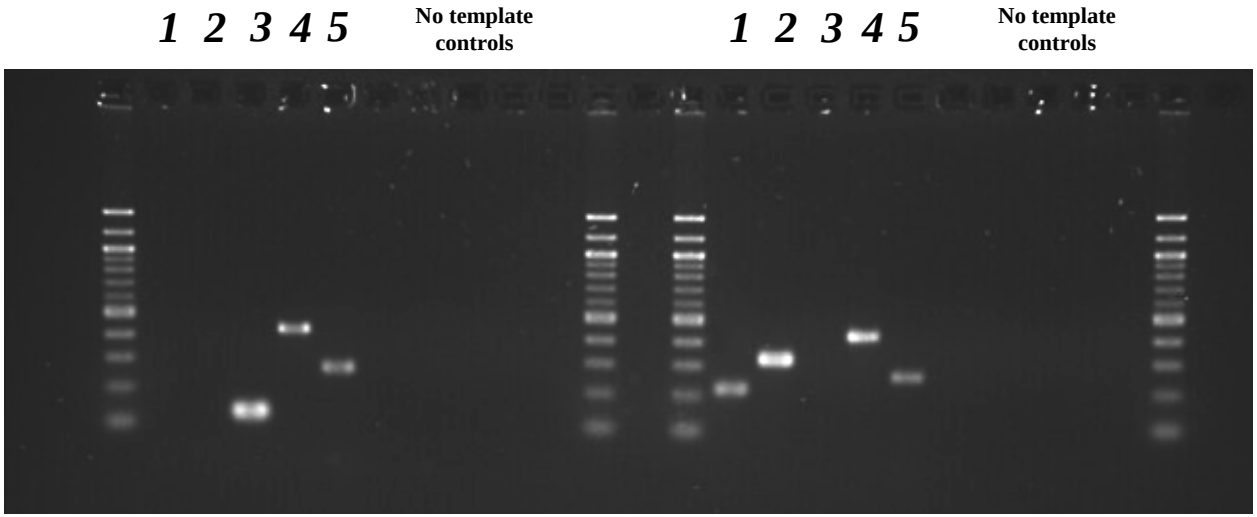


Figure 4.1. *BG1-5* PCR products amplified with locus-specific primers from RNA isolated from ileum (left) and liver (right) tissue. For each tissue the B-G PCR products and no template negative controls are included for comparison. Product sizes are as indicated by the 100 bp ladder.

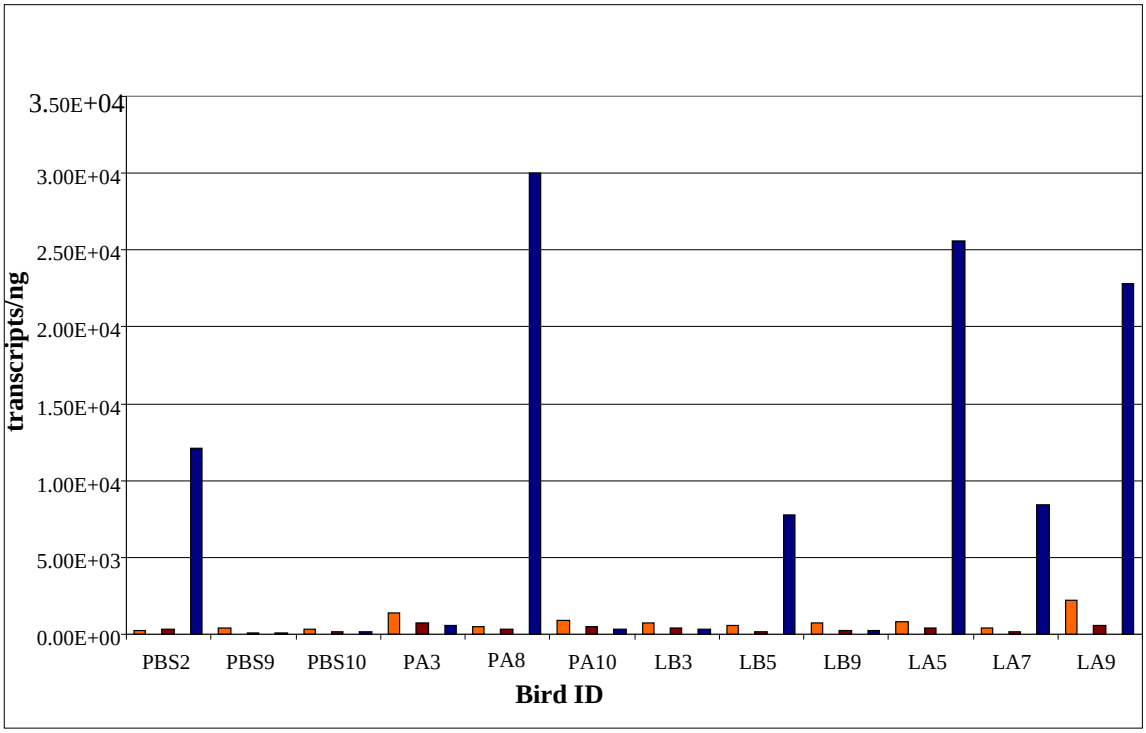


Figure 4.2. Gene RNA transcript numbers determined by qRT-PCR using gene-specific standard curves. *RNA2D*, *BG5* and *BG4* levels represented by orange, red and dark blue, respectively.

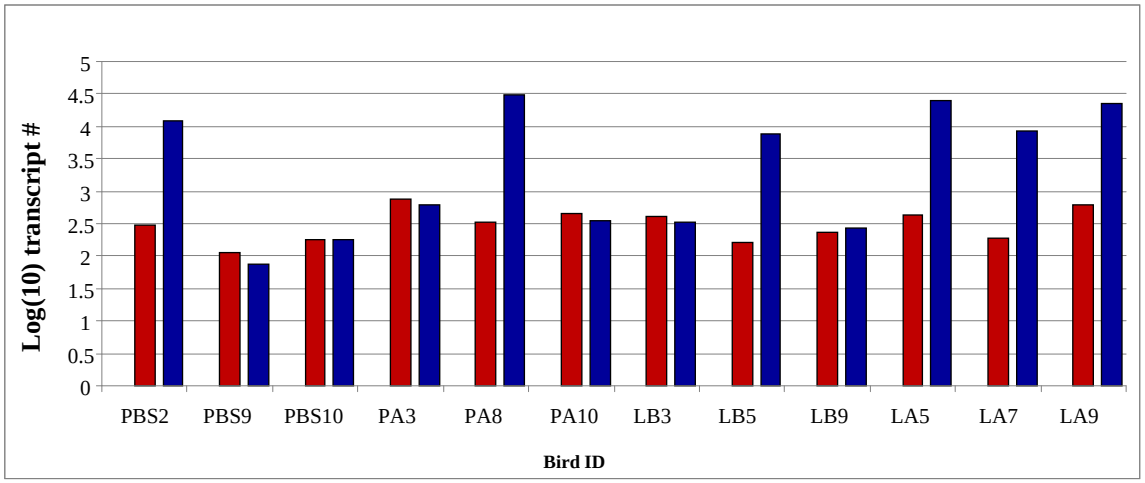
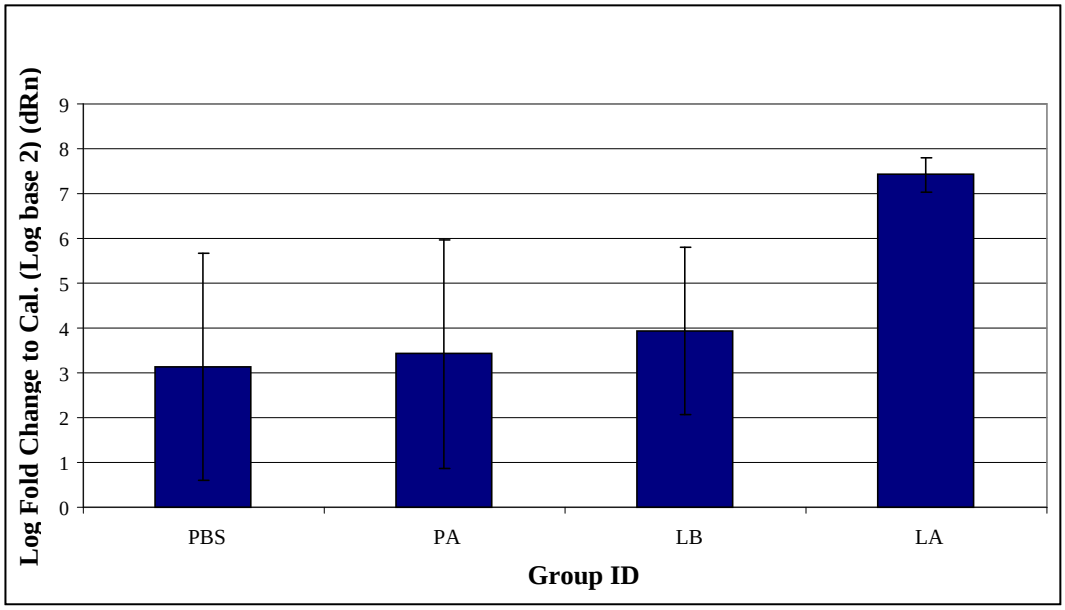


Figure 4.3. Log fold change in *BG5* and *BG4* gene RNA transcript number observed in liver tissue across treatment groups normalized to *RNA2D*.

A.



B.

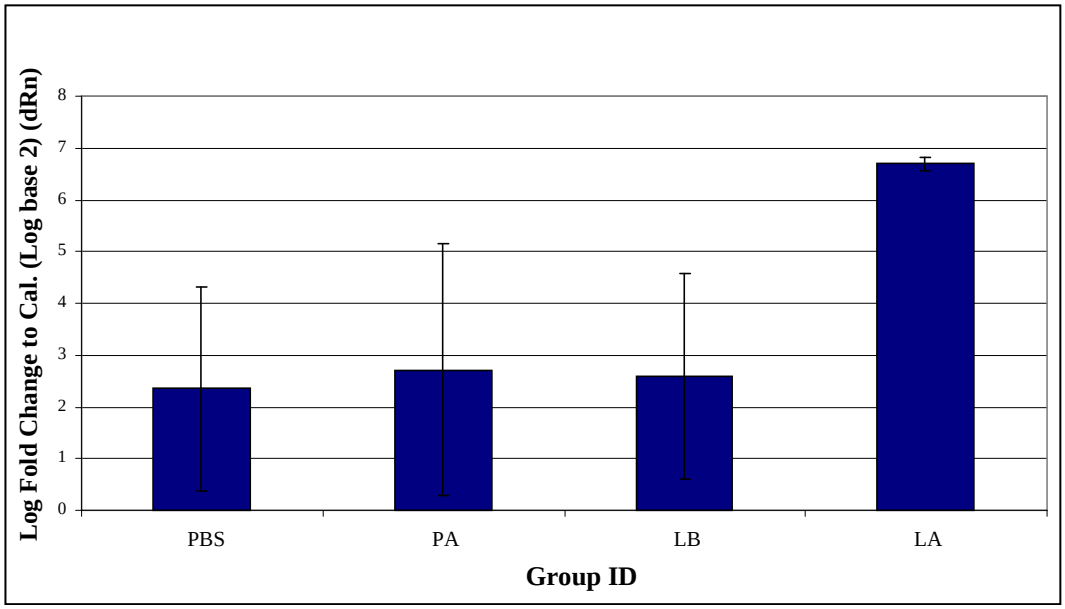


Figure 4.4 A-B. Average expression levels of *BG4* in liver tissue across treatment groups normalized to *RNA2D* (A) *and BG5* (B).

Chapter 5

SUMMARY AND FUTURE WORK

This study was designed to further characterize the B-G genes in the MHC *B*-locus of the turkey (*Meleagris gallopavo*). Direct 454 sequencing of a clone from the CHORI-260 BAC library was used to extend the turkey MHC *B*-locus sequence, identifying two additional B-G genes in the species. These B-G genes show similar exon patterns as those of the chicken and the three previously described B-G genes of the turkey, and through mRNA sequencing are shown to be functional. The two new B-G genes in the extended *B*-locus were compared to the previously identified B-G genes, which show cytoplasmic domains of variable lengths between genes. Variation between cytoplasmic domains has been linked to differences in function between other receptor families, such as those in the CHIR and KIR families (Chapter 3), which like the B-G family, also contain Ig-V like domains. The turkey BAC ends within 3' sequence of an additional B-G gene, which has not yet been fully sequenced. Future studies aim to complete the sequence of the turkey MHC B-locus in order to identify all B-G genes. Given the high sequence homology between genes, this will reduce the chance of amplifying multiple loci in diversity studies.

Examination of interspecies phylogenetic relationships and intraspecies variation in the signal pepide region (exon 1) of B-G genes provides new insights into the pattern of evolution and importance of these unique members of the avian MHC. The similarity in the position of these genes among avian species supports an initial duplication of these loci in a common ancestor. Comparisons of the turkey B-G genes support a birth-death

model of evolution acting in tandem with recombination processes such as gene conversion, unequal crossover and exon shuffling. Locus-specific segregation suggests divergence in function of these molecules. Analysis of variation within the signal peptide sequence confirmed multiple alleles at each B-G gene and found two clusters of polymorphism among the turkey B-G genes.

Variation between the B-G genes was compared between populations of individuals and between groups (wild, heritage and domestic) of birds to aid in elucidation of possible immune related differences. Identifying individual and group splice variants and differences in regulatory motifs in the cytoplasmic tail using RNA is crucial to completely characterizing diversity of the B-G genes, given their potential signaling function. The polymorphic nature of these genes, particularly evident in the signal peptide and cytoplasmic tail, may have effects on the protein structure and function of these genes. Other future studies aim to investigate variation in the IgV region and splice variants in the cytoplasmic tail lengths both between individuals and among groups of turkeys. Variation in the IgV region may affect ligand binding properties and differences in cytoplasmic tail length may, as in the chicken, include in the presence of ITIMs on certain molecules. It will be necessary to examine several cell types to characterize the cellular distribution of these proteins and determine locus-specific differences in expression. Likewise, protein studies are needed to elucidate the behavior of the signal peptide and functional properties of other regions of the B-G proteins, including the IgV region and the cytoplasmic tail. The ligands of these receptors remain to be discovered, but there is evidence of B-G proteins interacting with other MHC molecules, and other Ig-like

receptors have proven essential for Ig binding and crosstalk between the innate and adaptive immune system.

Variability of *BG5* was investigated using a population of wild turkeys based on previous data indicating it as the most variable of the described turkey B-G genes (Chapter 3). Southern blots of the B-G region indicate levels of diversity exceeding those of the Class IIβ genes, but whether this is due to individual variations in gene sequences or differences in copy number cannot be determined by the banding patterns. *BG5* investigations demonstrated the pitfalls of investigating gene families with high sequence homology. The unique amplicons sequenced in select individuals were most similar to *BG5*, and may indicate a duplicated locus or copy number variation present in this wild population. The total number of B-G genes has yet to be determined in the turkey and copy number variations between domestic and wild individuals are likely. Thus it cannot be established if the unique alleles (or loci) observed are exclusive to wild birds. Regardless, RFLP-SNP haplotyping proved a more informative measure of MHC variation than the Southern Blot method. Furthermore, MHC variation in the Eastern wild turkeys sampled in Wisconsin was intermediary between that of commercial birds and wild turkeys from other areas of the United States. This corroborates evidence supporting a loss of MHC diversity in commercial birds and may indicate a loss of MHC diversity at certain loci resulting from the reintroduction effort in Wisconsin. Measuring MHC diversity using individuals located across a wider geographical range in Wisconsin and comparing this to individuals in Missouri, the parent population, would more accurately test whether MHC variation was in fact lost as a result of the reintroduction bottleneck.

This study provides the first evidence for expression of multiple B-G genes in a panel of tissues and the first description of *BG3* expression in any tissue. Multiple B-G genes are co-expressed in a variety of tissues with evidence of exclusive *BG2* expression in whole blood and intestinal epithelium the only tissue in which *BG3* expression was observed. Given the presence of multiple B-G genes in the turkey and the variability in B-G tissue expression, gene-specific immune function in particular tissues is likely. This knowledge can be used to generate further hypotheses regarding the function of these molecules. Turkey B-G expression in immune related cells, such as thrombocytes, mast cells. macrophages, lymphocytes and leukocytes has yet to be characterized, but is likely to provide important insights into the function of this gene family. It would be interesting to test differences in expression in various immune tissues (i.e. spleen, bursa), resulting from viral and bacterial infections, such as avian influenza, *salmonella*, Newcastle disease and avian pneumovirus to aid in delineation of B-G gene function related to immunity. Expression changes in response to vaccination to viral diseases would also provide information about the function of these genes.

B-G expression in liver tissue of domestic turkeys in response to aflatoxin challenge and L*actobacillus* treatment show generally low levels of expression with the exception of *BG4*. In contrast to the consistent, low expression levels observed in the other liver B-G genes, *BG4* showed wide, intergroup variation in three of the four treatment groups. The difference in expression pattern suggests a variant function of *BG4* compared to the other B-G genes in the liver, but more individuals are needed to make definitive conclusions.

In birds supplemented with *Lactobacillus* and challenged with aflatoxin (LA group), consistent up-regulation of *BG4* occurred, possibly resulting from the neutralizing effect of *Lactobacillus* on aflatoxin. This effect is particularly obvious when expression levels of *BG4* and *BG5* are compared. The results of this study demonstrate a *BG4* response to the Lactobacillus + aflatoxin treatment. This is interesting, given the presence of at least three other B-G genes in that tissue, which show high sequence homology. It cannot be ruled out that the function of *BG4* is unrelated to immune response.

Future aflatoxin challenge studies measuring B-G expression changes in additional tissues such as the small intestine, where *BG4* is known to be expressed, and using more animals will help elucidate the signaling function of this molecule and may provide insights to the function of *BG3*. Epigenetic alterations, regulatory DNA sequence elements in the promoter regions or motifs located within gene variants may cause differences in expression among individuals and provide future targets of study. Sequence homology between the 5' UTR and upstream regions of the B-G genes is low and promoters for these genes have not yet been described in any species. Given the close proximity of the B-G genes in the *B*-locus as well as the extended region, it is possible for one promoter to control each gene complex. *In silico* analysis using software designed to identify promoter regions has not defined any potential promoter sites within upstream regions of the predicted translational start sites of the 5 described turkey B-G genes. Using 5' RACE, the transcriptional start sites of each gene can be identified, which may aid in defining promoters and other regulatory regions influencing B-G gene

expression. Identification of possible sequence variations in these regions through cloning and subsequent transfection into turkey cell lines may further clarify functional differences between genes.

Clearly, classification of B-G genes as immunoglobulin-like superfamily members and evidence of cell surface expression indicate a receptor or adhesion function. Their location in the MHC further suggests an immune function. The effect of the highly polymorphic family of B-G receptor molecules on the innate and adaptive immune responses await further investigation, but their multiplicity and linkage with the class I and class II molecules makes them an interesting and unique area of study. Further investigations characterizing an immune expression profile in specific cell types and protein studies will help further characterize the structure and function of this intriguing family.

REFERENCES

van der Aa LM, Jouneau L, Laplantine E, Bouchez O, Van Kemenade L, Boudinot P (2012) FinTRIMs, fish virus-inducible proteins with E3 ubiquitin ligase activity. *Dev Comp Immunol*, 36, 433-41.

Abdel-Wahhab MA, Aly SE (2003) Antioxidants and radical scavenging properties of vegetable extracts in rats fed aflatoxin-contaminated diet. *J Agric Food Chem,* 51, 2409–2414.

Afanassieff M, Goto RM , Ha J, Sherman MA, Zhong L, Auffray C, Coudert F, Zoorob R, Miller MM (2001) At least one class I gene in restriction fragment pattern-Y (*Rfp-Y*), the second MHC gene cluster in the chicken, is transcribed, polymorphic, and shows divergent specialization in antigen binding region. *J Immunol,* 1666, 3324-3333.

Agudo R, Alcaide M, Rico C, Lemus JA, Blanco G, Hiraldo F, Donazar JA (2011) Major histocompatibility complex variation in insular populations of the Egyptian vulture: inferences about the roles of genetic drift and selection. *Mol Ecol,* 20, 2329-2340.

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *PNAS USA*, 101, 3490-3494.

Alcaide M, Lemus JA, Blanco G, Tella JL, Serrano D, Negro JJ, Rodriguez A,Garcia-Montijano M (2010) MHC diversity and differential exposure to pathogens in kestrals (Aves: *Falconidae*). *Mol Ecol,* 19, 691-705.

Andersen CL, Jensen FL, Orntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res*, 64, 5245-5250.

Arafa AS, Bloomer RJ, Wilson HR, Simpson CF, Harms RH (1981) Susceptibility of various poultry species to dietary aflatoxin. *Br Poult Sci*, 22, 431-436.

Aspanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Crit Rev Immunol*, 17, 179-224.

Axelsson E, Webster MT, Smith NGC, Burt DW, Ellegren H (2005) Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Res*, 15, 120-125.

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.

Bacon LD, Hunt HD, Cheng HH (2001) Genetic resistance to Marek's disease. *Curr Top Microbiol Immunol*, 255, 121-141.

Bacon LD, Witter RL (1993) Influence of B-haplotype on the relative efficacy of Marek's disease vaccines of different serotypes. *Avian Dis*, 37, 53-59.

Bacon LD, Witter RL (1995) Efficacy of Marek's disease vaccines in MHC heterozygous chickens: MHC congenic x inbred line F1 matings. *J Hered*, 86, 269-273.

Ballingall KT, Rocchi MS, McKeever DJ, Wright F (2010) Trans-species polymorphism and selection in the MHC class II DRA genes of domestic sheep. *PLOS One*, 5:e11402.

Barrow AD, Trowsdale J (2008) The extended human leukocyte receptor complex: diverse ways of modulating immune responses. *Immunol Rev*, 224, 98-123.

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

Bennett RS, Nezworski J, Velayudhan BT, Nagaraja KV, Zeman DH, Dyer N, Graham T, Lauer DC, Njenga MK, Halvorson DA (2004) Evidence of avian pneumovirus spread beyond Minnesota among wild and domestic birds in central North America. *Avian Dis,* 48, 902-908.

Bernard CCA, Johns TG, Slavin A, Ichikawa M, Ewing C, Liu J, Bettadapura J (1997) Myelin oligodendrocyte glycoprotein: a novel candidate autoantigen in multiple sclerosis. *J Mol Med,* 75, 77-88.

Bikle DD, Munson S, Komuves L (1996) Zipper protein, a B-G protein with the ability to regulate actin/myosin interactions in the intestinal brush border. *J Biol Chem*, 271, 9075-9083.

Blount WP (1961) Turkey “X” disease. *J Brit Turk Found*, 1, 52, 55, 61.

Bollmer, JL, Hull JM, Ernest HB, Sarasola JH, Parker PG (2011) Reduced MHC and neutral variation in the Galapagos hawk, an island endemic. *BMC Evol Biol*, 11, 143.

Bonneaud C, Pérez-Tris J, Federici P, Chastel O, Sorci G (2006) Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution*, 60, 383-389.

Bonneaud C, Richard M, Faivre B, Westerdahl H, Sorci G (2005) An MHC class I allele associated to the expression of T-dependent immune response in the house sparrow. *Immunogenetics*, 57, 782-789.

Bonneaud C, Sorci G, Morin B, Westerdahl H, Zoorob R, Wittzell H (2004) Diversity of Mhc class I and IIB genes in house sparrows (*Passer domesticus*). *Immunogenetics,* 55, 855-865.

Borrego F, Ulbrecht M, Weiss EH, Coligan JE, Brooks AG (1998) Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *J Exp Med*, 187, 813-818.

Braud VM, Allan DS, O'Callaghan CA, Soderstrom K, D'Andrea A, Ogg GS, Lazetic S, Young NT, Bell JI, Philips JH, Lanier LL, McMichael AJ (1998) HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature*, 391, 795-799.

Braud VM, Jones EY, McMichael A (1997) The human major histocompatibility complex class Ib molecule HLA-E binds signal sequence-derived peptides with primary anchor residues at positions 2 and 9. *Eur J Immunol*, 27, 1164-1169.

Briles WE, Goto RM, Auffray C, Miller MM (1993) A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. *Immunogenetics,* 37, 408-414.

Briles WE, McGibbon WH (1948) Heterozygosity of inbred lines of chickens at two loci affecting cellular antigens. *Genetics*, 33, 605.

Briles WE, McGibbon WH, Irwin MR (1950) On multiple alleles effecting cellular antigens in the chicken. *Genetics,* 35, 633–652.

Brower L, Barr I, van de Pol M, Burke T, Komdeur J, Rechardson DS (2010) MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Mol Ecol,* 19, 3444-3455.

Buchholz R, Jones Dukes MD, Hecht S, Findley AM (2004) Investigating the turkeys'snood' as a morphological marker of heritable disease resistance. *J Anim Breed Genet*, 121, 176-185.

Capello D, Martini M, Gloghini A, Cerri M, Rasi S, Deambrogi C, Rossi D, Spina M, Tirelli U, Larocca LM, Carbone A, Gaidano G (2008) Molecular analysis of immunoglobulin variable genes in human immunodeficiency virus-related non-Hodgkin's lymphoma reveals implications for disease pathogenesis and histogenesis. *Haematologica*, 93, 1178-1185.

Chaves LD, Faile GM, Hendrickson JA, Mock KE, Reed KM (2011) A locus-wide approach to assessing variation in the avian MHC: the *B*-locus of the wild turkey. *Heredity*, 107, 40-49.

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, Krueth SB, Reed KM (2009a) Defining the turkey MHC: sequence and genes of the B locus. *J Immunol*, 183, 6530-6537.

Chaves LD, Ostroski BJ, Reed KM (2009b) 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.

Commane D, Hughes R, Shortt C, Rowland IR (2005) The potential mechanisms involved in the anti-carcinogenic action of probiotics. *Mut Res*, 591, 276–289.

Cutler G, Marshall LA, Chin N, Baribault H, Kassner PD (2007) Significant gene content variation characterizes the genomes of inbred mouse strains. *Genome Res*, 17, 1743-1754.

Dalgaard TS, Vitved L, Skjodt K, Thomsen B, Labouriau R, Jensen KH, Juul-Madsen HR (2005) Molecular characterization of major histocompatibility complex class I (B-F) mRNA variants from chickens differing in resistance to Marek's disease. *Scandinavian Journal of Immunolog*, 62, 259-270.

Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg Le Ann, Bouffard P, Burt DW, Crasta O, Crooijmans RP, Cooper K,Coulombe RA, De S, Delany ME, Dodgson JB, Dong JJ, Evans C, Frederickson KM, Flicek P, Florea L, Folkerts O, Groenen MA,Harkins TT, Herrero J, Hoffmann S, Megens HJ, Jiang A, de Jong P, Kaiser P, Kim H, Kim KW, Kim S, Langenberger D, Lee MK,Lee T, Mane S, Marcais G, Marz M, McElroy AP, Modise T, Nefedov M, Notredame C, Paton IR, Payne WS, Pertea G, Prickett D,Puiu D, Qioa D, Raineri E, Ruffier M, Salzberg SL, Schatz MC, Scheuring C, Schmidt

CJ, Schroeder S, Searle SM, Smith EJ,Smith J, Sonstegard TS, Stadler PF, Tafer H, Tu ZJ, Van Tassell CP, Vilella AJ, Williams KP, Yorke JA, Zhang L, Zhang HB, Zhang X, Zhang Y, Reed KM (2010) Multi-platform next generation sequencing of the domestic turkey (*Meleagris gallopavo*): Genome assembly and analysis. *PLoS Biol*, 8:e1000475. doi:10.1371/journal.pbio.1000475.

Da-Silva FA, Oluwafemi F (2009) Removal of aflatoxins by viable and heat-killed *Lactobacillus* species from fermented maize. *J Appl Biosci*, 16, 871-876.

Debenham SL, Hart EA, Ashurst JL, Howe KL, Quail MA (2005) Genomic sequence of the class II region of the canine MHC: comparison with the MHC of other mammalian species. *Genomics*, 85, 48-59.

Delarbre C, Jaulin C, Kourilsky P, Gachelin G (1992) Evolution of the major histocompatibility complex: a hundred-fold amplification of MHC class I genes in the African pigmy mouse *Nannomys setulosus*. *Immunogenetics*, 37, 29-38.

Deshpande A, Wheeler CM, Hunt WC, Peyton CL, White PS, Valdez YE**,** Nolan JP (2008) Variation in HLA class I antigen-processing genes and susceptibility to human papillomavirus type 16-associated cervical cancer. *J Infect Dis*, 197, 371-381.

Diaz GJ, Murcia HW, Cepeda SM (2010) Bioactivation of aflatoxin B1 by turkey liver microsomes: responsible cytochrome P450 enzymes. *Br Poult Sci*, 51, 828-837.

Eimes JA, Bollmer JL, Whittingham LA, Johnson JA, van Oosterhout C, Dunn PO (2011) Rapid loss of MHC class II variation in a bottlenecked population is explained by drift and loss of copy number variation. *J Evol Biol*, 24, 1847-1856.

Ekblom R, Saether A, Grahn M, Fiske P, Kalass JA, Hoglund J (2004) Major histocompatibility complex variation and mate choice in a lekking bird, the great snipe (*Gallinago media*). *Mol. Ecol*, 13, 3821-3828.

El Nazami H, Mykkannen H, Kankaanpan P, Salminen, Ahokas J (2000) Ability of *Lactobacillus* and *Propionibacterium* strains to remove aflatoxin B1 from the chicken duodenum. *J Food Prot*, 63, 549-552.

Emara MG, Nestor KE, Bacon LD (1993) The turkey major histocompatibility complex: characterization by mixed lymphocyte, graft-versus-host splenomegaly, and skin graft reactions. *Poult Sci*, 72, 60-66.

Ewald SJ, Livant EJ (2004) Distinctive polymorphism of chicken B-FI (major histocompatibility complex class I) molecules. *Poult. Sci*, 83, 600-605.

Freeman C, Meguerdichian M, Wheelwright N, Sollecito S (2003) Social pairing and female fidelity predicted by restriction fragment length polymorphism similarity at the major histocompatibility complex in a songbird. *Mol Ecol*, 12, 3077-3083.

Freeman-Gallant CR, Johnson EM, Saponara F, Stanger M (2002) Variation at the major histocompatibility complex in Savannah sparrows. *Mol Ecol*, 11, 1125–1130.

Gack MU (2011) TRIMming Flavivirus infection. *Cell Host Microbe*, 10, 175-177.

Gao J, Liu K, Liu H, Blair HT, Li G, Chen C, Tan P, Ma RZ (2010) A complete DNA sequence map of the ovine major histocompatibility complex. *BMC Genomics*, 11, 466.

Gehad AE, Mashaly MM, Siegel HS, Dunnington EA, Siegel PB (1999) Effect of genetic selection and MHC haplotypes on lymphocyte proliferation and interleukin-2 like

activity in chicken lines selected for high and low antibody production against sheep red blood cells. *Vet Immunol Immunopathol*, 68, 13-24.

Goto RM, Afanassieff M, Ha J, Iglesias GM, Ewald SJ, Briles WE, Miller MM (2002) Single-strand conformation polymorphism (SSCP) assays for major histocompatibility complex B genotyping in chickens. *Poult Sci*, 81, 1832-1841.

Goto R, Miyada CG, Young S, Wallace B, Abplanalp H, Bloom S, Briles WE, Miller MM (1988) Isolation of a cDNA Cl1 from the B-G subregion of the chicken histocompatibility (B) complex. *Immunogenetics*, 27, 102-109.

Goto R, Wang Y, Taylor Jr R, Wakenell P, Hosomichi K, Shiina T, Blackmore C, Briles E, Miller M (2009) BG1 has a major role in MHC-linked resistance to malignant lymphoma in the chicken. *PNAS USA*, 106, 16740-16745.

Goyal SM, Lauer D, Friendshuh K, Halvorson DA (2003) Seroprevalence of avian pneumovirus in Minnesota turkeys. *Avian Dis*, 47, 700-706.

Haeri M, Read LR, Wilkie BN, Sharif S (2004) Identification of peptides associated with chicken major histocompatibility complex class II molecules of B21 and B19 haplotypes. *Immunogenetics*, 56, 854-859.

Hala K, Plachy J, Schulmannova J (1981) Role of B-G-region antigen in the humoral immune response to the B-F-region antigen of chicken MHC. *Immunogenetics*, 14, 393-401.

Hatakeyama S (2011) TRIM proteins and cancer. *Nat Rev Cancer*, 11, 792-804.

Hathout AS, Mohamed SR, El-Nekeety AA, Hassan NS, Aly SE, Abdel-Wahhab MA (2011) Ability of *Lactobacillus casei* and *Lactobacillus reuteri* to protect against oxidative stress in rats fed aflatoxins-contaminated diet. *Toxicon,* 58, 179-186.

Hegde RS, Bernstein HD (2006) The surprising complexity of signal sequences. *Trends Biochem Sci*, 31, 563-571.

Hepkema BG, Blankert JJ, Albers GA, Tilanus MG, Egberts E, van der Zijpp AJ, Hensen EJ (1993) Mapping of susceptibility to Marek's disease within the major histocompatibility (B) complex by refined typing of White Leghorn chickens. *Anim Genet,* 24, 283-287.

Ho CS, Lunney JK, Ando A, Rogal-Gaillard C, Lee JH, Shook LB, Smith DM (2009) Nomenclature for factors of the SLA system, update 2008. *Tissue Antigens,* 73, 307-315.

Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK, Nishibori M, Inoko H, Hanzawa K, Shiina T (2008) Contribution of mutation, recombination, and gene conversion to chicken *MHC-B* haplotype diversity. *J Immunol,* 181, 3393-3399.

Hosomichi K, Shiina T, Suzuki S, Tanaka M, Shimizu S, Iwamoto S, Hara H, Yoshida Y, Kulski JK, Inoko H, Hanzawa K (2006) The major histocompatibility complex (Mhc) class IIB region has greater genomic structural flexibility and diversity in the quail than the chicken. *BMC Genomics,* 7, 322.

Hsu IC, Metcalf R A, Sun T, Welsh JN, Wang NJ, Harris CC (1991) Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature*, 350, 427−428.

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, 355, 167-170.

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *PNAS USA*, 86, 958-962.

Hunkapiller T, Hood L (1989) Diversity of the immunoglobulin gene superfamily. *Adv Immunol*, 44, 1-63.

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.

Jarvi SI, Goto RM, Briles WE, Miller MM (1996) Characterization of Mhc genes in a multigenerational family of ring-necked pheasants. *Immunogenetics*, 43, 125-135.

Jarvi SI, Goto RM, Gee GF, Briles WE, Miller MM (1999) Identification, inheritance, and linkage of B-G-like and MHC class I genes in cranes. *J Hered*, 90, 152-159.

Jarvi SI, Tarr CL, Mcintosh CE, Atkinson CT, Fleischer RC (2004) Natural selection of the major histocompatibility complex (Mhc) in Hawaiian honeycreepers (Drepanidinae). *Mol Ecol*, 13, 2157-2168.

Jensen LF, Hansen MM, Mensberg KL, Loeschcke V (2008) Spatially and temporally fluctuating selection at non-MHC immune genes: evidence from TAP

polymorphism in populations of brown trout (*Salmo trutta*, L.). *Heredity*, 100, 79-91.

Jirjis FF, Noll SL, Halvorson DA, Nagaraja KV, Shaw DP (2002) Pathogenesis of avian pneumovirus infection in turkeys. *Vet Pathol*, 39, 300-310.

Johnson JA, Bellinger MR, Toepfer JE, Dun P (2004) Temporal changes in allele frequencies and low effective population size in greater prairie-chickens. *Mol Ecol*, 13, 2617-2630.

Kamara D, Gyenai KB, Geng T, Hammade H, Smith EJ (2007) Microsatellite marker-based genetic analysis of relatedness between commercial and heritage turkeys (*Meleagris gallopavo*). *Poult Sci*, 86, 46-49.

Kaufman J, Milne S, Göbel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, 401, 923-925.

Kaufman J, Salomonsen J, Skodt K (1989) B-G cDNA Cl1s have multiple small repeats and hybridize to both chicken MHC regions. *Immunogenetics*, 30, 440-451.

Kaufman J, Salomonsen J, Skjodt K, Thorpe D (1990) Size polymorphism of chicken major histocompatibility complex-encoded B-G molecules is due to length variation in the cytoplasmic heptad repeat region. *PNAS USA*, 87, 8277-8271.

Kaufman J, Skojdt K, Salomonsen J (1991) The B-G multigene family of the chicken major histocompatibility complex. *Crit Rev Immunol*, 11, 113-143.

Kawai T, Akira S (2011) Regulation of innate immune signaling pathways by the tripartite motif (TRIM) family proteins. *EMBO Mol Med*, 3, 513-27.

Klein PJ, Buckner R, Kelly J, Coulombe Jr RA (2000) Biochemical basis for the extreme sensitivity of turkeys to aflatoxin B(1). *Toxicol Appl Pharmacol*, 165, 45-52.

Koch C, Skjødt K, Toivanen A, Toivanen P (1983) New recombinants within the MHC (B-complex) of the chicken. *Tissue Antigens*, 21, 129-137.

Kumanovics A, Takada T, Lindahl KF (2006) Genomic organization of the mammalian MHC. *Annu Rev Immunol*, 21, 629-657.

Lahtinen SJ, Haskard CA, Ouwehand AC, Salminen SJ, Ahokas JT (2004) Binding of aflatoxin B1 to cell wall components of *Lactobacillus rhamnosus* strain GG. *Food Add & Contam: Part A*, 21, 158-164.

Lamont SJ (1998) The chicken major histocompatibility. *Rev. Sci. Tech. Off. Int. Epiz*, 17, 128-142.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 20. *Bioinformatics*, 23, 2947-2948.

Laun K, Coggill P, Palmer S, Sims S, Ning Z, Ragoussis J, Volpi E, Wilson N, Bek S, Ziegler A, Volz A (2006) The leukocyte receptor complex in chicken is characterized by massive expansion and diversification of immunoglobulin-like loci. *PLoS Genetics*, 2, e73.

Lee YK, El-Nezami H, Haskard CA, Gratz S, Puong KY, Salminen S, Mykkänen H (2003) Kinetics of adsorption and desorption of Aflatoxin B1 by viable and nonviable bacteria. *J Food Prot*. 66, 426-430.

Li J, Johnson LW, Livant EJ, Ewald SJ (1999) The MHC of a broiler chicken line: serology, B-G genotypes, and B-F/B-LB sequences. *Immunogenetics*, 49, 215-24.

Li Z, Nestor KE, Saif YM, Anderson JW, Patterson RA (2001) Effect of selection for increased body weight in turkeys on lymphoid organ weights, phagocytosis, and antibody responses to fowl cholera and Newcastle disease-inactivated vaccines. *Poult Sci*, 80, 689-694.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451-1452.

Liu W, Miller MM, Lamont SJ (2002) Association of MHC class I and class II gene polymorphisms with vaccine or challenge response to *Salmonella enteritidis* in young chicks. *Immunogenetics*, 54, 582-590.

Loiseau C, Richard M, Garnier S, Chastel O, Julliard R, Zoorob R, Sorci G (2009) Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Mol Ecol*, 18, 1331-1340.

Loiseau C, Zoorob R, Garnier S, Birard J, Federici P, Julliard R, Sorci G (2008) Antagonistic effects of a MHC class I allele on malaria-infected house sparrows. *Ecol Lett*, 11, 258-265.

Lombard Z, Dalton DL, Venter PA, Williams RC, and Bornman L (2006) Association of HLA-DR, -DQ, and vitamin D receptor alleles and haplotypes with tuberculosis in the Venda of South Africa. *Hum Immunol*, 67, 643-654.

Longenecker BM, Mosmann TR (1980) Restricted expression of an MHC alloantigen in cells of the erythroid series: a specific marker for erythroid differentiation. *J Supramolecular Structure,* 13, 395-400.

Maenaka K, Jones EY (1999) MHC superfamily structure and the immune system. *Curr Opin Struct Biol,* 9, 745-753.

Marusina K, Iyer M, Monaaco JJ (1997) Allelic variation in the mouse TAP-1 and TAP-2 transporter genes. *J Immunol,* 158, 5251-5256.

Martin AM, Kulski JK, Witt C, Pontarotti P, Christiansen FT (2002) Leukocyte Ig-like receptor complex (LRC) in mice and men. *Trends Immunol,* 23, 81-88.

Mikami T, Onuma M, Hayashi TTA, Narita M, Okada K, Fujimoto Y (1975) Pathogenic and serologic studies of Japanese quail infected with JM strain of Marek's disease herpes virus. *J Natl Cancer Inst,* 54, 607-614.

Miller HC, Allendorf F, Daugherty CH (2010) Genetic diversity and differentiation at MHC genes in island populations of tuatara (*Spenodon spp*.). *Mol. Ecol,* 19, 3894-3908.

Miller HC, Lambert DM (2004) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand Robins (Petroicidae). *Mol. Ecol,* 13, 3709-2721.

Miller MM, Abplanalp H, Goto R (1988) Genotyping chickens for the B-G subregion of the major histocompatibility complex using restriction fragment length polymorphisms. *Immunogenetics,* 28, 374-379.

Miller MM, Goto R, Abplanup H (1984) Analysis of the B-G antigens of the chicken MHC by two-dimensional gel electrophoresis. *Immunogenetics*, 20, 373-385.

Miller MM, Goto R, Bernot A, Zoorob R, Auffray C, Bumstead N and Briles WE (1994) Two MHC class I and two Mhc class II genes map to the chicken Rfp-Y system outside the B complex. *PNAS USA*, 91, 4397-4401.

Miller MM, Goto R and Clark S (1982) Structural characterization of developmentally expressed antigenic markers on chicken erythrocytes using monoclonal antibodies. *Dev Biol*, 94, 400-414.

Miller MM, Goto RM, Taylor RL Jr, Zoorob R, Auffray C, Briles RW, Briles WE, Bloom SE (1996) Assignment of Rfp-Y to the chicken major histocompatibility complex/NOR microchromosome and evidence for high-frequency recombination associated with the nucleolar organizer region. *PNAS USA*, 93, 3958–3962.

Miller MM, Goto R, Young S, Chirivella J, Hawke D, Miyada CG (1991) Immunoglobulin variable-region-like domains of diverse sequence within the major histocompatibility complex of the chicken. *PNAS USA*, 88, 4377-4381.

Miller MM, Goto R, Young S, Hardy J (1990) Antigens similar to major histocompatibility complex B-G are expressed in the intestinal epithelium in the chicken. *Immunogenetics*, 32, 45-50.

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.

Mock KE, Theimer TC, Wakeling BF, Rhodes JR OE, Greenberg DL, Keim P (2001) Verifying the origins of a reintroduced population of Gould's wild turkey. *J Wildl Manage*, 65, 871-879.

Morgan TJ, Lillehoi HS, Sanders BG, Kline K (1990) Characterization and developmental expression of the chicken B-G heterodimer. *Dev and Comp Immun*, 14, 425-437.

Moore CB, M John, IR James, FT Christiansen, CS Witt and SA Mallal (2002) Evidence of HIV-1 adaptation to HLA-Restricted immune responses at a population level. *Science*, 296, 1439-1443.

Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen MA, Crooijmans RP, Megens HJ, Zhang H, Okimoto R, Vereijken A, Jungerius A, Albers GA, Lawley CT, Delany ME, MacEachern S, Cheng HH (2008) Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *PNAS USA*, 105, 17312-17317.

Munguia A, Federspiel MJ (2008) Efficient subgroup C avian sarcoma and leukosis virus receptor activity requires the IgV domain of the Tvc receptor and proper display on the cell membrane. *J Virol*, 82, 11419-11428.

Nadachowska-Brzyska K, Zielinski P, Radwan J, Babik W (2011) Interspecific hybridization increases MHC class II diversity in two sister species of newts. *Mol Ecol*, 17, 4801-4811.

Nei, M (1986) Definition and estimation of fixation indices. *Evolution*, 40, 643-645.

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *PNAS*, 94, 7799-806.

Nei M, Rooney A (2005) Concerted and birth-and-death evolution of multigene families. *Ann Rev Genet*, 39, 121-152.

Nestor KE, Noble DO, Zhu NJ, Moritsu Y (1996) Direct and correlated responses tolong-term selection for increased body weight and egg production in turkeys. *Poult Sci*, 75, 1180-1191.

Nestor KE, Saif YM, Zhu J, Noble DO, Patterson RA (1996) The influence of major histocompatibility complex genotypes on resistance to *Pasteurella multocida* and Newcastle disease virus in turkeys. *Poult Sci*, 75, 29-33.

Niens M, Jarrett RF, Hepkema B, Nolte IM, van den Berg A (2007) HLA-A*02 is associated with a reduced risk and HLA-A*01 with an increased risk of developing EBV+ Hodgkin lymphoma. *Blood*, 10, 3310-3315.

Nishibori M, Nakaki S, Tsudzuki M, Yamamoto Y (2000) Utility of three restriction fragment length polymorphism probes for genotyping of the chicken major histocompatibility complex class IV region. *Poult Sci*, 79, 305-311.

Noda Y, Okada Y, Saito N, Setou M, Xu Y, Zhang Z, Hirokawa N (2001) *KIFC3*, a microtubule minus end-directed motor for the apical transport of annexin XIIIb-associated triton-insoluble membranes. *J Cell Biol*, 155, 77-88.

Ozato K, Shin DM, Chang TH, Morse HC 3rd (2008) TRIM family proteins and their emerging roles in innate immunity. *Nat Rev Immunol*, 8, 849-860.

Page RD (2002) Visualizing phylogenetic trees using TreeView. *Curr Protoc Bioinformatics*, Chapter 6, Unit 6.2.

Pandey I, Chauhan SS (2007) Studies on production performance and toxin residues in tissues and eggs of layer chickens fed on diets with various concentrations of aflatoxin AFB1. *Br Poult Sci*, 48, 713-723.

Petersen TN, Brunek S, von-Heijne G and Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, 8(10): 785-786.

Pham-Dinh D, Della Gaspera B, Kerlero de Rosbo N, Dautigny A (1995) Structure of the human myelin/oligodendrocyte glycoprotein gene and multiple alternative spliced isoforms. *Genomics*, 29, 345–352.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96, 7-21.

Pink JL, Droege W, Hala K, Miggiano VC, Ziegler A (1977) A three-locus model for the chicken major histocompatibility complex. *Immunogenetics*, 5, 203-216.

Qureshi MA, Heggen CL, Hussain I (2000) Avian macrophage: effector functions in health and disease. *Dev and Comp Immun*, 24, 103-119.

Quist CF, Bounous DI, Kilburn JV, Nettles VF, Wyatt RD (2000) The effect of dietary aflatoxin on wild turkey poults. *J Wild Dis*, 36, 436-444.

Ratajewski M, Walczak-Drzewiecka A, Salkowska A, Dastych J (2011) Aflatoxins upregulate CYP3A4 mRNA expression in a process that involves the PXR transcription factor. *Toxicol Lett*, 205, 146-153.

Rawal S (2010) Mechanisms of the extreme sensitivity of turkeys to aflatoxin B1. *All Graduate Theses and Dissertations*. Paper 568, http://digitalcommons.usu.edu/etd/568.

Rawal S, Coulombe Jr RA (2011) Metabolism of aflatoxin B1 in turkey liver microsomes: the relative roles of cytochromes P450 1A5 and 3A37. *Toxicol Appl Pharmacol*, 254, 349-354.

Reed KM (2009) Turkey. *In* Cockett NE, Kole C, eds, Genome Mapping and Genomics in Domestic Animals. Springer, New York.

Reed KM, Bauer MM, Monson MM, Benoit B, Chaves LD, O'Hare TH, Delany M. (2011) Defining the turkey MHC: identification of expressed class I- and class IIB-like genes independent of the *MHC-B*. *Immunogenetics*, 63, 753-771.

Reed KM, Chaves LD, Garbe JR, Da Y, Harry DE (2003) Allelic variation and genetic linkage of avian microsatellites in a new turkey population for genetic mapping. *Cytogenet Genome Res*, 102, 331-339.

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.

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

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132, 385-386.

Ruby T, Bed'Hom B, Wittzell H, Morin V, Oudin A, Zoorob R (2004) Characterization of a cluster of TRIM-B30.2 genes in the chicken MHC B locus. *Immunogenetics*, 57, 116-128.

Saarinen M, Ekman P, Ikeda M, Virtala M, Grönberg A, Yu DT, Arvilommi H, Granfors K (2002) Invasion of Salmonella into human intestinal epithelial cells is modulated by HLA-B27. *Rheumatology*, 41, 651-657.

Sacco RE, Rimler RB, Ye X, Nestor KE (2001) Identification of new major histocompatibility complex class II restriction fragment length polymorphisms in a closed experimental line of Beltsville Small White turkeys. *Poult Sci*, 80, 1109-1111.

Saif YM, Nestor KE (2002) Increased mortality in turkeys selected for increased body weight following vaccination with a live Newcastle disease virus vaccine. *Avian Dis*, 46, 505-508.

Saito N, Okada Y, Noda Y, Kinoshita Y, Kondo S, Hirokawa N (1997) *KIFC2* is a novel neuron-specific C-terminal type kinesin superfamily motor for dendritic transport of multivesicular body-like organelles. *Neuron*, 18, 425-438.

Salomonsen J, Dunon D, Skjodt K, Thorpe D, Vainio O, Kaufman J (1991) Chicken major histocompatibility complex-encoded B-G antigens are found on many cell types that are important for the immune system. *Proc Natl Acad Sci*, 88, 1359-1363.

Salomonsen J, Skjodt K, Crone M, Simonsen M (1987) The chicken erythrocyte-specific MHC antigen. Characterization and purification of the B-G antigen by monoclonal antibodies. *Immunogenetics*, 25, 373-382.

Sarson AJ, Parvizi P, Lepp D, Quinton M, Sharif S (2008) Transcriptional analysis of host responses to Marek's disease virus infection in genetically resistant and susceptible chickens. *Anim* Genet, 39, 232-240.

von Schantz TV, Wittzell H, Goransson G, Grahn M, Persson K (1996) MHC genotype and male ornamentation: genetic evidence for the Hamilton-Zuk model. *Proc. R. Soc. Lond*, 263, 265-271.

Schierman LW, Nordskog AW (1961) Relationship of Blood Type to histocompatibility in chickens. *Science*, 34, 1008-1009.

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W (2000) PipMaker - A web server for aligning two genomic DNA sequences. *Genome Res*, 10, 577-586.

Sena L, Schneider MP, Brenig BB, Honeycutt RL, Honeycutt DA, Womack JE, Skow LC (2011) Polymorphism and gene organization of water buffalo MHC-DQB genes show homology to the BoLA DQB region. *Anim Genet*, 42:378-385.

Sheehy MJ (1991) T cell defined HLA epitopes and T cell receptor polymorphism in insulin dependent diabetes mellitus. *Baillieres Clin Endocrinol Metab,* 5, 341-355.

Shiina T, Briles WE, Goto RM, Hosomichi K, Yanagiya K, Shimizu S, Inoko H, Miller MM (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, Beck S, Kulski J, Inoko H (2004) Comparative genomic analysis of two avian (Quail and Chicken) MHC regions. *J Immunol*, 172, 6751-6763.

Siddle HV, Kreiss A, Eldridge MD, Noonan E, Clarke CJ, Pyecroft S, Woods GM, Belov K (2007) Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *PNAS USA*, 104, 16221-16226.

Simonsen M, Hala K, Nicolaisen EM (1980) Linkage disequilibrium of MHC gene in the chicken I. The B-F and B-G loci. *Immunogenetics*, 10, 103-112.

Sironi L, Lazzari B, Ramelli P, Stella A, Mariani P (2008) Avian TAP genes: detection of nucleotide polymorphisms and comparative analysis across species. *Genet Mol Res*, 7, 1267-1281.

Smith IA, Knezevic BR, Ammann JU, Rhodes DA, Aw D, Palmer DB, Mather IH, Trowsdale J (2010) BTN1A1, the mammary gland butyrophilin, and BTN2A2 are both inhibitors of T cell activation. *J Immunol*, 184, 3514-3525.

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front. Zool*, 2, 16.

Spensley PC (1963) Aflatoxin, the active principle in turkey 'X' disease. *Endeavor*, 22, 75-79.

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc R Soc*, 277, 979-988.

Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzen A, Winer R, Knight J, Vogel JH, Aken B, Andersen O, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlesen BO, Mourn T, Skage M, Berg PR, Gjoen T, Kuhl H, Thorsen J, Malde K, Reinhardt R, du L, Johansen SD, Searle S, Lien S, Nilsen F, Jonassen I, Omholt SW, Stenseth NC, Jakobsen KS (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature*, 477, 207-210.

Strandh M, Lannefors M, Bonadonna F, Westerdahl H (2011) Characterization of MHC class I and II genes in a subantarctic seabird, the blue petrel, *Halobaena caerulea* (Procellariiformes). *Immunogenetics*, 63, 653-666.

Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet*, 73, 1162-1169.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrijm in haplotype inference and missing-data imputation. Am J Hum Gen, 76, 449-462.

Stephens M, Smith N, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*, 68, 978-989.

Svejgaard A, Lobitz Jr WC (1974) HL-A histocompatibility antigens and skin diseases. *Br J Dermatol*, 91, 237-241.

Tamura K, Dudley J, Nei M, Kumar S (2007) *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*, 24, 1596-1599.

Tanaka-Matsuda, Ando A, Roga-Gaillard C, Chardon P, Uenishi (2009) Differences in the number of loci of swine leukocyte antigen classical class I genes among haplotypes. Genomics, 93, 261-273.

Thorsby E (1997) Invited anniversary review: HLA associated diseases. *Hum Immunol*, 53, 1-11.

Traherne JA, Horton R, Roberts AN, Miretti MM, Hurles ME, Stewart CA, Ashurst JL, Atrazhev AM, Coggill P, Palmer S, Almeida J, Sims S, Wilming LG, Rogers J, de Jong PJ, Carrington M, Elliott JF, Sawcer S, Todd JA, Trowsdale J, Beck S (2006) Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLOS Genetics*, 2:e9.

Trowsdale J (2011) The MHC, disease and selection. *Immunology Letters*, 137, 1-8.

Uchil PD, Quinlan BD, Chan WT, Luna JM, Mothes W (2008) TRIM E3 Ligases Interfere with Early and Late Stages of the Retroviral Life Cycle. *PLoS Pathog*, 4, 316.

Uni Z, Sklan D, Haklay N, Yonash N, Heller D (1995) Response of three class-IV major histocompatibility complex haplotypes to *Eimeria acervulina* in meat-type chickens. *Br Poult Sci*, 36, 555-561.

Unoki M, Okutsu J, Nakamura Y (2003) Identification of a novel human gene, *ZFP91*, involved in acute myelogenous leukemia. *Int J Oncol*, 22, 1217-1223.

Viertlboeck BC, Crooijmans RP, Groenen MA TW Göbel (2004) Chicken IG-like receptor. B2, a member of a multigene family, is mainly expressed on B lymphocytes, recruits both Src homology 2 domain containing tyrosine phosphatase (SHP)-1 and SHP-2, and inhibits proliferation. *J Immunol*, 173, 7385-7393.

Viertlboeck BC, Schweinsberg S, Schmitt R, Herberg FW, Göbel TW (2009) The chicken leukocyte receptor complex encodes a family of different affinity FcY receptors. *J Immunol*, 182, 6985-6992.

Vilhelmová M, Miggiano VC, Pink JR, Hála K, Hartmanová J (1977) Analysis of the alloimmune properties of a recombinant genotype in the major histocompatibility complex of the chicken. *Eur J Immunol*, 7, 674-679.

Walker BA, Hunt LG, Sowa AK, Skodjt K, Gobel TW, Lehner PJ, Kaufman J (2011) The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *PNAS USA*, 108, 8396-8401.

Wan Q, Zeng C, Ni X, Pan H and Fang S (2011) Giant Panda genomic data provide insight into the birth-and-death process of mammalian major histocompatibility complex class II genes. *Plos One*, 4:e4147\

Weibking TS, Ledoux DR, Bermudez AJ, Rottinghaus GE (1994) Individual and combined effects of feeding Fusarium moniliforme culture material, containing known levels of fumonisin B1, and aflatoxin B1 in the young turkey poult. *Poult Sci*, 73, 1517-1525.

Westerdahl H, Waldenström J, Hansson B, Hasselquist D, von Schantz T, Bensch S (2005) Associations between malaria and MHC genes in a migratory songbird. *Proc Biol Sci*, 272, 1511-1518.

Westerdahl H, Wittzel H, von Schantz T (2000) Mhc diversity in two passerine birds: no evidence for a minimal essential Mhc. *Immunogenetic*, 52, 92-100.

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.

Wise S (1987) The Wild Turkey. Wisconsin Deptartment of Natural Resources, Bureau of Wildlife Management.

Yamashiro H, Yoghizaki S, Tadaki T, Egawa K, Seo N (2010) Stimulation of human butyrophilin 3 molecules results in negative regulation of cellular immunity. *J Leukoc Biol*, 88, 757-767.

Yang WX, Sperry AO (2003) C-terminal kinesin motor KIFC1 participates in acrosome biogenesis and vesicle transport. Biol Reprod, 69, 1719-1729.

Yarru LP, Settivari RS, Antoniou E, Ledoux DR, Rottinghaus GE (2009) Toxicological and gene expression analysis of the impact of aflatoxin B_1 on hepatic function of male broiler chicks. *Poul Sci*, 88, 360-371.

Yonash N, Heller ED, Hillel J, Cahaner A (2000) Detection of RFLP markers associated with antibody response in meat-type chickens: haplotype/genotype, single-band, and multiband analyses of RFLP in the major histocompatibility complex. *J Hered*, 91, 24-30.

Zagalska-Neubauer M, Babik W, Stuglik M, Gustaffson L, Cichon M and Radwan J (2010) 454 sequencing reveals the extreme complexity of the class II major histocompatibility complex in the collard flycatcher. *BMC Evol Biol*, 10, 395.

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, 95.