# An examination of the genetic diversity of the major histocompatibility complex in commercial and domestic chicken breeds

BY

**Corie N. Darrington**

Department of Animal Science

McGill University, Montreal

April 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree

of

Master of Science

In

Animal Science

**Table of Contents**

# Abstract

The major histocompatibility complex (MHC) is a group of highly polymorphic, closely linked genes found in the genome of all jawed vertebrates. Genes in this region are involved in immune function, and have been found to influence transplantation, autoimmune disease, and disease resistance in different species. In the chicken, the MHC is relatively small and compact, with the core region being referred to as the "minimal essential MHC". Haplotypes in the MHC-B region have been linked to resistance and susceptibility to many important diseases, which have major impacts on poultry health. Understanding the variation present in this region can be valuable for further investigation of the relationship between haplotypes and disease resistance, as well as broader understanding of the immune function of the chicken.

The first part of this study used a SNP panel to evaluate the MHC-B diversity in the Chantecler chicken. This breed is native to Quebec, Canada, and is distinguished by its strong resistance to extreme cold temperatures. Despite having previously gone through a near extinction event, resulting in low population sizes, considerable MHC diversity was observed within the 4 populations tested. A total of 8 SNP haplotypes were observed, with 2 of those being unique to the Chantecler breed. This study found that like other chicken breeds, Chantecler is able to generate and maintain diversity. The discovery of unique haplotypes contributes to growing knowledge of the variability of the MHC-B region in the chicken. It also brings attention to a unique chicken breed, showing the Chantecler is a viable candidate for further genetic studies, with regards to the MHC, as well as other genomic regions.

The second part of this study aimed to compare the diversity of the MHC region of the wild Red Junglefowl, the ancestral breed of the modern chicken, with domestic and commercial chicken breeds. Using whole genome sequence data and the Genome Analysis Toolkit (GATK) best practices pipeline, total diversity for chromosome 16, the MHC-B region, and the core region were evaluated. It was found that the Red Junglefowl had the highest level of diversity, and the white layer line had the lowest level. It was found that the core region, which contains the class I and II genes, was more diverse than the whole MHC-B region, with an average of 25 SNPs per kB as compared to 18 SNPs per kB. Within the coding region of these genes, over 50% of the SNPs showed moderate to high impact, indicating that much of the polymorphism present in the region has influence over the genes. Of those genes, the class I BF genes were the most diverse. Within these genes, over 50% of SNPs were located in exons 2 and 3, with 80% having moderate to high impact. These exons code for the peptide binding pocket of MHC molecules, which has been linked to disease resistance. The high levels of diversity observed will provide a basis for further exploration into the association between MHC variation and immune function.

## Résumé

Le complexe majeur d'histocompatibilité (CMH) est un groupe de gènes très polymorphes et étroitement liés, présents dans le génome de tous les vertébrés à mâchoires. Les gènes de cette région sont impliqués dans la fonction immunitaire et ont une influence sur la transplantation, les maladies auto-immunes et la résistance aux maladies chez différentes espèces. Chez les poulets, le CMH est relativement petit et compact, le noyau étant appelé « CMH minimal essentiel ». Les haplotypes de la région du CMH-B ont été associés à la résistance et à la sensibilité à de nombreuses maladies importantes, qui ont des répercussions majeures sur la santé des volailles. La compréhension de la variation présente dans cette région peut s'avérer précieuse pour approfondir les recherches sur la relation entre les haplotypes et la résistance aux maladies, ainsi que pour mieux comprendre la fonction immunitaire du poulet.

La première partie de cette étude a utilisé un panel de SNP pour évaluer la diversité du CMH-B chez les poulets Chantecler. Cette race est originaire du Québec, au Canada, et se distingue par sa forte résistance aux au froid extrême. Bien qu'elle ait été victime d'une quasi-extinction, ce qui a entraîné une diminution de la taille des populations, une diversité considérable du CMH a été observée dans les quatre populations testées. Au total, 8 haplotypes SNP ont été observés, dont 2 sont propres à la race Chantecler. Cette étude a montré que, comme d'autres races de poulets, la race Chantecler est capable de générer et de maintenir la diversité. La découverte d'haplotypes uniques contribue à accroître les connaissances sur la variabilité de CMH-B région chez le poulet. Elle attire également l'attention sur une race de poulet unique, montrant que le Chantecler est un bon candidat pour des études génétiques plus approfondies, en ce qui concerne le CMH, ainsi que d'autres régions génomiques.

La deuxième partie de cette étude visait à comparer la diversité de la région du CMH du Red Junglefowl sauvage, la race ancestrale du poulet moderne, avec des races de poulets domestiques et commerciales. En utilisant les données de la séquence du génome entier et le pipeline des meilleures pratiques GATK, la diversité totale du chromosome 16, la région CMH-B et le noyau ont été évalués. Il a été constaté que le Red Junglefowl présentait le niveau de diversité le plus élevé, et que la lignée blanche de ponte présentait le niveau le plus bas. Il a été constaté que la région centrale, qui contient les gènes de classe I et II, était plus diversifiée que l'ensemble de la région du CMH-B, avec une moyenne de 25 SNP par kB contre 18 SNP par kB. Dans la région codante de ces gènes, plus de 50 % des SNP avaient un impact modéré ou élevé, ce qui indique qu'une grande partie du polymorphisme présent dans la région a une influence sur les gènes. Parmi ces gènes, les gènes BF de classe I étaient les plus diversifiés. Dans ces gènes, plus de 50 % des SNP étaient situés dans les exons 2 et 3, 80 % d'entre eux ayant un impact modéré à élevé. Ces exons codent pour la poche de liaison peptidique des molécules du CMH, qui a été associée à la résistance aux maladies. Les niveaux élevés de diversité observés serviront de base à une exploration plus approfondie de l'association entre la variation du CMH et la fonction immunitaire.

# Acknowledgments

First, I would like to thank my supervisor Dr. Xin Zhao for his guidance and support throughout my Master's, and for allowing me this opportunity to learn and conduct this research. I would also like to thank the members of my committee, Dr. Roger Cue and Dr. Alexander Yitbarek, for their expertise and continuous support.

I would also like to thank the members of my lab for their support and encouragement. I am grateful for all the help they have provided and for allowing me to learn even more. A special thanks to Dr. Janet E. Fulton of Hy-Line International for her collaborative efforts in both executing research and data analysis. I would also like to thank Jean-Marc Larivière for assisting with my research project and for his encouragement as I have worked on my program.

Lastly, I wish to thank the members of my family and my friends for encouraging me to pursue this degree and supporting me from far away, as well as the friends I have made throughout my program. To my brother Scott, thank you for being one of my best friends and editing my manuscript. To my Mom, thank you will never be enough, I love you.

## Contribution of Authors

Corie Darrington is the M.Sc. candidate who designed and performed all experiments in consultation with her principal supervisor and committee. She collected and analyzed data and wrote the manuscript.

Dr. Xin Zhao is the thesis supervisor who guided the research conducted. He assisted in designing and performing the experiments, as well as the preparation of the manuscripts.

Haijiao Lin assisted in the sample collection. Jean-Marc Larivière assisted in sample collection and reviewed chapter 3 of the manuscript. Dr. Janet E. Fulton assisted in performing experiments, data analysis, and reviewed chapter 3 of the manuscript.

# List of Tables

# List of Figures

# List of Abbreviations

BL- Brown layer

BR- Broiler

BTN- Butyrophilin

GATK- Genome Analysis Toolkit

HLA- Human leukocyte antigen

H-2- Histocompatibility complex 2

LAO/LAAO- L-amino acid oxidase

MHC- Major histocompatibility complex

NK- Natural killer

NKC- Natural killer complex

NKT- Natural killer T cell

OR- Olfactory receptor

RFLP- Restriction fragment length polymorphism

RJF- Red Junglefowl

SILK- Silkie

SNP- Single nucleotide polymorphism

SSCP- Single strand conformation polymorphism

SRCR- Scavenger receptor cysteine rich

TAP- Transporter associated with antigen presentation

TCR- T cell receptor

TRIM- Tripartite motif

VEP- Variant effect predictor

WL- White layer

WLH- White Leghorn

VNTR- Variable number tandem repeat

ZNP- Zinc finger protein

# 1. Introduction

First discovered in mice during transplant experiments, the major histocompatibility complex (MHC) is a large locus of closely linked polymorphic genes found in vertebrate DNA. These genes code for cell surface proteins essential to the immune system. Among these genes are those that encode the proteins for the MHC class I and class II molecules, which are highly polymorphic and serve specific roles in immune responses, particularly antigen presentation. Within different species, MHC has been linked to transplant responses, autoimmune diseases, and disease resistance.

In chickens, the MHC is located on chromosome 16 in 2 separate regions, MHC-B and MHC-Y. Both regions contain class I and class II molecules, along with other genes that have been found to contribute to immune response. Compared to the human MHC, also referred to as the human leukocyte antigen (HLA), which is composed of over 200 genes spanning approximately 4000 kb (kilobases), the chicken MHC is relatively simple and compact, with the B region being composed of 46 genes spanning approximately 240 kb. Nearly all these genes have mammalian counterparts, and as such, the B region in particular is often referred to as the "minimal essential MHC" (Kaufman et al., 1999).

As modern chickens have evolved from the Red Junglefowl, genetic variation among the species has increased greatly, with the emergence of numerous breeds that vary in both genotypic and phenotypic traits. Variation can be caused by mutation and recombination but is also influenced by processes such as gene flow and random mating. Genetic variation allows natural selection to take place, therefore driving evolution. One way of defining this variation is

through haplotypes, which are sets of tightly linked DNA variations, or polymorphisms, located on a single chromosome. Haplotypes in the MHC-B are particularly interesting, as they have been found to exhibit strong resistance or susceptibility to a wide range of viral, bacterial, and parasitical diseases (Miller and Taylor, 2016).

Currently, a SNP panel, developed by Chazara et al. (2010) and later updated by Fulton et al. (2016a), is used for defining haplotypes in the MHC-B region. Since its development, the SNP panel has been used to examine MHC variability across a wide variety of breeds, including wild Red Junglefowl (Nguyen-Phuc et al., 2016), Finnish Landrace (Fulton et al., 2017), Silkie (Tarrant et al., 2020), Argentinian Campero INTA (Iglesias et al., 2019, 2021), Korean native breeds (Manjula et al., 2020), Sri Lankan breeds (Manjula et al., 2021), and heritage breeds in the United States and Canada (Fulton et al., 2016b). In all these studies, multiple MHC haplotypes were identified, including previously defined types, as well as novel haplotypes. As more unique and diverse breeds are examined, the more can be understood about the nature of the region.

Comparison across breeds can provide valuable insight into the nature of the MHC region. Different populations face different factors that may influence the evolution of genes found in this region, resulting in the emergence of different haplotypes with different characteristics. This thesis aims to define and compare the diversity of the MHC-B region among different chicken breeds in an effort to better understand the evolutionary history of individual breeds and the species as a whole.

# 2. Literature Review

## 2.0 Major Histocompatibility Complex

### 2.1 History

Evidence of the major histocompatibility complex (MHC) was first discovered in mice by Gorer (1936) while doing experiments with antigen II and tumor grafts. Mice lacking antigen II were quick to reject tumors, whereas tumors grew well in mice with it. Later, experiments on tumor resistance would connect antigen II to the H locus, resulting in the discovery of the *histocompatibility locus 2 (H-2),* which encoded strong or major histocompatibility antigens as compared to other H loci that encoded weaker antigens, leading to the H-2 locus becoming the major histocompatibility locus in mice (Gorer et al., 1948). Later, Jean Dausset (1958) found that there were antibodies present in the sera of patients who had received multiple blood transfusions, which reacted with the leukocytes of other individuals. The alloantigens that were found are credited as the first evidence of the human MHC, which is typically referred to as the human leukocyte antigen (HLA).

The first evidence of the major histocompatibility complex, specifically MHC-B, in poultry was an alloantigen system displayed on red blood cells, which was first reported by Briles et al., (1950). The link between MHC and this alloantigen system was later confirmed by Schierman and Nordskog (1961) through the use of skin grafting. They discovered that when the B blood group locus types of both the donor and recipient were matched, the graft was more readily accepted. Since this discovery, the distinction between 2 regions, MHC-B and MHC-Y,

has been made (Miller et al., 1994; Miller et al., 1996), alongside efforts to define and map the genes present (Shiina et al., 2007).

## 2.2 Structure

There are 2 major proteins that make up MHC molecules, class I and class II. These molecules are polymorphic, and both assist in antigen presentation. While there are many similarities between the 2 classes, differences in structure help distinguish the 2 groups from one another. Also included in the makeup of MHC are class III molecules. Unlike class I and class II, the structure and function of class III molecules is poorly defined. It is often referred to as the complement system, and the genes included can vary across species.

### *Class I*

MHC class I molecules are expressed on nearly all nucleated cells and platelets. Class I molecules' main function is to present endogenous antigens, signaling the body's cells to start an immune response. They are divided into classical and non-classical groups, with classical molecules being tasked with the presentation of antigens to CD8+ T cells, and non-classical exhibiting limited polymorphism and various other functions, which include the ability to interact with CD8+ T cells, NKT (natural killer T) cells, and NK (natural killer) cells (Nakamura et al., 2019).

Class I molecules are heterodimers with a polymorphic chain composed of three domains, α1, α2, and α3, as well as a β2 microglobulin protein. The α1 and α2 domains make up a peptide binding cleft that allows the molecule to bind an intracellularly digested peptide and thereby take on a more stable structure. Polymorphisms among class I molecules can determine

the peptide-binding specificity, along with the amino acids on the 8 stranded antiparallel β sheets that help make up the peptide binding cleft, resulting in variation in regard to peptide binding (Kulski and Inoko, 2006).

When non-self-cells, such as virus-infected or tumor cells, undergo proteolysis and form endogenous protein antigens, MHC class I molecules are tasked with presenting the peptides that are then derived by binding them to the peptide-binding cleft. These molecules will bind peptides that are approximately 8-11 amino acids in length (Lundegaard et al., 2008). Displaying the peptides from non-self-cells will activate CD8+ T cells through the T cell receptor (TCR), allowing them to attack and prevent further growth of the cells. NK cells are also able to be activated by cells which express low levels of MHC class I molecules, allowing them to directly kill target cells, as well as activate dendritic cells, which further induce immune response (Mocikat et al., 2003).

***Class II***

MHC class II molecules are expressed on antigen presenting cells, including B cells, dendritic cells, and macrophages. The main function of the class II molecules is to present exogenous antigens in order to signal an immune response. Like class I, class II molecules are also heterodimers; however, they are only comprised of 2 domains, an α chain (α1 and α2) and a β chain (β1 and β2). The α1 and β1 parts of the domains make up the peptide binding cleft, and similar to class I, the molecule will become more stable when it binds to a peptide.

When cells take in exogenous foreign target cells, such as bacteria or fungi, a peptide is formed by lysosomal degradation, which is then presented by the class II molecules. The

peptides that are bound are longer in length than those bound by class I, typically ranging from 15 to 30 amino acids (Srinivasan et al., 1993). The presented peptide is then recognized by CD4+ T cells through the TCR. Following activation, cytokines are released to stimulate Th1 and Th2 cells, which in turn activate CD8+ cytotoxic T cells, NK cells, and B cells (Nakamura et al., 2019). The activation of these cells can then lead to cell apoptosis or opsonization, mediated by a complement system or an antibody.

### 2.3 MHC in Humans

The HLA complex is located on the short arm of chromosome 6, covering approximately 4 megabases. The region containing class I genes is located closer to the telomeric end, while the region containing class II genes is closer to the centromeric end. These 2 groups are separated by class III genes. The HLA locus is one of the most polymorphic in the human genome, resulting in a large haplotype. However, HLA has been shown to demonstrate strong linkage disequilibrium (Hviid and Christiansen, 2005; Evseeva et al., 2010), resulting in the non-random association of alleles happening at multiple loci. This often results in specific combinations at each locus, which allows for certain gene polymorphisms to be predicted with a high probability.

The 283 identified loci of the HLA genomic region can be classified as either protein-coding genes, non-coding RNA, small nucleolar RNA, or pseudogenes (Shiina et al., 2016). Among these are the loci that correspond to MHC class I: HLA-A, -B, and -C. As well as the loci that correspond to MHC class II: HLA-DR, -DQ, -DP, -DM, and -DO. The highly polymorphic nature of the system results in multiple alleles at each individual locus. Diversity among these loci can have influential effects on the immune system and how it functions.

### 2.3.1 Function

The various roles the HLA plays in the immune system have been heavily researched since its discovery. During his research on blood transfusion patients, Dausset (1958) correctly predicted that HLA would be important for tissue transplants. Aside from the role it plays in organ transplantation, HLA can also be a determinant for susceptibility or resistance to infectious or autoimmune disease.

#### *Transplantations*

The first discovery of MHC came during skin grafting experiments on mice, and the matching of antigen types of transplant donors and recipients for optimum histocompatibility is something that is still being researched today. In an incompatible transplant, the HLA antigens will recognize the transplanted cells as foreign and set them up to be eliminated by T cells. This immune response can result in hyperacute rejection, acute rejection, or chronic rejection of the transplanted tissue or organ.

Due to the polymorphic nature of HLA, the probability of a perfect match is low, even among family members, with siblings only having a 25% chance of sharing the same haplotype (Ayala Garcia et al., 2012). However, a perfect match is not required for HLA testing. This is largely due to the fact that not all of the HLA antigens play a role in transplant rejection. Instead, it has been found that the HLA-A, HLA-B, and HLA-DR have the largest effect on transplantations, and so these are the loci that are generally most important to match. Mismatches in HLA-DR are most important in the first 6 months, HLA-B mismatch effects emerge in the first 2 years following transplantation, and mismatches of HLA-A can have deleterious effects in the long-term survival of a transplant (Mahdi 2013).

***Relationship to disease***

Due to its role in the immune system, the relationship of HLA to disease is of great interest. While the relationship of HLA to susceptibility and resistance of infectious diseases is not as strong as it is in chickens, it has been found to be highly associated with the development of certain autoimmune diseases. In particular, studies have focused on MHC class II molecules and the role they play in initiating autoimmune responses.

Autoimmune diseases occur when a person's immune system generates autoantibodies that attack the healthy self-cells as if they were foreign cells. Examples include type I diabetes, multiple sclerosis, rheumatoid arthritis, Graves' disease, and systemic lupus erythematosus. These diseases are often chronic, with age at onset and manifestations varying. Prevalence of autoimmune diseases have also been found to vary among certain ethnic groups, as well as gender, with women typically being disproportionally affected as compared to men (Cooper and Stroehla, 2003).

The likelihood of developing an autoimmune disorder is caused by a number of factors, including genetic, social, and climatic. It can also depend on a person's age, gender, and health background. However, genetic predisposition, which is often dependent on the presence of certain MHC class II alleles, significantly increases the risk of development. Studies have found that in all the previously mentioned diseases, as well as others, certain MHC alleles have demonstrated positive association with the development of disease in an individual. Variants of the *HLA-DRB1* gene have shown positive association with multiple sclerosis across different populations, and the *0401, *0404, *0405, and *0101 alleles have been found in nearly every patient with rheumatoid arthritis (Zakharova et al., 2019). *HLA-DQA1* and *HLA-DQB1* are often

inherited with *HLA-DRB1* due to strong linkage disequilibrium and are often involved in the risk of developing autoimmune diseases. It is also worth noting that *HLA-DRB1*07* is protective against many diseases, including Graves' disease, type I diabetes, rheumatoid arthritis, and multiple sclerosis (Simmonds and Gough, 2007).

### 2.4 MHC in Mice

The polymorphic mouse *H-2* region shares many similarities to its HLA counterpart. Because of this, the mouse has served as the premier model for MHC based studies and has been valuable to research into gene expression and peptide presentation (Kumanovics et al., 2003). The basic structure is similar to the HLA with the regions' order being class I, class III, class II, although the size of the regions can vary. This basic structure is actually observed among most mammals, including primates, such as the macaque (Shiina et al., 2016). There are also many framework genes that have remained conserved across species and are orthologous, but paralogy has also been observed within the class I region (Kumanovics and Lindahl, 2004).

### 2.5 MHC in Poultry

Compared to the mammalian major histocompatibility complex, chicken MHC is rather simple and compact but still contains the essential counterparts of genes found in the mammalian MHC. As such, the B locus in particular is often referred to as the "minimal essential MHC" (Kaufman et al., 1999). Because of this simplicity, stronger associations with resistance and susceptibility to infectious disease are found with chicken MHC. It is located on chromosome 16 and includes polymorphic class I and class II molecules, similar to those found in HLA, localized in 2 separate regions, MHC-B and MHC-Y.

Like its human counterpart, the chicken MHC is also highly polymorphic and has a large haplotype. Haplotypes in the MHC-B region specifically have been connected to disease resistance, and alleles at particular loci in the MHC-B region have shown major genetic effect in relation to infectious diseases (Miller and Taylor, 2016). Less is known about MHC-Y in comparison to MHC-B, but some specific genes in this region have been shown to be polymorphic and dynamically expressed, alloimmunogenic (Hunt et al., 2006) or weak alloantigens (Pharr et al., 1996). Distinguishing MHC-Y from MHC-B has allowed researchers to discover more about its properties, but it is still relatively unknown compared to widely researched MHC-B.

### 2.5.1 MHC-Y

On chromosome 16, the MHC-Y region is separated from the MHC-B region by an array of 41 nucleotide (PO41) tandem repeats. Originally it was thought to be part of the MHC-B region, but inconsistencies were observed among defined haplotypes (Briles et al., 1993). Following these observations, 2 MHC class I genes, 2 MHC class II genes, and a C-type lectin gene, all of which were originally assigned to the MHC-B region, were reassigned to MHC-Y and localized on chromosome 16 (Miller et al., 1994; Miller et al., 1996). Like MHC-B, different haplotypes are also observed in MHC-Y. The 2 regions have also been found to assort independently, due to “an intervening region that supports highly frequent recombination” (Miller and Taylor, 2016).

Assembly of the full sequence of the MHC-Y region is still progressing, but so far, published data gives evidence for 5 main gene families: class I, class II, C-type lectin-like, LENG9, and zinc finger proteins (ZNP) (Rogers et al., 2003; Goto et al., 2022). Mapping of this

region is difficult, as there is a high density of repetitive sequences, with numerous genes having multiple copies. Like MHC-B, the genes located in MHC-Y are thought to contribute to immune functions, but the specific functions of these genes are still being researched. The unusual makeup of the MHC-Y region distinguishes it from other classic MHC regions.

Also located on chromosome 16 is a section of olfactory receptor (OR) and scavenger receptor genes (SRCR). Analyses of SRCR gene restriction fragments have provided evidence that this region is polymorphic and that the SRCR/OR genes segregate in linkage with MHC-Y (Miller et al., 2013). Multiple types of OR genes are present in this region, including those of the *OR14J1* family. *OR14J1* genes are found in the extended region of human MHC, and the linkage of these genes with MHC-Y class I genes is similar to that in humans, suggesting a long-term association between OR genes and MHC. If there is a connected function between the linked SRCR genes and MHC-Y, it has yet to be fully discovered, but SRCR genes have been linked with various functions in the immune system (Miller et al., 2013).

Very few studies have been done regarding MHC-Y haplotypes and their variation and specific links to disease resistance. Previously, typing was done using restriction fragment patterns in southern hybridization, which has been reliable, but lengthy and complex. Zhang et al. have developed a simpler method, using short tandem repeat regions (2020). A simpler method of identifying haplotypes will be helpful in better researching the contributions of MHC-Y to disease resistance and immune response.

### 2.5.2 MHC-B

Shiina et al., (2007) mapped the full MHC-B region and classified the properties of the genes found therein, allowing for a clearer view of the region as a whole. The MHC-B region can be divided into 5 different sub-regions: BG region, *TRIM/Blec* region, class I and II region, complement gene region, and CD1 gene region (Figure 2.1). While not every gene has been defined in regard to its functions and influence, it is understood that the majority of genes contribute to immune function in some way (Miller and Taylor, 2016). The majority of genes located in the MHC-B region also have mammalian homologues and are thought to function similarly. When looking at disease resistance and susceptibility, the haplotype variation of the MHC-B region has been found to have strong influence.

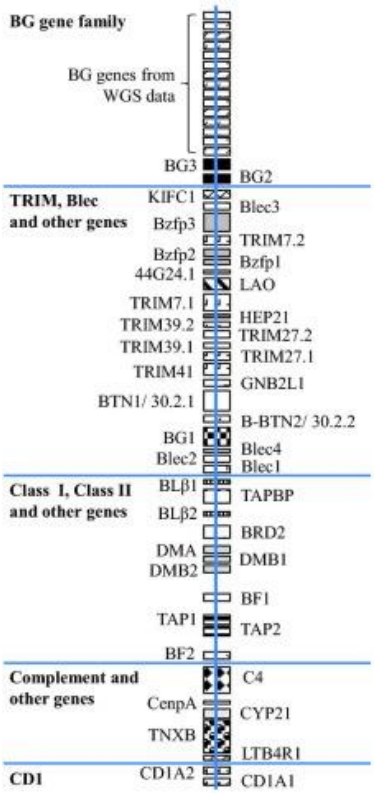


Figure 2.1: Gene map for the MHC-B region of *Gallus gallus.* Adapted from Miller and Taylor, 2016.

#### 2.5.2.1 BG Genes

Located upstream in the MHC-B region is a grouping of BG genes. These genes are distinguished by three structural elements: a single extracellular immunoglobulin-like domain of the V-region, a type 1 transmembrane domain, and a cytoplasmic tail composed of small domains (Miller et al., 1991). Indels and simple polymorphisms in the BG gene family can help distinguish haplotypes, with a significant variation of gene copy among haplotypes (Salomonsen et al., 2014). Despite the fact that BG genes have shown high degrees of specificity, only *BG1* (which is located further from the other BG genes) has a defined role in terms of immunity, whereas the roles other BG genes play in have yet to be defined. Individual BG genes have shown high degrees of specificity and have also been found to be expressed on many types of cells involved in immunity, including T and B cells (Salomonsen et al., 1991). The presence of BG genes among these cells indicates that some molecules play an important role in immunity, even if it is still unknown.

#### 2.5.2.2 Trim/Blec Genes

The *TRIM/Blec* region of the MHC-B is an area of 25 genes that are nearly all thought to contribute to immunity, specifically innate immunity. Some are more well understood and assessed to be "expressed genes," whereas others are thought as candidates to be expressed, or pseudogenes. The *TRIM* genes are thought to provide defense against viral infections, whereas the *Blec* genes are connected to natural killer cells, specifically working as receptors (Miller and Taylor, 2016). Also included in this region are ZNP, butyrophilins (BTN), and a handful of other miscellaneous genes.

### *Trim Genes*

Tripartite motif (TRIM) proteins play critical roles in both innate and adaptive immunity. In the chicken MHC-B region, there is a TRIM cluster, containing 7 TRIM genes. The TRIM genes in this region include TRIM7.2, TRIM7.1, TRIM39.2, TRIM27.2, TRIM39.1, TRIM27.1, and TRIM41. Many of the TRIM genes found in the chicken have been conserved and are homologous to the TRIM genes found in humans, chimpanzees, rats, mice, and other species. In humans, TRIM27 and TRIM39 are both located in the class I MHC region, whereas TRIM7 and TRIM41 are located on chromosome 5, alongside the gene GNB2L1, which is homologous to the guanine nucleotide binding protein found near the TRIM region in chickens (Ruby et al., 2005). This indicates that at some point during the evolutionary process, a mammalian specific breakpoint occurred, resulting in remodeling of the MHC region. In humans and mice, the polymorphism and function of TRIM genes have been studied, resulting in the discovery of associations with immune response and autoimmune conditions (Jia et al., 2021). However, very little is known about the polymorphisms and functions of these genes in the chicken (Kaufman, 2022). The defined immune connections in other species, as well as their location in the MHC-B region and close proximity to the class I and II genes, does lead to the hypothesis that the TRIM genes are involved in important immune processes.

### *Blec Genes*

BNK and Blec, also referred to as Blec2 and Blec1, are located next to each other, just before the class I and class II region in the MHC-B. These 2 genes are both considered C-type lectin genes, but they differ in structure. BNK is also polymorphic where Blec is monomorphic (Rogers and Kaufman, 2008). The presence of these 2 genes in the MHC-B region is interesting

as their mammalian orthologs are found in other genomic regions. BNK is an ortholog of NKRP1, and Blec is an ortholog of LLT1, both of which are located in the natural killer gene complex (NKC) in mammals (Rogers et al., 2005, Viertlboeck et al., 2008). This may indicate that this gene pair originated in the MHC region but translocated elsewhere during evolution (Rogers et al., 2005).

Immune functions for these 2 genes and their functions in the chicken have been studied more than other genes in the region. They are thought to contribute more to innate immunity, alongside the TRIM genes. BNK is considered to be a natural killer cell receptor, potentially working with ligands encoded in the class I genes (Miller and Taylor, 2016). As mentioned, it is polymorphic, exhibiting a high level of diversity, and is well expressed in lymphoid tissues (Rogers et al., 2005). The more monomorphic Blec has been found to be upregulated following cell activation and, like BNK, is well expressed in lymphoid tissues but is more conserved across haplotypes (Rogers and Kaufman, 2008).

### *Zinc Finger Proteins*

Three ZNP are found in the MHC-B region. They are B-locus zinc finger protein 1 (BZNP1), B-locus zinc finger protein 2 (BZNP2), and zinc finger protein 692 (ZNF692) (also referred to as BZNP3). Currently, BZNP2 is not currently annotated by NCBI and is classified as a pseudogene. The role of these genes in the chicken is not well documented. However, they do share identity with zinc finger proteins found in humans, so there is a likelihood that they function similarly.

Zinc finger proteins are a class of protein that show high specificity and are able to bind to specific sequences of DNA. This level of specificity means that these proteins have a wide

range of molecular function, influencing the regulation of numerous cellular processes throughout the body in the form of DNA regulation, RNA packaging, protein folding and assembly, and lipid binding (Razin et al., 2012). ZNP play an especially significant role in tissue homeostasis, acting as regulators for cell proliferation, differentiation, and apoptosis (Cassandri et al., 2017). ZNPs have also been connected to processes related to cancer onset and progression, with different genes playing roles in carcinogenesis, metastasis formation, and tumorigenesis. Aside from cancers, development of other neurodegenerative diseases in humans has also been linked to ZNPs. Mutations in or overexpression of certain genes have been identified as potential influences for the development of diseases such as Alzheimer's, Parkinson's, spinal muscular atrophy, and congenital heart diseases.

### *Butyrophilin Genes*

As members of the B7 gene family, BTN and butyrophilin-like (BTNL) genes play various roles in immune function. These genes are glycoproteins that consist of 2 immunoglobulin domains, IgC and IgV, and a transmembrane region, with most also containing the conserved B30.2 (PRY-SPRY) domain (Malinowska et al., 2017). BTN and BTNL genes are expressed on both lymphoid and non-lymphoid tissues and have been found to function as both stimulatory and inhibitory co-regulators of T cells (Abeler-Dorner et al., 2012). The BG genes found in chickens are closely related to BTN genes and are thought to share similar functions, however they differ in structure. In the MHC-B region, a cluster of BG genes make up the BG region, while a single gene, BG1, is found closer to the BF/BL region alongside a BTN gene (B-BTN1) and a second gene referred to as both B-BTN2 and TRIM10. Both of these genes were

predicted by bioinformatics and share similarities to rat TRIM39 and human TRIM27, respectively (Shiina et al., 2007).

As mentioned, most BTN genes contain the B30.2 domain, which is composed of the PRY and SPRY domains. This domain is also present in most TRIM genes and mediates protein-protein interactions, with many proteins involved in important signaling pathways (D'Cruz et al., 2012). This shared domain is part of the reason that the B-BTN2 gene is also referred to as TRIM10, but overall, it remains relatively unknown. How these genes function in relation to the chicken immune system is also relatively unknown. Currently, hypotheses regarding their nature can be made based on the structures available and comparisons to genes studied in the human and other mammalian species.

***Other Genes***

The following three genes, 44G24.1, RACK1, and IL4I1, are located in the MHC-B region, near the TRIM genes, but are not part of any gene family found in the region.

Very little is known about the 44G24.1 gene in chickens. It is also referred to as histone H2B-like and is located between the BG gene family and the TRIM family. No significant homology has been found between the 44G24.1 gene and other genes (Shiina et al., 2007), and as such, little can be deduced about its possible functions. Miller and Taylor (2016) list it as potentially having a DAXX (death domain associated protein) domain, which is associated with apoptosis, but more research needs to be done to confirm this information.

The guanine nucleotide binding gene, GNB2L1, is located between the TRIM region and the BTN genes. It is also referred to as RACK1, receptor for activated C kinase 1. It is conserved

across species and has been found to play roles in cell migration, apoptosis, proliferation, and innate immune responses (Adams et al., 2011). This gene has a WD40 motif protein that gives it a strong binding capacity, allowing it to interact with a wide range of proteins. In chickens, it has been found that polymorphisms in the gene can influence antibody response against Newcastle disease virus (Manzila et al., 2022). Another study also found that the RACK1 gene can interact with infectious bursal disease virus polymerases and proteins, influencing the virus-induced apoptosis and viral replication (Lin et al., 2015).

Interleukin 4 induced 1 gene (IL4I1) is located between 44G24.1 and the TRIM region. In chickens, it is also referred to l-amino acid oxidase (LAO or LAAO) based on the conserved domain. LAAOs are present in the secretory fluids of many animals, and the LAAO expressed on human macrophages and dendritic cells is the human IL4I1 gene (Kasai et al., 2015). LAAOs have been found to exhibit strong antimicrobial activities, including the regulation of cell proliferation, growth arrest, apoptosis, and necrosis, through the production of hydrogen peroxide (Kasai et al., 2015). One study found that IL4I1 was expressed and induced on chicken macrophages, granulocytes, and CD4 and γδ T-lymphocytes following *Salmonella enteritidis* infection (Elsheimer-Matulova et al., 2020). This study also found that, while IL4I1 may not be directly involved in bactericidal activity, it may be involved in controlling inflammatory and signaling response.

#### 2.5.2.3 Class I and II Genes

The class I and II region is perhaps the best understood region of the MHC-B. Since the general functions of class I and class II genes have been well defined, research has focused on the specific specializations of genes. There are 2 class I genes, *BF1* and *BF2*, that are

fundamentally similar in structure but differ in function. There are also 2 class II genes, BLB1 and BLB2. Other genes located near the class I and II genes include TAP, TAPBP (Tapasin), and DM genes. These genes serve to help with mediation of class I and class II molecules in regard to peptide antigen loading (Miller and Taylor, 2016). The BRD2 gene is also located within this region, however it is not classified as a classical MHC gene, nor does it have any strong connections to the class I and II genes beyond the basic immune functions it exhibits. Figure 2.2 shows the layout of this region.

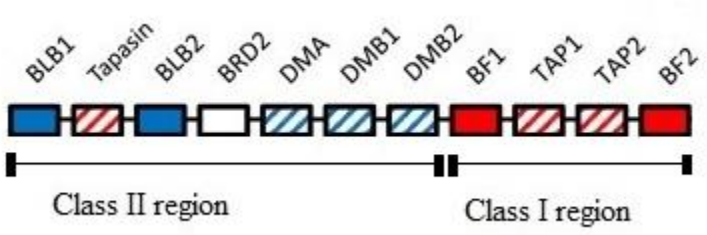


Figure 2.2: Layout of the MHC-B class I and class II gene regions. Classical MHC genes are indicated by solid boxes, with class I genes shown in red, and class II genes in blue. Class I and II adjacent genes are indicated by striped boxes, BRD2 is represented by a solid white box. Adapted from Kaufman, 2018.

### *Class I- BF1 and BF2*

In the MHC-B region of chickens, there are 2 class I genes, BF1 and BF2, and it has been found that the more polymorphic BF2 is dominantly expressed. Studies have also shown that the peptide motif of BF2, also referred to as BFIV or the major locus can vary across individuals (Butter et al., 2013). Variation in the peptide motif of chicken BF2 is achieved through a remodeling process that has not been documented in mammals (Chappell et al., 2015). Through this process, this class I molecule can either have a fastidious peptide motif or a promiscuous one, allowing it to associate with either a small or large repertoire of antigens (Kaufman, 2018).

While BF2 is found to be dominantly expressed, BF1 molecules are still thought to play important roles in immune function. BF1 has distinct polymorphism, resembling that of the human HLA-C class I locus (Ewald and Livant, 2004). In the alpha helix of the alpha 1 domain, BF1 alleles have a conserved locus-specific motif, whereas the region is more variable in BF2 alleles (Kim et al., 2018). In the human HLA-C alleles, the amino acid residues that align with this conserved domain are able to interact with the killer cell immunoglobulin-like receptors (KIR) expressed by natural killer (NK) cells (Boyington et al., 2000). Kim and colleagues (2018) have found that glycoproteins encoded by BF1 appear to interact with NK cells in ways that resemble HLA-C glycoproteins. Further research into the interaction of BF1 cells with NK cells is necessary, as different viruses have different methods of evading NK cells, which may influence the efficacy of BF1 cells.

### *TAP1 and TAP2*

Located between the 2 class I genes are 2 transporters associated with antigen presentation genes (TAP1 and TAP2). In other species, including humans and many placental mammals, the TAP genes are located further from the class I genes, in the class II region, and are found to be monomorphic and able to supply a wide variety of peptides that class I molecules can then present (Momburg et al., 1994). The close relation of the TAP genes to the class I genes in chickens has resulted in highly polymorphic TAP genes, which is a feature unique to the chicken. TAP molecules are part of the ATP (adenosine triphosphate)-binding cassette (ABC) family, composed of 2 membrane-spanning domains and 2 nucleotide binding domains (Walker et al., 2005). These molecules are part of the antigen-processing pathway, where they are responsible for loading processing and loading peptides onto class I molecules.

The closeness of the TAP genes with the class I genes may be responsible for the dominant expression of BF2 (Walker et al., 2011). Potential co-evolution between these polymorphic genes has resulted in unique combinations of alleles across haplotypes, leading to the high levels of specificity previously observed (Kaufman, 2015). Peptide transport in chickens is haplotype specific, with the translocation specificity matching or mirroring the motifs of the dominant BF2 molecules (Walker et al., 2011). Co-evolution of the TAP and class I genes has previously been observed in the rat (Joly et al., 1998), but not at the same level as has been observed in the chicken.

### *Class II- BLB1 and BLB2*

Within the class II region are 2 polymorphic classical class II beta chain BLB genes that flank the tapasin gene. The dominantly expressed "major" gene, BLB2, belongs to the B-LBII family, while the poorly expressed "minor" gene, BLB1, can belong to either the B-LBII or B-LBVI family, with variation being present across haplotypes (Jacob et al., 2000). BLB1 and BLB2 are closely related; however, BLB2 is more dominantly expressed, similar to BF2.

### *DM Genes*

Included in the class II region are three DM genes, DMA, DMB1, and DMB2. These genes are non-classical class II genes and similar to the TAP genes, assist in the formation of classical class II peptide complexes and antigen processing (Fling et al., 1994). The DM molecules protect the classical class II molecules from denaturation in acidic compartments, control the loading of antigenic peptides, and stabilize the peptide complex (Chazara et al., 2011). Like other genes in the MHC region, there is homology with the DM genes of chickens

and mammals (Shiina et al., 2007). They have the features of class II genes, with an MHC II α1 or β1 domain, an IgC1 α2 or β2 domain (with the α domains present in DMA and the β domains present in DMB1 and DMB2), and disulfide bridges composed of cysteines (Chazara et al., 2011). Like other MHC genes, polymorphism for the DM genes has been described but is lower than other genes, with specific domains being conserved, allowing for any class II molecules to bind to any DM molecules (Chazara et al., 2011).

Similar to the class I system, the peptide loading DM genes are closely related to the classical MHC class II genes. However, the relation is not as well understood. In general, BLB2 is more widely expressed than BLB1, similar to the class I BF genes. DMB2 is also expressed alongside BLB2 with high levels, while DMB1 is expressed alongside BLB1, mainly in the spleen and intestine (Parker and Kaufman, 2017). The expression of these 2 sets implies that there may be 2 class II systems functioning in the chicken MHC. The BLB2/DMB2 combination acts in a more general manner in many tissues, whereas the BLB1/DMB1 combination acts in a more specialized manner and is expressed mainly in the intestine (Parker and Kaufman, 2017).

Coevolution among the class II genes, similar to that observed in the class I region, has been proposed, particularly to describe why BLB2 is dominantly expressed over BLB1 (Parker and Kaufman, 2017). The expression of the BLB/DM pairs, as well as DMB2 being more similar to the mammalian DMB sequences and containing the appropriate class II elements (Chazara et al., 2011), supports this model. Previous models speculated that this region evolved as the result of genetic hitchhiking following selection of the closely linked genes in the class I region (Shiina et al., 2007), but the polymorphisms observed are more suggestive of selective pressure acting on the genes (Chazara et al., 2011).

### *TAPBP*

Although it is located in the class II region, tapasin, or TAPBP, interacts with the TAP and BF2 genes to form a peptide loading complex central to antigen presentation and processing. Like the TAP genes, tapasin is also highly polymorphic in a way that supports the dominant expression of a single class I gene (BF2). Tapasin creates a sort of bridge between TAP and class I molecules, allowing for peptide transportation to the class I molecules at optimized levels (Tan et al., 2002). Like mammalian TAP genes, mammalian tapasin is located further from the class I genes and is monomorphic, with no obvious functional differences among alleles. Tapasin is also included in the co-evolution theories surrounding the TAP and class I genes. While it is separated from the TAP and BF genes by a handful of class II genes, it is linked to the dominantly expressed class I gene, which has led to high levels of diversity (van Hateren et al., 2013). While recombination has been described in this region (Fulton et al., 2016), it occurs at lower levels compared to other regions, resulting in optimal combinations of TAP, class I, and tapasin genes staying together.

Overall, tapasin enhances the MHC I peptide loading process by localizing and stabilizing MHC I molecules at the peptide import site, stabilizing the TAP transporter, and increasing the rate, extent, and discrimination properties of peptides, ensuring the MHC I molecules become loaded with high affinity peptides (van Hateren et al., 2013). The extent to which these characteristics are observed varies from allele to allele. While it appears that all MHC class I molecules benefit from the presence of tapasin in some way, some benefit more greatly than others. For example, certain BF2 alleles are similar in sequence but use different peptide loading mechanisms. One allele is able to self-load peptides efficiently and gains some

enhancement from tapasin, but ultimately functions the same in the absence. On the other hand, another allele benefits greatly from the presence of tapasin and is more dependent on it for efficient peptide loading (van Hateren et al., 2013).

***BRD2***

Bromodomain-containing protein 2 (BRD2), also referred to as RING3, is also found in the class II gene region, although it is not classified as a class II gene in any way. Bromodomain proteins are able to alter chromatin status and control gene expression, contributing to epigenetic mechanisms that influence immune functions, among other bodily systems (Denis, 2010). Unlike the genes surrounding it, BRD2 does not have obvious roles in immunity. In chickens, BRD2 proteins have been found to interact with the matrix (M) protein of Newcastle Disease virus in a way that results in the downregulation of BRD2, allowing viral RNA synthesis and transcription (Duan et al., 2020). The functions of BRD2 related to gene transcription regulation, chromatin remodeling, cell proliferation, and apoptosis, as well as how it interacts with other proteins is being researched more, specifically in regard to Newcastle Disease virus (Zhou et al., 2021).

#### 2.5.2.4 Complement Genes

The complement subregion of MHC-B is also referred to as the class III region. It contains 5 genes, 4 of which are conserved. These are C4, CENP-A, CYP21, and TNXB. Of these, C4 has a strong connection to the immune system, specifically acting as a mediator for inflammation. The fifth gene, *LTB4R1,* also thought to play a role in inflammation, has only been located in the MHC of avian species (Shiina et al., 2007). Most of these genes have not been

studied in the chicken, and so the basis of understanding comes from previous studies on the mammalian counterparts.

Complement 4 (C4), which marks the start of the class III region, is the equivalent of the mammalian C4B gene. Another C4 gene maps to chromosome 1 in the chicken. This is the equivalent of the mammalian C4A gene and has been found to be better expressed than the MHC C4 (Kaufman, 2022). The C4 gene is an essential component of the complement system, a part of the innate immune system that enhances the abilities of antibodies and phagocytic cells. The complement system plays roles in microbial defense, elimination of foreign pathogens, and tissue homeostasis. Deficiency or dysfunction in the C4 gene has been linked to increased susceptibility to infection and predisposition for autoimmune diseases (Wang and Lui, 2021).

CYP21, also referred to as steroid-21 hydroxylase, belongs to the Cytochrome P450 superfamily of enzymes. Members of this family are involved in different physiological processes, including host defense mechanisms. In total, there are 45 CYP genes found in the chicken genome, but CYP21 is the only one found in the MHC region (Ren et al., 2019). The enzymes created by the CYP21 gene are involved in biosynthesis of aldosterone and cortisol. Mutations in the human homolog have been associated with congenital adrenal hyperplasia (Haider et al., 2013).

Tenascin XB (TNXB) is another highly conserved gene in the class III region. This is a large gene and is unusual as it overlaps with CYP21. It produces the glycoprotein TNX, which is found expressed in different tissues, but mainly in connective tissues. TNX plays roles in both

collagen binding and muscle metamorphosis, and deficiencies in the gene in humans has been linked to Ehlers-Dano syndrome, a hereditary connective tissue disorder (Miller, 2021).

#### 2.5.2.5 CD1 Genes

Two CD1 genes are located following the class III region. Referred to as either CD1A1 and CD1A2, or CD1-1 and CD1-2, these 2 genes are members of an ancient gene family, present in mammals and birds. This is another set of genes that is located outside of the MHC region in mammals, and like the BLK-Blec gene set, it is thought to have been present in the primordial MHC and translocated during evolution (Salomonsen et al., 2005). CD1 genes are also thought to be as ancient as MHC class I and II genes, and they also play important roles in the immune system, specifically against bacteria. These genes present lipid-containing antigens to T and NKT cells, with high levels of specificity and the capacity to distinguish between self and non-self (Salomonsen et al., 2005).

While the structures of both CD1 molecules in the chicken have been resolved, less is known about them in comparison to their mammalian counterparts. One interesting difference between the avian and mammalian CD1 genes is the presence of a small, more primordial antigen-binding pocket present in the avian CD1-2 gene (Zajonc et al., 2008). The binding groove in this unique pocket is very small, to the point that it is most likely only able to accommodate fatty acids or single alkyl chain lipids. This is not seen in the CD1-1 gene, which has a larger dual-pocket, dual-cleft binding pocket that is more similar to the mammalian counterparts (Dvir et al., 2010). CD1 genes also share similarities to class I and class II genes, with the structure being like that of class I, and the antigen loading systems being like that of

class II (Rogers and Kaufman, 2016). Polymorphism is also low in these genes, particularly in CD1-2, as compared to other genes in the MHC region (Miller et al., 2005).

### 2.5.3 MHC-B Haplotype

As chickens evolved, genetic variation was taking place, resulting in many breeds with many different characteristics. One way of defining genetic variation is through defining haplotypes. Haplotypes are a closely linked set of DNA variations, located within a specific region on a single chromosome. Due to their closeness, these variations are typically inherited together, which results in the conservation of a specific sequence that can then be passed down (National Human Genome Research Institute, 2019).

Although the MHC region in chickens is comparatively small, it is still highly polymorphic, which has resulted in high levels of variability. This variability is important to understanding disease resistance and the specific roles of genes in the MHC. Due to their genetic independence, MHC-B and MHC-Y haplotypes are not linked, playing different roles in the immune response. Much more is known about the MHC-B haplotypes, and they are what have been connected to specific disease resistance and susceptibility in chickens. As the systems for haplotyping become more refined and reliable, these connections will be better understood and can be used in regard to breeding and genomic selection for the benefit of populations.

In 1982, an international exchange of alloantisera and blood samples of chickens resulted in the first standardization of MHC-B haplotypes, with 27 types being serologically defined (Briles et al., 1982). This was accomplished by testing the alloantisera in hemagglutination assays in order to identify haplotypes. However, this method was limited due to the fact that the

alloantisera is rarely haplotype specific and cross-reactivity can occur, making identification of haplotypes much more difficult. The haplotype definition was also only done on defined lines and breeds, leaving the haplotypes of non-domesticated chicken breeds, as well as many broiler breeds, poorly defined (Fulton et al., 2016a).

To better study the differences in haplotypes, methods including restriction fragment length polymorphisms (RFLP), single strand conformation polymorphism (SSCP), and complex variable number tandem repeat (VNTR) have been developed to further confirm and identify MHC-B haplotypes (Miller et al., 1988; Goto et al., 2002; Fulton et al., 2006). VNTR using the LEI0258 allele has been particularly useful in simplifying and enhancing the ability to assign MHC-B haplotypes (Fulton et al., 2006). However, it has proven to not be fully sufficient on its own for typing, as serologically unique haplotypes have been found to have the same LEI0258 allele size.

In 2010, Chazara et al. (2010) developed the first SNP panel for MHC-B typing. It was later updated by Fulton et al. (2016a). This panel uses 90 SNPs across the 46 genes of the MHC-B and initially resulted in the identification of 78 haplotypes (including previously defined and novel) separated into 22 families, based on SNP similarities. Haplotypes that were identified using this panel are classified as "BSNP" haplotypes. For those initially identified haplotypes, this is followed by a letter to indicate the family the haplotype belongs to, based on neighbor-joining topology, and a two-digit number to distinguish from other members of the same group. If the size of the LEI0258 allele is known, it is included in parentheses. Also included in the parentheses is the serological B type, if known. One example of this is the BSNP-A02 (357;B75), where 357 shows the size of the LEI0258 allele, and B75 corresponds to the

previously defined B haplotype. For novel haplotypes defined after this initial group, the BSNP moniker remains, but is then followed by an abbreviation to indicate the breed where the haplotype was found, and a number, along with the LEI0258 size, if known (i.e.: BNSP-Kr01(295)) (Manjula et al., 2020).

This panel was also able to identify recombinants, which better helps with the understanding of diversity and the nature of recombination in the region. The patterns observed using the panel were consistent with previously defined haplotypes and helped clarify relationships between haplotypes. The development of this panel has allowed for quicker and more reliable haplotyping of the region. The MHC-B diversity of numerous breeds have been defined through the use of this panel, which has allowed for a better understanding of the region, as well as the genetic history of the chicken.

SNP analysis is an effective method for measuring genetic variation. SNPs are a common genetic variation present in a large portion of a population, and have been linked to different traits, including those which can influence an individual's relationship with different diseases. Aside from defining the haplotype of the MHC-B, SNP analysis can also be useful in determining the behavior of different alleles in the region. Understanding the associations between MHC-B and disease resistance will help improve the welfare of poultry, as well as uncover more about the functions of the immune system. SNPs are one way to do this, and the ease and reliability of SNP analysis makes it a useful tool for genetic studies.

#### 2.5.3.1 Haplotypes and Disease Resistance

Disease is a major concern to both large commercial poultry flocks and smaller native breed flocks. Depending on location, the prevalence and concern of disease can vary, as well as

present strains. This may play a role in the distribution of MHC-B haplotypes across breeds. Selection may have taken place in breeds which have been exposed to certain pathogens, resulting in the presence or absence of haplotypes. The standardization of MHC-B haplotypes has led to the associations between MHC-B haplotypes with disease to be studied extensively. The development of the SNP panel and subsequent discovery of novel haplotypes will continue to supply the research into this area. Understanding the behavior of haplotypes would be beneficial for breeding for disease resistance and improvement of flock health and immunity.

Table 2.1 shows the standardized MHC-B haplotypes associated with resistance and susceptibility in 6 diseases and pathogens of concern to the poultry industry. These include 3 viral diseases, Marek's disease, Infectious Bursal disease, and infectious bronchitis, along with 3 types of bacteria, *Staphylococcus, Salmonella,* and *E. coli.* While some haplotypes, like B2, show resistance for multiple diseases, there are others that show resistance towards one but susceptibility to another, such as B12 and B21. Strength of resistance can also vary, particularly in regard to heterozygosity. In the case of infectious bursal disease, a B15B2 heterozygous haplotype showed little bursal damage, whereas the B15B12 haplotype showed moderate bursal damage, indicating that one copy of B2 was as effective as two copies, but one copy of B12 was only half as effective as two copies (Butter et al., 2013). Similarly, in those chickens infected with *E.coli*, the heterozygous B13B21 showed more resistance than the susceptibility homozygous B21 haplotype, but less than resistant homozygous B13 haplotype (Cavero et al., 2009).

| | Marek's Disease | Infectious Bursal Disease | Infectious Bronchitis | *Staphylococcus* | *Salmonella* | *E.coli* |
|---|---|---|---|---|---|---|
| Resistant Haplotype | B2, B21 | B2, B12 | B2, B5 | - | BC | B13 |
| Susceptible Haplotype | B19 | B15 | B12, B19 | B4, B12 | B15, B18 | B21 |
| Reference | Kaufman and Venugopal, 1998 | Butter et al., 2013 | Da Silva et al., 2017 | Joiner et al., 2005 | Cotter et al., 1998 | Cavero et al., 2009 |

Table 2.1: Association of resistant and susceptible standardized MHC-B haplotypes with selected bacterial and viral diseases.

#### 2.5.3.2 Haplotypes and Class I and II Genes

Haplotype variation within the class I and II MHC genes has been a topic of interest, as this variation has been found to influence disease resistance and susceptibility. The BF2 gene in particular is thought to contribute greatly to the behavior of specific haplotypes. Zhang and coworkers (2012) compared the structures of the BF2 alleles for the B21 haplotype, which is known to confer resistance to a range of pathogens, to the B4 haplotype, which is known to be more susceptible to pathogens. The peptides from the B21 allele were all found to be 10mers and 11mers, allowing them to act more promiscuously as compared to the other peptides, which were almost exclusively octamers (Zhang et al., 2012).

While there may be an obvious advantage of the BF2 allele being more promiscuous, as observed in the B21 haplotype, the B4 haplotype is not rare among chicken populations, indicating that some selective pressure has been present, resulting in a high gene frequency

(Zhang et al., 2012). The fastidious nature of the B4 alleles may be responsible for its susceptibility to certain pathogens, but it may also confer high resistance to important pathogens. Heterozygote advantage in the MHC-B region may also play a role in the survival of the B4 haplotype. Heterozygous BF2 alleles may be able to present more and different peptides, leading to favorable combinations of haplotypes that remain present in populations.

Expression of BF1 and BF2 genes has also been found to differ across haplotypes, although BF2 expression is consistently higher (O'Neill et al., 2009). The linked nature of genes in the MHC-B region, and the high levels of polymorphism, results in variation across haplotypes that may not be easily quantified. The difference in expression levels of the BF1 and BF2 genes indicates that they have separate evolutionary histories. Where the BF2 gene was more impacted by selective pressures related to pathogen resistance, the BF1 gene shows signs of not being under selection and instead acquiring allele changes as a result of neutral changes and genetic drift (Shaw et al., 2007). This pattern is also observed with the major (BLB2) and minor (BLB1) class II genes (Shaw et al., 2007). It is also worth noting, in those haplotypes where the BF2 molecule has been found to be more promiscuous, such as B2, BF1 expression is higher, appearing more stable than other alleles and fastidious in nature (Kaufman, 2015).

Additionally, variation in the TAP genes can affect the expression of class I glycoproteins, so there may be differences in BF1 function across MHC-B haplotypes, according to the TAP variant present (Kim et al., 2018). Furthermore, at least 2 haplotypes (B14, B15) lack an identifiable BF1 gene, as the result of insertions and rearrangements (Shaw et al., 2007). The presence of a BF1 pseudogene in some haplotypes may indicate that the functions of the BF1 gene are not essential.

Comparison of human and chicken MHC class I molecules further confirms the notion that these genes have a large responsibility for associations with disease resistance. In humans, there are 6 class I molecules, allowing for a higher probability of binding to a peptide that will be more or less resistant against an invading pathogen, thereby creating an overall broader protection (Kaufman, 2014). As a result, there are very few instances of strong associations of the human MHC class I with resistance to infectious pathogens. Instead, stronger associations to autoimmune diseases are connected to human MHC class I.

While there isn't as much information about the haplotype variation observed in the class II genes, an interesting difference between the BF and BLB genes is that for every non-recombinant MHC haplotype, there are unique BLB gene variants (Hosomichi et al., 2008). High allelic variation has also been documented for the DM genes, also located in the class II region (Chazara et al., 2011), and it was found that there were as many unique gene haplotypes as MHC-B haplotypes for the class II genes. However, with the development of the more sensitive SNP panel used for MHC-B haplotyping (Fulton et al., 2016) and definition of hundreds of new haplotypes, this may not fully be the case. Regardless, the variation in the class II alleles is still much higher than that observed in the class I alleles.

### 2.5.4 Whole Genome Sequence Data

The first draft of the chicken genome was released by the International Chicken Genome Sequencing Consortium in 2004. Despite its wide use in vertebrate and avian genetic studies, the genome has taken a long time to complete, with different iterations missing various microchromosomes. The Vertebrate Genome Project has released the most current version (Galgal7), which was completed using a trio-binning approach (Rhie et al., 2021), but it still is

not considered to be a full picture of the chicken genome. The lack of full coverage sequence data previously made genomic studies difficult. Chromosome 16, which is a microchromosome and the chromosome in which the MHC was found, took a longer time to sequence, which resulted in methods including trisomy mapping being used to assign genes to the chromosome (Miller et al., 2013).

As coverage has improved for chromosome 16 and the genome as a whole, more research is able to be performed. While the MHC-Y region has yet to be fully sequenced, the MHC-B region has been well mapped (Shiina et al., 2007), and the more complete sequence is highly beneficial for studies regarding the region. Sequence data can allow for deeper analysis of genes and the variation present within. It can also be valuable for examining linkage between genes, the presence of selection signatures, and the level of heterozygosity within the genome. Comparison between different sequences can also help uncover the genetic history and ancestral background of specific regions.

### 2.6 MHC in other avian species

Generally, avian MHC tends to be overall smaller and more compact, containing fewer genes overall as compared to human and other mammalian MHC. Where there is some basic structure observed across most mammalian MHC regions, the same cannot be said for avian species. This is in part due to lack of full sequence mapping for the region across species, but there are also high levels of variation observed in those species where more data is available. Of the avian species that have been studied, the chicken has the most well-defined and well-studied MHC region, but research has been done on other galliform and passerine species as well.

### *Galliformes and Anseriformes*

The MHC region of galliform species shows high levels of similarity. The expression of a single dominant class I MHC gene observed in the chicken has been observed in 2 different Galliformes, the turkey (Monson et al., 2013) and the Japanese quail (Ye et al., 1999), as well as the mallard (Mesa et al., 2004), which is from the family Anseriformes, closely related to Galliformes. Also like the chicken, the TAP genes in the turkey, Japanese quail, and mallard are all located next to the dominantly expressed MHC class I gene. This further substantiates the coevolution theory that has resulted in dominant expression. One could speculate that, like the chicken, these other avian species with a dominantly expressed class I gene may also exhibit haplotype correlations with disease resistance and susceptibility.

Like the chicken, the turkey MHC is also divided into MHC-B and MHC-Y regions (Chaves et al., 2007). The MHC-B region is similar to the chicken in both size and gene organization, showing a high level of conservation between the 2 species. One of the most notable differences between the 2 species is that turkeys possess three BG and class II B loci, whereas chickens possess one BG and 2 class II B loci (Chaves et al, 2009). A notable inversion at the TAPBP and TAP1-TAP2 gene block also distinguishes the turkey MHC from the chicken. The similarities between genes and high level of conservation between the turkey and chicken further suggest that the turkey MHC may also exhibit close associations to disease resistance and susceptibility as has been observed with the chicken.

The Japanese quail MHC was sequenced by Ye et al. (1999), who found that, like chickens, the quail MHC is a densely packed genomic region. The same basic gene family

members are found in the quail, but the quail has an expanded number of duplicated genes (Shiina et al., 2004). These include class I, class II, NK, lectin, and BG genes. Regardless of these duplications, the gene sequences of the class I and class II genes still show a close relation between the quail and the chicken.

Comparison of ducks and chickens may provide more information on the primordial MHC and how it has evolved as lines have diverged. Overall, the organization of the duck MHC differs from that of the chicken, indicating that a translocation event possibly occurred, resulting in different structures (Moon et al., 2005). The duck MHC also contains 5 MHC class I alleles, but like the chicken, only one is dominantly expressed (Mesa et al., 2004). How these differences influence the overall functions of the MHC in the duck, and how they differ from the chicken, has yet to be determined.

### *Passerines*

Compared to the MHC of Galliform species, the MHC of passerine species, which includes sparrows, finches, and other perching birds, has been found to exhibit extreme levels of diversity, particularly in the class I gene family. O'Conner et al. (2016) conducted a study on twelve passerine species in which phylogenetic analysis of the MHC class I genes was performed. In this study, they found an extremely broad range of MHC class I gene numbers across the species, giving a glimpse into the wide breadth of diversity present within the order. It was found that species had between 7 and 37 MHC class I alleles, with individuals possessing anywhere from 4 to 19 MHC class I gene copies. Compared to the aforementioned galliform species, this is an extreme difference in MHC composition and may be part of the reason passerines are able to adapt to a wide variety of habitats. Multiple gene copies that are highly

polymorphic may broaden the range and capability of the immune system, allowing the birds to cope with a large diversity of pathogens.

Drews and Westerdahl (2019) found that, unlike Galliformes and other non-passerine species, the passerine species siskin (*Spinus spinus*) did not have a single dominantly expressed MHC class I allele. First, like the previously mentioned study, they found that gene copy number differed among the individuals tested, ranging from 6 to 16 MHC class I alleles, which includes both classical and non-classical alleles. Expression of these alleles was then measured, and it was found that the majority of the alleles were expressed at some level, and that individuals had 3 to 5 highly expressed alleles. This high expression suggests that these genes play an important role in adaptive immunity and contribute to the broad immune capabilities of these birds. Further research is needed to determine if this pattern of expression is true for passerines or not, as well as the implications and reasoning behind this level of expression.

The location of the TAP genes in the mentioned passerine species has not yet been determined. Evidence from the zebra finch genome indicate that they are likely not located near each other but separated by other genes (Ekblom et al., 2011). If this is the case for other passerines, it may explain why there are high numbers of MHC class I gene copies and why individuals express multiple copies at high levels. Since the class I genes appear to function more like humans and other mammals, as opposed to other birds like the chicken, it would make sense that the arrangement of the genes is similar, where the TAP and class I genes are separated and have little influence over the evolution of the other.

## 2.7 Objectives

The MHC is a region of great interest for the contributions the genes located within to the immune system. In chickens, the link between MHC-B haplotype and disease resistance is of major interest. Examination of the genes in this region, and how they vary among different breeds will be valuable in understanding how variation affects immune function, and how this variation can be used to improve the immune welfare of poultry. Unlike other domestic animals, the ancestral breed of the modern chicken, the Red Junglefowl, is still alive and allows for easy genetic analysis and comparisons. Since domestication, numerous chicken breeds have been developed worldwide, including those used for commercial production, heritage breeds, and indigenous breeds. Each breed of chicken has a unique history which has impacted its genetic identity and sets it apart from other breeds.

Comparison between chicken breeds can be useful for better understanding of the species and its unique genomic regions, including the MHC. Commercial chicken breeds have been subjected to intense artificial selection in order to create high producing lines, which has resulted in some negative side effects. Heritage chicken breeds are also typically maintained in closed breeding populations, but do not face the same issues as commercial lines. Indigenous breeds are less subjected to any sort of breeding intervention, and often have unique features as the result of specific geographical factors that have influenced selection. Being able to compare the MHC region of different breeds will provide insight into the genomic history of the region, as well as provide a basis for future studies into immune health.

The overarching objective of this thesis was to examine the diversity in the MHC-B region of the chicken. The first objective was to define the MHC-B diversity present in the

Chantecler chicken breed. To accomplish this, a high-density SNP panel (Fulton et al., 2016a) was used to haplotype the region, and the results were compared to other previously analyzed chicken breeds. The second objective was to compare the MHC-B diversity of the wild Red Junglefowl with domestic and commercial chicken breeds. This was accomplished through the use of bioinformatics tools which allowed for SNP analysis of whole genome sequence data.

## 3. MHC-B diversity in the Chantecler chicken

### 3.1 Abstract

The major histocompatibility complex (MHC) is a highly polymorphic cluster of genes that contributes to immune response. Located on chromosome 16, the chicken MHC has been found to have great influence over disease resistance and susceptibility. Through the use of a high-density SNP panel that encompasses the MHC-B region, haplotypes can be easily identified, allowing for further genetic analyses. This study aims to use an MHC-B SNP panel to evaluate the MHC-B variability in the Chantecler breed. This breed is native to Quebec, Canada, and is a dual-purpose breed known for its strong resistance to extreme cold temperatures. The Chantecler breed faced a near extinction event in the 1970s, which most likely resulted in a genetic bottleneck and loss of diversity. Despite this, SNP haplotype diversity was observed among 4 Chantecler populations. A total of 8 haplotypes were observed. Of these haplotypes, 6 were previously defined in other breeds, and the other 2 were unique to the Chantecler. Within the populations, the number of haplotypes ranged from 5 to 7, with 3 haplotypes, including the novel BSNP-Chant01, being present in all the groups. This study shows that there is reasonable diversity in the MHC-B region of the Chantecler breed and further contributes to understanding the variability of this regions in chickens.

### 3.2 Introduction

The Chantecler is a breed of chicken that originated in the province of Québec, Canada. In the early 1900s, Brother Wilfrid Chatelain, a monk from the Cistercian Abbey in Oka, QC, wanted to create a dual-purpose chicken that could be used for both meat and egg production,

and was hardy enough to withstand the intense cold temperatures of Canada (The Livestock Conservancy, 2022). Numerous heritage breeds were crossed to create the Chantecler. These included Dark Cornish, White Leghorn, Rhode Island Red, White Wyandotte, and White Plymouth Rock. The resulting Chantecler breed is characterized by a small to nonexistent wattle and a small cushion comb, which help minimize frostbite damage, as well as tight but fluffy white plumage. It should be noted that a second Canadian breed, the Partridge Chantecler, shares similar characteristics but was created to be more free ranging and has a different genetic background than the Chantecler.

Chantecler chicken populations are few and far between. The breed is rare to the point of nearly facing extinction in the 1970s but was brought back thanks to the efforts of small-scale farms. However, the breed is still categorized as "critical" by the American Livestock Breeds Conservancy (The Livestock Conservancy, 2022). The goal of this research is to characterize the diversity of the MHC-B region in Chantecler populations. This will provide genetic information that will be useful for genetic conservation, as well as improving welfare and production value of the breed. Additionally, information on the genetic variation present in the Chantecler will be useful for comparisons of variation among different breeds, providing further insight into an important region in the chicken genome.

Since its development, a SNP panel (Fulton et al., 2016a) has been used to examine MHC variability across a wide variety of chicken breeds, including wild Red Junglefowl (Nguyen-Phuc et al., 2016), Finnish Landrace (Fulton et al., 2017), Silkie (Tarrant et al., 2020), Argentinian Campero INTA (Iglesias et al., 2019, 2021), Korean native breeds (Manjula et al., 2020), Sri Lankan breeds (Manjula et al., 2021), heritage breeds, and commercially utilized elite

layer lines in the United States and Canada (Fulton et al., 2016b). In all of these studies, multiple MHC haplotypes were identified, including some found in multiple breeds, as well as breed-specific novel haplotypes. The use of the MHC-B SNP panel allows greater identification of the MHC diversity across breeds and can identify new recombinant haplotypes that were not easily detected by previous.

The purpose of this study was to describe the diversity present in the MHC-B region of the Chantecler chicken and compare it to the diversity of other breeds. The previous near extinction event the breed experienced most likely resulted in a genetic bottleneck and loss of diversity. Because of this, it was expected that while there would be some small measure of diversity in the region, it would not be as substantial as that which has been observed in other unique breeds.

## 3.3 Materials and Methods

### 3.3.1 Sample Collection and Genomic DNA

For this study, four distinct populations of Chantecler chickens were utilized and raised at the Institut de Technologie Agroalimentaire de Québec La Pocatière. Parental birds were gathered from 4 independent sources, bred on the campus, and returned to their respective farms once the fertilized eggs were laid. A total of 100 birds were hatched, with the number of samples ranging from 16-40 per population. Animals were raised in accordance with regulatory agency guidelines published by the Canadian Council on Animal Care (CCAC), and all procedures followed agency guidelines.

Whole blood samples were collected from the brachial vein. Approximately 10 μL was then transferred to FTA Elute micro cards (Qiagen, Hilden, Germany) and allowed to dry at room temperature before transportation. Genomic DNA (gDNA) was then extracted from the dried blood spots on the FTA cards following the manufacturer's instructions.

### 3.3.2 Genotyping, haplotype identification, and nomenclature

SNP genotyping of the MHC-B region was done using the panel previously described by Fulton et al. (2016a). Briefly, a subset of 90 SNPs (from MHCJ6 to MHC178) that cover the region between 30,189 and 204,933 bp of the MHC sequence defined by Shiina et al. (2007, GenBank AB268588) were used for the genotyping procedure. Genotyping for each SNP was done using the PCR-based KASP (LGC Biosearch Technologies, Middlesex, UK) procedure, which uses allele-specific competitive amplification and fluorescence-based endpoint reads (Semagn et al., 2013). Further details, including the specific primers can be found in Fulton et al., 2016a.

The 2-step procedure described by Fulton et al. (2016a) was used to identify haplotypes. Individuals that were homozygous for all SNPs were identified first, and an initial set of haplotypes was defined. From there, heterozygous SNP patterns were compared to the initial haplotypes. Previously identified haplotypes were subtracted, and the remaining SNP patterns were used to define additional haplotypes that were present in the population. All haplotypes that were identified in the Chantecler populations were then compared against haplotypes reported in previous studies that used the same SNP panel (Fulton et al., 2016a, Fulton et al., 2016b,

Nguyen-Phuc et al. 2016, Fulton et al., 2017, Iglesias et al., 2019 & 2021, Tarrant et al., 2020, Manjula et al., 2020 & 2021).

## 3.4 Results

### 3.4.1 MHC-B haplotype variation and frequency

The MHC-B haplotype information found in the 4 Chantecler populations is summarized in Table 3.1. There were a total of 8 haplotypes found across all 4 populations, with a range of 4-6 found within each line. Comparison of the specific SNP-defined haplotypes shows that 6 of these had been previously reported in other breeds (Standard Haplotype) while 2 of them were unique to the Chantecler populations tested.

| Population | N | Total Haplotype | Standard Haplotypes | Unique Haplotypes | No. homozygous | No. heterozygous |
|---|---|---|---|---|---|---|
| 1 | 40 | 5 | 3 | 2 | 15 | 25 |
| 2 | 26 | 7 | 5 | 2 | 13 | 13 |
| 3 | 16 | 4 | 3 | 1 | 7 | 9 |
| 4 | 18 | 6 | 5 | 1 | 6 | 12 |
| Total | 100 | 8 | 6 | 2 | | |

Table 3.1: Chantecler populations evaluated, the sample size, and number of haplotypes found in each population. The number of homozygous and heterozygous individuals in each population is also indicated.

Figure 3.1 shows the MHC-BSNP haplotypes found in the Chantecler populations. The 6 standard haplotypes found are: BSNP-A09A, BSNP-C02, BSNP-C05A/B, BSNP-K02, BSNP-

R01, and BSNP-V03 as defined by Fulton and co-workers (2016a, 2016b). Two additional haplotypes, unique to the Chantecler, were also found and named BSNP-Chant01 and BSNP-Chant02. Only 3 of the haplotypes were common to all 4 Chantecler populations: BSNP-C02 and BSNP-V03, both of which were previously identified in heritage broilers, and the novel BSNP-Chant01.

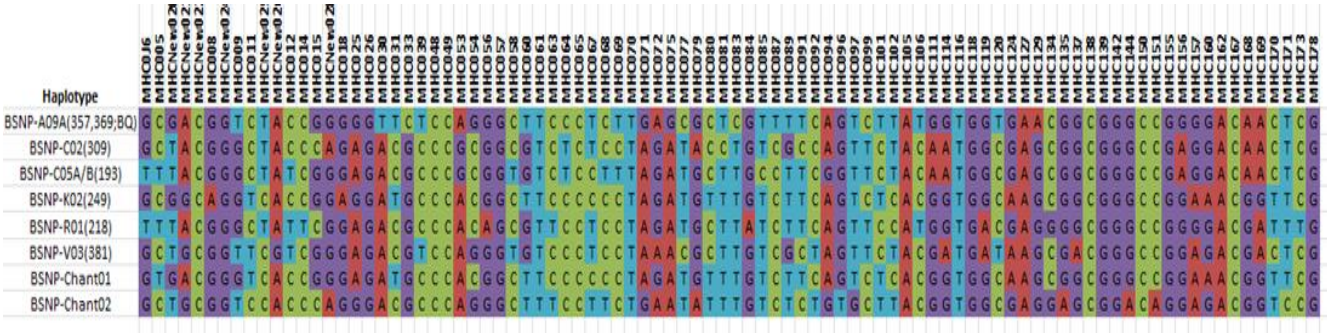


Figure 3.1: MHC-BSNP haplotypes found in Chantecler chicken populations. The 2 novel Chantecler haplotypes are highlighted in yellow. For each haplotype, the allele variant found for each SNP is indicated. For the previously defined haplotypes, the VNTR marker LEI0258 allele sizes (bp) are in parentheses. VNTR: Variable number tandem repeat.

The frequencies of each haplotype in the 4 Chantecler populations are summarized in Table 3.2. The BSNP-Chan01 haplotype had the highest frequency overall (37%), with the highest percentages found in populations 1 and 2 (56% and 35% respectively) in the table. The haplotype frequencies also reveal that each population had 2 to 3 haplotypes at a frequency of more than 0.10, with additional haplotypes occurring at frequencies less than 0.10. The haplotype with the lowest frequency, BSNP-K02, occurred only once in population 1.

| MHC-B SNP haplotype | Population1 | Population 2 | Population 3 | Population 4 | Total Observed |
|---|---|---|---|---|---|
| BSNP-A09A | | 0.02 | | 0.08 | 4 |
| BSNP-C02A/B | 0.1 | 0.08 | 0.66 | 0.39 | 47 |
| BSNP-C05 | | 0.06 | | 0.14 | 8 |
| BSNP-K02 | 0.013 | | | | 1 |
| BSNP-R01 | | 0.21 | 0.22 | 0.11 | 22 |
| BSNP-V03 | 0.26 | 0.27 | 0.06 | 0.03 | 38 |
| BSNP-Chant01 | 0.56 | 0.35 | 0.06 | 0.25 | 74 |
| BSNP-Chant02 | 0.06 | 0.02 | | | 6 |
| | | | | | 200 |

Table 3.2: Summary of MHC-B SNP haplotypes and their frequencies in the Chantecler populations. Total observations reflect the number of observations for each haplotype in the entire data set.

### 3.4.2 Comparison to other breeds

A previous study reported the SNP haplotypes in populations of broiler and heritage chicken breeds (Fulton et al., 2016b). Table 3.3 shows the haplotypes shared between the Chantecler breed and the breeds examined in this study. BSNP-C05 is represented by C05A and C05B. While these 2 haplotypes are identical in regard to the SNP panel, they were previously found to have variation in the upstream region, which was later excluded from the panel (Fulton et al., 2016a). Both iterations are shown in the table as they were discovered in different breeds, but in the Chantecler, they are referred to collectively as BSNP-C05.

| | A09A | C02 | C05A | C05B | K02 | R01 | V03 |
|---|---|---|---|---|---|---|---|
| Barred Plymouth Rock | | | | | + | | |
| Broiler | + | + | + | | + | | + |
| Epileptic Synthetic Line | + | | | + | + | | |
| New Hampshire | | | | + | | + | |
| Red Junglefowl | + | | | | | | |

Table 3.3: Previously reported haplotypes in other chicken breeds as well as Chantecler breed.

### 3.5.3 Novel Haplotypes

There were 2 novel haplotypes found in the Chantecler populations, BSNP-Chant01 and BSNP-Chant02 (Figure 3.1). BSNP-Chant01 is potentially a recombinant haplotype involving BSNP-K02, as they share identity for 73 consecutive SNPs starting at MHC026 until the final SNP, MHC178 (Figure 3.2). However, the second parental haplotype was not identified in the population, indicating it has either been lost since the recombination event or was not sampled in the populations. Lacking the second parental haplotype in the data means that the location of the recombination cannot be determined, but the substantial portion of shared identity between BSNP-Chant01 and BSNP-K02 gives confidence that BSNP-Chant01 is a recombinant.

The putative recombinant spot (between MHC25 and MHC26) is one that was previously identified by Fulton et al. (2016a) as a recombination hotspot. Comparison of the first 17 SNPs of BSNP-Chant01 to previously reported haplotypes does not reveal any strong candidates for the second parental haplotype. However, 9 consecutive SNPs in this group, from MHCNew024 to MHCNew028, also share identity with BSNP-K02. This brings the overall similarity of the SNPs of the 2 haplotypes to 92.22%. However, the occurrence of 2 SNP changes in a row (at

MHC18 and MHC25) adds confidence to the likelihood of the recombination not including the upstream section. Instead, the commonality between these SNPs may just be because the second parental haplotype and the K02 haplotype are similar in that region.

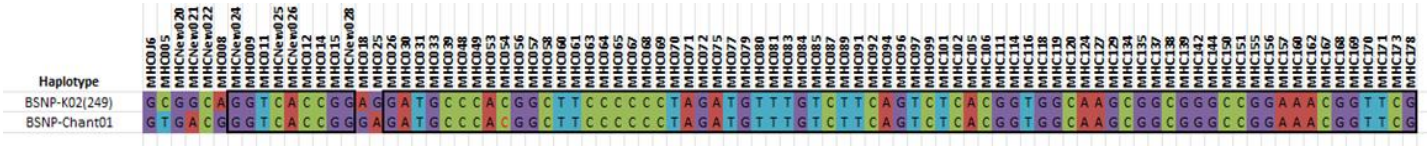


Figure 3.2: Possible recombinant haplotype BSNP-Chant01 aligned with BSNP-K02. Crossover regions are indicated as overlapped black rectangles.

During the SNP genotyping, MHC119 and MHC120 in BSNP-Chant01 heterozygous individuals showed unexpected results. In the example of Chant01/V03 heterozygote, the PCR process should show an even mix of G and A SNPs for MHC119, and C and T for MHC120. However, the Chant01 bases (G and C) exhibited much higher levels than V03 bases (A and T). The reason for this is unknown. Further examination of the sequence data of the region would be necessary to determine the full extent of variation in this region. Further testing of the BSNP-Chant01 haplotype may also be useful to determine its associations with disease resistance.

Compared to the haplotypes defined in the populations, BSNP-Chant02 does not share any portions of identity and thus at this time cannot be considered as a potential recombinant. Additional sampling of the MHCB haplotypes found in the Chantecler may reveal the existence of possible parental haplotypes.

### 3.5 Discussion

Despite experiencing a near extinction event, which likely resulted in a genetic bottleneck and loss of diversity, the Chantecler breed still shows diversity in the MHC region. Regardless of

a smaller sample size, 6 previously defined haplotypes and 2 unique haplotypes were present. While there is some commonality of haplotypes across the populations, they occurred at different frequencies, indicating diversity within the breed. Outside of the class I and II genes, the MHC-B region has been found to evolve rapidly (Shiina et al., 2007), resulting in high levels of haplotype variation. Segregated breeds and populations will face different factors, such as pathogens, that will influence selection in the MHC region, resulting in differences among the haplotypes of different breeds. The discovery of recombination being more prevalent in the region than previously thought (Fulton et al., 2016a), will also increase haplotype variation across breeds.

We did not detect any shared haplotypes between the Chantecler and its parental breeds, based on previously reported data (Fulton et al., 2016b). However, there are links among the breeds. Of the 8 haplotypes found in Chantecler chickens, 5 were previously reported in broiler lines (Table 3). Many of the heritage breeds that were used to create the Chantecler were also used in the creation of various broiler lines. Plymouth Rock and Cornish lines in particular were widely used for broiler creation (Crawford, R.D. 1990). Based on the above-mentioned information, it can be speculated that the Chantecler is more closely related to broilers than other breeds, including layer breeds. However, caution must be exercised. The MHC-B region is only a small part of the chicken genome. While the class I and II genes found in the core region show evidence for being linkage disequilibrium (Kaufman, 2018), the evidence that genes outside of this region evolve at a more rapid rate (Shiina et al., 2007) and the presence of multiple recombinant hotspots in the region (Fulton et al., 2016a) show that the region is highly variable. Therefore, MHC-B haplotype or sequence data may not be sufficient for measuring genetic relatedness among breeds. The SNP panel also only characterizes a small portion of the MHC-B

region and SNPs in the region that are excluded from the panel could vary among two breeds even if they share the same MHC-B SNP panel-based haplotype.

Broiler lines and Chanteclers have undergone separate selection processes that resulted in the breeds recognized today. Different characteristics were favored for either breed, which has resulted in distinct genetic differences. Considerations of how the selection of one trait may influence other aspects of the animal may not have been fully taken during the creation process. During the creation of these breeds, the knowledge and technology was not available to consider these traits and characteristics on a genetic level. Even with understanding of MHC-B haplotype and its influence on the immune system, selection processes may still have influenced which haplotypes survived. Depending on the geographic region where different lines were created, and the diseases most prevalent there, natural selection and artificial selection both may hold some responsibility for the haplotypes that are present in these lines today. It can then be hypothesized that the haplotypes found in both broilers and Chantecler may have positive or neutral properties, which have resulted in the haplotypes remaining present in both lines, despite their separate histories.

While the SNP panel does give evidence that the Chantecler may be more closely related to broilers than layer breeds, one of main breeds used in the creation of the Chantecler was the White Leghorn, a white layer heritage breed. A chicken diversity panel created by Malomane and coworkers (2019) shows that white layers, including White Leghorns, are clustered separately from brown layers and broilers, with a fair amount of genetic distance between the two groups. Despite its ancestral connection to the White Leghorn, the Chantecler may be preliminarily clustered closer to the broiler and brown layers, based on the data from the SNP panel. However, comparison of the whole genome sequencing of the MHC-B region between the

parental heritage breeds and the Chantecler chickens would be necessary to further reveal the relatedness among these breeds and uncover where the Chantecler falls on the phylogenetic tree of chicken breeds.

Both of the novel haplotypes discovered in the Chantecler populations provide an interesting look into the genetic history of the breed. The high frequency with which the BSNP-Chant01 haplotype occurs (Table 3.2) indicate that the genetic events that resulted in the formation of this haplotype occurred earlier in the history of the breed. This idea is further supported when coupled with the single occurrence of the closely related BSNP-K02 haplotype. There are a couple possibilities for how the BSNP-Chant01 haplotype originated. First, the BSNP-K02 went through recombination with another haplotype that is responsible for the identity of the first 17 SNPs in the BSNP-Chant01 haplotype. The second haplotype here has either disappeared or has yet to be identified in the breed. As previously mentioned, the Chantecler breed experienced a near extinction event, which could have resulted in a population bottleneck and loss of genetic diversity, which may include the second haplotype. The haplotype may also be missing as a result of small sampling numbers. Second, multiple recombination events involving the BSNP-K02 haplotype took place. This would explain why there is a small portion of shared identity prior to the major portion. The beginning of the 9 SNP sequence has also been described as a recombinant hotspot, so it is not unreasonable to speculate on multiple recombination events. However, identifying a potential second parental haplotype in this case would be difficult. Outside of the 2 shared regions, there are only 8 SNPs, split into a group of 6 and a group of 2. The likelihood of multiple haplotypes showing identity is high due to the small

SNP region. As is stands, a second haplotype cannot be identified as the potential second parental haplotype.

The last possibility is that BSNP-Chant01 is not a recombinant haplotype but instead is the result of variation in the BSNP-K02 haplotype. The MHC-B region is highly variable and its connection to the immune system results in high selective pressure acting on the genes present. This is clearly evidenced by the large number of haplotypes defined by the panel. This includes haplotypes that share high levels of similarity but vary slightly, in some cases by only a single SNP. There is a possibility that, over the course of time, SNPs in the BSNP-K02 haplotype varied, and those variants were either neutral or positive and, as a result, persisted to the point that the BSNP-Chant01 haplotype emerged. 5 SNPs differ between the 2 haplotypes, meaning that the BSNP-K02 haplotype would only have had to go through 5 variants at most to evolve to BSNP-Chant01 haplotype.

Of these three possibilities, none can be completely accepted or ruled out. While the 4 populations sampled came from independent sources, the overall sample size for this testing was small and may not fully reflect the true genetic diversity of the Chantecler chicken. Again, the previous events that led to the breed facing extinction could have resulted in diversity being lost through events like population bottleneck, resulting in the loss of second parental haplotype for the BSNP-Chant01 haplotype.

While the BSNP-Chant01 haplotype was found at a high frequency across all the tested populations, the BSNP-Chant02 occurred only 6 times. Unlike BSNP-Chant01, it shows no strong similarity to any previously defined haplotypes, making it even more unique. Looking at the occurrences of this haplotype, 2 hypotheses can be drawn. First, this is a haplotype that has

emerged recently in the population, which explains why it was only found in 2 of the 4 populations tested. Second, this haplotype has been present longer in the breed's history but has been selected against and lost in some populations. Like the BSNP-Chant01 and its parental haplotypes, the BNSP-Chant02 may have also been impacted by the population bottleneck the breed previously experienced. Coupled with the source flocks for sampling being maintained in small sizes, this points to the possibility that BSNP-Chant02 has been present longer in the history of the breed, but different circumstances have resulted in low occurrences in this sampling set.

All 6 occurrences of the BSNP-Chant02 haplotype occur in heterozygous individuals, with 3 being V03/BSNP-Chant02 and 3 being BSNP-Chant01/BSNP-Chant02. This may indicate that the BSNP-Chant02 haplotype on its own may not be as favorable as others, further substantiating the idea that the BSNP-Chant02 haplotype has been selected against. However, heterozygosity is not uncommon in the MHC-B region and may be advantageous. High levels of heterozygosity were reported in previous tested populations, including the Korean native breeds (Manjula et al., 2020) and the wild Red Junglefowl (Nyugen-Phuc et al., 2016). Heterozygosity may come with advantages in the MHC-B region. A study done by Butter et al. (2013) found that, when challenged with infectious bursal disease virus, chickens with the B2 and B12 haplotypes showed very little bursa damage when infected with IBVD, whereas the B15 haplotype experienced high bursa damage. Heterozygous chickens were then challenged with the virus, and the B15B2 heterozygous haplotype showed little damage, whereas the B15B12 haplotype showed moderate damage. In this case, one copy of B2 was as effective as 2 copies, but one copy of B12 was only half as effective as 2 copies. With different haplotypes associated

with different diseases, heterozygous MHC-B haplotypes that confer partial resistance to multiple diseases, as opposed to a homozygous haplotype that confers strong resistance, may be more favorable. In the case of the wild Red Junglefowl, this idea is further substantiated. The exposure to a wide and dynamic repertoire of pathogens means that the MHC in the Red Junglefowl is likely under intense balancing selection (Hedrick, 1998). As a result, the variability and diversity measured in these populations is much larger than that measured in domestic populations.

The distinct properties of the Chantecler chicken make it an interesting subject for genetic studies. Categorizing the variation of the MHC-B region in the Chantecler chicken is another contribution to the growing efforts to understand the true nature of the polymorphic MHC region in chickens. Discovery of 2 novel haplotypes provides opportunity for further studies that investigate the relationship with MHC haplotypes and disease resistance. The unique history and physical characteristics of the Chantecler breed makes it a good candidate for pangenome sequencing to identify and define the genetic variants within this distinctive breed.

# 4. Comparison of the MHC-B diversity of commercial and domestic chicken breeds

## 4.1 Abstract

The highly polymorphic major histocompatibility complex (MHC) plays an important role in the immune function of the chicken. Haplotypes of this region have been linked to resistance and susceptibility of many important diseases. Analysis of region, particularly the class I and II MHC genes, is beneficial as it can uncover the variation and diversity which impact the region's function. This study aimed to evaluate the polymorphism of the MHC-B region in the wild Red Junglefowl and compare it to the diversity present in commercial and domestic chicken lines. Whole genome sequence data was evaluated using the GATK best practices pipeline for 6 chicken breeds, and the Red Junglefowl exhibited the highest level of diversity, and the white layer population showed the lowest. Within all the breeds, diversity was highest in the class I and II region (average of 25 SNPs per kB). Of those genes, the BF2 gene was the most diverse, followed by the BF1 and BLB1 genes. The Ensembl variant effect predictor (VEP) tool also found that over 50% of SNPs found in the coding region have moderate to high impact, based on the calculated variant consequence. The identification of the abundant polymorphisms present in the region will be useful for further research into the association of MHC variation with immune function.

## 4.2 Introduction

Domestication of the modern chicken has played an important role in agriculture and food production. As the popularity of chicken as protein source continues to grow worldwide, so

does the industry. Estimates of the global poultry industry put the total amount of poultry produced at 101 million tons, with a value of around US$140 billion (USDA Foreign Agricultural Services, 2022). Four major regions, the United States, Brazil, China, and the EU, are responsible for 60% of the world production (Owusu-Apenten and Vieira, 2022). Development of both the layer and broiler chickens used in commercial production was the result of artificial selection. Traits related to egg laying and meat production were prioritized to create the lines that are mass produced and used for global consumption.

This intensive selection has led to increased inbreeding and loss of genetic variation, both of which are of major concern to the industry (Muir et al., 2008). Lack of genetic diversity in commercial agriculture may be detrimental to the livelihood of species, with some important traits being lost as others are favored in selection. Indigenous or rare breed populations, as well as ld populations that are less influenced by artificial selection can provide valuable sources of genetic information. In turn, this can be used to maintain and improve genetic diversity in commercial lines, resulting in improvement of production and overall quality of life.

Still found in its native ranges, covering a sizable region across Southeast and Central Asia (Brisbin et al., 2002), the Red Junglefowl (*Gallus gallus*) is considered to be the main ancestor of the modern chicken (*Gallus gallus domesticus*). While there is evidence of genetic contribution from other wild junglefowl species (Eriksson et al., 2008), the majority of the nuclear chicken genome can be traced back to the Red Junglefowl (Granevitze et al., 2007). Wild Red Junglefowl populations have been found to exhibit high levels of genetic diversity (Granevitze et al., 2007; Nguyen-Phuc et al., 2016; Malomane et al., 2021), especially when compared to commercial layers and broilers, which often display high levels of inbreeding and

gene fixation. Being able to compare modern chicken breeds to their ancestral origins is a unique opportunity that can greatly benefit the poultry industry. The ability to track changes in the genome can help uncover important traits and properties of various genes of interest. It also allows for reintegration of lost variation, which can assist in improving modern production lines, as well as maintaining unique breeds.

Variation in the genes located in major histocompatibility complex (MHC)-B region has been studied extensively since its discovery and continues to be a major focus of research, due to its strong correlations with the immune system and disease resistance. A SNP panel developed by Fulton et al., (2016a) is the current standard for haplotyping of the region and has been used to characterize the variation present in many different chicken breeds including the Red Junglefowl (Nguyen-Phuc et al., 2016). This study detailed high levels of diversity, with a total of 310 unique haplotypes in 398 samples. While the SNP panel has been used in different populations, including both commercial and indigenous breeds, no other breed has reported levels of variation nearly as high as the Red Junglefowl.

How much of the variation that has been described is impactful to the overall immune function of chickens is still being discovered. The purpose of this study was to examine the variation present in the MHC-B region of the Red Junglefowl, present in the form single nucleotide polymorphisms (SNPs) and compare that variation to other commercial and domestic chicken breeds. By looking at the genetic makeup of the ancestral species, we hoped to uncover insights into the evolutionary history of the domestic chicken, which can then provide information for future studies concerning MHC variation and immune function.

## 4.3 Materials and Methods

### 4.3.1 Sample Selection

Raw whole genome sequence data for 6 chicken breeds were used. Samples for Red Junglefowl (RJF, n=12), White Leghorn (WLH, n=12), and Silkie (SILK, n=12) breeds came from ChickenSD database. This data was previously collected by Wang et al. (2020). Samples for white layers (WL, n=12), brown layers (BL, n=12), and broiler (BR, n=12) breeds came from a previous study by Qanbari et al (2019) (Genbank accession: PRJEB30270).

### 4.3.2 Sequence Alignment and Variant Calling

Chicken reference genome assembly (Galgal7 version) was downloaded from NCBI. SAMtools (Samtools 1.10) was used to create an index of the reference sequence (Li et al., 2009). Trimmomatic (Trimmomatic 0.39) was used to obtain clean reads, with the following parameters applied for quality control: LEADING:3 TRAILING:3 SLIDINGWINDOW:5:20 MINLEN:36 (Bolger et al., 2014). The Burrows-Wheeler Alignment tool (BWA-0.7.17) was used to map reads to the reference genome with the command "mem -t 10 -M" (Li and Durban, 2010). Following alignment, the data was piped to SAMtools where it was converted to a BAM file and sorted according to the reference genome. The Genome Analysis Toolkit (GATK-4.2.6.1) MarkDuplicates module was used to remove duplicate reads (McKenna et al., 2010). The GATK HaplotypeCaller module was then used to convert the BAM files to GVCF and then switched to VCF files. SNPs and INDEL were called only for the 16$^{th}$ chromosome (NC_052547.1), where the MHC region is located. The GATK CombineGVCFs module was then used to combine the VCF files for all the samples into one file. Following combination, the

GATK GenotypeGVCFs module was used to perform joint genotyping on the data set. The GATK SelectVariants module was used to create two datasets, one for SNPs and one for INDEL, and the VariantFIltration module was used for hard filtering for both. For SNPs, the following criteria were used to exclude potential false variants: QUAL < 30.0 || QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 3.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0. For INDEL, the parameters are as follows: QUAL < 30.0 || QD < 2.0 || FS > 200.0 || SOR > 10.0 || MQRankSum < −12.5 || ReadPosRankSum < −20.0.

### 4.3.3 SNP Annotation and Analysis

VCF files were run through the Ensembl Variant Effect Predictor (VEP) tool to annotate and determine the effects of the variants (McLaren et al., 2016). The default parameters set by VEP were used for analysis, which included Ensembl/GENCODE transcript databases as the comparison datasets. Because VEP uses multiple transcripts for annotation, many SNPs are categorized under different genes. For analysis purposes, Ensemble canonical transcripts were used, and SNPs were assigned to genes based on location and indication of exon/intron location within genes, rather than just the symbol as indicated on the VEP results. This restricted view allowed for clearer analysis of SNP behavior in the genes of interest. Only those SNPs determined to be located within the genes were used for analysis.

When determining the effects of variants, VEP assigns impact levels based on the calculated variant consequence. Consequences are determined based on the transcripts available in the Ensembl databases. Impacts can be classified as either modifier, low, moderate, or high. For this study modifier impacts include 5 prime UTR variants, 3 prime UTR variants, and intron

variants. Variants with low impacts include synonymous variants, splice region variants, splice donor region variants, and splice polypyrimidine variants. Missense variants are classified as moderate impact variants, and high impact variants are stop gained variants, start lost variants, and splice acceptor variants. VEP also provides the affected codons and amino acids in the coding sequence and will state if the variant already exists.

## 4.4 Results

### 4.4.1 SNP Diversity

For each of the breeds in question, the combined VCF file featuring the 12 samples was run through the VEP tool. Table 4.1 shows the counts for total variants for each breed at three different levels. First, the total number of variants for the entire chromosome 16 is shown. Second is the total number of variants for the whole MHC-B region. This covers the genes in the region from KIFC1 to CD1C (NC_052547.1:2061452 to NC_052547.1:2294442). The third region is the core MHC-B region, also referred to as the class I and II region, due to the presence of the class I and II genes. This region covers the genes from BLB1 to BF2. The percentage of each region represented by the SNPs is also shown. Based on gene information from NCBI, the size for chromosome 16 (NC_052547.1) is 2,706,039 bp, the MHC-B region is approximately 227,467 bp, and the class I and II region is 43,934 bp.

| | Chromosome 16<br>Total SNP | <br>% | MHC-B Region<br>Total SNP | <br>% | Class I and II<br>Total SNP | <br>% |
|---|---|---|---|---|---|---|
| Brown Layer | 23701 | 0.876 | 4282 | 1.882 | 1115 | 2.538 |
| Broiler | 31427 | 1.162 | 4825 | 2.121 | 1310 | 2.982 |
| Red Junglefowl | 37136 | 1.372 | 5664 | 2.49 | 1520 | 3.46 |
| Silkie | 29558 | 1.092 | 4126 | 1.814 | 1061 | 2.415 |
| White Layer | 16023 | 0.592 | 2317 | 1.019 | 612 | 1.393 |
| White Leghorn | 22262 | 0.823 | 3603 | 1.584 | 875 | 1.992 |
| Average | 26685 | 0.986 | 4136 | 1.818 | 1082 | 2.463 |

Table 4.1: Total SNP variant counts for each breed at the level of chromosome 16, MHC-B region, and Class I and II regions. The percentage of each region represented by the SNPs is also included.

Diversity was the highest in the Red Junglefowl breed in all three regions, followed by the broiler. The Silkie and brown layer breeds had similar levels of diversity for the MHC-B and class I and II regions, but the Silkie showed a higher diversity for chromosome 16. The White Leghorn and white layer breeds exhibited the lowest diversity. For all breeds, SNP frequency was the highest within the class I and II region, with an average of 25 SNPs per kB. The MHC-B region also exhibited higher frequency than chromosome 16 as a whole, with an average of 18 SNPs per kB, as compared to 10 SNPs per kB. The frequency of SNPs found in this region reflect just how diverse the MHC is, as SNPs have been found to occur on average once every 1,000 nucleotides.

### 4.4.2 Class I and II Gene Region

The class I and II region cover the genes from BLB1 to BF2. Included are the class I genes (BF1 and BF2), the class II genes (BLB1 and BLB2), TAP genes (TAP1, TAP2, and TAPBP), DM genes (MHCDMA, DMB1, and DMB2), and BRD2. The region is known to be highly diverse, with variation in the BF2 gene in particular thought to be responsible for the strong associations of the chicken MHC-B with disease (Zhang et al., 2012). Among the genes in this region, BF2 showed the highest diversity overall, with an average of 47 SNPs per kb. The SNP frequency was highest in this gene for five of the six breeds, and tied for highest in the BR breed, with BF1. BF1 and BLB2 followed with averages of 35 and 28 SNPs per kb, respectively. The least diverse genes were BRD2, and TAPBP, with averages of 11, and 16 SNPs per kb, respectively (Figure 4.1). The class I and II genes exhibiting the highest levels is in line with their roles in antigen presentation.

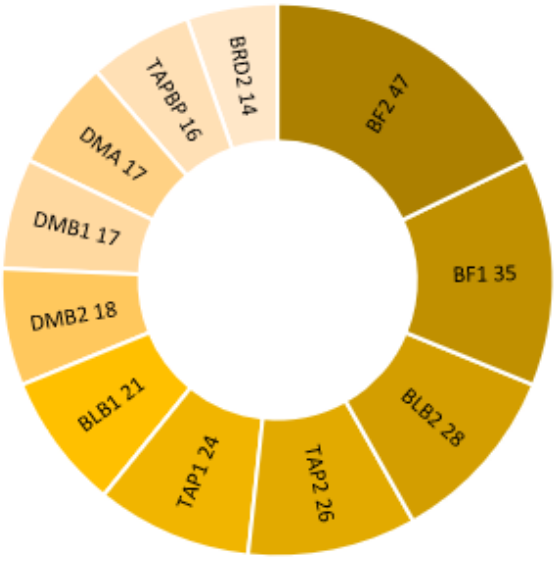


Figure 4.1: The number of SNPs per kb for 11 genes. The SNP rates from low to high in a clockwise direction, starting with BF2.

Within this region, the RJF and BR lines again showed the highest diversity, whereas the WLH and WL showed the lowest overall diversity. Across the 11 genes, SNP per kb averages for each population were as follows: RJF- 33.27, BR- 28.61, BL- 25.12, SILK- 24.71, WLH- 18.87, WL-14.69. Table 4.2 shows the SNP frequency for each of the 11 genes for each breed.

| | BLB1 | TAPBP | BLB2 | BRD2 | DMA | DMB1 | DMB2 | BF1 | TAP1 | TAP2 | BF2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BL | 0.021 | 0.021 | 0.023 | 0.014 | 0.016 | 0.02 | 0.02 | 0.044 | 0.02 | 0.021 | 0.057 |
| BR | 0.022 | 0.018 | 0.03 | 0.016 | 0.023 | 0.019 | 0.023 | 0.05 | 0.031 | 0.033 | 0.05 |
| RJF | 0.033 | 0.022 | 0.041 | 0.02 | 0.019 | 0.026 | 0.026 | 0.047 | 0.035 | 0.039 | 0.059 |
| SILK | 0.026 | 0.015 | 0.025 | 0.017 | 0.019 | 0.018 | 0.017 | 0.033 | 0.026 | 0.024 | 0.05 |
| WL | 0.015 | 0.011 | 0.025 | 0.007 | 0.011 | 0.006 | 0.008 | 0.02 | 0.014 | 0.015 | 0.03 |
| WLH | 0.011 | 0.011 | 0.026 | 0.011 | 0.016 | 0.014 | 0.017 | 0.017 | 0.021 | 0.024 | 0.038 |
| Average | 0.021 | 0.016 | 0.028 | 0.014 | 0.017 | 0.017 | 0.018 | 0.035 | 0.024 | 0.026 | 0.047 |

Table 4.2: SNP frequencies for 11 class I and class II genes.

In the class I and II region as a whole, the percentage of SNPs occurring in exons and introns was 52.6% and 47.4%, showing a closely even split. Of the eleven genes, 6 were found to have higher occurrences of SNPs in the exons (BLB1, BLB2, DMB1, BF1, TAP2, BF2), while the remaining 5 had higher occurrences of intron SNPs (TAPBP, BRD2, DMA, DMB2, TAP1). Within the functional regions (Figure 4.2a and 4.2b), 55% of the SNPs were modifiers (intron variant, 3 prime UTR, 5 prime UTR), 22% had low impact (synonymous variants, splice region variants), 23% had moderate impact (missense variants), and 0.3% having high impact (stop gained, start lost, splice acceptor). This shows the variation among SNPs, and the different impacts they can have on the genes in which they are located.

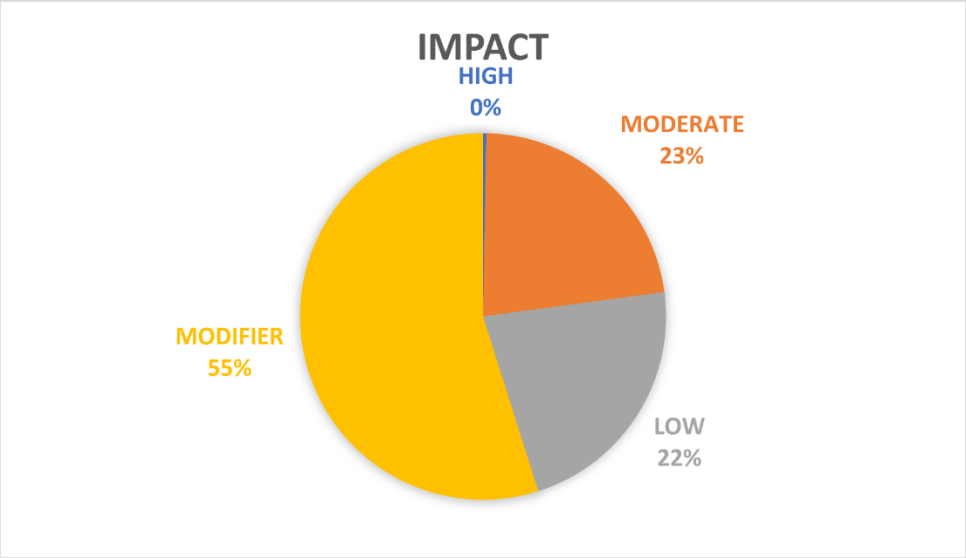


Figure 4.2a: Average calculated impact of SNPs found in the functional regions of the class I and II genes.

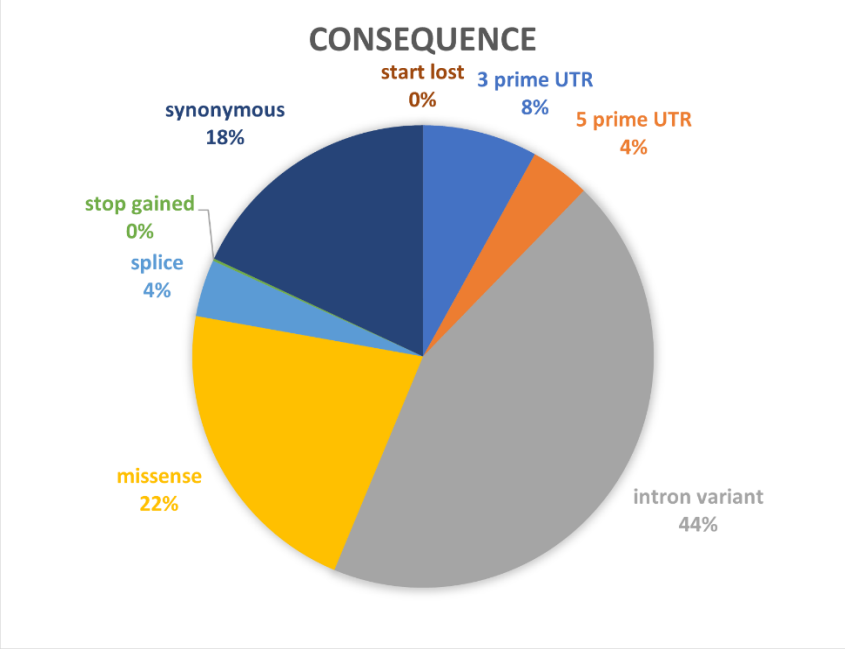


Figure 4.2b: Average calculated consequence of SNPs found in the functional regions of the class I and II genes.

Within the coding regions, more than half the SNPs had moderate to high impacts (54%). The majority of these variants were missense variants, with a small amount of stop gain, start

lost, and splice acceptor variants. The remaining 46% of SNPs were synonymous or other slice region variants, which have low impacts (Figure 4.3).

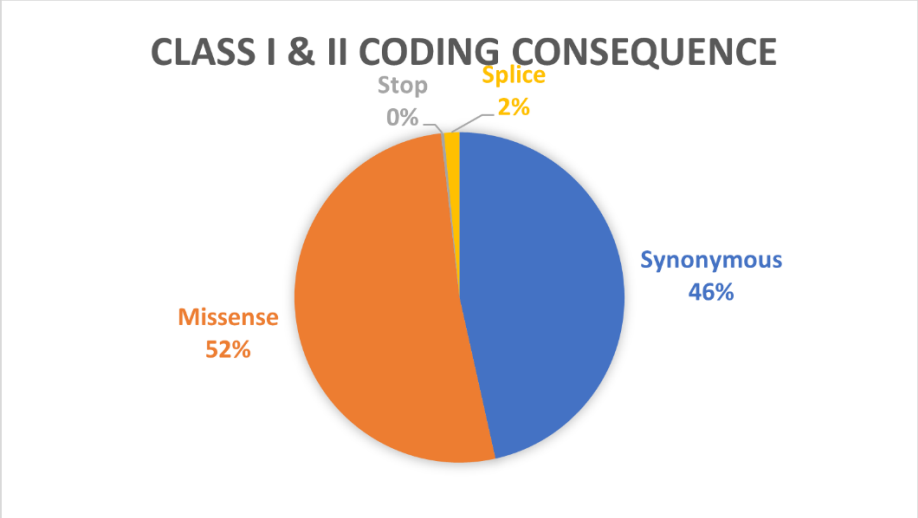


Figure 4.3: Average percentage of consequences of SNPs located in the coding region of class I and II genes.

Within the two class I genes (BF1 and BF2) approximately 50% of the SNPs were located within exons 2 and 3. For both of these genes, this distribution is quite disproportional. An even distribution for BF1, which has 8 exons, would 12.5% per exon. For BF2, which has 7 exons, it would be 14.3%. Across all 6 populations, this high disproportionate distribution is observed, with the RJF and BR breeds measuring around 60% for the BF1 gene. The only instance where this is not the case is BF1 gene in the WLH, which showed only 30% of SNPs in the two exons of interest. The SNPs in the two exons of these two genes also exhibit higher impact, with ~80% of SNPs being missense variants, which have moderate impact, and ~20% being synonymous variants, with low impact (Figure 4.4). The increase in variation in these genes may be linked to disease resistance, as well as their roles in antigen presentation.

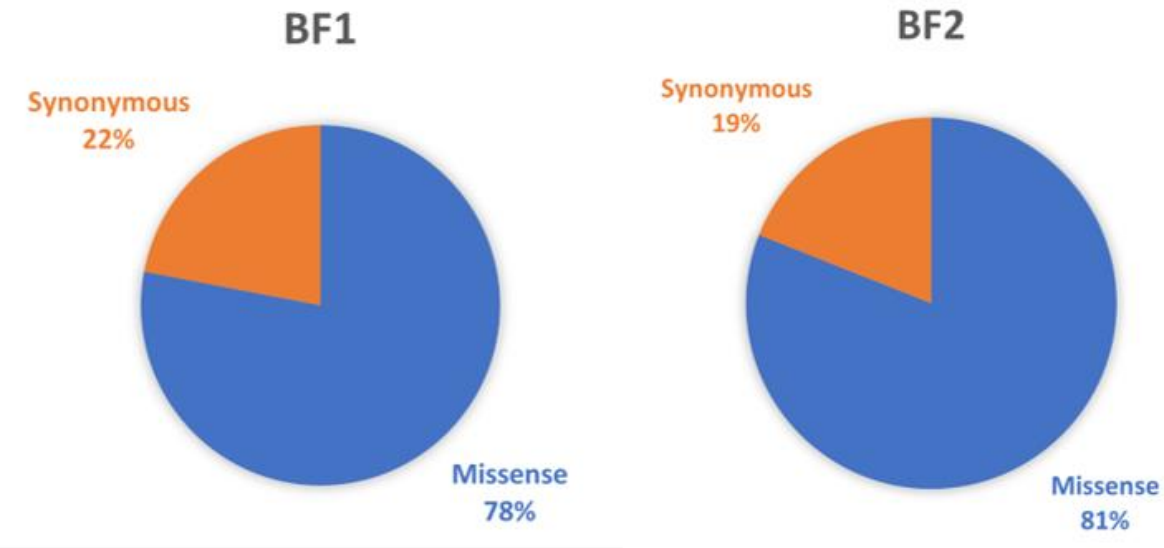


Figure 4.4: Average percentage of consequences of SNPs located within exon 2 and 3 of the BF1 and BF2 genes.

## 4.5 Discussion

Overall, the Red Junglefowl shows the greatest level of diversity. This is in line with a previous study of the MHC-B region (Nyugen-Phuc et al., 2016). This study used a SNP panel (Fulton et al., 2016a), to define haplotypes for the extended MHC-B region and found incredibly high levels of diversity for the Red Junglefowl as compared to other commercial and domestic chicken breeds. The wild nature of the Red Junglefowl means that it has not been subjected to artificial selection, which may in turn influence the overall genetic diversity. Red Junglefowl are also exposed to a wide range of pathogens and parasites that may influence the balancing selection that is thought to maintain the diversity in the MHC region (Hess and Edwards, 2002). A similar study was done with Silkie populations (Tarrant et al., 2020). While not nearly as diverse as the Red Junglefowl, considerable levels of diversity for the MHC-B region were observed for the Silkie breed. Similar observations can be made here, with considerable diversity

defined for Silkies, but still less than the Red Junglefowl. The samples for this study were collected from a commercial breeder colony, whereas the samples available from ChickenSD (used for this study) were taken from local populations in the Jiangxi province in China. Further information about both sources is not available, but the differences between local populations versus commercial breeder colonies may be an interesting topic for future studies.

Diversity was the lowest in the white layer and White Leghorn breeds, with the white layers showing total SNP counts less than half of what was observed in the Red Junglefowl. The white layers used for this study were originally established from White Leghorn lines (Qanbari et al., 2019), so it is expected that they may exhibit some similarities in terms of diversity. However, there is still a significant difference between these two breeds, with white layers showing the lower overall diversity, and White Leghorns showing similar levels with the brown layer breeds. Lower levels of diversity in the MHC-B region have been reported in populations of heritage breeds, including the White Leghorn (Fulton et al., 2016b), and the progenitor of the brown layer population used in the study, the Rhode Island Red. It is not uncommon to observe lower genetic diversity in breeds like these, as they are typically maintained as closed populations, which restricts gene flow. In the cases of the white and brown layer breeds, artificial selection also plays a role in the overall diversity levels. As different traits were prioritized during the creation of the lines, and closed breeding used to maintain said lines, genetic diversity has suffered as a consequence (Muir et al., 2008).

The broiler breed exhibited the second highest diversity. The higher levels of diversity in the broiler versus layer lines is also in line with the previously mentioned study (Fulton et al., 2016b). Compared to the heritage breeds sampled, more MHC-B haplotypes were observed in

the broiler breed. This creates an interesting divide between broiler and layer lines. Both groups originated from heritage chicken breeds, with some shared origins, particularly between broiler dam lines and brown layers. As a result, broiler lines have also been found to suffer from overall loss of genetic diversity similar to layer lines (Muir et al., 2008). Further use of a SNP panel to observe genetic diversity and compare the regions in the genome most affected for these lines may show the differences that have arisen as the result of different breeding histories, and why diversity in the MHC-B region may be higher in the broiler and brown layer lines as compared to the white layer line.

The MHC-B region as a whole is known to be highly diverse, with the core, class I and II regions in particular known to show high levels of polymorphism. This was observed in this study, with the overall diversity being higher in the core region. The high levels of diversity in class I and II genes are in line with their roles in antigen presentation. The BF and TAP genes have been the focus of many chicken MHC studies, due to their important roles as antigen presenters. The high diversity observed is consistent with these roles and previous studies of these genes. Of the two BF genes, BF2 has been found to be dominantly expressed, and is thought to play a key role in the strong associations between MHC-B haplotype and disease resistance. This dominant expression of the BF2 gene has been thought to come about as the result of co-evolution with the TAP genes, resulting in unique allelic combinations across haplotypes, leading to high levels of specificity observed among different haplotypes (Kaufman, 2015). High polymorphism within these genes can contribute to the effective antigen presentation, as well as the host's broad-spectrum resistance to pathogens. Understanding the diversity of these regions will result in better understanding of immune functions.

Exons 2 and 3 in BF genes encode the α1 and α2 domains of MHC class I molecules, which play important roles in disease resistance (Hunt and Fulton, 1998; Kaufman et al., 1999). These domains make up the binding cleft, which allows the molecule to bind with peptides, thereby taking on a more stable structure. Differences in the makeup of these domains allow the MHC molecules to bind more fastidiously or promiscuously with antigens, impacting the ability of the immune system to fight against specific diseases (Zhang et al., 2012; Kaufman, 2018). A majority of SNPs located in the BF1 and BF2 genes are located within these two exons. Of these SNPs around 80% have a moderate impact. Not only are these genes showing high levels of polymorphism, which has been previously observed, but that polymorphism has impact, which may be responsible for different behaviors of BF genes which have been observed in different haplotypes.

A similar study was conducted by a Chinese group (Yuan et al., 2021) just like our current study. In terms of gene diversity, some similar results were observed, with BF2, TAP2 (which was the fourth most diverse in this study), and BF1 as the most diverse genes. However, in that study the least diverse genes were the two class II BLB genes (Yuan et al., 2021), while our study reported higher levels of diversity for all the genes, in terms of SNP per kb. The difference may be in part due to different reference genomes used (Galgal6 for Yuan et al., 2021, Galgal7 for our current study), as well as the genetic differences between the breeds studied. Twenty-one breeds were used for the Chinese study, of which 18 were native Chinese breeds, which have been previously shown to exhibit high levels of diversity (Han et al., 2013), resulting in higher levels of diversity overall being observed. In that study, the remaining three breeds (which included Red Junglefowl and two inbred lines), showed lower diversity overall, but

would have less impact on the averages as there were 21 breeds in total. In this study, the WL and WLH breeds both exhibit low overall diversity but have a larger impact on the overall diversity as they make up a third of the breeds used in this study. Therefore, the average diversity for the 6 breeds is lowered, creating a wider difference between the diversity levels of the breeds of the two studies.

Regardless of the differences of the results, both studies show that the MHC-B region shows high levels of polymorphism, with differences observed across breeds. The differences in polymorphism may contribute to the immune function and ability of a host to exhibit broad spectrum disease resistance. It was also observed that over 50% of the polymorphisms observed in the coding region have moderate to high impact on the genes in which they are located. The breeds less affected by artificial selection tend to have higher levels of diversity, suggesting that their exposure to natural selection forces has influenced the density of genetic diversity present. The differences observed across the breeds show that different breeding histories have impacted the MHC-B region. It also shows that the region is able to maintain and possibly generate diversity, as shown by the commercial breeds exhibiting some level of diversity, even though they are in closed breeding systems.

The identification of this large wealth of genetic variants provide data for further research into the associations of MHC variability with immune traits. The abundance of SNPs present across the MHC-B region provides many potential molecular markers for genetic analysis. The linked nature of the MHC genes may also be useful in understanding the nature of MHC-B haplotypes. The SNP panel that is currently used for haplotyping (Fulton et al., 2016a) uses 90 SNPs across the B region. While none of these SNPs are located within the class I and II genes,

some can be found in the surrounding genes in the region, including the TAP genes. If these SNPs can be linked to SNPs within the BF genes, particularly those in exon 2 and 3, it may provide a basis for predicting haplotype behavior, particularly in regard to disease resistance.

This study investigated the class I and II region of the chicken MHC-B. The 6 breeds studied showed varying levels of diversity, in line with their unique breeding histories. Over 50% of SNPs found in the coding region were impactful, showing that the polymorphisms present in the region may be responsible for variation among the genes, which in turn may influence immune function. One example of this is the disproportionate distribution of impactful SNPs in the 2$^{nd}$ and 3$^{rd}$ exons of the BF genes. Changes in these exons may influence the ability of alleles to interact with different pathogens, which may manifest as strong associations with disease resistance. As these variants are better understood, they can be used to help improve poultry welfare in terms of breeding for disease resistance, development of treatments and vaccinations, and preservation of unique genetic diversity.

# 5. General Discussion

As one of the most polymorphic regions in the chicken genome, the major histocompatibility complex is an area of great interest. The roles various genes play in immune function, and the strong correlations to disease resistance have driven research efforts to better understand the region. With the production of poultry increasing worldwide, improving production is a major focus. Understanding of the genetic variation of this region can help with those efforts, in both local, small-scale villages and farms, as well as for larger, commercial production facilities.

The development of the SNP panel (Fulton et al., 2016a) for haplotyping of the MHC-B region has tremendously expanded the view of this region. Through this panel, the previously standardized set of haplotypes has expanded to more fully reflect just how variable the region is. The relative ease and low cost of KASP- based haplotyping has also allowed for more and more breeds to be investigated. Early work on the MHC-B region focused mainly on the White Leghorn breed, with later studies starting to use broilers and other lines (Briles and Briles, 1982; Goto et al., 2002). The diversity of the region cannot be fully understood through a single breed or two, especially when said breeds are typically kept in closed breeding systems, often subjected to artificial selection, which can impact genetic diversity.

Since its development, the SNP panel for MHC-B haplotyping has been used on a wide array of global chicken breeds. Asian breeds including Silkies (Tarrant et al., 2020), Korean native breeds (Manjula et al., 2020), Sri Lankan breeds (Manjula et al., 2021), and the wild Red Junglefowl (Nguyen-Phuc et al., 2016), as well as the Finnish Landrace breed (Fulton et al.,

2017), and Argentinian Campero INTA breed (Iglesias et al., 2019; Iglesias et al., 2021), have all been examined, alongside heritage and broiler breeds from North America (Fulton et al., 2016b). In each of these studies, unique MHC-B diversity was observed in each breed. While there are haplotypes that are common and found across many breeds, novel haplotypes have been defined in each breed. This is also the case for the Chantecler breed, which was haplotyped using the SNP panel in this study.

Like the other breeds mentioned, both previously identified and novel haplotypes were found in the Chantecler populations sampled. This is significant as the breed previously went through a near extinction event, most likely resulting in a population bottleneck and loss of genetic diversity. It is also not a very prominent heritage breed, so the overall population size is small. Regardless, unique MHC diversity was observed. Multiple heritage chicken breeds were used in the creation of the Chantecler, and it appears that the original sources provided much in the way of MHC-B diversity, and the breed was able to maintain that diversity. The emergence of 2 novel haplotypes also demonstrates the ability of the region to generate new diversity. With similar observances recorded across different breeds, it's clear there is much unique variation to be studied in the region.

The SNP panel used for haplotyping uses 90 SNPs spanning around 200,000 base pairs. This is only a fraction of the variants present within the MHC-B region. For the purposes of determining haplotypes, the 90 SNPs used are sufficient, as the haplotypes determined from the panel are consistent with previously defined haplotypes, but also because a level of linkage is observed within the region. Because of this linkage, it is likely that within a population many of the SNPs not included in the panel will be the same for a haplotype. While the purpose of the

panel is determining the haplotype of the MHC-B region, and not about defining variation within genes in the region, it may be a useful tool when looking at effects of variation in the region.

The use of whole genome sequence data can provide a much broader look at the genetic diversity of the MHC genes and be helpful in determining just how impactful said diversity is. Through the use of bioinformatics tools, SNPs in the MHC-B region are able to be located and annotated to examine what effects they may have. Within the 6 breeds examined in this study, varying levels of diversity were shown. However, some properties of the diversity were consistent across the populations.

First, the frequency of SNPs was consistently higher in the core MHC-B region as compared to the extended MHC-B region, and the frequency was higher for the MHC-B region as compared to the whole of chromosome 16. The core region contains the MHC class I and class II genes integral to antigen presentation, and it is these genes that are thought to correlate to the strong associations of MHC with disease resistance (Kaufman, 2018). With coevolution between the class I BF and TAP genes thought to be taking place (Kaufman, 2015), and similar ideas presented for the class II BLB and DM genes (Parker and Kaufman, 2017), unique allelic combinations are emerging between the gene pairs, which result in high levels of specificity observed across haplotypes, and high levels of diversity in the region.

For all the breeds examined, the most variable gene was BF2. This is in line with the role this gene plays in antigen presentation, and the correlation to disease resistance that has been previously observed (Hunt and Fulton, 1998). The 2$^{nd}$ and 3$^{rd}$ exon of this gene code for the domains that form the peptide binding cleft of the MHC molecule, and it has been found that the nature of this binding pocket has a high impact on the ability of the molecule to bind peptides. Depending on the makeup of the molecule, it can either act in a promiscuous way, binding to a

wide range of antigens, or in a more fastidious way, binding strongly to only specific antigens (Zhang et al., 2012; Kaufman, 2018). In the case of the promiscuous alleles, a more broad-spectrum disease resistance is observed, with more or less protection conferred against most pathogens, similar to the behavior of human class I alleles. In the case of the fastidious alleles, it either confers susceptibility or resistance to pathogens. Advantages are present for both types of alleles, as evidenced by the presence of haplotypes associated with both behaviors. In regions where certain diseases may be more prevalent, haplotypes associated fastidious alleles that confer strong protection against said disease may be preferred. In the case of commercial flocks, where disease management is highly controlled, a more broad-spectrum haplotype may be preferred.

It was also found that within the coding regions of the MHC-B region, around 50% of the SNPs were impactful with either moderate or high effects on the genes. This percentage was consistent across all the breeds, demonstrating that the present diversity is impactful even in those breeds where diversity may be lower overall. The variation observed within the breeds is also in line with diversity that was measured using the SNP panel. Overall, the Red Junglefowl showed the highest levels of diversity (Nguyen-Phuc et al., 2016), as to be expected due to the wild nature of the breed. The broiler line showed considerable diversity, and more than that of the layer lines and the heritage White Leghorn breed showed low levels (Fulton et al., 2016b). Finally, the Silkie showed moderate diversity, similar to the broiler and brown layer (Tarrant et al., 2020).

The identification of so many impactful variants provides a wealth of potential molecular markers for future studies. As efforts are made to understand the impact of variation within these genes, and how they may influence the immune function, the nature of different haplotypes can

also be better understood. With the identification of so many new haplotypes, the next step is understanding their behavior. In the initial development of the SNP panel, 78 MHC-B SNP haplotypes were defined and grouped into 22 families (Fulton et al., 2016a). Further expansion of this grouping to include the other defined haplotypes, and then exploration into the variation and diversity within those groupings will be valuable to genetic studies involving the MHC-B region and disease.

# 6. Conclusion

This study aimed to describe the diversity present in the major histocompatibility complex of the chicken. Through the use of a SNP panel, diversity was defined for the Quebec native Chantecler chicken. Novel haplotypes were discovered, contributing to the growing efforts to better understand the variation and polymorphic nature of this important region. Comparisons between different breeds and different haplotypes will be beneficial in further investigation of the relationship between MHC and disease resistance. This study also shows the potential for the Chantecler breed to be the focus of future studies. The unique features that set it apart from other breeds, as well as the identification of unique variation present in the genome show it would be a good candidate for further genetic studies.

The comparison of the MHC of the Red Junglefowl with domestic chicken breeds show that the unique breeding histories of these breeds have influence on the diversity of the region. Regardless of the divergence of breeds, similar observations were made across the populations, in regard to the consequences and impacts of said consequences. Even in those populations where the diversity was lower, the variation observed was still impactful to the same degree as the populations with higher diversity. This shows that the polymorphic nature of the MHC region has influence over the genes involved, which can then impact the immune function of the chicken.

The identification of diversity in the MHC region provides a basis for further work considering the link between variation and immune function. This study shows a prominent level of variation across different chicken breeds, which have been influenced by different selective

pressures. As this variation is better understood, it can be used to improve genetic welfare of chicken breeds, develop new treatments and vaccinations for important diseases, preserve genetic diversity within breeds, and uncover the unique features that influence the immune functions of the chicken.

# 7. References


Abeler-Dörner, L., Swamy, M., Williams, G., Hayday, A. C., & Bas, A. (2012). Butyrophilins: an emerging family of immune regulators. *Trends in Immunology*, *33*(1), 34–41.

A, C., Travers, P., Walport, M., & Shlomchik, M. J. (2014). *The major histocompatibility complex and its functions*. Nih.gov; Garland Science.

Adams, D. R., Ron, D., & Kiely, P. A. (2011). RACK1, A multifaceted scaffolding protein: Structure and function. *Cell Communication and Signaling*, *9*(1), 22.

Ayala García, M. A., González Yebra, B., López Flores, A. L., & Guaní Guerra, E. (2012). The major histocompatibility complex in transplantation. *Journal of Transplantation*, *2012*, 1–7.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Boyington, J. C., Motyka, S. A., Schuck, P., Brooks, A. G., & Sun, P. D. (2000). Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature*, *405*(6786), 537–543.

Briles, W. E., & Briles, R. W. (1982). Identification of haplotypes of the chicken major histocompatibility complex (B). *Immunogenetics*, *15*(5), 449–459.

Briles, W. E., Bumstead, N., Ewert, D. L., Gilmour, D. G., Gogusev, J., Hala, K., Koch, C., Longenecker, B. M., Nordskog, A. W., Pink, J. R. L., Schierman, L. W., Simonsen, M., Toivanen, A., Toivanen, P., Vainio, O., & Wick, G. (1982). Nomenclature for chicken major histocompatibility (B) complex. *Immunogenetics*, *15*(5), 441–447.

Briles, W. E., Goto, R. M., Auffray, C., & Miller, M. M. (1993). A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. *Immunogenetics*, *37*(6).

Briles, W. E., McGibbon, W. H., & Irwin, M. R. (1950). On multiple alleles effecting cellular antigens in the chicken. *Genetics*, *35*(6), 633–652.

Brisbin, I. L., Peterson, A. T., Okimoto, R., & Amato, G. (2002). Characterization of the genetic status of populations of Red Junglefowl. *JOURNAL-BOMBAY NATURAL HISTORY SOCIETY*, *99*(2), 217-223.

Butter, C., Staines, K., van Hateren, A., Davison, T. F., & Kaufman, J. (2013). The peptide motif of the single dominantly expressed class I molecule of the chicken MHC can explain the

response to a molecular defined vaccine of infectious bursal disease virus (IBDV). *Immunogenetics*, *65*(8), 609–618.

Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., Melino, G., & Raschellà, G. (2017). Zinc-finger proteins in health and disease. *Cell Death Discovery*, *3*(1), 1–12. https://doi.org/10.1038/cddiscovery.2017.71

Cavero, D., Schmutz, M., Philipp, H. C., & Preisinger, R. (2009). Breeding to reduce susceptibility to Escherichia coli in layers. *Poultry Science*, *88*(10), 2063–2068.

*Chantecler Chicken*. (n.d.). The Livestock Conservancy. https://livestockconservancy.org/heritage-breeds/heritage-breeds-list/chantecler-chicken/

Chappell, P. E., Meziane, E. K., Harrison, M., Magiera, Ł., Hermann, C., Mears, L., Wrobel, A. G., Durant, C., Nielsen, L. L., Buus, S., Ternette, N., Mwangi, W., Butter, C., Nair, V., Ahyee, T., Duggleby, R., Madrigal, A., Roversi, P., Lea, S. M., & Kaufman, J. (2015). Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. *ELife*, *4*.

Chaves, L. D., Krueth, S. B., & Reed, K. M. (2007). Characterization of the turkey MHC chromosome through genetic and physical mapping. *Cytogenetic and Genome Research*, *117*(1-4), 213–220.

Chaves, L. D., Krueth, S. B., & Reed, K. M. (2009). Defining the turkey MHC: sequence and genes of the B locus. *The Journal of Immunology*, *183*(10), 6530–6537.

Chazara, O., Tixier-Boichard, M., Morin, V., Zoorob, R., & Bed'Hom, B. (2011). Organisation and diversity of the class II DM region of the chicken MHC. *Molecular Immunology*, *48*(9-10), 1263–1271.

Chazara, O., Fulton, J., Juul-Madsen, H., Chang, C., & Bed'Hom, B. (2010). High resolution chicken MHC genotyping using a SNP panel. *Proceedings of the 32nd Conference of the International Society for Animal Genetics, July 26- 30th, Edinburgh, UK*, p.138.

Cooper, G. S., & Stroehla, B. C. (2003). The epidemiology of autoimmune diseases. *Autoimmunity Reviews*, *2*(3), 119–125.

Cotter, P. F., Taylor, R. L., & Abplanalp, H. (1998). B-complex associated immunity to Salmonella enteritidis challenge in congenic chickens. *Poultry Science*, *77*(12), 1846–1851.

Crawford, R. D. (1990). Poultry genetic resources: evolution, diversity, and conservation. *Developments in Animal and Veterinary Sciences (Netherlands)*.

da Silva, A. P., Hauck, R., Zhou, H., & Gallardo, R. A. (2017). Understanding immune resistance to infectious bronchitis using major histocompatibility complex chicken lines. *Avian Diseases*, *61*(3), 358–365.

*Data & Analysis | USDA Foreign Agricultural Service*. (2019). Usda.gov. https://www.fas.usda.gov/data

Dausset, J. (1958). Iso-leuco-anticorps. *Acta Haematologica*, *20*(1-4), 156–166.

D'Cruz, A. A., Babon, J. J., Norton, R. S., Nicola, N. A., & Nicholson, S. E. (2012). Structure and function of the SPRY/B30.2 domain proteins involved in innate immunity. *Protein Science*, *22*(1), 1–10.

Denis, G. V. (2010). Bromodomain coactivators in cancer, obesity, type 2 diabetes, and inflammation. *Discovery Medicine*, *10*(55), 489–499.

Drews, A., & Westerdahl, H. (2019). Not all birds have a single dominantly expressed MHC-I gene: Transcription suggests that siskins have many highly expressed MHC-I genes. *Scientific Reports*, *9*(1).

Duan, Z., Han, Y., Zhou, L., Yuan, C., Wang, Y., Zhao, C., Tang, H., & Chen, J. (2020). Chicken bromodomain-containing protein 2 interacts with the Newcastle disease virus matrix protein and promotes viral replication. *Veterinary Research*, *51*(1).

Dvir, H., Wang, J., Ly, N., Dascher, C. C., & Zajonc, D. M. (2010). Structural basis for lipid-antigen recognition in avian immunity. *The Journal of Immunology*, *184*(5), 2504–2511.

Ekblom, R., Stapley, J., Ball, A. D., Birkhead, T., Burke, T., & Slate, J. (2011). Genetic mapping of the major histocompatibility complex in the zebra finch (Taeniopygia guttata). *Immunogenetics*, *63*(8), 523–530.

Elsheimer-Matulova, M., Polansky, O., Seidlerova, Z., Varmuzova, K., Stepanova, H., Fedr, R., & Rychlik, I. (2020). Interleukin 4 inducible 1 gene (IL4I1) is induced in chicken phagocytes by Salmonella Enteritidis infection. *Veterinary Research*, *51*(1).

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., Wright, D., Jungerius, A., Vereijken, A., Randi, E., Jensen, P., & Andersson, L. (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics*, *4*(2), e1000010.

Evseeva, I., Nicodemus, K. K., Bonilla, C., Tonks, S., & Bodmer, W. F. (2010). Linkage disequilibrium and age of HLA region SNPs in relation to classic HLA gene alleles within Europe. *European Journal of Human Genetics*, *18*(8), 924–932.

Ewald, S. J., & Livant, E. J. (2004). Distinctive polymorphism of chicken B-FI (major histocompatibility complex class I) molecules. *Poultry Science*, *83*(4), 600–605.

Fling, S. P., Arp, B., & Pious, D. (1994). HLA-DMA and -DMB genes are both required for MHC class II/peptide complex formation in antigen-presenting cells. *Nature*, *368*(6471), 554–558.

Fulton, J. E., Berres, M. E., Kantanen, J., & Honkatukia, M. (2017). MHC-B variability within the Finnish Landrace chicken conservation program. *Poultry Science*, *96*(9), 3026–3030.

Fulton, J. E., Juul-Madsen, H. R., Ashwell, C. M., McCarron, A. M., Arthur, J. A., O'Sullivan, N. P., & Taylor, R. L. (2006). Molecular genotype identification of the Gallus gallus major histocompatibility complex. *Immunogenetics*, *58*(5-6), 407–421.

Fulton, J. E., McCarron, A. M., Lund, A. R., Pinegar, K. N., Wolc, A., Chazara, O., Bed'Hom, B., Berres, M., & Miller, M. M. (2016a). A high-density SNP panel reveals extensive diversity, frequent recombination and multiple recombination hotspots within the chicken major histocompatibility complex B region between BG2 and CD1A1. *Genetics Selection Evolution*, *48*(1).

Fulton, J. E., Lund, A. R., McCarron, A. M., Pinegar, K. N., Korver, D. R., Classen, H. L., Aggrey, S., Utterbach, C., Anthony, N. B., & Berres, M. E. (2016b). MHC variability in heritage breeds of chickens. *Poultry Science*, *95*(2), 393–399.

Gorer, P. A. (1936). The detection of a hereditary antigenic difference in the blood of mice by means of human group a serum. *Journal of Genetics*, *32*(1), 17–31.

Gorer, P. A., Lyman, S., & Snell, G. D. (1948). Studies on the genetic and antigenic basis of tumour transplantation. Linkage between a histocompatibility gene and "fused" in mice. *Proceedings of the Royal Society of London. Series B - Biological Sciences*, *135*(881), 499–505.

Goto, R., Afanassieff, M., Ha, J., Iglesias, G., Ewald, S., Briles, W., & Miller, M. (2002). Single-strand conformation polymorphism (SSCP) assays for major histocompatibility complex B genotyping in chickens. *Poultry Science*, *81*(12), 1832–1841.

Goto, R. M., Warden, C. D., Shiina, T., Hosomichi, K., Zhang, J., Kang, T. H., Wu, X., Glass, M. C., Delany, M. E., & Miller, M. M. (2022). The Gallus gallus RJF reference genome reveals an MHCY haplotype organized in gene blocks that contain 107 loci including 45 specialized, polymorphic MHC class I loci, 41 C-type lectin-like loci, and other loci amid hundreds of transposable elements. *G3 Genes|Genomes|Genetics*, *12*(11).

Granevitze, Z., Hillel, J., Chen, G. H., Cuc, N. T. K., Feldman, M., Eding, H., & Weigend, S. (2007). Genetic diversity within chicken populations from different continents and management histories. *Animal Genetics*, *38*(6), 576–583.

Haider, S., Islam, B., D'Atri, V., Sgobba, M., Poojari, C., Sun, L., Yuen, T., Zaidi, M., & New, M. I. (2013). Structure–phenotype correlations of human CYP21A2 mutations in congenital adrenal hyperplasia. *Proceedings of the National Academy of Sciences*, *110*(7), 2605–2610.

Han, B., Lian, L., Qu, L., Zheng, J., & Yang, N. (2013). Abundant polymorphisms at the microsatellite locus LEI0258 in indigenous chickens. *Poultry Science*, *92*(12), 3113–3119.

Hedrick, P. W. (1998). Balancing selection and MHC. *Genetica*, *104*(3), 207–214.

Hess, C. M., & Edwards, S. V. (2002). The Evolution of the major histocompatibility complex in birds. *BioScience*, *52*(5), 423.

Hosomichi, K., Miller, M. M., Goto, R. M., Wang, Y., Suzuki, S., Kulski, J. K., Nishibori, M., Inoko, H., Hanzawa, K., & Shiina, T. (2008). Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. *The Journal of Immunology*, *181*(5), 3393–3399.

Hunt, H. D., & Fulton, J. E. (1998). Analysis of polymorphisms in the major expressed class I locus ( B-F IV) of the chicken. *Immunogenetics*, *47*(6), 456–467.

Hunt, H. D., Goto, R. M., Foster, D. N., Bacon, L. D., & Miller, M. M. (2006). At least one YMHCI molecule in the chicken is alloimmunogenic and dynamically expressed on spleen cells during development. *Immunogenetics*, *58*(4), 297–307.

Hviid, T. V. F., & Christiansen, O. B. (2005). Linkage disequilibrium between human leukocyte antigen (HLA) Class II and HLA-G—possible implications for human reproduction and autoimmune disease. *Human Immunology*, *66*(6), 688–699.

Iglesias, G. M., Beker, M. P., Remolins, J. S., Canet, Z. E., Librera, J., Cantaro, H., Maizon, D. O., & Fulton, J. E. (2021). MHC-B variation in maternal and paternal synthetic lines of the Argentinian Campero INTA chicken. *Poultry Science*, *100*(8), 101253.

Iglesias, G. M., Canet, Z. E., Cantaro, H., Miquel, M. C., Melo, J. E., Miller, M. M., Berres, M. E., & Fulton, J. E. (2019). Mhc-B haplotypes in "Campero-Inta" chicken synthetic line. *Poultry Science*, *98*(11), 5281–5286.

Jacob, J. P., Milne, S., Beck, S., & Kaufman, J. (2000). The major and a minor class II β-chain (B-LB) gene flank the Tapasin gene in the B-F/B-L region of the chicken major histocompatibility complex. *Immunogenetics*, *51*(2), 138–147.

Jia, X., Zhao, C., & Zhao, W. (2021). Emerging roles of MHC class I region-encoded E3 ubiquitin ligases in innate immunity. *Frontiers in Immunology*, *12*.

Joiner, K. S., Hoerr, F. J., van Santen, E., & Ewald, S. J. (2005). The avian major histocompatibility complex influences bacterial skeletal disease in broiler breeder chickens. *Veterinary Pathology*, *42*(3), 275–281.

Joly, E., Le Rolle, A-F., Gonzélez, A. L., Mehling, B., Stevens, J., Coadwell, W. J., Hünig, T., Howard, J. C., & Butcher, G. W. (1998). Co-evolution of rat TAP transporters and MHC class I RT1-A molecules. *Current Biology*, *8*(3), 169–180.

Kasai, K., Ishikawa, T., Nakamura, T., & Miura, T. (2015). Antibacterial properties of l-amino acid oxidase: mechanisms of action and perspectives for therapeutic applications. *Applied Microbiology and Biotechnology*, *99*(19), 7847–7857.

Kaufman, J. (2014). The avian MHC. *Avian Immunology*, 149–167.

Kaufman, J. (2015). Co-evolution with chicken class I genes. *Immunological Reviews, 267*(1), 56–71.

Kaufman, J. (2018). Generalists and specialists: a new view of how MHC class I molecules fight infectious pathogens. *Trends in Immunology, 39*(5), 367–379.

Kaufman, J. (2022). The avian major histocompatibility complex. *Avian Immunology*, 135–161.

Kaufman, J., & Venugopal, K. (1998). The importance of MHC for Rous sarcoma virus and Marek's disease virus—Some Payne-ful considerations. *Avian Pathology*, *27*(sup1), S82–S87.

Kaufman, J., Milne, S., Göbel, T. W. F., Walker, B. A., Jacob, J. P., Auffray, C., Zoorob, R., & Beck, S. (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, *401*(6756), 923–925.

Kim, T., Hunt, H. D., Parcells, M. S., van Santen, V., & Ewald, S. J. (2018). Two class I genes of the chicken MHC have different functions: BF1 is recognized by NK cells while BF2 is recognized by CTLs. *Immunogenetics*, *70*(9), 599–611.

Kulski, J. K., & Inoko, H. (2006). Major histocompatibility complex (MHC) Genes. *ELS*.

Kumanovics, A., & Lindahl, K. (2004). Good copy, bad copy: choosing animal models for HLA-linked diseases. *Current Opinion in Genetics & Development*, *14*(3), 258–263.

Kumánovics, A., Takada, T., & Lindahl, K. F. (2003). Genomic organization of the mammalian *Mhc*. *Annual Review of Immunology*, *21*(1), 629–657.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Lin, W., Zhang, Z., Xu, Z., Wang, B., Li, X., Cao, H., Wang, Y., & Zheng, S. J. (2015). The association of receptor of activated protein kinase C 1(RACK1) with infectious bursal disease virus viral protein VP5 and voltage-dependent anion channel 2 (VDAC2) inhibits apoptosis and enhances viral replication. *Journal of Biological Chemistry*, *290*(13), 8500–8510.

Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., & Nielsen, M. (2008). NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Research*, *36*(Web Server issue), W509-512.

Mahdi, B. M. (2013). A glow of HLA typing in organ transplantation. *Clinical and Translational Medicine*, *2*(1), 6.

Malinowska, M., Tokarz-Deptuła, B., & Deptuła, W. (2017). Butyrophilins: an important new element of resistance. *Central European Journal of Immunology*, *42*(4), 399–403.

Malomane, D. K., Simianer, H., Weigend, A., Reimer, C., Schmitt, A. O., & Weigend, S. (2019). The SYNBREED chicken diversity panel: a global resource to assess chicken diversity at high genomic resolution. *BMC Genomics*, *20*(1).

Malomane, D. K., Weigend, S., Schmitt, A. O., Weigend, A., Reimer, C., & Simianer, H. (2021). Genetic diversity in global chicken breeds in relation to their genetic distances to wild populations. *Genetics Selection Evolution*, *53*(1).

Manjula, P., Fulton, J. E., Seo, D., & Lee, J. H. (2020). Major histocompatibility complex B variability in Korean native chicken breeds. *Poultry Science*, *99*(10), 4704–4713.

Manjula, P., Fulton, J. E., Seo, D., & Lee, J. H. (2021). Comparison of major histocompatibility complex- *B* variability in Sri Lankan indigenous chickens with five global chicken populations using MHC- *B* SNP panel. *Animal Genetics*, *52*(6), 824–833.

Manzila, I., Lestari, P., Sartika, T., Priyatno, T. P., Indriani, R., Nugroho, K., & Terryana, R. T. (2022). Polymorphisms and associations of the RACK1 genes with antibody response to Newcastle disease in KUB chickens. *THE SECOND INTERNATIONAL CONFERENCE on GENETIC RESOURCES and BIOTECHNOLOGY: Harnessing Technology for Conservation and Sustainable Use of Genetic Resources for Food and Agriculture*. https://doi.org/10.1063/5.0075622

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 122.

Mesa, C., Thulien, K., Moon, D., Veniamin, S., & Magor, K. (2004). The dominant MHC class I gene is adjacent to the polymorphic TAP2 gene in the duck, Anas platyrhynchos. *Immunogenetics*, *56*(3). https://doi.org/10.1007/s00251-004-0672-3

Miller, M. M., Abplanalp, H., & Goto, R. (1988). Genotyping chickens for the B-G subregion of the major histocompatibility complex using restriction fragment length polymorphisms. *Immunogenetics*, *28*(5), 374–379.

Miller, M. M., & Taylor, R. L. (2016). Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. *Poultry Science*, *95*(2), 375–392.

Miller, M. M., Goto, R., Bernot, A., Zoorob, R., Auffray, C., Bumstead, N., & Briles, W. E. (1994). Two Mhc class I and two Mhc class II genes map to the chicken Rfp-Y system outside the B complex. *Proceedings of the National Academy of Sciences*, *91*(10), 4397–4401.

Miller, M. M., Goto, R. M., Taylor, R. L., Zoorob, R., Auffray, C., Briles, R. W., Briles, W. E., & Bloom, S. E. (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. *Proceedings of the National Academy of Sciences*, *93*(9), 3958–3962.

Miller, M. M., Goto, R., Young, S., Chirivella, J., Hawke, D., & Miyada, C. G. (1991). Immunoglobulin variable-region-like domains of diverse sequence within the major

histocompatibility complex of the chicken. *Proceedings of the National Academy of Sciences*, *88*(10), 4377–4381.

Miller, M. M., Robinson, C. M., Abernathy, J., Goto, R. M., Hamilton, M. K., Zhou, H., & Delany, M. E. (2013). Mapping genes to chicken microchromosome 16 and discovery of olfactory and scavenger receptor genes near the major histocompatibility complex. *Journal of Heredity*, *105*(2), 203–215.

Miller, M. M., Wang, C., Parisini, E., Coletta, R. D., Goto, R. M., Lee, S. Y., Barral, D. C., Townes, M., Roura-Mir, C., Ford, H. L., Brenner, M. B., & Dascher, C. C. (2005). Characterization of two avian MHC-like genes reveals an ancient origin of the CD1 family. *Proceedings of the National Academy of Sciences*, *102*(24), 8674–8679.

Miller, W. L. (2021). Tenascin-X—discovery and early research. *Frontiers in Immunology*, *11*.

Mocikat, R., Braumüller, H., Gumy, A., Egeter, O., Ziegler, H., Reusch, U., Bubeck, A., Louis, J., Mailhammer, R., Riethmüller, G., Koszinowski, U., & Röcken, M. (2003). Natural killer cells activated by MHC class I low targets prime dendritic cells to induce protective CD8 T cell responses. *Immunity*, *19*(4), 561–569.

Molee, A., Kongroi, K., Kuadsantia, P., Poompramun, C., & Likitdecharote, B. (2015). Association between single nucleotide polymorphisms of the major histocompatibility complex class II gene and newcastle disease virus titre and body weight in leung hang khao chickens. *Asian-Australasian Journal of Animal Sciences*, *29*(1), 29–35.

Momburg, F., Roelse, J., Howard, J. C., Butcher, G. W., Hämmerling, G. J., & Neefjes, J. J. (1994). Selectivity of MHC-encoded peptide transporters from human, mouse and rat. *Nature*, *367*(6464), 648–651.

Monson, M. S., Mendoza, K. M., Velleman, S. G., Strasburg, G. M., & Reed, K. M. (2013). Expression profiles for genes in the turkey major histocompatibility complex B-locus. *Poultry Science*, *92*(6), 1523–1534.

Moon, D. A., Veniamin, S. M., Parks-Dely, J. A., & Magor, K. E. (2005). The MHC of the duck (Anas platyrhynchos) contains five differentially expressed class I genes. *The Journal of Immunology*, *175*(10), 6702–6712.

Muir, W. M., Wong, G. K.-S., Zhang, Y., Wang, J., Groenen, M. A. M., Crooijmans, R. P. M. A., Megens, H.-J., Zhang, H., Okimoto, R., Vereijken, A., Jungerius, A., Albers, G. A. A., Lawley, C. T., Delany, M. E., MacEachern, S., & Cheng, H. H. (2008). Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *Proceedings of the National Academy of Sciences*, *105*(45), 17312–17317.

Nakamura, T., Shirouzu, T., Nakata, K., Yoshimura, N., & Ushigome, H. (2019). The Role of Major histocompatibility complex in organ transplantation- donor specific anti-major histocompatibility complex antibodies analysis goes to the next stage -. *International Journal of Molecular Sciences*, *20*(18), 4544.

National Human Genome Research Institute. (2019). *Haplotype*. Genome.gov. https://www.genome.gov/genetics-glossary/haplotype

Nguyen-Phuc, H., Fulton, J. E., & Berres, M. E. (2016). Genetic variation of major histocompatibility complex (MHC) in wild Red Junglefowl (Gallus gallus). *Poultry Science*, *95*(2), 400–411.

O'Connor, E. A., Strandh, M., Hasselquist, D., Nilsson, J.-Å., & Westerdahl, H. (2016). The evolution of highly variable immunity genes across a passerine bird radiation. *Molecular Ecology*, *25*(4), 977–989.

O'Neill, A. M., Livant, E. J., & Ewald, S. J. (2009). The chicken BF1 (classical MHC class I) gene shows evidence of selection for diversity in expression and in promoter and signal peptide regions. *Immunogenetics*, *61*(4), 289–302.

Owusu-Apenten, R., & Vieira, E. R. (2022). *Elementary Food Science*. Springer Nature.

Parker, A., & Kaufman, J. (2017). What chickens might tell us about the MHC class II system. *Current Opinion in Immunology*, *46*, 23–29.

Pharr, G. T., Gwynn, A. V., & Bacon, L. D. (1996). Histocompatibility antigen(s) linked to Rfp-Y (Mhc-like) genes in the chicken. *Immunogenetics*, *45*(1), 52–58.

Qanbari, S., Rubin, C.-J., Maqbool, K., Weigend, S., Weigend, A., Geibel, J., Kerje, S., Wurmser, C., Peterson, A. T., Brisbin, I. L., Preisinger, R., Fries, R., Simianer, H., & Andersson, L. (2019). Genetics of adaptation in modern chicken. *PLOS Genetics*, *15*(4), e1007989.

Razin, S. V., Borunova, V. V., Maksimenko, O. G., & Kantidze, O. L. (2012). Cys2His2 zinc finger protein family: classification, functions, and major members. *Biochemistry. Biokhimiia*, *77*(3), 217–226.

Ren, J., Yang, L., Li, Q., Zhang, Q., Sun, C., Liu, X., & Yang, N. (2019). Global investigation of cytochrome P450 genes in the chicken genome. *Genes*, *10*(8), 617.

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., & Haase, B. (2021).

Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*(7856), 737–746.

Rogers, S. L., & 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*(8), 461–475.

Rogers, S. L., & Kaufman, J. (2016). Location, location, location: the evolutionary history of CD1 genes and the NKR-P1/ligand systems. *Immunogenetics*, *68*(8), 499–513.

Rogers, S. L., Göbel, T. W., Viertlboeck, B. C., Milne, S., Beck, S., & Kaufman, J. (2005). Characterization of the chicken C-type lectin-like receptors B-NK and B-lec suggests that the NK complex and the MHC share a common ancestral region. *The Journal of Immunology*, *174*(6), 3475–3483.

Rogers, S., Shaw, I., Ross, N., Nair, V., Rothwell, L., Kaufman, J., & Kaiser, P. (2003). Analysis of part of the chicken Rfp-Y region reveals two novel lectin genes, the first complete genomic sequence of a class I α-chain gene, a truncated class II β-chain gene, and a large CR1 repeat. *Immunogenetics*, *55*(2), 100–108.

Ruby, T., Bed'Hom, B., Wittzell, H., Morin, V., Oudin, A., & Zoorob, R. (2005). Characterisation of a cluster of TRIM-B30.2 genes in the chicken MHC B locus. *Immunogenetics*, *57*(1-2), 116–128.

Salomonsen, J., Chattaway, J. A., Chan, A. C. Y., Parker, A., Huguet, S., Marston, D. A., Rogers, S. L., Wu, Z., Smith, A. L., Staines, K., Butter, C., Riegert, P., Vainio, O., Nielsen, L., Kaspers, B., Griffin, D. K., Yang, F., Zoorob, R., Guillemot, F., & Auffray, C. (2014). Sequence of a complete chicken BG haplotype shows dynamic expansion and contraction of two gene lineages with particular expression patterns. *PLoS Genetics*, *10*(6), e1004417.

Salomonsen, J., Dunon, D., Skjødt, 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. *Proceedings of the National Academy of Sciences*, *88*(4), 1359–1363.

Salomonsen, J., Sørensen, M. R., Marston, D. A., Rogers, S. L., Collen, T., van Hateren, A., Smith, A. L., Beal, R. K., Skjødt, K., & Kaufman, J. (2005). Two CD1 genes map to the chicken MHC, indicating that CD1 genes are ancient and likely to have been present in the primordial MHC. *Proceedings of the National Academy of Sciences*, *102*(24), 8668–8673.

Schierman, L. W., & Nordskog, A. W. (1961). Relationship of blood type to histocompatibility in chickens. *Science*, *134*(3484), 1008–1009.

Semagn, K., Beyene, Y., Warburton, M. L., Tarekegne, A., Mugo, S., Meisel, B., Sehabiague, P., & Prasanna, B. M. (2013). Meta-analyses of QTL for grain yield and anthesis silking interval in 18 maize populations evaluated under water-stressed and well-watered environments. *BMC Genomics*, *14*(1), 313.

Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. (2004). *Nature*, *432*(7018), 695–716.

Shaw, I., Powell, T. J., Marston, D. A., Baker, K., van Hateren, A., Riegert, P., Wiles, M. V., Milne, S., Beck, S., & Kaufman, J. (2007). Different evolutionary histories of the two classical class I genes *BF1* and *BF2* illustrate drift and selection within the stable MHC haplotypes of chickens. *The Journal of Immunology*, *178*(9), 5744–5752.

Shiina, T., Blancher, A., Inoko, H., & Kulski, J. K. (2016). Comparative genomics of the human, macaque and mouse major histocompatibility complex. *Immunology*, *150*(2), 127–138.

Shiina, T., Briles, W. E., Goto, R. M., Hosomichi, K., Yanagiya, K., Shimizu, S., Inoko, H., & Miller, M. M. (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. *The Journal of Immunology*, *178*(11), 7162–7172.

Shiina, T., Shimizu, S., Hosomichi, K., Kohara, S., Watanabe, S., Hanzawa, K., Beck, S., Kulski, J. K., & Inoko, H. (2004). Comparative genomic analysis of two avian (quail and chicken) MHC regions. *The Journal of Immunology*, *172*(11), 6751–6763.

Simmonds, M., & Gough, S. (2007). The HLA region and autoimmune disease: associations and mechanisms of action. *Current Genomics*, *8*(7), 453–465.

Srinivasan, M., Domanico, S. Z., Kaumaya, P. T. P., & Pierce, S. K. (1993). Peptides of 23 residues or greater are required to stimulate a high affinity class II-restricted T cell response. *European Journal of Immunology*, *23*(5), 1011–1016.

Tan, P., Kropshofer, H., Mandelboim, O., Bulbuc, N., Hämmerling, G. J., & Momburg, F. (2002). Recruitment of MHC class I molecules by tapasin into the transporter associated with antigen processing-associated complex is essential for optimal peptide loading. *The Journal of Immunology*, *168*(4), 1950–1960.

Tarrant, K. J., Lopez, R., Loper, M., & Fulton, J. E. (2020). Assessing MHC-B diversity in Silkie chickens. *Poultry Science*, *99*(5), 2337–2341.

van Hateren, A., Carter, R., Bailey, A., Kontouli, N., Williams, A. P., Kaufman, J., & Elliott, T. (2013). A mechanistic basis for the co-evolution of chicken tapasin and major

histocompatibility complex class I (MHC I) proteins. *Journal of Biological Chemistry, 288*(45), 32797–32808.

Viertlboeck, B. C., Wortmann, A., Schmitt, R., Plachý, J., & Göbel, T. W. (2008). Chicken C-type lectin-like receptor B-NK, expressed on NK and T cell subsets, binds to a ligand on activated splenocytes. *Molecular Immunology*, *45*(5), 1398–1404.

Walker, B. A., Hunt, L. G., Sowa, A. K., Skjødt, K., Göbel, T. W., Lehner, P. J., & Kaufman, J. (2011). The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *Proceedings of the National Academy of Sciences, 108*(20), 8396–8401.

Walker, B. A., van Hateren, A., Milne, S., Beck, S., & Kaufman, J. (2005). Chicken TAP genes differ from their human orthologues in locus organisation, size, sequence features and polymorphism. *Immunogenetics*, *57*(3-4), 232–247.

Wang, H., & Liu, M. (2021). Complement C4, infections, and autoimmune diseases. *Frontiers in Immunology*, *12*. 694928.

Wang, M.-S., Thakur, M., Peng, M.-S., Jiang, Y., Frantz, L. A. F., Li, M., Zhang, J.-J., Wang, S., Peters, J., Otecko, N. O., Suwannapoom, C., Guo, X., Zheng, Z.-Q., Esmailizadeh, A., Hirimuthugoda, N. Y., Ashari, H., Suladari, S., Zein, M. S. A., Kusza, S., & Sohrabi, S. (2020). 863 genomes reveal the origin and domestication of chicken. *Cell Research, 30*(8), 693–701.

Ye, X., Zhu, J., Velleman, S., Bacon, W., & Nestor, K. (1999). Analysis of genetic polymorphisms in the major histocompatibility complex of Japanese quail. *Poultry Science, 78*(1), 8–11.

Yuan, Y., Zhang, H., Yi, G., You, Z., Zhao, C., Yuan, H., Wang, K., Li, J., Yang, N., & Lian, L. (2021). Genetic diversity of MHC B-F/B-L region in 21 chicken populations. *Frontiers in Genetics*, *12*. 710770.

Zajonc, D. M., Striegl, H., Dascher, C. C., & Wilson, I. A. (2008). The crystal structure of avian CD1 reveals a smaller, more primordial antigen-binding pocket compared to mammalian CD1. *Proceedings of the National Academy of Sciences*, *105*(46), 17925–17930.

Zakharova, M. Y., Ю, З. М., Belyanina, T. A., А, Б. Т., Sokolov, A. V., В, С. А., Kiselev, I. S., С, К. И., Mamedov, A. E., & Э, М. А. (2019). The contribution of major histocompatibility complex class II genes to an association with autoimmune diseases. *Acta Naturae*, *11*(4), 4–12.

Zhang, J., Chen, Y., Qi, J., Gao, F., Liu, Y., Liu, J., Zhou, X., Kaufman, J., Xia, C., & Gao, G. F. (2012). Narrow groove and restricted anchors of MHC class I molecule BF2*0401

plus peptide transporter restriction can explain disease susceptibility of B4 chickens. *The Journal of Immunology*, *189*(9), 4478–4487.

Zhang, J., Goto, R. M., & Miller, M. M. (2020). A simple means for MHC-Y genotyping in chickens using short tandem repeat sequences. *Immunogenetics*, *72*(5), 325–332.

Zhou, L., Han, Y. F., Yuan, C., & Duan, Z. Q. (2021). Screening and bioinformatics analysis of cellular proteins interacting with chicken bromodomain-containing protein 2 in DF-1 cells. *British Poultry Science*, *62*(6), 810–819.