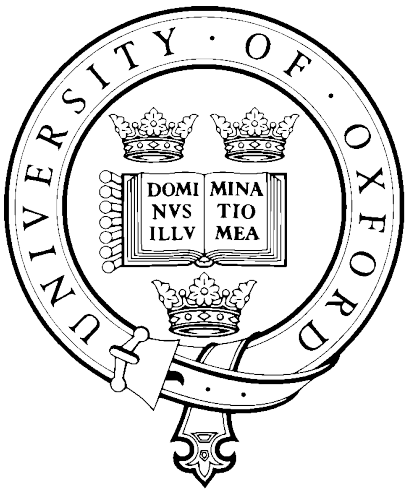
The secret in their *Mhc*:
Variation and selection in a free-living population of great tits



Irem Sepil

Jesus College

Thesis submitted for the degree of

Doctor of Philosophy

University of Oxford

Trinity Term 2012

Abstract

The secret in their *Mhc*: Variation and selection in a free-living population of great tits

DPhil thesis, submitted by Irem Sepil, Jesus College, Trinity Term 2012

Understanding the genetic basis of fitness differences has been a major goal for evolutionary biologists over the last two decades. Although there are many studies investigating how natural selection can promote local adaptation, few have succeeded to find the link between genotype and fitness of the phenotype. Polymorphic genes of the major histocompatibility complex (*Mhc*) are excellent candidates for such associations as they are a central component of the vertebrate immune system, playing an important role in parasite resistance, and hence can have direct effects on survival of their bearers. Although associations between *Mhc* and disease resistance are frequently documented, the epidemiological basis of the host-parasite interaction is often lacking and few studies have investigated the role that *Mhc* genes play in individual variation in fitness; thus comparatively little is known about the fitness consequences of *Mhc* in wild populations. Furthermore, the majority of work to date has involved testing associations between *Mhc* genotypes and disease. However, the mechanism by which any direct selection on the *Mhc* acts, depends on how genotypes map to the functional properties of *Mhc* molecules.

The aim of this thesis was to characterize *Mhc* alleles in terms of their predicted functional properties and to investigate whether and how selection operates on *Mhc* class I functional variation using the great tit (*Parus major*) population at Wytham Woods as a model host species. Through a comprehensive characterization effort and the use of 454 pyrosequencing platform, I performed a detailed analysis of genetic variation at *Mhc* class I exon 3 and grouped alleles with similar antigen-binding affinities into supertypes to classify functionally distinct *Mhc* types. There was extreme complexity at the *Mhc* class I of the great tit both in terms of allelic diversity and gene number. A total of 862 alleles were detected from 857 individuals; the highest number yet characterized in a wild bird species. The functional alleles were clustered into 17 supertypes; there was clear evidence that functional alleles were under strong balancing selection.

To understand the role of *Mhc* in disease resistance, I examined the linkage between *Mhc* supertypes, *Plasmodium* infection and great tit survival, and showed that certain functional variants of *Mhc* confer resistance to two divergent *Plasmodium* parasite species that are common in the environment. I further investigated the fitness consequences of functional variation at *Mhc*, using mark-recapture methods and long-term breeding data; and tested the hypotheses that selection: (i) maximizes *Mhc* diversity; (ii) optimizes *Mhc* diversity, or (iii) favours specific functional variants. I found that the presence of three different supertypes was associated with three different components of individual fitness: adult survival, annual recruitment probabilities and lifetime reproductive success. In contrast, there was no evidence for a selective advantage of *Mhc* functional diversity, either in terms of maximal or optimal supertype diversity. Finally, I explored the role that *Mhc* plays in female mate choice decisions and examined the reproductive fitness consequences of *Mhc*-dependent mating patterns. There was little evidence to suggest that functional dissimilarity at *Mhc* has any influence on female mate choice decisions or that dissimilarity at *Mhc* affects the reproductive output of the social pair.

Overall, this thesis provides strong support for the suggestion that selection favours specific functional variants of *Mhc,* possibly as a result of supertype-specific resistance or susceptibility to parasites that exert strong selective pressures on their hosts; whereas there is no support for selection favouring maximal or optimal *Mhc* diversity. More importantly it demonstrates that functional variants of *Mhc* class I loci are an important determinant of individual fitness in natural populations.

Acknowledgements

First and foremost, I sincerely would like to thank my supervisor Ben Sheldon. It has been the most challenging and rewarding five years of my life and I couldn't have done it without your guidance and brilliance. My DPhil had an unusual start, where I had to disappear for a year even before I started, and had to change my subject following my return. From that time on you have been extremely supportive and understanding; I deeply appreciate the freedom you gave me and that you let me follow my instincts regarding my DPhil subject. Your intellectual input to this work and to my development as a scientist has been invaluable. I am also grateful for your reliable calmness, optimism and clarity of thought, especially during these last few months, even when I had doubts about meeting my strict thesis deadline. It has been a privilege to be your student and I couldn't have wished for a better supervisor.

I am indebted to many members of the EGI for contributing to my research, education and happiness over the past few years. Thanks to Shelly Lachish, who ended up being my unofficial co-supervisor and good friend. I couldn't have completed this thesis without your intellectual input and emotional support. I feel very lucky that you came to Oxford. Thanks to Hooman Moghadam for patiently producing bioinformatic scripts so that I could deal with the massive 454 data, without which any analyses in this thesis would be possible. Thanks to Reinder Radersma for contributions to the work presented here and for coping with my endless questions over the last few weeks. Thanks to Andy Gosler and Matt Wood who patiently trained me in bird ringing, a skill that I will value always. Thanks to all those who helped in the fieldwork that went into this thesis – to Amy Hinks, Ella Cole and Ada Grabowska-Zhang, the whole great tit team for their much needed support during those ridiculously early mornings and long days. To Tobias Uller, Olof Hellgren, Simon Evans and Jo Chapman for happily answering many a naïve question about statistics and other things. I owe a great deal to all members of the EGI for creating such an enjoyable, supportive and intellectually stimulating environment to work in - thank you all very much.

I would also like to sincerely thank my co-authors outside the EGI. Special thanks to Elise Huchard for hosting me in Göttingen and sharing her expertise on *Mhc*, and for the continued support over the past year. Also to Anna Santure for helping me with SNP analysis.

Thanks to all my friends in Oxford who have made living here such a pleasure. Many of you have already left Oxford but I am confident that our friendship can only grow stronger despite the distances. I would like to specially thank Becky Dean, Claire Salisbury, Sandra Bouwhuis, Marta Szulkin, Katia Schörle, Campbell Allen and Tom Gheysens for their friendship and support. To Hanne Lovlie, thanks for brightening up many days in Oxford either with your presence or with the postcards and chocolates you sent from Sweden. To Julie Collet, thanks for being such a cheerful and fun friend and for being a part of many fond memories from Glastonbury to Paris. To Caroline Isaksson for being my lovely artsy partner from West End theaters to Camden jazz dinners and for sharing many girlie talk, Abba songs and tipsy giggles. To Sarah Knowles for being a great support throughout my PhD, and for continuing to encourage, and amuse me. To Eric Trottier and Julie Morand-Ferron, it has been a blast knowing you; recording and singing with you are memories I will never forget. To Nicole Milligan and Gökçe Pulcu, thanks for sharing all the highs and lows of PhD life and life in general, you have both been great support in the last two years.

A massive thanks to Camille Bonneaud, who took me under her wing during my first encounter with evolutionary biology. You are the reason I ended up in Oxford and I will always be grateful. Several people deserve thanks for being fantastic friends over the last two decades and encouraging me throughout my studies even from far distances. Naz, Aslı, Elif, Selma, Sena, Ebru, Sinan, Berna, Talya, Claudia, Ilker and Fero you have always been there for me one way or another, thanks for being an important part of my life. My family: Müjdat, Tansu, Ayşegül, Korhan and Bora you have always had faith in me and your support throughout this PhD has been invaluable.

I thank my mom, Canan, and dad, Mehmet, for their love and support that has enabled me to start and finish this PhD. Mom, you have been my role model for all my life and your enthusiasm in genetics led me to take this path. I hope I can be half of the mother you are to me to my soon to born daughter. Dad, you are a constant source of support and inspiration for me and to see how proud and happy you are when you talk about my work and Darwin's Galapagos expedition made me believe in myself and in what I do. I love you both.

Finally, I dedicate this thesis to my husband Kaan and daughter Sera. Kaan, you are the best thing that ever happened to me in my life and I owe you every success, every joy. You have helped me with this thesis in every possible way throughout these four years and I see it as a product of our hard work, not just mine. I thank you and love you with all my heart. I am looking forward to the new journey that awaits us with our baby girl, Sera.

Contents

Chapter 1

Introduction

Introduction

Detecting the molecular signature of natural selection has become a major interest for evolutionary biologists over the last decades (Ellegren & Sheldon 2008). Numerous studies have been conducted to identify the genetic locus of selection, and evidence suggests that a high percentage of genes have been subject to positive selection (reviewed in Ford 2002; Ellegren 2008). However, many of these studies suffer from the lack of knowledge on the genes' function, and solely base their evidence on statistical tests that fail to rule out alternative interpretations, leaving conceptual flaws (Hughes 2008). This limitation also eliminates the possibility to formulate and test hypothesis explaining why the gene is under selection; thus the actual source of selection remains unknown (Ford 2002). Therefore our understanding of how natural selection can promote adaptation and its genetic basis is still based on very few genes and further work on these candidate genes has the potential to give new insight on the causes and consequences of selection. For example more than 150 genes that affect mammalian pigmentation have been characterized and this data is extensively being used to determine how selection shape phenotypic and genetic variation (Hubbard *et al.* 2010). Likewise Pitx1, the major locus controlling pelvic reduction in sticklebacks, has been used to study parallel evolution across different populations, species and genera (Shapiro *et al.* 2006).

The major histocompatibility complex (*Mhc*) is a multigene family that has been studied intensely since the late 1980s with its structure and function having been characterised in detail. This attention is mainly due to the medical importance of this gene family in humans for understanding the biological basis of tissue transplantation, autoimmune diseases and pathogen susceptibility (Hedrick 1994). *Mhc* was first defined in the 1940's by mouse geneticist and transplant immunologist George Davis Snell, who shared the 1980 Nobel Prize in Physiology or Medicine "for their discoveries concerning genetically determined structures on the cell surface that regulate immunological reactions". Those aspects of *Mhc* genes that interest immunologists also intrigue evolutionary biologists and for the last two decades *Mhc* has

frequently been used as a model system for studying adaptively important genetic diversity (Hess & Edwards 2002).

The *Mhc* ought to be an excellent candidate for selection studies as it is the central component of the vertebrate adaptive immune system, playing an important role in parasite resistance, and hence with the potential to have direct effects on the survival of its bearers (Trowsdale & Parham 2004). Moreover, *Mhc* genes are known to be the most polymorphic gene group in vertebrates and it is believed that this diversity is maintained by selection from parasites (parasite-mediated selection) and through mate choice (sexual selection) (Doherty & Zinkernagel 1975; Yamazaki *et al.* 1976; Apanius *et al.* 1997). Interest in understanding *Mhc* genes in non-model vertebrates is growing significantly as more studies demonstrate implications of their diversity in the context of parasite resistance and susceptibility (Briles *et al.* 1977; Hill *et al.* 1991; Penn *et al.* 2002; Oliver *et al.* 2009; Loiseau *et al.* 2011), survival (Paterson *et al.* 1998; Langefors *et al.* 2001; Arkush *et al.* 2002; Brouwer *et al.* 2010; Worley *et al.* 2010), mate choice (Potts *et al.* 1994; Wedekind *et al.* 1995; Schwensow *et al.* 2008; Eizaguirre *et al.* 2009b), conservation (Hughes 1991; van Oosterhout *et al.* 2007; Siddle *et al.* 2010; Agudo *et al.* 2012) and even speciation (Eizaguirre *et al.* 2009a). Below I summarize the structure and function of *Mhc* genes, the most relevant features of *Mhc* that are the focus of this research, the study population I used and finally the central questions addressed by this work.

Function and structure of *Mhc* genes

Mhc genes encode glycoproteins that deliver foreign and self-peptides to the cell surface to enable self and non-self identification by T-cells (Klein 1986). Foreign antigens (parasites) enter cells either by infection or through phagocytosis of antigen-presenting cells. In the cells, the antigens get degraded into peptides (5-10 amino acids long) that can be recognized by an *Mhc* molecule, and following identification the *Mhc* molecule binds to the peptide and transports it to

the cell surface for presentation. If a T-cell binds to the *Mhc*-peptide complex, it initiates a cascade of immune responses (Potts & Wakeland 1990).

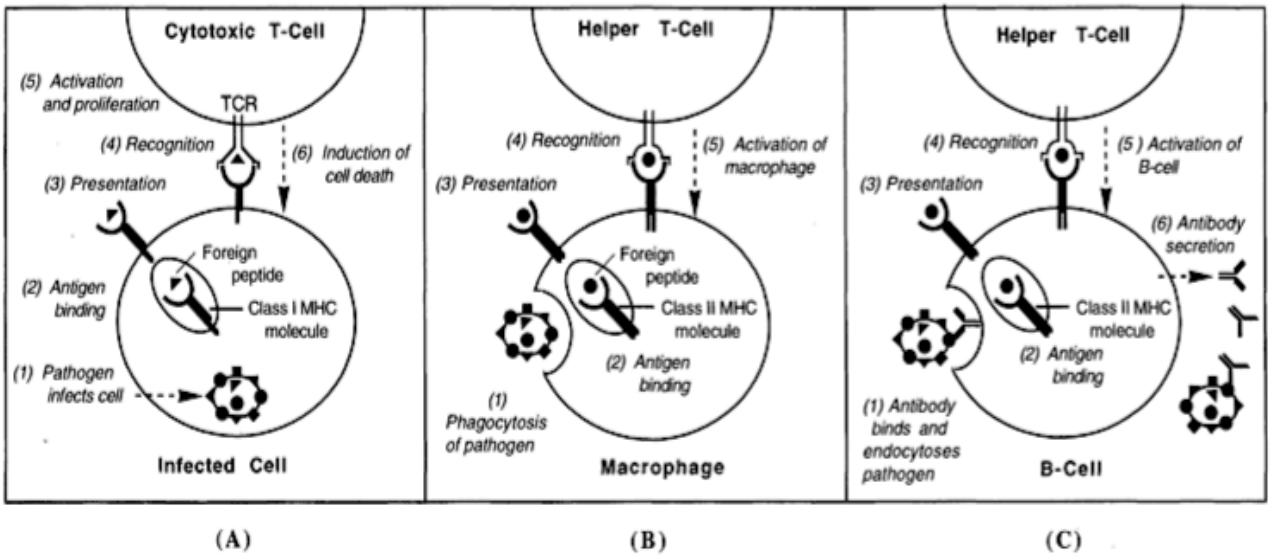
The crucial role of T-cells in triggering an immune response requires toleration of self-peptides so that T-cells won't attack their own cells, given that *Mhc* molecules present both self and foreign-peptides on the cell surface. Each T-cell contains a highly specific receptor (TCR) that distinguishes only a particular combination of *Mhc* molecule and peptide antigen. The recognition site of the TCR is generated in the thymus after a complex selection process. In the initial positive selection, all progenitor T-cells with receptors capable of binding *Mhc*-peptide complexes are selected for, whereas in the final negative selection process the receptors that recognize *Mhc*-self peptides are eliminated, thus self-tolerance is achieved (Allen 1994; Germain 1994). The surviving mature T-cells leave the thymus and migrate to the periphery where they only identify *Mhc*-foreign peptide complexes (Potts & Wakeland 1990). *Mhc* molecules are composed of two main parts; the immunoglobin domain, anchoring the molecule in the surface of cells, and the basket shaped receptor, or antigen-binding site (ABS) (Edwards & Hedrick 1998). An individual *Mhc* molecule can recognize a few peptides only, which is determined by the amino acid composition in the ABS.

There are two major subfamilies of *Mhc* that differ both functionally and structurally; class I and class II genes (Figure 1.1). Class I genes are expressed on the cell surface of almost all nucleated somatic cells and present peptides from intracellular pathogens (e.g. viruses, some protozoa) to cytotoxic-T cells. The recognition of *Mhc* class I-peptide complexes by the cytotoxic T-cell leads to T-cell activation and proliferation that induces destruction of the infected cell (reviewed in Bjorkman *et al.* 1987; Hughes & Yeager 1998).

Class II genes, in contrast, are only expressed on the surface of antigen-presenting cells of the immune system (macrophages, B cells, activated T-cells), and the *Mhc* molecules present peptides from extracellular pathogens (e.g. bacteria, helminths) to helper-T cells. When the helper T-cell recognizes the *Mhc*-foreign peptide complex, it binds to the complex and releases

cytokines to trigger an immune response, including the production of antibodies (reviewed in Hughes & Yeager 1998; Penn & Potts 1999).

Figure 1.1 - The functioning of (A) *Mhc* class I and (B, C) *Mhc* class II molecules (figure reproduced from Penn & Potts 1999).



Despite the functional similarity in immune recognition and structural similarity of the peptide presenting molecules, the genomic organization of *Mhc* differs widely across vertebrates. The mammalian *Mhc* is a gene-dense region that spans approximately four megabases of DNA and is divided into regions with similar function. Class I and class II genes are considered 'classical' *Mhc* genes to indicate their polymorphic, highly expressed structure and antigen processing and presenting function. Class III genes contain a selection of diverse immune and non-immune genes whose presence varies between species. In sharp contrast to mammals with their multiple class I and II loci, chicken (*Gallus gallus domesticus*) has the "minimal essential *Mhc*" a term referring to its surprisingly small, densely packed *Mhc* region with only two class I and II genes (Kaufman *et al.* 1999; Hess & Edwards 2002). Recent studies have shown that "minimal essential *Mhc*" does not hold for other avian orders as a larger (and sometimes much larger) number of class I and II genes have been detected in passerines (Westerdahl 2007).

The number of *Mhc* genes per subfamily can differ greatly between species and can show no sequence orthology between orders or even genera (e.g. class I gene number and organization differences between human and chimpanzees). The evolution of *Mhc* is explained by the "birth and death model of evolution". Following divergence from a common ancestral line, *Mhc* changed rapidly through gene duplication, polymorphism and gene conversion, and these new genes were either maintained, deleted or became non-functional by deleterious mutations (Nei *et al.* 1997). Therefore *Mhc* harbours both diverse functional genes and pseudogenes and these phenomena explain the variation among closely related species (reviewed in Kelley *et al.* 2005). The recent in depth characterization of zebra finch *Mhc* demonstrated the complexity of this region, as more than ten *Mhc* class I and class II loci located on different chromosomes were detected (Balakrishnan *et al.* 2010; but see Ekblom *et al.* 2011). Moreover a large number of the genes were suspected to be pseudogenes and only one *Mhc* class I locus appeared to be functional, contrasting to the pattern observed in other passerine species, where many functional class I loci were described (Westerdahl *et al.* 1999; Bonneaud *et al.* 2004; Promerová *et al.* 2009; Schut *et al.* 2011). This thesis focuses largely on *Mhc* class I variation and selection in a free-living population of great tits.

Testing for historical selection on *Mhc* genes

A large number of studies, from a broad range of vertebrate species, have found evidence for balancing selection, a form of natural selection that favours persistence of multiple alleles in a population, acting on *Mhc* (Hedrick & Thomson 1983). The involvement of *Mhc* molecules in protein-protein recognition and host-parasite interactions gives rise to selection favouring repeated nucleotide changes in the ABS codons and this particular pattern of selection is statistically easy to detect (Hughes 2007). Neutral theory predicts a constant rate of nucleotide substitution with time, whereas selection is expected to increase the occurrence and retention of adaptive nucleotide substitutions. One test is commonly being used to detect long-term effects

of selection on *Mhc* sequence data: the ratio of nonsynonymous to synonymous substitution (dN/dS ratio test). Mutations that do not affect the amino acid composition are called synonymous (silent) mutations (dS), and are regarded as selectively neutral; thus the rate of synonymous substitution and the rate of mutation are expected to be equal. Mutations that do affect the amino acid composition are called non-synonymous mutations (dN) and are possibly under selection. Whenever an advantageous non-synonymous mutation is retained by selection a dN/dS>1 should be observed, due to an increase in the rate of evolution compared to the neutral rate. If *Mhc* ABS diversity is maintained by balancing selection then beneficial non-synonymous mutations would be retained and a dN/dS>1 would be expected (Hill & Hastie 1987). In 1988, Hughes & Nei detected an excess of nonsynonymous mutations in the ABS codons of *Mhc* class I genes and suggested that the region was under balancing selection. Since then, many studies have used the test to provide evidence for the action of selection. Thus, in a meta-analysis Bernatchez & Landry (2003) reviewed 78 studies that have investigated *Mhc* in non-model species and showed that 48 of them applied the dN/dS ratio test from which all but one found the expected dN/dS>1.

However these results cannot be interpreted as evidence for selection acting on contemporary populations. In a review paper Garrigan & Hedrick (2003) ran computer stimulations to investigate the length of time needed to gain a selective signal using dN/dS ratio test and observed that the signal would develop over 10.000 generations. They also investigated the length of time required to lose a significant ratio and reached a confidence interval of 19-74 million generations. The authors came to the conclusion that accumulation of a significant dN/dS ratio would require a considerable time, but that when dN/dS>1 is reached it would take much longer for the signal to disappear even in the absence of selection. However, in this study point mutations were considered as the only mechanism generating sequence variation at the *Mhc* and the influence of gene conversion in generating new *Mhc* haplotypes was overlooked, thus the requisite amount of time to acquire a significant dN/dS ratio is not necessarily clear. Gene conversion can generate *Mhc* variation by transferring small DNA segments within or

across duplicated loci (Geliebter & Nathenson 1988). Yet it is statistically difficult to differentiate gene conversion from accumulation of point mutations, and little is known about the relative roles of the two mechanisms in generating *Mhc* variation in natural populations (Ohta 1991; Sato *et al.* 2011). Nevertheless, Spurgin *et al.* (2011) has recently shown that gene conversion is the predominant mechanism generating *Mhc* haplotypes in the Berthelot's pipit (*Anthus berthelotii*), a recently founded bird population. Moreover it is highly likely that the variation in *Mhc* gene copy number within multiple species is due to gene conversion (Malago-Trillo *et al.* 1998). Therefore, for a comprehensive understanding of *Mhc* evolution, it is important to test for gene conversion and the presence of historical selection consecutively. Also the dN/dS ratio test should solely be seen as evidence for past selection that has acted at some unknown time, potentially far in the past, for an unknown period. Hence, while the dN/dS ratio test presents powerful evidence for the operation of balancing selection on *Mhc* it fails to provide any information on the cause or timing of selection.

Tests that are used for detecting selection in contemporary populations utilize information on allele frequencies and mutation distributions within and between populations, thus provide powerful evidence for ongoing evolutionary processes. In particular, contrasting the differentiation among populations both at the *Mhc* and neutral markers by conventional statistics (e.g. F_{st}) has become one of the most powerful tests to determine the local strength of selection causing adaptive variation. Contemporary selection can also be revealed by examining measures of host fitness in relation to *Mhc* characteristics (reviewed in Bernatchez & Landry 2003; Piertney & Oliver 2006; Spurgin & Richardson 2010). Furthermore tests designed to detect balancing selection in current generations can be used to identify the present causes of selection. Theory predicts that *Mhc* polymorphism is maintained by the mechanisms of parasite-driven or sexual selection, and a strong empirical framework has been established supporting both mechanisms. However an alternative but not mutually exclusive mechanism has also been proposed to explain the extreme variation in *Mhc* genes. According to the idea of 'Associative Balancing Complex' evolution, recessive deleterious mutations accumulate in close proximity

of *Mhc* genes because they are rarely expressed and purifying selection is ineffective with low recombination rates (van Oosterhout 2009). Selection against the deleterious mutations at *Mhc*-linked regions can operate alongside parasite driven or sexual selection to maintain *Mhc* polymorphism.

Characterization and genotyping of *Mhc* loci

The complexity of *Mhc* structure, and the variation in its organisation among closely related species, pose a great challenge for *Mhc* characterization and genotyping in non-model study species. Especially in species where *Mhc* genes have undergone rapid duplication, gene conversion and recombination, the genes are mostly tightly linked to each other and alleles are shared among loci, making it almost impossible to identify and isolate independent locus; therefore simultaneous amplification of multiple loci becomes necessary (Babik *et al.* 2009).

Multilocus typing can be a difficult task and its reliability depends on a number of factors:

(1) Thorough background work must be carried out initially, in order to have a sufficient understanding of gene and allelic diversity; and ideally all allelic variants of interest should be amplified. Inaccurate characterization would bias any functional interpretation, particularly when highly expressed alleles are absent or underrepresented in the typing stage.

(2) The presence of pseudogenes together with functional loci necessitates the differentiation between expressed and nonfunctional alleles. An understanding of allele expression could be attained by RNA extraction and cDNA reverse transcription in a subset of individuals, and this information would significantly improve the analysis of sequence data. Alternatively historical selection tests can be used to separate the functional alleles from the non-functional ones.

(3) Designing the ideal primer pair for *Mhc* genotyping is of great importance. The primers should be complimentary to a conserved region so that all the alleles in the gene pool would be amplified; and they should target the ABS or most of the highly variable region specifically. Additional steps should also be taken for minimizing PCR artefacts.

Establishing an accurate genotyping method for complex *Mhc* systems is also a demanding task. Indirect typing methods that do not provide the nucleotide sequences, and instead rely on physical separation of alleles, have commonly been used to date. Denaturing gel gradient electrophoresis (DGGE), single-stranded conformational polymorphism (SSCP) and reference strand mediated conformational polymorphism (RSCA) are the most popular conformation-based mutation detection methods and can detect sequence variants according to their differential mobility (reviewed in Lenz *et al.* 2009). However, these genotyping methods have limited resolution, require extensive initial optimization, necessitate further characterization via cloning-sequencing and become more problematic and unreliable when alleles from multiple loci are being amplified (reviewed in Babik 2010).

The application of Next Generation Sequencing (NGS) methods for *Mhc* screening has recently been introduced as an exciting alternative to the conservative typing methods as they could significantly increase the efficiency and resolution of *Mhc* genotyping. The 454 pyrosequencing technology allows genotyping of hundreds of individuals, at several loci, in parallel. The most pronounced novelty of this technology is that it includes an emulsion polymerization chain reaction (emPCR) step in which each sequence is isolated before sequencing. This approach is similar to sequencing clonally amplified products but uses a cell-free environment, thus eliminates PCR artefact formation that occurs during bacteria cloning. Furthermore the use of primers that include unique nucleotide barcodes allow for individual identification, hence each read can be reliably attributed to its original sample. Because this technique also generates large numbers of sequences (perhaps > 1 million reads in a single run) it enables the representation of each allele, despite their rarity or low copy number. One major

disadvantage of the ultra-high-throughput NGS method is its error-prone sequencing technology. However this problem can be resolved by employing a strict, empirical quality control to distinguish true alleles from sequencing artefacts and the repeatability of the method can be checked by running a fraction of the samples in duplicate (Babik 2010; Galan *et al.* 2010).

Since its recent development, 454 pyrosequencing has been used a few times to genotype complex multilocus *Mhc* systems in non-model vertebrates (Babik *et al.* 2009; Galan *et al.* 2010; Kloch *et al.* 2010; Zagalska-Neubauer *et al.* 2010; Spurgin *et al.* 2011; Radwan *et al.* 2012). Each study succeeded in discriminating true alleles from artificial sequence variants by developing stepwise procedures for data analysis and variant validation, despite the high number of artefacts produced. The high repeatability and reliability of their results were confirmed for all but one, by cloning or by running samples in duplicates. In one study, that by Zagalska-Neubauer *et al.* (2010), the repeatability of duplicate samples was low, because the coverage of reads was not sufficient for reliable genotyping. However this did not prevent full characterization of diversity at species level.

These studies demonstrate that 454 technology can be employed for accurate typing of complex *Mhc* loci. Reliable genotyping of multilocus *Mhc* alleles is, however, still not sufficient for determining what equates to functionally important alleles (Spurgin & Richardson 2010). Certain *Mhc* alleles share common structural and functional features, and selection on the *Mhc* is likely to act via the functional properties of the underlying genes. Hence, treating alleles as if they are equally distinct in terms of their phenotypic effects is questionable: sometimes small sequence differences will result in large functional differences (e.g. polymorphisms altering the aminoacids at the ABS) and vice versa. Therefore, bioinformatic approaches have been proposed to identify the antigen-binding affinities of alleles, based on the physicochemical properties of their ABS, and to cluster alleles with similar predicted functional effects into groups, or 'supertypes' (Doytchinova & Flower 2005). The clustering of functionally similar *Mhc* alleles into supertypes has been used in the development of epitope-based vaccines (Sette

et al. 2002). Likewise, Trachtenberg *et al.* (2003) showed that rare human leukocyte antigen (HLA) supertypes confer a strong advantage in responding to HIV infection, independent of the contribution of single alleles. The biological relevance of grouping *Mhc* alleles with similar antigen-binding affinities into supertypes is being supported by a growing number of human and non-human primate studies (Lund *et al.* 2004; Schwensow *et al.* 2008; Huchard *et al.* 2010b). A central goal of this thesis was therefore to accurately characterize and genotype *Mhc* class I exon 3 of the great tit and to estimate the phenotypic effects of alleles, by using 454 pyrosequencing technology and clustering alleles into functional supertypes, to conduct a detailed study of selection acting on the functional diversity of great tit *Mhc*.

Mhc and associations with parasites

The critical role of *Mhc* genes in immune function led to the suggestion that parasite-mediated selection is the driving force maintaining diversity at *Mhc* loci. Thus *Mhc* genes are commonly being used as a model system for studying how parasite-mediated selection operates and three hypotheses have been put forward to explain how parasites can maintain diversity: negative frequency dependence, fluctuating selection and heterozygote advantage.

The *negative frequency dependent selection hypothesis* proposes a fitness advantage for rare allele carriers, because transmission efficiency of parasites increases with the frequency of susceptible hosts and common alleles are expected to be under the greatest pressure from parasites (Bodmer 1972; Slade & McCallum 1992). Hence, rare *Mhc* alleles are likely to offer greater resistance to parasites until the allele increases in frequency. The *fluctuating selection hypothesis*, on the other hand, states that spatiotemporal heterogeneity of parasites lead to differential selection on *Mhc* at the global scale (Hill *et al.* 1991). Variable selection pressures, determined by the abundance and diversity of parasites in space and time, are likely to contribute to the enormous allelic diversity at *Mhc* (Hedrick 2002). Lastly, the *heterozygote advantage hypothesis* proposes that high allelic diversity is advantageous for individuals facing

heterogeneous pathogenic pressures, as they are able to respond to a greater variety of parasites. Therefore *Mhc* heterozygotes are expected to have higher fitness than homozygotes; and as a result, more *Mhc* alleles can persist in the population (Doherty & Zinkernagel 1975; Hughes & Nei 1988). Still, several studies have found that individuals possessing an intermediate number of *Mhc* alleles, rather than the maximum; had higher relative fitness (Wegner *et al.* 2003, 2008; Madsen & Ujvari 2006; Kalbe *et al.* 2009; Kloch *et al.* 2010). This pattern is consistent with the model proposed by Nowak *et al.* (1992) that suggests reduced fitness for individuals with high *Mhc* diversity, resulting from excessive T-cell elimination during negative selection in thymus.

Studies encompassing a range of organisms including model species (humans: Thursz *et al.* 1997; mice: Penn *et al.* 2002; McClelland *et al.* 2003) as well as natural populations of non-model organisms (fish: Dionne *et al.* 2009; birds: Bonneaud *et al.* 2006b; Loiseau *et al.* 2011; mammals: Tollenaere *et al.* 2008; Oliver *et al.* 2009) have attempted to differentiate the relative roles of these mechanisms in maintaining *Mhc* diversity, by investigating the associations between specific *Mhc* alleles or allelic diversity and resistance or susceptibility to parasitic infections. However Spurgin & Richardson (2010) have recently demonstrated that in many cases the same predictions arise from the three mechanisms, and that most studies that have inferred one mechanism of selection have not fully considered the alternative options. Especially in systems where multiple loci are simultaneously being amplified, differentiating between mechanisms of parasite-mediated selection is rather difficult and resolving their relative importance may be almost impossible (Spurgin & Richardson 2010). Still, such disease association studies are valuable for understanding the genetic basis of variation in infection.

Despite recent advances in genetic technologies, detecting disease resistance loci remains difficult in non-model species (Amos *et al.* 2011); hence well-studied genetic markers like *Mhc* are great candidates for furthering our understanding of wildlife immunogenetics. Although numerous studies have explored links between *Mhc* and disease prevalence, often, the ecological/epidemiological basis of the host-parasite interaction is lacking; this lack makes the interpretation of associations between parasite prevalence and *Mhc* difficult. Especially

discriminating between alleles for qualitative resistance (alleles that provide protection against infection), quantitative resistance (alleles that suppress the development of infection) and susceptibility is challenging, even though such discrimination is crucial to understand the dynamics of complex host-parasite relationships. For example, positive associations between *Mhc* alleles and parasite prevalence might either be the result of susceptibility alleles that make individuals more prone to infection, or a consequence of quantitative-disease resistance alleles that limit the deleterious effects of the disease without eradicating it completely (Westerdahl *et al.* 2011). In order to interpret positive associations accurately, studies should consider the epidemiological and ecological context of host-parasite interactions.

A number of studies to date have investigated *Mhc*-based avian malaria resistance, by examining the associations between *Mhc* class I alleles and *Plasmodium* infection (Westerdahl *et al.* 2005, 2011; Bonneaud *et al.* 2006b; Loiseau *et al.* 2008, 2011). Avian malaria parasites, like their human-borne counterparts, are intracellular pathogens that invade host erythrocytes in the bloodstream, making it highly likely that *Mhc* class I molecules of avian hosts would play a role in the recognition of the peptides derived from malaria parasites and initiate cell destruction. Consequently, significant associations between *Mhc* alleles and parasite prevalence were reported in all these studies, however interpretation of the results varied considerably among work. Moreover, data on environmental determinants of avian malaria risk and knowledge on fitness consequences of *Plasmodium* parasites were scarce in these passerine systems. A second major aim of this thesis was therefore to address these problems and incorporate a detailed investigation of avian malaria infection and analysis of *Mhc* class I functional diversity, to understand the role that *Mhc* play in determining host resistance and susceptibility to *Plasmodium* infections.

Previous work on the tit population has revealed that two divergent *Plasmodium* parasite species, *P. relictum* and *P. circumflexum*, are common in the study site; and that there is pronounced spatial variation in the distribution of the two *Plasmodium* species in the two sympatric, closely-related host species, great tits and blue tits (*Cyanistes caeruleus*) (Wood *et*

al. 2007; Knowles *et al.* 2011; Lachish *et al.* 2012). *P. relictum* infections are effectively randomly distributed in space, whereas *P. circumflexum* infections exhibit pronounced spatial structuring that is stable over years in both species (Lachish *et al.* 2012). Moreover the two *Plasmodium* species differ substantially in their impacts on the blue tit host. While *P. relictum* infections are linked with reproductive costs (Knowles *et al.* 2010), *P. circumflexum* infections are associated with reduced survival, particularly during the acute stage of infection (Lachish *et al.* 2011); hence it is possible that these parasites might similarly exert differential selection pressures on great tit hosts.

Effect of *Mhc* on measures of fitness

Studies that relate *Mhc* diversity to parasite regime assume that the parasites under question exert strong selection on their hosts, though this assumption is generally not confirmed. Surprisingly little is known about the fitness consequences of *Mhc* variation in wild populations, since most work investigate parasite resistance using assays that do not assess the fitness costs of parasite infection (reviewed in Penn 2002). As mentioned before, selection can either maximize *Mhc* diversity (heterozygote advantage), optimize *Mhc* diversity (heterozygote advantage) or favour specific functional variants (negative frequency dependence or fluctuating selection). However most disease resistance studies examine host resistance against just one or few parasites only, even though the hypotheses concerning maximal or optimal allelic diversity are particularly important for individuals facing heterogeneous pathogenic pressures (Penn *et al.* 2002; McClelland *et al.* 2003). Hence, documenting associations between parasite prevalence and *Mhc* do not allow a straightforward inference as to how *Mhc* affects individual fitness and how selection operates to maintain *Mhc* allelic diversity. A number of studies have addressed these issues and investigated the role that *Mhc* play in individual variation in survival or lifetime reproductive success (LRS), to determine *Mhc*-fitness correlations (e.g. Paterson *et al.* 1998; Sauermann *et al.* 2001; Wegner *et al.* 2008; Eizaguirre *et al.* 2009b; Kalbe *et al.* 2009; Brouwer

et al. 2010; Worley *et al.* 2010; Thoß *et al.* 2011; Radwan *et al.* 2012). However, the vast majority of these studies were carried out in semi-natural settings, for example in field enclosure systems or large indoor arenas; providing stable and benign environments where selection is unlikely to operate as it would in nature. Whether and how *Mhc* effect individual survival and LRS in natural settings is much less well known, even though such knowledge is crucial to understand the processes underlying *Mhc* evolution. Two studies to date have investigated *Mhc*-linked host survival in natural populations and found positive associations between certain *Mhc* alleles and life expectancy (Paterson *et al.* 1998; Brouwer *et al.* 2010). Yet, in both these studies fitness was assessed in terms of individual survival only, while measures of reproductive performance and particularly LRS were not taken into account; hence the link between *Mhc* genes and individual fitness might have been underestimated. A third major aim of this thesis was therefore to combine a detailed investigation of host survival, reproductive performance and LRS with analysis of *Mhc* class I functional diversity, to understand the role that *Mhc* play in host fitness within a single wild population.

Mhc and female mate choice decisions

Mhc-dependent sexual selection has been proposed as a secondary mechanism maintaining *Mhc* diversity (Brown & Eklund 1994). Sexual selection acts continuously from the level of mating to the level of offspring production; thus different mechanisms have been proposed for the stage at which *Mhc*-dependent selection might be operating. *Mhc*-based mate choice (reviewed in Milinski 2006; Piertney & Oliver 2006; Huchard *et al.* 2010a), *Mhc*-dependent sperm selection (Wedekind *et al.* 1996; Rulicke *et al.* 1998; Yeates *et al.* 2009) and *Mhc*-related selective abortion (Hedrick & Thomson 1988; Ober 1998) are three mechanisms that have been explored; however *Mhc*-based mate selection has to date received the most attention. There are two main hypotheses explaining how *Mhc*-based mate choice can act. The *complementary genes hypothesis* states that females seek genes that complement their own genes in order to optimize

the genetic diversity, and hence fitness of their offspring (Zeh & Zeh 1996). Preference for *Mhc* dissimilar mates could act to increase *Mhc* diversity of the progeny so that the offspring could cope with more parasites and have enhanced immunocompetence (heterozygote advantage), or the offspring could be provided with a 'moving target' against rapidly evolving or recently introduced parasites (negative frequency dependence or fluctuating selection) (Penn & Potts 1999). However if maximal parasite resistance is achieved by an intermediate number of functional *Mhc* variants (Wegner *et al.* 2008); then mate choice should favour optimal *Mhc* dissimilarity. Moreover, *Mhc*-dependent disassortative mating could also be an adaptation to avoid mating with kin and to prevent inbreeding depression in the offspring. There is increasing support for the suggestion that *Mhc* constitution is olfactorily perceptible in a variety of taxa (Reusch *et al.* 2001; Olsson *et al.* 2003; Leinders-Zufall *et al.* 2004; Milinski *et al.* 2005; Radwan *et al.* 2008); hence individual odour profiles are predicted to be the cues for *Mhc* dissimilarity.

The alternative, but not exclusive *good genes hypothesis* states that female mate choice should be based on male quality regardless of the female's own genotype, so that the female would guarantee resource gain or genetic benefits for the offspring, thus maximize her own reproductive success (Hamilton & Zuk 1982). Condition-dependent traits such as sexual secondary characters are predicted to be the cues for male quality, and females could either prefer more *Mhc* diverse or optimally diverse males, or seek mates carrying specific *Mhc* types. Studies encompassing a wide range of organisms from fish (Forsberg *et al.* 2007; Garner *et al.* 2010) to mammals (Schwensow *et al.* 2008), including reptiles (Olsson *et al.* 2005; Miller *et al.* 2009) and birds (Bonneaud *et al.* 2006a; Juola & Dearborn 2012) have reported *Mhc*-dependent mating preferences (reviewed in Huchard *et al.* 2010a). However *Mhc*-based mate choice does not appear to be a general phenomenon since several studies failed to present evidence for *Mhc*-dependent mating decisions despite the inclusion of detailed behavioural and molecular data (Paterson & Pemberton 1997; Westerdahl 2004; Sommer 2005; Huchard *et al.* 2010a). Moreover, the genetic benefits associated with *Mhc*-based mate choice have rarely been

assessed in wild populations, even though documenting such patterns is crucial to understand the relevance of *Mhc* for mate choice in nature (but see Brouwer *et al.* 2010; Huchard *et al.* 2010a). Likewise, only few studies to date have controlled for effects of background relatedness and spatial population structure while investigating *Mhc*-based disassortative mating preferences, although both measures can substantially influence the observed mate choice patterns (Richardson *et al.* 2005; Miller *et al.* 2009; Szulkin *et al.* 2009). Thus a final goal of the work in this thesis was to rectify these problems, by applying spatially explicit randomizations and incorporating *Mhc* dissimilarity-relatedness in a single model, to test for *Mhc*-based disassortative mating in a wild population; and to examine the reproductive fitness consequences of *Mhc*-dependent mating patterns.

Study site and species

Most of the fieldwork for this thesis was conducted at Wytham Woods, a c. 380 ha continuous mixed deciduous woodland, near Oxford, UK (51°46'N, 1°20'W). The Edward Grey Institute (EGI) has been monitoring the great tit population breeding in nestboxes throughout Wytham Woods continuously since the early 1960s. There are 1020 nestboxes scattered at variable densities around the woodland that are suitable for great tit breeding attempts; each year between 250 to 500 pairs breed in Wytham Woods. The woodland is predominantly made up of the following tree species: oak (*Quercus spp.*), hazel (*Corylus avellana*), sycamore (*Acer pseudoplatanus*), ash (*Franxinus excelsior*) and beech (*Fagus sylvatica*). There are also relatively open areas comprising of bramble (*Rubus fruticosa*), elder (*Sambucus nigra)*, bracken (*Pteridium aquilinum*), blackthorn (*Prunus spinosa*) and hawthorn (*Crataegus monogyna*) (Savill *et al.* 2010). To facilitate management and research, the woodland is divided into nine sections. A map of Wytham Woods showing the borders of the nine sections and the location of nestboxes is presented in Figure 1.2a.

The great tit is a small, short-lived, hole-nesting passerine bird species that resides in Wytham Woods all year around (Figure 1.2b). They take readily to nestboxes and in this part of the UK have a synchronous, annual breeding season (April to June), and are single-brooded. Great tits usually lay between 8-12 eggs, which are incubated solely by the female. Eggs hatch approximately 13 days after incubation begins, and both parents provision the brood for 19-21 days until the young fledge (Gosler 1993). The fledglings maintain close contact with parents for a further 13-25 days; and once reaching independence they disperse from the natal area, with dispersal being female biased (Verhulst *et al.* 1997; Dingemanse *et al.* 2003). Natal dispersal is the major dispersal event in their life-history and the distances differ between the sexes, with females dispersing 49% further than males on average (median and IQR: males 528m (298-931); females 788m (456-1338)) (Szulkin & Sheldon 2008). Thereafter adults disperse only short distances between breeding events, hence display high breeding site fidelity following postnatal dispersal.

Figure 1.2 - Study site and study species: (A) a map of Wytham Woods and (B) an adult great tit in Wytham Woods.

(Photo credit for B. Nicole Milligan)

Thesis outline and aims

In this thesis, I used a combination of detailed genetic analysis and long-term breeding data to explore whether and how selection operates on *Mhc* class I functional variation in the great tit. Much of the work presented here involves the application of 454-pyrosequencing method for *Mhc* genotyping, and characterization of the functional properties of *Mhc* via a 'supertyping' approach, to assess selection on *Mhc* phenotype in the wild.

The thesis consists of four data chapters, as well as a general introduction (Chapter 1) and discussion (Chapter 6) to place the work presented here within the broader context of *Mhc* studies addressing selection in contemporary populations and *Mhc* research in non-model vertebrate species. In Chapter 2, I undertook a comprehensive characterization effort to detect *Mhc* class I exon 3 variation in the great tit, and used 454 pyrosequencing for genotyping. Then I clustered *Mhc* alleles into functional supertypes, to determine the biologically meaningful *Mhc* variants of each individual. In Chapter 3, I investigated associations among patterns of avian malaria infection and *Mhc* class I supertypes, to understand the role that *Mhc* play in determining host resistance and susceptibility to *Plasmodium* infections. In Chapter 4, I tested a number of hypotheses that relate *Mhc* diversity to fitness, to estimate the fitness consequences of *Mhc* supertypes for both survival and lifetime reproductive success in a wild population. Finally, in Chapter 5, I explored the role that *Mhc* class I plays in female mate choice decisions to understand whether selection favours preference for dissimilar *Mhc* supertypes.

The specific questions I aimed to answer were:

- How variable are *Mhc* class I genes in the great tit both in terms of allelic diversity and gene number, and how does genotypic variation relate to phenotypic effects? (Chapter 2)
- Does functional diversity at *Mhc* influence resistance or susceptibility to avian malaria infections? (Chapter 3)
- What are the effects of *Mhc* functional diversity on adult survival, reproductive performance and lifetime reproductive success in a wild population? (Chapter 4)
- Does functional compatibility at *Mhc* matter for females during mate choice decision and why? (Chapter 5)

Author contributions

Author contributions to each of the four data chapters contained in this thesis

	I. Sepil	H. K. Moghadam	E. Huchard	A. E. Hinks	S. Lachish	R. Radersma	A. W. Santure	I. De Cauwer	J. Slate	B. C. Sheldon
Chapters co-authored	**2, 3, 4, 5**	**2**	**2**	**3**	**3, 4**	**5**	**5**	**5**	**5**	**2, 3, 4, 5**
Contribution										
Writing of chapter, conducting analyses	2, 3, 4, 5									
Field data collection	2, 3, 4, 5			2, 3, 4, 5						
Mhc characterization and genotyping	2, 3, 4, 5									
Providing tools for bioinformatics		2								
Derived variables using GIS software				3						
Running mark-recapture methods					4					
SNP genotyping							5	5	5	
Running randomization tests						5				
Calculation of IBS relatedness							5			
Guidance on analyses			2		3, 4					2, 3, 4, 5
Comments on drafts		2	2		3, 4					2, 3, 4, 5
Generation of ideas, intellectual input	2, 3, 4, 5	2	2		3, 4	5	5			2, 3, 4, 5

References

Agudo, R., Carrete, M., Alcaide, M., Rico, C., Hiraldo, F. & Donazar, J.A. (2012). Genetic diversity at neutral and adaptive loci determines individual fitness in a long-lived territorial bird. *Proceedings of the Royal Society B: Biological Sciences,* doi: 10.1098/rspb.2011.2606.

Allen, P.M. (1994) Peptides in positive and negative selection: a delicate balance. *Cell*, 76, 593-596.

Amos, W., Driscoll, E. & Hoffman, J.I. (2011). Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proceedings of the Royal Society B: Biological Sciences*, 278, 1183-1188.

Apanius ,V., Penn, D., Slev, P.R., Ruff, L.R., Potts, W.K. (1997) The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, 17, 179–224.

Arkush, K.D., Giese, A.R., Mendonca, H.L., McBride, A.M., Marty, G.D. & Hedrick, P.W. (2002) Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. *Canadian Journal of Fisheries and Aquatic Sciences,* 59, 966–975.

Babik, W. (2010). Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources*, 10, 237-251.

Babik, W., Taberlet, P., Ejsmond, M.J. & Radwan, J. (2009). New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources*, 9, 713-719.

Balakrishnan, C., Ekblom, R., Völker, M., Westerdahl, H., Godinez, R., Kotkiewicz, H., *et al.* (2010). Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biology*, 8, 29.

Bernatchez, L. & Landry, C. (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, 16, 363-377.

Bjorkman, P., Saper, M., Samraoui, B. & Bennett, W. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature*, 329, 506-512.

Bodmer, W. (1972). Evolutionary significance of the HL-A system. *Nature*, 237, 139-183.

Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H. & Sorci, G. (2006a). Complex Mhc-based mate choice in a wild passerine. *Proceedings of the Royal Society B: Biological Sciences*, 273, 1111-1116.

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

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

Briles, W.E., Stone, H.A. & Cole, R.K. (1977) Marek's disease: effects of B histocompatibility alloalleles in resistant and susceptible chicken lines. *Science*, 195, 193–195.

Brouwer, L., Barr, I., van de Pol, M., Burke, T., Komdeur, J. & Richardson, D.S. (2010). MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, 19, 3444-3455.

Brown, J. & Eklund, A. (1994). Kin Recognition and the Major Histocompatibility Complex: An Integrative Review. *American Naturalist*, 143, 435-461.

Dingemanse, N.J., Both, C., van Noordwijk, A.J., Rutten, A.L. & Drent, P.J. (2003). Natal dispersal and personalities in great tits (*Parus major*). *Proceedings of the Royal Society B: Biological Sciences*, 270, 741-747.

Dionne, M., Miller, K.M., Dodson, J.J. & Bernatchez, L. (2009). MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1555-1565.

Doherty, P. & Zinkernagel, R. (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, 256, 50-52.

Doytchinova, I.A. & Flower, D.R. (2005). In silico identification of supertypes for class II MHCs. *Journal of Immunology*, 174, 7085-7095.

Edwards, S.V. & Hedrick, P.W. (1998). Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology & Evolution*, 13, 305-311.

Eizaguirre, C., Lenz, T.L., Traulsen, A. & Milinski, M. (2009a). Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. *Ecology Letters*, 12, 5-12.

Eizaguirre, C., Yeates, S.E., Lenz, T.L., Kalbe, M. & Milinski, M. (2009b). MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology*, 18, 3316-3329.

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, 523-530.

Ellegren, H. (2008). Comparative genomics and the study of evolution by natural selection. *Molecular Ecology*, 17, 4586-4596.

Ellegren, H. & Sheldon, B.C. (2008). Genetic basis of fitness differences in natural populations. *Nature*, 452, 169-175.

Ford, M.J. (2002). Applications of selective neutrality tests to molecular ecology. *Molecular Ecology*, 11, 1245-1262.

Forsberg, L.A., Dannewitz, J., Petersson, E. & Grahn, M. (2007). Influence of genetic dissimilarity in the reproductive success and mate choice of brown trout - females fishing for optimal MHC dissimilarity. *Journal of Evolutionary Biology*, 20, 1859-1869.

Galan, M., Guivier, E., Caraux, G., Charbonnel, N. & Cosson, J.-F. (2010). A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*, 11, 296.

Garner, S.R., Bortoluzzi, R., Heath, D.D. & Neff, B.D. (2010) Sexual conflict inhibits female mate choice for MHC dissimilarity in Chinook salmon. *Proceedings of the Royal Society B: Biological Sciences,* 277, 885–894.

Garrigan, D. & Hedrick, P.W. (2003). Detecting Adaptive Molecular Polymorphism: Lessons from the MHC. *Evolution*, 57, 1707-1722.

Geliebter, J. & Nathenson, S.G. (1988) Microrecombinations Generate Sequence Diversity in the Murine Major Histocompatibility Complex: Analysis of the K^{bm3} , K^{bm4}, K^{bm10} and K^{bmJJ} Mutants. *Molecular and Cellular Biology,* 8, 4342-4352.

Germain, R.N. (1994) MHC-dependent antigen processing and peptide presentation: Providing ligands for T-lymphocyte activation. *Cell,* 76, 287-299.

Gosler, A.G. (1993). *The Great Tit*. Paul Hamlyn, London.

Hamilton, W.D. & Zuk, M. (1982) Heritable true fitness and bright birds: a role for parasites? *Science*, 218, 384-387.

Hedrick, P.W. & Thomson, G. (1983) Evidence for balancing selection at HLA. *Genetics,* 104, 449–456.

Hedrick, P.W. & Thomson, G. (1988) Maternal-fetal interactions and the maintenance of histocompatibility polymorphism. *Genetics*, 19, 205-212.

Hedrick, P.W. (1994). Evolutionary Genetics of the Major Histocompatibility Complex. *The American Naturalist*, 143, 945-964.

Hedrick, P.W. (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution,* 56, 1902–1908.

Hess, C.M. & Edwards, S.V. (2002). The evolution of the major histocompatibility complex in birds. *Bioscience*, 52, 423–431.

Hill, A.V.S., Allsopp, C.E.M., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., *et al.* (1991). Common West African HLA antigens are associated with protection from severe malaria, *Nature*, 352, 595-600.

Hill, R.E. & Hastie, N.D. (1987) Accelerated evolution in the reactive centre regions of serine protease inhibitors. *Nature,* 326, 96–99.

Hubbard, J.K., Uy, J.A.C., Hauber, M.E., Hoekstra, H.E. & Safran, R.J. (2010). Vertebrate pigmentation: from underlying genes to adaptive function. *Trends in Genetics*, 26, 231-239.

Huchard, E., Knapp, L.A., Wang, J., Raymond, M. & Cowlishaw, G. (2010a). MHC, mate choice and heterozygote advantage in a wild social primate. *Molecular Ecology*, 19, 2545-2561.

Huchard, E., Raymond, M., Benavides, J., Marshall, H., Knapp, L.A. & Cowlishaw, G. (2010b). A female signal reflects MHC genotype in a social primate. *BMC Evolutionary Biology*, 10, 96.

Hughes, A. L. (1991) MHC polymorphism and the design of captive breeding programs. *Conservation Biology,* 5, 249–251.

Hughes, A.L. (2007). Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity*, 99, 364-373.

Hughes, A.L. & Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, 32, 415-435.

Hughes, A. & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, 335, 167-170.

Hughes, A.L. (2008). The origin of adaptive phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 13193-13194.

Juola, F. A. & Dearborn, D.C. (2012). Sequence-based evidence for major histocompatibility complex-disassortative mating in a colonial seabird. *Proceedings of the Royal Society B: Biological Sciences,* 279. 153-162

Kalbe, M., Eizaguirre, C., Dankert, I., Reusch, T.B.H., Sommerfeld, R.D., Wegner, K.M., *et al.* (2009). Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings of the Royal Society B: Biological Sciences,* 276, 925-934.

Kaufman, J., Milne, S., Göbel, T.W., Walker, B.A., Jacob, J.P., Auffray, C., *et al.* (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, 401, 923-925.

Kelley, J., Walter, L. & Trowsdale, J. (2005). Comparative genomics of major histocompatibility complexes. *Immunogenetics*, 56, 683-695.

Klein, J. (1986). *Natural History of the Major Histocompatibility Complex*. Wiley, New York.

Kloch, A., Babik, W., Bajer, A., Siński, E. & Radwan, J. (2010). Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole Myodes glareolus. *Molecular Ecology*, 19 Suppl 1, 255-265.

Knowles, S.C.L., Palinauskas, V. & Sheldon, B.C. (2010). Chronic malaria infections increase family inequalities and reduce parental fitness: experimental evidence from a wild bird population. *Journal of Evolutionary Biology*, 23, 557-569.

Knowles, S.C.L., Wood, M.J., Alves, R., Wilkin, T.A., Bensch, S. & Sheldon, B.C. (2011). Molecular epidemiology of malaria prevalence and parasitaemia in a wild bird population. *Molecular Ecology*, 20, 1062-1076.

Lachish, S., Knowles, S.C.L., Alves, R., Wood, M.J. & Sheldon, B.C. (2011). Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. *Journal of Animal Ecology*, 80, 1196-1206.

Lachish, S., Knowles, S.C.L., Alves, R., Sepil, I., Davies, A., Lee, S. *et al.* (2012). Spatial determinants of infection risk in a multi-species avian malaria system. *Ecography,* 35, 1-12.

Langefors, A., Lohm, J., Grahn, M., Andersen, Ø. & von Schantz, T. (2001) Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proceedings of the Royal Society B: Biological Sciences,* 268, 479–485.

Leinders-Zufall, T., Brennan, P., Widmayer, P., S, P.C., Maul-Pavicic, A., Jäger, M., *et al.* (2004). MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science*, 306, 1033-1037.

Lenz, T.L., Eizaguirre, C., Becker, S. & Reusch, T.B.H. (2009). RSCA genotyping of MHC for high-throughput evolutionary studies in the model organism three-spined stickleback *Gasterosteus aculeatus*. *BMC Evolutionary Biology*, 9, 57.

Loiseau, C., Zoorob, R., Garnier, S., Birard, J., Federici, P., Julliard, R., *et al.* (2008). Antagonistic effects of a Mhc class I allele on malaria-infected house sparrows. *Ecology Letters*, 11, 258-265.

Loiseau, C., Zoorob, R., Robert, A., Chastel, O., Julliard, R. & Sorci, G. (2011). *Plasmodium relictum* infection and MHC diversity in the house sparrow (*Passer domesticus*). *Proceedings of the Royal Society B: Biological Sciences*, 278, 1264-1272.

Lund, O., Nielsen, M., Kesmir, C., Petersen, A.G., Lundegaard, C., Worning, P., *et al.* (2004). Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics*, 55, 797-810.

Madsen, T. & Ujvari, B. (2006). MHC class I variation associates with parasite resistance and longevity in tropical pythons. *Journal of Evolutionary Biology*, 19, 1973-1978.

Malago-Trillo, E., Zaleska-Rutczynska, Z., McAndrew, B., Vincek, V., Figueroa, F., Sultmann, H. *et al.* (1988) Linkage Relationships and Haplotype Polymorphism Among Cichlid *Mhc* Class II *B* Loci. *Genetics,* 149, 1527-1537.

McClelland, E.E., Penn, D.J. & Potts, W.K. (2003). Major Histocompatibility Complex Heterozygote Superiority during Coinfection. *Infection & Immunity*, 71, 2079-2086.

Milinski, M. (2006). The Major Histocompatibility Complex, Sexual Selection, and Mate Choice. *Annual Review of Ecology, Evolution, and Systematics*, 37, 159-186.

Milinski, M., Griffiths, S., Wegner, K.M., Reusch, T.B.H., Haas-Assenbaum, A. & Boehm, T. (2005). Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 4414-4418.

Miller, H.C., Moore, J.A., Nelson, N.J. & Daugherty, C.H. (2009). Influence of major histocompatibility complex genotype on mating success in a free-ranging reptile population. *Proceedings of the Royal Society B: Biological Sciences,* 276, 1695-1704.

Nei, M., Gu, X. & Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 7799-7806.

Nowak, M.A., Tarczy-Hornoch, K. & Austyn, J.M. (1992). The optimal number of major histocompatibility complex molecules in an individual. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 10896-10899.

Ober, C., Hyslop, T., Elias, S., Weitkamp, L.R., Hauck, W.W. (1998) Human leukocyte antigen matching and fetal loss: results of a 10-year prospective study. *Human Reproduction*, 13, 33–38.

Ohta, T. (1991) Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. *Proceedings of the National Academy of Sciences of the United States of America*, 88, 6716-6720.

Oliver, M.K., Telfer, S. & Piertney, S.B. (2009). Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). *Proceedings of the Royal Society B: Biological Sciences*, 276, 1119-1128.

Olsson, M., Madsen, T., Nordby, J., Wapstra, E., Ujvari, B. & Wittsell, H. (2003). Major histocompatibility complex and mate choice in sand lizards. *Proceedings of the Royal Society B: Biological Sciences,* 270 Suppl , S254-S256.

Olsson, M., Madsen, T., Wapstra, E., Silverin, B., Ujvari, B. & Wittzell, H. (2005). MHC, health, color, and reproductive success in sand lizards. *Behavioral Ecology and Sociobiology*, 58, 289-294.

van Oosterhout, C., Smith, A.M., Hänfling, B., Ramnarine, I.W., Mohammed, R.S. & Cable, J. (2007). The guppy as a conservation model: implications of parasitism and inbreeding for reintroduction success. *Conservation Biology*, 21, 1573-1583.

van Oosterhout, C. (2009) A new theory of MHC evolution: beyond selection on the immune genes. *Proceedings of the Royal Society B: Biological Sciences,* 276, 657-665.

Paterson, S. & Pemberton, J.M. (1997). No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population. *Proceedings of the Royal Society B: Biological Sciences*, 264, 1813-1819.

Paterson, S., Wilson, K. & Pemberton, J. (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proceedings of the National Academy of Sciences of the United States of America*, 95, 3714-3719.

Penn, D.J. (2002) The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology,* 108, 1–21.

Penn, D.J., Damjanovich, K. & Potts, W.K. (2002). MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 11260-11264.

Penn, D.J. & Potts, W.K. (1999). The Evolution of Mating Preferences and Major Histocompatibility Complex Genes, *American Naturalist,* 153, 145-164.

Piertney, S.B. & Oliver, M.K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96, 7-21.

Potts, W.K. & Wakeland, E.K. (1990). Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution*, 5, 181-187.

Potts, W.K., Manning, C.J., Wakeland, E.K. (1994) The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 346, 369–78.

Promerová, M., Albrecht, T. & Bryja, J. (2009). Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). *Immunogenetics*, 61, 451-461.

Radwan, J., Tkacz, A. & Kloch, A. (2008). MHC and Preferences for Male Odour in the Bank Vole. *Ethology*, 114, 827-833.

Radwan, J., Zagalska-Neubauer, M., Cichoń, M., Sendecka, J., Kulma, K., Gustafsson, L., *et al.* (2012). MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Molecular Ecology,* 21, 2469-2479.

Reusch, T.B., Häberli, M.A., Aeschlimann, P.B. & Milinski, M. (2001). Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature*, 414, 300-302.

Richardson, D.S., Komdeur, J., Burke, T. & von Schantz, T. (2005) MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. *Proceedings of the Royal Society B: Biological Sciences*, 272, 759–767.

Rulicke, T., Chapuisat, M., Homberger, F.R., Macas, E., Wedekind, C. (1998). MHC-genotype of progeny influenced by parental infection. *Proceedings of the Royal Society B: Biological Sciences*, 265, 711–716.

Sato, A., Tichy, H., Grant, P.R., Grant, B.R., Sato, T. & O'hUigin, C. (2011) Spectrum of MHC Class II Variability in Darwin's Finches and Their Close Relatives. *Molecular Biology and Evolution*, 28, 1943-1956.

Sauermann, U., Nürnberg, P., Bercovitch, F., Berard, J., Trefilov, A., Widdig, A., *et al.* (2001). Increased reproductive success of MHC class II heterozygous males among free-ranging rhesus macaques. *Human Genetics*, 108, 249-254.

Savill, P., Perrins, C., Kirby, K. & Fisher, N. (2010). *Wytham Woods: Oxford's Ecological Laboratory*. Oxford University Press, New York.

Schut, E., Aguilar, J.R.-de, Merino, S., Magrath, M.J.L., Komdeur, J. & Westerdahl, H. (2011). Characterization of MHC-I in the blue tit (*Cyanistes caeruleus*) reveals low levels of genetic diversity and trans-population evolution across European populations. *Immunogenetics*, 63, 531-542.

Schwensow, N., Eberle, M. & Sommer, S. (2008). Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proceedings of the Royal Society B: Biological Sciences*, 275, 555-564.

Sette, A., Newman, M., Livingston, B., McKinney, D., Sidney, J., Ishioka, G., *et al.* (2002). Optimizing vaccine design for cellular processing, MHC binding and TCR recognition. *Tissue Antigens*, 59, 443-451.

Shapiro, M., Bell, M. & Kingsley, D. (2006). Parallel genetic origins of pelvic reduction in vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 13753-13758.

Siddle, H.V., Marzec, J., Cheng, Y., Jones, M. & Belov, K. (2010). MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. *Proceedings of the Royal Society B: Biological Sciences*, 277, 2001-2006.

Slade, R.W. & McCallum, H.I. (1992). Overdominantvs. Frequency-Dependent Selectionat MHC Loci. *Genetics*, 123, 861-862.

Sommer, S. (2005). Major histocompatibility complex and mate choice in a monogamous rodent. *Behavioral Ecology and Sociobiology*, 58, 181-189.

Spurgin, L.G. & Richardson, D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277, 979-988.

Spurgin, L.G., Van Oosterhout, C., Illera, J.C., Bridgett, S., Gharbi, K., Emerson, B.C. *et al.* (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, 20, 5213–5225.

Szulkin, M. & Sheldon, B.C. (2008). Dispersal as a means of inbreeding avoidance in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences*, 275, 703-711.

Szulkin, M., Zelazowski, P., Nicholson, G. & Sheldon, B.C. (2009). Inbreeding avoidance under different null models of random mating in the great tit. *Journal of Animal Ecology*, 78, 778-788.

Thoß, M., Ilmonen, P., Musolf, K. & Penn, D.J. (2011). Major histocompatibility complex heterozygosity enhances reproductive success. *Molecular Ecology*, 20, 1546-1557.

Thursz, M.R., Thomas, H.C., Greenwood, B.M. & Hill, A. (1997). Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nature Genetics*, 17, 11-12.

Tollenaere, C., Bryja, J., Galan, M., Cadet, P., Deter, J., Chaval, Y., *et al.* (2008). Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: insights from multivariate analyses and population genetics. *Journal of Evolutionary Biology*, 21, 1307-1320.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E., *et al.* (2003). Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, 9, 928-935.

Trowsdale, J. & Parham, P. (2004). Mini-review: defense strategies and immunity-related genes. *European Journal of Immunology*, 34, 7-17.

Verhulst, S., Perrins, C. & Riddington, R. (1997). Natal Dispersal of Great Tits in a Patchy Environment. *Ecology*, 78, 864-872.

Wedekind, C., Seebeck, T., Bettens, F., Paepke, A.J. (1995) MHC-dependent mate preferences in humans. *Proceedings of the Royal Society B: Biological Sciences*, 260, 245–249.

Wedekind, C., Chapuisat, M., Macas, E., Rulicke, T. (1996) Nonrandom fertilization in mice correlates with the MHC and something else. *Heredity,* 77, 400–409.

Wegner, K.M., Kalbe, M., Kurtz, J., Reusch, T.B.H. & Milinski, M. (2003). Parasite selection for immunogenetic optimality. *Science*, 301, 1343.

Wegner, K.M., Kalbe, M., Milinski, M. & Reusch, T.B. (2008). Mortality selection during the 2003 European heat wave in three-spined sticklebacks: effects of parasites and MHC genotype. *BMC Evolutionary Biology*, 8, 124.

Westerdahl, H. (2004). No evidence of an MHC-based female mating preference in great reed warblers. *Molecular Ecology*, 13, 2465-2470.

Westerdahl, H. (2007). Passerine MHC: genetic variation and disease resistance in the wild. *Journal of Ornithology*, 148, 469-477.

Westerdahl, H., Asghar, M., Hasselquist, D. & Bensch, S. (2011). Quantitative disease resistance: to better understand parasite-mediated selection on major histocompatibility complex. *Proceedings of the Royal Society B: Biological Sciences*, 279, 577-584.

Westerdahl, H., Waldenström, J., Hansson, B., Hasselquist, D., von Schantz, T. & Bensch, S. (2005). Associations between malaria and MHC genes in a migratory songbird. *Proceedings of the Royal Society B: Biological Sciences*, 272, 1511-1518.

Westerdahl, H., Wittzell, H. & von Schantz, T. (1999). Polymorphism and transcription of Mhc class I genes in a passerine bird, the great reed warbler. *Immunogenetics*, 49, 158-170.

Wood, M.J., Cosgrove, C.L., Wilkin, T.A., Knowles, S.C.L., Day, K.P. & Sheldon, B.C. (2007). Within-population variation in prevalence and lineage distribution of avian malaria in blue tits, *Cyanistes caeruleus*. *Molecular Ecology*, 16, 3263-3273.

Worley, K., Collet, J., Spurgin, L.G., Cornwallis, C., Pizzari, T. & Richardson, D.S. (2010). MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, 19, 3064-3075.

Yamazaki, K., Boyse, E.A., Mike, V., Thaler, H.T., Mathieson, B.J., Abbott, J., Boyse, J. & Zayas, Z. (1976). Control of mating preferences in mice by genes of the major histocompatibility complex. *Journal of Experimental Medicine*, 144, 1324–1335.

Yeates, S.E., Einum, S., Fleming, I.A., Megens, H.J., Stet, R.J.M., Hindar, K., Holt, W.V., Van Look, K.J.W. & Gage, M.J.G. (2009) Atlantic salmon eggs favour sperm in competition that have similar major histocompatibility alleles. *Proceedings of the Royal Society B: Biological Sciences*, 276, 559-566.

Zagalska-Neubauer, M., Babik, W., Stuglik, M., Gustafsson, L., Cichoń, M. & Radwan, J. (2010). 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC Evolutionary Biology*, 10, 395.

Zeh, J.A. & Zeh, D.W. (1996) The evolution of polyandry I: intragenomic conflict and genetic incompatability. *Proceedings of the Royal Society B: Biological Sciences,* 263, 1711–1717.

Chapter 2

Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system

BMC Evolutionary Biology (2012) **12,** 68.

Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system

Irem Sepil[1], Hooman K Moghadam[1], Elise Huchard[2,3], Ben C Sheldon[1]

1. Edward Grey Institute, Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, Oxford, UK
2. Behavioral Ecology and Sociobiology Unit, German Primate Centre, Kellnerweg 4, Göttingen 37077, Germany
3. CRC 'Evolution of Social Behavior', Georg-August University, Kellnerweg 6, Göttingen 37077, Germany

Abstract

Background

The critical role of Major Histocompatibility Complex (*Mhc*) genes in disease resistance and their highly polymorphic nature make them exceptional candidates for studies investigating genetic effects on survival, mate choice and conservation. Species that harbor many *Mhc* loci and high allelic diversity are particularly intriguing as they are potentially under strong selection and studies of such species provide valuable information as to the mechanisms maintaining *Mhc* diversity. However comprehensive genotyping of complex multilocus systems has been a major challenge to date with the result that little is known about the consequences of this complexity in terms of fitness effects and disease resistance.

Results

In this study, we genotyped the *Mhc* class I exon 3 of the great tit (*Parus major*) from two nestbox breeding populations near Oxford, UK that have been monitored for decades. Characterization of *Mhc* class I exon 3 was adopted and bidirectional sequencing was carried using the 454 sequencing platform. Full analysis of sequences through a stepwise variant validation procedure allowed reliable typing of more than 800 great tits based on 214,357 reads; from duplicates we estimated the repeatability of typing as 0.94. A total of 862 alleles were detected, and the presence of at least 16 functional loci was shown - the highest number characterized in a wild bird species. Finally, the functional alleles were grouped into 17 supertypes based on their antigen-binding affinities.

Conclusions

We found extreme complexity at the *Mhc* class I of the great tit both in terms of allelic diversity and gene number. The presence of many functional loci was shown, together with a pseudogene family and putatively non-functional alleles; there was clear evidence that functional alleles were under strong balancing selection. This study is the first step towards an in-depth analysis of this gene complex in this species, which will help understanding how parasite-mediated and sexual selection shape and maintain host genetic variation in nature. We believe that study systems like ours can make important contributions to the field of evolutionary biology and emphasize the necessity of integrating long-term field-based studies with detailed genetic analysis to unravel complex evolutionary processes.

Background

Genes of the *Mhc* encode cell-surface proteins responsible for the recognition and presentation of foreign antigens to T-lymphocytes, which then initiate an immune response against pathogens [1]. The *Mhc* is known to be the most variable gene group in vertebrates, both in

terms of allelic diversity and gene number [2]. Numerous studies have suggested that selection from parasites is the major factor driving this polymorphism, whereas the role of sexual selection has also been demonstrated in several species [3-5]. Interest in understanding *Mhc* genes is growing significantly as more studies demonstrate implications of their diversity in the context of survival [6], mate choice [7], conservation [8] and even speciation [9]. The number of *Mhc* genes can differ greatly between and within species, and may show no sequence orthology between genera or orders – e.g. class I gene number and organization differences between human and chimpanzees (reviewed in [10]). Differences in the number and organization of *Mhc* genes are explained by a birth and death model of evolution [11]. According to this model, new genes are formed by duplication events and, while some retain their function in the genome, others are inactivated or deleted [12]. Because an individual *Mhc* molecule can recognize and bind to a few antigens only, which is determined by the amino acid composition in their antigen-binding site (ABS), gene duplications and polymorphism at the ABS are thought to be adaptations enabling individuals to respond to a greater variety of antigens [13].

The mammalian *Mhc* is a gene-dense region that is divided into subfamilies with similar function. Class I and class II genes are considered 'classical' *Mhc* genes to indicate their polymorphic, highly expressed, structure; and antigen processing and presenting function. In sharp contrast to mammals with multiple class I and II loci, the avian model organism, the chicken has a "minimal essential *Mhc*" a term referring to its surprisingly small, densely packed *Mhc* region with only two class I and II genes [14,15]. However, recent studies have shown that "minimal essential *Mhc*" does not hold for other avian orders, as a larger number of class I and II genes have been detected in passerines [16]. The recent in-depth characterization of zebra finch *Mhc* demonstrated that more than ten *Mhc* class I and class II loci are located on different chromosomes in this species ([17], but see [18]). Many of the genes were suspected to be pseudogenes and only one *Mhc* class I locus appeared to be transcribed, contrasting to the pattern observed in other passerine species, where many functional class I loci have been

described [19-22]. Similarly the recent characterization of *Mhc* class II loci exon 2 in the collared flycatcher revealed the presence of at least nine expressed (and a larger number of pseudogene) loci, confirming the complexity and variable evolutionary dynamics of this region in Passeriformes [23].

The presence of multiple loci and high allelic diversity in passerines make them intriguing species for *Mhc* studies as they are potentially under strong selection, and thus can provide valuable information on the mechanisms maintaining *Mhc* polymorphism. However the complexity of multilocus *Mhc* genes also poses a great challenge for accurate genotyping. In species where the *Mhc* has undergone rapid duplication, gene conversion and recombination, the genes are often tightly linked to each other and alleles are shared among loci, making it almost impossible to identify and isolate independent loci. Therefore simultaneous amplification of multiple loci becomes necessary in such systems [24].

It is surprising that until recently the importance of accurate genotyping has been somewhat neglected in species harbouring complex multilocus *Mhc* systems [25]. Initial characterization of *Mhc* genes is essential to fully understand the architecture of the system and to design primers that would potentially amplify all the alleles of interest, so that precise genotyping could be achieved [26]. Moreover the fact that some alleles might be nonfunctional further complicates the situation and obligates the differentiation of functional alleles from the non-functional ones through historical selection tests and screening of cDNA libraries. The method chosen for *Mhc* genotyping is also of great importance. There are a few indirect genotyping techniques that have been widely used to date although these methods require extensive initial optimization, often necessitate further characterization via cloning and sequencing, and are still insufficient when genotyping complex multilocus systems (reviewed in [27,28]). The application of Next Generation Sequencing (NGS) methods for *Mhc* screening has recently been introduced as a promising alternative to the conservative typing methods and has the potential to provide high resolution, accurate and large-scale genotyping, thus solving three problems that have plagued previous studies in this field [24,28,29]. Especially in projects

where sample sizes are high, 454-pyrosequencing may outperform the other methods both in terms of the time it takes and per sample cost. One major drawback of the application of high-throughput NGS is its error-prone sequencing technology: therefore strict quality control is crucial to distinguish true alleles from sequencing artefacts (reviewed in [25]). To date relatively few published studies have used 454-pyrosequencing method for *Mhc* genotyping in non-model vertebrates; all succeeded in identifying real and artificial sequences. The repeatability of the results was confirmed for all but one, by cloning or running a fraction of the samples in duplicates [24,28,30]. In one study, that by Zagalska-Neubauer et al. [23], the repeatability of the duplicate samples was low, because the coverage of reads was not sufficient for reliable genotyping. However this did not prevent full characterization of diversity at species level.

These recent studies demonstrate that 454 or similar technologies can be employed for accurate typing of complex *Mhc* loci. Reliable genotyping of multilocus *Mhc* alleles is, however, still not sufficient for determining what equates to functionally important alleles. *Mhc* alleles commonly share the same ABS aminoacids and have similar antigen-binding motifs; hence their allelic differences are not always functionally relevant (discussed in [26]). Therefore a growing body of evidence across taxa highlights the importance of clustering *Mhc* alleles into functional supertypes and considering these supertypes as the unit of selection [31-34]. For instance Trachtenberg et al. [35] showed that rare human leukocyte antigen (HLA) supertypes confer a strong advantage in responding to HIV infection, independent of the contribution of single alleles.

Here, we describe an approach for characterizing and genotyping *Mhc* class I loci exon 3 in natural great tit (*Parus major*) populations. Genes of *Mhc* class I have been shown to influence avian malaria resistance [36-39], female mate choice [40] and extra-pair paternity [41,42] in other passerine species. Initially we undertook a comprehensive characterization effort to detect the diversity of allelic variants at exon 3, and used 454 pyrosequencing for genotyping. Using this technique we genotyped a fragment of *Mhc* class I exon 3 of several

hundred great tits simultaneously; and applied a stepwise variant validation procedure to separate true alleles (the procedure was based on the methods defined by Zagalska-Neubauer et al. [23] and Galan et al. [28]). Then we differentiated the functional alleles from the nonfunctional ones, first by investigating the presence of stop codons in sequences and examining the phylogenetic relationship between allelic clusters, and secondly through historical selection tests. Finally we clustered the *Mhc* alleles into functional supertypes, to determine the biologically meaningful *Mhc* variants of each individual. Our main aim was to describe an approach for accurate *Mhc* characterization and typing, a crucial prerequisite for studies investigating the role and importance of complex *Mhc* loci in non-model vertebrates.

Methods

Samples and DNA extractions

We analyzed 1496 first year or adult great tits, sampled between 2006–2010 in Wytham Woods and Bagley Woods nest-box breeding populations. Both populations have been monitored continuously since the 1960s and 1990s respectively, and are located within a few km of Oxford, UK. All great tits were ringed with aluminium bands for individual recognition. Blood was collected by wing or jugular venipuncture under UK Home Office licence (PPL 30/2409), and stored in SET Buffer at −80°C until DNA extraction. Total genomic DNA was extracted using standard ammonium acetate method and stored in AE Buffer (Qiagen). DNA concentration was measured using a Picogreen assay (Quant-iT Picogreen dsDNA Assay Kit, Invitrogen) and samples were diluted to 5-20 ng/µl when necessary.

Development of primers

We used degenerate primers HN34 and HN45 [43], as well as primers T3F and GTJSF in the 3′ direction and GBT3R and GTJSR in the 5′ direction, identified from the alignment of tit sequences available in NCBI (great tits [GenBank: AF346821– AF346832], green-backed tits

[GenBank: EF446972–EF446988] and blue tits [GenBank: AM232705– AM232717]) to isolate complete and partial exon 3 sequences. Then we designed specific primers within this exon using the consensus from these sequences (see Figure 2.1 and Table 2.1 for the sequence and location of each primer). *Mhc* class I exon 3 was chosen as the target region, as it bears most of the ABS. PCR reactions were run using Platinum® Taq DNA Polymerase Kit (Invitrogen). PCR amplification was performed in 25 µl reactions, with the following final concentrations: 1x PCR Buffer, 1.5 mM MgCl2, 0.5 µM of each primer (forward and reverse), 0.5 units of Platinum Taq Polymerase (Invitrogen) and 0.125 mM of each dNTP. 2 µl of extracted DNA (5-20 ng/µl) was added to this mix. The reaction was run for 30 cycles at 95°C for 45 s, 57°C-65°C for 45 s and 72°C for 45 s on a GeneAmp PCR System 9700 thermal cycler (Applied Biosystems). PCR products were purified using MinElute 96 UF PCR Purification Kit (Qiagen), ligated in plasmid vector and transformed to bacteria using pCR8/GW/TOPO TA Cloning Kit (Invitrogen). Each primer pair was tested on four individuals – two males (first year and adult) and two females (first year and adult) that were sampled in different areas of the population. Twelve clones from each individual were randomly selected and reamplified using the same PCR conditions. The amplicons were sequenced directly by dye terminator cycle sequencing (BigDye version 3.1) and loaded on an ABI PRISM 310 Automated Sequencer (Applied Biosystems). The sequences were edited in Sequencher version 4.2 (GeneCode) [44] and aligned using BioEdit Sequence Alignment Editor [45].

After isolating a large number of variable *Mhc* class I sequences, we used the consensus from the obtained sequences to design a degenerate primer pair (MHCD-F and MHCD-R) that would potentially amplify all the expressed alleles that are present in the great tit. By using degenerate primers we aimed to minimize the risk of missing out a fraction of allelic variation that might bias the results of the studies to follow. The primers amplified a 212–221 basepair fragment (without primers) of *Mhc* class I exon 3 (78%) in the great tit. Four individuals were amplified with the primers under the same PCR conditions and 65°C for annealing temperature. The PCR products were cleaned, cloned and sequenced as described above.

RNA extraction, cDNA synthesis, RT PCR

Total RNA was extracted from four great tit blood samples for expression analysis. Blood samples collected via jugular venipuncture were immediately placed into RNAprotect Animal Blood Tubes (Qiagen) for RNA stabilization and were stored at 4°C overnight. RNA isolation was done using RNeasy Protect Animal Blood Kit (Qiagen) and reverse transcription was carried out with Omniscript Reverse Transcription Kit (Qiagen). First strand cDNA was synthesized in a final volume of 20 µl with the following conditions: 1x Buffer RT, 1 µM of Oligo-dT primer, 10 units of RNase inhibitor, 4 units of Omniscript Reverse Transcriptase, 10.5 mM of each dNTP and 1 µl of template RNA. The mixture was incubated for 60 min at 37°C and a 2 µl aliquot of the finished reverse-transcription reaction was used for Two-Tube RT-PCR. We used four different primer pairs for RT-PCR (Table 2.1) and adjusted the same PCR conditions that we described for genomic DNA amplification. Again the PCR products were cleaned, cloned and sequenced.

Tagged primer design; amplicon preparation, dilution and pooling

454 pyrosequencing was performed on 1532 genomic DNA samples from 1492 great tits, collected between 2006–2010 in Wytham and Bagley Woods; and cDNA libraries of four Wytham Wood great tits. The individuals with cDNA libraries were not included in the genomic DNA sequencing panel, since the RNA stabilization step prevented DNA isolation from these samples. Forty genomic DNA samples were randomly chosen and ran in duplicates to estimate the repeatability of the results. In order to maximize the throughput for 1536 great tit samples from the bidirectional sequencing run, we used the 16 regions of the Pico Titer Plate individually (the maximum physical separation currently available). Therefore the samples were divided into 16 sets of 96 individuals and individuals in each region were differentiated by the use of multiplex identifiers (MID).

MHCD-F and MHCD-R were chosen as the template-specific primers since the aim was to amplify all the functional allelic variation present in the population. Forward (5'-

CGTATCGCCTCCCTCGCGCCATCAG - MID - TTMYGGCTGTGACCTCCTG-3') and reverse (5'-CTATGCGCCTTGCCAGCCCGCTCAG - MID -TTGCGCTYCAGCTCTTTC - 3') fusion primers were designed by adding a GS FLX Titanium Primer sequence and a 10bp MID sequence to the 5' end of the template specific primers (underlined) [46]. The MID's are the sequence tags that allow individual identification; when both forward and reverse primers are tagged, individual identification can easily be achieved by selecting different combinations of MID's for each sample. Hence a small number of MID's were used to uniquely barcode a large number of individuals [28]. We designed ten forward and ten reverse fusion primers using the first ten MID's from the Standard 454 Set. These MID's have been engineered by Roche to avoid misassignment of reads and they are tolerant to several errors [46]. The fusion primers were purified using HPLC to minimize the risk of mispriming events [47]. Recently, Lenz and Becker [48] compared the effects of different PCR conditions on *Mhc* artefact formation and showed that simple adjustments like decreasing PCR cycle number or increasing elongation time within each cycle reduced the number of artefacts. As proposed in [48], we tested the minimum number of cycles needed in this study by agarose gel electrophoresis, and found that bands were visible after 28 cycles. Therefore we adjusted the PCR cycle number to 28 and elongation time to 1 minute, to minimize the formation of PCR artefacts. Amplifications were run in a final volume of 25 µl, including 1x FastStart Buffer #2, 0.4 µM of forward and reverse titanium fusion primers, 0.5 units of FastStart Taq DNA Polymerase (Roche), 0.2 mM of each dNTP and 2 µl of extracted DNA (initial concentration of ~5-20 ng/µl) [49]. PCR conditions included an initial denaturation at 94°C for 3 min, followed by 28 cycles of denaturation at 94°C for 15 sec, annealing at 59°C for 45 sec and extension at 72°C for 1 min with a final extension step at 72°C for 8 min.

PCR products were purified using the MinElute 96 UF PCR Purification Kit (Qiagen) and viewed on agarose gels to estimate the concentration of the amplicons. Kloch et al. [30] found this method to be cheaper, less-time consuming and as satisfactory as Nanodrop measurements. 96 individuals with different tag combinations were pooled together in

approximate equimolar quantities, and a single pool was prepared for each region of the Pico Titer Plate. The pools were then sent for bidirectional 454 pyrosequencing using GS FLX Titanium chemistry at Genomic Services, Wellcome Trust Centre for Human Genetics, University of Oxford.

Bioinformatics and data processing

The experiment was run twice for bidirectional sequencing and the results were merged prior to analysis. The output was initially analyzed using jMHC, a software package specifically designed for *Mhc* amplicon analysis [50]. jMHC allows users to specify the template-specific primers, sequence tags, and the individuals represented by each tag. We used the software to extract the reads bearing complete primers and sequence tags; and to generate a table of all variants, and the number of variants represented in each individual. Therefore jMHC sorted out the reads and assigned them to the individuals, and then removed the sequences lacking complete primers or tags, and sequences that had ambiguous base pairs (Ns).

There are three potential sources of error during PCR and pyrosequencing: (i) PCR-generated mutations; (ii) PCR-generated chimeras and (iii) 454 sequencing errors. Most of the 454 sequencing errors constitute over- or under-calls and are low in frequency, hence they are easy to identify [24]. In contrast PCR-generated errors are harder to distinguish as they might occur early in amplification, can have higher frequencies, and the same chimeras may be produced repeatedly (discussed in [24,25,28]). Considering the high number of artefacts generated during these processes, an efficient quality control was crucial to reliably differentiate real alleles from artefacts. We applied a stepwise variant validation procedure to detect true alleles (Table 2.2); the method is broadly based on the procedures defined by Zagalska-Neubauer et al. [23] and Galan et al. [28].

Variant validation was based on six assumptions: (1) Sequences that have deletions or substitutions shifting the reading frame should be regarded as artefacts; (2) 454 sequencing artefacts should have low copy numbers in the whole dataset and on a per individual basis; (3)

PCR-generated mutations and chimeras could have high copy numbers on per individual bases but should co-occur with the allele or allele pair from which they originated [23,48], and should have low copy numbers in the whole dataset; (4) Verified sequences should be represented at least twice in an individual; (5) Verified sequences should be found in at least two individuals; and (6) Reliability of individual genotypes should be ensured by requiring a minimum number of sequences per individual.

Recently, Galan et al. [28] developed a probabilistic model to validate *Mhc* variants and to determine the confidence level of genotyping for each individual. The confidence level (*f*) depended on the values of *r*, *n* and *m*; *r* being the minimum copy number of each true variant, *n* the total number of sequences and, *m* the maximal number of variants for the gene; while the program permitted the presence of 1–4 loci. Although the number of *Mhc* class I loci in great tit was unknown and most likely exceeded 4 loci, we used the program 'Negative Multinomial' to estimate the minimum number of sequences required per individual for reliable genotyping and calculated the number of sequences that are necessary for amplifying all the variants at least twice ($r = 2$), giving a confidence level (f) of 0.95 [28]. We found that the minimum number of sequences (n) required per individual was 9, 23, 39 and 55 for 1, 2, 3 and 4 loci (m) respectively. The increase in sequence number was linear with respect to locus number; therefore we estimated that 151–199 sequences would be necessary to genotype 10–13 loci. Previous passerine studies demonstrated the presence of 5–10 *Mhc* class I loci [17,20,43]; hence we chose 200 reads for minimum coverage and two copies per individual for keeping a variant in our study. In Galan et al. [28] the minimum copy number of each true variant was set as three ($r = 3$) and the confidence level was set as 99.9% ($f = 0.999$), hence the minimal number of sequences required for reliable genotyping was much higher, compared to our estimates. Although it would have been ideal to apply the same criteria considering its robustness, in the present study we modified the values because the average number of reads per individual was lower than expected. Therefore we lowered the minimum copy number of each true variant to two ($r = 2$) and the confidence level to 95% ($f = 0.95$) to balance the tradeoff between read

quality and size of the dataset. Keeping variants that occur twice in an individual is relatively conservative, but it ensures the retention of real variants and allows the elimination of artefacts in the following steps. It can also be argued that assumptions 4 and 5 are not strictly true and might lead to the removal of rare variants. However we believe that it is better to be conservative and risk missing a small fraction of real alleles, rather than risking the inclusion of artefacts. Therefore our approach was in line with the standard two-PCR criterion – to be verified a sequence should be retrieved at least twice from independent PCR reactions - that *Mhc* studies traditionally use.

The sequence validation procedure had five steps (see Table 2.2 and Figure 2.2):

(1) Reads that did not match the expected allele sizes were eliminated (Assumption 1)

(2) Reads that occurred only one, two or three times in the whole dataset were removed (Assumption 4 and 5)

(3) Individuals with fewer than 200 reads were removed (Assumption 6)

(4) Artefacts were identified at the level of the whole dataset by applying the following criteria:

Maximum per amplicon frequency (MPAF) was determined for each unique variant [23]. All the variants that had MPAF equal to 0.01 (for these variants the highest within individual frequency was 1%) were examined to detect whether they were real alleles or artefacts. The decision was based on the three individuals that had the highest MPAF for the variant in question. If the variant could be explained as a chimera of more common alleles, or if it differed by a single basepair from more common alleles in all three individuals, it was eliminated as an artefact. Otherwise it was retained as a real allele (Assumption 2 and 3). A total of 82 out of 88 variants (93.2%) with 0.01 MPAF were detected as artefacts and deleted. Then we randomly selected 50 variants with MPAF smaller than 0.01 to check their status; all were identified as artefacts. Therefore we deleted all the variants below 0.01 MPAF. The variants above 0.01 MPAF were examined and the following pattern was detected: 83.4% of the variants (226 out of

271) that had MPAF between 0.011-0.015 were detected as artefacts; 34.5% of the variants (30 out of 87) that had MPAF between 0.016-0.02 were detected as artefacts; but only 2.8% of the variants (2 out of 72) that had MPAF between 0.021-0.025 MPAF were detected as artefacts. The rest of the variants were not checked to prevent false negatives – assigning real alleles as artefacts due to sequence similarities. However we randomly chose 50 variants with MPAF larger than 0.025 and all were identified as real alleles. In total we found and removed 340 artefacts on the scale of the whole dataset. Of these, 96.5% (328 artefacts) had two, three or four copies within the individual, whereas 3.5% (12 artefacts) had five or six copies. Therefore variants that have two, three and four copies within an individual represented the "grey area" as these sequences were likely to be artefacts.

(5) Artefacts were distinguished at the individual level by applying the following criteria:

All single copy variants within an individual were deleted (Assumption 4). All variants represented by two, three and four copies were examined to detect whether they could be explained as a chimera of more common variants or whether they differed by a single basepair from more common variants, and were deleted if the answer was yes (Assumption 3). Variants with more than four copies were retained as true alleles.

The fourth and fifth steps of the sequence validation procedure were based on the method developed by Zagalska-Neubauer et al. [23]. These steps involved visual analysis of variants that have low read number in the whole dataset and on a per individual basis, and were critical for distinguishing true alleles from artefacts. Although we increased the risk of introducing artefacts in Step 3 by choosing 200 reads for minimum coverage and two copies per individual for keeping a variant, the detailed inspection of variants in step 4 and 5 enabled us to detect and eliminate errors, hence we believe that the reliability of the genotyping method improved substantially following these two steps.

Phylogeny construction

A phylogenetic tree of great tit *Mhc* class I exon 3 variants were constructed using Neighbour-Joining method and Tamura-Nei model. The tree was rooted with a chicken (*Gallus gallus*) *Mhc* class I sequence [GenBank: AY234770] and the reliability of the branches was tested with 1000 bootstrap replicates. The phylogenetic tree was constructed using MEGA 5 [51]. The nucleotide diversity between sequences was analyzed using DnaSP ver. 5.10.00 [52].

Tests of historical selection

The strength of historical selection acting on great tit *Mhc* was tested using a likelihood ratio modelling approach. Two models of codon evolution were compared: the nearly neutral model (M7) where $dN/dS < 1$; and the positive selection model (M8) where the presence of $dN/dS > 1$ sites are allowed. If M8 fits the data better than M7, positively selected sites (PSS) were identified through the Bayes empirical Bayes (BEB) procedure. The analysis was computed using CodeML, implemented in the software PAML [53]. The frequencies of dN and dS were then estimated using a second approach: Nei-Gojobari method and Jukes-Cantor correction. The two approaches incorporate different evolutionary models; hence codon based Z-test of selection was conducted on (a) PSS, (b) non-PSS, (c) ABS, (d) non-ABS and (e) all sites, for comparative purposes [31]. The aminoacids corresponding to ABS were detected by superimposing the chicken major *Mhc* class I sequences and assuming concordance at the ABS [54]. The Z-test of selection estimates dN-dS and computes a 1- tailed test to determine if $dN > dS$. For sites predicted to be in contact with antigens, a significant $dN > dS$ (positive selection) would be expected. Analyses were computed using MEGA 5 [51].

Definition of supertypes

Mhc alleles with similar antigen-binding motifs were clustered into supertypes as proposed by Doytchinova and Flower [55]. Initially the aminoacid sequences of all PSS were aligned and the rest of the sites were removed [31,32]. Then each PSS aminoacid was characterized by five

physicochemical descriptor variables: *z1* (hydrophobicity), *z2* (steric bulk), *z3* (polarity), *z4* and *z5* (electronic effects) and translated into a matrix [56]. This matrix was subjected to K-means clustering algorithm and model selection for identifying genetic clusters, and discriminant analysis of principal components (DAPC) for describing the clusters using 'adegenet' package in R [57,58]. Initially, the function 'find.clusters' was used; K-means was run sequentially with increasing number of groups, and different clustering solutions were compared via Bayesian Information Criterion (BIC) values (reviewed in [58]). The number of clusters was then chosen based on the graph of BIC values for increasing number of clusters. The optimal number of supertypes was defined as the minimal number of clusters after which the BIC decreases by a negligible amount and was indicated by an elbow in the curve of BIC values as a function of cluster number. Once the number of clusters was chosen, we applied DAPC to visualize the relationship between the supertypes. The function 'dapc' was used to perform a discriminant analysis on the retained principal components (PCs); and a DAPC scatterplot was displayed to illustrate the differences between the clusters using the first two PCs [58].

Results

Characterization of great tit Mhc class I exon 3

Amplification of genomic DNA using nine different primer pair combinations yielded three distinct PCR products that differed at the length of exon 3 (282 bp, 276 bp or 273 bp). Alleles of similar sizes generally grouped together within the phylogenetic tree, however they were not organized in well supported clusters so it was not possible to identify exact gene families. In total 66 highly divergent alleles were retrieved from 12 great tits (nucleotide diversity = 0.114 ± 0.004). Only three of these alleles were clearly non-functional: their amino acid translation bore two stop codons, and possibly belonged to a pseudogene, as they formed a monophyletic cluster (Figure S2.1). Some individuals had up to nine alleles so we estimated that (on the basis of these data alone) the great tit had at least five *Mhc* class I loci, of which one is a pseudogene.

In order to understand which allelic clusters were functional and expressed, we amplified transcribed *Mhc* alleles from mRNA extracted from blood. Contrary to our expectations, the cDNA sequences were not confined to any allelic cluster and were distributed throughout the phylogeny (Figure S2.1). Moreover we found that even the pseudogene alleles were being transcribed, as they were present in the cDNA library. Since it was not possible to determine an allelic subgroup of interest for *Mhc* genotyping, we designed the degenerate primers MHCD-F and MHCD-R that would potentially amplify all the functional variation present in great tits. The primers were specifically designed to prevent a good match with the pseudogene alleles, so that the coverage of non-functional sequences could be limited while amplifying the rest of the alleles. The PCR products were 221, 215 and 212 basepair long and covered the major part of *Mhc* class I exon 3 (78%). The primers succeeded in amplifying alleles from each well-supported allelic cluster, suggesting that full scale sequencing of *Mhc* class I alleles could be attained with their use.

454 Genotyping

Genomic DNA from 1492 individuals, 40 of them amplified twice, and cDNA libraries from four individuals were amplified using the ten forward and ten reverse Titanium Fusion Primers we designed. Bidirectional sequencing was employed using the 16 regions of a Pico Titer Plate gasket. Overall, the experiments generated 638,501 reads. jMHC removed the imperfect sequences reducing the read number to 439,284 (68.8% of the previous step) with a mean of 286 reads per individual and a median of 252. A high level of variation was observed in the number of reads per individual (standard deviation of read number = 173; range = 4–1219). While more than one quarter of the variation in read number was introduced by the 454 region on which a sample was run (due to optimization problems experienced by the sequencing facility), some variation was also introduced by the concentration differences between purified PCR products.

First we eliminated reads that were not 221, 215 or 212 basepair long, since the characterization phase allowed us to define the expected allele sizes, leaving 366,231 reads (83.4% of the previous step) and 83,495 unique variants. Reads that occurred only once, twice or three times in the whole database were then eliminated, reducing the dataset to 332,135 reads (90.7% of the previous step) and 4,674 unique variants. At this stage we removed all the individuals that had fewer than 200 reads, as their genotypes were considered incomplete and thus unreliable. Lastly, artefacts were differentiated at the level of the whole dataset and the individual, resulting in retention of only 862 variants as true alleles. Overall, 871 samples passed our criteria: 857 individuals, 12 duplicates and two cDNA amplicons; the final genotypes were based on 214,357 reads (64.5% of the previous step).

Among the 40 samples genotyped twice only 12 duplicates passed our variant validation criteria, and the reliability of the experiment was calculated based on these samples. We removed the pseudogene alleles that bore a stop codon prior to calculation, because the primers were designed to avoid amplification of such alleles and they were functionally irrelevant. For each duplicate we calculated the agreement between the genotypes after the 3rd, 4th and 5th step of the variant validation procedure, to verify the efficiency of the method (see Table 2.3). We also calculated the agreement between the genotypes following supertype classification (see Table 2.3, last column), since supertypes will be considered as the unit of selection in subsequent analyses investigating the effects of *Mhc* variation on parasite burdens, survival, reproductive success and mate choice in this system. On average the repeatability of the duplicates increased from 0.34 to 0.94 between step 3 and step 5; 22 alleles were common in the duplicate, while 1.4 alleles occurred only in one duplicate. Of the 12 individuals six had identical genotypes following variant validation. A repeatability score was calculated for each duplicate and the scores were averaged to estimate the repeatability of the experiment. Hence the repeatability of the genotyping method was calculated as 94%. The lack of significant correlation between read number and allele number per individual, above the threshold of 200 reads (Figure 2.3) further supported the reliability of the experiment ($R^2 = 0.0038$, $p = 0.07$).

The agreement between genotypes increased from 0.94 to 0.96 after supertype classification; on average 10 supertypes were common in the duplicate, whereas 0.4 supertypes occurred in one duplicate. Of the 12 individuals eight had identical genotypes following supertyping and we found no correlation between the read number and supertype number per individual (R^2 = 0.0006, p = 0.45) (Figure S2.2).

Sequence diversity

In this study a total of 862 *Mhc* class I alleles [GenBank: JQ034624 - JQ035485] were detected, to our knowledge the highest number characterized in a wild bird species (Figure 2.4). The sequences were highly divergent with 133 polymorphic nucleotides, representing 60% (133/221) of the sites. Nucleotide diversity and the average number of nucleotide differences were π = 0.106 ± 0.001 and 22.59 ± 0.23 respectively. Of 66 alleles we retrieved in the characterization stage, fifty were present among these sequences (76%). Within the sample of 862 alleles, 39 were non-functional alleles bearing stop codons: 36 of these alleles had two stop codons at the same location, were highly similar to each other and formed a paraphyletic group suggesting the presence of a pseudogene family (Figure S2.3). Moreover these sequences formed a monophyletic cluster with 68 alleles that did not bear stop codons, but showed a dissimilar pattern of divergence from the rest of the alleles. These 68 sequences are referred as Group 1, while the remaining 758 sequences are referred as Group 2 (see Figure 2.4). The members of Group 1 were 215 basepairs long, similar to each other, and structurally similar to the pseudogene alleles variants (Figure S2.3). The segregating sites within Group 1 were considerably different than the segregating sites in Group 2, implying that separate evolutionary forces might be shaping this variation and that Group 1 sequences might be representing non-functional variants. Alleles of Group 2 were, on the other hand, highly divergent from each other, had similar segregating sites and varied in their size (221, 215 and 212 basepair long). Three non-functional alleles bearing stop codons (of the 39) were within Group 2, but unlike the members of the pseudogene family these sequences bore only one stop codon and were not

clustered together, implying that they evolved independent from each other (Figure S2.1). These alleles were probably formed recently as a result of mutation accumulations, and did not belong to a pseudogene.

Two cDNA amplicons passed our variant validation criteria and 42 alleles were retrieved in total. We checked whether the cDNA alleles were confined to particular allelic clusters; however, in line with our previous findings, the sequences were found throughout the phylogenetic tree (including the pseudogene family) and did not fall into any cluster. Hence the sequence variants obtained from cDNA were not useful in differentiating functional groups from the non-functional ones.

Signatures of historical selection

Tests of historical selection were performed separately on Group 1 and Group 2 alleles in order to clarify whether they were under different selective pressures. First we used a likelihood ratio modelling approach to compare the two models of codon evolution (nearly neutral versus positive selection) and selected the model that best fitted the data (see Table 2.4 for the results of the likelihood ratio tests). For Group 1 the model allowing a class of sites to be under positive selection (M8) fitted the data better than the null nearly neutral model (M7). Therefore M7 was rejected in favour of M8, and three sites were found to be under positive selection according to Bayes empirical analysis ($P < 0.05$ for all three sites). However none of these sites were identical to the chicken ABS, and only one was situated close to an ABS with a distance of two aminoacids (Figure 2.5a). For Group 2 it was not possible to run the analysis on all of the alleles because the number of sequences exceeded the capacity of CodeML software. Therefore we selected all the sequences that were present in more than 20 individuals as being common alleles, confirmed that they were equally distributed throughout the phylogeny (Figure S2.4) and ran the analyses on these 117 sequences. Again M8 fitted the data better and nine sites were detected as PSS ($P < 0.05$ for all nine sites). Six PSS were identical to the chicken ABS, suggesting full agreement (Figure 2.5b). The remaining three PSS had high aminoacid

variability and might have been sites of peptide or T cell binding. None of the PSS identified in Group 2 were identical with the PSS in Group 1, confirming that these two groups are under different selection pressures, and hence probably differ in their functionality.

We also performed codon based Z-test of selection on Group 1 and Group 2 separately (Table 2.5). Here we checked whether PSS, non-PSS, chicken ABS, chicken non-ABS and all sites were under selection to validate the results of the previous analysis. As expected, none of the sites were found to be under positive selection in Group 1, whereas both PSS and chicken ABS (sites predicted to be involved in antigen recognition) were positively selected in Group 2.

Supertypes

Group 2 sequences were clustered in functional supertypes, using the nine sites detected as PSS. We excluded the alleles bearing stop codons and the Group 1 sequences because they were considered non-functional. We created a matrix for the 775 Group 2 alleles based on their PSS physicochemical properties; and subjected the matrix to K-means clustering algorithm and model selection. The optimal number of clusters was indicated by a change in the slope of BIC decrease (the elbow in the curve) and was identified as 17 (Figure S2.5). Allele number per supertype ranged from 11 to 87 alleles with a mean of 44 (Figure 2.6a), and the population-wide frequency of supertypes ranged from 0.27 to 0.99 (Figure 2.6b). Once the number of clusters was chosen, we applied DAPC to visualize the relationship between the 17 supertypes. For this analysis 12 principal components (PCs) were kept to retain 93% of the variation. The DAPC scatterplot summarizes the differences between the clusters using the first two PCs (Figure 2.7).

Estimating the number of Mhc class I loci

The maximum number of alleles per individual was found to be 37 with a mean of 23.8 ± 3.9, suggesting that there are at least 19 *Mhc* class I loci in the great tit (Table 2.6). However some of these alleles belonged to a pseudogene, whereas others were considered non-functional (Group 1 alleles). We found a maximum of 32 functional alleles per individual with a mean of

19.5 ± 3.7. Therefore we estimate that the great tit has at least 16 functional loci. When functionally similar alleles were clustered in discrete supertypes, each individual had between 6–16 supertypes with a mean of 10.4 ± 1.6; the number of supertypes was approximately normally distributed in the population. Based on the data, the number of pseudo and putatively non-functional genes were also estimated. These numbers are probably an underestimate as the primers were designed to prevent a good fit with the alleles bearing a stop codon, and the putatively non-functional alleles were similar to the pseudogene alleles. Nevertheless, we found a maximum of six pseudogene alleles and six putatively non-functional alleles per individual, pointing to the presence of at least three loci in each group.

Discussion

In this study, characterization of great tit *Mhc* class I exon 3 was adopted and large-scale, high-resolution genotyping was carried out by the use of 454 pyrosequencing method. A total of 862 alleles with varying sizes and high nucleotide diversity were detected in 857 great tits demonstrating that this species has a polymorphic, multilocus *Mhc* region; tests of historical selection implied this polymorphism was mainly maintained by balancing selection. Full analysis of sequences through a stepwise variant validation procedure allowed reliable typing of 871 samples and duplicates confirmed the repeatability of the genotyping method. Lastly, the effective *Mhc* repertoire of each great tit was described by clustering alleles with similar antigen-binding affinities into 17 supertypes.

Our characterization effort proved to be effective as it enabled us to define the allelic groups of interest and the sequences to prevent (pseudogenes) during 454 pyrosequencing. High diversity among the potentially functional alleles led to the design of a degenerate primer pair, however attention was paid to minimize amplification of alleles with stop codons. Consequently, functional sequences constituted the major fraction of the reads (87%) that were retained following variant validation procedure, whereas the number of reads belonging to

putatively non-functional genes were much lower. This highlights the importance of carrying out a thorough background work prior multilocus typing, not only to attain an initial understanding of gene and allelic diversity, but also to design the ideal primers that target the expressed alleles only. For instance, Zagalska-Neubauer et al. [23] used 454 pyrosequencing to type the *Mhc* class II region of the collared flycatcher and the experiment successfully generated a mean of 541 reads per individual. However the coverage was insufficient for reliable genotyping since the primers extensively amplified the putative *Mhc* pseudogenes, greatly decreasing the coverage of expressed alleles.

We retrieved 76% of the alleles identified during the characterization stage. These were randomly distributed in the phylogeny and represented each allelic group, suggesting that we successfully amplified the allelic groups of interest. Few of the alleles that were not retrieved belonged to Group 1 or the pseudogene group as expected. The rest of the alleles identified at the initial characterization stage were most likely PCR-generated errors that were not differentiated at the characterization stage as they were either 1 basepair different from more common alleles or they could be explained as a chimera of more common alleles. Lenz and Becker [48] showed that PCR generated artefacts could constitute up to 25 percent of the alleles when no approach is taken to reduce artefact formation. Although we adjusted the PCR conditions during amplification for 454 genotyping, we used the standard conditions during characterization. Hence it would not be surprising to have generated a relatively high frequency of artefacts during the cloning –sequencing processes.

We used the 16 regions of the Pico Titer Plate and pooled 96 great tit amplicons in each region to maximize the number of individuals sequenced in the experiment. However from the total of 1536, only 871 samples (56%) were successfully genotyped following the variant validation procedure. The efficiency of the experiment could have been improved by optimising the sample number for a full plate run [28]. In amplicon sequencing, the depth of coverage is tightly linked to the number of amplicons being pooled. Therefore by pooling a smaller number of amplicons, we could have genotyped more individuals. However, it is rather difficult to

determine the optimal number of samples to use in a 454 run, especially in cases where the gene copy number is initially unknown. Alternatively, our method could have been improved by calculating the concentration of each amplicon using a Nanodrop, in order to assure the pooling of equimolar quantities [24]. This way the variation in read number per sample could have been minimized and possibly more individuals would have passed the minimum coverage threshold. Due to the error-prone sequencing technology of NGS, we applied a five-step variant validation procedure to differentiate real alleles from PCR/sequencing artefacts. It can be argued that some rare alleles might have been missed in few individuals (especially in the ones that had read numbers slightly exceeding 200) while these steps were being applied. However we believe this is unlikely to be a very large effect given that no significant correlation between read number and allele number per individual was found. Moreover the consistency of the genotyping method and the variant validation criteria was confirmed by the high repeatability (0.94) of duplicates, and by the substantial increase in the repeatability measures between the 3rd and 5th step of variant validation procedure.

Still, of the 12 duplicates only six had identical genotypes following variant validation. Therefore it is plausible to suggest that the quality of the genotyping method could have been improved by modifying the variant validation procedure. As mentioned in the methods section, we tried to maintain a balance between the quality and size of the dataset whilst setting up the genotyping criteria; hence a better experimental design would have maximized both measures. The inconsistency among samples in six of the duplicates was mostly due to removal of real alleles, because the variants in question had low frequency on per individual basis and were highly similar to more common variants. Although the removal of rare alleles poses a problem for accurate genotyping, the application of supertype classification alleviated the issue to some extent, since highly similar alleles were grouped into the same supertype and the contribution of single alleles became irrelevant following supertyping. Hence, of the 12 duplicates eight had identical genotypes after supertyping, no correlation was found between the read number and supertype number per individual and the agreement between duplicates was calculated as 0.96.

To date few *Mhc* studies carried out an empirical assessment of genotyping error via running duplicates in non-model vertebrates [24], therefore it is hard to assess how good the agreement is, but we believe that in the context of ecological studies, this level of repeatability is sufficient.

Four great tit cDNA libraries were constructed from mRNA sequences extracted from blood, and the libraries were used to identify transcribed allelic clusters. However a few cDNA sequences were retrieved from each allelic cluster, so it was not possible to isolate any group as more functionally important. Surprisingly, pseudogene alleles that bore stop codons were also found in the cDNA libraries. This emphasized the fact that classical sequencing methods are limited and even problematic if used for expression analysis, as they only inform on the presence or absence of transcripts. Such data are insufficient, because even non-functional genes could be transcribed in low levels unless a mutation at the promotor region halts the transcription completely. Therefore an approach based on expression level quantification (for instance using real-time PCR) should be adopted in order to differentiate highly expressed, hence functionally important alleles [26]. Assessment of gene expression through cDNA coverage has started to find wider application in NGS methods [59].

Alternatively, the presence of non-functional cDNA sequences can also be explained by genomic DNA contamination during RNA extraction. However we believe this is unlikely, because the method used for RNA purification (RNeasy Protect Animal Blood) included a DNase digestion step. Although mRNA data did not inform us on allelic expression, the phylogenetic tree of the class I alleles and the historical selection tests significantly improved our understanding of functional alleles. The Group 1 sequences formed a monophyletic cluster with pseudogene alleles and codon based Z-test of selection suggested none of the putative antigen-binding sites were positively selected. Still we detected a weak and non-significant signal for the positive selection of Group 1 ABS. Such weak signals are expected, as it has been shown to take around 19–74 million generations for a positive dN-dS signal to disappear even in the absence of selection [60]. Using a likelihood ratio modelling approach three sites were found to be under positive selection in Group 1, yet these were not within highly variable sites.

In contrast all the putative antigen-binding amino acids were under positive selection in Group 2, in addition to three highly variable sites. Moreover Z –test of selection suggested both ABS and PSS were under great selective pressure. These results implied that Group 2 sequences were actively involved in antigen recognition and have been under selection from a variety of parasites; whereas Group 1 sequences were non-functional, hence the observed variation possibly accumulated as a result of relaxed selection.

In total 755 alleles were considered functional and individual great tits possessed 9 to 32 putative expressed alleles. A discrepancy in the number of alleles per individual is common in *Mhc* studies across taxa, due to variation in gene copy number within species and allele-sharing between loci [24,31,61,62]. We cannot exclude the possibility that some alleles might have been missed at the PCR stage, although we believe this to be relatively unlikely considering that we gathered extensive sequence information during characterization, and amplifications were performed using a degenerate primer pair. Similarly sized alleles did not form monophyletic groups implying that 3bp insertions and deletions occurred several times in the history of the gene, independent of each other. The important contribution of indels to *Mhc* genomic diversity has already been shown in chicken [63]. Moreover a comparative study between human and chimpanzee *Mhc* proposed indels as the major driving force for genomic divergence [64]. Alternatively, this pattern might also be an effect of recombination and gene conversion, processes shown to be frequent in passerine *Mhc* systems [20,23].

The alleles of Group 2 were clustered into 17 functional supertypes based on the physicochemical properties of their PSS. We adopted a bioinformatic approach that was described by Doytchinova and Flower [55] to determine peptide specificity; however an experimental approach would have been ideal. For instance, a recent study established computational antigen-binding prediction algorithms based on empirical datasets, and showed an evolutionary advantage for allele pairs that are more divergent in recognizing a broader range of potential antigens [65]. Moreover the biological relevance of grouping *Mhc* alleles with similar antigen-binding affinities into supertypes is supported by a growing number of human

and non-human primate studies [33,35,66,67]. In this study we utilized a newly proposed method for identifying genetic clusters (reviewed in [58]). This method is an advance on previous methods, as it does not require arbitrary clustering decisions [32], but uses K-means clustering algorithm and model selection approach to compute associated summary statistics.

We found evidence for the presence of at least 16 functional loci in the great tit. This constitutes the highest number of expressed *Mhc* class I alleles/loci identified in a passerine species. However we believe it is likely that this complexity is not specific to great tits and that similar patterns can be found in other passerines. Studies on great reed warbler and scarlet rosefinch have already revealed high polymorphism and the existence of more than 5–6 functional *Mhc* class I loci in these species, although these studies had lower sample sizes (248 and 120 respectively) and used conformation-based mutation detection methods that rely on physical separation of alleles, without providing allele sequence information [20,43]. The utilization of indirect typing methods and motif-specific primers can be a rewarding approach in species where functional alleles are well described and confined to an allelic subset. However in species where the *Mhc* structure is complex and the number of co-amplifying alleles is high, like in the great tit, it is impossible to differentiate variants based on their migratory patterns and it is inevitable to underestimate the allelic diversity as a result of peak overlapping [28]. Moreover, because indirect typing methods require cloning-sequencing effort to reveal the nucleotide information, hundreds of clones must be sequenced to obtain a good estimate of allelic composition in such complex systems [25]. Therefore it is feasible to suggest that the employment of 454-pyrosequencing is ideal for passerine species that harbour many *Mhc* loci and high allelic diversity, as it has the potential to reveal the extent of *Mhc* complexity. In a recent study 454 technology was used for *Mhc* genotyping Atlantic cod and it was shown that the species possess more than 100 *Mhc* class I loci, a greatly expanded number compared to other teleost fish [68]. This extraordinary expansion of class I genes was explained by the absence of class II loci; this might represent a compensatory mechanism adapted by the Atlantic

cod immune system. This work illustrates the great flexibility in *Mhc* genomic organization among closely related species.

Conclusions

We found extreme complexity at the *Mhc* class I exon 3 of the great tit following a broad characterization and large-scale, high-resolution genotyping effort. The presence of at least 16 functional loci was shown, together with a pseudogene family and putatively non-functional alleles. There was clear evidence that expressed alleles were under strong balancing selection and these alleles were grouped into 17 functional supertypes based on their antigen-binding affinities. This study represents the first step of an in-depth analysis of the genetic basis of disease resistance, aimed at better understanding how parasite-mediated and sexual selection shape and maintain host genetic variation in nature. We believe that this approach will find wide application among evolutionary biologists in the near future and will allow advancements in our understanding of complex genetic regions.

Acknowledgements

We are grateful to the Wytham and Bagley fieldworkers who collected the data and materials for this study. We thank Judith Mank and Camille Bonneaud for useful comments on laboratory protocols and Alicia Davies for laboratory assistance. We also thank Wieslaw Babik, Colin Garroway, John Mulley, Jordi Paps, Kaan Oz and two anonymous reviewers for helpful discussion and advice. This work was funded by NERC grants (NER/A/S/2002/00877 and NE/F005725/1) to BCS.

References

1. Potts WK, Wakeland EK: **Evolution of diversity at the major histocompatibility complex.** *Trends Ecol Evol* 1990, **5:**408–412.2.

2. Edwards SV, Hedrick PW: **Evolution and ecology of MHC molecules: from genomics to sexual selection.** *Trends Ecol Evol* 1998, **13:**305–311.

3. Bernatchez L, Landry C: **MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years?** *J Evol Biol* 2003, **16:**363–377.

4. Piertney SB, Oliver MK: **The evolutionary ecology of the major histocompatibility complex.** *Heredity* 2006, **96:**7–21.

5. Milinski M: **The Major Histocompatibility Complex, Sexual Selection, and Mate Choice.** *Annu Rev Ecol Evol Syst* 2006, **37:**159–186.

6. Worley K, Collet J, Spurgin LG, Cornwallis C, Pizzari T, Richardson DS: **MHC heterozygosity and survival in red junglefowl.** *Mol Ecol* 2010, **19:**3064–3075.

7. Schwensow N, Eberle M, Sommer S: **Compatibility counts: MHC-associated mate choice in a wild promiscuous primate.** *Proceedings of the Royal Society B-Biological Sciences* 2008, **275:**555–564.

8. Siddle HV, Marzec J, Cheng Y, Jones M, Belov K: **MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer.** *Proceedings of the Royal Society B-Biological Sciences* 2010, **277:**2001–2006.

9. Eizaguirre C, Lenz TL, Traulsen A, Milinski M: **Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes.** *Ecol Lett* 2009, **12:**5–12.

10. Kelley J, Walter L, Trowsdale J: **Comparative genomics of major histocompatibility complexes.** *Immunogenetics* 2005, **56:**683–695.

11. Nei M, Gu X, Sitnikova T: **Evolution by the birth-and-death process in multigene families of the vertebrate immune system.** *Proc Natl Acad Sci U S A* 1997, **94:**7799–7806.

12. Nei M, Rooney AP: **Concerted and birth-and-death evolution of multigene families.** *Annu Rev Genet* 2005, **39:**121–152.

13. Trowsdale J, Parham P: **Mini-review: defense strategies and immunity-related genes.** *Eur J Immunol* 2004, **34:**7–17.

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

15. Hess CM, Edwards SV: **The evolution of the major histocompatibility complex in birds.** *Bioscience* 2002, **52:**423–431.

16. Westerdahl H: **Passerine MHC: genetic variation and disease resistance in the wild.** *J Ornithol* 2007, **148:**469–477.

17. Balakrishnan C, Ekblom R, Völker M, Westerdahl H, Godinez R, Kotkiewicz H, Burt DW, Graves T, Griffin DK, Warren WC, Edwards SV: **Gene duplication and fragmentation in the zebra finch major histocompatibility complex.** *BMC Biol* 2010, **8:**29.

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

19. Schut E, Aguilar JR, Merino S, Magrath MJL, Komdeur J, Westerdahl H: **Characterization of MHC-I in the blue tit (*Cyanistes caeruleus*) reveals low levels of genetic diversity and trans-population evolution across European populations.** *Immunogenetics* 2011, **63:**531–542.

20. Promerová M, Albrecht T, Bryja J: **Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*).** *Immunogenetics* 2009, **61:**451–461.

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

22. Westerdahl H, Wittzell H, von Schantz T: **Polymorphism and transcription of Mhc class I genes in a passerine bird, the great reed warbler.** *Immunogenetics* 1999, **49:**158–170.

23. Zagalska-Neubauer M, Babik W, Stuglik M, Gustafsson L, Cichoń M, Radwan J: **454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher.** *BMC Evol Biol* 2010, **10:**395.

24. Babik W, Taberlet P, Ejsmond MJ, Radwan J: **New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system.** *Mol Ecol Resour* 2009, **9:**713–719.

25. Babik W: **Methods for MHC genotyping in non-model vertebrates.** *Mol Ecol Resour* 2010, **10:**237–251.

26. Spurgin LG, Richardson DS: **How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings.** *Proceedings of the Royal Society B-Biological Sciences* 2010, **277:**979–988.

27. Lenz TL, Eizaguirre C, Becker S, Reusch TBH: **RSCA genotyping of MHC for high-throughput evolutionary studies in the model organism three-spined stickleback *Gasterosteus aculeatus*.** *BMC Evol Biol* 2009, **9:**57.

28. Galan M, Guivier E, Caraux G, Charbonnel N, Cosson J-F: **A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies.** *BMC Genomics* 2010, **11:**296.

29. Wegner KM: **Massive parallel MHC genotyping: titanium that shines.** *Mol Ecol* 2009, **18:**1818–1820.

30. Kloch A, Babik W, Bajer A, Siński E, Radwan J: **Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*.** *Mol Ecol* 2010, **19**(Suppl 1):255–265.

31. Huchard E, Weill M, Cowlishaw G, Raymond M, Knapp La: **Polymorphism, haplotype composition, and selection in the Mhc-DRB of wild baboons.** *Immunogenetics* 2008, **60:**585–598.

32. Schwensow N, Fietz J, Dausmann KH, Sommer S: **Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate.** *Heredity* 2007, **99:**265–277.

33. Lund O, Nielsen M, Kesmir C, Petersen AG, Lundegaard C, Worning P, Sylvester-Hvid C, Lamberth K, Røder G, Justesen S, Buus S, Brunak S: **Definition of supertypes for HLA molecules using clustering of specificity matrices.** *Immunogenetics* 2004, **55:**797–810.

34. Sette A, Sidney J: **Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism.** *Immunogenetics* 1999, **50:**201–212.

35. Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E, Funkhouser R, Fugate M, Theiler J, Hsu YS, Kunstman K, Wu S, Phair J, Erlich H, Wolinsky S: **Advantage of rare HLA supertype in HIV disease progression.** *Nat Med* 2003, **9:**928–935.

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

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

38. Loiseau C, Zoorob R, Robert A, Chastel O, Julliard R, Sorci G: **Plasmodium relictum infection and MHC diversity in the house sparrow (*Passer domesticus*).** *Proceedings of the Royal Society B-Biological Sciences* 2011, **278:**1264–1272.

39. Westerdahl H, Asghar M, Hasselquist D, Bensch S: **Quantitative disease resistance: to better understand parasite-mediated selection on major histocompatibility complex.** *Proceedings of the Royal Society B-Biological Sciences* 2011, 1–8. doi:10.1098/rspb.2011.0917.

40. Bonneaud C, Chastel O, Federici P, Westerdahl H, Sorci G: **Complex Mhc-based mate choice in a wild passerine.** *Proceedings of the Royal Society B-Biological Sciences* 2006, **273:**1111–1116.

41. Richardson DS, Komdeur J, Burke T, von Schantz T: **MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler.** *Proceedings of the Royal Society B-Biological Sciences* 2005, **272:**759–767.

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

43. Westerdahl H, Wittzell H, von Schantz T, Bensch S: **MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE.** *Heredity* 2004, **92:**534–542.

44. : *Sequencher® version 5.0 sequence analysis software, Gene Codes Corporation*. MI USA: Ann Arbor; [http://www.genecodes.com.].

45. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl Acids Symp Ser* 1999, **41:**95–98.

46. Technical Bulletin Genome Sequencer FLX System: *Amplicon Fusion Primer Design Guidelines for GS FLX Titanium Series Lib-A Chemistry*. 2009:1–3.

47. Technical Bulletin Genome Sequencer FLX System: *Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Basic MID Set*. 2009:1–11.

48. Lenz TL, Becker S: **Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci–implications for evolutionary analysis.** *Gene* 2008, **427:**117–123.

49. GS FLX Titanium Series: *Amplicon Library Preparation Method Manual*. 2009:1–6.

50. Stuglik MT, Radwan J, Babik W: **jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing.** *Mol Ecol Resour* 2011, **11:**739–742.

51. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28:**2731–2739.

52. Rozas J, Rozas R: **DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data.** *Computer Applications in the Biosciences: CABIOS* 1995, **11:**621–625.

53. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24:**1586–1591.

54. Wallny H-J, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, Skjødt K, Vainio O, Vilbois F, Wiles MV, Kaufman J: **Peptide motifs of the single dominantly expressed class I**

molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proc Natl Acad Sci U S A* 2006, **103:**1434–1439.

55. Doytchinova IA, Flower DR: **In silico identification of supertypes for class II MHCs.** *J Immunol* 2005, **174:**7085–7095.

56. Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S: **New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids.** *J Med Chem* 1998, **41:**2481–2491.

57. R Development Core Team: **R: A Language and Environment for Statistical Computing**. *R Foundation for Statistical Computing, Vienna, Austria*. {ISBN} 2011:3-900051-07-0 [http://www.R-project.org].

58. Jombart T, Devillard S, Balloux F: **Discriminant analysis of principal components: a new method for the analysis of genetically structured populations.** *BMC Genet* 2010, **11:**94.

59. Lerner HRL, Fleischer RC: **Prospects for the use of next-generation sequencing methods in ornithology.** *Auk* 2010, **127:**4–15.

60. Garrigan D, Hedrick PW: **Detecting adaptive molecular polymorphism: lessons from the MHC.** *Evolution* 2003, **57:**1707–1722.

61. Miller HC, Lambert DM: **Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae).** *Immunogenetics* 2004, **56:**178–191.

62. van Oosterhout C, Joyce Da, Cummings SM, Blais J, Barson NJ, Ramnarine IW, Mohammed RS, Persad N, Cable J: **Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (Poecilia Reticulata).** *Evolution* 2006, **60:**2562.

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

64. Anzai T, Shiina T, Kimura N, Yanagiya K, Kohara S, Shigenari A, Yamagata T, Kulski JK, Naruse TK, Fujimori Y, Fukuzumi Y, Yamazaki M, Tashiro H, Iwamoto C, Umehara Y, Imanishi T, Meyer A, Ikeo K, Gojobori T, Bahram S, Inoko H: **Comparative sequencing of**

human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. *Proc Natl Acad Sci U S A* 2003, **100:**7708–7713.

65. Lenz TL: **Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism.** *Evolution* 2011, **65:**2380–2390.

66. Huchard E, Raymond M, Benavides J, Marshall H, Knapp LA, Cowlishaw G: **A female signal reflects MHC genotype in a social primate.** *BMC Evol Biol* 2010, **10:**96.

67. Schwensow N, Fietz J, Dausmann K, Sommer S: **MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate.** *Evol Ecol* 2007, **22:**617–636.

68. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzén A, Winer R, Knight J, Vogel J-H, Aken B, Andersen Ø, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlsen BO, Moum T, Skage M, Berg PR, Gjøen T, Kuhl H, Thorsen J, Malde K, Reinhardt R, Du L, Johansen SD, Searle S, Lien S, Nilsen F, Jonassen I, Omholt SW, Stenseth NC, Jakobsen KS: **The genome sequence of Atlantic cod reveals a unique immune system.** *Nature* 2011, **477:**7–10.

Tables and Figures

Table 2.1 - Primers used for amplifying *Mhc* class I exon 3 in great tits.

Primer	Sequence (5′ - > 3′)	PCR	Reference
HN34	CCATGGGTCTCTGTGGGTA	gDNA	[39]
HN45	CCATGGAATTCCCACAGGAA	gDNA	[39]
T3F	TCCACACCATACAGCGAGTT	gDNA, cDNA	present study
GBT3R	TTTACGCTCCAGCTCTTTCC	gDNA, cDNA	present study
GTJSF	CGGCTGTGACCTCCTGTCC	gDNA	present study
GTJSR	ATTCYGGGCAGATGTGCTT	gDNA	present study
S2F	CACACCTCACAGTGGCTTTA	gDNA, cDNA	present study
S2R	CAGCTCTTTCTGCCCATATC	gDNA, cDNA	present study
L2F	CAATGGCTTTATGGCTGTGAC	gDNA, cDNA	present study
L2R	CCAGCTCTTTCCTCCCGTAT	gDNA, cDNA	present study
JSG2F	AGTTTCCGGCTGTGACCTC	gDNA	present study
JSG2R	GCCCGTATTCGACGTATTTC	gDNA	present study
JSG3F	CATACAGTGGCTTTATGGCTGT	gDNA	present study
JSG3R	CCCGTATCCGGTGTATTTCC	gDNA	present study
MHCV-F	CCCAGGTCTCCACACCATAC	gDNA	present study
MHCV-R	AGCTCTTTCCGCCCGTATT	gDNA	present study
MHCD-F	TTMYGGCTGTGACCTCCTG	gDNA, cDNA	present study
MHCD-R	TTGCGCTYCAGCTCTTTC	gDNA, cDNA	present study

The primers were developed in the order from top to bottom (with the exception of HN34-45). MHCD-F and MHCD-R are the final primers designed for 454 pyrosequencing. F indicates forward and R indicates reverse primers. gDNA indicates amplification of genomic DNA and cDNA indicates amplification of cDNA libraries with the specified primer.

Table 2.2 - Rationale for each step of the variant validation procedure.

	Variant validation procedure	**Rationale**
1	Remove variants that don't match the expected allele size (212, 215, 221 bp)	Variants that have deletions/substitutions shifting the reading frame probably result from sequencing errors (Assumption 1)
2	Remove variants that have less than four copies in the whole dataset	Variants represented once in an individual probably result from sequencing errors (Assumption 4) and variants represented only in one individual probably result from PCR errors (Assumption 5)
3	Remove individuals with less than 200 reads	A low number of reads per individual might lead to incomplete genotyping, thus the results would be unreliable (Assumption 6). The minimum number of reads required per individual is estimated using the probability distribution plotted by Galan et al. [28]
4	Remove variants that have MPAF lower than 0.01	Variants represented rarely in the whole dataset probably result from sequencing errors (Assumption 2)
	Remove variants that have MPAF between 0.01 - 0.025 if they can be explained as a chimera or a single basepair mutation	Variants represented rarely in the whole dataset but more frequently in per individual bases probably result from PCR errors if the parental sequences are also present (Assumption 3)
5	Remove variants that have a single copy per individual	Variants represented once in an individual probably result from sequencing errors (Assumption 4)
	Remove variants that have less than five copies per individual if they can be explained as a chimera or a single basepair mutation	Variants represented two, three or four times within an individual probably result from PCR errors if the parental sequences are present (Assumption 3). The threshold for PCR errors is estimated from the distribution of artefacts in the previous step

Table 2.3 - Repeatability measures for each duplicate pair after 3rd, 4th and 5th step of the variant validation procedure and following supertype classification.

	read no	**verified alleles / ST**				**unverified alleles / ST**				**repeatability**			
		S3	**S4**	**S5**	**ST**	**S3**	**S4**	**S5**	**ST**	**S3**	**S4**	**S5**	**ST**
duplicate-1	528 - 668	31	28	28	9	120	10	0	0	0.21	0.74	1	1
duplicate-2	387 - 418	29	29	28	11	71	14	0	0	0.29	0.67	1	1
duplicate-3	322 - 501	22	22	22	13	54	6	0	0	0.29	0.79	1	1
duplicate-4	471 - 507	26	25	22	9	71	7	0	0	0.27	0.78	1	1
duplicate-5	267 - 484	21	21	21	8	40	6	0	0	0.34	0.78	1	1
duplicate-6	276 - 456	21	20	20	10	52	4	0	0	0.29	0.83	1	1
duplicate-7	288 - 319	23	23	22	10	32	4	1	0	0.42	0.85	0.96	1
duplicate-8	208 - 455	21	21	20	10	50	13	2	0	0.3	0.62	0.91	1
duplicate-9	272 - 289	27	26	23	10	34	3	3	1	0.44	0.89	0.88	0.91
duplicate-10	240 - 471	21	21	18	10	66	20	3	1	0.24	0.51	0.86	0.91
duplicate-11	230 - 236	24	24	21	11	26	6	4	1	0.48	0.8	0.84	0.92
duplicate-12	240 - 570	20	20	19	10	50	22	4	2	0.29	0.48	0.83	0.83
Average	*287 - 448*	*24*	*23*	*22*	*10*	*56*	*9.6*	*1.4*	*0.4*	*0.32*	*0.73*	*0.94*	*0.96*

ST – supertypes; read no – the number of reads per sample. Verified alleles/supertypes stand for the number of identical alleles/supertypes between the duplicate samples. Unverified alleles/supertypes represent the number of alleles/supertypes that were found in one of the samples and indicate non-identical genotypes. Repeatability was calculated as the ratio of verified alleles/supertypes to the total number of alleles/supertypes for each duplicate. A repeatability value of 1 signifies identical genotypes. For a duplicate each measurement was calculated four times - after the 3rd, 4th and 5th steps of the variant validation procedure, and following supertype classification. The average values for each measure are provided in the bottom row.

Table 2.4 - Results of the likelihood ratio tests for different models of codon evolution and estimated parameter values.

Model	**lnL**	**ΔAIC**	**Parameters**
		Group 1	
M0 – one ω	-2975.2	454.8	ω=0.53
M7 – nearly neutral with β	-2788.9	100.2	
M8 – positive selection with β ($\omega_0 \leq 1$, $\omega_1 > 1$)	-2737.8	Best	p_0=0.83, p_1=0.17, ω_1=2.92
		Group 2	
M0 – one ω	-1154.5	68.5	ω=0.74
M7 – nearly neutral with β	-1124	25.5	
M8 – positive selection with β ($\omega_0 \leq 1$, $\omega_1 > 1$)	-1110.2	Best	p_0=0.93, p_1=0.07, ω_1=5.26

ω - dN/dS; nearly neutral with β - for all sites dN/dS ≤ 1 and the beta distribution approximates ω variation; positive selection - a proportion of sites evolves with dN/dS > 1; lnL - Log-likelihood value; ΔAIC - the difference between the value of the AIC of a given model and the best model; p_0 - proportion of sites with dN/dS ≤ 1, p_1 - proportion of positively selected sites (dN/dS > 1), ω_1 - estimated value of ω for sites under positive selection.

Table 2.5 - Results of codon based Z- test of selection for Group 1 and Group 2.

	Group 1				
	n	**dN**	**dS**	**Z**	**P**
ABS	18	0.171±0.135	0.016±0.03	1.244	0.108
Non-ABS	194	0.067±0.014	0.112±0.034	-1.218	1
PSS	9	0.152±0.101	0.038±0.039	1.493	0.069
Non-PSS	203	0.065±0.014	0.115±0.036	-1.254	1
All	212	0.075±0.015	0.103±0.032	-0.788	1
	Group 2				
	n	**dN**	**dS**	**Z**	**P**
ABS	18	0.578±0.132	0.134±0.068	3.694	**<0.001**
Non-ABS	203	0.054±0.011	0.128±0.032	-2.106	1
PSS	27	0.492±0.092	0.069±0.051	5.801	**<0.001**
Non-PSS	194	0.043±0.009	0.136±0.034	-2.686	1
All	221	0.085±0.017	0.126±0.029	-1.158	1

The rates of non-synonymous mutations (dN) and synonymous mutations (dS) were computed using Nei-Gojobari method and Jukes-Cantor correction. The standard errors were obtained through 1000 bootstraps replicates. Z test of selection estimated dN-dS (indicated as Z) and computed a1-tailed test to determine whether dN > dS (indicated as P). n is the number of nucleotides representing antigen-binding sites (ABS), non antigen-binding sites (Non-ABS), positively selected sites (PSS), non positively selected sites (Non-PSS) and all sites in Group 1 and Group 2. ABS were detected by superimposing the chicken *Mhc* class I sequences and assuming concordance. Significant p-values are bold.

Table 2.6 - Summary of *Mhc* genotyping results and estimation of minimum loci number.

	range	mean	s.d.	number of loci
Number of alleles per individual	12 - 37	23.76	3.98	19
Number of functional alleles per individual	9 - 32	19.48	3.74	16
Number of supertypes per individual	6 - 16	10.36	1.62	-

s.d.- standard deviation. Minimum loci number was estimated using the maximum allele number for each class.

Figure 2.1 - Schematic overview of the location of primers used in the study.

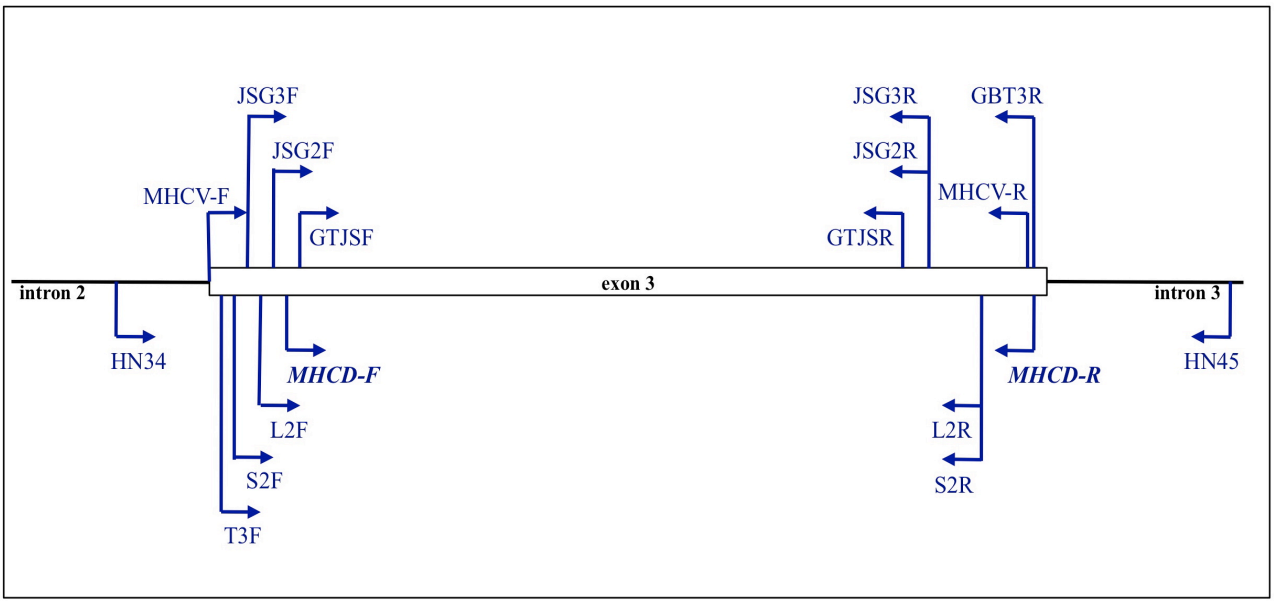


Figure 2.2 - Flow chart of the stepwise variant validation procedure.

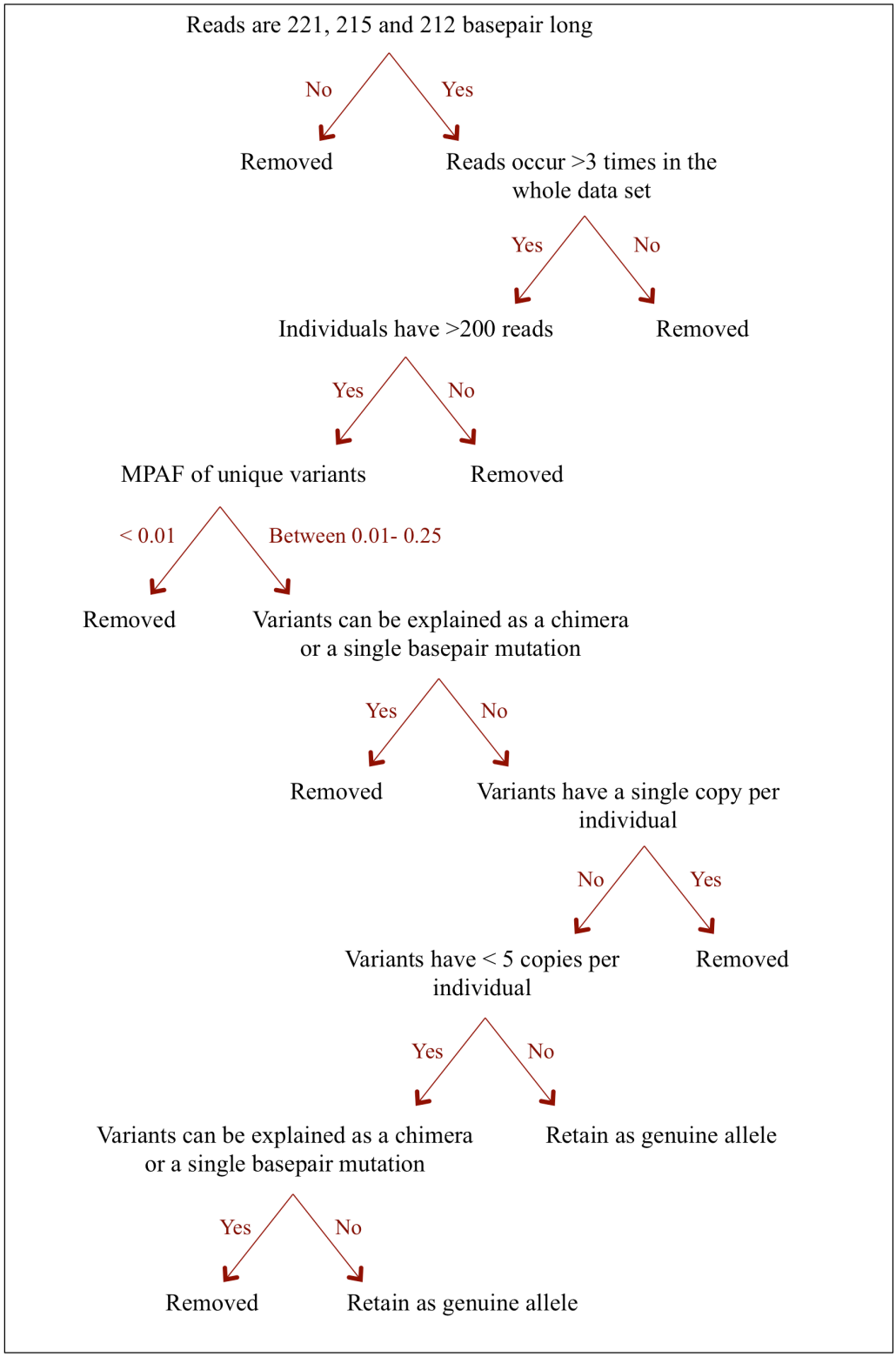


Figure 2.3 - Variation in allele number per individual with increasing read number.

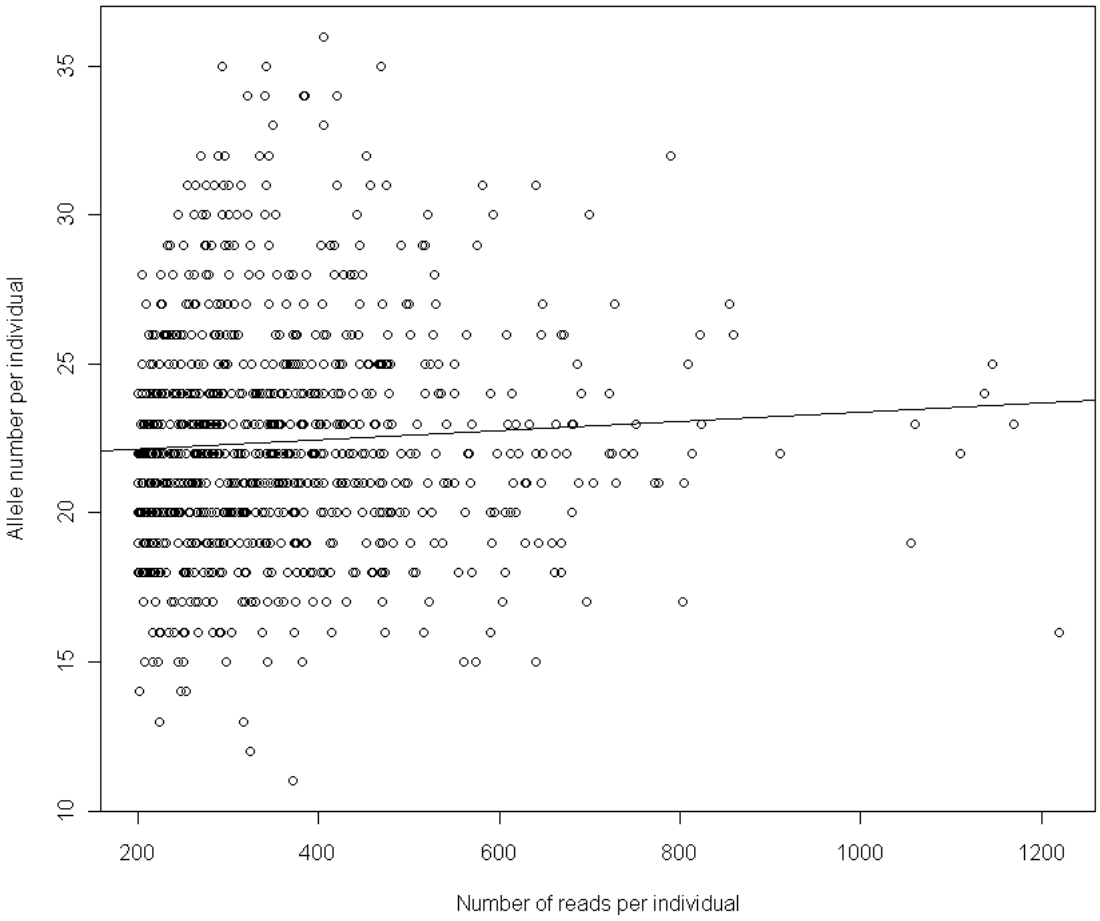
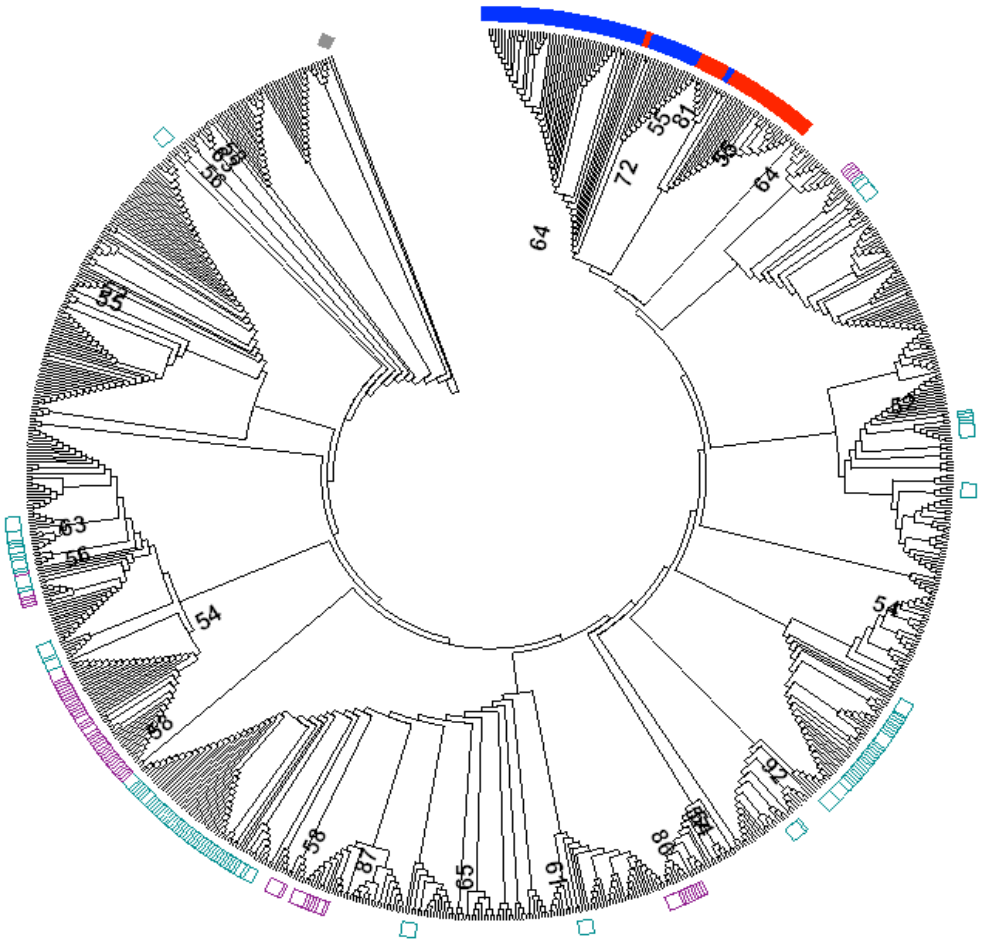
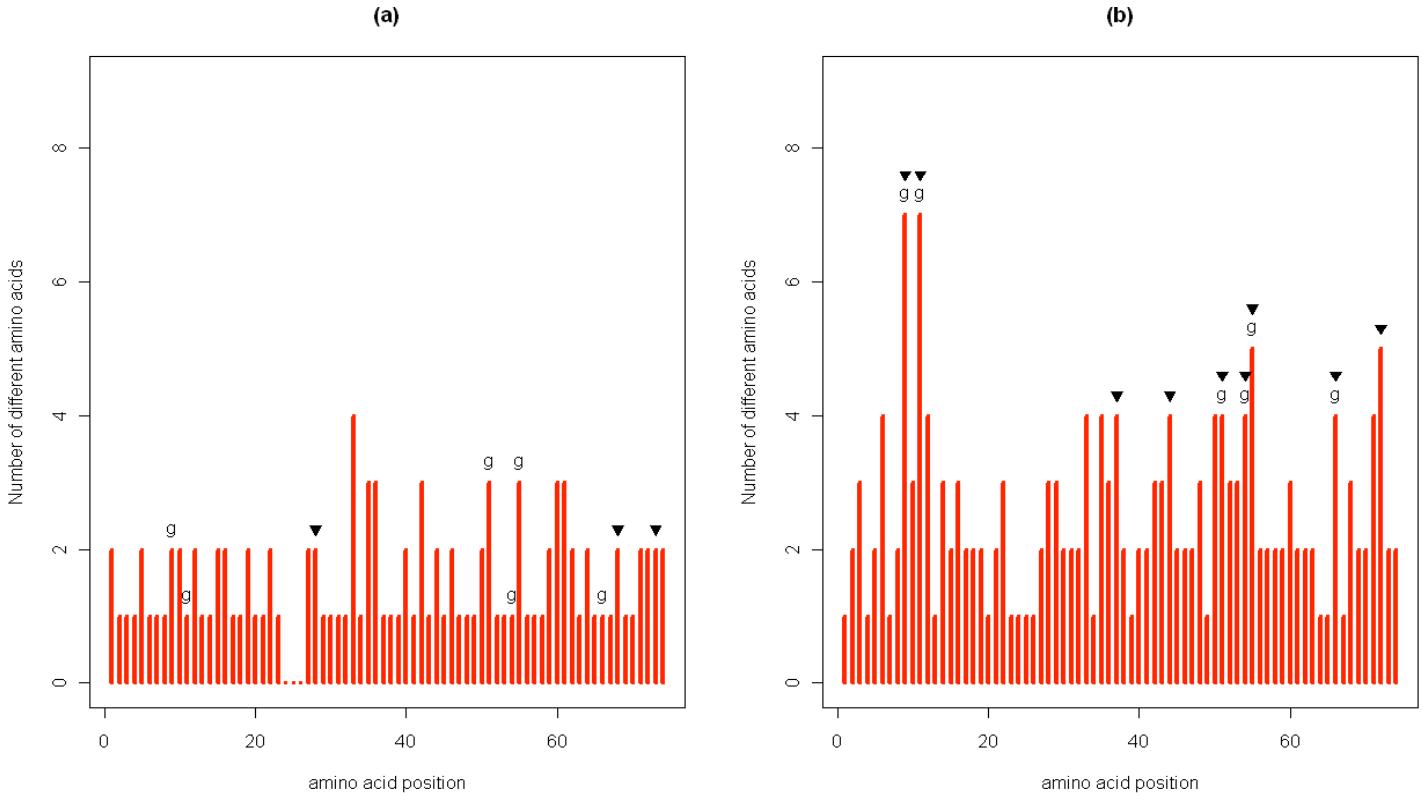


Figure 2.4 - Phylogenetic tree of great tit *Mhc* class I exon 3 sequences.



The tree was constructed using Neighbour-Joining method and Tamura-Nei model. The sequences differed at the length of exon 3 being either 221bp, 215bp or 212bp. The tree was rooted with a chicken (*Gallus gallus*) *Mhc* class I sequence [Genbank: AY234770] and the reliability of the branches was tested with 1000 bootstrap replicates. Bootstrap supports for the major clades are indicated with numbers. Pseudogene alleles bearing stop codons are marked in red and putatively non-functional alleles (Group 1) are marked in blue. The rest of the alleles represent Group 2 (except the chicken *Mhc* class I sequence). The chicken *Mhc* allele is marked in grey. The sequences that are 221-basepair long are indicated by purple squares and the sequences that are 215-basepair long are indicated by green squares. The 212-basepair sequences are not marked.

Figure 2.5 - Aminoacid variation plot for (a) Group 1 and (b) Group 2 alleles.



Chicken antigen-binding sites (ABS) are indicated with the letter 'g', whereas positively selected sites (PSS) are indicated with black triangles. In the Group 1 plot there are no aminoacids between the positions 24-26, because these alleles were 212-basepair long; hence had a nine-basepair deletion at this location.

Figure 2.6 - Plot of (a) *Mhc* allele number per supertype and (b) frequency distribution of supertypes.

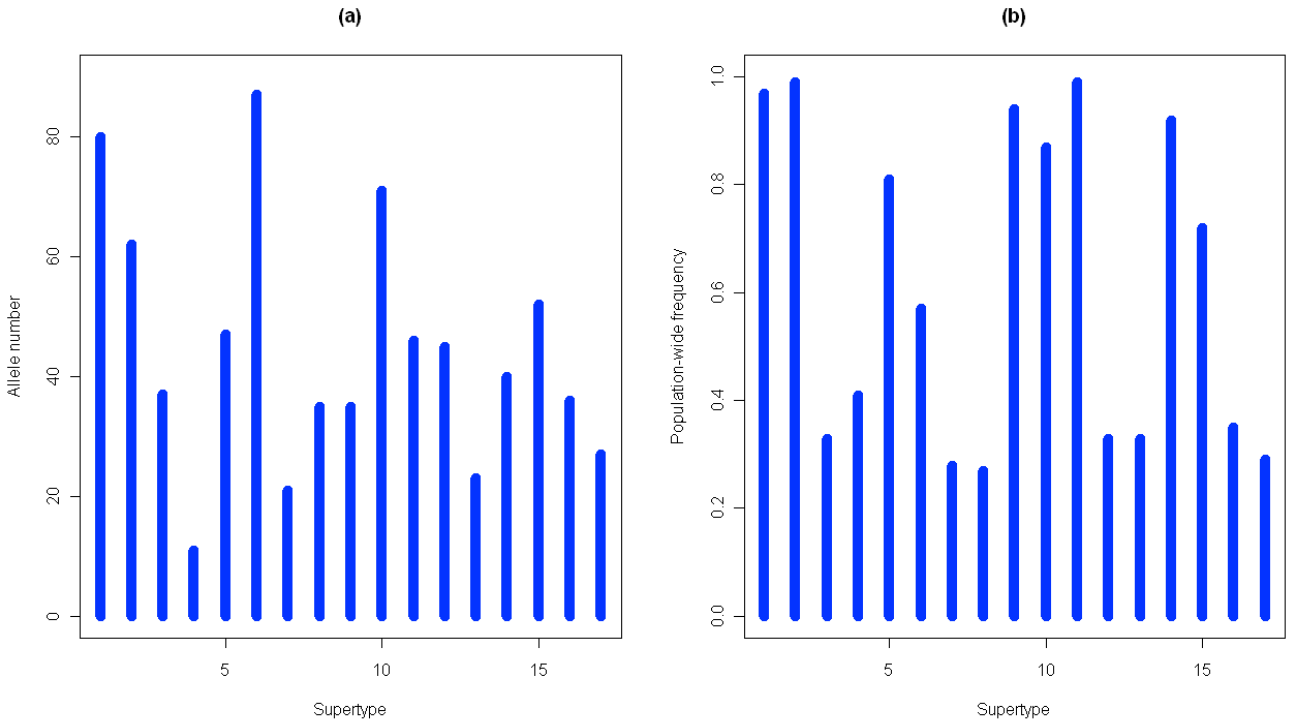
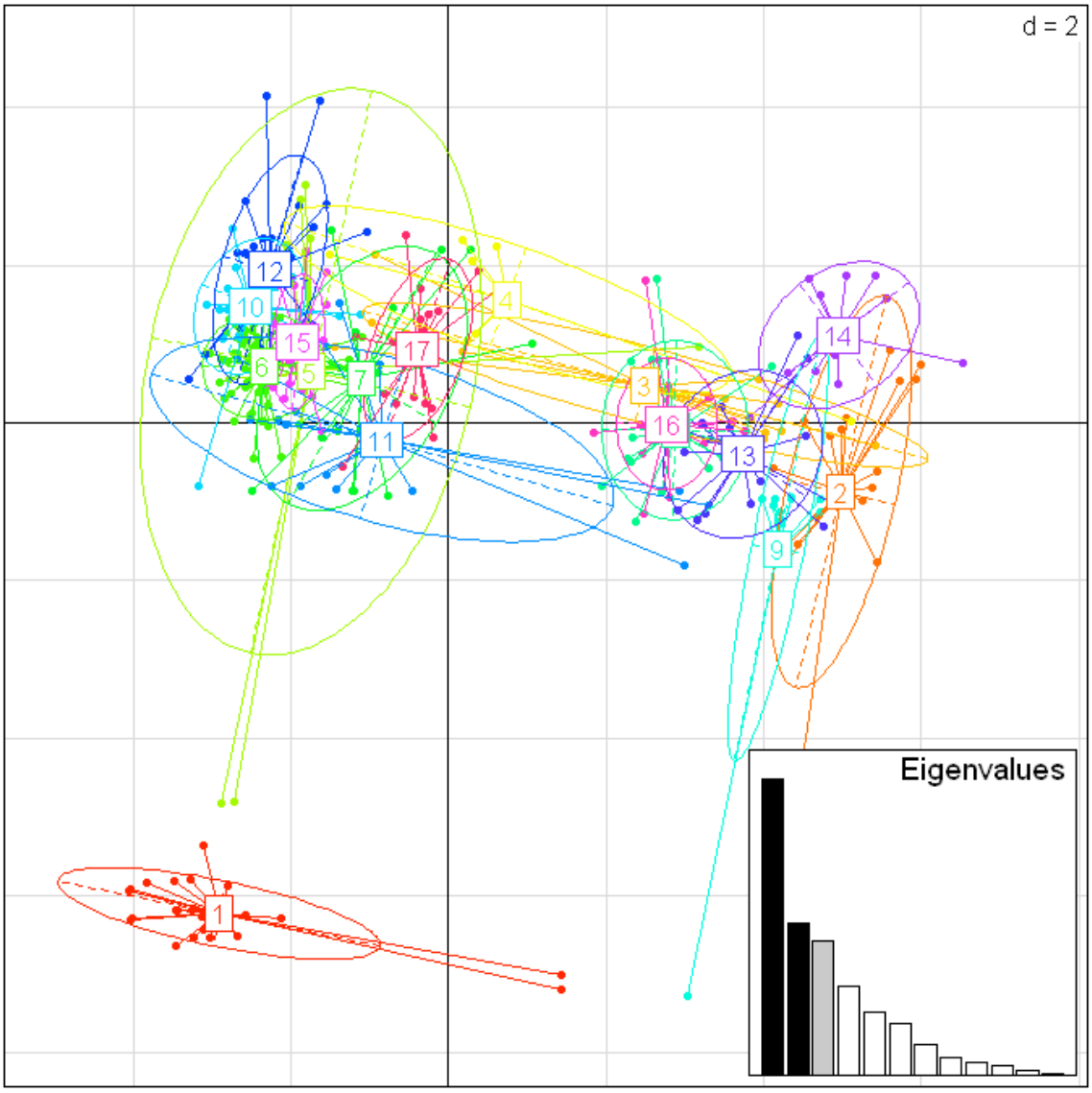


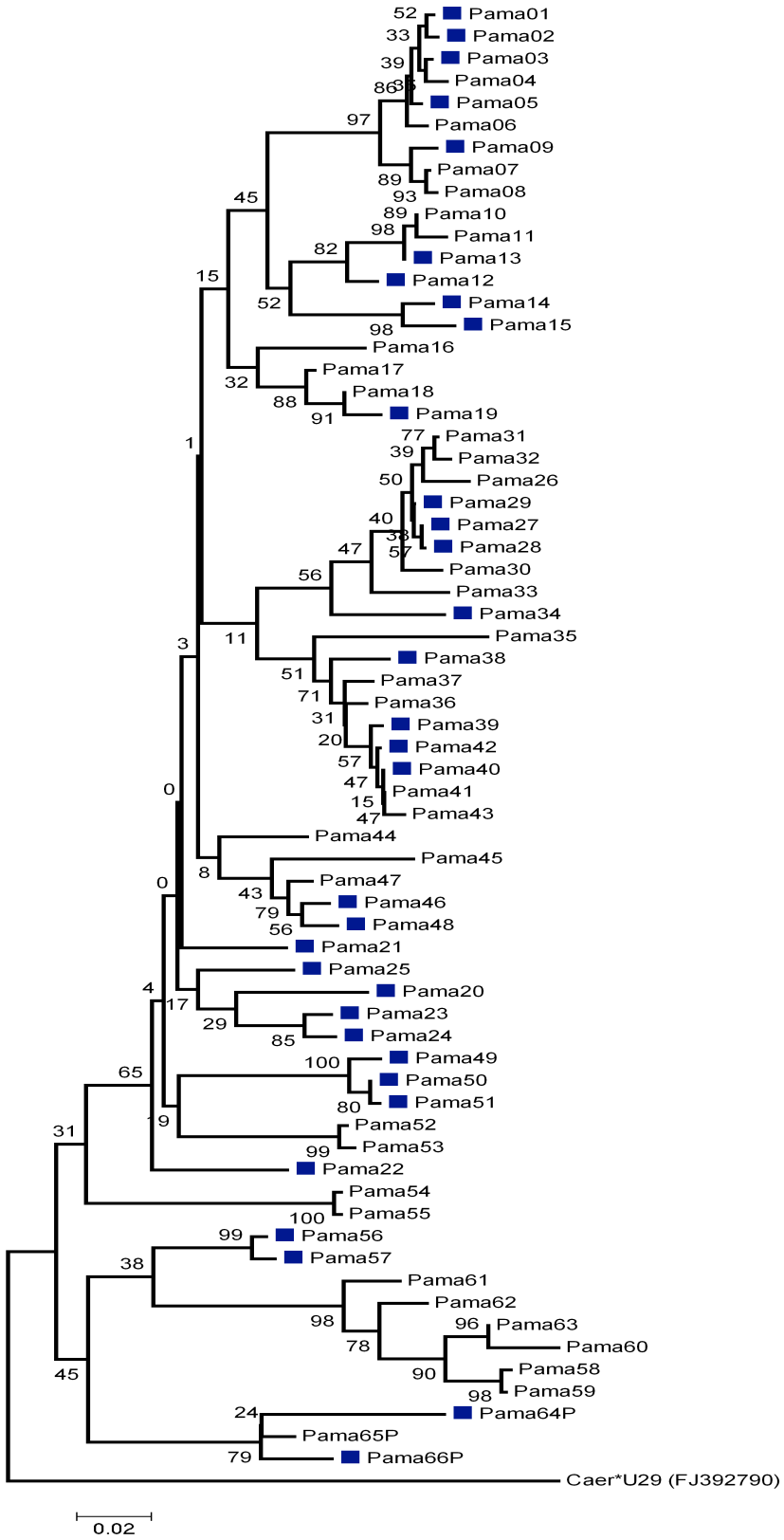
Figure 2.7 - DAPC scatterplot of the 17 *Mhc* supertypes.



12 PCs and three discriminant functions (dimensions) were retained during analyses, to describe the relationship between the clusters. The scatterplot show only the first two PCs (d=2) of the DAPC of *Mhc* supertypes. The bottom right graph illustrates the variation explained by the 12 PCs. Each allele is represented as a dot and the supertypes as ellipses.

Supplementary Information

Figure S2.1 - Phylogenetic tree of the *Mhc* class I sequences identified during characterization.



The tree was constructed using Neighbour-Joining method, Jukes-Cantor model and rooted with a Scarlet Rosefinch (*Carpodacus erythrinus*) *Mhc* class I exon 3 sequence [Genbank: FJ392790]. The interior branch numbers refer to bootstrap values with 2000 replications. Blue squares indicate sequences retrieved from cDNA. Three sequences adjacent to the Scarlet Rosefinch *Mhc* allele are the pseudogene alleles and are indicated with the letter 'P'.

Figure S2.2 - Variation in supertype number per individual with increasing read number.

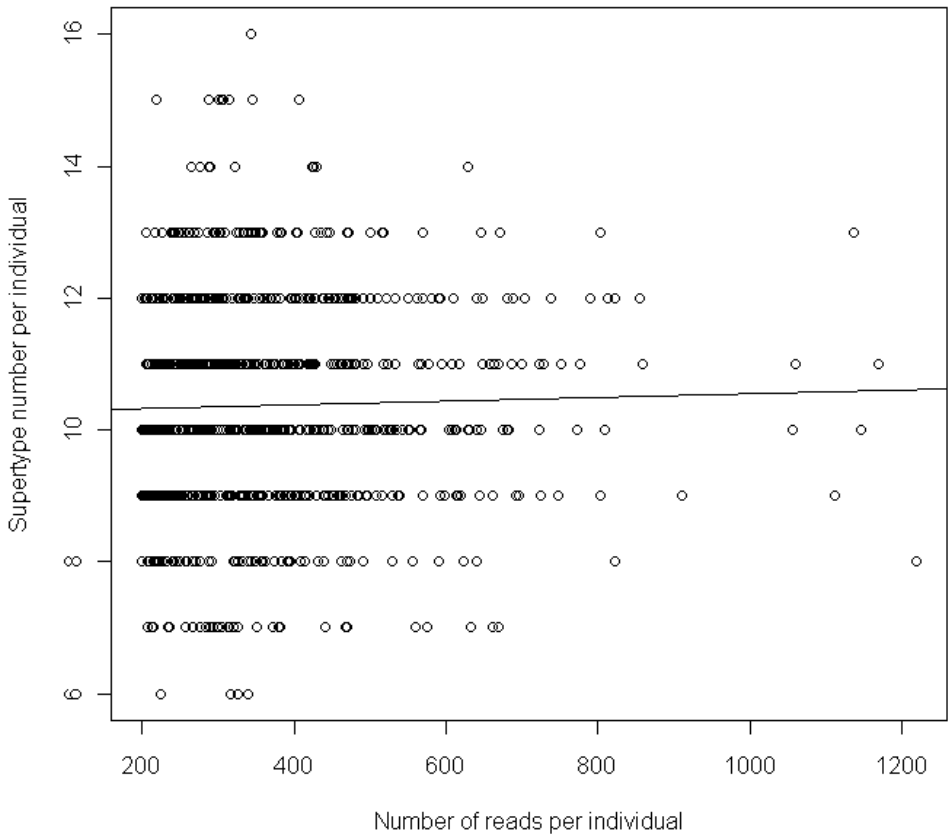
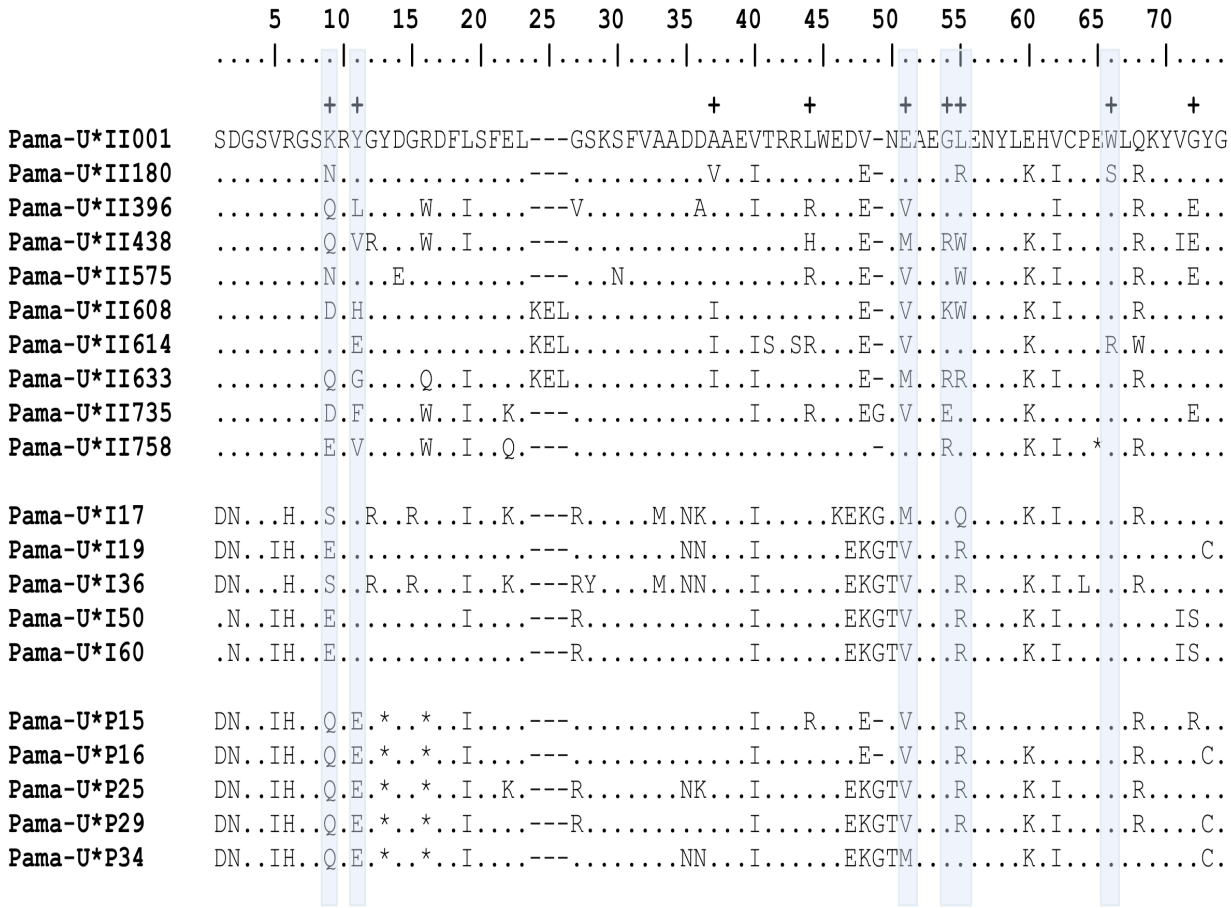
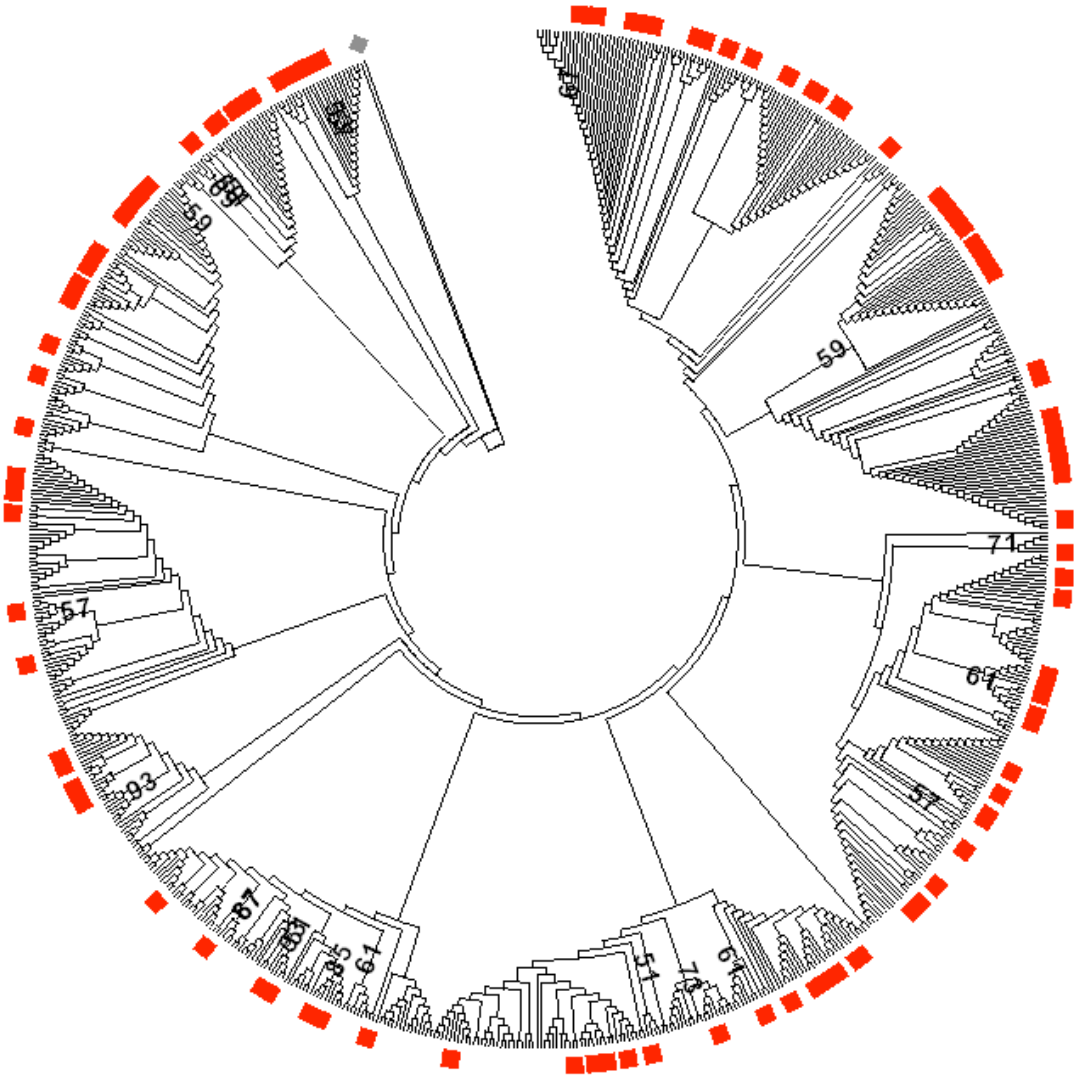


Figure S2.3 - Aminoacid sequences of a representative set of great tit *Mhc* class I alleles (exon 3).



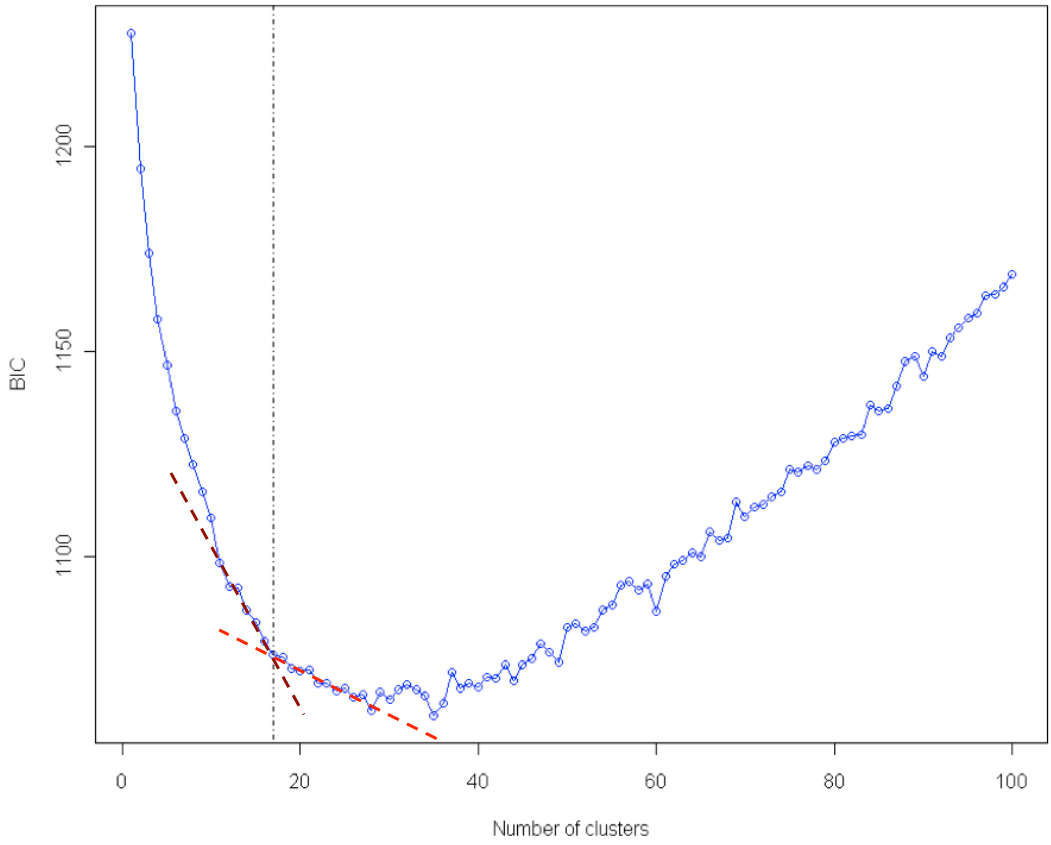
The first set of alleles (Pama-U*II001-II758) are members of Group 2; the second set of alleles (Pama-U*I17-I60) are members of Group 1; and the third set of alleles (Pama-U*P15-P34) are members of the pseudogene. Identity with Pama-U*II001 is indicated with dots, differences are shown by letter substitutions, and gaps are shown by dashes. Stop codons are shown by stars. Chicken antigen-binding sites (ABS) are shaded with blue, while Group 2 positively selected sites (PSS) are indicated with '+'. Pama-U*II758 is one if the three non-functional alleles within Group 2.

Figure S2.4 - Phylogenetic tree of functional Group 2 sequences.



The tree was constructed using Neighbour-Joining method, Jukes-Cantor model and rooted with a chicken (*Gallus gallus*) *Mhc* class I sequence [Genbank: AY234770]. The reliability of the branches was tested with 1000 bootstrap replicates. Bootstrap supports for major clades are indicated with numbers. The alleles that were present in more than 20 individuals are marked in red. The chicken *Mhc* allele is marked in grey.

Figure S2.5 - Graph of BIC values for increasing number of clusters.



The optimal number of supertypes was identified as 17 (marked with a vertical, black dashed-line) as it indicated the elbow in the curve of BIC values. After 17 clusters the slope of BIC decrease dropped notably. The change in the slope at the 17th cluster is indicated with two, red dashed-lines.

Chapter 3

Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population

Proceedings of the Royal Society B: Biological Sciences (in press)

Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population

Irem Sepil[1], Shelly Lachish[1], Amy E Hinks[1], Ben C Sheldon[1]

1. Edward Grey Institute, Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, Oxford, UK

Abstract

Major histocompatibility complex (*Mhc*) genes are believed to play a key role in the genetic basis of disease control. Although numerous studies have sought links between *Mhc* and disease prevalence, many have ignored the ecological and epidemiological aspects of the host-parasite interaction. Consequently, interpreting associations between prevalence and *Mhc* has been difficult, while discriminating alleles for qualitative resistance, quantitative resistance and susceptibility remains challenging. Moreover, most studies to date have quantified associations between genotypes and disease status, overlooking the complex relationship between genotype and the properties of the *Mhc* molecule that interacts with parasites. Here, we address these problems and demonstrate avian malaria (*Plasmodium*) parasite species-specific associations with functional properties of *Mhc* molecules (*Mhc* supertypes) in a wild great tit (*Parus major*) population. We further show that correctly interpreting these associations depends crucially on understanding the spatial variation in risk of infection and the fitness effects of infection. We report that a single *Mhc* supertype confers qualitative resistance to *P. relictum*, while a different *Mhc* supertype confers quantitative resistance to *P. circumflexum* infections. Furthermore, we demonstrate common functional properties of *Plasmodium* resistance alleles in passerine birds, suggesting this is a model system for parasite-*Mhc* associations in the wild.

Keywords: major histocompatibility complex (*Mhc*); avian malaria; *Plasmodium*; supertype; resistance; *Plasmodium*; great tit (*Parus major)*

Introduction

Infectious disease is a major driver of ecological and evolutionary processes within wild populations, and characterizing the genetic component of host immunity is crucial to understand the genetic basis of variation in infection and to assess the evolutionary impacts of diseases [1]. Despite recent advances in genetic technologies, detecting novel resistance loci remains difficult and our understanding of immunogenetics in wild populations is still based on a small number of genes, for example, *Mhc* genes [2] and interleukin 2 genes [3] (though see [4–6] for new markers being developed in wildlife-disease studies). Hence, further work on these candidate genes has the potential to provide valuable information to explain the variation in disease susceptibility in natural populations [7]. In particular, the critical role of *Mhc* genes in triggering an immune response make them excellent candidate markers for disease resistance studies, and *Mhc*-dependent disease resistance has been reported across taxa in natural populations of non-model species [8–10].

Early empirical evidence of a link between specific *Mhc* alleles and disease resistance was provided by the classic paper of Hill *et al.* [11], which suggested that specific *Mhc* alleles conferred protection from severe malaria infections (*Plasmodium spp.)* among African children. Avian malaria parasites, like their human-borne counterparts, are intracellular pathogens that invade host erythrocytes in the bloodstream. Thus, it is highly likely that the *Mhc* class I molecules of avian hosts recognize the peptides derived from malaria parasites and initiate cell destruction. Avian malaria parasites have been shown to compromise host fitness in both immunologically naïve populations [12], as well as in species with which they are assumed to have shared a long evolutionary history [13–15], and are thus likely to exert strong selection on

their hosts. These findings suggest that *Mhc*-linked malaria resistance should also exist in wild avian hosts, a premise supported by the results of several studies in passerine birds [16–20].

However, although significant associations between *Mhc* alleles and avian malaria resistance have been reported, the interpretation of these associations varied considerably in all these studies because both positive and negative associations were observed. Alleles that are negatively correlated with host infection are easily interpreted as disease-resistance alleles that provide qualitative protection against the parasite[17,19]. Positive correlations between *Mhc* alleles and host infection are more difficult to explain, because selection against susceptibility alleles should eliminate them from populations. Potentially, such positive associations may arise as a consequence of pleiotropy, whereby an allele that confers susceptibility to infection simultaneously confers protection to other pathogens [18]. Alternatively, they may arise as a consequence of quantitative disease resistance alleles that limit the deleterious effects of disease without eradicating it completely. For example, in great reed warbler (*Acrocephalus arundinaceus*) populations a positive relationship between a specific *Mhc* allele and *Plasmodium* infection was interpreted as quantitative resistance because parasitaemia (parasite load) was lower in individuals carrying the allele [21]. Selection might favour such alleles if individuals carrying them were better able to suppress acute infection and hence experienced improved survival relative to those without it, generating a positive relationship between presence of this allele and host infection. Another mechanism that may result in a link between *Mhc* genotype and avian malaria infection is heterozygote advantage [22,23]: possessing more *Mhc* alleles may allow more effective parasite presentation [2,9,24]. Alternatively, maximal *Mhc* diversity may lead to excessive T-cell elimination during negative selection in the thymus [25,26]; hence individuals possessing an intermediate, optimal number of *Mhc* alleles may have higher fitness [8,27].

Determining whether specific alleles confer susceptibility or quantitative resistance for the host is crucial for understanding the dynamics of complex host-parasite relationships. Using measures of parasitaemia as a surrogate measure of tolerance to initial infection makes a

number of untested assumptions regarding the relationship of this variable to host immunocompetence. Exploring the presence of pleiotropic effects, on the other hand, is very difficult, given the need to test for and diagnose many different potential pathogens. An alternative approach to discriminate advantageous alleles from disadvantageous alleles is to investigate the fitness consequences of the parasite strain (e.g. malaria) in terms of host survival rates [28–30]. If a mortality cost for infection during the acute stage can be shown in the population, then a positive association between prevalence of infection and the immuno-allele would be interpreted as quantitative disease resistance. Investigating the fitness consequences of *Mhc* alleles in terms of host survival rates will also discriminate advantageous *Mhc* alleles from disadvantageous alleles [31,32]. However, determining such patterns is a major challenge in wild population studies and none of the studies cited above were able to assess survival of hosts in relation to their *Mhc* or disease status (but see [28,31,33]).

Difficulties in understanding the mechanisms by which *Mhc* alleles confer susceptibility or resistance may also stem from other factors. First, risk of infection, particularly for vector-borne diseases, varies extensively at both large and small spatial scales [34–36]. Such factors, if not controlled for, may confound or conceal relevant patterns between *Mhc* and disease status [7]. Moreover, the phenotypic effects of *Mhc* alleles are rarely taken into account [37], despite the fact that it is the properties of the antigen-binding sites (ABS) that determine the interactions between *Mhc* molecules and parasites. Indeed, a growing body of evidence across taxa highlights the importance of grouping *Mhc* alleles into supertypes (defined based on functional properties of the ABS) in order to classify functionally distinct *Mhc* types [38,39].

In this study, we investigate whether *Mhc* class I supertypes are associated with *Plasmodium* infection and host survival in a population of great tits infected with two divergent *Plasmodium* parasite species, *P. relictum* and *P. circumflexum* [40], while controlling for confounding factors, such as local risk of infection and host related effects. Previous studies in this population on the closely related sympatric host species, blue tits (*Cyanistes caeruleus*), have shown that the two *Plasmodium* species differ substantially in their spatial distributions

and impacts [15,35]. While *P. circumflexum* infections are associated with reduced survival, particularly during the acute stage of infection [15], *P. relictum* infections are linked with reproductive costs [14]. Here, we aimed to understand the role that *Mhc* play in determining host resistance and susceptibility to *Plasmodium* infections; and tested the three hypotheses that relate *Mhc* diversity to disease resistance: (i) the maximal diversity hypothesis; (ii) the optimal diversity hypothesis, and (iii) effect of specific *Mhc* supertypes. Moreover we aimed to determine whether specific *Mhc* supertypes confer susceptibility or quantitative resistance to *Plasmodium* infections.

Materials and Methods

Study Population and sampling

We studied the relationship between avian malaria, *Mhc* supertypes and fitness in 576 adult great tits from a nest-box breeding population sampled over two years (2008 and 2009 breeding seasons, April-June), and monitored for survival to 2011. 117 individuals were sampled in both years; hence 693 samples were screened for *Plasmodium* infections. This population has been monitored continuously since the early 1960s and is located in Wytham Woods near Oxford, UK (51°46'N, 1°20'W). The great tit is a small, year-round resident short-lived passerine bird species. Between 250 and 450 great tit pairs breed in the study population annually, and display high breeding site fidelity following postnatal dispersal. Breeding birds were captured during the nestling phase, either within the nest box by hand or using traps, or with mist nets in front of the nest entrance. All adults and nestlings are processed and ringed with aluminium bands for individual recognition. Blood was collected by wing or jugular venipuncture and stored in SET Buffer at -80°C until DNA extraction.

Screening of Plasmodium infections

Total DNA was measured using a Picogreen assay (Quant-iT Picogreen dsDNA Assay Kit, Invitrogen, Grand Island, USA) and diluted to a concentration of 2ng/µl for *Plasmodium* quantification [35]. A quantitative polymerase chain reaction (qPCR)-based assay was employed for measuring parasitaemia. The primers L9 (5'-AAACAATTCCTAACAAAACAGC-3') and New R (5'-ACATCCAATCCATAATAAAGCA-3') were chosen to ensure *Plasmodium*-specific amplification and to differentiate the two *Plasmodium* species [14]. The two *Plasmodium* species are a clade of lineages based on sequence data from approximately 450bp of the mitochondrial cytochrome b gene fragment [34]. Standard curves were created using *Plasmodium* pSGS1 lineage, to calculate the parasitaemia in each sample. qPCR amplifications were run in a final volume of 25µl. Each DNA sample was analyzed thrice and *Plasmodium* parasitaemia was scored as the mean value. 25 samples were randomly chosen, re-quantified and re-amplified to estimate the repeatability of the method. Full details regarding qPCR screening are provided in Knowles *et al.* [14] .

454 pyrosequencing of Mhc class I genes

The *Mhc* characterizations and genotyping methodology used in this paper has been described in Sepil *et al.* [41] and complete details of all molecular protocols can be found therein; brief details are presented here. Bidirectional 454 pyrosequencing was performed on 1536 great tit samples using the 16 lanes of a Pico Titer Plate gasket. We used 10 forward and 10 reverse fusion primers to amplify a 212-221 basepair fragment of great tit *Mhc* class I exon 3. Amplifications were run in a final volume of 25µl and purified as described in Sepil *et al.* [41]. Samples with different primer combinations were pooled together in approximate equimolar quantities, a single pool was prepared for each lane, and the pools were sent for bidirectional 454 pyrosequencing.

High numbers of artefacts can be generated during PCR and 454 pyrosequencing; therefore we adopted a 5-step variant validation procedure to distinguish real alleles from PCR/sequencing errors. From the 638,501 reads that the experiments generated, final genotypes were based on 214,357 reads. 857 individuals passed our reliability criteria, of which 576 were captured in an extensive sampling of the breeding population in 2008-2009 and screened for avian malaria infection and hence were used in this study. A total of 862 *Mhc* class I alleles (Genbank: JQ034624 - JQ035485) were detected; 755 alleles were classified as functional and the presence of at least 16 functional loci was shown. There was clear evidence that the functional alleles were under strong balancing selection based on the detection of positively selected sites using a likelihood ratio test [41]. Lastly, functional alleles were translated into a matrix based on the physicochemical amino acid properties of their positively selected sites [42], and the matrix was subjected to a K-means clustering algorithm [43]; the optimal number of supertypes was identified as 17. The association between specific *Mhc* supertypes and disease prevalence was assessed for 13 supertypes; the remaining four supertypes identified (Supertypes 1, 2, 9 and 11) were removed from analyses as they were nearly fixed in the population with frequencies higher than 95% [41].

Measures of local infection risk

Spatial analysis was based on the GIS-derived measures of tit nest box locations [44]. Breeding tits are territorial and forage in the vicinity of their nests [45]; hence the nestbox coordinates give an accurate representation of an individual's location during and either side of the breeding season, when transmission is most likely [46]. Previous work on the Wytham Woods tit populations has shown that there is pronounced spatial variation in the distribution of the two *Plasmodium* species [34,35,47]. The local risk of infection with either *P. relictum* or *P. circumflexum* was obtained for each *Mhc* typed great tit in 2008 and 2009 and calculated as the prevalence of infected great tits within a 500m buffer (based on disease cluster distances

obtained in Wood *et al.* [34]; Lachish *et al.* [47]) of the focal individuals' nestbox. The analysis was done using GIS software (MapInfo Professional v7.8, Stamford, USA).

Statistical analyses

Statistical associations between *Mhc* supertypes and the probability of infection with either *P. relictum* or *P. circumflexum* were tested using generalized linear models with binomial errors and a logit link. If two years' data were available for an individual (which was the case for 20% of individuals), one was randomly excluded from the analysis, although inclusion of all the samples in the dataset did not qualitatively change the results. The probability of being infected was assessed in relation to the following covariates: the presence of each of the 13 *Mhc* supertypes described above, the total number of supertypes and alleles an individual possessed, the square of the total number of supertypes and alleles, local infection risk, year, sex, age class (categorized as one year of age, two years of age and three years or older) and a quadratic age class term (colinearity among explanatory variables was low; see the electronic supplementary material, Table S3.1). Supertype number, its square, allele number, its square, age class and its quadratic function were mean-centered before inclusion in the model, by subtracting the mean from each data point, to reduce the covariance between the terms.

We used Akaike's information criterion (AIC) to determine the combination of variables that best explained the data with minimal parameters. Model selection was performed by backward stepwise elimination and the fit of each new model was assessed by comparison of AIC values. Terms were eliminated from the model when their removal reduced AIC by at least 2 and were retained if their removal increased AIC by at least 2. Where the removal of a term resulted in a model with an approximately equal fit (i.e. a change in AIC of <2) the model with fewer terms was considered the most parsimonious model [48]. To confirm the validity of the minimum model, removed variables were added individually to assess any potential improvement in the model fit. Estimates of effect sizes (odd-ratios) were calculated for each *Mhc* supertype to confirm the strength of association between *Mhc* and probability of infection.

All analyses were conducted using R 2.12.1 [49].

The above analyses revealed a positive relationship between the presence of one *Mhc* supertype (supertype 6) and infection with *P. circumflexum* (see Results), suggesting that this supertype confers either susceptibility or quantitative resistance to *P. circumflexum* infection. To differentiate between these two possibilities, we conducted further analyses to investigate whether the parasite has a mortality cost during acute stage infection and whether the supertype conferred a survival advantage to its carriers (which would favour the idea that this supertype confers quantitative resistance to infection).

First, as proposed by Westerdahl *et al.* [21], we assessed whether *P. circumflexum* parasitaemia varied according to the presence of the *Mhc* supertype, using a linear mixed effects model. Again, if two years data were available for an individual, one was randomly excluded from the analysis. Parasitaemia was log transformed [ln (x+1)] to meet assumptions of constant variance and normal errors. Since parasitaemia values were calculated through comparison with standard curve values (see details above), the particular standard used was treated as a random factor to eliminate any bias arising from the use of different standards. Second, we assessed whether host survival probability varied between individuals that either possessed or did not possess supertype 6, as a function of individual infection status and the local risk of *P. circumflexum* infection. We reasoned that, if supertype 6 conferred quantitative resistance to *P. circumflexum* infection, then uninfected hosts that lack the supertype would have lower survival in high-risk areas (where they are at greater risk of becoming infected) than in low-risk areas. This should not be the case for infected birds, as the majority of these individuals would already have survived the brief acute stage of infection and thus harbour only chronic infections [50]. Moreover, we wished to assess whether host survival varied as a function of the local risk of *P. circumflexum* infection, since such a relationship would support a mortality cost for this parasite in this species. We predicted that uninfected hosts would have lower survival in high-risk areas than in low-risk areas, if *P. circumflexum* infection has a mortality cost during the acute stage. The absence of a pattern for uninfected birds would either indicate no or weak survival cost of

P. circumflexum infection or imply that some of the uninfected birds have already had and cleared infection, concealing the identification of any relevant pattern. We determined host survival by scoring whether each individual was recaptured in the subsequent breeding seasons (using data on recaptures, including 2011) or not. As annual recapture rates of great tits in the breeding season are high in Wytham Woods (higher than 0.80 [51]), this provides a reasonably robust index of survival rates. Analyses were performed using a generalized linear model with binomial errors and a logit link, and the starting model included the three variables (presence of supertype 6, infection status and local risk of *P. circumflexum* infection) and their interaction. Third, we assessed whether the frequency of the *Mhc* supertype varied between first year and older individuals as a function of the local risk of *P. circumflexum* infection. We predicted that the frequency of supertype 6 should be higher in older individuals, particularly in areas with elevated risk of infection, as a result of the selective disappearance of individuals that lack this supertype. We separated high and low infection risk areas by scoring whether an individual's nest box was located within 500m of the River Thames or not; *P. circumflexum* infection risk is substantially elevated in areas within 500m of this large water body [34,35,47]. We assessed whether the frequency of the *Mhc* supertype varied as a function of host age (categorized as first year and older), and *P. circumflexum* infection risk, using a generalized linear model with binomial errors and a logit link, with both terms and their interaction included in the starting model. Calculating local infection risk as the percentage of infected great tits within a 500m buffer, rather than using an arbitrary-cut of point, did not quantitatively change the results of the analysis (results not shown). Lastly, we explored the effects of supertype/allelic diversity, *Mhc* supertype 6 and their interaction on *P. circumflexum* prevalence. We predicted that if *Mhc* supertype 6 conferred susceptibility, individuals with lower supertype or allelic diversity would be more prone to *P. circumflexum* infection as a result of homozygosity (as in Worley *et al.* [30]).

In a final separate analysis, we explored whether the associations we found between the two *Mhc* supertypes and the probability of *Plasmodium* infections (see Results) were

comparable with the findings of earlier studies. Of five studies investigating a link between passerine *Mhc* class I and *Plasmodium* infection, only two studies, both on house sparrows (*Passer domesticus*), presented sequence information for *Mhc* alleles identified as being linked to P. *relictum* infection (GenBank: EU715815 - EU715817 and EF429132) [18,19]. The similarity of these house sparrow alleles to the *Mhc* supertypes examined in this study at their putative antigen-binding sites (to infer functional similarity) was assessed by combining these additional four alleles with our 755 functional class I alleles and re-running the K-means clustering algorithm [41]. Hence, in this analysis we aimed to determine whether *Mhc* alleles that are associated with avian malaria in house sparrows cluster, in terms of their functional properties, with alleles of the two *Mhc* supertypes linked with *Plasmodium* infection in great tits.

Results

Overall 55% of individuals were infected with *Plasmodium*; *P. circumflexum* prevalence (37.8%) was significantly higher than *P. relictum* prevalence (19.8%) ($\chi^2_1 = 44.89$, $P < 0.0001$). Fifteen individuals (2.7%) were co-infected with the two *Plasmodium* species and were included in both sets of analyses. Parasitaemia of infected individuals did not differ between the two *Plasmodium* species ($\chi^2_1 = 1.76$, $P = 0.184$, ΔAIC=-0.3); the repeatability of the parasitaemia data was 0.779 (Pearson's product-moment correlation, $p < 0.001$), similar to previous estimates of repeatability from blue tits (r = 0.71) [14].

Our analyses revealed there to be significant associations between two different *Mhc* supertypes and the probability of infection with *P. relictum* and *P. circumflexum*. The probability of *P. relictum* infection was negatively associated with the presence of supertype 17 (Figure 3.1a, electronic supplementary material, Table S3.2), such that individuals lacking the supertype were twice as likely as individuals carrying it to become infected with *P. relictum* (odds ratio: 0.51, 95% CI: 0.39-0.66, P=0.009, electronic supplementary material, Figure S3.1),

indicative of qualitative resistance. No other supertype was significantly associated with probability of *P. relictum* infection; nor was *P. relictum* infection related to the total number of supertypes or total number of alleles an individual possessed (either linearly or as a quadratic function; Table S3.2, Figure S3.1). The probability of *P. relictum* infection increased significantly with age and local infection risk, but did not vary between the sexes or the year of the study (Table S3.2).

In contrast, results of the generalized linear model for *P. circumflexum* infections in hosts revealed there was a significant positive association between the probability of *P. circumflexum* infection and the presence of supertype 6 (Figure 3.1b, electronic supplementary material, Table S3.3). Individuals carrying the supertype were 58% more likely to be infected with *P. circumflexum* (odds ratio: 1.58, 95% CI: 1.11-2.27, P=0.009, electronic supplementary material, Figure S3.2) than were the individuals that did not possess this *Mhc* supertype. No other supertype was significantly associated with *P. circumflexum* infection, nor was there any significant relationship between the total number of supertypes or total number of alleles an individual possessed (either linearly or as a quadratic function) and its infection status (Table S3.3, Figure S3.2). As in the *P. relictum* analysis above, the probability of *P. circumflexum* infection did not vary between the sexes, nor between the years of the study, but did vary significantly with *P. circumflexum* infection risk and host age, with older individuals more likely to be infected (Table S3.3).

The above analyses showed that individuals carrying *Mhc* supertype 6 were more likely to be infected with *P. circumflexum*, suggesting that supertype 6 may confer either susceptibility or quantitative resistance to *P. circumflexum* infection. Contrary to our predictions *P. circumflexum* parasitaemia did not vary between individuals that did or did not possess supertype 6 ($\chi^2_1 = 0.22$, $P = 0.641$, ΔAIC=-1.78), nor did survival probabilities vary between individuals with and without supertype 6 (see the electronic supplementary material, Table S3.4). Moreover we found no significant interaction between the presence of supertype 6 and supertype/allelic diversity in terms of *P. circumflexum* infection (supertype 6*supertype

diversity, $\chi^2_1 = 0.72$, $P = 0.4$, ΔAIC= -1.28; supertype 6*allelic diversity, $\chi^2_1 = 0.43$, $P = 0.51$, ΔAIC= -1.57). However, our results showed that uninfected, but not infected, hosts experienced lower survival rates in high infection risk areas than in low-risk areas, indicating that this pathogen entails a mortality cost for hosts during the acute stage of infection (see Figure 3.2, Table S3.4). In addition, as predicted, our analyses revealed that the frequency of supertype 6 varied between first year and older individuals as a function of *P. circumflexum* infection risk (local infection risk*host age, $\chi^2_1 = 4.55$, $P = 0.033$, ΔAIC=+2.55). The frequency of supertype 6 increased with age in high-risk areas: suggesting that individuals possessing this supertype are better able survive the lethal acute stage of infection, and that selection progressively increases the frequency of this supertype (Figure 3.3). Since none of these findings lends support for the susceptibility hypothesis, the evidence favours quantitative resistance.

Finally, we assessed the functional similarity of the *Mhc* class I alleles associated with *Plasmodium* prevalence in passerines, by running K-means clustering algorithm on the combination of four house sparrow *Mhc* alleles and 755 great tit *Mhc* alleles. Two of the house sparrow alleles (EU715815 and EF429132) clustered with 18 great tit alleles, of which 14 were originally designated supertype 17, while a third house sparrow allele (EU715816) clustered with 31 great tit alleles, of which 18 were supertype 6 in our original clustering (see the electronic supplementary material, Figure S3.3). The fourth allele (EU715817) clustered with four great tit alleles that were originally designated supertype 4 (Figure S3.3). Hence, three of the four alleles identified as being associated with malaria in another bird species cluster statistically, in terms of their functional properties, with those identified as being malaria-associated in the present study.

Discussion

Investigating associations between *Mhc* and parasite prevalence is a common means of studying genetically determined disease resistance in wild animals. Until recently, positive associations

between *Mhc* alleles and parasite prevalence had been taken as evidence of susceptibility to disease, while the potential for quantitative resistance (immuno-alleles that reduce the development of infection) has been largely neglected (but see [21]). In this study, we incorporated a detailed investigation of avian malaria infection and analysis of *Mhc* class I genes in a wild great tit population, to understand the role that *Mhc* genes play in determining host resistance and susceptibility to *Plasmodium* infections. We found that the presence of two *Mhc* supertypes (defined based on functional properties of the antigen-binding sites) was significantly associated with the probability of host infection with two congeneric *Plasmodium* species, but in contrasting fashions. The direction of the association for one *Mhc* supertype was indicative of *Mhc*-linked qualitative resistance to *P. relictum* infection, with individuals lacking that supertype twice as likely to be infected. However, the functional role of the second *Mhc* supertype, in regard to *P. circumflexum* infection, was more difficult to assess. Of the four hypotheses we tested, two analyses supported quantitative resistance, while two analyses provided equivocal results. Therefore, our results more strongly support the idea that supertype 6 may confer quantitative resistance to *P. circumflexum* infection. Hence, the findings of this study imply that different *Mhc* supertypes can confer both qualitative and quantitative resistance to different *Plasmodium* species in a single host population.

Different types of associations between immuno-alleles and the probability of infection are common for different levels of infection severity [21]. Virulent parasites are likely to induce mortality during the acute stage of infection and are difficult to suppress completely [52]. Hence, alleles that confer quantitative resistance (i.e. that limit the deleterious effects of infection but do not prevent infection) would be beneficial to hosts that are primarily exposed to virulent parasites. Conversely, qualitative resistance alleles that prevent parasite establishment (i.e. prevent infection) would be beneficial to hosts that are exposed to more benign parasites, as these parasites are unlikely to cause mortality and are easier to suppress [53]. Using a multistate modeling framework, Lachish *et al.* [15] showed that infection with *P. circumflexum* was associated with reduced survival in blue tits, compared to infection with *P. relictum*, suggesting

that *P. circumflexum* is more virulent than *P. relictum*. In line with this finding, we found that uninfected great tits experienced lower survival rates in areas where *P. circumflexum* infection risk is high, suggesting a mortality cost for this parasite in great tits as well. Our results showing *Mhc*-linked qualitative resistance to *P. relictum* infection by great tits carrying supertype 17, and the possibility of *Mhc*-linked quantitative resistance to *P. circumflexum* infection for individuals carrying supertype 6, are thus in agreement with the above rationale and may be explained by the differing virulence of these two parasites.

Consequently, we investigated whether uninfected hosts lacking supertype 6 had lower survival rates in high-risk areas, but found no such relationship. Although we would expect supertype 6 to confer a survival advantage for its carriers, the highest mortality from infection is likely to occur when birds are first exposed to the parasite, as juveniles [15]. Hence the absence of an association between supertype 6 and survival rates of adults in the present study is perhaps not unexpected. Moreover, recent analyses have shown that individuals possessing supertype 6 had significantly higher lifetime reproductive success (defined as the total number of recruits produced over a lifetime) and offspring recruitment probabilities at the brood level, implying a survival advantage for juveniles carrying this *Mhc* supertype 6 [32]. This finding supports the suggestion that supertype 6 confers quantitative resistance to *P. circumflexum* infection. In the present study, we showed that the frequency of supertype 6 increased with age in high infection risk areas, further supporting the idea that supertype 6 confers selective advantage for individuals that are likely to contract the disease. The negative association found between supertype 6 frequency and host age in low-risk areas (Figure 3.3), might indicate that there is a cost associated with carrying a disease-resistance supertype [54]. If this is true, then the absence of supertype 6 might be more beneficial for great tits breeding in areas where the vector transmitting *P. circumflexum* is absent. Hence, it seems possible that the selective advantage of this supertype depends on local infection risk and there might be differential selection for the supertype at this small spatial scale.

The results of this study lend support to hypotheses suggesting a fitness advantage for carriers of specific *Mhc* types. Therefore it is plausible to suggest that the extraordinary level of *Mhc* diversity observed in this population might be maintained through the mechanisms of negative-frequency dependent selection [55], fluctuating selection [56] or a combination of the two mechanisms. We found no support for hypotheses that link maximal or optimal *Mhc* diversity with individual fitness [20,27]; neither the number of supertypes, nor the number of alleles had an effect on *Plasmodium* infection. However it is worth noting that we were not able to test heterozygote advantage at a particular locus [22], since multiple loci (at least 16 loci) were analyzed simultaneously. Therefore we cannot discount the possibility that associations between disease prevalence and heterozygosity at certain loci might have been overlooked.

In the present study we used a 'supertyping' approach to determine the functional properties of the *Mhc* alleles. To date, the majority of work has assessed links between *Mhc* genotype and disease resistance, however much of selection on *Mhc* is likely to act via the phenotypic effects of the underlying genes. Here, we adopted the bioinformatic and statistical methods described by Doytchinova and Flower [42] and Jombart *et al.* [43] to define the physicochemical properties of the positively selected sites of each allele and to cluster *Mhc* alleles with similar peptide specificities into supertypes. Therefore, by analyzing the relationship between *Mhc* supertypes and *Plasmodium* infection, we aimed to predict the functional effects of *Mhc* supertypes. There is increasing support for the biological relevance of the supertyping approach [57,58]. For instance, Trachtenberg *et al.* [38] revealed an advantage of rare human leukocyte antigen (HLA) supertypes in HIV disease progression, independent of the contribution of single alleles. Therefore, we believe this approach is both appropriate and justified. Here, we utilized a newly proposed method for *Mhc* genotyping [43], which is an advance on previous methods, as it does not require arbitrary clustering decisions, but uses K-means clustering algorithm and model selection approach to compute associated summary statistics [41]. The supertype clusters changed significantly following the addition of the four house sparrow alleles. The house sparrow alleles differed substantially from the great tit alleles

and included amino acids not found amongst the great tit variants. Hence, the apparent instability of the clusters is likely an outcome of this difference. Nevertheless, the methodology can be further developed to improve the precision of identifying distinct clusters; experimental approaches can be used instead of bioinformatics approaches to determine the peptide specificities of *Mhc* molecules [59]. Moreover, it is worth noting that our amplifications did not cover a polymorphic section of *Mhc* exon 3 that affects the antigen-binding capabilities of the alleles. Amplifying the entire exon would be beneficial for future studies [see 60].

An important and novel finding from this study is that the alleles that are linked with *Plasmodium* infections in two passerine species (great tits and house sparrows) have similar antigen-binding affinities. Although population-specific associations between *Mhc* alleles and *Plasmodium* species have been reported in house sparrow populations [17,19], the functional differences between these alleles were not assessed, and our results imply that different alleles linked with disease may be similar at their antigen-binding sites even in unrelated host species. The functional similarity across species suggests that *Mhc*-linked malaria resistance can be a valuable area for further research to understand the genetic basis of variation in infection in wild systems. However, as this study shows, such work must be embedded in an understanding of the ecological and epidemiological processes affecting the host-parasite interaction.

Acknowledgements

We are very grateful to the Wytham fieldworkers who collected the data for this study. We thank Alicia Davies and Simon Lee for laboratory assistance, Tobias Uller, Colin Garroway, Kaan Oz and two anonymous reviewers for helpful discussion and advice. This work was partially funded by NERC grants (NER/A/S/2002/00877 and NE/F005725/1) to BCS.

References

1 Acevedo-Whitehouse, K. & Cunningham, A. A. 2006 Is MHC enough for understanding wildlife immunogenetics? *Trends Ecol. & Evol.* **21**, 433–438. (doi:10.1016/j.tree.2006.05.010)

2 Oliver, M. K., Telfer, S. & Piertney, S. B. 2009 Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (Arvicola terrestris). *Proc. R. Soc. Lond. B* **276**, 1119–1128. (doi:10.1098/rspb.2008.1525)

3 Turner, A. K., Begon, M., Jackson, J. A., Bradley, J. E. & Paterson, S. 2011 Genetic diversity in cytokines associated with immune variation and resistance to multiple pathogens in a natural rodent population. *PLoS Genet.* **7**, e1002343. (doi:10.1371/journal.pgen.1002343)

4 Hellgren, O. & Sheldon, B. C. 2011 Locus-specific protocol for nine different innate immune genes (antimicrobial peptides: β-defensins) across passerine bird species reveals within-species coding variation and a case of trans-species polymorphisms. *Mol Ecol Resour.* **11**, 686–692. (doi:10.1111/j.1755-0998.2011.02995.x)

5 Tschirren, B., Råberg, L. & Westerdahl, H. 2011 Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents. *J. Evol. Biol.* **24**, 1232–1240. (doi:10.1111/j.1420-9101.2011.02254.x)

6 Grueber, C. E., Wallis, G. P., King, T. M. & Jamieson, I. G. 2012 Variation at innate immunity toll-like receptor genes in a bottlenecked population of a new zealand robin. *PloS One* **7**, e45011. (doi:10.1371/journal.pone.0045011)

7 Amos, W., Driscoll, E. & Hoffman, J. I. 2011 Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proc. R. Soc. Lond. B* **278**, 1183–1188. (doi:10.1098/rspb.2010.1920)

8 Madsen, T. & Ujvari, B. 2006 MHC class I variation associates with parasite resistance and longevity in tropical pythons. *J. Evol. Biol.* **19**, 1973–1978. (doi:10.1111/j.1420-9101.2006.01158.x)

9 Kekäläinen, J., Vallunen, J. A., Primmer, C. R., Rättyä, J. & Taskinen, J. 2009 Signals of major histocompatibility complex overdominance in a wild salmonid population. *Proc. R. Soc. Lond. B* **276**, 3133–3140. (doi:10.1098/rspb.2009.0727)

10 Kloch, A., Babik, W., Bajer, A., Siński, E. & Radwan, J. 2010 Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. *Mol. Ecol.* **19 Suppl 1**, 255–265. (doi:10.1111/j.1365-294X.2009.04476.x)

11 Hill, A. V. S. *et al*. 1991 Common West African HLA antigens are associated with protection from severe malaria. *Nature* **352**, 595 - 600. (doi:10.1038/352595a0)

12 Benning, T. L., LaPointe, D., Atkinson, C. T. & Vitousek, P. M. 2002 Interactions of climate change with biological invasions and land use in the Hawaiian Islands: Modeling the fate of endemic birds using a geographic information system. *Proc. Natl Acad. Sci. USA* **99**, 14246–14249. (doi:10.1073/pnas.162372399)

13 Marzal, A., de Lope, F., Navarro, C. & Møller, A. P. 2005 Malarial parasites decrease reproductive success: an experimental study in a passerine bird. *Oecologia* **142**, 541–545. (doi:10.1007/s00442-004-1757-2)

14 Knowles, S. C. L., Palinauskas, V. & Sheldon, B. C. 2010 Chronic malaria infections increase family inequalities and reduce parental fitness: experimental evidence from a wild bird population. *J. Evol. Biol.* **23**, 557–569. (doi:10.1111/j.1420-9101.2009.01920.x)

15 Lachish, S., Knowles, S. C. L., Alves, R., Wood, M. J. & Sheldon, B. C. 2011 Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. *J. Anim. Ecol.* **80,** 1196–1206. (doi:10.1111/j.1365-2656.2011.01836.x)

16 Westerdahl, H., Waldenström, J., Hansson, B., Hasselquist, D., von Schantz, T. & Bensch, S. 2005 Associations between malaria and MHC genes in a migratory songbird. *Proc. R. Soc. Lond. B* **272**, 1511–1518. (doi:10.1098/rspb.2005.3113)

17 Bonneaud, C., Pérez-Tris, J., Federici, P., Chastel, O. & Sorci, G. 2006 Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution* **60**, 383–389. (doi:10.1111/j.0014-3820.2006.tb01114.x)

18 Loiseau, C., Zoorob, R., Garnier, S., Birard, J., Federici, P., Julliard, R. & Sorci, G. 2008 Antagonistic effects of a Mhc class I allele on malaria-infected house sparrows. *Ecol. Lett.* **11**, 258–265. (doi:10.1111/j.1461-0248.2007.01141.x)

19 Loiseau, C., Zoorob, R., Robert, A., Chastel, O., Julliard, R. & Sorci, G. 2011 *Plasmodium relictum* infection and MHC diversity in the house sparrow (*Passer domesticus*). *Proc. R. Soc. Lond. B* **278**, 1264–1272. (doi:10.1098/rspb.2010.1968)

20 Radwan, J., Zagalska-Neubauer, M., Cichoń, M., Sendecka, J., Kulma, K., Gustafsson, L. & Babik, W. 2012 MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Mol. Ecol.* **21,** 2469-2479(doi:10.1111/j.1365-294X.2012.05547.x)

21 Westerdahl, H., Asghar, M., Hasselquist, D. & Bensch, S. 2011 Quantitative disease resistance: to better understand parasite-mediated selection on major histocompatibility complex. *Proc. R. Soc. Lond. B* **279,** 577-584. (doi:10.1098/rspb.2011.0917)

22 Doherty, P. & Zinkernagel, R. 1975 Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**, 50–52. (doi:10.1038/256050a0)

23 Takahata, N. & Nei, M. 1990 Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**, 967–978.

24 Froeschke, G. & Sommer, S. 2005 MHC class II DRB variability and parasite load in the striped mouse (*Rhabdomys pumilio*) in the Southern Kalahari. *Mol. Biol. Evol.* **22**, 1254–1259. (doi:10.1093/molbev/msi112)

25 Nowak, M. A., Tarczy-Hornoch, K. & Austyn, J. M. 1992 The optimal number of major histocompatibility complex molecules in an individual. *Proc. Natl Acad. Sci. USA* **89**, 10896–10899. (doi:10.1073/pnas.89.22.10896)

26 Woelfing, B., Traulsen, A., Milinski, M. & Boehm, T. 2009 Does intra-individual major histocompatibility complex diversity keep a golden mean? *Phil. Trans. R. Soc. B* **364**, 117–128. (doi:10.1098/rstb.2008.0174)

27 Kalbe, M., Eizaguirre, C., Dankert, I., Reusch, T. B. H., Sommerfeld, R. D., Wegner, K. M. & Milinski, M. 2009 Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proc. R. Soc. Lond. B* **276**, 925–934. (doi:10.1098/rspb.2008.1466)

28 Paterson, S., Wilson, K. & Pemberton, J. 1998 Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proc. Natl Acad. Sci. USA* **95**, 3714–3719. (doi:10.1073/pnas.95.7.3714)

29 Lohm, J., Grahn, M., Langefors, A., Andersen, Ø., Storset, A. & Von Schantz, T. 2002 Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. *Proc. R. Soc. Lond. B* **269**, 2029–2033. (doi:10.1098/rspb.2002.2114)

30 Worley, K., Collet, J., Spurgin, L. G., Cornwallis, C., Pizzari, T. & Richardson, D. S. 2010 MHC heterozygosity and survival in red junglefowl. *Mol. Ecol.* **19**, 3064–3075. (doi:10.1111/j.1365-294X.2010.04724.x)

31 Brouwer, L., Barr, I., Van de Pol, M., Burke, T., Komdeur, J. & Richardson, D. S. 2010 MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Mol. Ecol.* **19**, 3444–3455. (doi:10.1111/j.1365-294X.2010.04750.x)

32 Sepil, I., Lachish, S. & Sheldon, B. C. 2013 Mhc-linked survival and lifetime reproductive success in a wild population of great tits. *Mol. Ecol.* **22**, 384–396. (doi:10.1111/mec.12123)

33 De Assunção-Franco, M., Hoffman, J. I., Harwood, J. & Amos, W. 2012 MHC genotype and near-deterministic mortality in grey seals. *Sci. Rep.* **2**, 659. (doi:10.1038/srep00659)

34 Wood, M. J., Cosgrove, C. L., Wilkin, T. A., Knowles, S. C. L., Day, K. P. & Sheldon, B. C. 2007 Within-population variation in prevalence and lineage distribution of avian malaria in blue tits, *Cyanistes caeruleus*. *Mol. Ecol.* **16**, 3263–3273. (doi:10.1111/j.1365-294X.2007.03362.x)

35 Knowles, S. C. L., Wood, M. J., Alves, R., Wilkin, T. A., Bensch, S. & Sheldon, B. C. 2011 Molecular epidemiology of malaria prevalence and parasitaemia in a wild bird population. *Mol. Ecol.* **20**, 1062–1076. (doi:10.1111/j.1365-294X.2010.04909.x)

36 Sehgal, R. N. M. *et al.* 2011 Spatially explicit predictions of blood parasites in a widely distributed African rainforest bird. *Proc. R. Soc. Lond. B* **278**, 1025–1033. (doi:10.1098/rspb.2010.1720)

37 Spurgin, L. G. & Richardson, D. S. 2010 How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc. R. Soc. Lond. B* **277**, 979–988. (doi:10.1098/rspb.2009.2084)

38 Trachtenberg, E. *et al.* 2003 Advantage of rare HLA supertype in HIV disease progression. *Nat. Med.* **9**, 928–935. (doi:10.1038/nm893)

39 Huchard, E., Raymond, M., Benavides, J., Marshall, H., Knapp, L. A. & Cowlishaw, G. 2010 A female signal reflects MHC genotype in a social primate. *BMC Evol. Biol.* **10**, 96. (doi:10.1186/1471-2148-10-96)

40 Valkiunas, G. 2005 *Avian malaria parasites and other haemosporidia*. CRC Press, Boca Raton.

41 Sepil, I., Moghadam, H. K., Huchard, E. & Sheldon, B. C. 2012 Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evol. Biol.* **12**, 68. (doi:10.1186/1471-2148-12-68)

42 Doytchinova, I. & Flower, D. R. 2005 In silico identification of supertypes for class II MHCs. *J. Immunol.* **174**, 7085–7095.

43 Jombart, T., Devillard, S. & Balloux, F. 2010 Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94. (doi:10.1186/1471-2156-11-94)

44 Wilkin, T. A., Perrins, C. M. & Sheldon, B. C. 2007 The use of GIS in estimating spatial variation in habitat quality : a case study of lay-date in the Great Tit *Parus major*. *Ibis* **149**, 110–118. (doi:10.1111/j.1474-919X.2007.00757.x)

45 Stauss, M. J., Burkhardt, J. F. & Tomiuk, J. 2005 Foraging flight distances as a measure of parental effort in blue tits *Parus caeruleus* differ with environmental conditions. *J. Avian Biol.* **36**, 47–56. (doi:10.1111/j.0908-8857.2005.02855.x)

46 Cosgrove, C. L., Wood, M. J., Day, K. P. & Sheldon, B. C. 2008 Seasonal variation in *Plasmodium* prevalence in a population of blue tits *Cyanistes caeruleus*. *J. Anim. Ecol.* **77**, 540–548. (doi:10.1111/j.1365-2656.2008.01370.x)

47 Lachish, S., Knowles, S. C. L., Alves, R., Sepil, I., Davies, A., Lee, S., Wood, M. J. & Sheldon, B. C. 2012. Spatial determinants of infection risk in a multi-species avian malaria system. *Ecography* **35,** 1–12. (doi:10.1111/j.1600-0587.2012.07801.x)

48 Burnham, K. & Anderson, D. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag, New York.

49 R Development Core Team. 2011 R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*. , {ISBN} 3–900051–07–0, http://www.R–project.org.

50 Bensch, S., Waldenström, J., Jonzén, N., Westerdahl, H., Hansson, B., Sejberg, D. & Hasselquist, D. 2007 Temporal dynamics and diversity of avian malaria parasites in a single host species. *J. Anim. Ecol.* **76**, 112–122. (doi:10.1111/j.1365-2656.2006.01176.x)

51 Bouwhuis, S., Choquet, R., Sheldon, B. C. & Verhulst, S. 2012 The forms and fitness cost of senescence: age-specific recapture, survival, reproduction, and reproductive value in a wild bird population. *Am. Nat.* **179**, E15–E27. (doi:10.1086/663194)

52 Gandon, S. & Michalakis, Y. 2000 Evolution of parasite virulence against qualitative or quantitative host resistance. *Proc. R. Soc. Lond. B* **267**, 985–990. (doi:10.1098/rspb.2000.1100)

53 May, R. & Nowak, M. 1994 Superinfection, metapopulation dynamics, and the evolution of diversity. *J. Theor. Biol.* **170**, 95–114. (doi:10.1006/jtbi.1994.1171)

54 Sheldon, B. C. & Verhulst, S. 1996 Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology. *Trends Ecol. & Evol.* **5347**, 317–321. (doi:10.1016/0169-5347(96)10039-2)

55 Slade, R. W. & McCallum, H. I. 1992 Overdominantvs. Frequency-Dependent Selectionat MHC Loci. *Genetics* **132**, 861–862. (doi:10.2460/javma.240.8.931)

56 Hedrick, P. W. 2002 Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**, 1902–1908.

57 Bertoni, R., Sidney, J., Fowler, P., Chesnut, R. W., Chisari, F. V & Sette, A. 1997 Human histocompatibility leukocyte antigen-binding supermotifs predict broadly cross-

reactive cytotoxic T lymphocyte responses in patients with acute hepatitis. *J. Clin. Invest.* **100**, 503–513. (doi:10.1172/JCI119559)

58 Lund, O. *et al.* 2004 Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* **55**, 797–810. (doi:10.1007/s00251-004-0647-4)

59 Lenz, T. L. 2011 Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution* **65**, 2380–2390. (doi:10.1111/j.1558-5646.2011.01288.x)

60 Llaurens, V., McMullan, M. & van Oosterhout, C. 2012 Cryptic MHC polymorphism revealed but not explained by selection on the class IIB peptide-binding region. *Mol. Biol. Evol.* **29,** 1631-1644. (doi:10.1093/molbev/mss012)

Figures

Figure 3.1 - Probability of (a) *P. relictum* infection in great tit hosts (± SE) as a function of the presence of supertype 17; (b) *P. circumflexum* infection in great tit hosts (± SE) as a function of the presence of supertype 6.

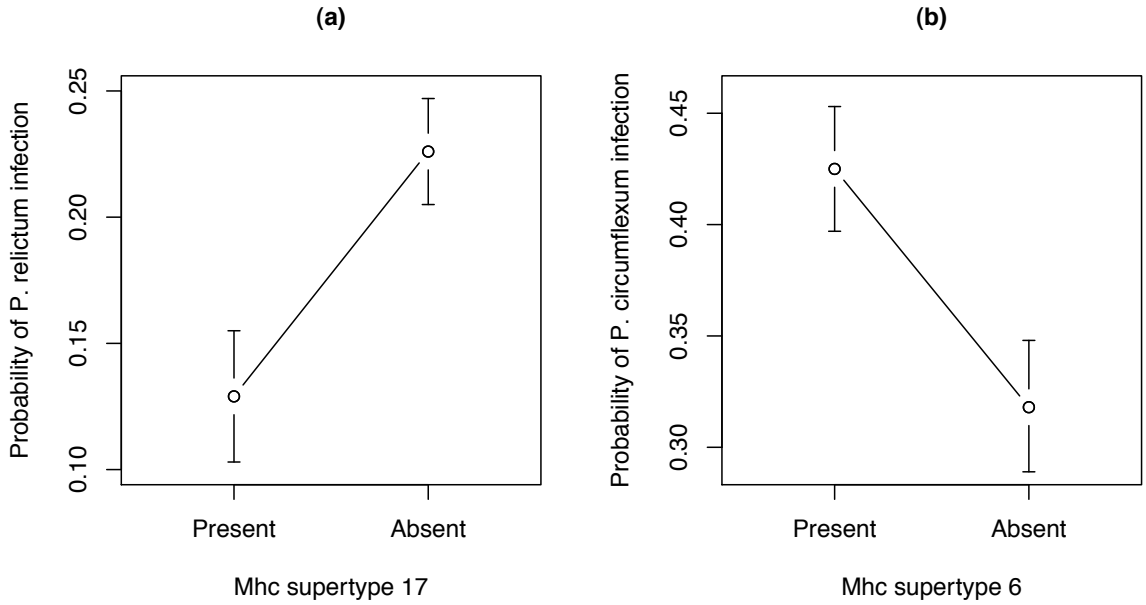


Figure 3.2 – Mean survival rates of (a) uninfected great tits (± SE) as a function of infection risk; (b) *P. circumflexum* infected great tits (± SE) as a function of infection risk.

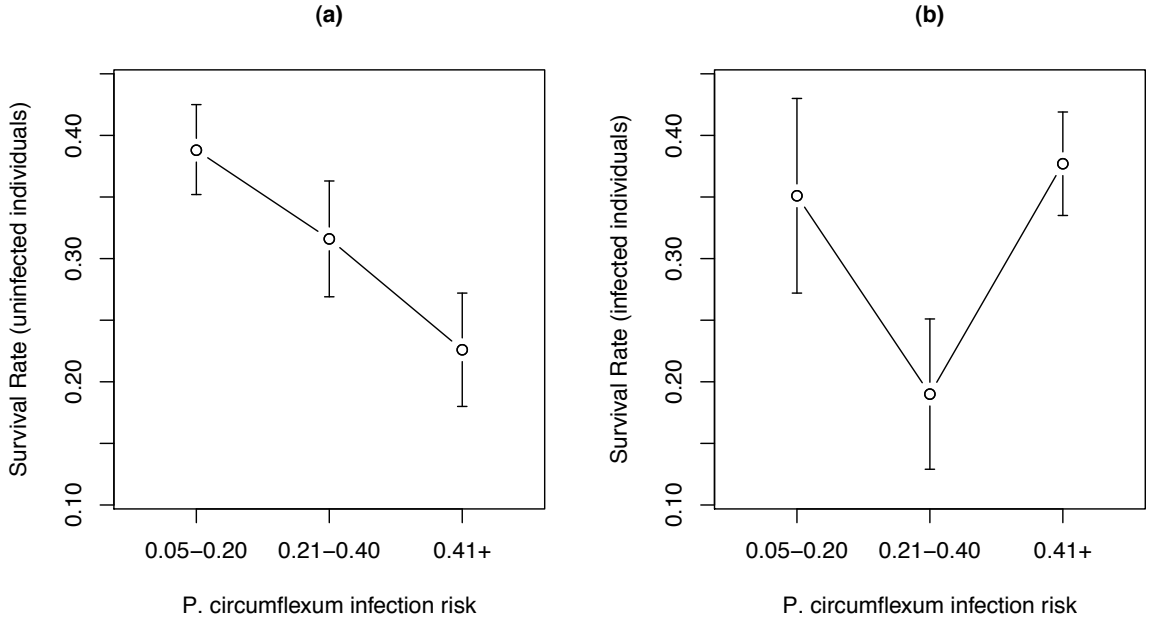


Figure 3.3 – Frequency of *Mhc* supertype 6 in great tit hosts (± SE) in high *P. circumflexum* infection risk areas (solid bars) and low *P. circumflexum* infection risk areas (dotted bars) as a function of host age.

Electronic Supplementary Material – Additional Tables and Figures

Table S3.1 - Pair-wise correlations of the explanatory variables.

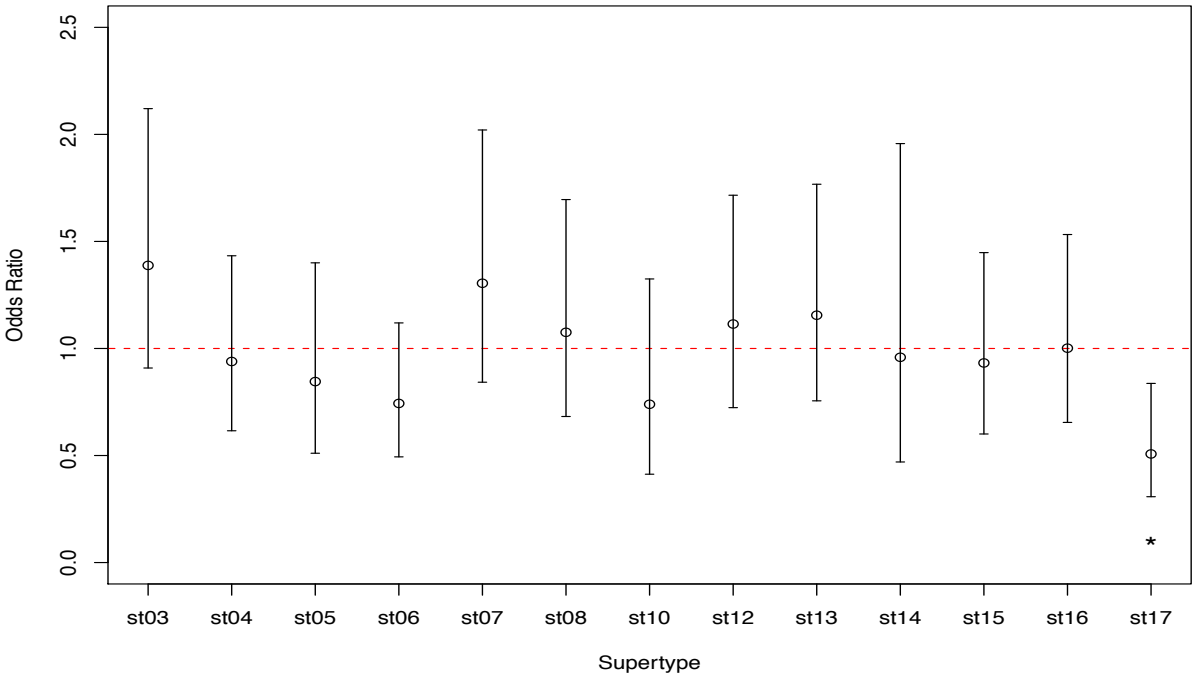
	ST3	ST4	ST5	ST6	ST7	ST8	ST10	ST12	ST13	ST14	ST15	ST16	ST17	ST.no	Allele.no	TIR	SIR	Year	Age
ST3	-																		
ST4	-0.03	-																	
ST5	0.01	-0.06	-																
ST6	-0.02	0.04	-0.01	-															
ST7	0.12	0.09	-0.06	-0.04	-														
ST8	0.01	0.01	0.05	-0.02	0.05	-													
ST10	-0.01	0.09	0.05	-0.07	0.11	0.03	-												
ST12	0.01	0.03	-0.01	-0.10	0.04	-0.10	-0.09	-											
ST13	0.06	-0.02	0.09	0.06	-0.03	0.06	-0.05	-0.03	-										
ST14	0.08	0.03	0.04	0.01	0.04	0.05	-0.01	0.03	-0.04	-									
ST15	0.01	0.11	0.12	-0.02	-0.02	-0.03	0.01	0.03	0.03	0.02	-								
ST16	0.09	-0.11	0.02	0.01	-0.04	-0.13	0.03	-0.03	0.07	-0.01	-0.01	-							
ST17	-0.12	-0.13	-0.01	-0.04	0.03	-0.16	0.06	0.01	-0.11	-0.03	0.06	-0.12	-						
ST.no	0.34	0.30	0.31	0.24	0.36	0.22	0.24	0.23	0.33	0.23	0.38	0.23	0.11	-					
Allele.no	0.18	0.22	0.25	0.11	0.26	0.09	0.22	0.11	0.09	0.19	0.24	0.13	0.03	0.61	-				
TIR	-0.01	-0.05	-0.02	0.02	0.03	0.03	0.02	-0.01	0.01	-0.07	0.03	-0.03	0.03	0.02	0.02	-			
SIR	0.02	-0.01	0.01	-0.02	0.01	-0.06	-0.05	-0.03	-0.01	0.04	-0.05	0.13	-0.06	-0.02	-0.02	-0.64	-		
Year	-0.01	-0.05	0.01	-0.02	0.16	-0.08	-0.08	0.03	-0.03	-0.08	-0.04	0.01	-0.02	-0.10	-0.04	0.04	-0.02	-	
Age	-0.04	-0.05	-0.01	-0.04	0.01	0.05	-0.03	0.06	-0.01	-0.01	-0.06	0.01	-0.02	-0.05	-0.02	0.10	-0.09	-0.06	-

Table S3.2 - Results of model selection based on AIC for generalized linear models with binomial errors examining the influence of *Mhc*, host age and sex, and local infection risk on the probability of *P. relictum* infection.

predictor	coefficient	SE	test statistic	P-value	ΔAIC
supertype 3	0.305	0.232	1.315	0.188	-0.29
supertype 4	-0.144	0.267	-0.539	0.590	-1.70
supertype 5	-0.169	0.298	-0.565	0.572	-1.68
supertype 6	-0.227	0.232	-0.979	0.328	-1.04
supertype 7	0.289	0.247	1.171	0.242	-0.64
supertype 8	-0.134	0.281	-0.476	0.634	-1.77
supertype 10	-0.169	0.337	-0.502	0.616	-1.76
supertype 12	-0.093	0.309	-0.301	0.763	-1.91
supertype 13	-0.076	0.459	-0.166	0.868	-1.97
supertype 14	-0.195	0.419	-0.464	0.642	-1.79
supertype 15	-0.099	0.303	-0.328	0.743	-1.89
supertype 16	-0.218	0.243	-0.896	0.370	-1.19
supertype 17	**-0.650**	**0.263**	**-2.470**	**0.014**	**+4.60**
total number of supertypes	-0.019	0.093	-0.209	0.835	-1.96
total number of supertypes^2	-0.038	0.031	-1.210	0.226	-0.50
total number of alleles	-0.011	0.029	-0.378	0.705	-1.86
total number of alleles^2	0.009	0.005	1.839	0.066	+1.21
local infection risk	**5.385**	**1.554**	**3.465**	**0.001**	**+10.34**
year	-0.120	0.232	-0.518	0.605	-1.74
sex	0.397	0.217	1.828	0.068	+1.36
age class	**0.563**	**0.135**	**4.168**	**<0.001**	**+15.74**
age class^2	-0.189	0.229	-0.826	0.409	-1.32

(Test statistics reported are z-statistics, with P-values of coefficients calculated from Wald tests. ΔAIC is the change in the AIC value after removal of that term from the model. Terms shown in bold are the variables retained in the minimum adequate model.)

Figure S3.1 – Estimates of effect size (odds ratios with 95% CI) of each *Mhc* supertype on the probability of *P. relictum* infection.



Odds ratios above one indicate positive associations between *Mhc* supertypes and infection. Odds ratios below one indicate negative associations between *Mhc* supertypes and infection.

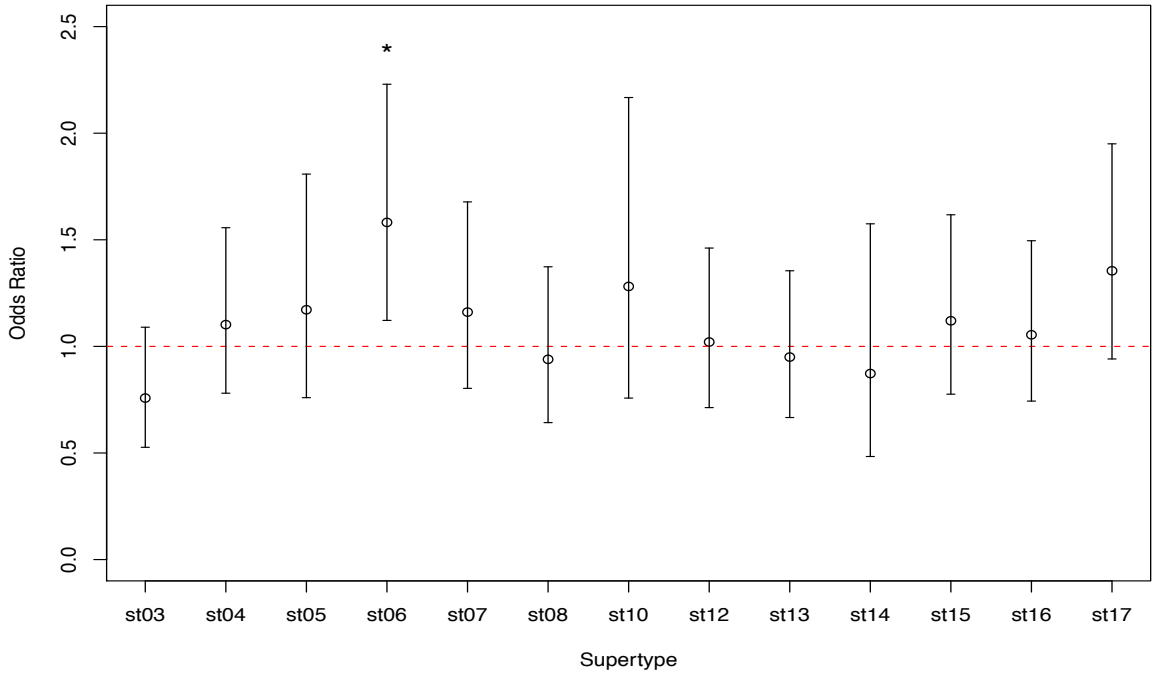
* odds ratio is significantly different from one.

Table S3.3 - Results of model selection based on AIC for generalized linear models with binomial errors, examining the influence of *Mhc*, host age and sex, and local infection risk on the probability of *P. circumflexum* infection.

predictor	coefficient	SE	test statistic	P-value	ΔAIC
supertype 3	0.283	0.400	0.706	0.480	-1.50
supertype 4	0.216	0.207	1.043	0.297	-0.92
supertype 5	0.283	0.276	1.027	0.304	-0.93
supertype 6	**0.553**	**0.197**	**2.807**	**0.005**	**+6.03**
supertype 7	0.213	0.244	0.870	0.384	-1.24
supertype 8	0.343	0.307	1.120	0.263	-0.74
supertype 10	0.221	0.328	0.672	0.501	-1.54
supertype 12	0.165	0.225	0.734	0.463	-1.47
supertype 13	0.180	0.258	0.696	0.486	-1.52
supertype 14	0.322	0.381	0.846	0.398	-1.28
supertype 15	0.173	0.249	0.697	0.486	-1.51
supertype 16	0.183	0.209	0.877	0.380	-1.23
supertype 17	0.325	0.210	1.548	0.12	+0.39
total number of supertypes	-0.101	0.085	-1.197	0.232	-0.56
total number of supertypes^2	-0.002	0.030	-0.076	0.940	-1.99
total number of alleles	0.048	0.026	1.849	0.064	+1.43
total number of alleles^2	-0.006	0.005	-1.284	0.199	-0.30
local infection risk	**4.369**	**0.455**	**9.600**	**<0.001**	**+105.85**
year	0.098	0.208	0.470	0.638	-1.78
sex	0.197	0.200	0.986	0.324	-1.03
age class	**0.273**	**0.121**	**2.248**	**0.025**	**+3.08**
age class^2	-0.143	0.209	-0.686	0.493	-1.53

(Test statistics reported are z-statistics, with P-values of coefficients calculated from Wald tests. ΔAIC is the change in the AIC value after removal of that term from the model. Terms shown in bold are the variables retained in the minimum adequate model.)

Figure S3.2 – Estimates of effect size (odds ratios with 95% CI) of each *Mhc* supertype on the probability of *P. circumflexum* infection.



Odds ratios above one indicate positive associations between *Mhc* supertypes and infection. Odds ratios below one indicate negative associations between *Mhc* supertypes and infection.

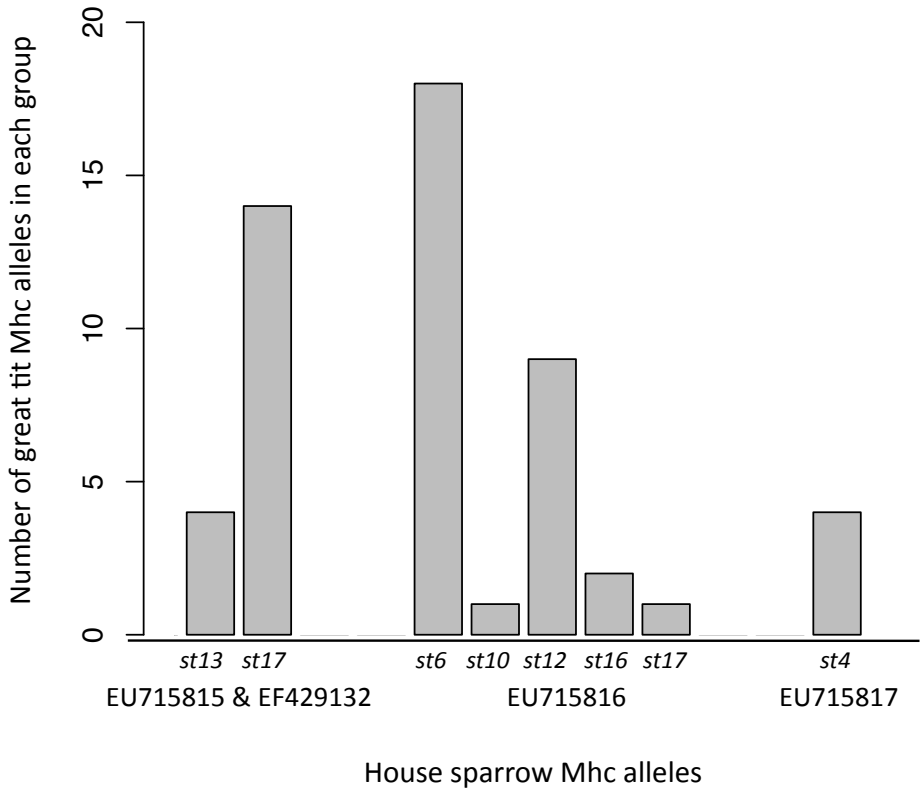
* odds ratio is significantly different from one.

Table S3.4 – Results of model selection based on AIC for generalized linear models with binomial errors, examining the influence of *P. circumflexum* infection risk, host infection status and presence of *Mhc* supertype 6 on host survival rates.

Predictor	Coefficient	SE	Test Statistic	P-value	ΔAIC
LIR	**-0.428**	**0.453**	**-0.945**	**0.344**	**+1.09 ***
IS	**0.103**	**0.202**	**0.509**	**0.611**	**+0.45 ***
supertype 6	0.119	0.181	0.659	0.510	-1.57
LIR * IS	**1.900**	**0.940**	**2.022**	**0.043**	**+2.19**
LIR * supertype 6	0.851	0.961	0.885	0.376	-1.22
IS * supertype 6	-0.221	0.376	-0.588	0.557	-1.65
LIR * IS * supertype 6	-0.429	1.939	-0.221	0.825	-1.95

'LIR' = local infection risk, 'IS' = infection status. Test statistics reported are z-statistics, with P-values of coefficients calculated from Wald tests. ΔAIC is the change in the AIC value after removal of that term from the model. Terms shown in bold are the variables retained in the minimum adequate model. Values indicated by * are the change in AIC for removal of all instances of that term from the model.

Figure S3.3 – Plot showing the great tit *Mhc* supertypes (and number of great tit *Mhc* alleles) that clustered with each of the four house sparrow *Mhc* alleles following K-means clustering. 'st' indicates supertype.



Chapter 4

Mhc-linked survival and lifetime reproductive success in a wild population of great tits

Molecular Ecology (2013) **22,** 384-396.

Mhc-linked survival and lifetime reproductive success in a wild population of great tits

Irem Sepil[1], Shelly Lachish[1], Ben C Sheldon[1]

1. Edward Grey Institute, Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, Oxford, UK

Abstract

Major histocompatibility complex (*Mhc*) genes are frequently used as a model for adaptive genetic diversity. Although associations between *Mhc* and disease resistance are frequently documented, little is known about the fitness consequences of *Mhc* variation in wild populations. Further, most work to date has involved testing associations between *Mhc* genotypes and fitness components. However, the functional diversity of the *Mhc*, and hence the mechanism by which selection on *Mhc* acts, depends on how genotypes map to the functional properties of *Mhc* molecules. Here, we test three hypotheses that relate *Mhc* diversity to fitness: (1) the maximal diversity hypothesis; (2) the optimal diversity hypothesis, and (3) effect of specific *Mhc* types. We combine mark-recapture methods with analysis of long-term breeding data to investigate the effects of *Mhc* class I functional diversity (*Mhc* supertypes) on individual fitness in a wild great tit (*Parus major*) population. We found that the presence of three different *Mhc* supertypes was associated with three different components of individual fitness: survival, annual recruitment and lifetime reproductive success (LRS). Great tits possessing *Mhc* supertype 3 experienced higher survival rates than those that did not, whereas individuals with *Mhc* supertype 6 experienced higher LRS and were more likely to recruit offspring each year. Conversely, great tits that possessed *Mhc* supertype 5 had reduced LRS. We found no evidence

for a selective advantage of *Mhc* diversity, either in terms of maximal or optimal supertype diversity. Our results support the suggestion that specific *Mhc* types are an important determinant of individual fitness.

Keywords: major histocompatibility complex (*Mhc*); survival; great tit (*Parus major)*; lifetime reproductive success (LRS), mark-recapture models; *Mhc* supertype

Introduction

Understanding the genetic basis of fitness differences has been a major goal for evolutionary biologists over the last two decades (Falconer & Mackay 1996; Ellegren & Sheldon 2008; Kruuk *et al.* 2008). Major histocompatibility complex (*Mhc*) genes are regarded as essential genes for individual fitness because, as the central component of the vertebrate adaptive immune system, they play an important role in parasite resistance (Trowsdale & Parham 2004), and thus can have direct effects on an individual's survival or reproductive success (Eizaguirre *et al.* 2009). *Mhc* genes encode cell-surface proteins responsible for the recognition and presentation of foreign antigens to T-lymphocytes, which then initiate an immune response against parasites (Potts & Wakeland 1990). The *Mhc* is the most polymorphic gene group in vertebrates, and the fundamental role of the *Mhc* in immune function and pathogen defense suggests the hypothesis that *Mhc* polymorphism is driven by the selective forces imposed by parasites and their interaction with their hosts (Takahata & Nei 1990; Edwards & Hedrick 1998; Bernatchez & Landry 2003; Piertney & Oliver 2006). Studies encompassing a range of organisms including model species (humans: Thursz *et al.* 1997; mice: Penn *et al.* 2002; McClelland *et al.* 2003) as well as natural populations of non-model organisms (fish: Dionne *et al.* 2009; birds: Bonneaud *et al.* 2006; Loiseau *et al.* 2011; mammals: Tollenaere *et al.* 2008; Oliver *et al.* 2009) have demonstrated associations between *Mhc* genotype and resistance or susceptibility to parasitic infections. Although such disease resistance studies have been crucial

for developing our understanding of wildlife immunogenetics, their findings do not allow a straightforward inference as to how *Mhc* affects individual fitness in wild populations.

Organisms are expected to encounter a diverse range of parasites in their natural environment, leading to the expectation of extreme diversity at *Mhc*, since different *Mhc* molecules bind different sets of peptides; hence specific *Mhc* alleles provide resistance to specific parasites (reviewed in Milinski 2006). Therefore, it has been proposed that high allelic diversity enables an immunogenic response to a greater variety of parasites and is thus advantageous for individuals facing heterogeneous pathogenic pressures (Doherty & Zinkernagel 1975). Despite this simple expectation, several studies have found that individuals possessing an intermediate number of *Mhc* alleles, rather than the maximum number, have higher relative fitness (Wegner *et al.* 2003; Madsen & Ujvari 2006; Wegner *et al.* 2008; Kalbe *et al.* 2009; Kloch *et al.* 2010). This pattern is consistent with the model proposed by Nowak *et al.* (1992) which suggests reduced fitness for individuals with high *Mhc* diversity resulting from excessive T-cell elimination during negative selection in thymus. Lastly, individual fitness may be dependent on resistance to the key parasites in the environment. Hence the carriers of specific *Mhc* alleles might have higher fitness in the population (Bodmer 1972; Slade & McCallum 1992).

Most disease resistance studies in natural populations examine host resistance against just one or few pathogens only, and do not assess the fitness consequences of *Mhc*-linked parasite resistance. Moreover, although multiple parasitic challenges are expected to occur in nature, only a few are likely to exert strong selective pressures on the host. Nevertheless, it is seldom known whether the parasites considered in *Mhc* studies include the key parasites in the environment. Hence rather little is known about *Mhc*-dependent fitness differences in wild populations (Brouwer *et al.* 2010). Since the hypotheses concerning maximal or optimal allelic diversity are particularly important for individuals facing heterogeneous pathogenic pressures, most work conducted to date fails to reveal whether allelic diversity is linked with individual fitness (Thoß *et al.* 2011). A final important consideration is that positive associations between

Mhc alleles and infection are difficult to interpret in natural populations, as they might be either the result of susceptibility alleles, or a consequence of quantitative-disease resistance alleles that limit the deleterious effects of the disease (Westerdahl *et al.* 2011; Sepil *et al.* submitted). Hence, investigating *Mhc*-fitness correlations is necessary for discriminating advantageous alleles from disadvantageous alleles.

An alternative, and potentially more robust approach to understanding how *Mhc* variation affects individual fitness is to investigate the role that the *Mhc* plays in individual variation in survival and lifetime reproductive success (LRS). To date, relatively few studies have investigated *Mhc*-fitness correlations, and many of those that have were carried out in semi-natural settings (e.g. Sauermann *et al.* 2001; Wegner *et al.* 2008; Eizaguirre *et al.* 2009; Kalbe *et al.* 2009; Worley *et al.* 2010; Thoß *et al.* 2011). Semi-natural settings are likely to provide stable and benign environments that differ markedly in parasitic fauna from natural conditions, and thus present situations where selection is unlikely to operate as it would in the wild. For this reason, there is still a clear need for studies investigating *Mhc*-linked survival and LRS in wild populations.

Unfortunately, determining *Mhc*-fitness correlations is a major challenge in the wild, as it necessitates long-term, individual based study systems where individuals are marked, breeding attempts are recorded, and the fate of offspring is known (Clutton-Brock & Sheldon 2010). For this reason, studies assessing *Mhc*-linked fitness in natural populations have, to date, been limited to few species. In a large, unmanaged population of Soay sheep (*Ovis aries L.*), the presence of specific *Mhc* alleles, but not individual heterozygosity, was the critical factor determining mortality in lambs and yearlings (Paterson *et al.* 1998). Similarly, in an island population of Seychelles warblers (*Acrocephalus sechellensis*), offspring carrying a specific allele had a fivefold longer life expectancy than offspring without this allele; there was also a survival advantage for juveniles with higher *Mhc* diversity, but this pattern was absent during the adult life stage (Brouwer *et al.* 2010). In both these studies fitness was assessed in terms of individual survival only, while measures of reproductive performance and integrated measures

such as LRS were not taken into account; consequently the link between *Mhc* genes and individual fitness is not necessarily clear. In a recent study Radwan *et al.* (2012) investigated the link between avian malaria infection, LRS, lifespan and *Mhc* class II diversity in an island population of collared flycatchers (*Ficedula albicollis*). Although they detected a negative relationship between the prevalence of blood parasites and functional *Mhc* diversity, they found no evidence for an association between lifespan, LRS and *Mhc*.

A final difficulty with much of the work to date is that it has assessed links between genotypic variation and fitness, despite the fact that selection on *Mhc* is likely to act via the phenotypic effects of the underlying genes (Spurgin & Richardson 2010). Although the aim of most *Mhc* studies has been to explain genetic diversity, treating alleles as if they are equally distinct in terms of their phenotypic effects is questionable: sometimes small sequence differences will result in large differences in the functional properties of a protein, and vice versa. Because the functional properties of *Mhc* molecules are quite well understood, it is possible to take an alternative approach and characterize alleles in terms of the predicted properties of their antigen-binding sites based on the amino-acid sequences inferred from the DNA sequences (Trachtenberg *et al.* 2003; Lenz 2011). Analysis of these properties can then address the similarity of alleles in terms of their predicted functional effects, and alleles with similar effects can be clustered into groups, or 'supertypes' (Schwensow *et al.* 2007; Huchard *et al.* 2008). Investigating the relationship between *Mhc* supertypes and fitness is thus an analysis of the link between expected *Mhc* functional properties and individual fitness.

In the present work we combine mark-recapture methods with analysis of long-term breeding data to examine the survival and reproductive fitness consequences of *Mhc* class I supertypes for great tits breeding within a long-term study population at Wytham Woods, near Oxford. We focused on the *Mhc* class I region, because *Plasmodium* spp., *Leucocytozoon* spp. and avian pox-virus infections are studied concurrently in this population of great tits, and it is highly likely that *Mhc* class I proteins present antigens from these intra-cellular parasites (Hughes & Yeager 1998; Penn & Potts 1999). Previous work on this population has found

extreme complexity at the *Mhc* class I loci of the great tit, both in terms of allelic diversity and gene number (Sepil *et al.* 2012). This work also revealed evidence that functional *Mhc* alleles were under balancing selection, implying that this species can provide valuable information on the processes by which *Mhc* effect individual fitness. We aimed to test the associations between *Mhc* class I variation, defined in terms of the functional diversity of this gene family, and fitness components in order to test the hypotheses that selection: (i) maximizes *Mhc* diversity; (ii) optimizes *Mhc* diversity, or (iii) favours specific functional variants.

Materials and Methods

Study Population

We investigated *Mhc*-fitness associations in a dataset comprising 618 great tits captured as part of a long-term study conducted at Wytham Woods, near Oxford, UK (51°46'N, 1°20'W). The c. 380 ha study site is continuous mixed deciduous woodland, in which 1020 nest boxes are distributed at variable densities. In this part of the UK, great tits have a synchronous, annual breeding season (April to June), and are single-brooded. Throughout the breeding season, all nestboxes were checked weekly to obtain records of lay date, clutch size, hatching date, brood size and number of fledglings. Breeding birds were captured between day 6 and 14 of the nestling phase, either within the nestbox by hand or using traps, or with mist nets in front of the nest entrance. All captured individuals, and all locally born nestlings that survive to 15 days of age, were marked with unique metal rings. Sex of individuals was determined based on the presence (male) or absence (female) of a brood patch, while age class (one year olds, two year olds, or older) was determined using plumage characteristics (Svensson 1992) and ringing records. Blood samples for *Mhc* genotyping were collected by wing or jugular venipuncture from breeding birds under UK Home Office license and stored in SET Buffer at -80°C until DNA extraction. Total genomic DNA was extracted using standard ammonium acetate method and stored in AE Buffer (Qiagen).

Mhc genotyping

All statistical analyses in this paper were performed on the genetic data published in Sepil *et al.* (2012), and more complete details of all molecular protocols can be found therein; briefer details are presented here. Bidirectional 454 pyrosequencing method was employed to amplify a 212-221 basepair fragment (without primers) of *Mhc* class I exon 3 (78%) in great tits, using the 16 lanes of a Pico Titer Plate gasket (Sepil *et al.* 2012). Individuals in each lane were differentiated by the use of 10 multiplex identifiers (MID), inserted in the forward and reverse fusion primers. Forward (5'-CGTATCGCCTCCCTCGCGCCATCAG- MID - TTMYGGCTGTGACCTCCTG-3') and reverse (5'-CTATGCGCCTTGCCAGCCCGCTCAG –MID - TTGCGCTYCAGCTCTTTC -3') fusion primers were composed of the GS FLX Titanium Primer sequence, a 10bp MID sequence and the template specific primer sequence (underlined). We used published (Westerdahl *et al.* 2004), as well as unpublished, primers to isolate complete and partial *Mhc* class I exon 3 sequences. Using the consensus from over 100 sequences, we designed a degenerate primer pair (the template-specific primers) that would potentially amplify all the expressed alleles that are present in the great tit (Sepil *et al.* 2012). *Mhc* class I exon 3 was chosen as the target region, as it contains the antigen-binding codons. Polymerase chain reaction (PCR) was performed as described in Sepil *et al.* (2012). Samples with different MID combinations were pooled together in approximate equimolar quantities, a single pool was prepared for each lane, and the pools were sent for bidirectional 454 pyrosequencing at Genomic Services, Wellcome Trust Centre for Human Genetics, University of Oxford.

Since high numbers of artefacts can be generated during PCR and 454 pyrosequencing, we adopted a 5-step variant validation procedure to differentiate true alleles from artefacts (for more details see Sepil *et al.* 2012). From the 638501 reads that the experiments generated, final genotypes were based on 214357 reads (34% of the initial read number); 857 individuals passed our reliability criteria and a total of 862 *Mhc* class I alleles (Genbank: JQ034624 - JQ035485) were detected. Of these 857 individuals, 618 were captured in an extensive sampling of the

breeding population in 2008 or 2009, and were hence used in the present paper. From 12 duplicates, the repeatability of the method was estimated as 0.94, indicating that the coverage was sufficient for reliable genotyping. Functional *Mhc* alleles were differentiated from non-functional alleles using allele sequence information, examining phylogenetic relationships and the application of historical selection tests (for more details see Sepil *et al.* 2012). Overall 755 alleles were classified as functional and the presence of at least 16 functional loci was shown, together with a pseudogene family and putatively non-functional alleles. Lastly, the functionally similar alleles were grouped into 17 supertypes, using the methods described by Doytchinova & Flower (2005) and Jombart *et al.* (2010); see Sepil *et al.* (2012). The association between specific *Mhc* supertypes and individual fitness components was assessed for 13 supertypes; the remaining four supertypes (supertypes 1, 2, 9 and 11) were removed from the analysis as they were almost fixed in the population with frequencies higher than 95%.

Associations between Mhc and survival rates

The standard Cormack–Jolly–Seber (CJS) framework was employed to estimate annual survival probabilities of the 618 genotyped great tits captured in 2008 and 2009. We used the parametric bootstrap procedure available in program MARK (White & Burnham 1999) to assess the fit of the CJS model (with time-dependent survival and recapture rates) to the data and to calculate the variance inflation factor ($\hat{c}$; observed deviance/mean deviance from bootstrap replicates), which indicates the degree of overdispersion in the data. There was no evidence for significant lack of fit of this general model (1000 bootstrap replicates, $P = 0.102$), nor any indication of overdispersion ($\hat{c} = 1.05$). Previous mark-recapture analyses of this population have revealed that great tit survival rates differ between males and females and with age, but vary in a similar manner over time for individuals (Bouwhuis *et al.* 2012). Recapture rates meanwhile, were shown to be time-invariant and sex-dependent (individuals are captured in association with active nests, and females are more readily captured). Accordingly, we began model selection by parameterising recapture rates to be sex-dependent (*p(sex)*) and allowing for time variation and

full sex- and age-effects in survival rates (*φ(age*sex+t)*, incorporating three age classes; one year olds, two year olds and individuals aged 3+ years). This model was much better supported by the data (see Table 4.1, model 4 vs. model 17) and became our starting model for subsequent model selection. To test the hypotheses of maximal and optimal *Mhc* diversity on great tit survival rates, we fitted two models, which allowed survival rates to vary either as a linear or non-linear (quadratic) function of the standardized total number of supertypes an individual possessed. Then, to assess the influence of specific *Mhc* genotypes on survival rates we added effects of each of the 13 *Mhc* supertypes described above to this base model. All models were fitted to the data using program E-SURGE (Choquet 2009). To ensure convergence of models on the global minima, models were run using repeated random initial values (multiple random option with N=8; Choquet 2007). Model selection was based on Akaike Information Criteria corrected for small sample size (AICc; Burnham & Anderson 2002). The model in the candidate model set with the lowest AICc value is the most parsimonious, 'best', model supported by the data, providing the best balance between bias due to under-fitting and lost precision due to over-fitting (Burnham & Anderson 2002). Models with differences in AICc values of more than two indicate that one model is considerably better supported by the data, while where models differed in AICc by less than two the model with fewer terms was considered the most parsimonious model (Burnham & Anderson 2002). Inspection of the confidence intervals of the beta coefficients (effect sizes) was used to confirm the strength of association between *Mhc* genotypes and survival rates in the top models.

Associations between Mhc and reproductive success

We examined associations between *Mhc* genotype and reproductive success in 1193 breeding attempts by 607 great tits that were captured in 2008 or 2009, and bred between the years 2004 and 2011 (11 individuals were excluded because of incomplete breeding data). We used generalized linear mixed effects models to investigate the effect of *Mhc* on the reproductive performance of great tits, measured at four consecutive stages of the breeding attempt. The

specific reproductive measures we considered were: (i) clutch size, modelled as a Poisson response with a log link; (ii) brood size, modelled as a Poisson response; (iii) the number of young fledged (fledging success), a Poisson response, adjusted for overdispersion; and (iv) recruitment success, a binary variable indicating whether any of the young recruited to the breeding population, modeled as a binomial response with a logit link. Recruitment of young for those individuals that had survived to breed in 2011 (N=26) was assessed using winter recapture records (records between September 2011 and March 2012) rather than recaptures from the breeding season and may thus be biased (as not all offspring that survive to breed are recaptured during their first winter, and not all offspring that survive the winter are recaptured as breeding birds). Data from these nesting attempts were nevertheless included to prevent a non-random bias regarding the birds that survived to breed in 2011, in comparison to the ones that did not. Removal of these breeding attempts from the dataset did not qualitatively change the results. Recruitment measured as in this paper only reflects fitness if emigration of young is independent of state (as suggested for the Wytham population by Verhulst *et al.* 1997 and Bouwhuis *et al.* 2009), and in this study independent of *Mhc* type. Although, we do not have data to relate emigration to *Mhc*, we have no evidence that local recruitment is biased with respect to *Mhc* genotype.

Models were performed using the R package 'lme4' with individual identity and year included as random effects to control for repeated observations of the same individual and temporal environmental heterogeneity. Starting models included the following fixed effects: the presence of each of the 13 *Mhc* supertypes described above (collinearity among supertypes was very low; phi coefficients ranged from 0.002 to 0.157), the total number of supertypes an individual possessed (*Mhc* diversity), and the square of the total number of supertypes, to test for possible nonlinear relationships between reproductive measures and *Mhc* diversity. Supertype number and its square were mean-centered before inclusion in the model. Local breeding density (measured as nestbox density in each of nine Wytham woodland sections) was fitted as a covariate in all starting models as it is known to significantly influence several

measures of reproductive performance in this species (Bouwhuis *et al.* 2009, 2010), while sex was included to account for sex-specific effects. In addition, we fitted clutch size as a covariate when modeling brood size, brood size as a covariate when modeling fledging success, and fledging success as a covariate when modeling recruitment success. Hence each analysis addresses the additive effect of *Mhc* supertype at that reproductive stage, controlling for any influence on preceding stages. We also repeated the same analyses without fitting the previous stage as a covariate to consider the total effect of *Mhc* at the reproductive stage.

We then investigated the effect of *Mhc* on LRS, defined as the total number of recruits produced over a lifetime. We used a generalized linear mixed effect model with Poisson errors adjusted for overdispersion, and included year of birth as a random effect to control for temporal environmental heterogeneity and cohort effects. The initial starting model included the following fixed effects: the presence of each of the 13 *Mhc* supertypes, the total number of supertypes an individual possessed, and the square of the total number of supertypes, sex and local breeding density. We also repeated the same analyses fitting the total number of breeding attempts recorded for the individual as a covariate, to account for differences in longevity among individuals.

We used Akaike's information criterion (AIC) to determine the combination of variables that best explained the data with minimal parameters. Model selection was performed by backward stepwise elimination and the fit of each new model was assessed by comparison of AIC values. Terms were eliminated from the model when their removal resulted in an improved fit (i.e. a <2 reduction in AIC, Burnham & Anderson 2002), and were retained if their removal resulted in an increase in AIC >2. Where the removal of a term resulted in a model with an approximately equal fit (i.e. a increase in AIC of <2) the model with fewer terms was considered the most parsimonious model (Burnham & Anderson 2002). To confirm the validity of the minimum model, removed variables were added individually to assess any potential improvement in the model fit.

Results

Associations between Mhc and survival rates

Results of mark–recapture analysis modelling the effect of *Mhc* on survival probabilities of great tits revealed little support for an association between *Mhc* diversity and survival (Table 4.1). Models including supertype number, either as a linear or quadratic covariate, received approximately equal support to the base model (Table 4.1, model 2 & model 6 vs. model 4), and the 95% confidence interval for the effect of *Mhc* diversity on survival rates included zero (Fig. 4.1, Table S4.1, Supporting information). Hence, there is little evidence to suggest that individuals that possess high or intermediate *Mhc* diversity are at an advantage in terms of survival.

In contrast, there was stronger support for models testing specific effects of *Mhc* supertypes on fitness components. The analyses revealed strong support for an effect of supertype 3 on great tit survival rates (Table 4.1; model 1 vs. model 4). The model including supertype 3 alone was 29 times better supported than the base model (Table 4.1, evidence ratio of model 4 and model 1 = 0.753 / 0.026 = 29.0), received 75% of the weight in the candidate set, and the 95% confidence intervals for the effect of supertype 3 on survival did not include zero (Fig. 4.1, Table S4.1, Supporting information). Individuals carrying supertype 3 had on average 60% greater probability of survival per year (odds ratio = 1.60 95% CI 1.17–2.18) than did those individuals that did not carry the supertype (Fig. 4.2). Although our analysis revealed that the model including supertype 13 received approximately equal support to the base model (Table 4.1, model 3 vs. model 4), its effect was weak, as the 95% confidence interval for its effect included zero (Fig. 4.1, Table 4.S1, Supporting information). There was no support for an effect of the remaining *Mhc* supertypes on great tit survival (Table 4.1).

Associations between Mhc and reproductive success

Results of the mixed effect models revealed no evidence that the reproductive performance of individuals varied according to *Mhc* supertype diversity. There was no effect of the total number of supertypes an individual possessed (maximal *Mhc* diversity), or of the square of the total number of supertypes (optimal *Mhc* diversity) on clutch size, hatching success, fledging success and recruitment success (Fig. 4.3a, Table 4.2, Table 4.S2 – 4.S4, Supporting information). Moreover, our analyses revealed no support for an association between specific *Mhc* supertypes and reproductive performance in terms of clutch size, hatching success and fledging success (Table 4.S2 – 4.S4, Supporting information). However, the results of the mixed effects models on recruitment success revealed a significant association between the probability of offspring recruitment and the presence of *Mhc* supertype 6 (Table 4.2, Fig. 4.3b), such that individuals carrying the supertype had on average 39% greater probability of producing a recruit (odds ratio = 1.39, 95% CI 1.07–1.80). Individuals carrying supertype 13 tended to have higher offspring recruitment probabilities (Table 4.2), although the effect was not substantial and the odds ratio confidence interval included one (odds ratio = 1.24, 95% CI 0.95–1.61). Individuals carrying supertype 5 tended to have lower offspring recruitment probabilities (Table 4.2); though again the effect was not substantial with the confidence interval for the odds ratio including one (odds ratio = 0.75, 95% CI 0.54–1.04). As expected, nest box density had a negative effect on clutch size and fledging success, and a positive effect on recruitment success, in line with earlier studies on this population (Szulkin *et al.* 2007; Bouwhuis *et al.* 2009). Reproductive performance did not vary between the sexes at any stage, but did vary significantly with the effect of the former stage (Table 4.2, Table 4.S2 – 4.S4, Supporting information). Removal of the former reproductive stages from the models, to consider the total effect of *Mhc* at the reproductive stage, did not qualitatively change the results of any analysis (results not shown).

Finally, results of the mixed effect models for LRS revealed little support for an association between *Mhc* diversity and the number of recruits produced over an individual's

lifetime. *Mhc*-diverse individuals tended to have higher number of recruits, however the effect was not substantial; and there was no effect of intermediate *Mhc* diversity (Table 4.3, Fig. 4.4a). Hence, we found little evidence suggesting that individuals that possess high or intermediate *Mhc* diversity are at an advantage in terms of offspring recruitment. However, our analyses revealed there to be significant associations between two different *Mhc* supertypes and LRS (Table 4.3). Individuals carrying *Mhc* supertype 6 had a 26% higher LRS (average ± SE: individuals with ST6: 0.97 ± 0.06 individuals without ST6: 0.77 ± 0.07), whereas individuals carrying *Mhc* supertype 5 had a 19% lower LRS (average ± SE: individuals with ST5: 0.85 ± 0.05, individuals without ST5: 1.04 ± 0.12) (Fig. 4.4b). No other supertype had a substantial effect on LRS (Table 4.3). Nest box density had a positive effect on recruit number, but it did not vary between the sexes (Table 4.3). Lastly, adding the total number of breeding attempts recorded for the individual as a covariate, did not qualitatively change the results, although unsurprisingly longevity became the strongest predictor of LRS (Table 4.S5, Supporting information).

Discussion

Understanding the adaptive importance of *Mhc* genes requires knowledge of their fitness effects on hosts in wild populations, but the complex nature of ecological systems poses a great challenge for studies investigating the causes and consequences of *Mhc* variation, due to the numerous confounding and interacting variables that determine fitness. In the present work, we investigated the role that the *Mhc* plays in host fitness in a wild great tit population, by incorporating analysis of *Mhc* class I supertypes, and a detailed investigation of host survival, reproductive performance and LRS, while controlling for confounding factors, such as spatiotemporal environmental heterogeneity and host-related effects. We found that the presence of three *Mhc* supertypes was associated with three different components of individual fitness: survival, annual recruitment probabilities and LRS. Great tits carrying *Mhc* supertype 3

had higher survival rates, while individuals carrying *Mhc* supertype 6 had higher LRS and were more likely to recruit an offspring at the brood level, and individuals with *Mhc* supertype 5 had lower LRS. In contrast we found no support for hypotheses of *Mhc* diversity linked to fitness.

Associations between *Mhc* genotypes and disease resistance or susceptibility are frequently reported, and assumed to be driven by selection imposed by parasites that negatively impact on host fitness. However, very few studies to date have investigated *Mhc*-linked host survival in wild populations (reviewed in Spurgin & Richardson 2010). In this study, using a robust mark-recapture modeling framework we were able to demonstrate a survival advantage for great tits carrying *Mhc* supertype 3, suggesting that the supertype might confer resistance against a specific parasite in the environment. Further studies are needed to determine the particular parasite associated with supertype 3 and to reveal the mechanism causing this supertype-specific survival advantage. As *Mhc* class I genes are expressed on the cell surface of almost all nucleated cells and present peptides from intracellular antigens, we expect it is most likely that *Mhc* supertype 3 will provide resistance against an intracellular parasite. *Haemoproteus* spp., *Leucocytozoon* spp. and avian pox infections have previously been recorded in this population, but little is known about their virulence in wild birds. Of particular interest, a novel form of avian pox has been spreading in this population since 2009, and is associated with reduced survival of infected great tits (Lachish *et al.* 2012); examining the potential for a link between avian pox infections and *Mhc* supertype 3 will be an exciting area for further research. Still, an exhaustive pathogen screening effort is likely to be required to identify the exact parasite involved.

Elsewhere, we have shown that two different *Mhc* supertypes, supertype 6 and supertype 17, are strongly associated with the prevalence of *Plasmodium circumflexum,* and *P. relictum* respectively, and suggested that supertype 6 confers quantitative resistance to *P. circumflexum* infections and supertype 17 confers qualitative resistance to *P. relictum* infections (Sepil *et al.* submitted). While *P. circumflexum* infections are associated with reduced survival, particularly during the acute stage of infection (Lachish *et al.* 2011; Sepil *et al.* submitted), *P.*

relictum infections are believed to be more benign, linked with reproductive costs in blue tits (Knowles *et al.* 2010a). Although we would thus expect *Mhc* supertype 6 to confer a survival advantage for its carriers, the highest mortality from infection is likely to occur when the birds are first exposed to the parasite, during the pre-breeding juvenile phase of life (Lachish *et al.* 2011). Hence the absence of an association between supertype 6 and survival rates of adults in the present study may not be surprising. Indeed, we found that great tits carrying supertype 6 had significantly higher LRS and offspring recruitment probabilities at the brood level, suggesting there is a survival advantage for juveniles that carry *Mhc* supertype 6. Therefore, the results presented here complement other studies on this system. The lack of an association between supertype 17 and reproductive success on the other hand, might simply reflect the fact that *P. relictum* infections have minor effects on their great tit hosts. Although experimental studies have provided evidence for reproductive costs of *P. relictum* infections in wild bird species (Knowles *et al.* 2010a), observational studies have mostly yielded inconclusive or negative results (Bensch *et al.* 2007; Marzal *et al.* 2008), suggesting that such effects might be too weak to detect in correlational studies. Taken together, these results illustrate the necessity of investigating the role that *Mhc* plays in individual variation in survival and LRS for inferring the fitness consequences of *Mhc*-linked parasite resistance.

We found that great tits carrying *Mhc* supertype 5 had significantly lower LRS, and that at the brood level, individuals carrying supertype 5 tended to have lower offspring recruitment probabilities, although the latter effect was not substantial. The negative relationship between the presence of a supertype and offspring recruitment suggests that *Mhc* supertype 5 might confer susceptibility to a parasite in the local environment. We suggest that this negative relationship could arise for two reasons. First, the reproductive effort of the breeding parents carrying supertype 5 might have been compromised due to parasite infection, which affected their probability of offspring recruitment. Evidence for reproductive costs under parental parasitic infection has been provided in several wild bird studies, suggesting that modulation of reproductive trade-offs by parasites is a common phenomenon with consequences for offspring

quality (Gustafsson *et al.* 1994; Descamps *et al.* 2009; Knowles *et al.* 2010b). A second possibility is that the offspring of individuals carrying *Mhc* supertype 5 might experience lower survival rates as a result of a mortality-inducing infection. Our mark-recapture modeling framework did not include juveniles so we were not able to investigate the link between supertype 5 and juvenile survival rates. However, support for the suggestion that *Mhc*-linked disease susceptibility can influence juvenile mortality has been provided by a recent study of semi-natural red junglefowl. Worley *et al.* (2010) found that an *Mhc* haplotype had detrimental effects on juvenile survival during an outbreak of coccidiosis, and that birds homozygous for this haplotype had lower survival than all other genotypes. To differentiate between these two possibilities requires identifying the specific parasite associated with *Mhc* supertype 5, which as discussed above, would necessitate an exhaustive pathogen screening effort. Characterising the *Mhc* of fledglings and assessing their survival rates would also facilitate this investigation.

Although we found *Mhc*-linked fitness effects in terms of host survival, recruitment success and LRS, results from the three breeding stages underlying reproductive success revealed no association between *Mhc* supertype and reproductive traits. Our data suggest that neither *Mhc* diversity (either in terms of maximal and optimal diversity), nor the presence of specific *Mhc* supertypes, have any effect on clutch size, hatching success and fledging success, at the brood level. Nevertheless, the effect of *Mhc* on these stages of the breeding attempt might become apparent if *Mhc* compatibility between the social pair is assessed, rather than the *Mhc* type of each parent. Preference for *Mhc* dissimilar mates could act to increase the *Mhc* heterozygosity of the progeny or it might be an adaptation to avoid mating with kin, preventing inbreeding depression in the offspring (Penn & Potts 1999). *Mhc*-based disassortative mating has been reported across taxa, both in model species (human: Ober *et al.* 1997; Chaix *et al.* 2008; mouse: Potts *et al.* 1991) and natural populations of non-model species (fish: Consuegra & Garcia de Leaniz 2008; reptiles: Olsson *et al.* 2003; mammals: Schwensow *et al.* 2008; birds: Freeman-Gallant *et al.* 2003). Investigating the potential for *Mhc*-based mating preferences in this population of great tits is the subject of ongoing work. It is worth noting however, that

brood level reproductive success measures do not reveal the overall contribution reproductive attempts make to parental fitness, as they only reflect the outcome of a single breeding attempt. Therefore LRS is a more suitable measure for assessing *Mhc*-fitness correlations, and an assessment of *Mhc* compatibility is impractical under this approach, as individuals commonly mate with multiple different partners over their lifetimes.

Two other studies that have investigated *Mhc*-linked host survival in wild populations found positive associations between certain *Mhc* alleles and life expectancy. In a large unmanaged population of Soay sheep, Paterson *et al.* (1998) showed that a specific *Mhc* allele significantly associated with strongyle parasite resistance in lambs and yearlings had substantial links with juvenile survival. Likewise, Brouwer *et al.* (2010) found that Seychelles warbler offspring carrying a specific *Mhc* allele had five times higher life expectancy than offspring without this allele, however, as in the present study, the putative pathogen assumed to be driving this pattern was not identified. The consistency between these two studies and the results of this paper suggest that, in natural populations host survival may be dependent on resistance to the 'key' parasite in the environment, even though individuals are expected to combat a diverse range of parasites. In contrast, Banks *et al.* (2010) detected increased survival for individuals heterozygous at a *Mhc*-linked microsatellite marker in a natural population of mountain brushtail possums (*Trichosurus cunninghami*). Similarly in a long-term study of free-living Alpine chamois (*Rupicapra rupicapra*), Schaschl *et al.* (2012) showed that male chamois survive longer if heterozygous at the *Mhc* class II DRB locus. We found no fitness advantage for individuals possessing maximal *Mhc* diversity, and no evidence for an intermediate, 'optimal', diversity since both supertype number and its square had no effect on survival, individual reproductive performance and LRS amongst breeding birds. Likewise, Radwan *et al.* (2012) found no evidence for an association between functional *Mhc* diversity and LRS in collared flycatchers. Our result should, however, be interpreted with caution, as we were not able to assess *Mhc*-linked juvenile survival, the life stage during which mortality is highest (Perrins 1965). Brouwer *et al.* (2010) showed that the survival advantage for Seychelles

warblers with high *Mhc* diversity was specific for juveniles, with the association absent for adults. Indeed, our results revealed a weak positive relationship between *Mhc* diversity, adult survival rates and LRS. As discussed above, investigating the effects of *Mhc* on juvenile survival will be necessary to provide further insight into the extent of *Mhc*-linked host fitness.

In the present paper, we investigated *Mhc*-fitness correlations in a species that harbors many *Mhc* class I loci; indeed, our analyses (Sepil *et al.* 2012) suggest that at least 16 functional loci are present in the great tit. Species with complex *Mhc* regions are particularly intriguing for selection studies as they provide valuable information on the selective forces maintaining *Mhc* diversity (Spurgin & Richardson 2010). However, this degree of complexity in the *Mhc* region makes it impractical to take a locus-specific approach to *Mhc* characterization. Moreover, in species with many *Mhc* loci, the genes are often tightly linked to each other and alleles are shared among loci. For these reasons, simultaneous analysis of multiple loci is likely to be a rewarding approach for understanding the selective forces acting on the *Mhc* region in such species. A notable disadvantage of this approach, however, is the fact that simple hypothesis such as that of heterozygote advantage at a particular *Mhc* locus cannot be tested. Hence, studies such as ours, which analyze multiple *Mhc* loci simultaneously, cannot discount the possibility that associations between fitness components and heterozygosity at certain loci may exist. Nevertheless, since overall genetic diversity will often reflect heterozygosity across multiple loci, several previous studies have used the total number of *Mhc* alleles an individual possesses to test the heterozygote advantage hypothesis in species with multilocus *Mhc* systems (see for example, Westerdahl *et al.* 2005; Loiseau *et al.* 2008; Kalbe *et al.* 2009). Many of these studies provide support for the hypothesis that pathogen resistance is maximised by an intermediate number of *Mhc* alleles (Reusch *et al.* 2001; Madsen & Ujvari 2006; ; Kloch *et al.* 2010), while a smaller number of such studies support the hypothesis that maximal genetic diversity predicts resistance to parasites (Westerdahl *et al.* 2005; Radwan *et al.* 2012).

In this paper, we took a different approach, involving characterizing *Mhc* variants in terms of their predicted protein phenotypes, and then pooling the variants into functionally

similar types. Hence, this approach investigates the effects of *Mhc* class I phenotypic diversity, rather than genetic diversity. Since selection on *Mhc* is likely to act via the phenotypic effects of the underlying genes (Spurgin & Richardson 2010), this approach is both justified and appropriate. For instance, Trachtenberg *et al.* (2003) revealed that human leukocyte antigen (HLA) supertypes alone confer an advantage in responding to HIV infection, independent of the contribution of single HLA alleles. Therefore, we believe that the supertyping approach is an advance on the traditional approaches that examine associations between *Mhc* allelic diversity/individual *Mhc* alleles and fitness measures. Our results based on this approach revealed little support for an association between overall *Mhc* diversity and individual fitness, either in terms of intermediate or maximal diversity. It is worth noting however, that the classification of *Mhc* alleles into supertypes could potentially be improved, since the partial exon 3 sequences did not cover the section ranging from amino acid positions 5-10, a highly polymorphic region that affects the antigen binding capabilities of alleles. Amplifying the entire exon is likely to be critical for future studies.

The results of this study instead lend support to hypotheses that suggest there is a selective advantage for individual *Mhc* types. The n*egative-frequency dependent selection hypothesis* proposes a coevolutionary arms race in which parasites and *Mhc* alleles fluctuate in frequency, such that *Mhc* diversity is maintained via a dynamic process (Slade & McCallum 1992). The *fluctuating selection hypothesis*, on the other hand, proposes that spatiotemporal heterogeneity of parasites leads to differential selection on *Mhc* types, such that the enormous diversity of parasites in space and time drives polymorphism at *Mhc* (Hedrick 2002). Previously we have shown that the *Mhc* class I region is under strong balancing selection in this population of great tits (Sepil *et al.* 2012), however we couldn't elucidate the exact selective mechanism maintaining *Mhc* diversity in this population. Extending our study over several more generations might enable differentiation between these two mechanisms.

To our knowledge, this study is the first to estimate the fitness consequences of functional *Mhc* variants for both survival and LRS in a wild population. Our findings show that

certain *Mhc* supertypes can have different effects on host fitness, possibly as a result of supertype-specific disease resistance or susceptibility, whereas we found no support for hypotheses that link *Mhc* diversity to fitness. Our study illustrates the importance of integrating long-term, individual based, studies of natural populations with detailed genetic analysis, to reveal the associations between *Mhc* functional variation and host fitness in complex ecological systems. Such studies are crucial for understanding of the processes underlying *Mhc* evolution, and elucidating the mechanisms by which selection operates to maintain *Mhc* allelic diversity in natural populations.

Acknowledgements

We are very grateful to the Wytham fieldworkers who collected the data and materials for this study. We thank Alicia Davies and Simon Lee for laboratory assistance. We also thank Sandra Bouwhuis, Kaan Oz and three anonymous reviewers for helpful discussion and advice. This work was partially funded by NERC grants (NER/A/S/2002/00877 and NE/F005725/1) to BCS.

References

Banks SC, Dubach J, Viggers KL, Lindenmayer DB (2010) Adult survival and microsatellite diversity in possums: effects of major histocompatibility complex-linked microsatellite diversity but not multilocus inbreeding estimators. *Oecologia*, **162,** 359-370.

Bensch S, Waldenström J, Jonzén N, *et al.* (2007) Temporal dynamics and diversity of avian malaria parasites in a single host species. *The Journal of Animal Ecology*, **76**, 112-122.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16**, 363-377.

Bodmer W (1972) Evolutionary significance of the HL-A system. *Nature*, **237,** 139-145.

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

Bouwhuis S, Charmantier A, Verhulst S, Sheldon BC (2010) Individual variation in rates of senescence: natal origin effects and disposable soma in a wild bird population. *Journal of Animal Ecology*, **79**, 1251-1261.

Bouwhuis S, Choquet R, Sheldon BC, Verhulst S (2012) The forms and fitness cost of senescence: age-specific recapture, survival, reproduction, and reproductive value in a wild bird population. *American Naturalist*, **179**, E15-E27.

Bouwhuis S, Sheldon BC, Verhulst S, Charmantier A (2009) Great tits growing old: selective disappearance and the partitioning of senescence to stages within the breeding cycle. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 2769-2777.

Brouwer L, Barr I, van de Pol M *et al.* (2010) MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, **19**, 3444-3455.

Burnham K, Anderson D (2002) *Model Selection and Multimodel Inference: a Practical Information-theoretic Approach*. Springer-Verlag, New York.

Chaix R, Cao C, Donnelly P (2008) Is mate choice in humans MHC-dependent? *PLoS Genetics*, **4**, e1000184.

Choquet R (2007) *E-SURGE 1.0 User's Manual*. Montpelier: CEFE.

Choquet R (2009) *Program E-SUREGE: A Software Application for Fitting Multievent Models. In: Thomspon DL, Cooch EG, Conroy MJ, Editors. Modelling Demographic Processes in Marked Populations.* New York: Springer.

Clutton-Brock T, Sheldon BC (2010) Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, **25**, 562-573.

Consuegra S & Garcia de Leaniz C (2008) MHC-mediated mate choice increases parasite resistance in salmon. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 1397-1403.

Descamps S, Gilchrist HG, Bêty J *et al.* (2009) Costs of reproduction in a long-lived bird: large clutch size is associated with low survival in the presence of a highly virulent disease. *Biology Letters*, **5**, 278-281.

Dionne M, Miller KM, Dodson JJ, Bernatchez L (2009) MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 1555-1565.

Doherty P, Zinkernagel R (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, **256**, 50-52.

Doytchinova IA, Flower DR (2005) In silico identification of supertypes for class II MHCs. *Journal of Immunology*, **174**, 7085-7095.

Edwards SV, Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology & Evolution*, **13**, 305-311.

Eizaguirre C, Yeates SE, Lenz TL *et al.* (2009) MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology*, **18**, 3316-3329.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**, 169-175.

Falconer D, Mackay T (1996) *Introduction to Quantitative Genetics*. Essex: Longman.

Freeman-Gallant CR, Meguerdichian M, Wheelwright NT, Sollecito SV (2003) Social pairing and female mating fidelity predicted by restriction fragment length polymorphism similarity at the major histocompatibility complex in a songbird. *Molecular Ecology*, **12**, 3077-3083.

Gustafsson L, Nordling D, Andersson MS *et al.* (1994) Infectious diseases, reproductive effort and the cost of reproduction in birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **346**, 323-331.

Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution,* **56,** 1902-1908.

Huchard E, Weill M, Cowlishaw G *et al.* (2008) Polymorphism, haplotype composition, and selection in the Mhc-DRB of wild baboons. *Immunogenetics*, **60**, 585-598.

Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics,* **32,** 415-435.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Kalbe M, Eizaguirre C, Dankert I *et al.* (2009) Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 925-934.

Kloch A, Babik W, Bajer A *et al.* (2010) Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus. Molecular Ecology*, **19 Suppl 1**, 255-265.

Knowles SCL, Palinauskas V, Sheldon BC (2010a) Chronic malaria infections increase family inequalities and reduce parental fitness: experimental evidence from a wild bird population. *Journal of Evolutionary Biology*, **23**, 557-569.

Knowles SCL, Wood MJ, Sheldon BC (2010b) Context-dependent effects of parental effort on malaria infection in a wild bird population, and their role in reproductive trade-offs. *Oecologia*, **164**, 87-97.

Kruuk LEB, Slate J, Wilson AJ (2008) New Answers for Old Questions: The Evolutionary Quantitative Genetics of Wild Animal Populations. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 525-548.

Lachish S, Knowles SCL, Alves R *et al.* (2011) Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. *Journal of Animal Ecology*, **80,** 1196-1206.

Lachish S, Lawson B, Cunningham AA, Sheldon BC (2012) The epidemiology of Paridae pox in a woodland setting. *Plos One* (In press)

Lenz TL (2011) Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution*, **65**, 2380-2390.

Loiseau C, Zoorob R, Garnier S *et al.* (2008) Antagonistic effects of *Mhc* class I allele on malaria-infected house sparrows. *Ecology Letters*, **11,** 258-265.

Loiseau C, Zoorob R, Robert A *et al.* (2011) *Plasmodium relictum* infection and MHC diversity in the house sparrow (*Passer domesticus*). *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1264-1272.

Madsen T, Ujvari B (2006) MHC class I variation associates with parasite resistance and longevity in tropical pythons. *Journal of Evolutionary Biology*, **19**, 1973-1978.

Marzal A, Bensch S, Reviriego M *et al.* (2008) Effects of malaria double infection in birds: one plus one is not two. *Journal of Evolutionary Biology*, **21**, 979-987.

McClelland EE, Penn DJ, Potts WK (2003) Major Histocompatibility Complex Heterozygote Superiority during Coinfection. *Infection and Immunity,* **71**, 2079-2086.

Milinski M (2006) The Major Histocompatibility Complex, Sexual Selection, and Mate Choice. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 159-186.

Nowak MA, Tarczy-Hornoch K, Austyn JM (1992) The optimal number of major histocompatibility complex molecules in an individual. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 10896-10899.

Ober C, Weitkamp LR, Cox N *et al.* (1997) HLA and mate choice in humans. *American Journal of Human Genetics*, **61**, 497-504.

Oliver MK, Telfer S, Piertney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1119-1128.

Olsson M, Madsen T, Nordby J *et al.* (2003) Major histocompatibility complex and mate choice in sand lizards. *Proceedings of the Royal Society B: Biological Sciences*, **270 Suppl** , S254-S256.

Paterson S, Wilson K, Pemberton J (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate

population (*Ovis aries L.*). *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 3714-3719.

Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11260-11264.

Penn DJ, Potts WK (1999) The Evolution of Mating Preferences and Major Histocompatibility Complex Genes. *American Naturalist*, **153**, 145-164.

Perrins C (1965) Population fluctuations and clutch size in the great tit, *Parus major*. *Journal of Animal Ecology*, 601–647.

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

Potts W, Manning C, Wakeland EK (1991) Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature*, **352**, 619-621.

Potts WK, Wakeland EK (1990) Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution*, **5**, 181-187.

Radwan J, Zagalska-Neubauer M, Cichoń M *et al.* (2012) MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Molecular Ecology,* **21**, 2469-2479.

Reusch TBH, Häberli MA, Aeschlimann PB, Milinski M (2001) Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature*, **414,** 300-302.

Sauermann U, Nürnberg P, Bercovitch F *et al.* (2001) Increased reproductive success of MHC class II heterozygous males among free-ranging rhesus macaques. *Human Genetics*, **108**, 249-254.

Schaschl H, Suchentruck F, Morris DL *et al.* (2012) Sex-specific selection for MHC variability in Alpine chamois. *BMC Evolutionary Biology*, **12,** 20.

Schwensow N, Eberle M, Sommer S (2008) Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 555-564.

Schwensow N, Fietz J, Dausmann KH, Sommer S (2007) Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity*, **99**, 265-277.

Sepil I, Moghadam HK, Huchard E, Sheldon BC (2012) Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evolutionary Biology*, **12**, 68.

Slade RW, McCallum HI (1992) Overdominant vs. Frequency-Dependent Selection at MHC Loci. *Genetics*, **123**, 861-862.

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 979-988.

Svensson L (1992) *Identification Guide to European Passerines*. Natural History Museum, Stockholm.

Szulkin M, Garant D, McCleery RH, Sheldon BC (2007) Inbreeding depression along a life-history continuum in the great tit. *Journal of Evolutionary Biology*, **20**, 1531-1543.

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, **124**, 967-978.

Thoß M, Ilmonen P, Musolf K, Penn DJ (2011) Major histocompatibility complex heterozygosity enhances reproductive success. *Molecular Ecology*, **20**, 1546-1557.

Thursz MR, Thomas HC, Greenwood BM *et al.* (1997) Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nature Genetics*, **17**, 11-12.

Tollenaere C, Bryja J, Galan M *et al.* (2008) Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: insights from multivariate analyses and population genetics. *Journal of Evolutionary Biology*, **21**, 1307-1320.

Trachtenberg E, Korber B, Sollars C *et al.* (2003) Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, **9**, 928-935.

Trowsdale J, Parham P (2004) Mini-review: defense strategies and immunity-related genes. *European Journal of Immunology*, **34**, 7-17.

Wegner KM, Kalbe M, Kurtz J *et al.* (2003) Parasite selection for immunogenetic optimality. *Science*, **301**, 1343.

Wegner KM, Kalbe M, Milinski M, Reusch TB (2008) Mortality selection during the 2003 European heat wave in three-spined sticklebacks: effects of parasites and MHC genotype. *BMC Evolutionary Biology*, **8**, 124.

Verhulst S, Perrins CM, Riddington R (1997) Natal dispersal of great tits in a patchy environment. *Ecology,* **78,** 864-872.

Westerdahl H, Wittzell H, von Schantz T, Bensch S (2004) MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. *Heredity*, **92,** 534-542.

Westerdahl H, Waldenström J, Hansson B *et al.* (2005) Associations between malaria and MHC genes in a migratory songbird. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 1511-1518.

Westerdahl H, Asghar M, Hasselquist D, Bensch S. (2011) Quantitative disease resistance: to better understand parasite-mediated selection on major histocompatibility complex. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 577-584.

White G, Burnham K (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, S120-S139.

Worley K, Collet J, Spurgin LG *et al.* (2010) MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, **19**, 3064-3075.

Data Accessibility

- DNA sequences: Genbank accessions JQ034624 - JQ035485

- 454 raw sequence data, individual by individual sequence data, Allele vs. supertype data, MARK input dataset, reproductive performance dataset and LRS dataset: DRYAD entry doi:10.5061/dryad.7pr69

Tables and Figures

Table 4.1 – Results of mark–recapture analysis modelling the effect of *Mhc* supertypes on annual survival probabilities (φ) of great tits.

No	Model	k	Deviance	AICc	ΔAICc	*w*
1	**Base + ST3**	**11**	**1116.181**	**1138.482**	**0.000**	**0.753**
2	Base + STNum	11	1121.913	1144.214	5.733	0.043
3	Base + ST13	11	1122.317	1144.618	6.136	0.035
4	Base φ (age*sex +t) *p* (sex)	10	1124.993	1145.244	6.762	0.026
5	Base + ST15	11	1122.945	1145.246	6.764	0.026
6	Base + STNum+STNum^2	12	1121.643	1146.000	7.518	0.018
7	Base + ST7	11	1123.919	1146.221	7.739	0.016
8	Base + ST17	11	1124.863	1147.164	8.682	0.010
9	Base + ST14	11	1124.923	1147.224	8.743	0.010
10	Base + ST10	11	1124.933	1147.234	8.752	0.009
11	Base + ST4	11	1124.953	1147.254	8.772	0.009
12	Base + ST16	11	1124.961	1147.262	8.780	0.009
13	Base + ST12	11	1124.977	1147.278	8.797	0.009
14	Base + ST6	11	1124.982	1147.283	8.801	0.009
15	Base + ST5	11	1124.991	1147.292	8.811	0.009
16	Base + ST8	11	1124.992	1147.293	8.811	0.009
17	CJS φ (t) p (t)	5	1144.857	1154.925	16.443	0.000

k - number of parameters; Δ AIC_c - increase in AIC_c compared to the top model; w - model weight; ST - supertype; STNum - supertype number; STNum^2 - square of supertype number; '*' indicates interaction between variables; '+' indicates additive effects; Goodness of fit tests were performed on the general Cormack-Jolly-Seber (CJS) model. 'age' denotes three age-classes (one year olds, two year olds, and older); 't' denotes temporal (yearly) variation. Recapture rates (p) were modelled as sex dependent in the base model.

Table 4.2 - Results of model selection based on AIC for mixed effect models with binomial errors, examining the influence of *Mhc*-linked factors on recruitment success.

Predictor	Coefficient	SE	Test Statistic	ΔAIC
supertype 3	0.038	0.179	0.214	-2.00
supertype 4	0.169	0.138	1.228	-1.00
supertype 5	-0.333	0.174	-1.908	+1.90
supertype 6	**0.338**	**0.138**	**2.449**	**+4.00**
supertype 7	0.191	0.177	1.084	0.00
supertype 8	0.194	0.153	1.273	-1.00
supertype 10	-0.342	0.228	-1.503	0.00
supertype 12	0.061	0.232	0.262	-2.00
supertype 13	0.235	0.142	1.658	+1.00
supertype 14	0.051	0.330	0.155	-2.00
supertype 15	0.054	0.199	0.270	-2.00
supertype 16	0.196	0.146	1.340	0.00
supertype 17	0.189	0.151	1.251	0.00
total number of supertypes	-0.082	0.058	-1.423	0.00
total number of supertypes^2	-0.025	0.017	-1.492	0.00
sex	0.023	0.139	0.168	-2.00
nest box density	**0.068**	**0.033**	**2.092**	**+3.00**
fledging success	**0.276**	**0.028**	**9.756**	**+126.00**

Test statistics reported are z-statistics. ΔAIC is the change in the AIC value after removal of that term from the model. Terms shown in bold are the variables retained in the minimum adequate model.

Table 4.3 - Results of model selection based on AIC for mixed effect models with Poisson errors and an overdispersion correction, examining the influence of *Mhc*-linked factors on lifetime reproductive success (LRS).

Predictor	**Coefficient**	**SE**	**Test Statistic**	**ΔAIC**
supertype 3	0.158	0.114	1.379	0.00
supertype 4	0.087	0.098	0.892	-1.20
supertype 5	**-0.243**	**0.113**	**-2.158**	**+2.90**
supertype 6	**0.294**	**0.094**	**3.114**	**+8.90**
supertype 7	-0.028	0.129	-0.219	-2.00
supertype 8	0.105	0.106	0.994	-0.90
supertype 10	-0.075	0.207	-0.363	-1.80
supertype 12	-0.046	0.142	-0.325	-1.90
supertype 13	0.120	0.099	1.208	-0.40
supertype 14	-0.134	0.167	-0.804	-1.30
supertype 15	-0.030	0.115	-0.257	-1.90
supertype 16	0.130	0.104	1.243	-0.50
supertype 17	0.096	0.103	0.926	-1.10
total number of supertypes	0.050	0.030	1.707	+1.20
total number of supertypes^2	-0.017	0.012	-1.373	+0.10
sex	-0.085	0.090	-0.943	-1.10
nest box density	**0.056**	**0.021**	**2.662**	**+5.70**

Test statistics reported are z-statistics. ΔAIC is the change in the AIC value after removal of that term from the model. Terms shown in bold are the variables retained in the minimum adequate model.

Figure 4.1 - Estimates of effect size (beta coefficients with 95% CI) of *Mhc* variables of interest (covariates that were better supported than the base model in Table 4.1) on the survival of great tits. Beta coefficients below zero indicate negative associations between *Mhc* and survival. * beta coefficient is significantly different from zero; ST – supertype.

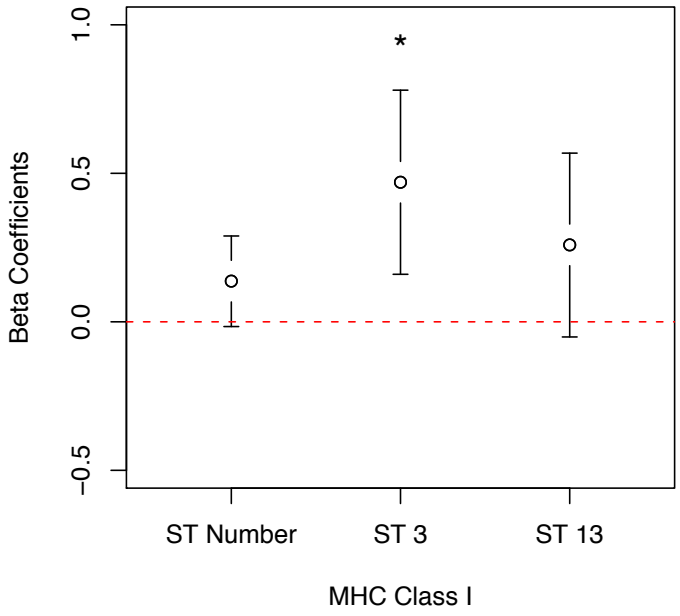


Figure 4.2 - Estimates of annual survival rates (± SE; averaged over age/sex classes) in great tits. Solid bars indicate individuals that carry *Mhc* supertype 3, while dotted bars indicate individuals lacking *Mhc* supertype 3.

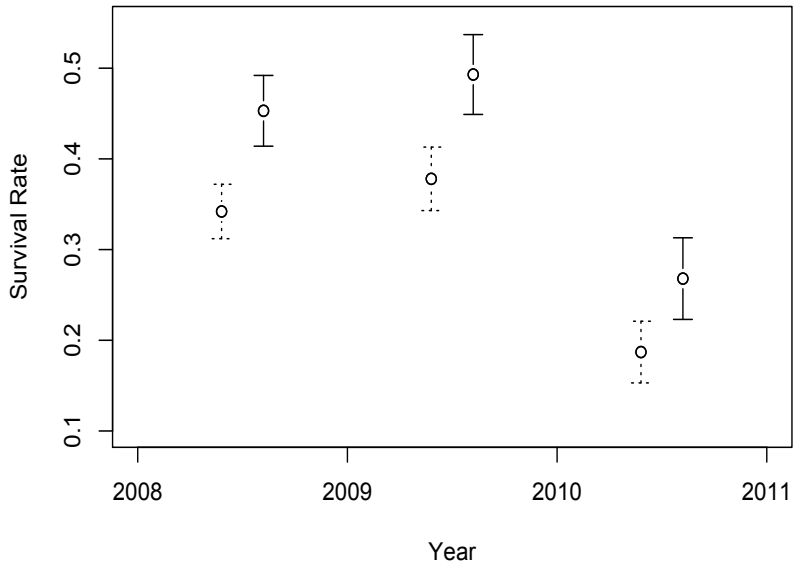


Figure 4.3 - Effects of (a) *Mhc* diversity, and (b) *Mhc* supertype 6 on annual offspring recruitment probabilities. Error bars are based on SE.

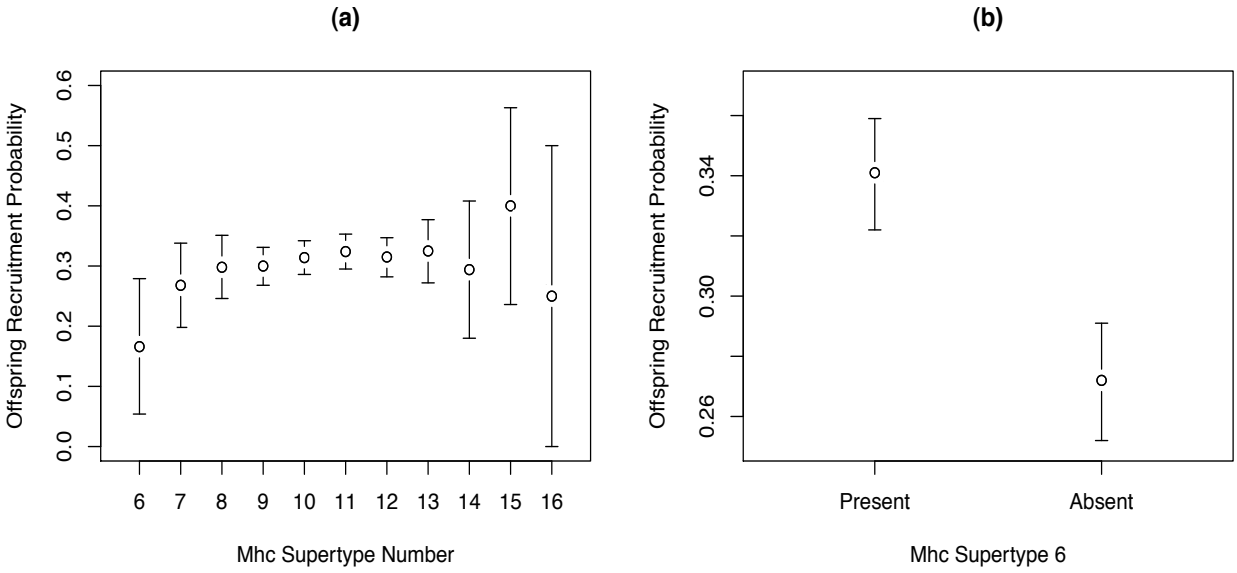
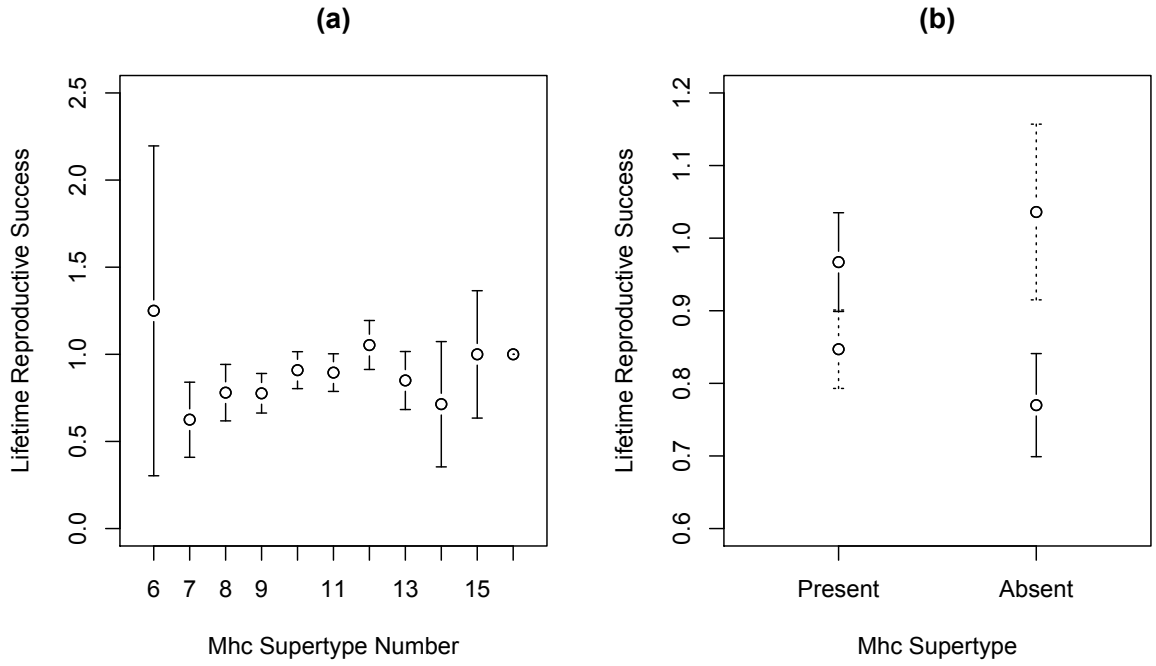


Figure 4.4 - Effect of (a) *Mhc* supertype number, and (b) *Mhc* supertype 6 (solid bars) and *Mhc* supertype 5 (dotted bars) on lifetime reproductive success (LRS). Error bars are based on SE.



Supplementary Information

Table S4.1 - Estimates of the beta coefficient for covariates of interest (variables of the models that were equally well or better supported than the base model in Table 4.1) on survival of great tits. Terms shown in bold are significantly different from zero.

Covariate	Beta estimate (SE)	CI-	CI+
ST3	**0.470**	**0.160**	**0.780**
STNum	0.137	-0.016	0.289
ST13	0.259	-0.051	0.568

ST - supertype; STNum - supertype number; CI – confidence intervals.

Table S4.2 – Results of model selection based on AIC for mixed effect models with Poisson errors, examining the influence of *Mhc*-linked factors on clutch size.

Predictor	**Coefficient**	**SE**	**Test Statistic**	**ΔAIC**
supertype 3	-0.003	0.027	-0.11	-2.00
supertype 4	0.019	0.020	0.95	-1.10
supertype 5	-0.031	0.025	-1.22	-0.50
supertype 6	0.005	0.022	0.22	-2.00
supertype 7	0.007	0.025	0.29	-1.90
supertype 8	0.006	0.023	0.28	-1.90
supertype 10	0.009	0.035	0.26	-1.90
supertype 12	0.008	0.022	0.35	-1.90
supertype 13	0.008	0.023	0.34	-1.90
supertype 14	0.014	0.037	0.39	-1.90
supertype 15	0.005	0.028	0.17	-2.00
supertype 16	0.005	0.021	0.25	-1.90
supertype 17	-0.002	0.041	-0.05	-2.00
total number of supertypes	-0.003	0.007	-0.45	-1.80
total number of supertypes^2	0.000	0.002	0.06	-2.00
sex	0.001	0.020	0.03	-2.00
nest box density	**-0.012**	**0.005**	**-2.54**	**+4.50**

Test statistics reported are z-statistics. ΔAIC is the change in the AIC value after removal of that term from the model. Terms shown in bold are the variables retained in the minimum adequate model.

Table S4.3 – Results of model selection based on AIC for mixed effect models with Poisson errors, examining the influence of *Mhc*-linked factors on hatching success.

Predictor	**Coefficient**	**SE**	**Test Statistic**	**ΔAIC**
supertype 3	0.004	0.025	0.181	-2.00
supertype 4	0.003	0.024	0.145	-2.00
supertype 5	0.002	0.034	0.052	-2.00
supertype 6	0.014	0.021	0.660	-1.60
supertype 7	0.008	0.026	0.327	-1.90
supertype 8	0.004	0.045	0.093	-2.00
supertype 10	0.042	0.033	1.276	-0.40
supertype 12	0.008	0.022	0.384	-1.80
supertype 13	0.006	0.025	0.238	-2.00
supertype 14	0.007	0.040	0.181	-1.90
supertype 15	0.000	0.023	-0.016	-2.00
supertype 16	0.009	0.023	0.397	-1.80
supertype 17	0.003	0.027	0.103	-2.00
total number of supertypes	-0.004	0.008	-0.470	-1.80
total number of supertypes^2	-0.001	0.003	-0.467	-1.80
sex	0.006	0.021	0.285	-1.90
nest box density	0.000	0.005	-0.062	-2.00
clutch size	**0.112**	**0.006**	**17.900**	**+289.2**

Test statistics reported are z-statistics. ΔAIC is the change in the AIC value after removal of that term from the model. Terms shown in bold are the variables retained in the minimum adequate model.

Table S4.4 – Results of model selection based on AIC for mixed effect models with Poisson errors and an overdispersion correction, examining the influence of *Mhc*-linked factors on fledging success.

Predictor	**Coefficient**	**SE**	**Test Statistic**	**ΔAIC**
supertype 3	-0.028	0.083	-0.334	-1.00
supertype 4	-0.006	0.092	-0.069	-2.00
supertype 5	0.018	0.101	0.181	-2.00
supertype 6	-0.007	0.138	-0.051	-2.00
supertype 7	-0.039	0.086	-0.458	-1.00
supertype 8	-0.042	0.086	-0.484	-1.00
supertype 10	0.065	0.114	0.566	0.00
supertype 12	-0.007	0.095	-0.077	-2.00
supertype 13	-0.017	0.086	-0.200	-1.00
supertype 14	-0.053	0.132	-0.402	-1.00
supertype 15	-0.037	0.085	-0.428	-1.00
supertype 16	-0.023	0.079	-0.289	-1.00
supertype 17	-0.045	0.078	-0.574	0.00
total number of supertypes	0.002	0.022	0.107	-1.00
total number of supertypes^2	0.002	0.009	0.252	-2.00
sex	0.039	0.071	0.553	0.00
nest box density	**-0.017**	**0.017**	**-1.039**	**+5.00**
hatching success	**0.153**	**0.018**	**8.451**	**+454.00**

Test statistics reported are z-statistics. ΔAIC is the change in the AIC value after removal of that term from the model. Terms shown in bold are the variables retained in the minimum adequate model.

Table S4.5 – Results of model selection based on AIC for mixed effect models with Poisson errors and an overdispersion correction, examining the influence of *Mhc*-linked factors on lifetime reproductive success (LRS).

Predictor	**Coefficient**	**SE**	**Test Statistic**	**ΔAIC**
supertype 3	-0.080	0.090	-0.882	-1.40
supertype 4	0.094	0.085	1.114	-1.10
supertype 5	**-0.241**	**0.096**	**-2.509**	**+2.90**
supertype 6	**0.218**	**0.080**	**2.710**	**+4.10**
supertype 7	0.004	0.157	0.027	-2.00
supertype 8	-0.017	0.095	-0.180	-1.90
supertype 10	-0.158	0.132	-1.197	-0.90
supertype 12	-0.135	0.087	-1.558	0.00
supertype 13	0.031	0.094	0.331	-1.90
supertype 14	-0.190	0.126	-1.515	-0.20
supertype 15	-0.126	0.094	-1.345	-0.60
supertype 16	0.084	0.111	0.760	-1.60
supertype 17	0.078	0.099	0.783	-1.50
total number of supertypes	0.026	0.026	1.027	-1.20
total number of supertypes^2	-0.019	0.010	-1.868	+1.00
sex	0.060	0.078	0.775	-1.50
nest box density	0.029	0.018	1.589	0.00
longevity	**0.553**	**0.032**	**17.305**	**+161.8**

Test statistics reported are z-statistics. ΔAIC is the change in the AIC value after removal of that term from the model. Terms shown in bold are the variables retained in the minimum adequate model.

Chapter 5

No evidence for *Mhc*-based disassortative mating or fitness benefits of *Mhc*-dependent mate choice in a wild population of great tits

No evidence for *Mhc*-based disassortative mating or fitness benefits of *Mhc*-dependent mate choice in a wild population of great tits

Irem Sepil[1], Reinder Radersma[1], Anna W Santure[2], Isabelle De Cauwer[2], Jon Slate[2], Ben C Sheldon[1]

1. Edward Grey Institute, Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, Oxford, UK

2. Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK

Abstract

Genes of the major histocompatibility complex (*Mhc*) are regarded as a potentially important target of mate choice due to the fitness benefits that may be conferred to the offspring. According to complementary genes hypothesis, females could either enhance the immunocompetence of their offspring or avoid inbreeding depression by mating with *Mhc*-dissimilar males. Here, we investigate whether selection favours a preference for dissimilar or optimally dissimilar *Mhc* supertypes (defined based on functional properties of the antigen-binding sites), and whether *Mhc* supertype dissimilarity predicts relatedness in a wild great tit population. In particular, we explore the role that *Mhc* class I plays in female mate choice decisions while controlling for relatedness and spatial population structure, and examine the reproductive fitness consequences of *Mhc* compatibility between mates. We found no evidence for the hypotheses that females select mates on the basis of either maximal or optimal *Mhc* dissimilarity. A weak correlation between *Mhc*-supertype sharing and relatedness suggests that *Mhc* dissimilarity at functional variants may not provide an effective index of relatedness.

Moreover the reproductive success of pairs did not vary with *Mhc* dissimilarity. Our results provide no support to the suggestion that selection favours, or that mate choice realises, a preference for complimentary *Mhc* types.

Keywords: major histocompatibility complex (*Mhc*); mate choice; complementary genes hypothesis; great tit (*Parus major)*; reproductive success; relatedness; spatial population structure

Introduction

Female mate choice is often invoked as a mechanism by which females improve their reproductive success and offspring quality (Mays & Hill 2004). Genes of the major histocompatibility complex (*Mhc*) are commonly linked to the evolution of mating preferences due to the potential genetic benefits that may be conferred to offspring through *Mhc*-based mate choice (Brown & Eklund 1994; Jordan & Bruford 1998; Penn & Potts 1999). *Mhc* genes are a central component of vertebrate adaptive immune system, encoding glycoproteins that deliver foreign and self-peptides to the cell surface to enable self and non-self identification by T-cells, which then initiate an immune response against pathogens (Klein 1986; Potts & Wakeland 1990).

There are two main hypotheses explaining how *Mhc*-based mate choice can act. The *good genes hypothesis* states that female mate choice should be based on male quality regardless of the female's own genotype, so that the female would guarantee resource gain or genetic benefits for the offspring, thus maximize her own reproductive success (Hamilton & Zuk 1982). In this study, however, we test the alternative suggestion of a preference for complimentary *Mhc* types. According to the *complementary genes hypothesis* females seek genes that complement their own genes in order to optimize genetic diversity, and thereby fitness of their offspring (Zeh & Zeh 1996). Clearly, *Mhc* can provide the variability required

for a genetically based recognition system, as it is the most polymorphic gene group in vertebrates and numerous studies have provided evidence for the influence of *Mhc* on individual odour profiles (Reusch *et al.* 2001; Olsson *et al.* 2003; Leinders-Zufall *et al.* 2004; Milinski *et al.* 2005; Radwan *et al.* 2008). Preference for *Mhc* dissimilar mates could act to maximize the *Mhc* diversity of the progeny so that the offspring could cope with more parasites and have enhanced immunocompetence (heterozygote advantage: Doherty and Zinkernagel 1975), or the offspring could be provided with a 'moving target' against rapidly evolving or recently introduced parasites (negative frequency dependence or fluctuating selection: Bodmer 1972; Slade & McCallum 1992; Hedrick 2002) (reviewed in Penn & Potts 1999; Milinski 2006; Bernatchez & Landry 2003; Piertney & Oliver 2006). However, maximal *Mhc* diversity may lead to excessive T-cell elimination during negative selection in the thymus (Nowak *et al.* 1992; Woelfing *et al.* 2009); and individuals possessing an intermediate number of *Mhc* alleles may be privileged under varying pathogenic pressures, having higher relative fitness (Madsen & Ujvari 2006; Kalbe *et al.* 2009). This would favour an intermediate, optimal, level of *Mhc* dissimilarity between mates (Reusch *et al.* 2001; Milinski 2006).

Mhc-dependent mating could also be an adaptation to avoid mating with a kin, *Mhc* genes acting as a marker for assessing relatedness between individuals (Potts & Wakeland 1990). Because *Mhc* molecules are crucial for distinguishing self and non-self at the molecular and cellular level, it has been suggested that the role of *Mhc* in recognition might have extended to higher biological levels, to enable kin and non-kin differentiation (Potts *et al.* 1994). According to the *inbreeding avoidance hypothesis* selection should favour a preference for maximal *Mhc* dissimilarity to avoid the deleterious consequences of kin mating (Brown 1983; Brown & Eklund 1994). Alternatively, a preference for intermediate dissimilarity could be under selection to optimize between outbreeding and inbreeding depression (Bonneaud *et al.* 2006).

Mhc-based disassortative mating preferences have been reported across taxa, both in model species (Potts *et al.* 1991; Ober *et al.* 1997; Chaix *et al.* 2008) and wild populations of

non-model species (Consuegra & de Leaniz 2008; Schwensow *et al.* 2008; Juola & Dearborn 2011- but see Huchard *et al.* 2010a). While a number of studies suggested that females target *Mhc per se*, by mating with *Mhc*-dissimilar (Landry *et al.* 2001; Miller *et al.* 2009) or optimally dissimilar males (Forsberg *et al.* 2007; Eizaguirre *et al.* 2009), others could not differentiate the role of inbreeding avoidance or *Mhc* diversity on mate choice decisions (Freeman-Gallant *et al.* 2003; Olsson *et al.* 2003; Setchell *et al.* 2010). However, few studies to date have investigated the genetic benefits associated with *Mhc*-based disassortative mating, even though documenting such patterns is crucial to understand the relevance of *Mhc* for mate choice in nature (but see Huchard *et al.* 2010a; Brouwer *et al.* 2010). Likewise, the method chosen for differentiating between the processes of preference for immunocompetence and inbreeding avoidance is also of great importance, and analyses that incorporate *Mhc* dissimilarity and relatedness in a single model are necessary to clarify whether *Mhc* per se is the target of mate choice (see Potts *et al.* 1994; Richardson *et al.* 2005; Thoß *et al.* 2011). Moreover, in order to assess female mate choice accurately, studies should consider males that are geographically close enough to the female as potential mates (given that she is likely to encounter them). Despite this, spatial population structure is only rarely taken into account in tests assessing *Mhc* assortative mating (but see Miller *et al.* 2009). Such factors, if not corrected for, may confound or conceal relevant patterns between *Mhc* and mate choice.

A final issue to consider in *Mhc*-mate choice association studies is that the majority of work to date has investigated the linkage between *Mhc* genotypes and mating preferences, although allelic differences between *Mhc* alleles are not always functionally relevant, especially when alleles share similar antigen binding motifs (Schwensow *et al.* 2008). To assess the functional diversity of *Mhc* alleles accurately, one should consider the predicted properties of their antigen-binding sites (Spurgin & Richardson 2010). An approach based on clustering alleles with similar effects into supertypes is increasingly being employed in human and nonhuman primate studies, and supertypes are considered as the unit of selection to differentiate the phenotypic effect of the underlying *Mhc* genes (Trachtenberg *et al.* 2003; Naugler & Liwski

2008; Huchard *et al.* 2010b). Therefore if females seek *Mhc*-dissimilar males in order to maximize *Mhc* genetic diversity and enhance the immunocompetence of the offspring, it is likely that a preference based on functional *Mhc* dissimilarity might be operating. However in the context of inbreeding avoidance, *Mhc*-based mate choice might depend on the alleles themselves, and not on the functionality of *Mhc* variants. Still, it is unclear whether a preference based on functional *Mhc* dissimilarity between mates could also be a reliable predictor of overall genetic similarity.

In the present work we investigate the complementary genes hypothesis in a wild population of great tits (*Parus major*). We combined *Mhc* class I supertype data with SNP genotypes of great tits to test whether females choose mates based on (i) maximally dissimilar *Mhc* supertypes, or (ii) optimally dissimilar *Mhc* supertypes, while controlling for relatedness between individuals. To account for the males that the female was likely to encounter, we controlled for spatial population structure in each model. We also tested whether *Mhc* supertype sharing was correlated with relatedness. Finally, in order to assess the relevance of *Mhc* for female mate choice, we investigate whether the initially tested mating patterns correlate with reproductive success.

Materials and Methods

Study Population

We investigated *Mhc*-based mate choice in a dataset comprising 793 great tits captured between 2006-2010 breeding seasons as a part of a long-term study conducted at Wytham Woods, near Oxford, UK (51°46'N, 1°20'W). The c. 380 ha study site is mixed deciduous woodland, in which 1020 nest boxes are scattered at variable densities. The great tit is a small, short-lived passerine bird species resident in the UK. In this part of their range great tits have a synchronous, annual breeding season (April to June), are single-brooded and both parents take care of the young until they fledge. Throughout the breeding season, all nest boxes were

checked weekly to obtain records of lay date, clutch size, hatching date, brood size and number of fledglings. Breeding birds were captured between day 6 and 14 of the nestling phase, either within the nestbox by hand or using traps, or with mist nets in front of the nest entrance. All adults and nestlings were processed and ringed with aluminium bands for individual recognition, and blood samples were collected from each adult under UK Home Office licence. Blood was collected by wing or jugular venipuncture and stored in SET Buffer at -80°C until DNA extraction. Total genomic DNA was extracted using standard ammonium acetate method and stored in AE Buffer (Qiagen).

Genetic analysis

Great tits were genotyped at the *Mhc* class I exon 3 loci using a bidirectional 454 pyrosequencing method. The methods for *Mhc* class I characterization and genotyping are described in detail in Sepil *et al.* (2012). Following genotyping, alleles with similar functional properties (based on the physicochemical amino acid properties of their positively selected sites, Doytchinova & Flower 2005) were grouped into 17 supertypes; full details regarding supertype identification are provided in Sepil *et al.* (2012). *Mhc* supertype similarity between females and males were calculated as the *Mhc*-supertype sharing value: the proportion of supertype sharing in a pair is twice the number of supertypes shared by two individuals, divided by the sum of the supertypes of each individual [$D = 2Fab / (Fa + Fb)$] (Wetton *et al.* 1987).

In addition to *Mhc* genotyping, individuals were genotyped, as part of a large genotyping study, at 9,193 single nucleotide polymorphisms (SNPs) on an Illumina iSelect BeadChip, according to the manufacturer's protocol. Following genotyping, the 9,193 SNPs were quality-checked manually in Genome Studio v2010.2 (Illumina). After excluding SNPs that that did not result in clearly distinctive genotypes, identical by state allele sharing (IBS) was calculated using PLINK (Purcell *et al.* 2007) across 8,673 SNPs for each pair of the 2,644 individuals successfully genotyped on the SNP chip (see van Bers *et al.* 2012 for further details on the development and genotyping of the SNP chip), and values for each of the individuals for

which *Mhc* genotypes were available (637 great tits) were extracted. IBS was chosen as a proxy to estimate relatedness because identical by descent (IBD) -based methods estimating relatedness are typically restricted to a few hundred unlinked loci (e.g. see Wang 2007). IBS measures were highly correlated (r = 0.89) with relatedness calculated from the social pedigree for this population (which will include cases of paternal mis-assignment due to extra-pair paternity), suggesting that IBS is an appropriate measure to estimate relatedness between breeding pairs.

Mhc and non-random mating

Randomization tests

We used randomization procedures to generate null models relating measures of genetic diversity among breeding individuals. We first performed a non-spatially explicit randomization in which *Mhc*-typed males were randomly allocated to *Mhc*-typed females within years (2006: 65 males and 34 females; 2007: 117 males and 47 females; 2008: 193 males and 93 females; 2009: 133 males and 68 females; 2010: 63 males and 10 females). All *Mhc*-typed males were included in the randomizations, whereas females were only included if they also had a *Mhc*-typed partner: hence there were more males than females in the analysis. To test whether females preferred males with maximal *Mhc* dissimilarity we calculated the average supertype sharing for observed mating pairs within the population as well as the randomized mating pairs for comparison. We predicted that average supertype sharing between mates would be significantly lower than in random pairs, if females preferred males with maximal *Mhc* dissimilarity. Then, to test whether females preferred males with optimal *Mhc* dissimilarity we calculated the variance in supertype sharing for observed mating pairs within the population and compared it with that for randomized mating pairs. Here we reasoned that, if females preferred males with optimal *Mhc* dissimilarity, then the variance in supertype sharing would be significantly lower in observed mating pairs than in randomized mating pairs. The same analysis was repeated for relatedness to determine whether females were more or less likely to

mate with kin. We predicted that average relatedness between mates would be significantly lower than in random pairs, if females preferred males with maximal *Mhc* dissimilarity to avoid the deleterious consequences of kin mating. Secondly, we predicted that if females preferred males with intermediate *Mhc* dissimilarity to optimize between outbreeding and inbreeding depression, then the variance in relatedness would be significantly lower in observed mating pairs than in randomized mating pairs. Lastly, to control for background relatedness while investigating non-random *Mhc* supertype dissimilarity between mates, we calculated standardized average supertype sharing and its variance by computing the difference between the z-scores for relatedness and supertype sharing. This was done to eliminate the influence of relatedness in *Mhc*-based mate choice decisions. We predicted that, if females preferred males with maximal *Mhc* dissimilarity only to enhance the immunocompetence of their offspring, then the standardized average supertype sharing between mates would be significantly lower than in random pairs. However, if females preferred males with optimal *Mhc* dissimilarity only to achieve maximal pathogen resistance for their offspring, then the standardized variance in supertype sharing would be significantly lower in observed mating pairs than in randomized mating pairs. In each test we performed 10 000 randomizations.

Accounting for spatial structure

Because both natal (Greenwood *et al.* 1979) and breeding dispersal distances (Harvey *et al.* 1979) of great tits are considerably smaller than the study area, we expected variation in the probability that two individuals to encounter and form a breeding pair, and hence the possibility that this was related to their degree of genetic similarity. We therefore preformed spatially explicit randomizations as a means to generate null models that incorporated such spatial effects. We used GIS-derived measures of tit nest box coordinates as the geographical locations of the breeding great tits (Wilkin *et al.* 2007). A grid of squares of a given size (see below) was projected over a map of the research area. For each iteration, this grid was randomly moved and rotated. Male identities were randomized within each square while the locations of individuals

were kept constant. Over a large number of iterations the probability that two individuals were selected as a pair decreased with distance and approached a bivariate normal distribution for latitude and longitude with the locations of the females as the mean.

We performed 10 000 randomizations at 12 different spatial scales; hence the area that the female could choose a mate from varied in each analysis. The spatial scales (length of the edges of the squares) that we considered were 1m, 50m, 100m, 200m, 400m, 600m, 800m, 1000m, 1500m, 2000m, 5000m and 10000m; and the average number of mates that was randomized per female at these distances were 1, 1.08, 1.42, 2.72, 6.46, 10.79, 15.37, 20.00, 31.26, 43.30, 91.22 and 114.12.

Mhc and reproductive success

We examined associations between *Mhc*-supertype sharing values and reproductive success in 252 pairs of great tits (each involving different pairs of individuals) that bred between the years 2006-2010. We used generalized linear mixed effects models to investigate the effect of pair *Mhc* dissimilarity on great tit reproductive success, measured at four consecutive stages of the breeding attempt. The specific reproductive measures we considered were: (i) clutch size, modelled as a Poisson response with a log link; (ii) brood size, modelled as a Poisson response; (iii) the number of young fledged (fledging success), a Poisson response, adjusted for overdispersion; and (iv) recruitment success, a binary variable indicating whether any of the young recruited to the breeding population, modeled as a binomial response with a logit link. The reproductive measures were assessed in relation to the *Mhc*-supertype sharing value of the pair and local breeding density (measured as nestbox densities at the nine Wytham woodland sections); the latter is known to significantly influence several measures of reproductive performance in the great tit (Bouwhuis *et al.* 2009, 2010). Year was included as a random effect in each model to control for temporal environmental heterogeneity. In addition, we fitted clutch size as a covariate when modeling brood size, brood size as a covariate when modeling fledging success, and fledging success as a covariate when modeling recruitment success; hence each

analysis addressed the additive effect of *Mhc* dissimilarity at that reproductive stage, controlling for any influence on preceding stages. We also repeated the same analyses without fitting the previous stage as a covariate to consider the total effect of parental *Mhc* dissimilarity at the reproductive stage. Lastly, we fitted relatedness as a covariate and repeated each analysis. Inclusion of relatedness reduced the dataset to 203 pairs (each involving different pairs of individuals), because in 49 pairs either the male or female of the pair had not been included in the SNP genotyping panel. Models were performed using the R package 'lme4'.

We used Akaike's information criterion (AIC) to determine the combination of variables that best explained the data with minimal parameters. Model selection was performed by backward stepwise elimination and the fit of each new model was assessed by comparison of AIC values. Terms were eliminated from the model when their removal resulted in an improved fit (i.e. a <2 reduction in AIC, Burnham & Anderson 2002), and were retained if their removal resulted in an increase in AIC >2. Where the removal of a term resulted in a model with an approximately equal fit (i.e. a change in AIC of <2) the model with fewer terms was considered the most parsimonious model (Burnham & Anderson 2002). To confirm the validity of the minimum model, removed variables were added individually to assess any potential improvement in the model fit.

Results

Mhc and non-random mating

Mhc supertype sharing values varied between 0.444-1.000, and SNP based IBS relatedness values varied between 0.745-0.800 among mated pairs (Figure 5.1). Results of the randomization tests yielded little support for an association between female mate choice and pair *Mhc* dissimilarity. There was no difference in mean *Mhc* supertype sharing (Figure 5.2a) or variance in *Mhc* supertype sharing (Figure 5.2b) between mated and randomly assigned pairs. However there was general tendency for randomly assigned mates to be more similar at the *Mhc*

as the spatial scale increased, which hints at a weak tendency to choose *Mhc* dissimilar mates among the observed matings (Figure 5.2a). Likewise, there was no difference in mean relatedness (Figure 5.2c) or variance in relatedness (Figure 5.2d) between mated and randomly assigned pairs. However, in contrast to mean *Mhc* supertype sharing, though also weakly in this dataset, randomly assigned mates tended to be less similar in terms of SNP-based IBS relatedness as the spatial scale increased, suggesting a weak influence of genetic population structure on the similarity of observed mating partners (Figure 5.2c). A similar pattern was observed when *Mhc* supertype sharing was assessed while controlling for relatedness (Figure 5.2e). Overall, the tendency for lower supertype sharing among observed pairs than those among random pairs when population structure is accounted for suggests a very weak tendency to avoid *Mhc* similar mates, but this effect is not statistically significant. Similarly, the variance in standardized *Mhc* supertype sharing (Figure 5.2f) was not significantly different than would be expected under random mating. Hence, there is, from these analyses, little evidence to suggest that females choose mates on the basis of maximal or optimal *Mhc* dissimilarity. Finally, we found a weak correlation between *Mhc* supertype sharing and SNP-based estimates of relatedness among mated pairs (r = 0.136, p = 0.053), implying that *Mhc* supertype dissimilarity is a poor predictor of kinship (Figure 5.3).

Mhc and reproductive success

Results of the mixed effect models revealed no evidence that the reproductive performance of pairs varied according to *Mhc* dissimilarity. There was no effect of *Mhc* supertype sharing on clutch size, hatching success, fledging success or recruitment success (Table 5.1). Reproductive performance did not vary as a function of nestbox density in these models, but did vary significantly with the effect of the former stage (Table 5.1). Adding relatedness as a covariate to the models did not change the results of any analysis. Likewise, removing the former reproductive stages from the models did not qualitatively change the results of any analysis (results not shown).

Discussion

Mhc genes are regarded as an important target of mate choice due to the genetic benefits that can be obtained by the offspring; however it is still unclear whether selection favors a preference for genetic compatibility to enhance immunocompetence or avoid inbreeding, especially in wild populations. Here, we investigated the role that *Mhc* class I functional diversity (*Mhc* supertypes) plays in female mate choice decisions while controlling for relatedness and spatial population structure, and also examined the reproductive fitness consequences of *Mhc*-dependent mating patterns in a wild great tit population. Our results offer little support for the hypotheses that females select males based on maximal or optimal *Mhc* supertype dissimilarity. At best, there was a weak tendency to choose males with a dissimilar *Mhc* supertype. Similarly, we did not detect any reproductive advantage for mating with *Mhc* dissimilar males. Moreover, *Mhc* supertype similarity was not a strong predictor of relatedness, implying that functional diversity at *Mhc* would be unlikely to act as a cue for kinship. Overall we found no support for the hypothesis of *Mhc* compatibility. However it is worth noting that a preference for genetic compatibility was only calculated in terms *Mhc* supertypes; therefore we cannot discount the possibility that *Mhc* allele dependent mate choice might have been overlooked.

The results of the randomization tests varied depending on the spatial scale chosen for encountering males. Because the breeding and foraging distances of animals are often expected to be smaller than study areas, it is important to control for spatial population structure, in order to provide a more direct assessment of female mate choice and to control for any bias that may arise as a result of spatial genetic structuring (Miller *et al.* 2009). For instance, our analyses revealed a trend towards preferring related individuals when all mated males were compared with randomized pairs, however the tendency disappeared when we restricted our comparisons to the males that were within few hundred meters (spatial scales 1 to 7). In the great tit, this result is expected, because limited natal dispersal causes related individuals to occur closer

together than nonrelated individuals while breeding (e.g. Szulkin & Sheldon 2008; Szulkin *et al.* 2009). Therefore our results highlight the importance of considering fine-scale genetic structure on small geographic distances, a phenomenon that has not been given much consideration in other studies investigating mating patterns (though see Szulkin *et al.* 2009).

Although a preference for maximally or optimally *Mhc*-dissimilar males have been documented in other avian species (Freeman-Gallant *et al.* 2003; Bonneaud *et al.* 2006; Juola & Dearborn 2011); we found no evidence to suggest that mate choice is based on genetic compatibility in this population of great tits, despite a slight tendency to mate with more *Mhc* dissimilar males. Moreover, there was no indication of a reproductive advantage linked with *Mhc* compatibility between mates, implying that *Mhc* diversity of the progeny may have little effect on their fitness. In agreement with this finding, recent analyses have shown no fitness advantage for individuals possessing maximal *Mhc* diversity and no evidence for an optimal diversity in terms of adult survival, individual reproductive performance and lifetime reproductive success (Sepil *et al.* 2013). Hence the absence of an association between *Mhc* compatibility, female mate choice and reproductive success in the present study is perhaps not unexpected. Likewise, selection for *Mhc*-based disassortative mating might be weak in the Wytham Woods great tit due to the outbred nature of the population. Although the fitness consequences of inbreeding depression are substantial, close inbreeding (i.e. mating with close kin where $f \geq 0.125$) occurs rarely, in fewer than 1.5% of all breeding events analysed over 41 years (Szulkin *et al.* 2007), and no evidence was found to suggest birds avoid mating with related partners (Szulkin *et al.* 2009); hence our findings might reflect the lack of inbreeding avoidance. Finally, our analysis revealed a weak correlation between *Mhc* supertype sharing and SNP-based IBS relatedness; therefore it is plausible to suggest that functional diversity at *Mhc* is unlikely to act as a cue for kinship. The mechanisms by which *Mhc* acts as a cue for kinship are not fully understood, but are more likely to involve detection of differences in the *Mhc* molecules produced by expressed genes (Howard 1977). In that case, comparison of the functional properties of these molecules (as done here) is more relevant than a comparison of

the sequence differences among alleles. However it is also possible that the *Mhc* alleles themselves might act as a cue for odour detection, influencing kin and non-kin differentiation (Singh *et al.* 1987). In such case one would expect *Mhc*-dependent mate choice based on alleles rather than supertypes to avoid mating with a kin. Therefore the lack of evidence for *Mhc*-based disassortative mating or fitness benefits of *Mhc*-dependent mate choice might be due to the consideration of supertypes. Investigating the role of *Mhc* alleles in great tit mating preferences is the subject of ongoing work.

A further possible explanation for the absence of a preference for *Mhc* dissimilar males is that females might be unable to discriminate male *Mhc* profile on the basis of odour. Experimental work on mammal and fish species have revealed compelling evidence that females use olfactory cues to evaluate the *Mhc* genotype of their prospective mating partners (Reusch *et al.* 2001; Milinski *et al.* 2005; Radwan *et al.* 2008; Setchell *et al.* 2010), however little is known about whether birds could use *Mhc*-related cues to choose compatible mates. Although birds have long been thought to be anosmic, a growing number of studies on different avian orders indicate that birds are able to use olfactory information in a variety of contexts (Balthazart & Taziaux 2009). Therefore, experimental studies investigating the link between *Mhc*-based mate choice and individual odour profiles are necessary to clarify whether birds can assess the *Mh*c content of potential partners using olfactory cues.

Mhc-based mating preferences may also target specific supertypes or male *Mhc* diversity, regardless of females own *Mhc*, as predicted by good genes hypothesis. Elsewhere we have shown that the possession of specific supertypes was strongly associated with adult survival and lifetime reproductive success (Sepil *et al.* 2013). However, in this study, we were unable to test preferences for certain supertypes because all the males in our dataset were part of a breeding pair and we did not screen non-mated males in the population. Sampling males that failed to form a breeding pair and examining the probability of pair formation as a function of male *Mhc* diversity and possession of specific supertypes would be ideal for testing good genes hypothesis and is an area that awaits further exploration.

Finally it should also be emphasized that although the great tit is a socially monogamous species, extra-pair copulations are frequent in the population; 13% of nestlings are reported to be extra-pair young (Patrick *et al.* 2011). Extra-pair copulations occur regularly in passerine bird species that form social mating pairs (Birkhead & Møller 1992; Griffith *et al.* 2002) and are considered as a strategy for females restricted in their choice of social mate, to enhance the genetic quality of offspring (reviewed in Jennions & Petrie 2000). Therefore, the absence of *Mhc*-based social mate choice in the great tit might be due to the occurrence of extra-pair copulations, if females engage in extra-pair mating to correct for any genetic disadvantage gained from paring up with a suboptimal male. Evidence for *Mhc*-based extra-pair copulations has been provided in several wild bird studies (Freeman-Gallant *et al.* 2003; Richardson *et al.* 2005; Promerová *et al.* 2011). Particularly, Richardson *et al.* (2005) showed that females were more likely to obtain extra-pair paternity when their social mate had low *Mhc* diversity, but failed to detect any association between *Mhc* and choice of social mate. Moreover, in a follow up study Brouwer *et al.* (2010) revealed a positive association between *Mhc* diversity and juvenile survival, and argued that this pattern was indicative of hidden genetic benefits gained through extra-pair copulations. In this study we couldn't examine *Mhc*-dependent extra pair mating preferences and their fitness benefits; hence the possibility of hidden genetic benefits for offspring was overlooked. Investigating the potential for *Mhc*-based extra-pair copulations in this population of great tits will be the subject of future work.

To our knowledge, this study is among the first to test *Mhc*-based non-random mating patterns while controlling for relatedness and population genetic structure; and to investigate the genetic benefits associated with *Mhc*-based disassortative mating in a wild population. Our results indicate that great tits do not prefer males that are functionally dissimilar at *Mhc* class I as their social mate and there is no fitness benefit associated with mate choice. Therefore we found no evidence supporting the suggestion that selection favours preference for complimentary *Mhc* types.

Acknowledgements

This paper is dedicated to the memory of Daisy Balogh who was a PhD student at the Institute of Zoology, London investigating the role of *Mhc* in mate choice decisions and its fitness consequences in the Wytham Woods great tit population. We are grateful to the Wytham fieldworkers who collected the data for this study. We thank Alicia Davies for laboratory assistance, Ada Grabowska-Zhang, Sandra Bouwhuis and Kaan Oz for helpful discussion and advice. This work was partially funded by NERC grants (NER/A/S/2002/00877 and NE/F005725/1) to BCS.

References

Balthazart, J. & Taziaux, M. (2009). The underestimated role of olfaction in avian reproduction? *Behavioural Brain Research*, 200, 248-259.

Bernatchez, L. & Landry, C. (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, 16, 363-377.

van Bers, N.E.M., Santure, A.W., van Oers, K., de Cauwer, I., Dibbits, B.W., Mateman, C., *et al.* (2012). The design and cross-population application of a genome-wide SNP chip for the great tit *Parus major*. *Molecular Ecology Resources*, 12, 753-770.

Birkhead, T.R. & Møller, A.P. (1992) *Sperm competition in birds: Evolutionary causes and consequences*. Academic Press, London

Bodmer, W. (1972). Evolutionary significance of the HL-A system. *Nature*, 237, 139-183.

Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H. & Sorci, G. (2006). Complex Mhc-based mate choice in a wild passerine. *Proceedings of the Royal Society B: Biological Sciences*, 273, 1111-1116.

Bouwhuis, S., Charmantier, A., Verhulst, S. & Sheldon, B.C. (2010). Individual variation in rates of senescence: natal origin effects and disposable soma in a wild bird population. *Journal of Animal Ecology*, 79, 1251-1261.

Bouwhuis, S., Sheldon, B.C., Verhulst, S. & Charmantier, A. (2009). Great tits growing old: selective disappearance and the partitioning of senescence to stages within the breeding cycle. *Proceedings of the Royal Society B: Biological Sciences*, 276, 2769-2777.

Brouwer, L., Barr, I., van de Pol, M., Burke, T., Komdeur, J. & Richardson, D.S. (2010). MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, 19, 3444-3455.

Brown, J. & Eklund, A. (1994). Kin Recognition and the Major Histocompatibility Complex: An Integrative Review. *American Naturalist*, 143, 435-461.

Brown, J.L. (1983). *Some paradoxical goals of cells and organisms: the role of the MHC In: Ethical questions in brain and behavior*. New York: Springer-Verlag; 111-124.

Burnham, K. & Anderson, D. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag, New York.

Chaix, R., Cao, C. & Donnelly, P. (2008). Is mate choice in humans MHC-dependent? *PLoS Genetics*, 4, e1000184.

Consuegra, S. & Garcia de Leaniz, C. (2008). MHC-mediated mate choice increases parasite resistance in salmon. *Proceedings of the Royal Society B: Biological Sciences*, 275, 1397-1403.

Doherty, P. & Zinkernagel, R. (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, 256, 50-52.

Doytchinova, I. & Flower, D.R. (2005). In silico identification of supertypes for class II MHCs. *Journal of Immunology*, 174, 7085-7095.

Eizaguirre, C., Yeates, S.E., Lenz, T.L., Kalbe, M. & Milinski, M. (2009). MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology*, 18, 3316-3329.

Forsberg, L.A. Dannewitz, J., Petersson, E. & Grahn, M. (2007). Influence of genetic dissimilarity in the reproductive success and mate choice of brown trout - females fishing for optimal MHC dissimilarity. *Journal of Evolutionary Biology*, 20, 1859-1869.

Freeman-Gallant, C.R., Meguerdichian, M., Wheelwright, N.T. & Sollecito, S.V. (2003). Social pairing and female mating fidelity predicted by restriction fragment length polymorphism similarity at the major histocompatibility complex in a songbird. *Molecular Ecology*, 12, 3077-3083.

Greenwood, P.J., Harvey, P.H. & Perrins, C.M. (1979). The role of dispersal in the Great Tit (*Parus major*): The causes, consequences and heritability of natal dispersal. *Journal of Animal Ecology*, 48, 123-142.

Griffith, S.C., Owens, I.P.F. & Thuman, K.A. (2002) Extra pair paternity in birds: a review of interspecific variation and adaptive function. *Molecular Ecology*, 11, 2195-2212.

Hamilton, W.D. & Zuk, M. (1982) Heritable true fitness and bright birds: a role for parasites? *Science*, 218, 384-387.

Harvey, P., Greenwood, P. & Perrins, C. (1979). Breeding area fidelity of great tits (*Parus major*). *Journal of Animal Ecology*, 48, 305-313.

Hedrick, P.W. (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution,* 56, 1902–1908.

Howard, J.C. (1977) H-2 and mating preferences. *Nature*, 266, 406–408.

Huchard, E., Knapp, L.A., Wang, J., Raymond, M. & Cowlishaw, G. (2010a). MHC, mate choice and heterozygote advantage in a wild social primate. *Molecular Ecology*, 19, 2545-2561.

Huchard, E., Raymond, M., Benavides, J., Marshall, H., Knapp, L.A. & Cowlishaw, G. (2010b). A female signal reflects MHC genotype in a social primate. *BMC Evolutionary Biology*, 10, 96.

Jennions, M.D. & Petrie, M. (2000) Why do females mate multiply? A review of the genetic benefits. *Biological Reviews*, 75, 21–64.

Jordan, W.C. & Bruford, M.W. (1998) New perspectives on mate choice and the MHC. *Heredity*, 81, 127–133.

Juola, F.A. & Dearborn, D.C. (2011). Sequence-based evidence for major histocompatibility complex-disassortative mating in a colonial seabird. *Proceedings of the Royal Society B: Biological Sciences,* 279. 153-162.

Kalbe, M., Eizaguirre, C., Dankert, I., Reusch, T.B.H., Sommerfeld, R.D., Wegner, K.M., *et al.* (2009). Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings of the Royal Society B: Biological Sciences*, 276, 925-934.

Klein, J. (1986). *Natural History of the Major Histocompatibility Complex*. Wiley, New York.

Landry, C., Garant, D., Duchesne, P. & Bernatchez, L. (2001). "Good genes as heterozygosity": the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proceedings of the Royal Society B: Biological Sciences*, 268, 1279-1285.

Leinders-Zufall, T., Brennan, P., Widmayer, Chandramani, P.S., Maul-Pavicic, A., Jäger, M., *et al.* (2004). MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science*, 306, 1033-1037.

Madsen, T. & Ujvari, B. (2006). MHC class I variation associates with parasite resistance and longevity in tropical pythons. *Journal of Evolutionary Biology*, 19, 1973-1978.

Mays, H.L. & Hill, G.E. (2004). Choosing mates: good genes versus genes that are a good fit. *Trends in Ecology & Evolution*, 19, 554-559.

Milinski, M. (2006). The Major Histocompatibility Complex, Sexual Selection, and Mate Choice. *Annual Review of Ecology, Evolution, and Systematics*, 37, 159-186.

Milinski, M., Griffiths, S., Wegner, K.M., Reusch, T.B.H., Haas-Assenbaum, A. & Boehm, T. (2005). Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 4414-4418.

Miller, H.C., Moore, J.A., Nelson, N.J. & Daugherty, C.H. (2009). Influence of major histocompatibility complex genotype on mating success in a free-ranging reptile population. *Proceedings of the Royal Society B: Biological Sciences,* 276, 1695-704.

Naugler, C. & Liwski, R. (2008). An evolutionary approach to major histocompatibility diversity based on allele supertypes. *Medical Hypotheses*, 70, 933-937.

Nowak, M.A., Tarczy-Hornoch, K. & Austyn, J.M. (1992). The optimal number of major histocompatibility complex molecules in an individual. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 10896-10899.

Ober, C., Weitkamp, L.R., Cox, N., Dytch, H., Kostyu, D. & Elias, S. (1997). HLA and mate choice in humans. *American Journal of Human Genetics*, 61, 497-504.

Olsson, M., Madsen, T., Nordby, J., Wapstra, E., Ujvari, B. & Wittsell, H. (2003). Major histocompatibility complex and mate choice in sand lizards. *Proceedings of the Royal Society B: Biological Sciences*, 270 Suppl, S254-S256.

Patrick, S.C., Chapman, J.R., Dugdale, H.L., Quinn, J.L. & Sheldon, B.C. (2011). Promiscuity, paternity and personality in the great tit. *Proceedings of the Royal Society B: Biological Sciences,* 279, 1724-1730.

Penn, D.J. & Potts, W.K. (1999). The Evolution of Mating Preferences and Major Histocompatibility Complex Genes, 153, 145-164.

Piertney, S.B. & Oliver, M.K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96, 7-21.

Potts, W., Manning, C. & Wakeland, E. (1991). Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature*, 352, 619-621.

Potts, W.K. & Wakeland, E.K. (1990). Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution*, 5, 181-187.

Potts, W.K., Manning, C.J., Wakeland, E.K. (1994) The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 346, 369–78.

Promerová, M., Vinkler, M., Bryja, J., Poláková, R., Schnitzer, J., Munclinger, P., *et al.* (2011). Occurrence of extra-pair paternity is connected to social male's MHC-variability in the scarlet rosefinch *Carpodacus erythrinus*. *Journal of Avian Biology*, 42, 5-10.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559-575.

Radwan, J., Tkacz, A. & Kloch, A. (2008). MHC and Preferences for Male Odour in the Bank Vole. *Ethology*, 114, 827-833.

Reusch, T.B., Häberli, M.A., Aeschlimann, P.B. & Milinski, M. (2001). Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature*, 414, 300-302.

Richardson, D.S., Komdeur, J., Burke, T. & von Schantz, T. (2005). MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. *Proceedings of the Royal Society B: Biological Sciences,* 272, 759-767.

Schwensow, N., Eberle, M. & Sommer, S. (2008). Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proceedings of the Royal Society B: Biological Sciences,* 275, 555-564.

Sepil, I., Moghadam, H.K., Huchard, E. & Sheldon, B.C. (2012). Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evolutionary Biology*, 12, 68.

Sepil, I., Lachish, S. & Sheldon, B.C. (2013). *Mhc*-linked survival and lifetime reproductive success in a wild population of great tits. *Molecular Ecology*, 22, 384-396.

Setchell, J.M., Charpentier, M.J.E., Abbott, K.M., Wickings, E.J. & Knapp, L.A. (2010). Opposites attract: MHC-associated mate choice in a polygynous primate. *Journal of Evolutionary Biology*, 23, 136-148.

Singh, P.B., Brown, R.E. & Roser, B. (1987) MHC antigens in urine as olfactory recognition cues. *Nature*, 327, 161.

Slade, R.W. & McCallum, H.I. (1992). Overdominantvs. Frequency-Dependent Selectionat MHC Loci. *Genetics*, 123, 861-862.

Spurgin, L.G. & Richardson, D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277, 979-988.

Szulkin, M., Garant, D., McCleery, R.H. & Sheldon, B.C. (2007). Inbreeding depression along a life-history continuum in the great tit. *Journal of Evolutionary Biology*, 20, 1531-1543.

Szulkin, M. & Sheldon, B.C. (2008). Dispersal as a means of inbreeding avoidance in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences*, 275, 703-711.

Szulkin, M., Zelazowski, P., Nicholson, G. & Sheldon, B.C. (2009). Inbreeding avoidance under different null models of random mating in the great tit. *Journal of Animal Ecology*, 78, 778-788.

Thoß, M., Ilmonen, P., Musolf, K. & Penn, D.J. (2011). Major histocompatibility complex heterozygosity enhances reproductive success. *Molecular Ecology*, 20, 1546-1557.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E., *et al.* (2003). Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, 9, 928-935.

Wang, J. (2007). Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetical Research*, 89, 135-153.

Wetton, J., Carter, R., Parkin, D. & Walters, D. (1987). Demographic study of a wild house sparrow population by DNA fingerprinting. *Nature*, 147-149.

Wilkin, T.A., Perrins, C.M. & Sheldon, B.C. (2007). The use of GIS in estimating spatial variation in habitat quality : a case study of lay-date in the Great Tit *Parus major*. *Ibis*, 149, 110-118.

Woelfing, B., Traulsen, A., Milinski, M. & Boehm, T. (2009) Does intra-individual major histocompatibility complex diversity keep a golden mean? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 117–128.

Zeh, J.A. & Zeh, D.W. (1996) The evolution of polyandry I: intragenomic conflict and genetic incompatability. *Proceedings of the Royal Society B: Biological Sciences,* 263, 1711–1717.

Tables and Figures

Table 5.1 – Results of model selection based on AIC for mixed effect models examining the influence of *Mhc* supertype sharing on four stages of reproductive performance.

Response variable	Predictors	Coefficient	SE	Test Statistic	ΔAIC
Clutch size	Supertype sharing	-0.012	0.230	-0.050	-1.99
	Density	-0.005	0.009	-0.590	-1.66
Hatching success	Supertype sharing	-0.089	0.239	-0.371	-1.86
	Density	0.003	0.010	0.288	-1.92
	Clutch size	**0.114**	**0.013**	**8.822**	**+71.21**
Fledging success	Supertype sharing	0.204	0.720	0.283	-1.50
	Density	-0.015	0.030	-0.491	-0.30
	Hatching success	**0.142**	**0.033**	**4.295**	**+118.7**
Recruitment success	Supertype sharing	-1.357	1.543	-0.879	-1.20
	Density	0.102	0.066	1.532	+0.30
	Fledging success	**0.261**	**0.058**	**4.473**	**+24.3**

Test statistics reported are z-statistics. ΔAIC is the change in the AIC value after removal of that term from the model. Terms shown in bold are the variables retained in the minimum adequate model.

Figure 5.1 – Frequency distributions of (a) *Mhc*-supertype sharing, (b) SNP based identical by state relatedness among observed mated pairs

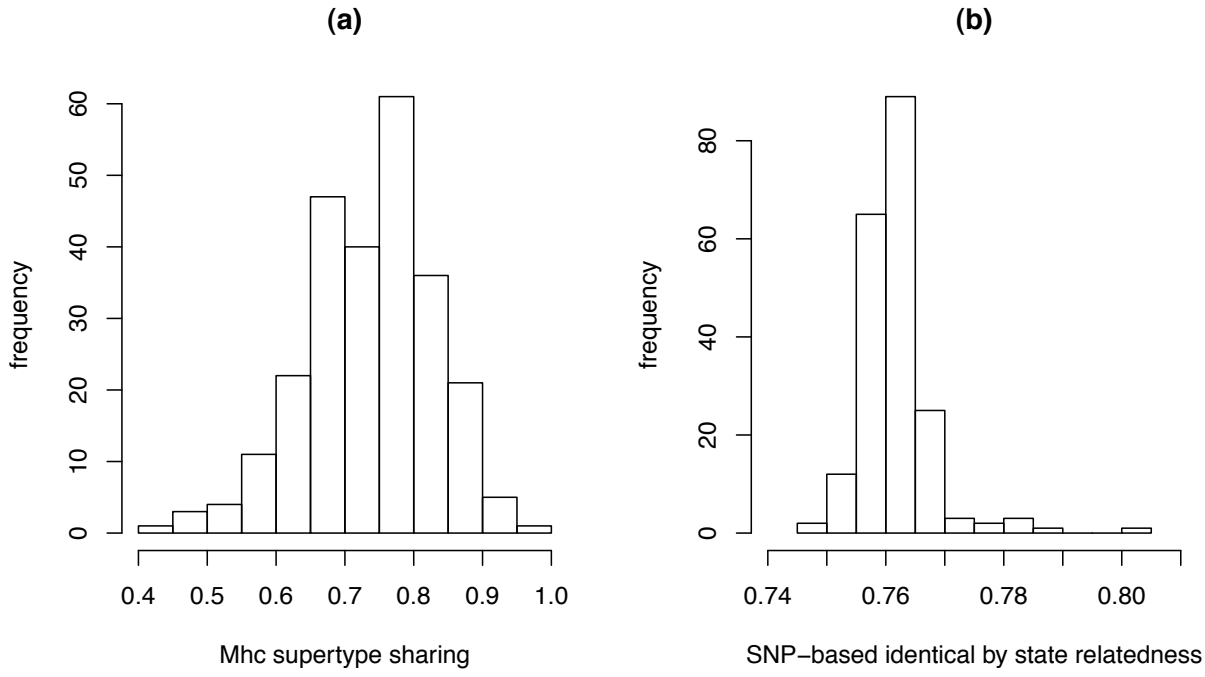


Figure 5.2 – Distribution of genetic parameters representing similarity between randomly selected pairs at a range of spatial scales, and at the entire population scale (W), compared with the observed value (scale 1) for mating adult great tits. (a) mean supertype sharing, (b) variance in supertype sharing, (c) mean relatedness based on SNP genotypes, (d) variance in relatedness, (e) standardized supertype sharing, (f) variance in standardized supertype sharing. Horizontal dashed lines indicate the 95% confidence interval.

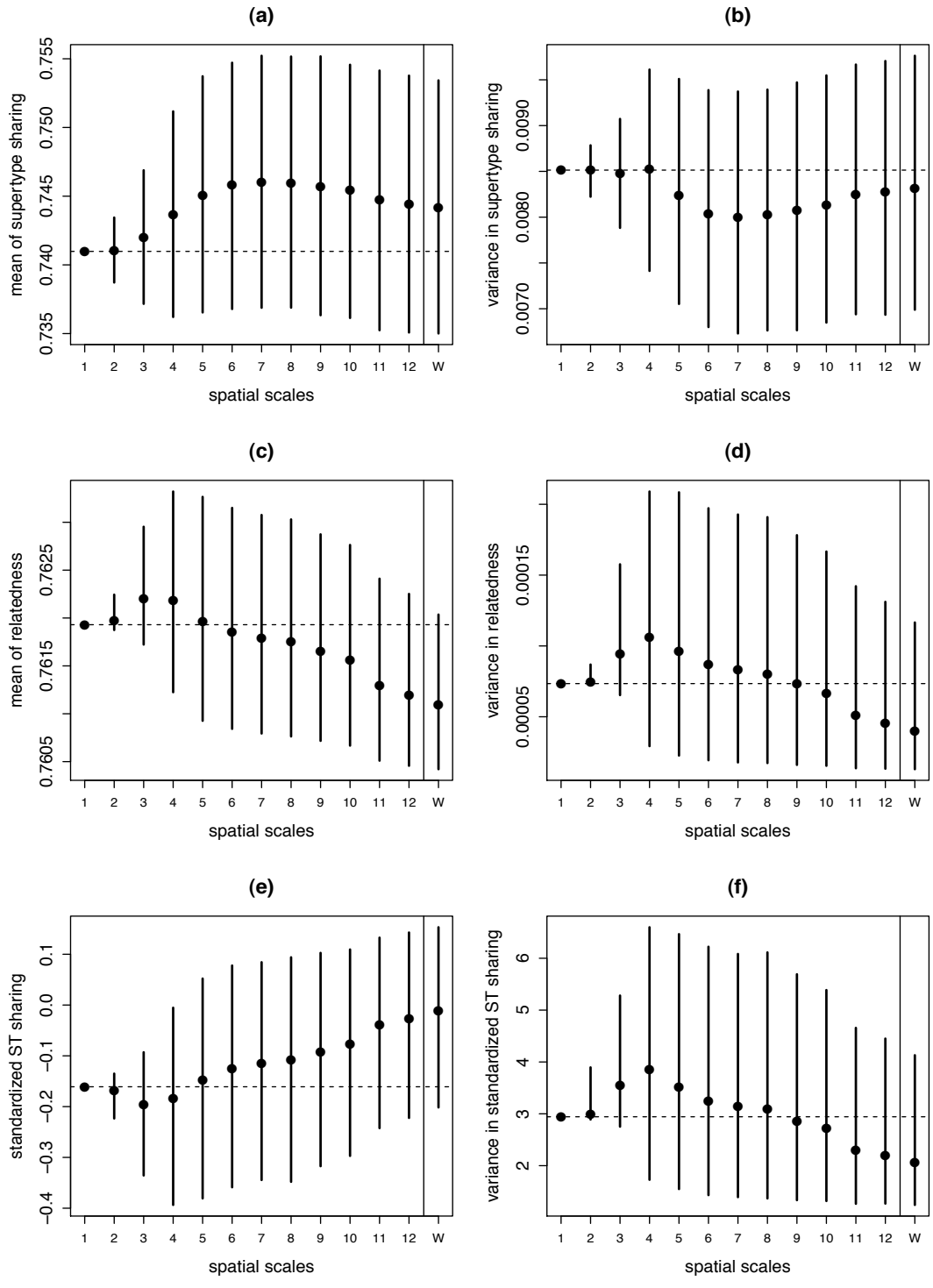
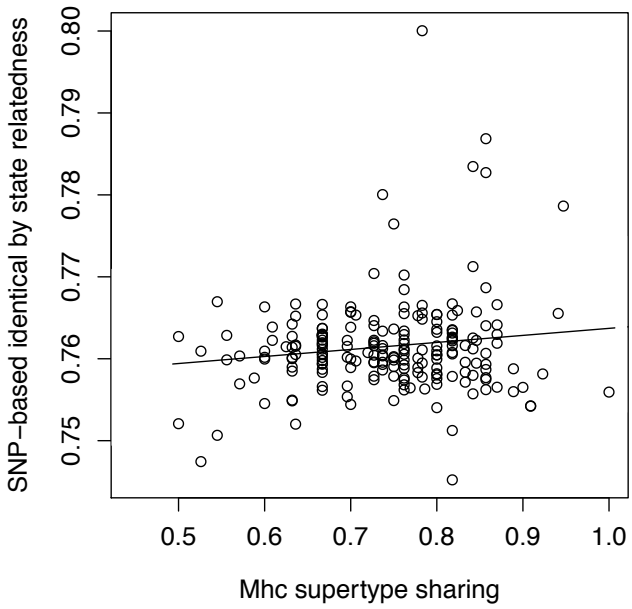


Figure 5.3 - Correlation between *Mhc* supertype sharing and SNP based identical by state relatedness for 203 mated pairs



Chapter 6

Discussion

Discussion

In this thesis I investigated whether and how selection operates on *Mhc* class I functional variation in the great tit, using a combination of detailed genetic analysis and long-term breeding data. In this final chapter, I will summarize the key findings from this thesis and discuss the questions they raise. I will also address limitations of the work presented here, and suggest avenues for future research to improve our understanding of *Mhc* evolution.

Summary of principal findings

In the first data chapter of this thesis (Chapter 2), I characterized and genotyped *Mhc* class I exon 3 of the great tit and estimated the effective *Mhc* repertoire of each individual by grouping alleles with similar antigen binding affinities into supertypes. The comprehensive characterization effort I undertook, and the application of a next generation sequencing method, allowed for a detailed investigation of *Mhc* class I variation in the great tit both in terms of allelic diversity and gene number. Furthermore, by assessing the presence of stop codons in sequences, examining the phylogenetic relationship between allelic clusters, and through historical selection tests, I was able to differentiate functional and putatively non-functional alleles. Reliable *Mhc* genotyping of more than 800 great tits revealed the presence of at least sixteen functional loci, together with a pseudogene family and putatively non-functional loci; there was clear evidence that functional alleles were under strong balancing selection. A total of 862 alleles were detected from 857 individuals; the highest number yet characterized in a wild bird species. Of these 755 alleles were classified as functional and were grouped into 17 supertypes. This degree of complexity is much larger than the classical avian model for *Mhc*, the chicken, however it is in line with studies detecting a large number of *Mhc* class I and class

II loci in passerines; demonstrating the evolutionary lability of genetic organisation of this complex.

Avian malaria parasites have emerged as a popular model system for studies examining *Mhc*-based disease resistance in wild populations. In Chapter 3, I investigated associations among patterns of avian malaria infection and *Mhc* class I supertypes, to understand the role that the *Mhc* plays in determining host resistance and susceptibility to *Plasmodium* infections. The analyses revealed there to be significant associations between two different *Mhc* supertypes and the probability of infection with two different *Plasmodium* species, *P. relictum* and *P. circumflexum*, yet in contrasting fashions. The direction of the association for supertype 17 was indicative of *Mhc*-linked qualitative resistance to *P. relictum* infection, with individuals lacking the supertype twice as likely to be infected. However, the functional role of the second *Mhc* supertype, supertype 6, in regard to *P. circumflexum* infection, was more difficult to assess. Here, by conducting three further analyses, I was able to show that acute stage *P. circumflexum* infection has a mortality cost and exerts strong selection on its great tit host; thus the positive association between supertype 6 and probability of infection was indicative of *Mhc*-linked quantitative resistance to *P. circumflexum* infection. This study was the first to provide empirical evidence showing that two different *Mhc* supertypes can confer qualitative and quantitative resistance to avian malaria in a single host population. Moreover, I demonstrated that the alleles linked with *Plasmodium* infections have similar antigen binding affinities in great tits and house sparrows (*Passer domesticus*), suggesting that this is a suitable model system for parasite-*Mhc* associations in the wild.

In Chapter 4, I combined mark-recapture methods with analysis of long-term breeding data to investigate the effects of *Mhc* supertypes on adult survival, reproductive performance and lifetime reproductive success, and found that the presence of three different supertypes was associated with three different components of individual fitness. While individuals carrying *Mhc* supertype 3 had higher adult survival rates, individuals possessing supertype 5 had lower lifetime reproductive success. Moreover, the analyses revealed that *Mhc* supertype 6, the

supertype linked with *Mhc*-based quantitative resistance to *P. circumflexum* infection in Chapter 3, conferred a selective advantage to its carriers, as great tits possessing this *Mhc* supertype experienced higher lifetime reproductive success and were more likely to recruit an offspring at the brood level. This study thus provided strong support to the suggestion that selection favours specific functional variants of *Mhc,* possibly as a result of supertype-specific resistance or susceptibility to parasites that exert strong selective pressures on their hosts. In contrast, I found no evidence for a selective advantage of *Mhc* functional diversity, either in terms of maximal or optimal supertype diversity.

In Chapter 5, I explored the role that *Mhc* class I functional diversity plays in female mate choice decisions while controlling for relatedness and spatial population structure, examined the reproductive fitness consequences of *Mhc*-dependent mating patterns and assessed whether *Mhc* supertype dissimilarity predicts relatedness. I found no evidence implying that females prefer more *Mhc* dissimilar or optimally dissimilar males to optimize the functional *Mhc* diversity of their offspring or to avoid mating with kin. Consequently, there was no reproductive advantage for mating with *Mhc* dissimilar males and only a weak correlation between *Mhc*-supertype sharing and relatedness.

Perspectives and future directions

Characterization and genotyping of Mhc loci in non-model vertebrates

The thorough background work and the application of 454 pyrosequencing technology revealed extreme complexity at the *Mhc* class I of the great tit; I found evidence for the presence of at least 16 functional loci. A high number of functional class I loci (between four to six) have also been described in other passerine species (Westerdahl *et al.* 1999; Bonneaud *et al.* 2004; Promerová *et al.* 2009; Schut *et al.* 2011), however this constitutes the highest number of expressed *Mhc* class I alleles and loci identified in passerines. Moreover it is likely that this complexity is not specific to great tits and that similar patterns can be observed in other

systems, if next generation sequencing (NGS) technologies were to be applied for genotyping. For instance, 454 technology was recently used for *Mhc* genotyping collared flycatcher (*Ficedula albicollis*) and it was shown that the species possess more than ten expressed *Mhc* class II loci (Zagalska-Neubauer *et al.* 2010; Radwan *et al.* 2012).

Although species that harbour many *Mhc* loci and high allelic diversity are intriguing as they are potentially under strong selection and studies of such species can provide valuable information as to the mechanisms maintaining *Mhc* diversity (Babik *et al.* 2009), the extreme complexity at the *Mhc* region also raises the question as to how confident we can be about whether we characterized and genotyped all the functional variants in the region. Despite undertaking comprehensive background work for characterization, placing special emphasis on primer design and applying a strict variant validation procedure to differentiate artefacts (Babik 2010), I still cannot exclude the possibility that some alleles might have been missed at the PCR stage, or that some rare alleles might have been removed while accounting for artefacts. Thus, it is reasonable to conclude that accurate characterization and genotyping of alleles from multiple, co-amplifying loci is still a major challenge in complex *Mhc* systems even when NGS methods are being employed, and future studies should prioritize improving experimental design and optimizing quality control measures. For instance, Zagalska-Neubauer *et al.* (2010) used 454 technology to type the *Mhc* class II loci of collared flycatcher, observed low repeatability among samples genotyped twice and concluded that the coverage of reads per individual was not sufficient for reliable genotyping. Consequently, in a follow up study the authors developed new primers, used the previous calculations to set up better genotyping criteria, and as a result found no discrepancies among replicates, reaching 100% repeatability (Radwan *et al.* 2012). Alternatively future studies might consider focusing on simplified *Mhc* regions, particularly species possessing one or two loci; however treating the results from species with simple *Mhc* structures as if they are applicable to complex systems is questionable (Spurgin & Richardson 2010).

Among the 1536 samples included in the experiment, only 871 were successfully genotyped by the end of the bioinformatics procedure. The rest of the samples were considered unreliable for having less than 200 reads; hence were removed from the dataset. The efficiency of the genotyping method could have been improved by optimising the sample number for a full plate run (Galan *et al.* 2010). However, determining the optimal number of samples to use in a 454 run might prove difficult, especially in cases where *Mhc* loci number is initially unknown. Alternatively, novel methods can be developed to make fuller use of sequence data. Here, I only applied a five-step variant validation procedure to differentiate real *Mhc* alleles from PCR/sequencing artefacts. Future studies can perhaps include Bayesian methods that account for uncertainty or integrate pedigree information, so that errors generated during PCR and pyrosequencing can be detected and discarded more effectively.

In the present study, I used a 'supertyping' approach to determine the functional properties of *Mhc* class I alleles. To date, the majority of work has assessed links between *Mhc* genotypes and measures of fitness, however much of selection on *Mhc* is likely to act via the phenotypic effects of the underlying genes. Because it is the properties of the antigen-binding sites that determine the interactions between *Mhc* molecules and parasites, I adopted the bioinformatic and statistical methods described by Doytchinova and Flower (2005) and Jombart *et al.* (2010) to define the physicochemical properties of the positively selected sites of each allele and to cluster *Mhc* alleles with similar peptide specificities into supertypes. Therefore by analyzing the relationship between *Mhc* supertypes and measures of fitness, I aimed to predict the functional effects of *Mhc* molecules. A growing body of evidence supports the biological relevance of the supertyping approach (Trachtenberg *et al.* 2003; Naugler & Liwski 2008; Schwensow *et al.* 2008). The choice of using *Mhc* supertypes can also be justified by statistical arguments, since detection of allele specific effects necessitates both high sample sizes and simple *Mhc* systems (Huchard *et al.* 2010). Here, I utilized a newly proposed method for *Mhc* supertyping (Jombart *et al.* 2010), which is an advance on previous methods, as it does not require arbitrary clustering decisions. However, it is also worth noting that the classification of

Mhc alleles into supertypes could potentially be improved, since the partial *Mhc* sequences did not cover the section ranging from amino acid positions 5-10, a highly polymorphic region that affects the antigen binding capabilities of alleles. Amplifying the entire exon is likely to be critical for future studies. Moreover, our supertyping approach discards the possibility that non-positively selected sites might also affect the antigen binding affinities of *Mhc* molecules or that other molecules (for example TAP) interact with the *Mhc* to determine binding properties (Walker *et al.* 2011). Likewise sites that are of importance to *Mhc* functioning might have been missed due to gene conversion, compromising the accuracy of our method. Hence, the methodology can be further developed to improve the precision at identifying distinct clusters. Alternatively, experimental approaches can be used instead of bioinformatic approaches to determine peptide specificities of *Mhc* molecules. For instance, a recent study established computational antigen-binding prediction algorithms based on empirical datasets, and showed an evolutionary advantage for allele pairs that are more divergent in recognizing a broader range of potential antigens (Lenz 2011).

Mhc and disease resistance studies

Studies investigating the genetic basis of variation in disease susceptibility should ideally be able to estimate or preferentially measure host-parasite encounter probabilities (Amos *et al.* 2011), and predict the fitness effects of infection on the host (Westerdahl *et al.* 2011), in order to avoid erroneous conclusions from correlational analyses in the wild. Although individual infection status and parasitaemia (parasite load) are two key parasitological variables used in evolutionary studies of host-parasite systems, especially for avian malaria research (e.g. Bensch *et al.* 2007; Ortego *et al.* 2007; Sehgal *et al.* 2011; Asghar *et al.* 2012), both measures can be influenced by a number of environmental factors and failure to discriminate such effects can add noise to any analyses exploring disease resistance and susceptibility. Previous work on the Wytham Woods tit population has revealed the central role of environmental factors in determining the probability of avian malaria infection (Wood *et al.* 2007; Knowles *et al.* 2011;

Lachish *et al.* 2012), hence I added local infection risk in the analyses to tease apart whether prevalence patterns reflected the lack of host-parasite encounter or host resistance to infection. However, future studies could adopt a more direct measure for estimating an individual's infection history. For instance, serological assays based on antibody detection can be employed to test whether an individual has ever been exposed to the parasite; this would allow an accurate assessment of qualitative resistance, since such data would permit comparison of disease status among exposed individuals only. Serological assays have mainly been used in experimental infection studies of avian malaria (Atkinson *et al.* 2001; Jarvi *et al.* 2002), yet Woodworth *et al.* (2005) investigated whether *Plasmodium* antibodies were present in wild-caught Hawaii amakihi (*Hemignathus virens*) and showed that a high proportion of individuals tested positive for antibodies, even though infection was not microscopically detectable. This work demonstrates that serological assays can be used for detecting *Plasmodium* antibodies in studies exploring *Mhc*-linked avian malaria resistance.

Understanding relative infection risk in the local scale allowed me to formulate and test hypotheses regarding the fitness effects of *P. circumflexum* parasites, due to the sedentary nature of great tits following postnatal dispersal. Therefore I was able to argue that the positive association between the presence of *Mhc* supertype 6 and probability of infection with *P. circumflexum* was indicative of quantitative resistance, since the analyses implied a mortality cost for *P. circumflexum* parasites during the acute stage of infection, in line with the findings of a recent study carried out on the closely related sympatric host species, blue tits (*Cyanistes caeruleus*) (Lachish *et al.* 2011). However, despite the fine-scale analysis I undertook, identifying the fitness consequences of this parasite strain was very difficult and I still cannott rule out the possibility that supertype 6 might confer disease susceptibility. Unfortunately, determining the fitness effects of the acute stage of infection remains a major challenge in wild populations (Valkiunas 2005); hence *Mhc*-disease prevalence patters are still difficult to explain. Ideally, future studies should determine the fitness consequences of parasites to their hosts at any stage of infection, in order to provide a clearer picture of how *Mhc* is linked with

probability of infection. Experimental infection of hosts in field enclosure systems that resemble natural conditions could be highly informative in this regard, thus seems like a promising approach that future studies could adapt.

Effect of Mhc on measures of fitness

A key aim of this thesis was to investigate whether functional diversity at *Mhc* had fitness consequences in terms of host survival and lifetime reproductive success. Although previous studies have demonstrated *Mhc*-fitness correlations in semi-natural settings (e.g. Sauermann *et al.* 2001; Wegner *et al.* 2008; Eizaguirre *et al.* 2009; Kalbe *et al.* 2009; Worley *et al.* 2010; Thoß *et al.* 2011), rather little is known about *Mhc* related fitness effects in the wild (but see Paterson *et al.* 1998; Brouwer *et al.* 2010). Determining *Mhc*-fitness correlations in natural populations has proven challenging and has been hampered by the limited number of suitable study systems where individuals are marked, breeding attempts are recorded, and the fate of the offspring is known (Clutton-Brock & Sheldon 2010). Linking functional variation at *Mhc* to measures of fitness in inherently complex natural populations also requires a level of experimental design and analytical rigour that is difficult to achieve due to the extent of confounding and interacting variables (Oliver & Piertney 2010). In the present work I tried to address this issue by controlling for environment-driven variation in fitness measures, in order to detect the influence of *Mhc* on reproductive performance and lifetime reproductive success in great tits. Likewise, the recapture and survival rates were parameterized to account for sex, age and time effects. Previous work on this population revealed that great tit survival rates differ between males and females, and with age (Bouwhuis *et al.* 2012); and that spatiotemporal heterogeneity has strong effects on a number of life-history traits (Bouwhuis *et al.* 2009, 2010). For instance clutch size and number of fledglings are negatively associated with breeding density in Wytham and both measures fluctuate substantially depending on the climatic conditions of the year; a simple correlational analysis failing to incorporate such variables could easily lead to erroneous conclusions. The detection of strong spatiotemporal effects acting at

small spatial scales highlights the importance of study designs and analytical frameworks that take account of such effects in order to accurately link measures of fitness and underlying genetic variation. Alternatively, studies in which individuals are sampled in homogenous environments and are compared within birth cohorts would be useful to limit the effects of such confounding factors; however treating the results from such simple designs as if they are applicable to complex systems might be unrealistic.

The work presented here suffered from a few limitations that could be improved upon in future studies. First, the mark-recapture modelling framework did not include juveniles so only the effects of *Mhc* on adult survival was analysed, although tits experience the highest mortality during the juvenile life stage, in the first few months after leaving the nest (Perrins 1965). *Mhc* genotyping fledglings to assess their survival rates would facilitate this investigation. Second, *Mhc*-fitness associations were examined among great tits captured in 2008 and 2009; a proportion of the great tits included in the research are still alive in 2012. To ensure that the results were not biased due to these few samples, I repeated recruitment analyses after removing individuals that were present in 2011 breeding season; the previously documented correlations remained unchanged. Ideally, future studies could investigate *Mhc*-fitness correlations among individuals that were monitored for their entire lifespan to prevent any bias in the results (e.g. Brouwer *et al.* 2010). Finally, although the study revealed correlations between specific *Mhc* supertypes and measures of fitness, possibly as a result of supertype-specific disease resistance and susceptibility, these results were generally obtained without any knowledge regarding the parasites involved (except the association between *Mhc* supertype 6 and infection with *P. circumflexum*). Therefore, an exhaustive pathogen screening effort is required to identify the exact parasite involved and to further our understanding of how selection acts on *Mhc*.

Mhc and female mate choice decisions

Results from the final part of this thesis indicated that functional compatibility at *Mhc* had no effect on female mate choice decisions. Likewise, the reproductive success of the pair was not influenced by the *Mhc* similarity of the social pair. A choice for *Mhc* dissimilar males is often invoked as a mechanism to increase the *Mhc* diversity of the offspring or to avoid inbreeding depression (Potts & Wakeland 1990; Brown & Eklund 1994). Overall, the findings of this thesis provided no evidence to support the hypothesis that *Mhc* diversity confers any advantage to the individual, since we failed to detect any link with avian malaria infection, host survival, reproductive performance and lifetime reproductive success. Moreover, previous work on this population of great tits have shown that although the fitness consequences of breeding with close kin are substantial, the occurrence of such events is rare, and there is no evidence to suggest that birds avoid mating with related partners, other than through dispersal (Szulkin *et al.* 2007, 2009). Therefore the absence of an association between *Mhc* compatibility, female mate choice and reproductive success in the present study is perhaps not unexpected.

In this thesis I demonstrated that the presence of particular *Mhc* supertypes have substantial effects on the survival, annual recruitment probabilities and lifetime reproductive success of its carriers. Hence, it is likely that selection might favour a preference for males carrying specific *Mhc* supertypes so that females would guarantee resource gain or genetic benefits for the offspring, thus maximize their own reproductive success, as predicted by the *good genes hypothesis*. Unfortunately, I was not able to assess whether females targeted mates possessing specific *Mhc* alleles, since the dataset was comprised of mated individuals only. Studies that sample the unmated males in the population are essential to test this possibility. However sampling unmated males in the Wytham great tit population might prove difficult, because each breeding season we fail to capture a small proportion of mated males; hence it would be challenging if not impossible to differentiate between unmated males and males that have a pair but escaped sampling. Moreover, the assessment of *Mhc*-based mate choice extended only as far as social mate preferences, even though extra pair matings occur quite

frequently in this population: 13% of nestlings are reported to be extra-pair young (Patrick *et al.* 2012). Ideally, future mate choice studies would investigate whether females engage in extra-pair matings to correct for any genetic disadvantage gained from pairing up with a suboptimal male.

Overall conclusion

This thesis provided a rare detailed analysis of the role that functional variation at *Mhc* class I plays in disease resistance, host survival, reproductive success and mate choice within a well-characterized wild bird population, improving our understanding of how selection operates on *Mhc* in nature. Initially I demonstrated that great tit have a multilocus, highly variable *Mhc* class I region that has historically been under strong balancing selection. Then I revealed that certain functional variants of *Mhc* (supertypes) confer resistance to two divergent *Plasmodium* parasite species that are common in the environment, thus suggested selection should favour the persistence of such supertypes. Likewise, I showed that the presence of particular *Mhc* supertypes have substantial effects on the survival, annual recruitment probabilities and lifetime reproductive success of its carriers; however found no evidence to suggest that functional dissimilarity at *Mhc* have any influence on female mate choice decisions, and that dissimilarity at *Mhc* affects the reproductive output of the social pair. Altogether the results of this thesis provide strong evidence to the hypothesis that selection targets specific functional variants of *Mhc*, whereas there was no support for selection favouring maximal or optimal *Mhc* diversity. More importantly this thesis demonstrates that functional variants of *Mhc* class I loci are an important determinant of individual fitness in natural populations.

References

Amos, W., Driscoll, E. & Hoffman, J.I. (2011). Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proceedings of the Royal Society B: Biological Sciences*, 278, 1183-1188.

Asghar, M., Westerdahl, H., Zehtindjiev, P., Ilieva, M., Hasselquist, D. & Bensch, S. (2012). Primary peak and chronic malaria infection levels are correlated in experimentally infected great reed warblers. *Parasitology*, 1-7.

Atkinson, C.T., Dusek, R.J. & Lease, J.K. (2001). Serological responses and immunity to superinfection with avian malaria in experimentally-infected Hawaii amakihi. *Journal of Wildlife Diseases*, 37, 20-27.

Babik, W. (2010). Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources*, 10, 237-251.

Babik, W., Taberlet, P., Ejsmond, M.J. & Radwan, J. (2009). New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources*, 9, 713-719.

Bensch, S., Waldenström, J., Jonzén, N., Westerdahl, H., Hansson, B., Sejberg, D., *et al.* (2007). Temporal dynamics and diversity of avian malaria parasites in a single host species. *Journal of Animal Ecology*, 76, 112-122.

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

Bouwhuis, S., Charmantier, A., Verhulst, S. & Sheldon, B.C. (2010). Individual variation in rates of senescence: natal origin effects and disposable soma in a wild bird population. *Journal of Animal Ecology*, 79, 1251-1261.

Bouwhuis, S., Choquet, R., Sheldon, B.C. & Verhulst, S. (2012). The forms and fitness cost of senescence: age-specific recapture, survival, reproduction, and reproductive value in a wild bird population. *American Naturalist*, 179, E15-E27.

Bouwhuis, S., Sheldon, B.C., Verhulst, S. & Charmantier, A. (2009). Great tits growing old: selective disappearance and the partitioning of senescence to stages within the breeding cycle. *Proceedings of the Royal Society B: Biological Sciences*, 276, 2769-2777.

Brouwer, L., Barr, I., van de Pol, M., Burke, T., Komdeur, J. & Richardson, D.S. (2010). MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, 19, 3444-3455.

Brown, J. & Eklund, A. (1994). Kin Recognition and the Major Histocompatibility Complex: An Integrative Review. *American Naturalist*, 143, 435-461.

Clutton-Brock, T. & Sheldon, B.C. (2010). Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, 25, 562-573.

Doytchinova, I.A. & Flower, D.R. (2005). In silico identification of supertypes for class II MHCs. *Journal of Immunology*, 174, 7085-7095.

Eizaguirre, C., Yeates, S.E., Lenz, T.L., Kalbe, M. & Milinski, M. (2009). MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology*, 18, 3316-3329.

Galan, M., Guivier, E., Caraux, G., Charbonnel, N. & Cosson, J.-F. (2010). A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*, 11, 296.

Huchard, E., Raymond, M., Benavides, J., Marshall, H., Knapp, L.A. & Cowlishaw, G. (2010). A female signal reflects MHC genotype in a social primate. *BMC Evolutionary Biology*, 10, 96.

Jarvi, S.I., Schultz, J.J. & Atkinson, C.T. (2002). PCR diagnostics underestimate the prevalence of avian malaria (*Plasmodium relictum*) in experimentally-infected passerines. *Journal of Parasitology*, 88, 153-158.

Jombart, T., Devillard, S. & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94.

Kalbe, M., Eizaguirre, C., Dankert, I., Reusch, T.B.H., Sommerfeld, R.D., Wegner, K.M., *et al.* (2009). Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings of the Royal Society B: Biological Sciences*, 276, 925-934.

Knowles, S.C.L., Wood, M.J., Alves, R., Wilkin, T. A, Bensch, S. & Sheldon, B.C. (2011). Molecular epidemiology of malaria prevalence and parasitaemia in a wild bird population. *Molecular Ecology*, 20, 1062-1076.

Lachish, S., Knowles, S.C.L., Alves, R., Sepil, I., Davies, A., Lee, S., Wood, M.J. & Sheldon, B.C. (2012) Spatial determinants of infection risk in a multi-species avian malaria system. *Ecography,* 35, 1–12.

Lachish, S., Knowles, S.C.L., Alves, R., Wood, M.J. & Sheldon, B.C. (2011). Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. *Journal of Animal Ecology*, 80, 1196-1206.

Lenz, T.L. (2011) Computational prediction of Mhc II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution*, 65, 2380-2390.

Naugler, C. & Liwski, R. (2008). An evolutionary approach to major histocompatibility diversity based on allele supertypes. *Medical Hypotheses,* 70, 933-937.

Oliver, M. & Piertney, S. (2010). Beyond splitting hares and rabbiting on about major histocompatibility complex complexity. *Molecular Ecology*, 19, 4099-4101.

Ortego, J., Cordero, P.J., Aparicio, J.M. & Calabuig, G. (2007). No relationship between individual genetic diversity and prevalence of avian malaria in a migratory kestrel. *Molecular Ecology*, 16, 4858-4866.

Paterson, S., Wilson, K. & Pemberton, J. (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proceedings of the National Academy of Sciences of the United States of America*, 95, 3714-3719.

Patrick, S.C., Chapman, J.R., Dugdale, H.L., Quinn, J.L. & Sheldon, B.C. (2012). Promiscuity, paternity and personality in the great tit. *Proceedings of the Royal Society B: Biological Sciences*, 279, 1724-1730.

Perrins, C. (1965). Population fluctuations and clutch size in the great tit, *Parus major*. *Journal of Animal Ecology*, 34, 601–647.

Potts, W.K. & Wakeland, E.K. (1990). Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution*, 5, 181-187.

Promerová, M., Albrecht, T. & Bryja, J. (2009). Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). *Immunogenetics*, 61, 451-461.

Radwan, J., Zagalska-Neubauer, M., Cichoń, M., Sendecka, J., Kulma, K., Gustafsson, L., *et al.* (2012). MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Molecular Ecology*, 21, 2469-2479.

Sauermann, U., Nürnberg, P., Bercovitch, F., Berard, J., Trefilov, A., Widdig, A., *et al.* (2001). Increased reproductive success of MHC class II heterozygous males among free-ranging rhesus macaques. *Human Genetics*, 108, 249-254.

Schut, E., Aguilar, J.R.-d., Merino, S., Magrath, M.J.L., Komdeur, J. & Westerdahl, H. (2011). Characterization of MHC-I in the blue tit (*Cyanistes caeruleus*) reveals low levels of genetic diversity and trans-population evolution across European populations. *Immunogenetics*, 63, 531-542.

Schwensow, N., Eberle, M. & Sommer, S. (2008). Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proceedings of the Royal Society B: Biological Sciences*, 275, 555-564.

Sehgal, R.N.M., Buermann, W., Harrigan, R.J., Bonneaud, C., Loiseau, C., Chasar, A., *et al.* (2011). Spatially explicit predictions of blood parasites in a widely distributed African

rainforest bird. *Proceedings of the Royal Society B: Biological Sciences*, 278, 1025-1033.

Spurgin, L.G. & Richardson, D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277, 979-988.

Szulkin, M., Garant, D., McCleery, R.H. & Sheldon, B.C. (2007). Inbreeding depression along a life-history continuum in the great tit. *Journal of Evolutionary Biology*, 20, 1531-1543.

Szulkin, M., Zelazowski, P., Nicholson, G. & Sheldon, B.C. (2009). Inbreeding avoidance under different null models of random mating in the great tit. *Journal of Animal Ecology*, 78, 778-788.

Thoß, M., Ilmonen, P., Musolf, K. & Penn, D.J. (2011). Major histocompatibility complex heterozygosity enhances reproductive success. *Molecular Ecology*, 20, 1546-1557.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E., *et al.* (2003). Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, 9, 928-935.

Valkiunas, G. (2005). *Avian malaria parasites and other haemosporidia*. CRC Press, Boca Raton.

Walker, B.A., Hunt, L.G., Sowa, A.K., Skjodt, K., Gobel, 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 of the United States of America*, 108, 8396-8401.

Wegner, K.M., Kalbe, M., Milinski, M. & Reusch, T.B. (2008). Mortality selection during the 2003 European heat wave in three-spined sticklebacks: effects of parasites and MHC genotype. *BMC Evolutionary Biology*, 8, 124.

Westerdahl, H., Asghar, M., Hasselquist, D. & Bensch, S. (2011). Quantitative disease resistance: to better understand parasite-mediated selection on major histocompatibility complex. *Proceedings of the Royal Society B: Biological Sciences*, 279, 577-584.

Westerdahl, H., Wittzell, H. & von Schantz, T. (1999). Polymorphism and transcription of Mhc class I genes in a passerine bird, the great reed warbler. *Immunogenetics*, 49, 158-170.

Wood, M.J., Cosgrove, C.L., Wilkin, T.A., Knowles, S.C.L., Day, K.P. & Sheldon, B.C. (2007). Within-population variation in prevalence and lineage distribution of avian malaria in blue tits, *Cyanistes caeruleus*. *Molecular Ecology*, 16, 3263-3273.

Woodworth, B.L., Atkinson, C.T., Lapointe, D.A., Hart, P.J., Spiegel, C.S., Tweed, E.J., *et al.* (2005). Host population persistence in the face of introduced vector-borne diseases: Hawaii amakihi and avian malaria. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 1531-1536.

Worley, K., Collet, J., Spurgin, L.G., Cornwallis, C., Pizzari, T. & Richardson, D.S. (2010). MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, 19, 3064-3075.

Zagalska-Neubauer, M., Babik, W., Stuglik, M., Gustafsson, L., Cichoń, M. & Radwan, J. (2010). 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC Evolutionary Biology*, 10, 395.