Major histocompatibility complex and mate choice in two passerine birds

Silje Larsen Rekdal

UiO Natural History Museum

Dissertation for the degree of Philosophiae Doctor

Natural History Museum
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

2020

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2255

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Photo: Bjørn Aksel Bjerke

In medio tutissimus ibis
(Ovid, *Metamorphoses*)

Contents

Summary

All organisms are continuously encountering pathogens, which they have to fight to avoid infectious diseases and fitness costs. In vertebrates, genes of the major histocompatibility complex (MHC) are crucial for pathogen resistance, by triggering immune responses against specific intracellular (MHC class I; MHCI) and extracellular (MHC class II; MHCII) pathogens, and are among the most polymorphic genes known. An intermediate optimum in individual MHC diversity is expected based on a trade-off between the number of pathogens recognized by the immune system and the depletion of T cells through negative selection in the thymus if the number of MHC alleles is excessively large. Consequently, there is strong selection to optimize MHC constitution in offspring, and whenever possible, there should be selection for choosing a mate with compatible MHC diversity, which provides offspring with an intermediate MHC repertoire. MHC-based mate choice has been studied in many taxa, with various results supporting preferences for mates with specific alleles, for intermediate and maximally MHC-diverse mates, and for mates with an intermediate and maximized MHC dissimilarity. Hence, there is no consensus across species, and many questions concerning MHC-based mate choice and its role in shaping MHC diversity remain unanswered.

This thesis explores MHC-based mate choice in two passerine species, the bluethroat (*Luscinia svecica*) and the willow warbler (*Phylloscopus trochilus*), in an extra-pair mating framework. Studying MHC-based extra-pair mate choice in passerines is useful for two main reasons. First, as the extra-pair male only contributes sperm, as opposed to the social male, it enables the exploration of indirect (*i.e.* genetic) benefits of mate choice. Second, MHC in passerines is highly polymorphic and polygenic, with an intriguingly extensive copy number variation. This remarkable diversity demands robust and sophisticated genotyping strategies in order to provide the resolution needed to test evolutionary and ecological hypotheses. The aim of **paper I** in this thesis was, therefore, to evaluate and highlight discrepancies arising from different sequencing platforms, primer set-ups and allele calling pipelines when genotyping MHC. While consistent genotypes were retrieved across strategies in the moderately variable bluethroat MHCI, we found that the genotyping strategy employed mattered greatly when applied to the hypervariable bluethroat MHCII. By using family data and replicates, we established a variant filtering scheme based on Sommer *et al.* (BMC Genomics, 14, 2013, 542). Although this is unlikely to describe the complete allelic repertoire in individuals, we believe this is a useful approach when genotyping highly variable MHC.

In **paper II**, we revealed non-random, self-referential female choice of extra-pair males that rendered offspring with an MHCII repertoire size close to the

population mean in the bluethroat. The number of unique, functional MHCII alleles ("PSS alleles"; based on positively selected sites) was closer to the population mean in extra-pair partners than in social pairs, and accordingly, extra-pair young were closer to the population mean number of PSS alleles than within-pair young. This paper consequently demonstrates extra-pair mate choice for intermediate MHCII diversity, with coherent results across adults and their offspring. A possible fitness benefit of having an intermediate number of MHCII alleles was explored in **paper III**, by testing for an association between the distance from the population mean number of PSS alleles and the response to injections of phytohemagglutinin (PHA) in bluethroat nestlings. We found that nestlings with MHCII diversity close to the population mean had an elevated swelling response to PHA. Together, **paper II** and **paper III** lend support to the hypothesis of an intermediate optimum in MHCII diversity in the bluethroat, and demonstrate an adaptive benefit of female extra-pair behaviour in this passerine bird.

In contrast, we did not find evidence of extra-pair mate choice for an intermediate MHCI diversity in the willow warbler in **paper IV**. There was also no support for the prediction that females socially paired to a suboptimal male will be more likely to partake in extra-pair mating. Rather, our results suggested two female mating strategies, where females with few "supertypes" (clusters of PSS alleles based on physiochemical properties of the amino acids) were more likely to engage in extra-pair copulations. This thesis thus points to variation in MHC-based mate choice across two passerine species and across the two major classes of MHC genes.

By taking advantage of new sequencing methodology, we achieved a high resolution in genotypes in hypervariable MHC, enabling us to test novel hypotheses about MHC-based mate choice in passerines. This thesis has contributed to increasing our knowledge of avian MHC-based mate choice and to the long-standing debate concerning adaptive benefits of female extra-pair mating in birds.

Acknowledgements

First and foremost, I wish to thank my supervisors, professor Jan T. Lifjeld, professor Arild Johnsen and lab manager Jarl Andreas Anmarkrud. I could not have been more satisfied with my supervisor team, and I highly appreciate each one of you. Thank you for being great academic role models – I feel honored and lucky to be guided through these years by you.

Jan; thank you for being enthusiastic, positive and encouraging. I have truly enjoyed being your student, and I am even more impressed by both biology itself and you after almost every discussion. Your way of keeping sight of the general concepts and having the biological relevance in mind, no matter how chaotic and complex I feel a project is, has really been educating and motivating for me.

Arild; your skillfulness and kindness have been invaluable during my PhD. I know it has been said by others before, but I need to repeat it: thank you for always having your door open. I hope you know how much this has meant for both the thesis and for me. Thank you for being understanding and wise, for being a role model in so many ways, and for all academic input. You have taught me so much.

Jarl Andreas; thank you for your high spirits and superb technical assistance! I am grateful for all your support and for your patience and help – regardless if my questions are large or awkwardly small. I am proud of being your first PhD student, and hope many others will get the chance.

Oddmund Kleven; thank you for conducting the microsatellite paternity analysis on the willow warblers, and for your thorough and quick responses.

Erica Leder; thank you for all your help, and especially thanks for helping me open the door to the world of bioinformatics. I appreciate your hospitality (and your interest in outdoor sports!) during my visits to Turku, Tartu and Tjärnö, which were all memorable trips.

Bjørn Aksel Bjerke; thank you for letting me use your excellent pictures in this thesis. They have really contributed to the visual appeal of the finished product.

Emma Whittington; thank you for proof-reading the introduction, and for bringing Tilly into my life.

I am sincerely grateful for having been part of SERG, also known as ”the bird group”, during the course of this PhD. Thank you for including and welcoming me, despite my continuous nagging about bread or tiny details on wording in a manuscript. It has been a pleasure and a large support to share big and small events in the everyday life with you! So - in addition to Jan, Arild, Jarl Andreas, Erica, Bjørn Aksel and Emma - thank you, Christian Aas, Diana Carneiro, Becky Cramer, Marie Danielsen, Gaute Grønstøl, Lars Erik Johannessen, Anna Kearns, Petter Zahl Marki, Taiwo Crossby Omotoriogun, Melissah Rowe, Audun Schrøder-Nielsen, Lisbeth

Thorbek and other affiliates for making SERG a marvelous group to spend my workdays with. Special thanks are due to Hanna Støstad, for being a dear and helpful office buddy for three years, and also to Gunnhild Marthinsen, who has been an empathic friend to trust when things are hard. I would also like to thank my other colleagues at NHM for contributing to a diverse and interesting place of employment, situated in the most wonderful scenery. I will truly miss the combination of a highly stimulating scientific environment and a friendly and inclusive workplace. To be happy at work not only makes the workdays more enjoyable, but I am convinced it has also increased my productivity - so thank you to all that have joined soccer, the "Blåstrupene" choir and other events.

Trude Magnussen; thank you for your friendship and countless strolls in the park. You are a fantastic friend, and I am so glad you have been at the museum these years. To the rest of "team shut up" (Siri Birkeland and José Cerca); thank you for great company during the last phase of the PhD. This would have been a lot harder without you.

Lastly, but most importantly, I wish to thank my family and friends for all support and for filling my life with so much joy. I am deeply indebted to my parents - thank you for always being there for me, for motivating me and for being a safe haven. This thesis would have been impossible without you. If I have ever managed something, it is because you equipped me for the tasks and supported me along the way. A word of cordial thanks also goes to my beloved grandmother, for believing in me (really) no matter what.

Hanne; thank you for understanding without the need of explaining, and thank you for always helping, even without doing too much. You and your honesty is a rescue line throughout my life.

To my daughter Eira; thank you for just being such an amazing little girl, and thank you for giving me perspectives. Martin; thank you for all your patience, for standing by my side through my frustrations, and for your endless efforts to say and do the right things. I look so much forward also to the next chapters in life with you!

Silje Rekdal

Oslo, January 2020

List of papers

The four papers listed below are included in this thesis. Paper I and paper II are published open-access, while paper III and IV are un-submitted manuscripts. For the two published papers, raw data and excessively large appendices in the form of excel spread sheets can be found online, as stated in data accessibility sections the respective papers. For paper III, supplementary material is printed in this thesis, except for the genotypes and the GenBank accession numbers (supplementary material S2 and S3) which are available upon request. Raw data, R-scripts and supplementary tables of individuals, primers, barcodes and genotypes (supplementary material S1, S3 and S6) are available upon request for paper IV. Other appendices for this paper are attached.

I) **Rekdal SL**, Anmarkrud JA, Johnsen A and Lifjeld JT (2018).
Genotyping strategy matters when analyzing hypervariable major histocompatibility complex - Experience from a passerine bird.
Ecology and evolution **8**(3): 1680-1692. doi: 10.1002/ece3.3757

II) **Rekdal SL**, Anmarkrud JA, Lifjeld JT and Johnsen A (2019).
Extra-pair mating in a passerine bird with highly duplicated major histocompatibility complex class II: Preference for the golden mean.
Molecular Ecology **28**(23): 5133-5144. doi: 10.1111/mec.15273

III) **Rekdal SL**, Anmarkrud JA, Lifjeld JT and Johnsen A.
Elevated PHA-response at an intermediate number of major histocompatibility complex class II alleles in bluethroat nestlings.
Manuscript prepared for submission to Biology Letters

IV) **Rekdal SL**, Anmarkrud JA, Kleven O, Lifjeld JT and Johnsen A.
Is extra-pair mating associated with the hypervariable major histocompatibility complex class I in willow warblers?
Manuscript

1. Introduction

To fight and resist infectious diseases is a ubiquitous challenge that all organisms face. Genes of the major histocompatibility complex (MHC) are of great importance to the vertebrate immune system, and there should thus be a selective pressure to optimize the genetic constitution at these loci in offspring. Given that the MHC genotypes vary substantially among potential mates, and that their genotypes can both be assessed and compared to the chooser's own genotype, the choosy sex is predicted to choose mates that will produce offspring with an optimal diversity at MHC (Milinski 2006). MHC genes are indeed among the most polymorphic genes known (Janeway *et al.* 2001a), and are hypervariable and highly duplicated in many passerine birds (Westerdahl 2007). While the underlying mechanisms are largely unknown, MHC-based mate choice has been demonstrated in many passerines - but previous results are equivocal and hampered by the difficulty of sequencing and genotyping MHC.

This thesis examines MHC-based mate choice in the context of extra-pair paternity in two passerine species, the bluethroat (*Luscinia svecica*) and the willow warbler (*Phylloscopus trochilus*). Mating systems with substantial occurrence of extra-pair paternity could be useful for highlighting the genetic benefits of mate choice, as the extra-pair male usually contributes only sperm and hence could be chosen for his genes alone (Trivers 1972, Brown 1997, Mays & Hill 2004). Genetic benefits for females engaging in extra-pair copulations are nevertheless not widely recognized, and debate over an adaptive function is still ongoing (Forstmeier *et al.* 2014, Brouwer & Griffith 2019). One such benefit could be to optimize the MHC constitution in offspring, because of its importance in an individual's disease resistance and hence survival (Milinski 2006).

Thus, this thesis focuses on MHC-based extra-pair mate choice, with the main aim of increasing the knowledge on its yet unresolved role in passerine birds, as well as to contribute to the understanding of female extra-pair behavior. A first step to be able to answer evolutionary questions involving MHC is to reveal the genotypes of the individuals, which is not trivial in hypervariable systems. The first goal of the thesis will be to highlight discrepancies that arise from using different genotyping strategies, as well as to establish a robust pipeline, taking advantage of the tremendous developments in sequencing technologies. The method will then be applied to test evolutionary hypotheses concerning MHC-based extra-pair mate choice in the bluethroat and willow warbler.

1.1. The function of MHCI and MHCII

MHC genes are found in jawed vertebrates (Janeway *et al.* 2001a), and are crucial for adaptive immunity. Classical MHC genes encode transmembrane glycoproteins which present pathogen-derived antigens to T cells and subsequently trigger immune responses against the specific pathogens. Two classes of MHC genes are relevant for this thesis; MHC class I (MHCI) and MHC class II (MHCII). For both classes, antigens are processed intracellularly to short fragments that are loaded on MHC molecules and presented to T cells (reviewed in Blum *et al.* 2013).

The gene products of MHCI are found on the surface of all nucleated cells, where they present *intra*cellular antigens to cytotoxic CD8$^+$ T cells (Zinkernagel & Doherty 1974). In the case of the binding of a CD8$^+$ T cell with the matching T cell receptor to the MHCI/antigen complex - and in the presence of co-stimulator signals and cytokines - the T cell will activate apoptosis in the target cell, either through the release of perforin and granzymes or through the Fas/FasL pathway (Coico & Sunshine 2015a). Consequently, MHCI is important in the killing of virus-infected and cancerous cells.

On the other hand, MHCII is essential for the fight against *extra*cellular pathogens (*i.e.* bacteria, parasites and free-floating viruses). Antigens derived from these pathogens are presented on MHCII molecules found on so-called professional antigen presenting cells (APC), such as B cells, macrophages and dendritic cells. The MHCII/antigen complex is bound by CD4$^+$ T cells exhibiting the corresponding T cell receptor. This signal is enhanced by the interaction between pairs of co-stimulator surface molecules and other adhesion molecules on the APC and CD4$^+$ T cell, and will subsequently lead to the activation of CD4$^+$ T cells. Cytokines released from innate immune cells will stimulate transcription factors causing the activated CD4$^+$ T cells to differentiate into T_H1, T_H2, T_H17 or T_{Reg} cells (Coico & Sunshine 2015a). These subsets of CD4$^+$ T cells will in turn release characteristic signature cytokines, which ultimately will lead to the destruction of the pathogens through effector cells like B cells and their antibodies, or macrophages (Janeway *et al.* 2001a, Coico & Sunshine 2015a).

In addition to enable the combat against specific pathogens, MHC plays critical roles in autoimmune diseases and in the body´s defence against cancer

(Trowsdale & Knight 2013). The significance of these genes is further emphasized by their vital influence on transplantation of tissues, from which they obtained their name.

1.1.1. MHC structure

The structural conformation of MHCI and MHCII molecules consists of two polypeptide chains with a total of four extracellular, immunoglobulin-like domains (Bjorkman *et al.* 1987a, Brown *et al.* 1993, Coico & Sunshine 2015b; see Figure 1).

MHCI molecules have one transmembrane heavy chain (~42 kDa) with three ~90 amino acids long extracellular domains (α_1, α_2, α_3) and a cytoplasmic tail (Bjorkman *et al.* 1987a, Coico & Sunshine 2015b). A light chain, the β_2-microglobulin (β_2m; ~12 kDa), is non-covalently bound to the heavy chain. While the membrane-proximal α_3-domain and the β_2m are predominantly conserved, the distal α_1 and α_2 are polymorphic. Together, they constitute the peptide-binding region (PBR), in which a "floor" consisting of a β-sheet is surrounded by α-helices creating a groove for antigen binding. This groove can bind one antigen-derived peptide with a length of 8-9 amino acid residues, and will only bind to specific peptides with high affinity (Matsumura *et al.* 1992). Most of the polymorphic amino residues in the MHCI molecules are located in this groove, and are mainly coded for in exon 3 (Bjorkman *et al.* 1987b, Hughes & Nei 1988). These residues determine the antigen specificity of the MHC by constituting pockets that anchor specific side chains of the antigen, while the remaining interactions between the MHC molecule and the antigen are primarily nonspecific (Matsumura *et al.* 1992). When the antigen is fixed in this groove, other residues of the peptide protrude outwards and are available for recognition by the T cell receptor (Matsumura *et al.* 1992).

Similarly, an antigen binding groove consisting of a β-sheet and two α-helices is present on MHCII molecules. This groove is more "open-ended" than the PBR on MHCI, and can fit antigens of more than 15 residues (Matsumura *et al.* 1992, Brown *et al.* 1993). While the MHCII anchoring pockets are slightly less specific than those found on MHCI, there are up to five such allele-specific pockets for MHCII as compared to two or three for MHCI (Matsumura *et al.* 1992, Stern *et al.* 1994, Rammensee 1995). The MHCII molecule further deviates from MHCI by being a dimerization of two relatively similar polypeptide chains (the α- and the β-chain; ~35 and 28 kDa) with two extracellular domains and a cytoplasmic tail each (Brown *et al.* 1993, Coico & Sunshine 2015b). The PBR is formed at the interface of the distal domains of these chains (between α_1 and β_1). The main bulk of the polymorphism found in the PBR are coded for in exon 2 of MHCIIB genes (Hughes & Nei 1989).

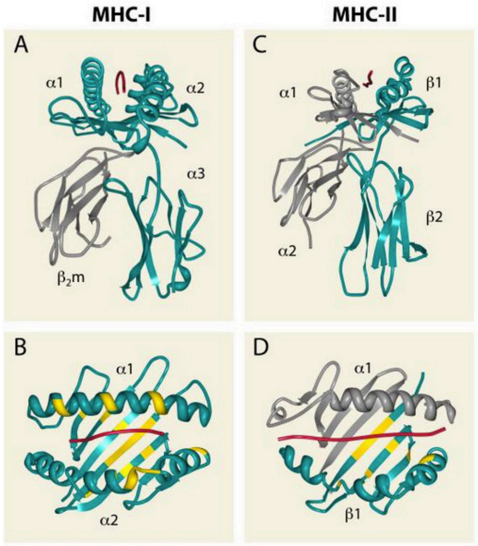


Figure 1: The three-dimensional structure of a complete MHCI molecule (**A**, side view) and the PBR of MHCI (**B**, seen from above), as well as an MHCII molecule (**C**) and the PBR of MHCII (**D**), with bound antigen (colored red). The heavy chain of MHCI and the β-chain of MHCII are colored teal, while the β_2-microglobulin of MHCI and α-chain of MHCII are in grey color. Polymorphic residues in the PBR for both molecules are highlighted in yellow (**B**, **D**). The model of MHCI (**A**, **B**) is based on the human MHCI allele HLA-A2 with residues 58-66 of the influenza matrix protein bound in the PBR, while the model of MHCII (**C**, **D**) is based on the human MHCII allele HLA-DR1 with residues 306-318 of the influenza hemagglutinin bound in the PBR. Reprinted from Blum *et al.* (2013) with permission (Copyright Clearance Center, license number 4599310116294).

1.1.2. Selection of T cells in thymus

Whereas the diversity in MHC molecules stems from multiple alleles, most of the immense T cell receptor diversity ($\sim 10^7$ in humans; Arstila *et al.* 1999) is generated by gene rearrangements in the thymus (Janeway *et al.* 2001b, Coico & Sunshine 2015c). In the thymus, only T cells with a T cell receptor capable of interacting with self-MHC are retained (positive selection; Kisielow *et al.* 1988). Autoimmunity is avoided by negative selection, *i.e.* T cells that bind with too high affinity to self-antigens undergo apoptosis (Kappler *et al.* 1987, Coico & Sunshine 2015c). Taking the extensive T cell receptor diversity into consideration, some of the T cells that have survived this thymic selection will likely bind with high affinity to a pathogen-derived peptide loaded on an MHC molecule, and in the presence of co-stimulatory signals, an immune response against the specific pathogen will be triggered (Coico & Sunshine 2015a).

1.1.3. Copy number variation

Because the MHC is both polygenic and polymorphic, while at the same time one MHC allele is able to bind a few different antigens, a large number of various pathogens can be recognized within an individual (Janeway *et al.* 2001a). Classical genes of the MHC are indeed considered the most polymorphic vertebrate genes (Janeway *et al.* 2001a), which reflects their role in adaptive immunity. On a macro-evolutionary scale, sharing of polymorphic allelic lineages has been described among species (Figueroa *et al.* 1988, Takahata & Nei 1990, Anmarkrud *et al.* 2010). As these ancestral polymorphisms can predate speciation events, they are referred to as trans-species polymorphism (Arden & Klein 1982, Klein 1987).

There is substantial variation in the number of classical MHC genes a species possesses. The evolutionary birth-and-death model of multi-gene families, in which gene duplication events create new MHC loci, is thought to give rise to this copy number variation (Nei *et al.* 1997). Aves is a great example of a taxon with extensive variation in copy number of MHC genes (Minias *et al.* 2018). Some species, found mainly among galliforms and birds of prey, have "a minimal essential MHC" (Kaufman *et al.* 1999, Minias *et al.* 2018), with only a couple of MHCI and MHCII loci described (*e.g.* Eimes *et al.* 2013). Many passerine species lie at the other extreme with many MHC genes (Westerdahl 2007), with the highest known to date being 33 MHCI loci in the sedge warbler (*Acrocephalus schoenobaenus*; Biedrzycka *et al.* 2017). Expression of multiple classical MHCI genes is demonstrated in siskins (*Spinus spinus*; Drews and Westerdahl 2019). The number of MHC alleles found within an individual also varies markedly *within* a species (*e.g.* Reusch *et al.* 2001, Woelfing *et al.* 2009, Lighten *et al.* 2014, O'Connor *et al.* 2016). However, as identical MHC variants possibly could be distributed across multiple MHC loci (*e.g.* Westerdahl *et al.* 2004), the term "allele" will in this sense mean a unique variant found within an individual and is not locus-specific, masking the true number of MHC loci in an individual.

1.1.4. Optimal number of alleles

The pronounced difference in intra-individual number of MHC alleles among and within species raises the questions of why there is such a high level of variation, and whether there exists an optimal number of alleles. On the one hand, if a pathogen could be presented to T cells by one MHC allele only, it could be beneficial to have as many alleles as possible to increase the number of pathogens one can fight off. Also, it increases the chance that a particular pathogen would be presented on an MHC molecule in an individual (Borghans *et al.* 2003, Woelfing *et al.* 2009, Westerdahl *et al.* 2012). More alleles may in addition increase the efficiency of an immune response, as more T cells could come into play if multiple epitopes from one pathogen could be presented by MHC molecules (Spurgin & Richardson 2010). However, there are likely selective forces that constrain the intra-individual number of MHC alleles. This is based on the notion that individuals possess only a small fraction of all alleles present in the population, even though MHC loci are capable of duplication and divergence (Woelfing *et al.* 2009, Lenz 2011). MHC duplication events are also often followed by "pseudogenization", further implying selective limitations on intra-individual diversity (Sawai *et al.* 2008, Lenz 2011).

Having a large number of MHC alleles may deplete the number of circulating T cells available for inducing an immune response, due to negative selection in the thymus (Vidović and Matzinger 1988, Nowak *et al.* 1992, Woelfing *et al.* 2009, Migalska *et al.* 2019, but see Borghans *et al.* 2003). There are also associations between certain MHC alleles and autoimmune diseases, which means that individuals with more alleles are running a greater risk of having such disorders (Gough & Simmonds 2007, Woelfing *et al.* 2009, Lenz *et al.* 2015). Lastly, having only a few MHC alleles may facilitate an effective immune response by ensuring that a sufficient number of T cells are activated, due to the increased concentration of specific peptide-MHC complexes (van den Berg & Rand 2003, Woelfing *et al.* 2009).

Borghans *et al.* (2003) argued that the selection pressure for having high MHC diversity is diminished by the flexibility in the peptide-binding of an MHC allele. As multiple antigens are derived from a given pathogen, and each MHC allele is able to bind a range of antigenic peptides (Sette *et al.* 1989, Matsumura *et al.* 1992), having an adequate number of different MHC alleles may merely be sufficient for combating a broad range of pathogens (Hughes & Nei 1989). The degenerate peptide-binding by MHC alleles has led to scientists grouping the alleles into supertypes with overlapping peptide-binding motifs (*e.g.* Sette & Sidney 1998, Trachtenberg *et al.* 2003, Doytchinova and Flower 2005), which might be a more biologically relevant level for selection to act on (Sette & Sidney 1999, Spurgin & Richardson 2010).

These arguments form the basis for the hypothesis that it is not the maximum number of MHC alleles *per se* that offers the greatest fitness benefit, but rather that there is an optimal, intermediate number of MHC alleles within an individual (Nowak *et al.* 1992, Milinski 2006, Kalbe *et al.* 2009, Woelfing *et al.* 2009, Spurgin & Richardson 2010). For instance, Wegner *et al.* (2003) found support for this hypothesis experimentally in three-spined sticklebacks (*Gasterosteus aculeatus*), by showing that individuals with intermediate MHC diversity had the highest fitness when exposed to multiple pathogens. Similar results have been found in other taxa as well, *e.g.* water pythons (*Liasis fuscus*; Madsen & Ujvari 2006) and bank voles (*Myodes glareolus*; Kloch *et al.* 2010). The number of alleles within an individual that constitutes the optimum may vary among species dependent on life history traits (*e.g.* Minias *et al.* 2017, Whittingham *et al.* 2018). One prediction is that this optimum is associated with the number of pathogens a species is exposed to (*e.g.* Milinski 2006). Following this train of thought, the number of MHC alleles per individual is for example expected to be higher in migrant species than in resident (Westerdahl *et al.* 2000, Freeman-Gallant *et al.* 2002, Minias *et al.* 2017, Minias *et al.* 2018, Whittingham *et al.* 2018). The higher pathogen load, due to increased transmission rates, in colonial breeding birds compared to solitary breeders could similarly lead to higher MHC diversity in the former (Minias *et al.* 2017). However, contrary to these predictions, Gangoso *et al.* (2012) found low MHC diversity in falcons despite being successful colonizers, birds of prey and including both migratory and colonial species (*e.g.* Eleonora's falcon (*Falco eleonorae*)). The authors suggest that other mechanisms, either in the innate or adaptive branch of the immune system, could have evolved to compensate for the low MHC diversity. Atlantic cod (*Gadus morhua*), for instance, lacks MHCII genes entirely, but has increased MHCI diversity and altered components of the innate immunity (Toll-like receptors), likely as compensation for the loss of MHCII (Star *et al.* 2011).

Further work on gadiform teleosts suggests that MHC may be important as "speciation genes", as a positive association between MHCI copy number and speciation rates has been identified (Malmstrøm *et al.* 2016). If the optimal MHC genotype is dependent on habitat, hybridization between locally adapted species will lead to offspring with an excessive MHC diversity. This could subsequently enhance postzygotic reproductive isolation and facilitate speciation through reinforcement of prezygotic barriers, with a more pronounced effect in species with many MHC loci (Eizaguirre *et al.* 2009a, Malmstrøm *et al.* 2016).

1.2. Selective forces acting upon MHC

Gene duplication, point mutation and inter-allelic gene conversion are the main genetic processes generating the vast array of different alleles found in a population (Nei *et al.* 1997, Edwards & Hedrick 1998, Spurgin *et al.* 2011, Promerová *et al.* 2013). The mechanisms that maintain MHC polymorphism can be divided into two main selection regimes: parasite-mediated selection and sexual selection (Piertney & Oliver 2006).

1.2.1. Parasite-mediated selection

Parasite-mediated selection can result in balancing selection through heterozygote overdominance, negative frequency-dependent selection or fluctuating selection (see Hedrick 1999, Sommer 2005, Spurgin and Richardson 2010). It is nevertheless difficult to disentangle the different selection mechanisms and determine their relative importance, as they may interact and are not mutually exclusive (Spurgin & Richardson 2010).

Overdominance

Heterozygote overdominance as a means of maintaining diversity at MHC was initially proposed by Doherty and Zinkernagel (1975), by realizing that heterozygotes may have increased "immunological surveillance function", *i.e.* that heterozygotes should be able to detect more pathogens (see also Hughes and Nei 1988, and Hughes and Nei 1989). A sub-model of the heterozygote overdominance model is the divergent allele advantage (Wakeland *et al.* 1990), which states that individuals with more divergent alleles will be able to bind a wider range of antigens on their MHC molecules, and thus fight more pathogens.

Negative frequency-dependent selection

Instead of focusing on heterozygosity *per se*, the negative frequency-dependent selection model takes the co-evolutionary arms race between pathogens and their hosts into consideration. The rationale behind this model is that pathogens will adapt to the most frequent alleles (Snell 1968, Bodmer 1972, Takahata & Nei 1990, Slade & McCallum 1992, Lively & Dybdahl 2000, Trachtenberg *et al.* 2003), and hence escape presentation on these MHC molecules. There will consequently be selection in the hosts for having rare MHC alleles, which in turn maintains MHC diversity (Slade & McCallum 1992).

The continuous co-evolution between parasites and hosts, where hosts evolve to recognize and fight pathogens, while the pathogens on their part evolve to escape recognition, has been termed “the Red Queen hypothesis”. This was first launched by Van Valen (1973), who described a “zero-sum game” between species and named it after the Red Queen in L. Carroll´s “Through the Looking-Glass” (1871); “*Now, here, you see, it takes all the running you can do, to keep in the same place*” (Figure 2).

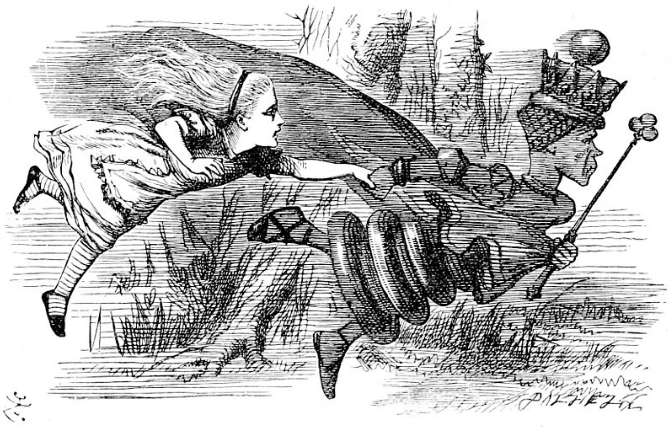


Figure 2: Alice in Wonderland and the Red Queen, illustrated by John Tenniel.

Fluctuating selection

The selection pressure from different pathogens can fluctuate in cycles in time and space (Hamilton & Zuk 1982). This can sustain MHC polymorphism (Hedrick *et al.* 1987, Hill *et al.* 1991, Edwards & Hedrick 1998), also in the absence of heterozygote overdominance and negative frequency-dependent selection (Hedrick 2002). This is because the specific MHC alleles which confer resistance against the prevailing pathogens will be subjected to positive Darwinian selection (Hughes & Nei 1989, Eizaguirre *et al.* 2012). When the pathogen regime is changing over time, different MHC alleles will be advantageous (Hedrick 2002). The intensity of the selection on the alleles will, therefore, fluctuate spatiotemporally, helping to maintain genetic diversity (Spurgin & Richardson 2010).

1.2.2. Sexual selection

As the generation time of many parasites is much shorter in many parasites than in their hosts, sexual reproduction and MHC-based mate choice can be a way for hosts to keep pace with the evolution of parasites in the "zero-sum game" (Ebert & Hamilton 1996, Penn & Potts 1999, Milinski 2006). Selective mate choice has further been described as an amplifier for pathogen-mediated selection (Spurgin & Richardson 2010), by having the potential to yield an optimal MHC constitution in offspring in their fight against the ever-changing pathogens (Penn & Potts 1999). An optimal constitution could in this respect be specific alleles to resist diseases (Briles *et al.* 1977, Hill *et al.* 1991, Bonneaud *et al.* 2005). Alternatively, but not mutually exclusive, MHC in offspring could be optimized through intra-individual *diversity*. This can be achieved by maximizing the diversity to increase the number of pathogens an individual can combat (Doherty & Zinkernagel 1975, Hughes & Nei 1988, Penn *et al.* 2002), or by making mate-choice decisions resulting in an intermediate optimal number of alleles as in section 1.1.4. Lastly, inbreeding avoidance could also be a cause of disassortative mate choice, in which MHC could play a central role; as MHC similarity between individuals is likely to increase with increasing relativeness, mate choice for dissimilar MHC could be a mechanism by which individuals avoid inbreeding (Egid & Brown 1989, Potts & Wakeland 1990, Potts *et al.* 1994). In contrast, extensive outbreeding can result in disruption of co-adapted genes, and mate choice for intermediate MHC dissimilarity might result from a trade-off between inbreeding and outbreeding depression (Bonneaud *et al.* 2006, Roberts 2009, Slade *et al.* 2019).

The processes underlying mate choice can be divided into three not mutually exclusive categories: good genes, male diversity and compatibility (Mays & Hill 2004, Kamiya et al. 2014).

Good genes

"Good genes" are specific alleles that have an additive genetic effect (Mays & Hill 2004, Neff & Pitcher 2005), and that will increase the fitness of the bearer (Hamilton & Zuk 1982). Importantly, in this framework, males possessing "good genes" will be attractive to all females (Mays & Hill 2004). Hamilton and Zuk (1982) proposed that one mechanism by which secondary sexual characters could signal male quality is via an association with parasite burden. Females thus may choose a male based on his genetic disease resistance as signalled in his display. This was further combined with the handicap principle suggested by Zahavi (1975), and explained by a negative feedback loop where testosterone could play a pivotal role by enhancing the secondary sexual characters and at the same time reducing the immunocompetence in a male (Folstad & Karter 1992). Individual males with "good immunological genes" could therefore have a reduced cost in terms of parasite infection, and tolerate an increased testosterone level and exaggerated sexual characters.

Evidence for mate choice based on specific MHC alleles that likely protect against prevailing pathogens has been found in some species (*e.g.* great snipes (*Gallinago media;* Ekblom *et al.* 2004) and sticklebacks (Eizaguirre *et al.* 2009b)). Further, associations have been identified between specific MHC alleles and sexually selected male traits (*e.g.* male size in great snipes (Ekblom *et al.* 2004), redness in sticklebacks (Jäger *et al.* 2007) and spur length in pheasants (*Phasianus colchicus* (von Schantz *et al.* 1996), supporting the hypothesis proposed by Hamilton and Zuk (1982). Eradication of genetic variation as a consequence of directional female choice is possibly avoided due to the co-evolutionary adaptive cycles in hosts and pathogens, where selected traits indicate the male's ability to withstand *current* parasites (Hamilton & Zuk 1982, Folstad & Karter 1992, Dunn *et al.* 2013).

Male diversity

Male MHC diversity itself can also be subject to sexual selection, in line with the suggestion of heterozygous males being of superior condition and having higher mating success (Brown 1997). Individuals with high MHC diversity will have increased disease resistance (Doherty & Zinkernagel 1975, Hughes & Nei 1988), and will also be more likely to carry advantageous alleles against several of the many prevailing pathogens (Dunn *et al.* 2013). Alternatively, females might prefer males with an intermediate MHC diversity (see section 1.1.4).

A meta-analysis by Kamiya *et al.* (2014) revealed female choice for MHC-diverse males in some taxa, although they could not delineate between selection for maximal or intermediate diversity. Ditchkoff *et al.* (2001) found that MHCII-heterozygote white-tailed deer (*Odocoileus virginianus*) had greater antler development; a secondary sexual character. Dunn *et al.* (2013) and Whittingham *et al.* (2015) found that sexually selected ornamental plumage traits in male yellowthroats (*Geothlypis trichas*) were positively associated with the number of unique MHCII alleles. The MHC diversity in extra-pair Seychelle warbler (*Acrocephalus sechellensis*) males, from which the females only obtain indirect genetic benefits, were higher than in their cuckolded rival (Richardson *et al.* 2005). This lends additional support to female choice for high MHC diversity in males.

In sticklebacks, gravid females preferred odor from males with many MHC alleles over males with few alleles, but not necessarily dissimilar from their own alleles (Reusch *et al.* 2001). Furthermore, stickleback females may evaluate males based on their nest quality, which was greatest in males possessing an intermediate number of MHCII alleles (Jäger *et al.* 2007). Similarly, large song repertoire in male song sparrows is associated with an intermediate MHC diversity (Slade *et al.* 2017).

Compatibility

Maybe the most extensively tested hypothesis with regard to MHC and mate choice is that of disassortative mating to enhance offspring diversity. By choosing a mate that is genetically dissimilar from oneself, inbreeding will be avoided and offspring will have an increased MHC repertoire for parasite resistance. This implies that there is no single "best male" in the population, and the male that is the "best fit" for one female, may not be the "best fit" for others (Brown 1997). Yamazaki *et al.* (1976) first demonstrated that mice preferred MHC-dissimilar mates, which has subsequently been suggested for a wide array of taxa, although with equivocal results (see Kamiya *et al.* 2014). The idea of MHC-dependent disassortative mate choice became popular after the study on humans by Wedekind *et al.* (1995), where female students not taking oral contraceptives preferred the odor of MHC-dissimilar men. The same disassortative pattern has been shown in fish (*e.g.* Landry *et al.* 2001), reptiles (*e.g.* Olsson *et al.* 2003) and birds (*e.g.* Freeman-Gallant *et al.* 2003, Strandh *et al.* 2012). Other studies demonstrate MHC-based mate choice for an intermediate dissimilarity within the couple (*e.g.* Bonneaud *et al.* 2006, Forsberg *et al.* 2007, Eizaguirre *et al.* 2009b, Griggio *et al.* 2011, Baratti *et al.* 2012).

If mate choice for compatible MHC genes is to take place, the choosing sex would have to not only be able to discriminate between possible mates' MHC constitution, but also to take their own into account. This might require complex sensory mechanisms, and might be less widespread than choice for diversity *per se* (Kamiya *et al.* 2014). The mode of female MHC-based mate choice may further be dependent on mating system, pathogen abundance and reproductive strategies (Mays & Hill 2004, Neff & Pitcher 2005, Winternitz *et al.* 2017). Also, females might use "nested rules" and choose based on compatibility among the males exhibiting "good genes", or the other way around (Aeschlimann *et al.* 2003, Freeman-Gallant *et al.* 2003, Mays & Hill 2004, Milinski 2006). While complicating the disentanglement of different selection mechanisms, this could help explain the inconsistent results found across taxa.

1.3. MHC and extra-pair paternity (EPP)

In order to comprehend both natural and sexual selection in species where extra-pair paternity (EPP) occurs, it is important to understand the genetic mating system (Brouwer & Griffith 2019). Exploring why females engage in extra-pair copulations offers a unique opportunity to obtain insight into the genetic benefits of mate choice, as they often only receive sperm from extra-pair males (EPM; Trivers 1972, Brown 1997, Mays & Hill 2004). The female might consequently choose an extra-pair male based on his genetic constitution only, while the choice of social (within-pair) mate also could be based on direct resources such as territorial quality and parental abilities (Birkhead & Møller 1992, Andersson 1994, Brown 1997), or could be the result of a hasty choice (Lindsay *et al.* 2019). Comparisons between within-pair and extra-pair mates, as well as between within-pair and extra-pair offspring, thus help elucidate the purely genetic benefits that might result from mate choice (Trivers 1972, Kempenaers et al. 1992, Johnsen et al. 2000, Foerster et al. 2003). Further, comparisons between social pairs or social males with broods with and without extra-pair young (EPY; sired by another male than the social male) could also contribute to reveal any benefits, as females might partake in extra-pair mating to "correct" for a suboptimal social mate choice (*e.g.* Freeman-Gallant *et al.* 2003, Schwensow *et al.* 2008).

The immense abundance and rapid evolution of parasites means genes involved in immunity and disease resistance are likely candidates for genetic benefits of mate choice (Milinski 2006). While studies have indicated increased fitness from extra-pair copulations (*e.g.* Kempenaers *et al.* 1992, Johnsen *et al.* 2000, Foerster *et al.* 2003), results when studying correlations between MHC diversity and extra-pair copulations are mixed and inconclusive.

Two studies on birds imply selection for male MHC diversity, by suggesting a negative association between within-pair male (WPM) MHC diversity and paternity loss: in the scarlet rosefinch (*Carpodacus erythrinus*), WPM with high MHC diversity were less cuckolded than WPM with low MHC diversity (Promerová *et al.* 2011), and female Seychelle warblers paired with WPM with low MHC diversity had an increased risk of obtaining EPP (Richardson *et al.* 2005). In this population of Seychelle warbler, Brouwer *et al.* (2010) found an association between MHC diversity and juvenile survival, which, combined with the study of Richardson *et al.* (2005), implies a genetic benefit of EPP through male diversity.

On the other hand, a study in Savannah sparrows (*Passerculus sandwichensis*) by Freeman-Gallant *et al.* (2003) lends support to the compatibility hypothesis, by showing that females with no EPY were paired to more MHC-*dissimilar* WPM than females with EPY. However, the opposite pattern was observed in the yellowthroat by

Bollmer *et al.* (2012); females paired with more MHC-dissimilar WPM had a greater proportion of EPY. The authors argue that this could arise from selection for intermediate MHC diversity, although they found no further evidence for this. Yet, while Winternitz *et al.* (2015) found no evidence for choice of EPM based on MHC dissimilarity in the scarlet rosefinch, the EPY had a smaller variance in the number of MHC alleles (but not supertypes) than within-pair young (WPY; sired by the social male). This finding is in line with selection for an intermediate number of MHC alleles, although this was indicated with offspring data only and not adults.

MHC diversity could further be linked to the degree of promiscuity (Westerdahl 2004, Gohli et al. 2013, Winternitz et al. 2013). Gohli et al. (2013) found that more promiscuous species had higher MHCII diversity across passerine birds, and hypothesized that female promiscuity in passerines could have evolved as a way to keep their immune system up with ever-changing pathogens. Variation in promiscuity could in this way be connected to the strength of selective pressure from pathogens and consequently to MHC diversity (Gohli et al. 2013). This suggests that extra-pair mating strategies could have evolved in species with high pathogen load, to obtain immunogenetic benefits (Lindsay *et al.* 2019).

1.4. Mechanisms for MHC-based female choice in birds

While MHC-based mate choice has been reported in a wide range of vertebrate species (*e.g.* Tregenza & Wedell 2000, Kamiya *et al.* 2014; see sections 1.2.2 and 1.3), the proximate mechanisms remain unclear. Most studies on this topic have been conducted in mammals (*e.g.* Penn 2002, Leinders-Zufall *et al.* 2004, Ziegler *et al.* 2005, Sturm *et al.* 2013), but studies suggesting mechanisms in birds are also emerging. Importantly, the idea that birds do not possess an olfaction sense has been refuted, and several recent studies have indicated odor as a genetic assessment mechanism (Caro *et al.* 2015). For instance, Strandh *et al.* (2012) found MHC-based mate choice in a procellariiform bird species (blue petrel (*Halobaena caerulea*)) with highly developed olfaction. The authors proposed an odor-mediated discrimination of individuals through uropygial (preen) gland secretions or plumage microbiota - either directly through differences in the odor-producing microbial community originating from the hosts' MHC genotype, or indirectly through their microbial metabolites. Further, they argued that because bacterial antigens are mainly presented on MHCII molecules, these mechanisms could lead to associations between MHCII genotype and individual odor and MHCII-based mate choice. This has gained support from studies by Leclaire *et al.* (2014) and Slade *et al.* (2016), demonstrating associations between MHCII-similarity and composition of preen secretions in kittiwakes (*Rissa tridactyla*) and song sparrows (*Melospiza melodia*), respectively. A recent study by Grieves *et al.* (2019) further suggested that song sparrows are able to discriminate individuals based on MHC dissimilarity and MHC diversity through preen oil. MHCII genotypes have also been found to correlate with plumage microbiota in the blue petrel, possibly due to bacterial breakdown of preen wax into volatile odorants (Leclaire *et al.* 2019).

In addition to precopulatory mate choice, there are postcopulatory opportunities for female choice, resulting in differential fertilization success among sperm (see for example Lenz *et al.* 2018). In birds, cryptic MHC-based female choice was implicated by Alcaide *et al.* (2012), who found transmission rate distortion of male MHCII-haplotypes in the lesser kestrel (*Falco naumanni*). Also, a greater number of sperm from MHCI-dissimilar males were trapped in the perivitelline layer of the egg compared to MHC-similar males in mating experiments in red jungelfowl (*Gallus gallus*; Løvlie *et al.* 2013). However, in order to understand the mechanism of this cryptic female choice, it is crucial to unravel whether MHC is expressed on sperm, and whether this occurs for MHCI and MHCII, and at the haploid or diploid level (Promerová *et al.* 2017).

1.5. Thesis aims and objectives

The overarching aim of this thesis is to test predictions concerning female mate choice for optimal MHC, realized through extra-pair copulations. This could help elucidate the genetic benefits of extra-pair copulations and mate choice, and reveal potential differences in female choice of WPM and EPM. Further, this thesis could conceivably provide new, and elaborate on existing, knowledge about MHC-based mate choice, which ultimately could enhance the general understanding of the co-evolutionary arms race between pathogens and the hosts.

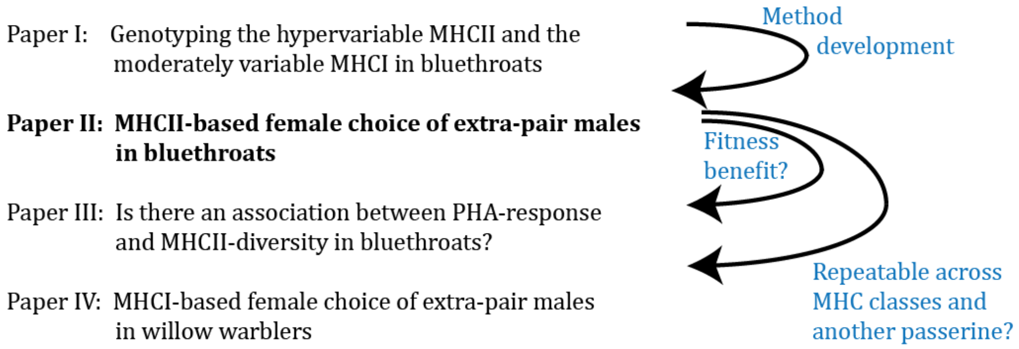
This thesis will be based on previously collected samples of bluethroats (*Luscinia svecica svecica*) and willow warblers (*Phylloscopus trochilus trochilus*); two passerine species with extreme MHC diversity and high levels of EPP. Genotyping of MHC will be conducted with next-generation sequencing, enabling insights into the allelic repertoire that were not previously possible.

However, as new sequencing methodology is still error-prone, sophisticated allele calling strategies are needed in order to approach true genotypes. Consistent allele calling is indeed a premise for using the data to answer questions related to behavior, ecology and evolution. **Paper I** is dedicated to comparing genotyping strategies of bluethroat MHCI and MHCII; two markers with medium and exceptionally high polymorphism, respectively. We will also aim to establish a robust genotyping pipeline which can be applied in the succeeding papers. Sequencing replicates and patterns of inheritance between parents and offspring will facilitate and help evaluate the genotyping.

In **paper II**, MHCII-based female mate choice for EPM is explored in a dataset consisting of bluethroat nests with the associated female, WPM and EPM, as well as WPY and EPY. The aim of this paper is to reveal if there is any pattern connected to MHCII diversity and female choice of EPM, *i.e.* if EPM or extra-pair partners have a more optimal MHCII constitution than WPM or social partners. A corresponding pattern when comparing WPY with EPY as found among established pairs, is expected. The dataset utilized in this study is part of the dataset in which Johnsen *et al.* (2000) described an enhanced immunocompetence in EPY as compared to both their maternal and paternal within-pair half siblings, which suggests that there is a benefit of extra-pair mating based on the compatibility of maternal and paternal genes. The prediction that bluethroat nestlings with MHCII diversity close to the population mean will have an increased response to injections of phytohemagglutinin (PHA) is tested in **paper III**. An association between these variables would suggest a fitness benefit of possessing intermediate MHCII diversity, as previous studies have demonstrated a correlation between PHA swelling response and fitness traits (*e.g.* Møller and Saino 2004).

Paper IV aims to investigate MHC-based female choice of EPM in a similar manner as **paper II**, but with focus on the other class of MHC and in another species: the highly diverse willow warbler MHCI. Inclusion of broods and social pairs also without EPY enables to test the prediction that females paired to a suboptimal male will be more likely to engage in extra-pair copulations.

With this thesis, I aim to contribute to the fields of MHC-related research and research on extra-pair mating. This will span different areas of the fields; including laboratory and bioinformatic approaches to establish an MHC genotyping pipeline (**paper I**), an evolutionary study on MHCII-based extra-pair mate choice in bluethroats (**paper II**), and a study on a possible fitness benefit of an intermediate MHCII diversity in the same system (**paper III**). The generality of the findings with regards to MHC-based extra-pair mate choice will be approached by conducting a similar study on MHCI in another passerine, using the same genotyping methods (**paper IV**). Support for non-random female extra-pair mate choice, as well as any fitness benefits of EPP indicated, could contribute to the long-standing debate on female benefits of extra-pair mating (*e.g.* Forstmeier *et al.* 2014). See Figure 3 for an overview of the papers in this thesis.



Paper I: Genotyping the hypervariable MHCII and the moderately variable MHCI in bluethroats

Paper II: MHCII-based female choice of extra-pair males in bluethroats

Paper III: Is there an association between PHA-response and MHCII-diversity in bluethroats?

Paper IV: MHCI-based female choice of extra-pair males in willow warblers

Figure 3: A schematic overview of the papers in this thesis, and how they relate to one another. I regard **paper II** as the principal work in this thesis, and have marked this paper in bold.

2. Methodological overview

2.1. Study species and sampling

The studies in this thesis are conducted on two passerine bird species; the bluethroat (*Luscinia svecica svecica*, Linnaeus, 1758; **paper I**, **paper II** and **paper III**) and the willow warbler (*Phylloscopus trochilus trochilus*, Linnaeus, 1758; **paper IV**), sampled in Øvre Heimdalen, Norway.

2.1.1. Bluethroat

The bluethroat is a small (12-25 g), sexually dichromatic and chiefly insectivore bird (Collar 2005) in the family Muscicapidae (Figure 4). The species is distributed in Eurasia and Alaska, with several subspecies described, often distinguished by plumage color of the central spot of the male throat (Collar 2005, Johnsen *et al.* 2006). The population studied in this thesis is of the subspecies *L. s. svecica*, in which the males have a chestnut-colored throat patch (Johnsen *et al.* 2006). This is an obligate migratory population which overwinters predominantly in areas in southwest-Asia (Collar 2005).

Figure 4: Bluethroat male. Photo: Bjørn Aksel Bjerke

This thesis utilized bluethroat samples collected in 1998, 1999 and 2014 by affiliates of SERG. Around 70 bluethroat males defend their territories in the Heimdalen field site (Johnsen *et al.* 2000), and in 1998 and 1999 a complete overview of the established territories and sampling of all males were attempted. Although socially monogamous within a breeding season, the bluethroat is a promiscuous species, and females often obtain extra-pair copulations from other males while breeding in the territory of their WPM (Krokene *et al.* 1996). This results in about 50% of nests containing one or more EPY, and about 26% of chicks being EPY (Johnsen & Lifjeld 2003). Thus, the chicks in one nest are often sired by multiple males, and the individual males can have offspring in multiple nests. Identification of the WPM for each female was done through field observations: color-banding of the males allowed us to establish the territory owners as well as observe mate guarding and paternal care behaviors. Determination of EPM, WPY and EPY was done based on microsatellites by Johnsen *et al.* (2000), Fossøy *et al.* (2008) and Sætre *et al.* (2018).

2.1.2. Willow warbler

The willow warbler is in contrast to the bluethroat an inconspicuous and sexually monomorphic passerine bird (Figure 5), belonging to the family Phylloscopidae. Willow warblers are highly abundant in Norway, with ~4 million breeding pairs (Bairlein 2006) and breeds throughout the country and in the rest of northern Eurasia, often in close connection to birch and other deciduous trees and shrubs (Bairlein 2006). The species is highly migratory, and the two subspecies found in Norway – the southern *P. t. trochilus* and the northern *P. t. acredula* – differ slightly in morphology and have separate migration routes (Hedenström & Petterson 1987, Bensch *et al.* 1999). The two subspecies exhibit a narrow migratory divide across Scandinavia, estimated at around 63°N in Sweden (Bensch *et al.* 1999, Bensch *et al.* 2009, Liedvogel *et al.* 2014) and 65°N in Norway (Støstad *et al.* 2016). *P. t. trochilus*, the subspecies to which the Heimdalen individuals belong, migrates southwest to sub-Saharan west-Africa (Hedenström & Petterson 1987, Chamberlain *et al.* 2000, Bensch *et al.* 2006). Similar to the bluethroat, the willow warbler is small (6-15 g) and primarily insectivorous (Bairlein 2006). The level of extra-pair copulations in the Heimdalen population is also comparable to that in the bluethroat; 45% of the nests were cuckolded and 33% of chicks were sired by a male other than the social male (Bjørnstad & Lifjeld 1997).

The willow warblers analyzed in this thesis were sampled in Heimdalen during the field seasons of 1993, 1996-1998 and 2001-2003. As for the bluethroat, social pairs were identified in the field. The paternity status (WPY/EPY) of the chicks was resolved based on microsatellites and comparisons with genotypes of the female and WPM, as part of **paper IV**.

Figure 5: Willow warbler in Heimdalen. Photo: Bjørn Aksel Bjerke

2.1.3. Sampling

Sampling took place at a field site of about 3 km^2 in the subalpine valley of Øvre Heimdalen, Øystre Slidre in Norway (61°25' N, 8°52' E, approximately 1100 meters above sea level; Figure 6). Shrubs of willows (*Salix* sp.), dwarf birch (*Betula nana*) and juniper (*Juniperis communis*) interspersed with open ground dominate the vegetation. Elements of bilberry (*Vaccinium myrtillus*) heaths, tall herb meadows and mire vegetation are also present, in addition to a belt of mountain birch (*Betula pubescens* var. *pumila*) forest in the south-facing slope (Østhagen & Egelie 1978).

Figure 6: The subalpine valley of Øvre Heimdalen, Øystre Slidre;
the field site for sampling of bluethroats and willow warblers for this thesis. Photo by myself

Adult birds were caught in mist nets. Playback of bluethroat or willow warbler song was used to attract males after territory establishment. After puncturing the *vena brachialis* using a small needle, a droplet of blood (maximum 25 μL) was collected by capillary tube and stored in 1 mL plastic vials for subsequent analyses. Morphological traits of adults were also measured (body mass, and tarsus and wing lengths). Bluethroat individuals were aged according to Svensson (1992).

Blood was sampled and nestling weighed 2-9 days after hatching. Occasionally, blood was collected from the *vena metatarsus* using the same technique described above. For **paper III**, PHA assays (see below) and a strict scheme of weighing was conducted.

Adult birds were ringed with a numbered metal ring (retrieved from Stavanger museum, Norway), in addition to a unique combination of three colored plastic rings. The plastic rings enabled identification of the individual birds in the field, and allowed us to determine the identity of the social pair male based on observations of mate guarding and parental care during nesting period.

2.1.4. PHA immune assay

In **paper III**, skin swelling in the wing following subcutaneous injections of phytohemagglutinin (PHA) was measured as an approximation of immune responsiveness (Goto *et al.* 1978, Martin *et al.* 2006, Tella *et al.* 2008, Vinkler *et al.* 2010, Vinkler & Albrecht 2011, Vinkler *et al.* 2014). This test is easily conducted in the field, benefiting from work on domestic fowl (*e.g.* Greaves *et al.* 1968, Stadecker *et al.* 1977, Goto *et al.* 1978) and adopted by immunoecological and evolutionary studies of wild birds (*e.g.* Johnsen *et al.* 2000, Hasselquist 2007, Audet *et al.* 2015).

PHA is a lectin produced by red kidney beans (*Phaseolus vulgaris*), likely involved in anti-herbivory and toxic to humans when eaten raw (Martin *et al.* 2006). The molecule binds to the T cell receptor, although not through the antigen binding site, and causes a large fraction of the T cells to undergo mitosis (Coico & Sunshine 2015a). Injections of PHA cause local cell damage and infiltration of immune cells, mainly lymphocytes, macrophages and eosinophil and basophil granulocytes, with a change in cellular composition according to time since injection (Stadecker *et al.* 1977, Goto *et al.* 1978, McCorkle *et al.* 1980). Initially, PHA-stimulated T cells might produce lymphokines that attract other immune cells (Stadecker *et al.* 1977, McCorkle *et al.* 1980), that again cause vasodilation, inflammation and swelling (Martin *et al.* 2006). The complex reaction to PHA, involving both the innate and adaptive branch of the immune system, has triggered a debate concerning the use of this test to measure cell-mediated adaptive immunocompetence (*e.g.* Martin *et al.* 2006, Vinkler *et al.* 2010, Vinkler *et al.* 2012). The observation of reduced PHA-response in thymectomized chicken, however, strongly suggests that this is a thymus-dependent response (Goto *et al.* 1978). Tella *et al.* (2008) further advocated the use of PHA as an indicator of adaptive T cell mediated immunocompetence due to their findings of a larger secondary than primary response to the mitogen, and proliferation of specific T cells important in adaptive immune responses.

In the PHA immune assay used in **paper III**, PHA dissolved in saline was injected subcutaneously in the wing of bluethroat chicks at day 5 (sensitizing injection in the outer, metacarpal region) and day 7 (treatment injection in the inner, ulnar region) post-hatching. Pure saline was injected in the other wing. The thickness of both wings were measured immediately before and 24 hours after the treatment injection. The PHA-response (in mm) was given as:

$$\Delta thickness_{PHA-treated\ wing} - \Delta thickness_{Saline-treated\ wing}$$

2.2. MHC sequencing and genotyping

DNA was extracted from collected blood samples. For both MHCI and MHCII, extensive gene duplications and high sequence similarity may severely complicate single-locus amplification (Westerdahl 2007). Thus, we did not use locus-specific primers, but rather a multilocus amplification approach. As allele calling consequently could be challenging, all samples were sequenced in duplicates with a unique barcode for each replicate. Family data further facilitated genotyping.

2.2.1. Sequencing

In **paper I**, sequencing of bluethroat MHCI (third exon) and MHCII (second exon of the β-chain) was conducted on an Ion Personal Genome Machine twice; once with a short, single index primer setup (SI; the barcode and barcode adapter included on the forward primer only), and once with a longer, dual index primer setup (DI; IonTorrent adapter, barcode, barcode adapter and a seven nucleotide spacer motif included on both the forward and the reverse primer). Additionally, MHCII was sequenced on an Illumina MiSeq machine, with Illumina Linker sequences, barcodes and a 1-7 nucleotide heterogenic spacer motif (Fadrosh *et al.* 2014) on both forward and reverse primer.

For the dataset used in **paper II** and **III**, sequencing of bluethroat MHCII was conducted on Illumina MiSeq with the same primer setup as used for the MiSeq sequencing in **paper I**. The willow warbler MHCI was sequenced on an Ion GeneStudio S5 system for **paper IV**, with a dual index barcode setup.

2.2.2. Allele calling and genotype establishment

Although next-generation sequencing offers great advances and enables studies not previously possible, the methods are error prone. Genotyping of highly polymorphic, multilocus markers - such as the bluethroat MHCII and willow warbler MHCI - is especially demanding. This is due to the massive amplicon sequencing of similar alleles making the separation of true alleles from sequencing artefacts more challenging (Babik *et al.* 2009). We utilized two general assumptions regarding sequencing artefacts: the artefacts should be less abundant than real alleles, and they should originate from real alleles (Babik *et al.* 2009). We accordingly developed a workflow with a strict filtering scheme (**paper I**), based on the pipeline from Sommer *et al.* (2013). Briefly, after quality filtering, barcode splitting and clustering of identical reads were conducted in the program jMHC (Stuglik *et al.* 2011). Further, we employed a cut-off threshold of 0.2%, in which variants with an intra-amplicon frequency of <0.2% were discarded for the amplicon in question, following Galan *et al.* (2010). Chimera detection was accomplished using UCHIME (Edgar *et al.* 2011). For each amplicon, the most frequent variant was scored as allele, and the remaining variants were divided into "=1bp" (one base pair) and ">1bp" (more than one base pair) variants, according to the difference in number of base pairs to their most similar, more frequent variant. As the "=1bp" variants are more likely to be sequencing artefacts, these were treated more strictly, and called as alleles only if present (and not removed during filtering steps) in the replicate amplicon. The ">1bp" variants were called as alleles if also present in the family (see **paper I**, **paper II** and **paper IV** for details). While we cannot rule out repeatable errors completely, the stringent filtering employed likely removes the majority of artefacts. In **paper I**, allele calling was additionally conducted by using the program AmpliSAS (Sebastian *et al.* 2016).

Paper I is based on nucleotide genotypes. However, in order to answer evolutionary and ecological questions, we focused in **paper II**, **III** and **IV** on amino acid residues under positive selection (PSS; positively selected sites), which presumably are important in antigen binding (Hughes & Nei 1989, Sepil *et al.* 2012). We identified these residues using the program CodeML in the package PAML (Yang 2007), and selected those that have previously been described as under positive selection in passerines (*e.g.* Balakrishnan *et al.* 2010). Using only these residues, we established PSS alleles from the nucleotide alleles. The PSS alleles were further grouped into functional supertypes by their physiochemical properties of the amino acids, in an attempt to group alleles that bind similar antigens (Sandberg *et al.* 1998, Doytchinova & Flower 2005, Sepil *et al.* 2012). This was done through discriminant analysis of principal components (DAPC) in the R package adegenet (Jombart 2008, Jombart *et al.* 2010). We further established genotypes for PSS alleles and supertypes for each individual included in **paper II**, **III** and **IV**.

Throughout this thesis, the different, putatively functional MHC variants are referred to as “alleles”, although they originate from several loci, in line with the main bulk of avian MHC literature. “Genotype” is correspondingly used as a characterization of all alleles co-amplified within an individual. After grouping the nucleotide variants into PSS alleles and supertypes, we chose to focus on the average and sum of the amino acid differences between all pairs of PSS alleles (**paper II**), as well as the *number* of PSS alleles and supertypes (*i.e.* allele count; **paper II**, **III** and **IV**). This is based on the assumption that the structural diversity in chemosignals important to MHC-based mate choice could be related to the number of MHC alleles expressed (Milinski *et al.* 2005, Boehm & Zufall 2006). Additionally, by clustering the PSS variants into supertypes, we aimed for encompassing the functional diversity in antigen binding among the PSS sequences based on their physiochemical properties, instead of using the number of amino acid differences. The results from **paper II** also made the PSS allele and supertype count the candidate variables when moving on to **paper III** and **IV.**

2.3. Statistical analyses

Comparisons between the strategies employed when genotyping bluethroat MHCI and MHCII in **paper I** (*i.e.* primarily primer setups and allele calling pipelines) were conducted primarily through errors revealed from the lack of Mendelian inheritance and through the number of alleles called by each strategy. Initially, “a combined genotype” for each individual was established, consisting of all alleles called for an individual across all strategies. In order to compare strategies, we calculated the average proportion of “the combined genotype” each strategy was able to identify.

While the sequencing and allele calling for MHC genotypes involved the use of state-of-the-art technologies, the statistical tests in **paper II** themselves are relatively simple. In this paper, we tested two main hypotheses concerning MHCII-based mate choice in bluethroats realized through extra-pair mating: 1) that there was a difference in *mean* MHC diversity between within-pair and extra-pair units (*i.e.* males, pairs and offspring), implying selection for maximized diversity, and 2) that there was a difference between the groups in their *distance* from the population mean number of PSS alleles and supertypes, in line with the framework of selection for an intermediate optimum. Differences in the *mean* number of PSS alleles and supertypes between within-pair and extra-pair units were explored with paired t-tests (paired within nests) and Welch’s t-test for unequal variances, as well as linear mixed models for offspring, controlling for the identity of the parents. We also established a linear model to see if there were any correlations between the number of PSS alleles in the female, and her WPM and EPM. We further calculated the *distance* from the population mean number of PSS alleles for all individuals, and likewise conducted paired t-tests, Welch's t-test and linear mixed models to test if there was any difference between within-pair and extra-pair units. Additionally, we also conducted Levene's test to test the equality of variances, *i.e.* to test if the number of PSS alleles or supertypes were more concentrated around an intermediate number in one group than the other. Pseudoreplication in the offspring data was avoided by randomly choosing one individual from each nest, with bootstrapping. Finally, we tested if the distance from the population mean number of unique PSS alleles in the social and extra-pair partners deviated from random expectations, by simulating random female choice among all males, and comparing the observed mean values against the random distribution.

In **paper III**, the association between PHA-response and intermediate MHCII diversity in bluethroats was tested using a linear mixed model. PHA-response was entered as the response variable, and the distance from population mean number of PSS alleles, as well as paternity, body mass on the day of the second PHA injection and mean temperature during the nestling period as predictor variables. Nest identity was used as a random factor. The model included within-nest centring (Van de Pol & Wright 2009) of the MHCII diversity variable, in order to consider the inter-correlation of MHCII diversity within nests and to separate the among-broods and within-brood effect.

Similar tests as in **paper II** were also conducted in **paper IV**, where we compared the differences in the mean number of PSS alleles and supertypes, in addition to the differences in their variances, between willow warbler WPY and EPY from broods with mixed paternity. EPM were not identified, but as the dataset included social pairs and broods without any EPY, we could test the hypothesis that females paired to a suboptimal male will be more likely to engage in extra-pair copulations. Comparisons were thus made between groups (defined by the presence of EPY in their associated brood) of social pairs, WPM, WPY and females, in a similar manner as for WPY and EPY from mixed paternity broods. A non-random distribution of EPY was also tested for, using a X^2-test and binomial exact test of goodness-of-fit. Lastly, we tested for an association between the proportion of EPY in the brood and the number of supertypes within the female, using a quasi-binomial generalized linear model.

3. Main findings and discussion

3.1. Genotyping strategy matters!

Paper I demonstrates that the strategy employed matters greatly when genotyping highly polymorphic markers, such as the bluethroat MHCII. The aim in this paper was two-fold; to establish a robust strategy for genotyping polymorphic MHC, and to highlight discrepancies that could arise when applying different strategies. Eight bluethroat individuals from two family groups were sequenced in replicates on the moderately diverse MHCI and the hypervariable MHCII on an Ion PGM, using a single and a dual index approach. Alleles were called through two pipelines: a modified filtering method based on Sommer *et al.* (2013) and the software AmpliSAS (Sebastian *et al.* 2016). MHCII was additionally sequenced on Illumina MiSeq. We obtained consistent results when genotyping the moderately variable bluethroat MHCI. In contrast, the hypervariable bluethroat MHCII rendered divergent genotypes across the different strategies.

For the extremely diverse bluethroat MHCII, the number and identity of alleles varied both according to sequencing platform, primer strategy and bioinformatic pipeline. The pedigree information from the two family groups indicated that there were few errors in inheritance pattern, and therefore suggested that the allelic drop-outs were method-specific. In general, more alleles were called using the single index primer structure and the modified Sommer-pipeline, than for the other strategies. In addition to platform-specific errors causing differences between the platforms (*e.g.* homopolymer errors are relatively common when sequencing on Ion Torrent), platform-specific adaptors and primer structure could cause deviations in genotype. For instance, a primer “tail” consisting of the spacer/barcode/adapter could cause an amplification bias where variants that have complementary bases to these sequences outside the target primer binding site preferentially will be amplified. Further, as amplified fragments have complementary sequences to the whole primer and tail, the primers will more easily anneal to and thus amplify these than the original DNA template. The dual index primers had a longer tail and, unlike with single index primes, this is attached to both the forward and reverse primer. Consequently, the aforementioned mechanisms leading to genotyping errors should be greater when using dual index than single index primers. We found, as expected, that there were more allelic drop-outs using the dual index-approach than the single index.

Discrepancies between the allele calling pipelines could be due to algorithmic differences such as clustering of similar alleles (AmpliSAS), the use of replicates and chimera handling. While the modified Sommer-pipeline is considerably more time consuming and more prone to human errors than the automated AmpliSAS, it is more flexible and allows the utilization of existing information such as a pedigree. Although not directly comparable, we obtained more consistent genotypes within the families by the modified Sommer-pipeline than AmpliSAS.

Importantly, the use of replicates and family data greatly improved genotyping. We can, however, not completely rule out repeatable errors and other artefacts, and - as evident from the deviating genotypes - we are not able to reveal the true allelic repertoire using these methods. Nevertheless, we believe that next-generation sequencing and sophisticated allele calling approaches offer genotypes that are highly useful for answering evolutionary questions, to a degree not previously possible.

This study reveals an immense diversity in the bluethroat MHCII (>28 loci), which is among the most polymorphic MHC described for passerine birds (Biedrzycka *et al.* 2017). Genotyping such diverse markers requires vigilance and robust approaches, and we caution against comparing results obtained through different methods. In this respect, our study contributes to highlight potential pitfalls that should be considered when genotyping hypervariable MHC, but also shows that consistent genotypes can be obtained through carefully designed sequencing schemes and allele calling. It thus lends support to the application of the described strategies, even though the true genotypes may not be uncovered.

3.2. Extra-pair mating for the MHCII golden mean

In **paper II**, we explored MHCII-based mate choice in the bluethroat, a species exhibiting both a high degree of EPP and extreme intra-individual MHCII polymorphism. We compared functional MHCII diversity in within-pair and extra-pair mates, pairs and offspring, and found that the number of unique, functional

MHCII alleles (PSS alleles) was significantly closer to the population mean in extra-pair partners than in social pairs. As expected from this finding, we also showed that the number of alleles in EPY was closer to the population mean than in WPY. This lends support for non-random female choice of EPM, and implies that females chose EPM in a self-referencing manner producing EPY with an MHCII diversity closer to an assumed optimum, compared to WPY.

This result is relevant to four topics within MHC research. First, it adds to the body of studies implying selection for an intermediate optimum of MHC diversity; the "golden mean" (Wegner *et al.* 2003, Kalbe *et al.* 2009, Woelfing *et al.* 2009). This optimum is presumably a trade-off between the benefits and costs linked to high MHC diversity (*e.g.* pathogen surveillance versus T-cell depletion and autoimmune disorders, as discussed in section 1.1.4). There is currently no consensus on mating preferences for MHC diversity or dissimilarity across taxa (Kamiya *et al.* 2014), which is likely complicated by different optima among species (*e.g.* Westerdahl *et al.* 2000, Minias *et al.* 2018, O'Connor *et al.* 2018) and the current state of the population. If, for instance, the true optimum is above the population mean, mate choice for maximized diversity could be falsely inferred (Kamiya *et al.* 2014). Correspondingly, Aeschlimann *et al.* (2003) simulated a population bottleneck, and found that female sticklebacks under this scenario chose MHC-dissimilar males, while this was not the case in an outbred, heterozygous population. With this in mind, it is important to study wild and outbred populations. **Paper II** contributes in this respect to the long-standing debate of MHC-based mate choice for optimality, under natural conditions.

Second, we found substantial evidence for compatibility, *i.e.* that the best choice each female can make, is based on her own genotype (Brown 1997). This means that not only will the female have to assess the genotypes of the males, but she will have to relate them to her own. As the resulting offspring are the product of both the maternal and paternal genomes, females should choose a compatible mate to optimize offspring MHC diversity (Trivers 1972, Penn & Potts 1999), given that she is capable of fulfilling such a complex task. Mate choice based solely on male MHC diversity may render suboptimal offspring, depending on the female's genotype (Brown 1997). While most studies on the proximate mechanisms are conducted on mammals (*e.g.* Leinders-Zufall *et al.* 2004, Boehm & Zufall 2006, Sturm *et al.* 2013), our study supports the emerging view that female birds are able to discriminate individuals based on similarity through odor (Leclaire *et al.* 2017, Grieves *et al.* 2019), possibly mediated through plumage microbiota (Leclaire *et al.* 2019) or preen gland secretions (Strandh *et al.* 2012, Leclaire *et al.* 2014, Slade *et al.* 2016).

Third, our study exemplifies how extra-pair copulations could be adaptive through altered MHC constitution in offspring. As this behavior incurs costs to females (*e.g.* loss of paternal care, increased risk of sexually transmitted diseases), explanations in terms of genetic benefits are needed (Brouwer & Griffith 2019). While other studies have shown such benefits (Johnsen *et al.* 2000, Foerster *et al.* 2003), results are far from conclusive (Forstmeier *et al.* 2014). The diversity and crucial importance of MHC in adaptive immunity make MHC a possible candidate for genetic benefits of mate choice (Milinski 2006). MHC diversity has indeed been found to increase with level of promiscuity (Gohli *et al.* 2013, Winternitz *et al.* 2013). However, even though there are indications of an association between MHC diversity and female choice of EPM (Freeman-Gallant *et al.* 2003, Richardson *et al.* 2005, Promerová *et al.* 2011, Winternitz *et al.* 2015), no clear answers have been provided by comparing within-pair and extra-pair units when studying MHC and EPP in single bird species (discussed in section 1.3). Reasons of this lack of evidence could be either an absence of a pattern or a failure to reveal a true association. As sequencing technology and bioinformatic tools are ever-improving, the ability to detect actual patterns increases, and moreover, the focus on an intermediate optimum help scientists to conduct other tests than to explore linear relationships. Thus, as the field of MHC-research moves forward, it paves the way and enables a higher resolution when studying evolutionary and behavioral questions. Further, this study indicates that signs of MHC-based mate choice could be masked if EPP is disregarded, and supports the hypothesis that females could use different “rules” when choosing social and extra-pair males (Brown 1997, Mays & Hill 2004).

Fourth, our study is in stark contrast to the hypothesis put forward by Dearborn *et al.* (2016), stating that MHC-based mate choice would have diminishing returns in species with many MHC loci, as the offspring in any scenario will inherit a multi-locus genotype. We found evidence for MHC-based mate choice in a species with one of the highest numbers of MHCII loci known to date (see **paper I**), although we unveiled a pattern of stabilizing selection rather than selection for a maximal number of alleles.

In conclusion, this study is a step towards understanding both MHC-based mate choice and the adaptive benefits of extra-pair mating. While selection for intermediate dissimilarity has been found in some species (*e.g.* Bonneaud *et al.* 2006, Forsberg *et al.* 2007, Baratti *et al.* 2012), it has to my knowledge not been shown previously in an extra-pair mating context with clear, coherent results across adults and offspring. However, this study did not include fitness and so we cannot say that an intermediate number of alleles does indeed constitute an optimum. This limitation became the starting point of **paper III**.

3.3. Elevated PHA-response at intermediate MHC diversity

The aim of **paper III** was to test for a fitness benefit of having an intermediate MHCII diversity. We utilized that a PHA-response had been measured for many of the bluethroat nestlings in the dataset used in **paper II** (Johnsen *et al.* 2000), and ran the analysis on the individuals that had both a PHA-response measure and an MHCII genotype from **paper II** (using the "PSS alleles", but referring to those as "functional MHCII alleles" throughout **paper III**). The PHA-response was used as fitness measure, as the swelling response to PHA-injections is positively correlated with fitness traits such as survival and recruitment (Tella *et al.* 2002, Møller & Saino 2004, Bowers *et al.* 2014). We found that the PHA-response was negatively associated with the distance from the population mean number of functional MHCII alleles among broods. This implies an increased PHA-response in individuals with an intermediate MHCII diversity. There was also a significant effect of paternity on the PHA-response in our model. This suggests that there is an additional effect of paternity on PHA-response that we do not account for when including the MHCII variable in the model.

Our data implies that there is a fitness benefit of having an intra-individual number of functional MHCII alleles close to the population mean - shown through an increased PHA-response - and supports an intermediate optimum in intra-individual MHCII diversity (Wegner *et al.* 2003, Milinski 2006, Woelfing *et al.* 2009). However, the relationship between PHA-response and the distance from the population mean number of functional MHCII alleles is likely not direct. While the association between specific MHC alleles and PHA-response has been shown in studies in birds (Taylor jr *et al.* 1987, Bonneaud *et al.* 2005), Gaigher *et al.* (2019) argue that the effect of MHC will likely be small and connected with low power to reveal any potential correlation. The PHA mitogen causes T-cell activation mainly without presentation on MHC molecules (Coico & Sunshine 2015a), which argues against any heavy influence of specific alleles on the PHA-response we measured. Further, although we exposed the individuals to a sensitizing injection of PHA, the response of the second injection is measured only 72 hours after this initial exposure. This might not be a sufficient amount of time for an adaptive response (Vinkler *et al.* 2014), implying that other components of the immune system than MHC are important for the swelling response, and that the significant relationship we observed between PHA-response and MHCII is indirect. In addition, Johnsen *et al.* (2000) found that EPY had an increased PHA-response compared to both maternal *and* paternal half-siblings raised by the social male, lending substantial support for the "compatible genes" hypothesis. Therefore, the elevated PHA-response in EPY is likely not caused by specific alleles (Garvin *et al.* 2006), but could reflect superior quality of any kind. These arguments suggest that individuals possessing an intra-individual number of functional MHCII alleles close to the population mean may be able to invest more in a costly immune response, possibly through higher immune cell activity (Martin *et al.* 2006).

This study supports a "golden mean" in MHCII diversity (Woelfing *et al.* 2009); *i.e.* that there is a fitness benefit of having an intermediate number of functional MHCII alleles (Wegner *et al.* 2003, Milinski 2006). An intermediate optimum in MHC diversity has previously been shown in studies on parasite resistance (Wegner *et al.* 2003, Madsen & Ujvari 2006, Kloch *et al.* 2010) and reproductive traits (Jäger *et al.* 2007, Kalbe *et al.* 2009), but to my knowledge, an intermediate MHC diversity has not formerly been associated with an increased PHA-response. With the notable exception of the study by Bonneaud *et al.* (2005), most studies on MHC diversity and PHA-response in wild species (see Gaigher *et al.* 2019) have, however, primarily tested linear associations, and significant effects of intermediate MHC diversity may have passed unnoticed. This paper thus emphasizes the importance of analyzing data also considering an intermediate optimum in order to reveal potential effects of MHC diversity. Using the methodology described herein, we were able to demonstrate an association between PHA-response and MHCII diversity using a total of samples well below the "required" sample size for detecting any effect of MHC on immunocompetence as stated by Gaigher *et al.* (2019).

This study is based on a subset of the dataset used in Johnsen *et al.* (2000) and Fossøy *et al.* (2008), who found that EPY have a higher PHA-response than their within-pair half-siblings. In **paper II**, we showed that the number of functional MHCII alleles in EPY is closer to the population mean, compared to WPY. The association between PHA-response and the distance from population mean number of MHCII alleles is thus as expected based on these studies. Combining the results from **paper II** and **paper III**, we find indications of a fitness benefit of extra-pair copulations, mediated through an intermediate MHCII diversity. Although in line with the "optimal MHC hypothesis" (see above), this deviates from many studies on genetic similarity and extra-pair mate choice in birds (see review by Brouwer and Griffith 2019). For instance, Fossøy *et al.* (2008) found that females were less similar to the EPM than the WPM on neutral genetic markers. Moreover, the studies of Fossøy *et al.* (2009) and Arct *et al.* (2019) both implied an increased PHA-response in offspring that have genetically dissimilar parents. In our study, an elevated PHA-response is correlated with *intermediate* MHC diversity; a pattern likely not observed if parents are highly dissimilar at MHC. Considering that the dataset used in our study overlaps to some degree with that used by Fossøy *et al.* (2008) and Fossøy *et al.* (2009), the effect we observed is likely not a by-product of effects observed at neutral markers. Nevertheless, paternity was significantly associated with PHA-response in our model, implying that we can partly, but not fully, explain this fitness benefit of extra-pair mating (*i.e.* the increased PHA-response) through MHC diversity. Although extra-pair mate choice is complex and several aspects of MHC-based mate choice remain unexplained, this study suggests a potential fitness benefit of extra-pair copulations through MHCII-based female choice.

3.4. MHCI and extra-pair mating in the willow warbler

In the last paper of this thesis (**paper IV**), MHCI-based extra-pair mate choice was explored in the willow warbler. We did not find any evidence of a general female extra-pair mate choice based on compatibility or male diversity at MHCI. Additionally, we found no support for the hypothesis that females socially paired to a suboptimal male will be more likely to partake in extra-pair mating to improve the MHCI constitution of their offspring by "correcting" for their suboptimal social mate choice (Petrie & Kempenaers 1998, Freeman-Gallant *et al.* 2003). Rather, the existence of two different female strategies was indicated, in which females with a low number of supertypes are more likely to have EPY. This is suggested by the result that females with EPY in their brood had fewer supertypes than genetically monogamous females, and a tendency for the number of supertypes in the female to be negatively correlated with the proportion of EPY in the brood. Additionally, there was a non-random distribution of EPY among broods; significantly more broods than expected contained either none or many EPY (see also Bjørnstad and Lifjeld 1997).

We had predicted that females paired to a suboptimal male will be more likely to have EPY, but found contrarily that social pairs with broods containing one or more EPY were significantly more concentrated around an intermediate number of unique supertypes in the pair, than both genetically monogamous social pairs and random pairs were. Although the tests on mate choice did not remain significant after controlling for multiple testing, implications of the results are nevertheless discussed.

The lack of any difference in MHCI diversity between EPY and WPY in broods with mixed paternity suggests that there was no difference in MHCI diversity between the WPM and EPM for the females with EPY in their broods affecting her choice. As the EPM were not included in the dataset, this could however not be tested directly in this study. Moreover, the lack of any difference in MHCI diversity between WPY from pure WPY broods and WPY from broods also containing EPY could imply that the choice of mates is indeed correcting for the smaller number of supertypes in females that are engaging in extra-pair copulations. One possibility is therefore that only females with few supertypes conduct MHC-based mate choice, and that they prefer social and extra-pair mates with many supertypes.

A study by Griggio *et al.* (2011) found that female house sparrows with few MHCI alleles preferred males with many alleles, while females with an intermediate or high number of alleles exhibited no such preference. Our results are in line with this pattern, suggesting that females with few MHCI supertypes conduct MHCI-based mate choice and also are more likely to engage in extra-pair copulations. Extra-pair mating could be costly, for instance due to loss of paternal care from the social male (Brouwer & Griffith 2019), and for females with a "sufficient" number of supertypes, any gains from extra-pair mating may not be enough to outweigh these costs. Conversely, another explanation could be that females with many supertypes are in better condition, and hence able to resist forced extra-pair mating attempts (Westneat & Stewart 2003).

We did not find any support for mate choice based on an intermediate number of PSS alleles in this study, unlike the results from **paper II**. Our results rather imply that it is beneficial to possess many supertypes. This could be explained by the hypothesis of divergent allele advantage (Wakeland *et al.* 1990, Lenz 2011), where individuals with divergent alleles have a selective advantage by being able to bind a broader range of antigens. Also, individuals with fewer supertypes will be less likely to possess specific supertypes. Studies on great tits (*Parus major*) have indeed found an association between specific supertypes and increased resistance against *Plasmodium* infections (Sepil *et al.* 2013a) as well as lifetime reproductive success and survival (Sepil *et al.* 2013b). This indicates that females with few supertypes will have the most to gain by engaging in extra-pair mating.

A lack of a consistent pattern in this study and **paper II** - where we did find evidence of MHCII-based extra-pair mate choice - could be caused by differences in the MHC classes and/or between the species. A likely source of differences between the MHC classes could be differences in their correlation with perceptible phenotypes. Preen oil is highly variable and is suggested to be a main source of chemical signaling in birds (Whittaker *et al.* 2010). As MHCII is important in the fight against extracellular bacteria while MHCI present intracellular pathogens to T cells, MHC-based discrimination through differences in the chemical properties of preen oil might only be possible via MHCII and not MHCI (Strandh *et al.* 2012). This could be due to differences in microbial communities in the preen gland associated with differences in MHCII genotypes (Strandh *et al.* 2012), which ultimately may alter the wax esters in the preen oil (Grieves *et al.* 2019). A correlation between MHCII and the chemical composition of preen oil is detected in both Charadriiformes and passerine species (Leclaire *et al.* 2014, Slade *et al.* 2016). Grieves *et al.* (2019) recently demonstrated that song sparrows (*Melospiza melodia*) are able to distinguish between preen oil from MHCII-similar and MHCII-dissimilar individuals, and from individuals with high and low MHCII diversity. For MHCI, ligands have been shown to be important in chemical signals in mammals and fish (*e.g.* Leinders-Zufall *et al.* 2004, Milinski *et al.* 2005), however, to my knowledge, no mechanisms have been demonstrated in birds.

There are also differences between the species (*i.e.* the bluethroat and the willow warbler) that possibly could be associated with differences in MHC-based mate choice as implied from **paper II** and **paper IV**. The bluethroat is a sexually dimorphic species with an extensive courtship where the males display their throat patch close to the female (Peiponen 1960, Johnsen & Lifjeld 1995). Of the two classical MHC classes, only the MHCII is hypervariable in the bluethroat, while MHCI shows a moderate diversity (see **paper I**). The willow warbler, on the other hand, shows extreme diversity at both markers (Westerdahl *et al.* 2000, O'Connor *et al.* 2016). Their courtship display does likely not occur in close proximity as in the bluethroat, and the females often solicit copulations (Arvidsson 1992, Bairlein 2006). The greater distance between males and females during courtship may impact the opportunity of conducting mate choice based on olfactory cues, *e.g.* based on MHC diversity reflected in preen gland secretions (see above and section 1.4), or it may reflect a lack of such mate choice. The species do, however, exhibit similarities in life history that might imply that they are exposed to a similar pathogen burden, which could affect the importance of an optimized MHC constitution. For instance, in our study site in Heimdalen, both species are highly migratory, show similar rates of EPP and are both insectivores (see section 2.1.). Alternatively, if the main purpose of MHC-based mate choice is inbreeding avoidance, then MHC-disassortative mate preferences should be stronger in species in which the risk of inbreeding is larger (Yamazaki *et al.* 1976, Potts *et al.* 1994, Penn & Potts 1999, Reusch *et al.* 2001). However, willow warblers are assumed not to be inbred (Jan Lifjeld, pers. comm.), and neither the Heimdalen bluethroat population runs a great risk of inbreeding, due to the low level of philopatry (Johnsen and Lifjeld, unpublished data). Associations between the abovementioned factors in the species' ecology and differences in MHC-based mate choice remain to be explored, and could provide a rewarding avenue for future research.

As we did not observe extra-pair mate choice based on MHCI in the willow warbler in line with the findings in **paper II**, this study intriguingly demonstrates an absence of consistency in MHC-based mate choice in MHC classes in two passerine species. This highlights the need for conducting similar studies in more species, in a comparative framework, in order to reveal any pattern of such mate choice and possible influencing factors. Nevertheless, this study could help shed light on one of the main unanswered questions regarding extra-pair paternity in birds, namely, why some but not all females have EPY (Brouwer & Griffith 2019).

4. Concluding remarks and future directions

Despite immense development in sequencing technologies and decades of research, many questions concerning MHC-based mate choice and extra-pair mating remain unanswered. This thesis has contributed to increasing our knowledge in these fields. First, we revealed the difficulties and discrepancies arising from different MHC genotyping strategies, emphasizing the need for careful planning and designing of sequencing and allele calling pipelines. While we were still unable to obtain true genotypes with absolute confidence, we were able to achieve an approximation for informative analyses. New methodologies have greatly improved the genotyping accuracy of highly polymorphic markers, like for the bluethroat MHCII and willow warbler MHCI. This enabled a closer examination of MHC-based mate choice than was previously possible in these hypervariable systems. By applying the established genotyping approach on large bluethroat and willow warbler datasets of adults and nestlings with known paternity, the results in this thesis highlight two different patterns of MHC-based extra-pair mate choice. For the bluethroat, we found support for non-random female choice of EPM based on compatibility and an intermediate optimum in MHCII diversity. For the willow warbler, two different female mating strategies were implied, where females with few MHCI supertypes were more likely to engage in extra-pair mating. This sheds light on the variation in patterns in MHC-based mate choice across species and MHC classes. Future studies should aim to further explore the causes of this variation in female mate choice strategy and the variation in the number of MHCI and MHCII loci among species. For instance, it would be interesting to conduct comparative studies testing for associations between pathogen exposure, MHC diversity, mating strategies and MHC-based extra-pair mate choice across passerine birds (see Minias *et al.* 2017, Whittingham *et al.* 2018 and Lindsay *et al.* 2019).

Logical extensions to the projects in this thesis could be inclusion of genetically monogamous social pairs in the bluethroat MHCII dataset and EPM in the willow warbler MHCI dataset. Additionally, mate choice based on bluethroat MHCI and willow warbler MHCII could be explored and compared to **paper II** and **paper IV**, to highlight species differences within MHC classes.

Establishing the proximate mechanisms is a crucial next step to more fully understanding MHC-based mate choice. There is continuous advancement on this matter also in passerine birds, *e.g.* by demonstrating that song sparrows can discriminate MHCII diversity and similarity through preen oil odor (Grieves *et al.* 2019). Considering the evidence in this thesis for MHC-based female mate choice in bluethroats, it would be interesting to characterize preen oil composition in this species and test for correlations with MHCII diversity, as well as to test for associations between its chemical distance and MHCII dissimilarity in male-female dyads (*e.g.* Leclaire *et al.* 2014, Slade *et al.* 2016). Similarities in preen secretions and microbiomes could also be compared between social pairs and extra-pair partners, with the prediction that the distribution will have a smaller variance in extra-pair partners than in social pairs. Furthermore, applying proteomic techniques on sperm cells isolated from the ejaculate could reveal whether MHC molecules are expressed on these cells, facilitating post-copulatory MHC-based female choice of compatible mates (Lenz *et al.* 2018).

Regardless of the mechanism, two steps in MHC-based mate choice are predicted (Aeschlimann *et al.* 2003): female choice for optimal MHC diversity, and assessment of male secondary sexual characters to reveal health and thus resistance alleles against prevailing pathogens. This thesis has not focused on specific alleles with an additive advantage in a “good gene” framework (*e.g.* Hamilton & Zuk 1982, Ekblom *et al.* 2004, Eizaguirre *et al.* 2009b), and many exciting projects are possible in this respect. For instance, it would be interesting to screen for common parasites in bluethroats and willow warblers (*e.g. Plasmodium* and *Haemoproteus*; Bensch & Åkesson 2003, Svoboda *et al.* 2015) and test for associations between infection status and specific MHC alleles or supertypes (*e.g.* Sepil *et al.* 2013a). This could be related to an extra-pair framework, testing for differences between WPM and EPM, or the probability of being cuckolded (see for instance Podmokła *et al.* 2015). Moreover, parasitic load could be used as a fitness measure, testing the hypothesis that individuals possessing an intermediate number of alleles will have the greatest resistance (Wegner *et al.* 2003).

The current era of genomics holds great potential for enhancing our knowledge of MHC and its evolution. Long-read sequencing provides a promising path forward, and could prove especially helpful in the characterization of haplotypes and linkage within the MHC region (reviewed by O’Connor *et al.* 2019). Development of locus-specific primers could resolve the long-awaited locus-specificity of alleles in non-model species (Canal *et al.* 2010). Together, these advances could move the field of avian MHC research towards identifying "true" genotypes and heterozygosity, which again will facilitate the use of MHC markers in evolutionary and ecological studies. Also, expression analyses and increased knowledge of the interaction between antigens and MHC molecules could lead to new insights, particularly into the relationship between fitness traits and MHC (Drews & Westerdahl 2019, O'Connor *et al.* 2019). Importantly, more specific insights into these aspects could help to improve our understanding of how selection acts on MHC, moving beyond approximations like "PSS" and the supertypes utilized in this thesis.

References

Aeschlimann P, Häberli M, Reusch T, Boehm T and Milinski M (2003). Female sticklebacks *Gasterosteus aculeatus* use self-reference to optimize MHC allele number during mate selection. *Behavioral Ecology and Sociobiology* **54**(2): 119-126.

Alcaide M, Rodríguez A, Negro JJ and Serrano D (2012). Male transmission ratio distortion supports MHC-linked cryptic female choice in the lesser kestrel (Aves: Falconidae). *Behavioral Ecology and Sociobiology* **66**(11): 1467-1473.

Andersson MB (1994). Sexual selection. 1st edition. West Sussex, Princeton University Press.

Anmarkrud J, Johnsen A, Bachmann L and Lifjeld J (2010). Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes. *Journal of Evolutionary Biology* **23**(6): 1206-1217.

Arct A, Drobniak SM, Mellinger S, Gustafsson L and Cichoń M (2019). Parental genetic similarity and offspring performance in blue tits in relation to brood size manipulation. *Ecology and Evolution* **9**(18): 10085-10091.

Arden B and Klein J (1982). Biochemical comparison of major histocompatibility complex molecules from different subspecies of *Mus musculus*: Evidence for trans-specific evolution of alleles. *Proceedings of the National Academy of Sciences* **79**(7): 2342-2346.

Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J and Kourilsky P (1999). A direct estimate of the human αβ T cell receptor diversity. *Science* **286**(5441): 958-961.

Arvidsson BL (1992). Copulation and mate guarding in the willow warbler. *Animal Behaviour* **43**(3): 501-509.

Audet J-N, Ducatez S and Lefebvre L (2015). The town bird and the country bird: Problem solving and immunocompetence vary with urbanization. *Behavioral Ecology* **27**(2): 637-644.

Babik W, Taberlet P, Ejsmond MJ and Radwan J (2009). New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources* **9**(3): 713-719.

Bairlein F (2006). Willow Warbler (*Phylloscopus trochilus*). Family Sylviidae (Old World Warblers). In: *Handbook of the Birds of the World.* Ed.: Del Hoyo J, Elliott A and Christie D. Barcelona, Lynx Edicions. Volume 11. Old World Flcatchers to Old World Warblers**:** 649.

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

Baratti M, Dessi-Fulgheri F, Ambrosini R, Bonisoli-Alquati A, Caprioli M, Goti E, Matteo A, Monnanni R, Ragionieri L and Ristori E (2012). MHC genotype predicts mate choice in the ring-necked pheasant *Phasianus colchicus*. *Journal of Evolutionary Biology* **25**(8): 1531-1542.

Bensch S, Andersson T and Akesson S (1999). Morphological and molecular variation across a migratory divide in willow warblers, *Phylloscopus trochilus*. *Evolution* **53**(6): 1925-1935.

Bensch S and Åkesson S (2003). Temporal and spatial variation of hematozoans in Scandinavian willow warblers. *Journal of Parasitology* **89**(2): 388-392.

Bensch S, Bengtsson G and Åkesson S (2006). Patterns of stable isotope signatures in willow warbler *Phylloscopus trochilus* feathers collected in Africa. *Journal of Avian Biology* **37**(4): 323-330.

Bensch S, Grahn M, Müller N, Gay L and Åkesson S (2009). Genetic, morphological, and feather isotope variation of migratory willow warblers show gradual divergence in a ring. *Molecular Ecology* **18**(14): 3087-3096.

Biedrzycka A, O'connor E, Sebastian A, Migalska M, Radwan J, Zając T, Bielański W, Solarz W, Ćmiel A and Westerdahl H (2017). Extreme MHC class I diversity in the sedge warbler (*Acrocephalus schoenobaenus*); Selection patterns and allelic divergence suggest that different genes have different functions. *BMC Evolutionary Biology* **17**(1): 159.

Birkhead TR and Møller AP (1992). Avian mating systems and sperm competition. Chapter 2 in: *Sperm competition in birds: evolutionary causes and consequences.* 1st edition. Ed.: Birkhead TR and Møller AP. London, Academic Press Limited**:** 17-42.

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL and Wiley DC (1987a). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**(6139): 506-512.

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL and Wiley DC (1987b). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329**(6139): 512-518.

Bjørnstad G and Lifjeld JT (1997). High frequency of extra-pair paternity in a dense and synchronous population of willow warblers *Phylloscopus trochilus*. *Journal of Avian Biology* **28**(4): 319-324.

Blum JS, Wearsch PA and Cresswell P (2013). Pathways of antigen processing. *Annual Review of Immunology* **31**: 443-473.

Bodmer WF (1972). Evolutionary significance of the HL-A system. *Nature* **237**(5351): 139-183.

Boehm T and Zufall F (2006). MHC peptides and the sensory evaluation of genotype. *Trends in Neurosciences* **29**(2): 100-107.

Bollmer JL, Dunn PO, Freeman-Gallant CR and Whittingham LA (2012). Social and extra-pair mating in relation to major histocompatibility complex variation in common yellowthroats. *Proceedings of the Royal Society of London B: Biological Sciences* **279**: 4778-4785.

Bonneaud C, Richard M, Faivre B, Westerdahl H and Sorci G (2005). An Mhc class I allele associated to the expression of T-dependent immune response in the house sparrow. *Immunogenetics* **57**(10): 782-789.

Bonneaud C, Chastel O, Federici P, Westerdahl H and Sorci G (2006). Complex *Mhc*-based mate choice in a wild passerine. *Proceedings of the Royal Society B: Biological Sciences* **273**(1590): 1111-1116.

Borghans JA, Noest AJ and De Boer RJ (2003). Thymic selection does not limit the individual MHC diversity. *European Journal of Immunology* **33**(12): 3353-3358.

Bowers EK, Hodges CJ, Forsman AM, Vogel LA, Masters BS, Johnson BG, Johnson LS, Thompson CF and Sakaluk SK (2014). Neonatal body condition, immune responsiveness, and hematocrit predict longevity in a wild bird population. *Ecology* **95**(11): 3027-3034.

Briles WE, Stone HA and Cole R (1977). Marek's disease: Effects of B histocompatibility alloalleles in resistant and susceptible chicken lines. *Science* **195**(4274): 193-195.

Brouwer L, Barr I, Van De Pol M, Burke T, Komdeur J and Richardson DS (2010). MHC-dependent survival in a wild population: Evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology* **19**(16): 3444-3455.

Brouwer L and Griffith SC (2019). Extra-pair paternity in birds. *Molecular Ecology* **28**(11): 4864-4882.

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL and Wiley DC (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**(6432): 33-39.

Brown JL (1997). A theory of mate choice based on heterozygosity. *Behavioral Ecology* **8**(1): 60-65.

Canal D, Alcaide M, Anmarkrud JA and Potti J (2010). Towards the simplification of MHC typing protocols: Targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*. *BMC Research Notes* **3**(1): 236.

Caro SP, Balthazart J and Bonadonna F (2015). The perfume of reproduction in birds: Chemosignaling in avian social life. *Hormones and Behavior* **68**: 25-42.

Carroll L (1871). Through the looking-glass, and what Alice found there. 1st edition 2009. Mayo, Evertype.

Chamberlain C, Bensch S, Feng X, Åkesson S and Andersson T (2000). Stable isotopes examined across a migratory divide in Scandinavian willow warblers (*Phylloscopus trochilus trochilus* and *Phylloscopus trochilus acredula*) reflect their African winter quarters. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267**(1438): 43-48.

Coico R and Sunshine G (2015a). Activation and function of T cells. Chapter 11 in: *Immunology: A short course.* 7th edition. Ed.: Coico R and Sunshine G. West Sussex, John Wiley & Sons.

Coico R and Sunshine G (2015b). How T cells recognize antigen: The role of the major histocompatibility complex. Chapter 9 in: *Immunology: A short course.* 7th edition. Ed.: Coico R and Sunshine G. West Sussex, John Wiley & Sons.

Coico R and Sunshine G (2015c). Biology of the T lymphocyte. Chapter 10 in: *Immunology: A short course.* 7th edition. Ed.: Coico R and Sunshine G. West Sussex, John Wiley & Sons.

Collar NJ (2005). Bluethroat (*Luscinia svecica).* Family Turdidae (Thrushes). In: *Handbook of the Birds of the World.* Ed.: Del Hoyo J, Elliott A and Christie D. Barcelona, Lynx Edicions. Vol 10. Cuckoo-shrikes to Thrushes**:** 748-749.

Dearborn DC, Gager AB, Mcarthur AG, Gilmour ME, Mandzhukova E and Mauck RA (2016). Gene duplication and divergence produce divergent MHC genotypes without disassortative mating. *Molecular Ecology* **25**(17): 4355-4367.

Ditchkoff SS, Lochmiller RL, Masters RE, Hoofer SR and Bussche RaVD (2001). Major-histocompatibility-complex-associated variation in secondary sexual traits of white-tailed deer (*Odocoileus virginianus*): Evidence for good-genes advertisement. *Evolution* **55**(3): 616-625.

Doherty PC and Zinkernagel RM (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**: 50-52.

Doytchinova IA and Flower DR (2005). In silico identification of supertypes for class II MHCs. *The Journal of Immunology* **174**(11): 7085-7095.

Drews A and Westerdahl H (2019). Not all birds have a single dominantly expressed MHC-I gene: Transcription suggests that siskins have many highly expressed MHC-I genes. *Scientific Reports* **9**(1): 1-11.

Dunn PO, Bollmer JL, Freeman-Gallant CR and Whittingham LA (2013). MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats. *Evolution* **67**(3): 679-687.

Ebert D and Hamilton WD (1996). Sex against virulence: The coevolution of parasitic diseases. *Trends in Ecology & Evolution* **11**(2): 79-82.

Edgar RC, Haas BJ, Clemente JC, Quince C and Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**(16): 2194-2200.

Edwards SV and Hedrick PW (1998). Evolution and ecology of MHC molecules: From genomics to sexual selection. *Trends in Ecology & Evolution* **13**(8): 305-311.

Egid K and Brown JL (1989). The major histocompatibility complex and female mating preferences in mice. *Animal Behaviour* **38**: 548-550.

Eimes J, Reed K, Mendoza K, Bollmer J, Whittingham L, Bateson Z and Dunn P (2013). Greater prairie chickens have a compact MHC-B with a single class IA locus. *Immunogenetics* **65**(2): 133-144.

Eizaguirre C, Lenz TL, Traulsen A and Milinski M (2009a). Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. *Ecology Letters* **12**(1): 5-12.

Eizaguirre C, Yeates SE, Lenz TL, Kalbe M and Milinski M (2009b). MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology* **18**(15): 3316-3329.

Eizaguirre C, Lenz TL, Kalbe M and Milinski M (2012). Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nature Communications* **3**: 621.

Ekblom R, Saether SA, Grahn M, Fiske P, Kålås JA and Höglund J (2004). Major histocompatibility complex variation and mate choice in a lekking bird, the great snipe (*Gallinago media*). *Molecular Ecology* **13**(12): 3821-3828.

Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM and Ravel J (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**(1): 6.

Figueroa F, Gúnther E and Klein J (1988). MHC polymorphism pre-dating speciation. *Nature* **335**(6187): 265-267.

Foerster K, Delhey K, Johnsen A, Lifjeld JT and Kempenaers B (2003). Females increase offspring heterozygosity and fitness through extra-pair matings. *Nature* **425**(6959): 714-717.

Folstad I and Karter AJ (1992). Parasites, bright males, and the immunocompetence handicap. *The American Naturalist* **139**(3): 603-622.

Forsberg LA, Dannewitz J, Petersson E and 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**(5): 1859-1869.

Forstmeier W, Nakagawa S, Griffith SC and Kempenaers B (2014). Female extra-pair mating: Adaptation or genetic constraint? *Trends in Ecology & Evolution***29**(8): 456-464.

Fossøy F, Johnsen A and Lifjeld JT (2008). Multiple genetic benefits of female promiscuity in a socially monogamous passerine. *Evolution* **62**(1): 145-156.

Fossøy F, Johnsen A and Lifjeld J (2009). Cell-mediated immunity and multi-locus heterozygosity in bluethroat nestlings. *Journal of Evolutionary Biology* **22**(9): 1954-1960.

Freeman-Gallant CR, Johnson EM, Saponara F and Stanger M (2002). Variation at the major histocompatibility complex in Savannah sparrows. *Molecular Ecology* **11**(6): 1125-1130.

Freeman-Gallant CR, Meguerdichian M, Wheelwright NT and 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**(11): 3077-3083.

Gaigher A, Burri R, San-Jose LM, Roulin A and Fumagalli L (2019). Lack of statistical power as a major limitation in understanding MHC-mediated immunocompetence in wild vertebrate populations. *Molecular Ecology*. **28**(23): 5115-5132.

Galan M, Guivier E, Caraux G, Charbonnel N and 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**(1): 296.

Gangoso L, Alcaide M, Grande JM, Muñoz J, Talbot SL, Sonsthagen SA, Sage G and Figuerola J (2012). Colonizing the world in spite of reduced MHC variation. *Journal of Evolutionary Biology* **25**(7): 1438-1447.

Garvin JC, Abroe B, Pedersen MC, Dunn PO and Whittingham LA (2006). Immune response of nestling warblers varies with extra-pair paternity and temperature. *Molecular Ecology* **15**(12): 3833-3840.

Gohli J, Anmarkrud JA, Johnsen A, Kleven O, Borge T and Lifjeld JT (2013). Female promiscuity is positively associated with neutral and selected genetic diversity in passerine birds. *Evolution* **67**(5): 1406-1419.

Goto N, Kodama H, Okada K and Fujimoto Y (1978). Suppression of phytohemagglutinin skin response in thymectomized chickens. *Poultry Science* **57**(1): 246-250.

Gough SCL and Simmonds MJ (2007). The HLA region and autoimmune disease: Associations and mechanisms of action. *Current Genomics* **8**(7): 453-465.

Greaves M, Roitt I and Rose ME (1968). Effect of bursectomy and thymectomy on the responses of chicken peripheral blood lymphocytes to phytohaemagglutinin. *Nature* **220**(5164): 293-295.

Grieves L, Gloor G, Bernards M and Macdougall-Shackleton E (2019). Songbirds show odour-based discrimination of similarity and diversity at the major histocompatibility complex. *Animal Behaviour* **158**: 131-138.

Griggio M, Biard C, Penn DJ and Hoi H (2011). Female house sparrows "count on" male genes: Experimental evidence for MHC-dependent mate preference in birds. *BMC Evolutionary Biology* **11**(1): 44.

Hamilton WD and Zuk M (1982). Heritable true fitness and bright birds: A role for parasites? *Science* **218**(4570): 384-387.

Hasselquist D (2007). Comparative immunoecology in birds: Hypotheses and tests. *Journal of Ornithology* **148**(2): 571-582.

Hedenström A and Petterson J (1987). Migration routes and wintering areas of willow warblers *Phylloscopus trochilus*(L.) ringed in Fennoscandia. *Ornis Fennica* **64**(4): 137-143.

Hedrick PW, Thomson G and Klitz W (1987). Evolutionary genetics and HLA: Another classic example. *Biological Journal of the Linnean Society* **31**(4): 311-331.

Hedrick PW (1999). Balancing selection and MHC. *Genetica* **104**(3): 207-214.

Hedrick PW (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**(10): 1902-1908.

Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, Mcmichael AJ and Greenwood BM (1991). Common West African HLA antigens are associated with protection from severe malaria. *Nature* **352**(6336): 595-600.

Hughes AL and Nei M (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**(6186): 167-170.

Hughes AL and Nei M (1989). Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proceedings of the National Academy of Sciences* **86**(3): 958-962.

Janeway CA, Travers P, Walport M and Shlomchik MJ (2001a). Antigen Presentation to T Lymphocytes. Chapter 5 in: *Immunobiology: The Immune System in Health and Disease.* 5th edition. Ed.: Janeway CA, Travers P, Walport M and Shlomchik MJ. New York, Garland Science.

Janeway CA, Travers P, Walport M and Shlomchik MJ (2001b). The Generation of Lymphocyte Antigen Receptors. Chapter 4 in: *Immunobiology: The Immune System in Health and Disease.* 5th edition. Ed.: Janeway CA, Travers P, Walport M and Shlomchik MJ. New York, Garland Science.

Johnsen A and Lifjeld JT (1995). Unattractive males guard their mates more closely: An experiment with bluethroats (Aves, Turdidae: *Luscinia s. svecica*). *Ethology* **101**(3): 200-212.

Johnsen A, Andersen V, Sunding C and Lifjeld JT (2000). Female bluethroats enhance offspring immunocompetence through extra-pair copulations. *Nature* **406**(6793): 296-299.

Johnsen A and Lifjeld JT (2003). Ecological constraints on extra-pair paternity in the bluethroat. *Oecologia* **136**(3): 476-483.

Johnsen A, Andersson S, Fernandez JG, Kempenaers B, Pavel V, Questiau S, Raess M, Rindal E and Lifjeld JT (2006). Molecular and phenotypic divergence in the bluethroat (*Luscinia svecica*) subspecies complex. *Molecular Ecology* **15**(13): 4033-4047.

Jombart T (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**(11): 1403-1405.

Jombart T, Devillard S and Balloux F (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics* **11**(1): 94.

Jäger I, Eizaguirre C, Griffiths S, Kalbe M, Krobbach C, Reusch T, Schaschl H and Milinski M (2007). Individual MHC class I and MHC class IIB diversities are associated with male and female reproductive traits in the three-spined stickleback. *Journal of Evolutionary Biology* **20**(5): 2005-2015.

Kalbe M, Eizaguirre C, Dankert I, Reusch TB, Sommerfeld RD, Wegner KM and Milinski M (2009). Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings of the Royal Society B: Biological Sciences* **276**(1658): 925-934.

Kamiya T, O'dwyer K, Westerdahl H, Senior A and Nakagawa S (2014). A quantitative review of MHC-based mating preference: The role of diversity and dissimilarity. *Molecular Ecology* **23**(21): 5151-5163.

Kappler JW, Roehm N and Marrack P (1987). T cell tolerance by clonal elimination in the thymus. *Cell* **49**(2): 273-280.

Kaufman J, Milne S, Göbel TW, Walker BA, Jacob JP, Auffray C, Zoorob R and Beck S (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **401**(6756): 923-925.

Kempenaers B, Verheyen GR, Van Den Broeck M, Burke T, Van Broeckhoven C and Dhondt A (1992). Extra-pair paternity results from female preference for high-quality males in the blue tit. *Nature* **357**(6378): 494.

Kisielow P, Teh HS, Blüthmann H and Von Boehmer H (1988). Positive selection of antigen-specific T cells in thymus by restricting MHC molecules. *Nature* **335**(6192): 730-733.

Klein J (1987). Origin of major histocompatibility complexes polymorphism: The trans-species hypothesis. *Human Immunology* **19**(3): 155-162.

Kloch A, Babik W, Bajer A, Siński E and 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**: 255-265.

Krokene C, Anthonisen K, Lifjeld JT and Amundsen T (1996). Paternity and paternity assurance behaviour in the bluethroat, *Luscinia s. svecica*. *Animal Behaviour* **52**(2): 405-417.

Landry C, Garant D, Duchesne P and Bernatchez L (2001). 'Good genes as heterozygosity': The major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**(1473): 1279-1285.

Leclaire S, Van Dongen WF, Voccia S, Merkling T, Ducamp C, Hatch SA, Blanchard P, Danchin É and Wagner RH (2014). Preen secretions encode information on MHC similarity in certain sex-dyads in a monogamous seabird. *Scientific reports* **4**: 6920.

Leclaire S, Strandh M, Mardon J, Westerdahl H and Bonadonna F (2017). Odour-based discrimination of similarity at the major histocompatibility complex in birds. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **284**(1846): 20162466.

Leclaire S, Strandh M, Dell'ariccia G, Gabirot M, Westerdahl H and Bonadonna F (2019). Plumage microbiota covaries with the major histocompatibility complex in blue petrels. *Molecular Ecology* **28**(4): 833-846.

Leinders-Zufall T, Brennan P, Widmayer P, Maul-Pavicic A, Jäger M, Li X-H, Breer H, Zufall F and Boehm T (2004). MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science* **306**(5698): 1033-1037.

Lenz TL (2011). Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution: International Journal of Organic Evolution* **65**(8): 2380-2390.

Lenz TL, Deutsch AJ, Han B, Hu X, Okada Y, Eyre S, Knapp M, Zhernakova A, Huizinga TW and Abecasis G (2015). Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. *Nature Genetics* **47**(9): 1085-1090.

Lenz TL, Hafer N, Samonte IE, Yeates SE and Milinski M (2018). Cryptic haplotype-specific gamete selection yields offspring with optimal MHC immune genes. *Evolution* **72**(11): 2478-2490.

Liedvogel M, Larson KW, Lundberg M, Gursoy A, Wassenaar LI, Hobson KA, Bensch S and Åkesson S (2014). No evidence for assortative mating within a willow warbler migratory divide. *Frontiers in Zoology* **11**(1): 52.

Lighten J, Oosterhout C, Paterson IG, Mcmullan M and Bentzen P (2014). Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). *Molecular Ecology Resources* **14**(4): 753-767.

Lindsay WR, Andersson S, Bererhi B, Höglund J, Johnsen A, Kvarnemo C, Leder EH, Lifjeld JT, Ninnes CE, Olsson M, Parker GA, Pizzari T, Qvarnström A, Safran RJ, Svensson O and Edwards SV (2019). Endless forms of sexual selection. *PeerJ* **7**: e7988.

Lively CM and Dybdahl MF (2000). Parasite adaptation to locally common host genotypes. *Nature* **405**(6787): 679-681.

Løvlie H, Gillingham MA, Worley K, Pizzari T and Richardson DS (2013). Cryptic female choice favours sperm from major histocompatibility complex-dissimilar males. *Proceedings of the Royal Society B: Biological Sciences* **280**(1769): 20131296.

Madsen T and Ujvari B (2006). MHC class I variation associates with parasite resistance and longevity in tropical pythons. *Journal of Evolutionary Biology* **19**(6): 1973-1978.

Malmstrøm M, Matschiner M, Tørresen OK, Star B, Snipen LG, Hansen TF, Baalsrud HT, Nederbragt AJ, Hanel R and Salzburger W (2016). Evolution of the immune system influences speciation rates in teleost fishes. *Nature Genetics* **48**: 1204-1210.

Martin LB, Han P, Lewittes J, Kuhlman JR, Klasing KC and Wikelski M (2006). Phytohemagglutinin-induced skin swelling in birds: Histological support for a classic immunoecological technique. *Functional Ecology* **20**(2): 290-299.

Matsumura M, Fremont DH, Peterson PA and Wilson IA (1992). Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science* **257**(5072): 927-934.

Mays JHL and Hill GE (2004). Choosing mates: Good genes versus genes that are a good fit. *Trends in Ecology & Evolution* **19**(10): 554-559.

Mccorkle F, Olah I and Glick B (1980). The morphology of the phytohemagglutinin-induced cell response in the chicken's wattle. *Poultry Science* **59**(3): 616-623.

Migalska M, Sebastian A and Radwan J (2019). Major histocompatibility complex class I diversity limits the repertoire of T cell receptors. *Proceedings of the National Academy of Sciences* **116**(11): 5021-5026.

Milinski M, Griffiths S, Wegner KM, Reusch TB, Haas-Assenbaum A and Boehm T (2005). Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proceedings of the National Academy of Sciences* **102**(12): 4414-4418.

Milinski M (2006). The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology, Evolution, and Systematics* **37**: 159-186.

Minias P, Whittingham LA and Dunn PO (2017). Coloniality and migration are related to selection on MHC genes in birds. *Evolution* **71**(2): 432-441.

Minias P, Pikus E, Whittingham LA and Dunn PO (2018). Evolution of copy number at the MHC varies across the avian tree of life. *Genome Biology and Evolution***11**(1): 17-28.

Møller AP and Saino N (2004). Immune response and survival. *Oikos* **104**(2): 299-304.

Neff BD and Pitcher TE (2005). Genetic quality and sexual selection: An integrated framework for good genes and compatible genes. *Molecular Ecology* **14**(1): 19-38.

Nei M, Gu X and 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* **94**(15): 7799-7806.

Nowak MA, Tarczy-Hornoch K and Austyn JM (1992). The optimal number of major histocompatibility complex molecules in an individual. *Proceedings of the National Academy of Sciences* **89**(22): 10896-10899.

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

O'Connor EA, Cornwallis CK, Hasselquist D, Nilsson J-Å and Westerdahl H (2018). The evolution of immunity in relation to colonization and migration. *Nature Ecology & Evolution* **2**(5): 841-849.

O’Connor EA, Westerdahl H, Burri R and Edwards SV (2019). Avian MHC evolution in the era of genomics: Phase 1.0. *Cells* **8**(10): 1152.

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

Peiponen V (1960). Verhaltensstudien am blaukehlchen [Behavior studies of the bluethroat]. *Ornis Fennica* **37**: 69-83.

Penn DJ and Potts WK (1999). The evolution of mating preferences and major histocompatibility complex genes. *The American Naturalist* **153**(2): 145-164.

Penn DJ (2002). The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. *Ethology* **108**(1): 1-21.

Penn DJ, Damjanovich K and Potts WK (2002). MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences* **99**(17): 11260-11264.

Petrie M and Kempenaers B (1998). Extra-pair paternity in birds: Explaining variation between species and populations. *Trends in Ecology & Evolution* **13**(2): 52-58.

Piertney S and Oliver M (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**(1): 7-21.

Podmokła E, Dubiec A, Arct A, Drobniak SM, Gustafsson L and Cichoń M (2015). Malaria infection status predicts extra-pair paternity in the blue tit. *Journal of Avian Biology* **46**(3): 303-306.

Potts WK and Wakeland EK (1990). Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution* **5**(6): 181-187.

Potts WK, Manning CJ and Wakeland EK (1994). The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: An experimental test. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **346**(1317): 369-378.

Promerová M, Vinkler M, Bryja J, Poláková R, Schnitzer J, Munclinger P and Albrecht T (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**(1): 5-10.

Promerová M, Králová T, Bryjová A, Albrecht T and Bryja J (2013). MHC class IIB exon 2 polymorphism in the Grey partridge (*Perdix perdix*) is shaped by selection, recombination and gene conversion. *PloS One* **8**(7): e69135.

Promerová M, Alavioon G, Tusso S, Burri R and Immler S (2017). No evidence for MHC class II-based non-random mating at the gametic haplotype in Atlantic salmon. *Heredity* **118**(6): 563-567.

Rammensee H-G (1995). Chemistry of peptides associated with MHC class I and class II molecules. *Current Opinion in Immunology* **7**(1): 85-96.

Reusch TB, Häberli MA, Aeschlimann PB and Milinski M (2001). Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature* **414**(6861): 300-302.

Richardson DS, Komdeur J, Burke T and 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**(1564): 759-767.

Roberts SC (2009). Complexity and context of MHC-correlated mating preferences in wild populations. *Molecular Ecology* **18**(15): 3121-3123.

Sandberg M, Eriksson L, Jonsson J, Sjöström M and Wold S (1998). New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *Journal of Medicinal Chemistry* **41**(14): 2481-2491.

Sawai H, Go Y and Satta Y (2008). Biological implication for loss of function at major histocompatibility complex loci. *Immunogenetics* **60**(6): 295-302.

Schwensow N, Fietz J, Dausmann K and Sommer S (2008). MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate. *Evolutionary Ecology* **22**(5): 617-636.

Sebastian A, Herdegen M, Migalska M and Radwan J (2016). amplisas: A web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources* **16**(2): 498-510.

Sepil I, Moghadam HK, Huchard E and 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**(1): 68.

Sepil I, Lachish S, Hinks AE and Sheldon BC (2013a). 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* **280**(1759): 20130134.

Sepil I, Lachish S and Sheldon BC (2013b). Mhc-linked survival and lifetime reproductive success in a wild population of great tits. *Molecular Ecology* **22**(2): 384-396.

Sette A, Buus S, Appella E, Smith JA, Chesnut R, Miles C, Colon SM and Grey HM (1989). Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proceedings of the National Academy of Sciences* **86**(9): 3296-3300.

Sette A and Sidney J (1998). HLA supertypes and supermotifs: A functional perspective on HLA polymorphism. *Current Opinion in Immunology* **10**(4): 478-482.

Sette A and Sidney J (1999). Nine major HLA class I supertypes account for the vast preponderance of HLA-A and-B polymorphism. *Immunogenetics* **50**(3-4): 201-212.

Slade J, Watson M, Kelly T, Gloor G, Bernards M and Macdougall-Shackleton E (2016). Chemical composition of preen wax reflects major histocompatibility complex similarity in songbirds. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **283**(1842): 20161966.

Slade J, Watson M and Macdougall-Shackleton E (2017). Birdsong signals individual diversity at the major histocompatibility complex. *Biology Letters* **13**(11): 20170430.

Slade JW, Watson MJ and Macdougall-Shackleton EA (2019). “Balancing” balancing selection? Assortative mating at the major histocompatibility complex despite molecular signatures of balancing selection. *Ecology and Evolution* **9**(9): 5146-5157.

Slade RW and Mccallum HI (1992). Overdominant vs. frequency-dependent selection at MHC loci. *Genetics* **132**(3): 861-862.

Snell G (1968). The H-2 locus of the mouse: Observations and speculations concerning its comparative genetics and its polymorphism. *Folia Biologica* **14**(5): 335-358.

Sommer S (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* **2**(1): 16.

Sommer S, Courtiol A and Mazzoni CJ (2013). MHC genotyping of non-model organisms using next-generation sequencing: A new methodology to deal with artefacts and allelic dropout. *BMC Genomics* **14**(1): 542.

Spurgin LG and Richardson DS (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences* **277**(1684): 979-988.

Spurgin LG, Van Oosterhout C, Illera JC, Bridgett S, Gharbi K, Emerson BC and Richardson DS (2011). Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology* **20**(24): 5213-5225.

Stadecker MJ, Lukic M, Dvorak A and Leskowitz S (1977). The cutaneous basophil response to phytohemagglutinin in chickens. *The Journal of Immunology* **118**(5): 1564-1568.

Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH and Sharma A (2011). The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**(7363): 207-210.

Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL and Wiley DC (1994). Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368**(6468): 215-221.

Strandh M, Westerdahl H, Pontarp M, Canbäck B, Dubois M-P, Miquel C, Taberlet P and Bonadonna F (2012). Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. *Proceedings of the Royal Society B: Biological Sciences* **279**(1746): 4457-4463.

Stuglik MT, Radwan J and Babik W (2011). jMHC: Software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Molecular Ecology Resources* **11**(4): 739-742.

Sturm T, Leinders-Zufall T, Maček B, Walzer M, Jung S, Pömmerl B, Stevanović S, Zufall F, Overath P and Rammensee H-G (2013). Mouse urinary peptides provide a molecular basis for genotype discrimination by nasal sensory neurons. *Nature Communications* **4**: 1616.

Støstad HN, Rekdal SL, Kleven O, Laskemoen T, Marthinsen G, Johnsen A and Lifjeld JT (2016). Weak geographical structure in sperm morphology across the range of two willow warbler *Phylloscopus trochilus* subspecies in Scandinavia. *Journal of Avian Biology* **47**(5): 731-741.

Svensson L (1992). Identification guide to European passerines. 4th edition. Thetford, United Kingdom, British Trust for Ornithology.

Svoboda A, Marthinsen G, Pavel V, Chutný B, Turčoková L, Lifjeld JT and Johnsen A (2015). Blood parasite prevalence in the Bluethroat is associated with subspecies and breeding habitat. *Journal of Ornithology* **156**(2): 371-380.

Sætre CLC, Johnsen A, Stensrud E and Cramer ER (2018). Sperm morphology, sperm motility and paternity success in the bluethroat (*Luscinia svecica*). *PloS One* **13**(3): e0192644.

Takahata N and Nei M (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**(4): 967-978.

Taylor Jr RL, Cotter P, Wing T and Briles W (1987). Major histocompatibility (B) complex and sex effects on the phytohaemagglutinin wattle response. *Animal genetics* **18**(4): 343-350.

Tella JL, Scheuerlein A and Ricklefs RE (2002). Is cell-mediated immunity related to the evolution of life-history strategies in birds? *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**(1495): 1059-1066.

Tella JL, Lemus JA, Carrete M and Blanco G (2008). The PHA test reflects acquired T-cell mediated immunocompetence in birds. *PLoS One* **3**(9): e3295.

Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E, Funkhouser R, Fugate M, Theiler J and Hsu YS (2003). Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine* **9**(7): 928-935.

Tregenza T and Wedell N (2000). Genetic compatibility, mate choice and patterns of parentage: invited review. *Molecular Ecology* **9**(8): 1013-1027.

Trivers R (1972). Parental investment and sexual selection. Chapter 7 in: *Sexual selection and the descent of man, 1871-1971.* 1st edition. Ed.: Campbell BG. Chicago, Aldine publishing company.

Trowsdale J and Knight JC (2013). Major histocompatibility complex genomics and human disease. *Annual Review of Genomics and Human Genetics* **14**: 301-323.

Van De Pol M and Wright J (2009). A simple method for distinguishing within-versus between-subject effects using mixed models. *Animal Behaviour* **77**(3): 753.

Van Den Berg HA and Rand DA (2003). Antigen presentation on MHC molecules as a diversity filter that enhances immune efficacy. *Journal of Theoretical Biology* **224**(2): 249-267.

Van Valen L (1973). A new evolutionary law. *Evolutionary Theory* **1**: 1-30.

Vidović D and Matzinger P (1988). Unresponsiveness to a foreign antigen can be caused by self-tolerance. *Nature* **336**(6196): 222-225.

Vinkler M, Bainová H and Albrecht T (2010). Functional analysis of the skin-swelling response to phytohaemagglutinin. *Functional Ecology* **24**(5): 1081-1086.

Vinkler M and Albrecht T (2011). Handling 'immunocompetence' in ecological studies: Do we operate with confused terms? *Journal of Avian Biology* **42**(6): 490-493.

Vinkler M, Schnitzer J, Munclinger P and Albrecht T (2012). Phytohaemagglutinin skin-swelling test in scarlet rosefinch males: Low-quality birds respond more strongly. *Animal Behaviour* **83**(1): 17-23.

Vinkler M, Svobodová J, Gabrielová B, Bainová H and Bryjová A (2014). Cytokine expression in phytohaemagglutinin-induced skin inflammation in a galliform bird. *Journal of Avian Biology* **45**(1): 43-50.

Von Schantz T, Wittzell H, Bose N, Grahn M and Persson K (1996). MHC genotype and male ornamentation: Genetic evidence for the Hamilton-Zuk model. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **263**(1368): 265-271.

Wakeland EK, Boehme S, She JX, Lu C-C, Mcindoe RA, Cheng I, Ye Y and Potts WK (1990). Ancestral polymorphisms of MHC class II genes: Divergent allele advantage. *Immunologic Research* **9**(2): 115-122.

Wedekind C, Seebeck T, Bettens F and Paepke AJ (1995). MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **260**(1359): 245-249.

Wegner KM, Kalbe M, Kurtz J, Reusch TB and Milinski M (2003). Parasite selection for immunogenetic optimality. *Science* **301**(5638): 1343-1343.

Westerdahl H, Wittzell H and Von Schantz T (2000). Mhc diversity in two passerine birds: No evidence for a minimal essential Mhc. *Immunogenetics* **52**(1-2): 92-100.

Westerdahl H (2004). No evidence of an MHC-based female mating preference in great reed warblers. *Molecular Ecology* **13**(8): 2465-2470.

Westerdahl H, Wittzell H, Von Schantz T and Bensch S (2004). MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. *Heredity* **92**(6): 534-542.

Westerdahl H (2007). Passerine MHC: Genetic variation and disease resistance in the wild. *Journal of Ornithology* **148**(2): 469-477.

Westerdahl H, Asghar M, Hasselquist D and Bensch S (2012). Quantitative disease resistance: To better understand parasite-mediated selection on major histocompatibility complex. *Proceedings of the Royal Society B* **279** (1728): 577-584.

Westneat DF and Stewart IR (2003). Extra-pair paternity in birds: Causes, correlates, and conflict. *Annual Review of Ecology, Evolution, and Systematics* **34**(1): 365-396.

Whittaker DJ, Soini HA, Atwell JW, Hollars C, Novotny MV and Ketterson ED (2010). Songbird chemosignals: Volatile compounds in preen gland secretions vary among individuals, sexes, and populations. *Behavioral Ecology* **21**(3): 608-614.

Whittingham LA, Freeman-Gallant CR, Taff CC and Dunn PO (2015). Different ornaments signal male health and MHC variation in two populations of a warbler. *Molecular Ecology* **24**(7): 1584-1595.

Whittingham LA, Dunn PO, Freeman-Gallant CR, Taff CC and Johnson JA (2018). Major histocompatibility complex variation and blood parasites in resident and migratory populations of the common yellowthroat. *Journal of Evolutionary Biology* **31**(10): 1544-1557.

Winternitz J, Minchey S, Garamszegi LZ, Huang S, Stephens P and Altizer S (2013). Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. *Proceedings of the Royal Society B: Biological Sciences* **280**(1769): 20131605.

Winternitz J, Abbate J, Huchard E, Havlíček J and Garamszegi L (2017). Patterns of MHC-dependent mate selection in humans and nonhuman primates: A meta-analysis. *Molecular Ecology* **26**(2): 668-688.

Winternitz JC, Promerova M, Polakova R, Vinker M, Schnitzer J, Munclinger P, Babik W, Radwan J, Bryja J and Albrecht T (2015). Effects of heterozygosity and MHC diversity on patterns of extra-pair paternity in the socially monogamous scarlet rosefinch. *Behavioral Ecology and Sociobiology* **69**(3): 459-469.

Woelfing B, Traulsen A, Milinski M and Boehm T (2009). Does intra-individual major histocompatibility complex diversity keep a golden mean? *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**(1513): 117-128.

Yamazaki K, Boyse E, Mike V, Thaler H, Mathieson B, Abbott J, Boyse J, Zayas Z and Thomas L (1976). Control of mating preferences in mice by genes in the major histocompatibility complex. *The Journal of Experimental Medicine* **144**(5): 1324-1335.

Yang Z (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**(8): 1586-1591.

Zahavi A (1975). Mate selection - A selection for a handicap. *Journal of Theoretical Biology* **53**(1): 205-214.

Ziegler A, Kentenich H and Uchanska-Ziegler B (2005). Female choice and the MHC. *Trends in Immunology* **26**(9): 496-502.

Zinkernagel RM and Doherty PC (1974). Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**(5450): 701-702.

Østhagen H and Egelie K (1978). The vegetation of the Øvre Heimdalen valley. *Ecography* **1**(2-3): 103-106.

Paper I - IV

I

Photo: Bjørn Aksel Bjerke

Received: 19 September 2017 | Revised: 30 November 2017 | Accepted: 4 December 2017

DOI: 10.1002/ece3.3757

ORIGINAL RESEARCH

WILEY Ecology and Evolution Open Access

Genotyping strategy matters when analyzing hypervariable major histocompatibility complex-Experience from a passerine bird

Silje L. Rekdal | **Jarl Andreas Anmarkrud** | **Arild Johnsen** | **Jan T. Lifjeld**

Natural History Museum, University of Oslo, Oslo, Norway

Correspondence
Silje L. Rekdal, Natural History Museum, University of Oslo, Oslo, Norway.
Email: silje.rekdal@nhm.uio.no

Funding information
Natural History Museum, Oslo, Norway

Abstract

Genotyping of classical major histocompatibility complex (MHC) genes is challenging when they are hypervariable and occur in multiple copies. In this study, we used several different approaches to genotype the moderately variable MHC class I exon 3 (MHCIe3) and the highly polymorphic MHC class II exon 2 (MHCIIβe2) in the bluethroat (*Luscinia svecica*). Two family groups (eight individuals) were sequenced in replicates at both markers using Ion Torrent technology with both a single- and a dual-indexed primer structure. Additionally, MHCIIβe2 was sequenced on Illumina MiSeq. Allele calling was conducted by modifications of the pipeline developed by Sommer et al. (BMC Genomics, 14, 2013, 542) and the software AmpliSAS. While the different genotyping strategies gave largely consistent results for MHCIe3, with a maximum of eight alleles per individual, MHCIIβe2 was remarkably complex with a maximum of 56 MHCIIβe2 alleles called for one individual. Each genotyping strategy detected on average 50%–82% of all MHCIIβe2 alleles per individual, but dropouts were largely allele-specific and consistent within families for each strategy. The discrepancies among approaches indicate PCR biases caused by the platform-specific primer tails. Further, AmpliSAS called fewer alleles than the modified Sommer pipeline. Our results demonstrate that allelic dropout is a significant problem when genotyping the hypervariable MHCIIβe2. As these genotyping errors are largely nonrandom and method-specific, we caution against comparing genotypes across different genotyping strategies. Nevertheless, we conclude that high-throughput approaches provide a major advance in the challenging task of genotyping hypervariable MHC loci, even though they may not reveal the complete allelic repertoire.

KEYWORDS
bluethroat, Illumina MiSeq, Ion Torrent, *Luscinia svecica*, major histocompatibility complex

1 | INTRODUCTION

The polymorphic and polygenic major histocompatibility complex (MHC) in many vertebrates is inherently difficult to genotype. The last decade has provided new sequencing platforms that may enable cheaper, faster, and more accurate and reproducible MHC genotyping (Babik, Taberlet, Ejsmond, & Radwan, 2009; Biedrzycka, Sebastian, Migalska, Westerdahl, & Radwan, 2017; Duke et al., 2015; Grogan, McGinnis, Sauther, Cuozzo, & Drea, 2016; Lighten, Oosterhout, & Bentzen, 2014). However, the task is still quite demanding in hypervariable study systems, partly due to the challenge of designing pipelines to separate true alleles from methodological artifacts (Babik, 2010; Grogan et al., 2016; Lighten et al., 2014; Sebastian, Herdegen, Migalska, & Radwan, 2016). Validation of true alleles through the use of replicate amplicons and pedigree information may however assist in quality control of the genotyping process of such systems (Gaigher et al., 2016; Grogan et al., 2016; Sommer, Courtiol, & Mazzoni, 2013; Zagalska-Neubauer et al., 2010).

Major histocompatibility complex genes are crucial to trigger adaptive immune responses in jawed vertebrates and are among the most polymorphic genes known (Janeway, Travers, Walport, & Shlomchik, 2001). MHC class I (MHCI) genes encode transmembrane glycoproteins found on the surface of most cells. Peptides derived from intracellular pathogens bind specifically to the peptide-binding region (PBR) of membrane-bound MHCI molecules and are presented to CD8$^+$ cytotoxic T cells. Similarly, peptides from extracellular pathogens are presented to CD4$^+$ helper T cells by MHC class II (MHCII) molecules, which are located on specialized antigen-presenting cells such as B cells, dendritic cells, and macrophages. Due to the specificity of the PBR, organisms with more MHC alleles are able to trigger an immune response against more pathogens. The polymorphism at MHC genes is believed to be influenced by several processes, including pathogen-mediated balancing selection and sexual selection (see reviews by Edwards and Hedrick (1998) and Piertney and Oliver (2006)).

In Aves, the structure of the MHC varies immensely. While chicken (*Gallus gallus*) is described as having a “minimal essential MHC” (Kaufman et al., 1999), many non-Galliform species exhibit an increased number of MHC loci (e.g., great snipe (*Gallinago media*; Ekblom, Grahn, & Höglund, 2003), blue petrel (*Halobaena caerulea*; Strandh, Lannefors, Bonadonna, & Westerdahl, 2011), and Eurasian coot (*Fulica atra*; Alcaide, Munoz, Martínez-de la Puente, Soriguer, & Figuerola, 2014)). In Passeriformes, the MHC genes are extensively duplicated and highly diverse, and pseudogenes are commonly found (Westerdahl, 2007). For instance, Bollmer, Dunn, Freeman-Gallant, and Whittingham (2012) detected a minimum of eight MHCI exon 3 (MHCIe3) and 23 MHCII β exon 2 (MHCIIβe2) loci in the common yellowthroat (*Geothlypis trichas*) using 454 sequencing, while Zagalska-Neubauer et al. (2010) revealed numerous pseudogenes as well as at least nine transcribed MHCIIβe2 loci in collared flycatcher (*Ficedula albicollis*). Further, O'Connor, Strandh, Hasselquist, Nilsson, and Westerdahl (2016) described MHCIe3 diversity in 12 passerine species, in which the minimum number of loci ranged from four in the bluethroat (*Luscinia svecica*) to 19 in the willow warbler (*Phylloscopus trochilus*). High intra-individual diversity was also found by Anmarkrud, Johnsen, Bachmann, and Lifjeld (2010), who used a traditional cloning and Sanger sequencing approach to identify 61 unique MHCIIβe2 alleles in 20 bluethroats and a minimum number of 11 functional loci.

For passerine birds, gene duplication and high-sequence similarity at MHC loci due to gene conversion preclude single-locus amplification when intron sequences are not known (Westerdahl, 2007). When performing PCR amplicon sequencing from multilocus gene targets, such as the MHC, several aspects may contribute to PCR-induced biases. For example, similarity to primer sequence, GC content in primer-binding sites and differences in secondary structures will influence the amplification success of the DNA template (Pawluczyk et al., 2015; Polz & Cavanaugh, 1998; Suzuki & Giovannoni, 1996). Hence, primer design is important in order to reduce PCR-introduced biases. Many researchers now use “phusion primers” when performing amplicon sequencing. These are primer sequences with platform-specific adapters, sample-specific index tags (barcode), and other sequence motifs added to the target gene sequence. These motifs will generate a “primer tail”. This tail may thus introduce amplification biases if it has a noncompatible GC pattern to the nucleotides surrounding the primer-binding motif, or if local secondary structures obstruct annealing of the primer.

Although Roche 454 pyrosequencing has been extensively applied in MHC studies on nonmodel organisms since 2009 (Babik et al., 2009), this platform is now being phased out and new technologies are applied. Of the available high-throughput sequencing platforms, Ion Torrent semiconductor sequencing and Illumina MiSeq paired-end sequencing are currently among the most appropriate alternatives for MHC genotyping due to read lengths and output. However, as for every sequencing method, these techniques are also subject to errors. In addition to substitution errors made by polymerases, chimera formation is common in multilocus PCR amplification (Kanagawa, 2003; Lenz & Becker, 2008). Further, homopolymer errors causing indels are abundant in Ion Torrent (Loman et al., 2012). Being aware of these pitfalls is essential, and it is important to take measures to minimize the impact of artifacts arising before or during sequencing. Thus, establishment of robust PCR approaches and allele-calling pipelines is crucial for separating artifacts from true alleles.

Two main assumptions are generally made in the processes of recognizing artifacts in MHC studies: Artifacts should be less common than real alleles across and within individuals, and they should originate from true alleles (Babik et al., 2009). Based on these assumptions, Babik et al. (2009) used per individual frequencies to identify a threshold below which artifacts should occur. This was further elaborated by Galan, Guivier, Caraux, Charbonnel, and Cosson (2010) who established two thresholds: T_1; the minimum number of reads *per sample* required for reliable genotyping, and T_2; intra-amplicon frequency corresponding to the minimum number of reads *per variant* to validate true alleles. However, as pure threshold approaches potentially misidentify alleles and artifacts (Lighten et al., 2014), stricter approaches are needed in complex MHC systems. In a MHCII study on flycatchers, Zagalska-Neubauer et al. (2010) used a 2-PCR-3-reads-in-each inclusion criteria, where the variants had to be present with at least three reads in two independent PCRs to be considered alleles. Further, in order to account for artifacts and allelic dropout, Sommer et al. (2013) established an expanded workflow for genotyping MHC in nonmodel organisms. Their allele-calling pipeline relies on amplicon replicates, artifact detection, and relative intra-amplicon frequencies after initial quality filtering of the sequencing reads. By combining this pipeline with a 1% threshold approach similar to Galan's T_2, Grogan et al. (2016)

genotyped MHC-DRB in ring-tailed lemurs (*Lemur catta*) across 454 and Ion Torrent platforms with consistent results. Introducing such a threshold of 0.4% also minimized the number of artifacts when genotyping the hypervariable MHCIe3 in sedge warblers (*Acrocephalus schoenobaenus*) using this pipeline (Biedrzycka et al., 2017).

Another MHC allele-calling pipeline based on a stepwise threshold clustering methodology was developed by Stutz and Bolnick (2014). In this pipeline, sequences are clustered based on similarity and filtered, attempting to reveal artifacts and add their depths to the putative alleles from which they are originating. By incorporating platform-specific error rates, Sebastian et al. (2016) have implemented this pipeline in the publicly available AmpliSAS tool.

Biedrzycka et al. (2017) compared four allele-calling strategies for genotyping sedge warbler MHCIe3 including both AmpliSAS and the original workflow established by Sommer et al. (2013). They found high agreement (>90%) between these pipelines at coverages above 2,000 reads but argue for the use of coverages of >5,000 reads due to the increased reliability.

In our study, we further compare allele calling by modifying the Sommer pipeline and AmpliSAS by the use of family data, in order to facilitate genotyping of bluethroat MHC. We thus compare different aspects of genotyping strategies (i.e., sequencing approaches and allele-calling pipelines) in a moderately variable MHC gene (MHCIe3) and a highly polymorphic MHC gene (MHCIIβe2) in the bluethroat, using two family sets (two offspring in each family, with their genetic parents). The MHCIe3 and MHCIIβe2 amplicons were sequenced on an Ion Personal Genome Machine™, applying two different primer tail approaches. Additionally, the MHCIIβe2 amplicons were sequenced on the Illumina MiSeq® platform. Using these approaches, we aimed to test whether platform or sequence motif and length of the primer tail would bias the outcome. We modified the pipeline of Sommer et al. (2013) and the downstream analyses of the output from AmpliSAS software (Sebastian et al., 2016), and used allelic inheritance patterns between parent and offspring genotypes as additional support for the results. Accordingly, using family data, we wished to establish a workflow for robust genotyping of bluethroat MHCIe3 and MHCIIβe2, which is a premise for the use of these markers in ecological and evolutionary analyses.

2 | MATERIALS AND METHODS

This study is based on DNA from blood samples of two offspring and their biological parents in two family groups of bluethroats (*L. svecica svecica*; Appendix S1). The eight individuals were sampled in the subalpine habitat of Øvre Heimdalen valley, Øystre Slidre, Norway (61°25′N, 8°52′E). Norwegian Animal Research Authority gave ethical permissions to the fieldwork (license 2014/53673 to AJ). Parentage was confirmed through a panel of microsatellites in another study (Sætre, Johnsen, Stensrud, & Cramer, unpublished data). DNA was extracted using E-Z® 96 Blood DNA Kit (Omega Bio-Tek Inc. [D1199-01]), following the protocol of the manufacturer.

2.1 | Sequencing

All amplicons (for explanation of terms, see Appendix S2) were amplified in duplicates, with a unique barcode identifier for each replicate. Primer sequences and binding sites are provided in the Supplementary material (Appendices S3 and S4). In order to minimize PCR artifacts, the number of PCR cycles was reduced to 25 (Lenz & Becker, 2008). For detailed description of the amplification and sequencing, see Appendix S5.

MHCIe3 was amplified using the primer pair MhcPasCI-FW and MhcPasCI-RV (Alcaide, Liu, & Edwards, 2013) and sequenced on an Ion PGM. Two primer structures were applied (see Figure 1); one in which barcode and barcode adapter were included only on the forward primer (single index; SI), and one including Ion Torrent adapter, barcode, barcode adapter, and seven nucleotide spacer motif on both forward and reverse primers (dual index; DI).

The primers MHCIIFihy-E2CF and MHCIIFihy-E2CR (Canal, Alcaide, Anmarkrud, & Potti, 2010) were used to amplify MHCIIβe2 by a similar SI and DI approach, and the amplicon sequencing was conducted on an Ion PGM. Additionally, MHCIIβe2 amplicons were generated by including Illumina Linker sequences, barcodes, and heterogeneity spacer motif (Fadrosh et al., 2014) on both forward and revers primers, and sequenced on Illumina MiSeq (see Figure 1).

2.2 | Allele calling

Allele calling was conducted through two pipelines: one based on a previously published pipeline by Sommer et al. (2013), and one based on the software AmpliSAS (Sebastian et al., 2016).

2.2.1 | Modified pipeline from Sommer et al. (2013)

A flowchart of this allele-calling method is outlined in Figure 2, while a detailed description and comments on the modifications are provided in Appendix S6. In short, paired MiSeq reads were merged using FLASH (Magoč & Salzberg, 2011), and raw reads from all datasets were quality filtered using standard UNIX commands and fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Barcode splitting and clustering of identical reads into variants were conducted by jMHC (Stuglik, Radwan, & Babik, 2011). Variants with less than three reads in any amplicon were discarded, as were amplicons with less than 500 reads in total. After this step, we established a threshold above which we expected to have included most true alleles and excluded most artifacts, based on the assumption that true alleles will amplify to a greater depth than artifacts will (Babik et al., 2009; Lighten et al., 2014). A cut-off threshold of 0.2% was conservatively inferred from visually recognizing a change in number of unique variants included at different values of cut-off (Figure 3). Hence, variants with intra-amplicon frequency of less than 0.2% were discarded for the respective amplicons.

The remaining variants across the whole dataset were aligned to previously published sequences (GenBank accession number KU169737–KU169747 (MHCI; O'Connor et al., 2016) and

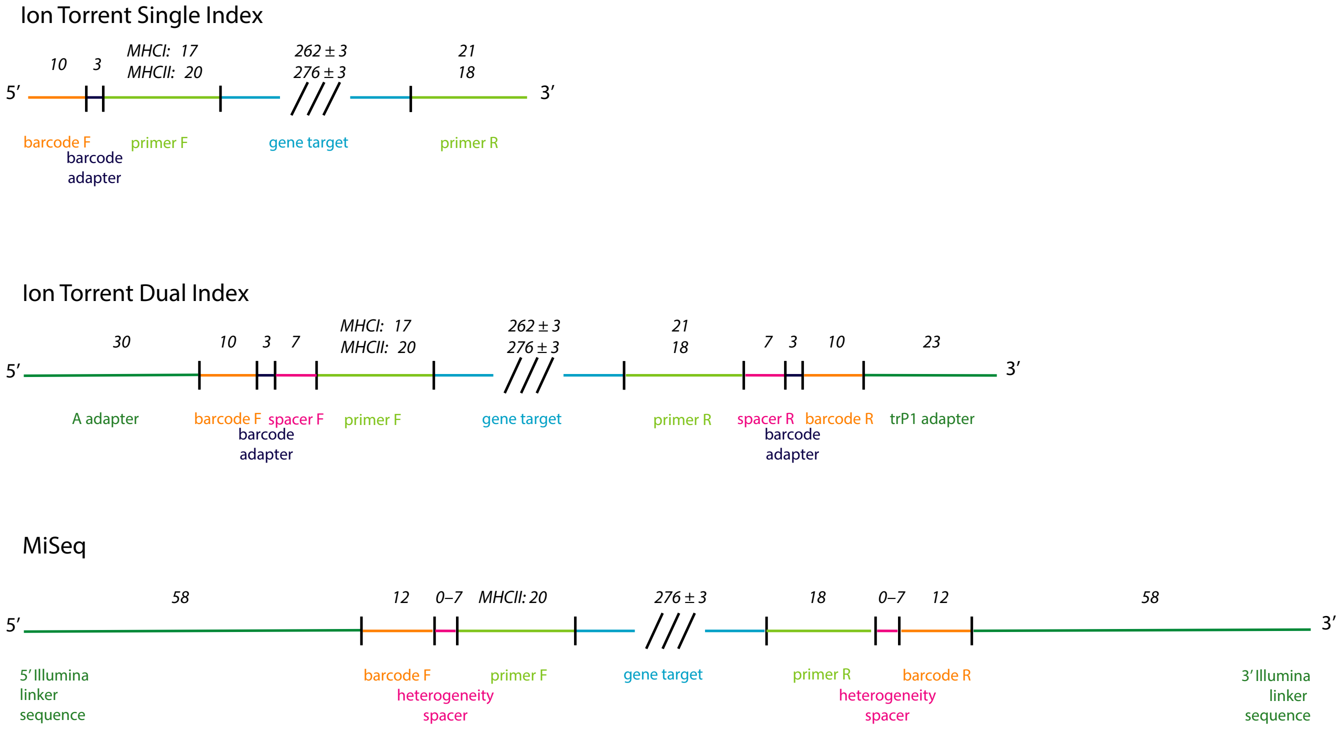


FIGURE 1 Primer setup used to amplify MHC class I exon 3 (MHCI) and MHC class II exon 2 (MHCII) in eight bluethroats. MHCI was sequenced in a single index run and in a dual index run on Ion Torrent. MHCII was additionally sequenced on Illumina MiSeq. The numbers are referring to the length of the respective parts of the primers

HQ539575–HQ539614 (MHCII; Gohli et al., 2013)) with ClustalW (Thompson, Higgins, & Gibson, 1994) in MEGA7 (Kumar, Stecher, & Tamura, 2016), and trimmed correspondingly. Variants with shift in reading frame, stop codon, or lacking the conserved residues Cys7 and Cys70 (MHCI; O'Connor et al., 2016) or Cys10 and Cys75 (MHCII; Gohli et al., 2013) were discarded. Further, chimera detection was carried out using UCHIME (Edgar, Haas, Clemente, Quince, & Knight, 2011).

While the most frequent variant within each amplicon was scored as an allele, the remaining variants were divided into "=1 bp" and ">1 bp" variants, according to the number of base pair differences to their most similar, more frequent variant, found by MEGA7. Amplicon replicates were then utilized to score artifacts (chimeric variants scored as chimera also in replicate or not found within replicate above 0.2% threshold; "=1 bp" variants not found within replicate above 0.2% threshold; ">1 bp" variant not present above 0.2% threshold in any other amplicon from individuals within the same family group). The remaining variants within each amplicon were scored as alleles (see Figure 2).

For offspring with one failed amplicon, family information was used as a substitute for the replicate. Here, "=1 bp" variant or a chimeric variant was scored as an allele if present in parental genotypes. ">1 bp" variants were called as alleles if found in any other family member.

2.2.2 | AmpliSAS pipeline

As a second allele-calling pipeline, we used the online tool suite AmpliSAT (Sebastian et al., 2016). After initial filtering (see Appendix S6), the datasets were explored in AmpliCHECK. Corresponding to the AmpliCHECK results, we set maximum number of alleles per amplicon to 60 for the algorithm implemented in AmpliSAS. Minimum amplicon depth was set to 500, while we used default platform-specific error rates for substitutions and indels. In-frame length was required for the dominant sequence within a cluster. Further, the frequency of the subdominant cluster with respect to the dominant frequency was changed from the default of 25% to 10%, in order to avoid clustering similar alleles with different amplification efficiencies (Biedrzycka et al., 2017). Lastly, variants with an intra-amplicon frequency of less than 0.20% or with a depth below three reads were discarded. Variants of lengths exhibiting frameshifts in relation to the expected length, noncoding variants, and chimeras were also discarded.

Duplicates of all individuals enabled validation of alleles based on the presence in the replicate sample. Hence, only variants scored in both amplicon replicates of an individual were called as alleles. Error rates were calculated as the percentage of putative alleles not found in both replicate runs (errors in replicates).

One offspring had however only one successful MHCII-MiSeq amplicon. Here, we called putative alleles if present in one or both parents in the same run.

2.3 | Comparing genotypes from modified Sommer pipeline and AmpliSAS

As the sequences were trimmed in the modified pipeline from Sommer et al. (2013) in order to match published sequences, the sequences from AmpliSAS were aligned and trimmed correspondingly

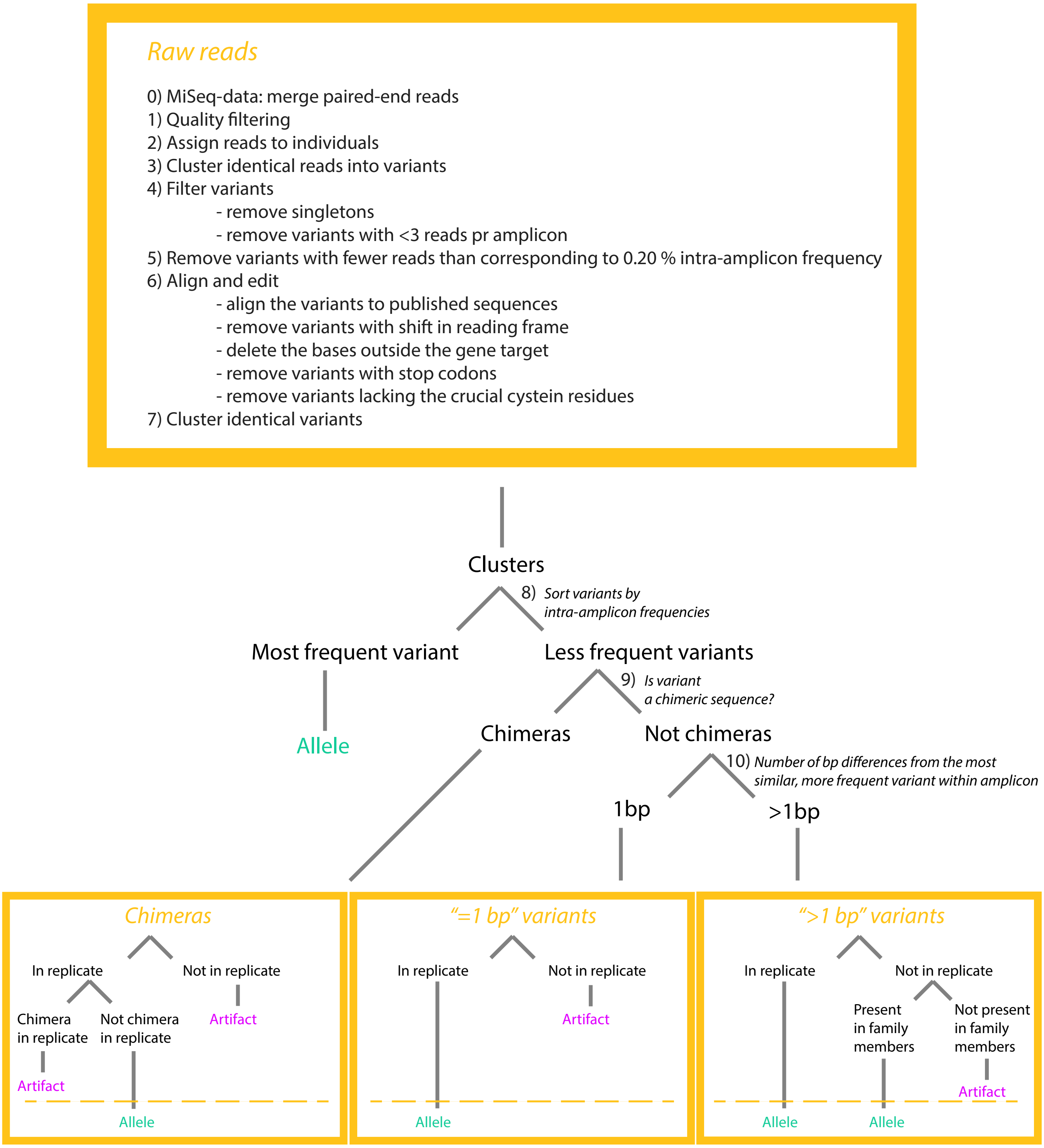


FIGURE 2 Flowchart over the pipeline modified from Sommer et al. (2013), conducted to genotype MHC class I exon 3 and MHC class II exon 2 in eight bluethroats

in MEGA7, and collapsed using fastx toolkit. Identical alleles obtained in different pipelines were given identical names using standard UNIX commands. The genotypes were then compared, both among the different primer setups and between the allele-calling pipelines, by counting shared alleles across methods within individuals. A saturation rate plot was established, to visualize the average proportion of all alleles genotyped for an individual each approach was able to genotype (Figure 4). The figure was made in R, version 3.2.5 (R Core Team 2016), with the package ggplot2 (Wickham, 2009).

Within each approach, the family data were evaluated in order to reveal lack of Mendelian inheritance, that is, variants scored

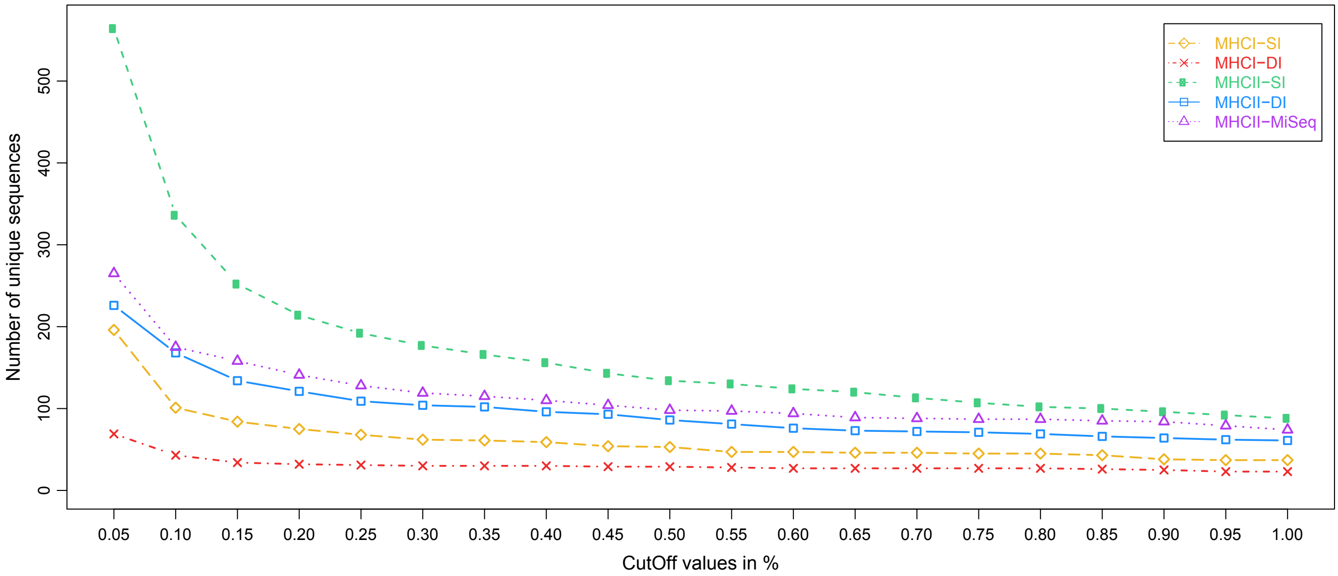


FIGURE 3 Number of unique sequences for different values of cut-off, found within each strategy (MHC class, primer approach, and platform) in the eight bluethroat individuals. Only variants having three reads or more in any amplicon are included. Short and low-quality reads are removed, but no further filtering was conducted before this step, in order to reveal any threshold where low-frequency artifacts are likely included

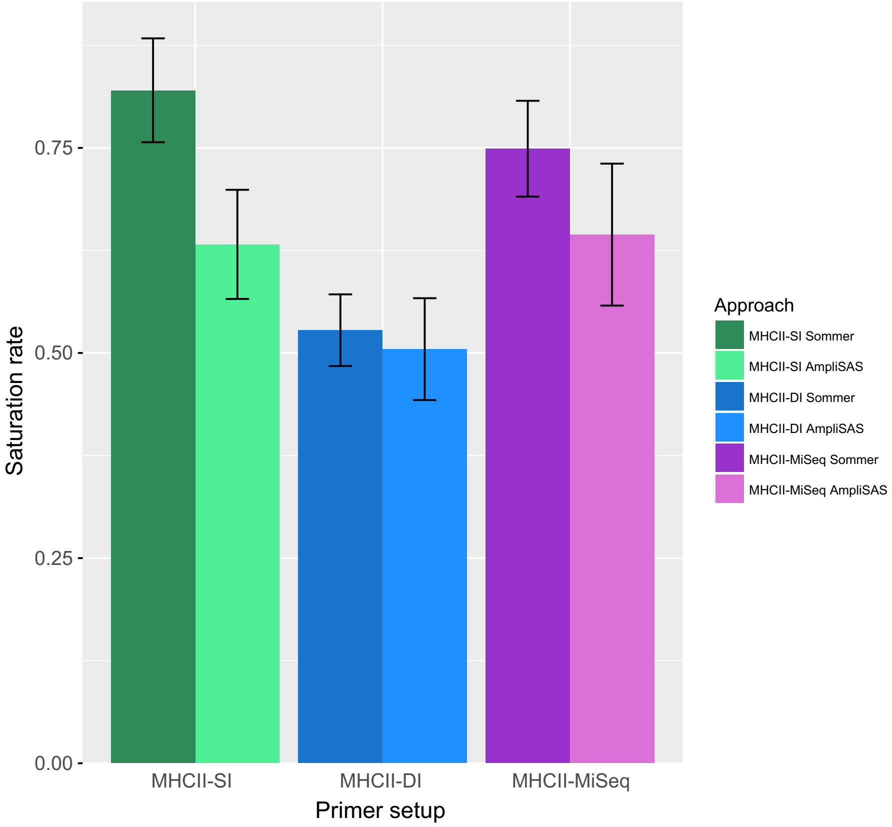


FIGURE 4 The saturation rate for each approach for genotyping MHC class II exon 2 in eight bluethroats, calculated as the average proportion of the individual “combined genotypes” each approach was able to genotype. The “combined genotype” was established for each individual by combining all alleles that were called in the individual using at least one approach

as alleles in offspring but not in any of its parent. Errors in pedigree were calculated as the percentage of alleles genotyped in offspring individuals but not found in any of their parents. Only offspring having no failed replicates were used for this purpose.

3 | RESULTS

3.1 | MHCIe3

After barcode splitting in both allele-calling pipelines, all MHCI-SI amplicons and 12 of 16 MHCI-DI amplicons had >5,000 reads per amplicon.

TABLE 1 Results from genotyping eight bluethroats at MHC class I exon 3 (MHCI), using two primer strategies (single index [SI] and dual index [DI]) on Ion Torrent

	MHCI-SI	MHCI-DI
Modified pipeline from Sommer et al. (2013)		
Reads per amplicon after jMHC (including only variants with ≥3 reads within any amplicon)	8,744–48,840	2,953–18,154
Total number of alleles	18	18
Alleles per individual	6.6 (5–8)	6.6 (5–8)
Unique amino acid sequences	14	14
Errors in pedigree	0	0
AmpliSAS		
Reads per amplicon	>5,000	3,002–>5,000
Reads per amplicon assigned to alleles	2,797–3,534	2,028–4,815
Total number of alleles	16	16
Alleles per individual	5.4 (4–7)	5.0 (4–7)
Alleles per amplicon	5.6 (4–7)	5.5 (4–8)
Unique amino acid sequences	13	13
Errors in replicates	3.4%	9.1%
Errors in pedigree	0	0

The datasets were genotyped through two pipelines; one using a modification of the pipeline published by Sommer et al. (2013) and one using the online tool AmpliSAS (Sebastian et al., 2016).

The remaining four MHCI-DI amplicons had >2,900 reads per amplicon.

Using the modified pipeline from Sommer et al. (2013), 18 unique alleles were identified in all eight individuals combined, with complete concordance for every individual between the MHCI-SI and the MHCI-DI primer setup. There were between five and eight alleles per individual across all samples (Appendix S7), corresponding to at least four MHCI-loci. The 18 alleles translated into 14 unique amino acid sequences (Table 1, Appendix S8).

Two alleles were recognized by the modified Sommer pipeline only. Of these, one allele was not found in both replicates of any individual by AmpliSAS, while the other allele was marked as "suspicious sequence" by AmpliCHECK and not outputted in the final result from AmpliSAS. All alleles identified with AmpliSAS were also called using the modified Sommer pipeline.

Genotyping with AmpliSAS thus identified 16 alleles in total (Table 1), with the same unique variants genotyped using MHCI-SI and MHCI-DI. Three alleles found in an individual by MHCI-SI were however not called as alleles in the same individual by MHCI-DI, as they were found in only one of the two MHCI-DI amplicons by AmpliSAS. Apart from this discrepancy, there was no disagreement between the results from the two primer setups (see Appendix S9).

There was no deviation from Mendelian inheritance of offspring alleles in either of the two allele-calling pipelines.

3.2 | MHCIIβe2

All Ion Torrent MHCII amplicons (MHCII-SI and MHCII-DI) and 12 of the 16 MHCII-MiSeq amplicons had >5,000 reads after barcode splitting in both pipelines. Excluding one failed sample (one replicate MHCII-MiSeq run of individual 69390), the remaining MHCII-MiSeq amplicons had each assigned >2,400 reads at this stage.

3.2.1 | Modified pipeline from Sommer et al. (2013)

Analyzing the data using the modified pipeline from Sommer et al. (2013), 117 unique alleles were identified across all sequencing strategies. The 117 alleles translated into 105 unique amino acid sequences. Less than 5% of the alleles found in the offspring were not found in any of their parents (Table 2).

Fewer alleles were called in every individual using the MHCII-DI sequencing approach, than in the two other approaches (averagely 35.6, 23.0, and 32.6 alleles were called per individual for the MHCII-SI, MHCII-DI, and MHCII-MiSeq data, respectively). The number of alleles per individual called within the separate sequencing approaches ranged from 17 to 47. When combining all three sequencing approaches, a maximum of 25 alleles were genotyped per individual by all three approaches, while a maximum of 53 alleles were genotyped per individual in at least one approach.

3.2.2 | AmpliSAS pipeline

Across all eight samples and all three sequencing approaches, 114 unique alleles were called using AmpliSAS, which translated into 102 unique amino acid sequences. For this allele-calling pipeline, 7–10 (6.4%–12.5%) pedigree errors were found in each of the primer approaches.

On average, the MHCII-DI approach yielded 22.0 alleles per individual, while 27.8 alleles were scored per individual in both the MHCII-SI and the MHCII-MiSeq approaches. As with the modified Sommer pipeline, AmpliSAS also called fewer alleles using MHCII-DI than MHCII-SI for most individuals (in all individuals except two; on average 5.75 more alleles were scored per individual with MHCII-SI than with MHCII-DI).

TABLE 2 Results from genotyping eight bluethroats at MHC class II exon 2 (MHCII), using two primer strategies (single index [SI] and dual index [DI]) on Ion Torrent as well as being sequenced on Illumina MiSeq

	MHCII-SI	MHCII-DI	MHCII-MiSeq
Modified pipeline from Sommer et al. (2013)			
Reads per amplicon after jMHC (including only variants with ≥3 reads within any amplicon)	6,495–19,886	14,777–68,524	2,436–10,553
Total number of alleles	96	74	105
Alleles per individual	35.6 (29–47)	23.0 (17–31)	32.6 (23–41)
Unique amino acid sequences	84	65	93
Errors in pedigree	2.1%	4.3%	3.4%
AmpliSAS			
Reads per amplicon	>5,000	>5,000	2,877–>5,000
Reads per amplicon assigned to alleles	3,251–4,445	3,754–4,681	2,611–4,721
Total number of alleles	75	74	94
Alleles per individual	27.8 (20–39)	22.0 (16–30)	27.8 (21–34)
Alleles per amplicon	31.6 (22–42)	23.5 (16–31)	28.5 (21–35)
Unique amino acid sequences	68	64	82
Errors in replicates	12.1%	6.4%	8.4%
Errors in pedigree	6.4%	10.9%	12.5%

The datasets were genotyped through two pipelines; one using a modification of the pipeline published by Sommer et al. (2013) and one using the online tool AmpliSAS (Sebastian et al., 2016).

Within the separate sequencing approaches, 16–39 alleles were found per individual (Table 2). Genotyping rendered a maximum of 20 alleles per individual found in all three sequencing approaches, and a maximum of 53 alleles per individual were found in at least one approach when combining the results.

3.2.3 | Comparing the approaches

We established a "combined genotype" for each individual consisting of all unique MHCIIβe2 alleles that were called in at least one of the approaches for that respective individual. The number of alleles per individual in the "combined genotype" ranged from 35 to 56, with an average of 43.5. Averaged over all individuals, the MHCII-SI primer setup followed by allele calling with the modified Sommer pipeline was able to retrieve the highest percentage of the "combined genotype" (i.e., saturation rate), as compared to the other approaches (Figure 4). The dual-indexed approaches had the lowest saturation rate, while within every primer setup, the modified Sommer pipeline had higher saturation rate than when allele calling using AmpliSAS. The maximum number of unique alleles in a combined genotype (i.e., 56 alleles) implies a minimum of 28 MHCIIβe2 loci in the bluethroat (see Appendix S10).

All unique alleles across all individuals found with AmpliSAS were also found using the modified Sommer pipeline, except for one allele which was also the only allele lacking a cysteine residue in position 75 when translated—which suggests that it is a nonfunctional allele (see Appendix S11). Within each individual, on average 80.1% (MHCII-SI), 94.4% (MHCII-DI), and 91.9% (MHCII-MiSeq) of the alleles were called by both the modified Sommer pipeline and AmpliSAS. For the MiSeq data, all alleles genotyped for each individual by AmpliSAS were also called by the modified Sommer pipeline. The latter pipeline genotyped on average 4.88 more MiSeq alleles per individual than AmpliSAS.

For both allele-calling pipelines, fewest unique alleles were genotyped across all samples using the MHCII-DI data, while the highest number of unique alleles was found in the MiSeq dataset. Also in terms of alleles per individual, the MHCII-DI run rendered fewest alleles for both pipelines (Table 2 and Figure 4).

4 | DISCUSSION

In this study, we aimed to establish a robust genotyping method for MHC class I exon 3 (MHCIe3) and MHC class II exon 2 (MHCIIβe2) in bluethroats. Simultaneously, we intended to highlight possible differences in MHC genotyping resulting from different sequencing platforms, primer design, and bioinformatic allele-calling pipelines. For the hypervariable MHCIIβe2, both the number and the identity of alleles varied among the abovementioned approaches. More consistent genotypes were obtained when analyzing MHCIe3, in which mainly bioinformatic pipeline but not primer structure influenced the results. Our use of family data and replicates was advantageous in order to validate alleles. We thus recommend including such data when analyzing highly polymorphic markers.

4.1 | Sources of variation among strategies: platforms

One of the main challenges when genotyping variable multilocus systems like MHC is to be able to separate real alleles from artifacts (Babik et al., 2009; Lighten et al., 2014). As specific sequencing platforms have typical error profiles (Duke et al., 2015; Loman et al., 2012; Quail et al., 2012)—for example, homopolymer errors are more prominent for Ion Torrent than Illumina—different platforms could render different genotypes (Sebastian et al., 2016). Variant filtering in subsequent allele-calling pipelines is however supposed to eradicate sequencing errors. Yet, this process might not work perfectly and could potentially result in high-depth, in-frame artifactual variants incorrectly being called as alleles. Nevertheless, the low number of alleles not following Mendelian inheritance makes this unlikely to be the only explanation for the observed differences in genotypes. Another reason for the discrepancy could be platform-specific adapter motifs in the primers (mechanisms for unequal amplification discussed below).

4.2 | Sources of variation among strategies: primers

The differences between primer approaches are better elucidated when comparing Ion Torrent runs, which are not confounded by differences in sequencing platforms. It is important to note that the MHCIe3 runs with different primer structure (MHCI-SI and MHCI-DI) yielded identical results within each allele-calling pipeline (except one individual genotyped with AmpliSAS; discussed later), while this was not the case for MHCIIβe2 (MHCII-SI and MHCII-DI).

The underlying mechanisms that possibly explain the observed MHCIIβe2 genotype discrepancies may involve differential amplification of alleles (e.g., Sommer et al., 2013). A "tail" of spacer/barcode/adapter in the primer sequence would preferentially amplify variants that have complementary bases to this tail outside the gene target. Such a tail is applied on both forward and reverse primers in the DI approach, while in contrast, the SI primers have only the barcode attached, and only to the forward primers (Ion Torrent adapters were ligated onto the amplicons after MHC amplification). As a result, we would expect a stronger effect and fewer alleles to be amplified using the DI approach. Indeed, we observed fewer alleles called for every individual by both bioinformatic pipelines using MHCII-DI than MHCII-SI, when disregarding alleles that exhibited Mendelian errors.

Amplified PCR fragments will have complementary sequences to the whole primer (including the tail), in contrast to the template, in which only the gene target is complementary to the primer sequence. The annealing affinity of the primers in the PCR will thus be higher to amplified fragments compared to template sequences. Because the DI primer setup consists of a longer tail on both forward and reverse primers, this "affinity effect" may create more bias and potential allelic dropouts in the MHCII-DI amplifications than in the MHCII-SI amplifications.

Lastly, the probability of secondary structure formation in the DI primer sequence is higher due to the longer primer tail. The combination of these effects might explain the lower number of observed alleles for the MHCII-DI approach compared to the MHCII-SI approach.

Our results demonstrate that the use of phusion primers can create allelic dropout in PCR amplifications of multilocus targets. Researchers should be aware of this potential pitfall and address this issue when sequencing polymorphic multilocus regions such as the MHC.

4.3 | Sources of variation among strategies: allele-calling pipelines

Within each MHCIIβe2 primer setup, we found distinct genotypes depending on the allele-calling pipelines used. First, one possible cause to the disparity between the modified Sommer pipeline and AmpliSAS is the clustering in the AmpliSAS algorithm. While all variants passing the filters are called as alleles in the modified Sommer pipeline, AmpliSAS cluster similar variants based on platform-specific error rates and relative frequencies of variants clustered together. Hence, the fewer alleles scored using AmpliSAS could be caused by erroneous clustering of low-frequency, true alleles to other true alleles. The discrepancy could also arise from high-frequency artifacts incorrectly being called as alleles in the modified Sommer pipeline. The latter is unlikely to be a general explanation because all except two of the 39 MHCIIβe2 alleles called using the modified Sommer pipeline and not AmpliSAS showed Mendelian inheritance. However, repeatable errors could be a cause of this pattern. Checking the "=1 bp"-variants in offspring revealed that many were instances where the "=1 bp" variant was found within one parent which lacked the "source" variant, while the "source" variant was found within the other parent which lacked the "=1 bp" variant (data not shown). Repeatable errors are thus likely not a major problem in the modified Sommer pipeline, although we cannot dismiss it completely. Also, within each MHCIIβe2 primer setup, not all of the alleles called only when using the modified Sommer pipeline had high sequence similarities to other alleles (see Appendix S12). Clustering of similar alleles may thus not account for all the instances in which AmpliSAS genotyped fewer alleles.

Second, the use of replicates could be an additional explanation for the higher number of alleles called when using the modified Sommer pipeline. In the AmpliSAS pipeline, we scored a variant as an allele in an individual if the AmpliSAS program genotyped it as an allele for both replicates. Variants found in only one of the replicates were however treated more carefully in the modified Sommer pipeline. Here, such variants were called as alleles if the variant in question was more than one base pair different from a more frequent variant within the same amplicon and concurrently found within other amplicons of the same family group. The underlying rationale is that these ">1 bp variants" are less likely to be sequencing errors, but if they are, it is unlikely that the same artifact is found within multiple amplicons from the same family group. Variants that are only one base pair different from a more frequent variant are managed more strictly and are required to be present in both replicates of an individual in order to be called as allele.

The use of replicates can also affect the interpretation of chimeras, and hence the number of alleles called. Whereas a variant that is scored as a putative allele in one amplicon but marked as a chimera in the replicate will be called as an allele in the modified Sommer pipeline, this variant will not be genotyped by AmpliSAS. This is because chimeras are removed from the output from AmpliSAS, and as the variant is then lacking from one of the two replicates, it will not be called as an allele.

While we reduced the number of PCR cycles to 25 in order to minimize PCR artifacts, other actions could be taken, for instance a prolonged elongation step or the introduction of a reconditioning step (Lenz & Becker, 2008). However, as we are comparing approaches that all apply the same PCR-protocol (see Appendix S5), this is not further evaluated here.

Third, there could be an effect of difference in coverage levels required for the two allele-calling pipelines. AmpliSAS is based on a subsampling of 5,000 reads from each amplicon as default, while the modified Sommer pipeline takes all reads within an amplicon into account. Biedrzycka et al. (2017) achieved high repeatability and reliability when using 5,000 reads per amplicon for genotyping a sedge warbler MHCI-dataset (with complexity similar to the MHCIIβe2 dataset in our study) by AmpliSAS, justifying the use of the default subsampling value. Implementing a minimum amplification efficiency of 0.2 (as in Biedrzycka et al. (2017)) and a maximum of 47 alleles per individual to the information in Figure S6 and Table S4 in Sommer et al. (2013), a minimum of 2,456 reads per amplicon are required to determine a complete genotype with at least three reads per allele (99.9% confidence level). As all MHCIIβe2 amplicons sequenced on the Ion Torrent initially had >5,000 reads each, while all except one (<500 reads) MHCII-MiSeq and MHCI amplicons had >2,800 reads, we chose to only exclude the one failed MiSeq amplicon. While the remaining amplicons thus would have sufficient coverage for genotyping using the modified Sommer pipeline, we also chose to keep these amplicons for AmpliSAS, because of the additional strength we get from including family data and replicates of each individual. The missing of a MHCIIβe2 allele in AmpliSAS could thus be due to low coverage, as three of the eight individuals had one MHCII-MiSeq amplicon replicate with coverage <5,000 (2,877–4,399) reads. Indeed, eight of the 11 MiSeq alleles found by the modified Sommer pipeline and not by AmpliSAS could be explained in this manner, where the alleles missing when genotyping in AmpliSAS are found in high-coverage amplicons but not in their lower-coverage replicates (data not shown). In other words, this implies that higher sequencing depth is required for the AmpliSAS pipeline compared to the modified Sommer pipeline in order to obtain the same accuracy when genotyping highly polymorphic loci. Bluethroat MHCIe3, which has relatively few loci (i.e., four; O'Connor et al., 2016), is likely not affected by coverage differences to the same extent as the polymorphic bluethroat MHCIIβe2. This is in line with Razali, O'Connor, Drews, Burke, and Westerdahl (2017), who found that MiSeq and 454 sequencing provided equal results despite differences in read depths when genotyping amplicons with low diversity, while the results were less consistent in amplicons with higher diversity.

4.4 | Comparing the approaches

Genotyping MHCIe3 yielded mainly consistent results across all approaches. Still, the modified Sommer pipeline revealed two additional unique alleles compared to the results from AmpliSAS. These alleles showed neither any deviation from Mendelian inheritance nor any signs of nonfunctionality, and we assume they are false negatives in the AmpliSAS pipeline. Further, within each allele-calling pipeline, only one individual exhibited differences in MHCI-SI and MHCI-DI genotypes (genotyped using AmpliSAS). The three "missing" alleles in the MHCI-DI genotype were however called in one of the MHCI-DI replicates and found in the unfiltered AmpliSAS results for both replicates. The lack of these MHCI-DI alleles could thus be caused by erroneous filtering in AmpliSAS.

For the hypervariable MHCIIβe2, the results were less consistent. Across all primer approaches, four unique alleles were found only when using the modified Sommer pipeline, while one unique allele was found only using AmpliSAS. It is worth mentioning that this latter allele was the only allele lacking the important cysteine residue in position 75 and hence is likely a pseudogene or an artifact. As cysteine residues in position 10 and 75 are included in the filtering steps of the modified Sommer pipeline, this allele would not be retained in the outputted genotype.

Within each primer setup, more MHCIIβe2 alleles were scored using the modified Sommer pipeline than AmpliSAS. This is evident also from the saturation rate plot (Figure 4; see also Appendix S13), where a higher percentage of the combined genotypes (i.e., alleles called within an individual using at least one approach) were called using the modified Sommer pipeline than using AmpliSAS. Most of these alleles are expected to be true positives, as almost all alleles called in the offspring also were genotyped for one or both of their parents (Table 2). Furthermore, there were fewer errors in pedigree for genotyping with the modified Sommer pipeline than with AmpliSAS. These numbers are however not directly comparable between the allele-calling pipelines, because of the use of family information to facilitate genotyping in the modified Sommer pipeline. The "errors in pedigree" (Table 2) will thus be biased towards the modified Sommer pipeline as compared to AmpliSAS. Yet, the lower number of errors in pedigree and the higher saturation rate still suggest that the modified Sommer pipeline could render more consistent and comprehensive results than AmpliSAS. The more automated allele calling through AmpliSAS is however both faster and less prone to human mistakes, which need to be taken into account when deciding upon which approach to use. Based on our results, we still recommend the use of the modified Sommer pipeline on highly polymorphic systems, especially when family data are available and the sequencing is not ultradeep (>5,000 reads in every amplicon).

Correspondingly, MHCII-SI (genotyped using the modified Sommer pipeline) had higher saturation rate and lower percentage of errors in pedigree compared to the other approaches and could be the preferred approach. However, single indexing requires substantially more barcodes when multiplexing a large number of individuals, and sequencing using dual-indexed primers on Illumina MiSeq as in this study (Figure 1) could thus be a cost-efficient option.

The difficulty of correctly genotyping highly polymorphic loci like bluethroat MHCIIβe2 also raises the issue of balancing false negatives against false positives. An approach that calls more alleles will likely also score some artifacts as alleles, while a "stricter" approach will be more prone to fail to genotype true alleles. The relative importance of this will likely be dependent on the research question (e.g., whether the study is on diversity, or associations between pathogens and MHC alleles). Although false positives such as repeatable errors cannot be completely ruled out, we believe that consistency in family data is an overall strength for the approach used. Hence, the opportunity for validation of alleles provided by family data is of great value, and using an approach that minimizes the errors in inheritance pattern could guide the choice of method.

4.5 | Polymorphism levels in MHCIe3 and MHCIIβe2 in bluethroats

For MHC class I, we detected maximum eight alleles per individual, implying minimum four MHCIe3 loci in bluethroats. This is in accordance with the results from O'Connor et al. (2016), who also reported four loci in the species, using a dual index approach and two different primer combinations. The same study revealed considerable diversity in the number of MHCIe3 loci across passerines, with bluethroats (four loci) and willow warblers (19 loci) at the extreme ends. This strengthens our findings of relative low complexity at MHCI in bluethroats, and that the genotyping at these loci likely is robust against variation in primer design and allele-calling approaches in the species.

The complexity recognized at MHCIIβe2 in this study is in stark contrast to the results from MHCIe3. Our study supports the findings of both Anmarkrud et al. (2010) and Gohli et al. (2013), where high levels of MHCIIβe2 polymorphism were detected in the bluethroats through cloning and Sanger sequencing. Although likely underestimated due to technical limitations, Anmarkrud et al. (2010) identified a minimum of 11 functional MHCIIβe2 loci. In this study, when combining all strategies, up to 56 alleles and thus a minimum of 28 MHCIIβe2 loci were described for one individual, testifying to the incredible diversity at this marker.

5 | CONCLUSION

Our results reveal that different genotyping strategies yield similar genotypes in bluethroat MHCIe3, a system with relatively low polymorphism. In contrast, caution needs to be exercised when sequencing highly complex markers such as the bluethroat MHCIIβe2. For bluethroat MHCIIβe2, our results demonstrate that genotyped alleles will be biased according to both primer design and allele-calling pipeline. Consequently, comparisons of results across approaches and studies are error prone in this polymorphic marker. However, the use of family data and replicates lends support to results found within each strategy and prove to be especially valuable for validation of alleles in the complex MHCIIβe2. As such, the methodology described herein could be useful for exploration of ecological and evolutionary relevant hypotheses relative to MHC variation, even though it does not necessarily describe the true repertoire of alleles within each individual.

ACKNOWLEDGMENTS

We are grateful to Lisbeth Thorbek for expert technical assistance in the laboratory, Camilla LoCascio Sætre for performing the DNA extraction, and to Jacek Radwan and an anonymous reviewer for valuable comments to the manuscript. The Norwegian Sequencing Center conducted the Illumina MiSeq sequencing. The project was funded by an internal grant at Natural History Museum, Oslo, Norway.

CONFLICT OF INTERESTS

The authors declare no conflict of interests.

AUTHOR CONTRIBUTIONS

SLR, JAA, AJ, JTL designed the study. JAA performed the laboratory experiments. SLR performed the bioinformatics analyses and drafted the manuscript. All authors reviewed the manuscript and approved the final version.

DATA ACCESSIBILITY

MHC alleles uncovered in this study have been imported to GenBank (accession nos. MF769960–MF769977 (MHCIe3) and accession nos. MF769842–MF769959 (MHCIIβe2)). All sequence data used in this study are uploaded to the NCBI Sequence Read Archive under BioProject ID: PRNA400123. Voucher specimen accession nos. and barcodes: See Appendix S1. Each voucher accession no. is searchable via the online Collection Explorer: http://nhmo-birds.collectionexplorer.org/accession.aspx.

ORCID

Silje L. Rekdal http://orcid.org/0000-0003-4593-1709

REFERENCES

Alcaide, M., Liu, M., & Edwards, S. V. (2013). Major histocompatibility complex class I evolution in songbirds: Universal primers, rapid evolution and base compositional shifts in exon 3. *PeerJ*, *1*, e86. https://doi.org/10.7717/peerj.86

Alcaide, M., Munoz, J., Martínez-de la Puente, J., Soriguer, R., & Figuerola, J. (2014). Extraordinary MHC class II B diversity in a non-passerine, wild bird: The Eurasian Coot *Fulica atra* (Aves: Rallidae). *Ecology and Evolution*, *4*, 688–698. https://doi.org/10.1002/ece3.974

Anmarkrud, J., Johnsen, A., Bachmann, L., & Lifjeld, J. (2010). Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes. *Journal of Evolutionary Biology*, *23*, 1206–1217. https://doi.org/10.1111/j.1420-9101.2010.01999.x

Babik, W. (2010). Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources*, *10*, 237–251. https://doi.org/10.1111/j.1755-0998.2009.02788.x

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. https://doi.org/10.1111/j.1755-0998.2009.02622.x

Biedrzycka, A., Sebastian, A., Migalska, M., Westerdahl, H., & Radwan, J. (2017). Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Molecular Ecology Resources*, *17*, 642–655. https://doi.org/10.1111/1755-0998.12612

Bollmer, J. L., Dunn, P. O., Freeman-Gallant, C. R., & Whittingham, L. A. (2012). Social and extra-pair mating in relation to major histocompatibility complex variation in common yellowthroats. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*, 4778–4785 rspb20121885. https://doi.org/10.1098/rspb.2012.1885

Canal, D., Alcaide, M., Anmarkrud, J. A., & Potti, J. (2010). Towards the simplification of MHC typing protocols: Targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*. *BMC Research Notes*, *3*, 236. https://doi.org/10.1186/1756-0500-3-236

Duke, J., Lind, C., Mackiewicz, K., Ferriola, D., Papazoglou, A., Derbeneva, O., & Monos, D. (2015). Towards allele-level human leucocyte antigens genotyping – Assessing two next-generation sequencing platforms: Ion Torrent Personal Genome Machine and Illumina MiSeq. *International Journal of Immunogenetics*, *42*, 346–358. https://doi.org/10.1111/iji.12213

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*, 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

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. https://doi.org/10.1016/S0169-5347(98)01416-5

Ekblom, R., Grahn, M., & Höglund, J. (2003). Patterns of polymorphism in the MHC class II of a non-passerine bird, the great snipe (*Gallinago media*). *Immunogenetics*, *54*, 734–741.

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, *2*, 6. https://doi.org/10.1186/2049-2618-2-6

Gaigher, A., Burri, R., Gharib, W., Taberlet, P., Roulin, A., & Fumagalli, L. (2016). Family-assisted inference of the genetic architecture of major histocompatibility complex variation. *Molecular Ecology Resources*, *16*, 1353–1364. https://doi.org/10.1111/1755-0998.12537

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. https://doi.org/10.1186/1471-2164-11-296

Gohli, J., Anmarkrud, J. A., Johnsen, A., Kleven, O., Borge, T., & Lifjeld, J. T. (2013). Female promiscuity is positively associated with neutral and selected genetic diversity in passerine birds. *Evolution*, *67*, 1406–1419. https://doi.org/10.1111/evo.12045

Grogan, K. E., McGinnis, G. J., Sauther, M. L., Cuozzo, F. P., & Drea, C. M. (2016). Next-generation genotyping of hypervariable loci in many individuals of a non-model species: Technical and theoretical implications. *BMC Genomics*, *17*, 204. https://doi.org/10.1186/s12864-016-2503-y

Janeway, C. A., Travers, P., Walport, M., & Shlomchik, M. J. (2001). The major histocompatibility complex and its functions. In C. A. Janeway, P. Travers, M. Walport & M. J. S (Eds.), *Immunobiology: The immune system in health and disease* (5th ed.). New York, NY: Garland Science. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK27156/. ISBN-10: 0-8153-3642-X

Kanagawa, T. (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering*, *96*, 317–323. https://doi.org/10.1016/S1389-1723(03)90130-7

Kaufman, J., Milne, S., Göbel, T. W., Walker, B. A., Jacob, J. P., Auffray, C., & Beck, S. (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, *401*, 923–925. https://doi.org/10.1038/44856

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*, 1870–1874. https://doi.org/10.1093/molbev/msw054

Lenz, T. L., & Becker, S. (2008). Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci—Implications for evolutionary analysis. *Gene*, *427*, 117–123. https://doi.org/10.1016/j.gene.2008.09.013

Lighten, J., Oosterhout, C., & Bentzen, P. (2014). Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology*, *23*, 3957–3972. https://doi.org/10.1111/mec.12843

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, *30*, 434–439. https://doi.org/10.1038/nbt.2198

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

O'Connor, E., Strandh, M., Hasselquist, D., Nilsson, J. Å., & Westerdahl, H. (2016). The evolution of highly variable immunity genes across a passerine bird radiation. *Molecular Ecology*, *25*, 977–989. https://doi.org/10.1111/mec.13530

Pawluczyk, M., Weiss, J., Links, M. G., Aranguren, M. E., Wilkinson, M. D., & Egea-Cortines, M. (2015). Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. *Analytical and Bioanalytical Chemistry*, *407*, 1841–1848. https://doi.org/10.1007/s00216-014-8435-y

Piertney, S., & Oliver, M. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, *96*, 7–21. https://doi.org/10.1038/sj.hdy.6800724

Polz, M. F., & Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, *64*, 3724–3730.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., & Gu, Y. (2012). A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, *13*, 341. https://doi.org/10.1186/1471-2164-13-341

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org. ISBN: 3-900051-07-0

Razali, H., O'Connor, E., Drews, A., Burke, T., & Westerdahl, H. (2017). A quantitative and qualitative comparison of illumina MiSeq and 454 amplicon sequencing for genotyping the highly polymorphic major histocompatibility complex (MHC) in a non-model species. *BMC Research Notes*, *10*, 346. https://doi.org/10.1186/s13104-017-2654-1

Sebastian, A., Herdegen, M., Migalska, M., & Radwan, J. (2016). amplisas: A web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources*, *16*, 498–510. https://doi.org/10.1111/1755-0998.12453

Sommer, S., Courtiol, A., & Mazzoni, C. J. (2013). MHC genotyping of non-model organisms using next-generation sequencing: A new methodology to deal with artefacts and allelic dropout. *BMC Genomics*, *14*, 542. https://doi.org/10.1186/1471-2164-14-542

Strandh, M., Lannefors, M., Bonadonna, F., & Westerdahl, H. (2011). Characterization of MHC class I and II genes in a subantarctic seabird, the blue petrel, *Halobaena caerulea* (Procellariiformes). *Immunogenetics*, *63*, 653–666. https://doi.org/10.1007/s00251-011-0534-8

Stuglik, M. T., Radwan, J., & Babik, W. (2011). jMHC: Software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Molecular Ecology Resources*, *11*, 739–742. https://doi.org/10.1111/j.1755-0998.2011.02997.x

Stutz, W. E., & Bolnick, D. I. (2014). Stepwise threshold clustering: A new method for genotyping MHC loci using next-generation sequencing technology. *PLoS One*, *9*, e100587. https://doi.org/10.1371/journal.pone.0100587

Suzuki, M. T., & Giovannoni, S. J. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology*, *62*, 625–630.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*, 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Westerdahl, H. (2007). Passerine MHC: Genetic variation and disease resistance in the wild. *Journal of Ornithology*, *148*, 469–477. https://doi.org/10.1007/s10336-007-0230-5

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag. ISBN 3319242776. https://doi.org/10.1007/978-0-387-98141-3

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. https://doi.org/10.1186/1471-2148-10-395

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Rekdal SL, Anmarkrud JA, Johnsen A, Lifjeld JT. Genotyping strategy matters when analyzing hypervariable major histocompatibility complex-Experience from a passerine bird. *Ecol Evol*. 2017;00:1–13. https://doi.org/10.1002/ece3.3757

Appendix legends

Rekdal et al.: Genotyping strategy matters when analyzing hypervariable MHC

Appendix S1: Bluethroat individuals

Overview over the individuals studied. The accession numbers refer to DNA Bank accession numbers at the Natural History Museum of Oslo, Norway.

Appendix S2: Terms

Explanation of central terms as they are used in this study.

Appendix S3: Primer sequences

The sequences and lengths of the primers used in this study.

Appendix S4: Primer binding sites

The binding sites of the MHCI and MHCII primers. Modified figures from Alcaide et al. (2013) and Canal et al. (2010).

Appendix S5: Methods: sequencing details

Details of the amplification and sequencing of MHC class I exon 3 and MHC class II exon.

Appendix S6: Methods: allele calling details

Details of the allele calling of MHC class I exon 3 and MHC class II exon 2, based on the pipeline developed by Sommer et al. (2013) and the software AmpliSAS (Sebastian et al. 2016).

Appendix S7: Genotypes, MHCIe3

Genotypes for the eight individuals for MHC class I exon 3.

Appendix S8: Sequence logo, MHCIe3

Sequence logo created using the online tool WebLogo (http://weblogo.berkeley.edu/logo.cgi; Crooks et al. (2004)) for MHCI. Unique MHCI alleles identified among all strategies were aligned to previously described bluethroat MHCI alleles (GenBank accession number KU169737-KU169747; O'Connor et al. (2016)) by ClustalW (Thompson et al. 1994) in MEGA7 (Kumar et al. 2016), and translated correspondingly. Antigen-binding sites, as indicated as sites where $d_N/d_S > 1$ in passerine MHCI exon 3 sequences by Balakrishnan et al. (2010), are marked with *. One sequence (LuSv_MHCIe3_SR_015) was omitted due to ambiguity in the placement of a deletion when aligned to the other alleles in MEGA.

Appendix S9: Network MHCIe3

A minimum spanning network (Bandelt et al. 1999) of the unique MHCI alleles identified in this study, made through the online software PopArt (http://popart.otago.ac.nz). The circles represent unique alleles found among the strategies, and their sizes are corresponding to the number of strategies they are found within. Additionally, they are color-coded according to the scheme in the lower right corner, implying in which strategies the alleles are found. The numbers are representing substitutions between alleles.

Appendix S10: Genotypes, MHCIIβe2

Genotypes for the eight individuals for MHC class II exon 2.

Appendix S11: Sequence logo, MHCIIβe2

Sequence logo created using the online tool WebLogo (http://weblogo.berkeley.edu/logo.cgi; Crooks et al. (2004)) for MHCII. Unique MHCII alleles identified among all strategies were aligned to previously described bluethroat MHCII alleles (GenBank accession number HQ539575-HQ539614;

Gohli et al. (2013)) by ClustalW (Thompson et al. 1994) in MEGA7 (Kumar et al. 2016), and translated correspondingly. Antigen-binding sites, as indicated as sites where $d_N/d_S > 1$ in passerine MHCII exon 2 sequences by Balakrishnan et al. (2010), are marked with *.

Appendix S12: Network MHCIIβe2

A minimum spanning network (Bandelt et al. 1999) of the unique MHCII alleles identified in this study, made through the online software PopArt (http://popart.otago.ac.nz). The circles represent unique alleles found among the strategies, and their sizes are corresponding to the number of strategies they are found within. Additionally, they are color-coded according to the scheme in the lower right corner, implying in which strategies the alleles are found. The numbers are representing substitutions between alleles.

Appendix S13: Venn diagram MHCIIβe2

Venn diagram showing the number of unique MHC II alleles found within each strategy, and the number of overlapping alleles among strategies. Made in R (version 3.2.5, R Core Team 2016), using the package VennDiagram (Chen & Boutros 2011).

References:

Alcaide, M., Liu, M., & Edwards, S. V. (2013). Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. *PeerJ,* 1, e86. doi: 10.7717/peerj.86

Balakrishnan, C. N., Ekblom, R., Völker, M., Westerdahl, H., Godinez, R., Kotkiewicz, H., . . . Warren, W. C. (2010). Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *Bmc Biology,* 8, 1.

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution,* 16, 37-48.

Canal, D., Alcaide, M., Anmarkrud, J. A., & Potti, J. (2010). Towards the simplification of MHC typing protocols: targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*. *BMC research notes,* 3, 236. doi: 10.1186/1756-0500-3-236

Chen, H., & Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC bioinformatics,* 12, 35.

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Research,* 14, 1188-1190.

Gohli, J., Anmarkrud, J. A., Johnsen, A., Kleven, O., Borge, T., & Lifjeld, J. T. (2013). Female promiscuity is positively associated with neutral and selected genetic diversity in passerine birds. *Evolution,* 67, 1406-1419. doi: 10.1111/evo.12045

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution,* 33, 1870-1874. doi: 10.1093/molbev/msw054

O'Connor, E., Strandh, M., Hasselquist, D., Nilsson, J. Å., & Westerdahl, H. (2016). The evolution of highly variable immunity genes across a passerine bird radiation. *Molecular Ecology*, 977-989. doi: 10.1111/mec.13530

Sebastian, A., Herdegen, M., Migalska, M., & Radwan, J. (2016). amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources,* 16, 498-510. doi: 10.1111/1755-0998.12453

Sommer, S., Courtiol, A., & Mazzoni, C. J. (2013). MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC genomics,* 14, 1. doi: 10.1186/1471-2164-14-542

Team, R. C. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org. ISBN: 3-900051-07-0

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22, 4673-4680. doi: 10.1093/nar/22.22.4673

Appendix S1: Bluethroat individuals

Identifyer this study	Accession number	Ring number	Sex	Life stage	Attribute	Family	Date collected
69390	NHMO-BI-69390		Unknown	Pullus	Chick	A	June 21th 2014
69391	NHMO-BI-69391		Unknown	Pullus	Chick	A	June 21th 2014
69392	NHMO-BI-69392	ED49187	Female	3K+	Mother	A	June 21th 2014
69393	NHMO-BI-69393	ED49188	Male	3K+	Father	A	June 21th 2014
69524	NHMO-BI-69524	ED49606	Female	3K+	Mother	B	July 1st 2014
69525	NHMO-BI-69525		Unknown	Pullus	Chick	B	July 1st 2014
69526	NHMO-BI-69526		Unknown	Pullus	Chick	B	July 1st 2014
69528	NHMO-BI-69528	ED49179	Male	2K+	Father	B	July 1st 2014

MHCI

	forward barcode amplicon A MHCI-SI	forward barcode amplicon B MHCI-SI
69390	TTCGTGATTC	TCTATTCGTC
69391	TTCCGATAAC	AGGCAATTGC
69392	TGAGCGGAAC	TTAGTCGGAC
69393	CTGACCGAAC	CAGATCCATC
69524	TCCTCGAATC	TCGCAATTAC
69525	TAGGTGGTTC	TTCGAGACGC
69526	TCTAACGGAC	TGCCACGAAC
69528	TTGGAGTGTC	AACCTCATTC

	forward barcode amplicon A MHCI-DI	reverse barcode amplicon A MHCI-DI	forward barcode amplicon B MHCI-DI	reverse barcode amplicon B MHCI-DI
69390	TACCAAGATC	GTTCTCCTTA	TACCAAGATC	GATCTTGGTA
69391	CAGAAGGAAC	GTTCTCCTTA	CAGAAGGAAC	GATCTTGGTA
69392	CTGCAAGTTC	GTTCTCCTTA	CTGCAAGTTC	GATCTTGGTA
69393	TAAGGAGAAC	GAATCCTCTT	TAAGGAGAAC	GTTCCTTCTG
69524	AAGAGGATTC	GAATCCTCTT	AAGAGGATTC	GTTCCTTCTG
69525	TACCAAGATC	GAATCCTCTT	TACCAAGATC	GTTCCTTCTG
69526	CAGAAGGAAC	GAATCCTCTT	CAGAAGGAAC	GTTCCTTCTG
69528	CTGCAAGTTC	GAATCCTCTT	CTGCAAGTTC	GTTCCTTCTG

MHCII

	forward barcode amplicon A MHCII-SI	forward barcode amplicon B MHCII-SI
69390	AAGAGGATTC	TCTAACGGAC
69391	TACCAAGATC	TTGGAGTGTC
69392	CAGAAGGAAC	TCTAGAGGTC
69393	CTGCAAGTTC	TCTGGATGAC
69524	TTCGTGATTC	TCTATTCGTC
69525	TTCCGATAAC	AGGCAATTGC
69526	TGAGCGGAAC	TTAGTCGGAC
69528	CTGACCGAAC	CAGATCCATC

	forward barcode amplicon A MHCII-DI	reverse barcode amplicon A MHCII-DI	forward barcode amplicon B MHCII-DI	reverse barcode amplicon B MHCII-DI
69390	AAGAGGATTC	GTTACCTTAG	AAGAGGATTC	GAATCCTCTT
69391	TACCAAGATC	GTTACCTTAG	TACCAAGATC	GAATCCTCTT
69392	CAGAAGGAAC	GTTACCTTAG	CAGAAGGAAC	GAATCCTCTT
69393	CTAAGGTAAC	GTTCTCCTTA	CTAAGGTAAC	GATCTTGGTA
69524	TAAGGAGAAC	GTTCTCCTTA	TAAGGAGAAC	GATCTTGGTA
69525	AAGAGGATTC	GTTCTCCTTA	AAGAGGATTC	GATCTTGGTA
69526	TACCAAGATC	GTTCTCCTTA	TACCAAGATC	GATCTTGGTA
69528	CAGAAGGAAC	GTTCTCCTTA	CAGAAGGAAC	GATCTTGGTA

	forward barcode amplicon A MHCII-MiSeq	reverse barcode amplicon A MHCII-MiSeq	forward barcode amplicon B MHCII-MiSeq	revserse barcode amplicon B MHCII-MiSeq
69390	CCATCACATAGG	CCTAAACTACGG	CCATCACATAGG	GTTACGTGGTTG
69391	GTGGTATGGGAG	CCTAAACTACGG	GTGGTATGGGAG	GTTACGTGGTTG
69392	ACTTTAAGGGTG	CCTAAACTACGG	ACTTTAAGGGTG	GTTACGTGGTTG
69393	GAGCAACATCCT	CCTAAACTACGG	GAGCAACATCCT	GTTACGTGGTTG
69524	TGTTGCGTTTCT	CCTAAACTACGG	TGTTGCGTTTCT	GTTACGTGGTTG
69525	ATGTCCGACCAA	CCTAAACTACGG	ATGTCCGACCAA	GTTACGTGGTTG
69526	CCTAAACTACGG	TGCAGATCCAAC	CCTAAACTACGG	TACCGCCTCGGA
69528	TGCAGATCCAAC	TGCAGATCCAAC	TGCAGATCCAAC	TACCGCCTCGGA

Appendix S2: Terms

Term used in this paper	Explanation
Read	A single, not necessarily unique, DNA-sequence derived from sequencing
Amplicon	DNA fragments amplified from a single marker in one PCR, from a single individual with a unique sequence tag
Variant	Cluster of identical reads
Allele	Variants that are passing all filters, and are believed to be a protein coding MHC sequence
Artefact	Variants that are not called as alleles in an individual, due to sequencing errors, shift in reading frame *et cetera*
Replicate	Replicates are amplicons derived from the same DNA extract, but that are amplified in independent PCRs

Appendix S3: Primer sequences

IonTorrent Single Index

General design	Forward:	5’-Barcode F| Barcode Adapter |Primer F-3’				
	Reverse:	5’-Primer R-3’				
Barcode adapter:	GAT					***3 bp***
Primer:	**MHCI:**	Target forward:	MhcPasCI-FW	CSCSCAGGTCTSCACAC	Alcaide *et al* . 2013	***17 bp***
		Target reverse:	MhcPASCI-RV	CWCARKAATTCTGYTCHCACC	Alcaide *et al* . 2013	***21 bp***
	MHCII:	Target forward:	MHCIIFihy-E2CF	CCGTGTCCTGCACACACAGC	Canal *et al* . 2010	***20 bp***
		Target reverse:	MHCIIFIhy-E2CR	GGGACASGCTCTGCCCCG	Canal *et al* . 2010	***18 bp***
Barcode 1	CTAAGGTAAC					***10 bp***
Barcode 2	TAAGGAGAAC					
Barcode 3	AAGAGGATTC					
Barcode 4	TACCAAGATC					
Barcode 5	CAGAAGGAAC					
Barcode 6	CTGCAAGTTC					
Barcode 7	TTCGTGATTC					
Barcode 8	TTCCGATAAC					
Barcode 9	TGAGCGGAAC					
Barcode 10	CTGACCGAAC					
Barcode 11	TCCTCGAATC					
Barcode 12	TAGGTGGTTC					
Barcode 13	TCTAACGGAC					
Barcode 14	TTGGAGTGTC					
Barcode 15	TCTAGAGGTC					
Barcode 16	TCTGGATGAC					
Barcode 17	TCTATTCGTC					
Barcode 18	AGGCAATTGC					
Barcode 19	TTAGTCGGAC					
Barcode 20	CAGATCCATC					

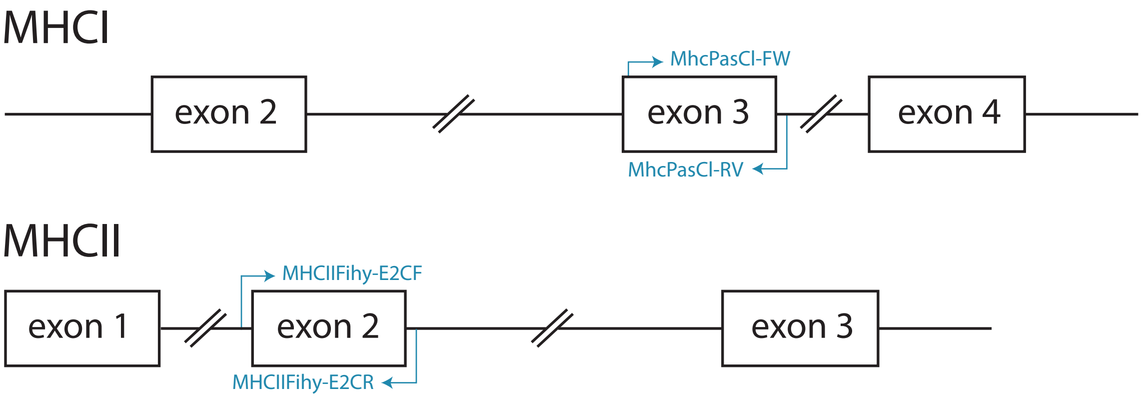
Ion Torrent Dual Index

General design	Forward:	5’-A adapter|Barcode F| Barcode Adapter |Spacer F|Primer F-3’				
	Reverse:	5’-trP1 adapter|Barcode R|Barcode adapter|Spacer R|Primer R-3’				
A adapter:	CCATCTCATCCCTGCGTGTCTCCGACTCAG					***30 bp***
trP1 adapter:	CCTCTCTATGGGCAGTCGGTGAT					***23 bp***
Barcode adapter:	GAT					***3 bp***
SpacerF	CCTGTGG					***7 bp***
SpacerR	CACTTCT					***7 bp***
Primer:	**MHCI:**	Target forward:	MhcPasCI-FW	CSCSCAGGTCTSCACAC	Alcaide *et al* . 2013	***17 bp***
		Target reverse:	MhcPASCI-RV	CWCARKAATTCTGYTCHCACC	Alcaide *et al* . 2013	***21 bp***
	MHCII:	Target forward:	MHCIIFihy-E2CF	CCGTGTCCTGCACACACAGC	Canal *et al* . 2010	***20 bp***
		Target reverse:	MHCIIFIhy-E2CR	GGGACASGCTCTGCCCCG	Canal *et al* . 2010	***18 bp***
Barcode F1	CTAAGGTAAC					***10 bp***
Barcode F2	TAAGGAGAAC					
Barcode F3	AAGAGGATTC					
Barcode F4	TACCAAGATC					
Barcode F5	CAGAAGGAAC					
Barcode F6	CTGCAAGTTC					
Barcode R1	GTTACCTTAG					
Barcode R2	GTTCTCCTTA					
Barcode R3	GAATCCTCTT					
Barcode R4	GATCTTGGTA					
Barcode R5	GTTCCTTCTG					

MiSeq

General design	Forward:	5’-Illumina Linker Sequence F|Barcode F|Heterogeneity Spacer F|Primer F-3’				
	Reverse:	5’-Illumina Linker Sequence R|Barcode R|Heterogeneity Spacer R|Primer R-3’				
Illumina Linker Sequence F:	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT					***58 bp***
Illumina Linker Sequence R:	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT					***58 bp***
Primer:	**MHCII:**	Target forward:	MHCIIFihy-E2CF	CCGTGTCCTGCACACACAGC	Canal *et al* . 2010	***20 bp***
		Target reverse:	MHCIIFIhy-E2CR	GGGACASGCTCTGCCCCG	Canal *et al* . 2010	***18 bp***
Barcode + Heterogenic spacer F1	CCTAAACTACGG					***12 bp + 0-7 bp***
Barcode + Heterogenic spacer F2	TGCAGATCCAAC					
Barcode + Heterogenic spacer F3	CCATCACATAGG					
Barcode + Heterogenic spacer F4	GTGGTATGGGAG	T				
Barcode + Heterogenic spacer F5	ACTTTAAGGGTG	T				
Barcode + Heterogenic spacer F6	GAGCAACATCCT	T				
Barcode + Heterogenic spacer F7	TGTTGCGTTTCT	GT				
Barcode + Heterogenic spacer F8	ATGTCCGACCAA	GT				
Barcode + Heterogenic spacer F9	AGGTACGCAATT	GT				
Barcode + Heterogenic spacer F10	ACAGCCACCCAT	CGA				
Barcode + Heterogenic spacer F11	TGTCTCGCAAGC	CGA				
Barcode + Heterogenic spacer F12	GAGGAGTAAAGC	CGA				
Barcode + Heterogenic spacer F13	GTTACGTGGTTG	ATGA				
Barcode + Heterogenic spacer F14	TACCGCCTCGGA	ATGA				
Barcode + Heterogenic spacer F15	CGTAAGATGCCT	ATGA				
Barcode + Heterogenic spacer F16	TACCGGCTTGCA	TGCGA				
Barcode + Heterogenic spacer F17	ATCTAGTGGCAA	TGCGA				
Barcode + Heterogenic spacer F18	CCAGGGACTTCT	TGCGT				
Barcode + Heterogenic spacer F19	CACCTTACCTTA	GAGTGG				
Barcode + Heterogenic spacer F20	ATAGTTAGGGCT	GAGTGG				
Barcode + Heterogenic spacer F21	GCACTTCATTTC	GAGTGG				
Barcode + Heterogenic spacer F22	TTAACTGGAAGC	CCTGTGG				
Barcode + Heterogenic spacer F23	CGCGGTTACTAA	CCTGGAG				
Barcode + Heterogenic spacer F24	GAGACTATATGC	CCTGGAG				
Barcode + Heterogenic spacer R1	CCTAAACTACGG					
Barcode + Heterogenic spacer R2	TGCAGATCCAAC					
Barcode + Heterogenic spacer R3	CCATCACATAGG					
Barcode + Heterogenic spacer R4	GTGGTATGGGAG	A				
Barcode + Heterogenic spacer R5	ACTTTAAGGGTG	A				
Barcode + Heterogenic spacer R6	GAGCAACATCCT	A				
Barcode + Heterogenic spacer R7	TGTTGCGTTTCT	TC				
Barcode + Heterogenic spacer R8	ATGTCCGACCAA	TC				
Barcode + Heterogenic spacer R9	AGGTACGCAATT	TC				
Barcode + Heterogenic spacer R10	ACAGCCACCCAT	CTA				
Barcode + Heterogenic spacer R11	TGTCTCGCAAGC	CTA				
Barcode + Heterogenic spacer R12	GAGGAGTAAAGC	CTA				
Barcode + Heterogenic spacer R13	GTTACGTGGTTG	GATA				
Barcode + Heterogenic spacer R14	TACCGCCTCGGA	GATA				
Barcode + Heterogenic spacer R15	CGTAAGATGCCT	GATA				
Barcode + Heterogenic spacer R16	TACCGGCTTGCA	ACTCA				
Barcode + Heterogenic spacer R17	ATCTAGTGGCAA	ACTCA				
Barcode + Heterogenic spacer R18	CCAGGGACTTCT	ACTCA				
Barcode + Heterogenic spacer R19	CACCTTACCTTA	TTCTCT				
Barcode + Heterogenic spacer R20	ATAGTTAGGGCT	TTCTCT				
Barcode + Heterogenic spacer R21	GCACTTCATTTC	TTCTCT				
Barcode + Heterogenic spacer R22	TTAACTGGAAGC	CACTTCT				
Barcode + Heterogenic spacer R23	CGCGGTTACTAA	CACTTCT				
Barcode + Heterogenic spacer R24	GAGACTATATGC	CACTTCT				

Appendix S4: Primer binding sites



Modified from Alcaide *et al.* (2013; MHCI) and Canal *et al.* (2010; MHCII)

Appendix S5: Methods: sequencing details

Sequencing of MHCI exon 3 and MHCII exon 2 in bluethroats

MHC class I exon 3 sequencing using IonTorrent

Single-indexed amplicons (MHCI-SI)

MHC class I exon 3 amplicons were generated using the MhcPasCI-FW and MhcPasCI-RV primer pair (Alcaide et al. 2013). The forward primers were combined with IonXpress barcodes and the GAT primer adapter motif (see Appendix S3). The amplicons were amplified with the following conditions: 1X reaction buffer, 0.8 mM dNTP, 0.5 µM of each primer and the Q5 polymerase (New England Biolabs) following the manufacturer's recommendations. Additionally, we included 1X Q5 High GC Enhancer supplied with the polymerase for this PCR reaction. The following thermal profile was applied: initial denaturation at 98 °C for 30 seconds, 25 cycles of denaturation 98 °C for 10 seconds, annealing at 59 °C for 20 seconds and elongation at 72 °C for 15 seconds, and a final elongation step at 72 °C for 2 minutes. Amplicons were inspected on agarose gels and concentrations were estimated on a Fragment Analyzer (Advanced Analytical) using the High Sensitivity Genomic DNA Analysis kit (Advanced Analytical). Equimolar amounts of each amplicon were pooled and the final library was prepared using the NEBNext® Fast DNA Library Prep Set for Ion Torrent (New England Biolabs) following the manufacturers recommendations. The amplicons were sequenced on a 316v2 chip prepared using Ion Chef (ThermoFischer) on an Ion Personal Genomic Machine (PGM) (ThermoFischer) together with the dual-indexed IonTorrent libraries and 12 amplicons not included in this project (60 amplicons in total).

Dual-indexed amplicons (MHCI-DI)

The dual-indexed amplicons were generated with the same gene target as for the single-indexed amplicons. Yet, the amplification primers were designed to include the following motifs: (i) IonTorrent A adapters on the forward primers and IonTorrent trP1 adapters on the reverse primers, (ii) A unique combination of IonXpress barcode on forward primers and reverse complement IonXpress barcode on reverse barcodes, (iii) barcode adapter motif GAT, (iv) a "spacer motif" of seven nucleotides, as described in Fadrosh et al. (2014). The motivation for including this motif was to mimic amplicons prepared for Illumina sequencing (see below) and to be able to distinguish single-indexed and dual-indexed MHC class I exon 3 amplicons, (v) Gene target primers.

The amplicons were obtained using the same PCR conditions and thermal profile as for the single-indexed amplicons. Amplicon concentrations was estimated on a Fragment Analyzer (Advanced Analytical) using the High Sensitivity Genomic DNA Analysis kit. Equimolar amounts of amplicons were pooled and purified using Agencourt AMPure XP beads (Beckman Coulter, Inc.) and sequenced directly on a 316v2 chip, prepared using Ion Chef (ThermoFischer) (together with the MHCI-SI and MHCII-DI libraries as described above), on an Ion PGM system (ThermoFischer).

MHC class II exon 2 amplicon sequencing using IonTorrent

Single-indexed amplicons (MHCII-SI)

For the single-indexed amplicons, we used the MHCIIFihy-E2CF and MHCIIFihy-E2CR primers described by Canal et al. (2010) as gene target. IonXpress barcodes were included together with the barcode adapter GAT on the forward primer sequences. PCR conditions and thermal profile were the same as for MHC class I exon 3, with the exception of annealing at 72 °C and exclusion of the Q5 High GC Enhancer. All amplicons were inspected on agarose gels. To prepare the sequencing library, we followed the same procedure as described for the MHC class I exon 3 single-indexed library. The library was sequenced along with four additional amplicons on a 316v2 chip prepared using the Ion OneTouch 2 system (ThermoFischer) using Ion PGM system (ThermoFischer).

Dual-indexed amplicons (MHCII-DI)

Dual-indexed primer strategy was the same as for the dual-indexed MHC class I exon 3 IonTorrent library, except for the use of MHCIIFihy-E2CF/MHCIIFIhy-E2CR (Canal et al. 2010) primer pair as gene target (see Appendix S3). PCR conditions and thermal profile were the same as for the MHC class II exon 2 single-indexed sequencing strategy. The final library was sequenced together with the MHC class I exon 3 amplicons as described above.

MHC class II exon 2 amplicon sequencing using Illumina MiSeq *(MHCII-MiSeq)*

The 16 amplicons for Illumina MiSeq sequencing was generated using a dual-indexed sequencing strategy, together with 560 other amplicons (manuscript in prep.) The PCR amplification primer setup followed Fadrosh et al. (2014), which included in both forward and reverse direction: (i) Illumina Linker Sequence, (ii) a 12 nucleotide barcode, (iii) a "heterogeneity spacer" motif of 0-7 nucleotides for optimization of MiSeq amplicon sequencing, (iv) gene target primers. Unique combinations of 24 forward and 24 reverse barcodes allowed for multiplexing of 576 amplicons. Gene target motif was the same as for the single-indexed MHC class II exon 2 primers (see Appendix S3). PCR conditions and thermal profile was as described for the IonTorrent MHCII strategies. Amplicons were inspected on agarose gels, equimolar amounts of amplicons were pooled, purified and sequenced on an Illumina MiSeq® platform using v3 chemistry.

References:

Alcaide, M., Liu, M., & Edwards, S. V. (2013). Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. *PeerJ,* 1, e86. doi: 10.7717/peerj.86

Canal, D., Alcaide, M., Anmarkrud, J. A., & Potti, J. (2010). Towards the simplification of MHC typing protocols: targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*. *BMC research notes,* 3, 236. doi: 10.1186/1756-0500-3-236

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome,* 2, 6. doi: 10.1186/2049-2618-2-6

Appendix S6: Methods: allele calling details

Allele calling of MHCI exon 3 and MHCII exon 2 in bluethroats

Modified pipeline from Sommer et al. (2013):

Initially, short reads (<150 base pair) were removed using the awk-command in UNIX (IonTorrent-data), and paired end reads were merged using FLASH v1.2.11 (Magoč and Salzberg (2011), MiSeq-data). Further raw read filtering was done using fastx toolkit v0.0.13/0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/), by removing reads where >5% of the bases had a Phred quality score below 20. The heterogenic spacer motifs in the MiSeq-data were removed by standard UNIX commands, in order to make the unique identifiers of equal lengths as demanded by jMHC (see below).

Of the 4.1 million raw reads obtained for IonTorrent run combining MHCI-SI, MHCI-DI and MHCII-DI (60 amplicons, see Appendix S5), 44.5 % passed the initial quality filtering. The MHCII-SI run rendered 3.5 million reads across all 20 amplicons sequences (see Appendix S5), of which 10.7% passed the initial quality filters. The MHCIIβe2 was also sequenced for the eight test-individuals at an Illumina MiSeq-platform together with a larger dataset consisting of duplicate samples of 280 additional individuals. FLASH combined 84.88% of the paired end raw reads and of these in total 9.9 million merged reads, 96.5% passed the quality filtering.

For all datasets, remaining reads were assigned to amplicons based on barcode, and clustered into variants using the software jMHC v1.6.1624 (Stuglik et al. 2011). Only variants with complete barcodes and three or more reads across the dataset were outputted. The variants were further evaluated in an Excel spreadsheet, and variants with less than three reads in any amplicon were discarded. Amplicons with less than 500 reads in total after this step were removed (applied to one MHCII-MiSeq amplicon).

One of the major modifications of the Sommer-pipeline was the implementation of a second frequency threshold above which most true alleles should occur, similar to Galan's T_2 (Galan et al. 2010). A central assumptions when genotyping MHC using PCR-based high throughput sequencing methods is that most artifacts should have lower sequencing depths than real alleles (Babik et al. 2009; Lighten et al. 2014). As a first exploratory step of separating real alleles from artifacts, we thus set cut-off values based on the distribution of unique variants at different intra-amplicon threshold levels for inclusion (Figure 3, main text). By visually inspecting the plot, we tried to find the balance between including true alleles with low amplification efficiency and excluding true artefacts. The greatest change in the number of unique variants included was observed below 0.10 % - 0.20 %, suggesting the inclusion of low frequency artefacts around these cut-off values. Because the remaining variants would be filtered based on other criteria in subsequent steps of the pipeline, we applied a threshold value of 0.20 % in the analyses as an approach of classifying "possible alleles". This value is close to the maximum per amplicon frequency of 0.18 % used by Biedrzycka et al. (2017) when examining a sedge warbler MHC dataset.

Further, every variant exceeding the threshold frequency in any amplicon was then aligned to published sequences from bluethroat (GenBank accession number KU169737-KU169747 (MHCIe3; O'Connor et al. (2016)) and HQ539575-HQ539614 (MHCIIβe2; Gohli et al. (2013)), using ClustalW (Thompson et al. 1994) in MEGA7 v7.0.14 (Kumar et al. 2016). All variants displaying shifts in reading frame or displaying stop codons when translated were removed, and bases outside the targeted 239 (MHCIe3) or 267 (MHCIIβe2) base pairs were deleted. Following O'Connor et al. (2016) and Gohli et al. (2013) respectively, we additionally removed variants not possessing Cys7 and Cys70 (MHCI) or

Cys10 and Cys75 (MHCII), motivated by the conserved function of these residues. Identical variants were collapsed using the fastx_collapser command in fastx toolkit, and their depths were summed up.

The variants were sorted based on their intra-amplicon frequency. The most frequent variant in every amplicon was scored as “allele” for the amplicon in question. Chimera detection was conducted on the included variants of each amplicon in UCHIME (Edgar et al. 2011), using the uchime_denovo command in USEARCH v8.1.1861 or v 7.0.1090. For the non-chimeric variants, the number of base pair differences to the most similar, more frequent variant was calculated in MEGA7. As variants with only one base pair difference from a more frequent variant are more likely to be an artefactual sequence, these “=1bp”-variants were treated more strictly in the subsequent workflow. This is in line with the recommendations for species with “closely related ‘putative alleles’” (Sommer et al. 2013), which we are expecting for the bluethroat (MHCII; see Anmarkrud et al. (2010)).

In the next step, amplicon replicates and variant status (*i.e.* chimeric, “=1bp” and “>1bp” variants) were used to identify artifacts. First, a variant was considered an artifact in four scenarios: (1) a chimeric variant not present in amplicon replicate; (2) a chimeric variant detected as a chimera in amplicon replicate; (3) a “=1bp” variant not present in amplicon replicate; or (4) a “>1bp” variant not present in any other amplicon from individuals within the same family group. The artifact scoring was done for all amplicons before proceeding in the workflow. The variant was considered present if it had a frequency above the intra-amplicon threshold, and not only by its mere presence with at least three reads (deviating from Sommer et al. (2013)).

Second, the remaining variants above the 0.2% cut-off threshold in each amplicon were scored as follows: If the “=1bp” and “>1bp” variants were present in the amplicon replicate, they were scored as alleles. However, if a “>1bp” variant was present in other amplicons within the family group albeit not in replicate, it was nevertheless scored as allele. Chimeric variants were scored as allele if it was present as a non-chimeric sequence in amplicon replicate.

The two latter steps are however deviating slightly from the protocol in Sommer et al. (2013), in which “>2bp variants” are scored if present in other amplicons across the whole dataset while not in the replicate. We changed the original “>2bp” intra-amplicon evaluation category to “>1bp” according to the recommendations in Sommer et al. (2013), due to the expectancy of high complexity in bluethroat MHCII. Further, the “>1bp variants” were compared against the family group and not across the whole data set, as this will ensure increased applicability if used on datasets of different sizes. This treatment of “>1bp variants” will affect the observed “errors in pedigree”, calculated as the percentage of alleles in offspring not found in any of its parents, which is thus not directly comparable between the two genotyping pipelines. Lastly, the “unclassified variants” and “putative alleles” categories from Sommer et al. (2013) were collapsed to “Alleles”, due to the presence of high frequency artefacts.

Dealing with individuals with one failed replicate

Family information can prove particularly useful when genotyping complex markers, by assuming all true alleles within an offspring will be present in one or both of its parents. The genotype of the parents can thus be utilized to avoid loss of data in cases where an offspring has one amplicon without sufficient coverage for genotyping. Accordingly, we established a genotyping regime for offspring having one failed amplicon within a strategy, for whom parentage was known and the biological parents were included in the dataset. In these instances, the successful amplicon was analyzed following the protocol of the modified Sommer-pipeline until the artefact and allele calling (after step 10, main text Figure 3). The alleles of the parents were then used as a substitute for the failed amplicon in the subsequent steps: if a “=1bp” variant or a chimeric variant was scored as an allele in one of the parents, it was genotyped also for the focal offspring. The “>1 bp” variants were treated in the same manner as for every individual, and scored if found within any other family member.

AmpliSAS

Initial filtering:

MHCI-SI, MHCI-DI and MHCII-DI

Due to size limitations of input files in AmpliCLEAN, the initial quality and length filtering of the raw data was done using fastx_toolkit, as described for the modified pipeline from Sommer et al. (2013). The resulting output file was used as input for AmpliSAS twice, once for genotyping MHCI (using the MHCI-SI and MHCI-DI amplicons) and once for genotyping MHCII (using the MHCII-DI amplicons).

MHCII-SI

In order to treat the input files identically, the raw read file was quality and length trimmed using fastx_toolkit as described for MHCII-DI above. A fasta file containing the variants scored as alleles for MHCII-DI was used an input “Allele file” in order to name the variants identically.

MHCII-MiSeq

In order to obtain an input file of accepted size for the online version of AmpliSAS, AmpliMERGE and AmpliCLEAN were replaced by alternative approaches: the software FLASH was used to merge the sequence pairs from the paired end MiSeq run, while fastx_toolkit was used to filter the reads based on quality, with the same parameters as for the modified pipeline from Sommer et al. (2013). The merged, cleaned file was gzipped and used as input for AmpliSAS. A fasta file containing alleles from the AmpliSAS run of MHCII-SI and MHCII-DI was used as “Allele file”.

Genotyping:

The online software suite AmpliSAS (Sebastian et al. 2016) was found at http://evobiolab.biol.amu.edu.pl/amplisat/index.php?amplisas. The respective input files were as stated above, and only variants scored in both amplicon replicates of an individual were called as alleles.

References:

Anmarkrud, J., Johnsen, A., Bachmann, L., & Lifjeld, J. (2010). Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes. *Journal of evolutionary biology,* 23, 1206-1217. doi: 10.1111/j.1420-9101.2010.01999.x

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. doi: 10.1111/j.1755-0998.2009.02622.x

Biedrzycka, A., Sebastian, A., Migalska, M., Westerdahl, H., & Radwan, J. (2017). Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Molecular Ecology Resources,* 17, 642-655. doi: 10.1111/1755-0998.12612

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics,* 27, 2194-2200. doi: 10.1093/bioinformatics/btr381

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, 1. doi: 10.1186/1471-2164-11-296

Gohli, J., Anmarkrud, J. A., Johnsen, A., Kleven, O., Borge, T., & Lifjeld, J. T. (2013). Female promiscuity is positively associated with neutral and selected genetic diversity in passerine birds. *Evolution,* 67, 1406-1419. doi: 10.1111/evo.12045

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution,* 33, 1870-1874. doi: 10.1093/molbev/msw054

Lighten, J., Oosterhout, C., & Bentzen, P. (2014). Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology,* 23, 3957-3972. doi: 10.1111/mec.12843

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics,* 27, 2957-2963. doi: 10.1093/bioinformatics/btr507

O'Connor, E., Strandh, M., Hasselquist, D., Nilsson, J. Å., & Westerdahl, H. (2016). The evolution of highly variable immunity genes across a passerine bird radiation. *Molecular Ecology*, 977-989. doi: 10.1111/mec.13530

Sebastian, A., Herdegen, M., Migalska, M., & Radwan, J. (2016). amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources,* 16, 498-510. doi: 10.1111/1755-0998.12453

Sommer, S., Courtiol, A., & Mazzoni, C. J. (2013). MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC genomics,* 14, 1. doi: 10.1186/1471-2164-14-542

Stuglik, M. T., Radwan, J., & Babik, W. (2011). jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Molecular Ecology Resources,* 11, 739-742. doi: 10.1111/j.1755-0998.2011.02997.x

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22, 4673-4680. doi: 10.1093/nar/22.22.4673

Appendix S7: Genotypes, MHCIe3

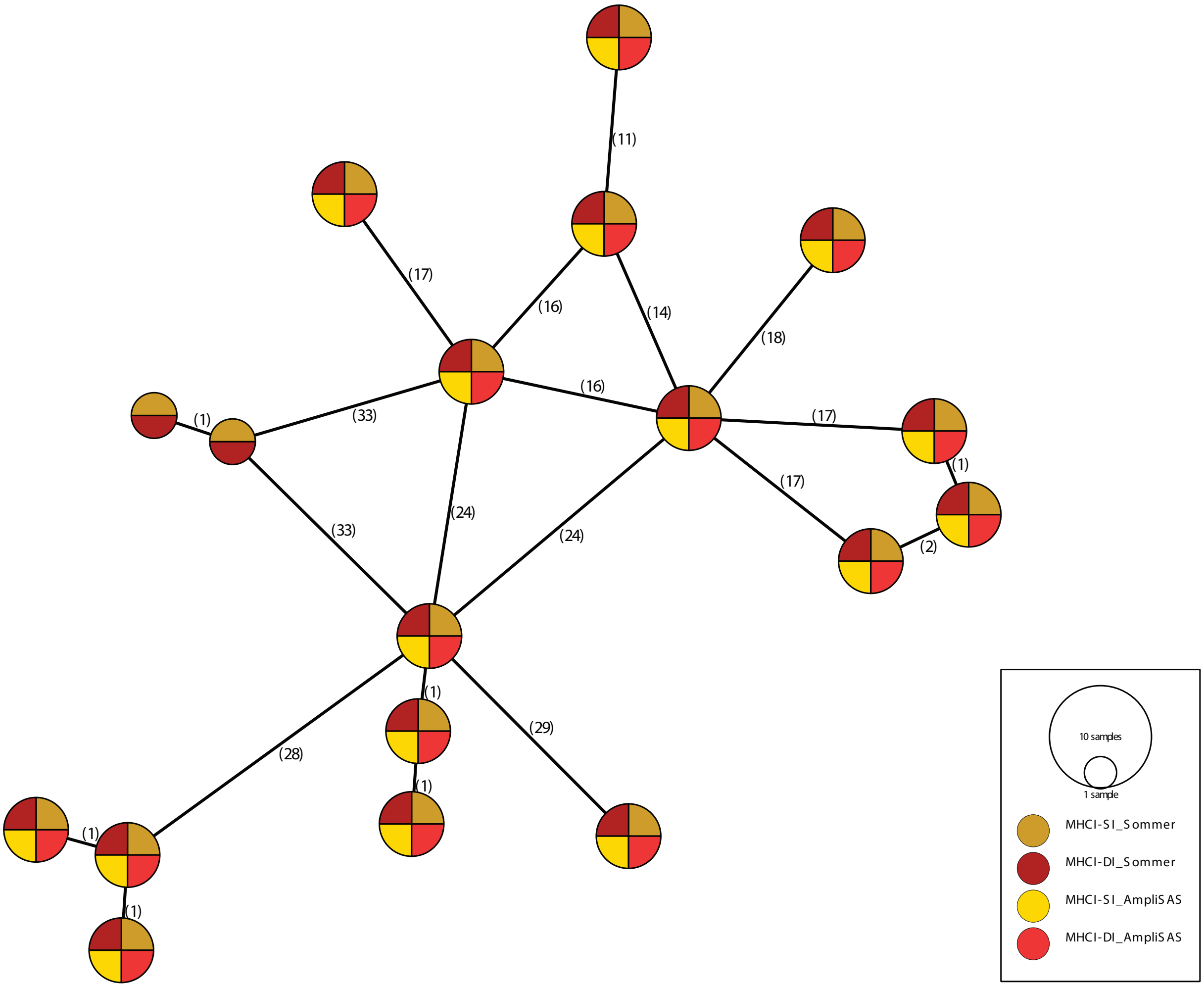
8 bluethroat individuals from Heimdalen, Norway

Individual	69392				69393				69390				69391			
Family role	Mother, family A				Father, family A				Offspring, family A				Offspring, family A			
Genotyping pipeline	SommerPipeline		AmpliSAS		SommerPipeline		AmpliSAS		SommerPipeline		AmpliSAS		SommerPipeline		AmpliSAS	
Primer strategy	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI
No of alleles	6	6	4	4	6	6	5	5	6	6	4	4	5	5	4	4
No of pedigree errors									0	0	0	0	0	0	0	0
AlleleName_LuSv_SR	Column25	Column4	Column1	Column2	Column72	Column5	Column3	Column12	Column92	Column6	Column13	Column14	olumn112	Column7	Column15	Column16
LuSv_MHCIe3_SR_002	Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_025																
LuSv_MHCIe3_SR_022																
LuSv_MHCIe3_SR_024																
LuSv_MHCIe3_SR_023	Allel	Allele	Allele	Allele									Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_017																
LuSv_MHCIe3_SR_020																
LuSv_MHCIe3_SR_001	Allel	Allele			Allel	Allele			Allel	Allele			Allel	Allele		
LuSv_MHCIe3_SR_031	Allel	Allele							Allel	Allele						
LuSv_MHCIe3_SR_015																
LuSv_MHCIe3_SR_019																
LuSv_MHCIe3_SR_004																
LuSv_MHCIe3_SR_007																
LuSv_MHCIe3_SR_003					Allel	Allele	Allele	Allele					Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_027					Allel	Allele	Allele	Allele					Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_012	Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele				
LuSv_MHCIe3_SR_021	Allel	Allele	Allele	Allele					Allel	Allele	Allele	Allele				
LuSv_MHCIe3_SR_013					Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele				

Individual	69524				69525				69526				69528			
Family role	Mother, family B				Offspring, family B				Offspring, family B				Father, family B			
Genotyping pipeline	SommerPipeline		AmpliSAS		SommerPipeline		AmpliSAS		SommerPipeline		AmpliSAS		SommerPipeline		AmpliSAS	
Primer strategy	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI	MHCI-SI	MHCI-DI
No of alleles	7	7	6	6	7	7	6	6	8	8	7	4	8	8	7	7
No of pedigree errors					0	0	0	0	0	0	0	0				
AlleleName_LuSv_SR	olumn132	Column8	Column17	Column18	olumn152	Column9	Column19	Column20	olumn172	Column10	Column21	Column22	olumn192	Column11	Column23	Column24
LuSv_MHCIe3_SR_002					Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_025									Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_022	Allel	Allele	Allele	Allele					Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_024	Allel	Allele	Allele	Allele					Allel	Allele	Allele	Allele				
LuSv_MHCIe3_SR_023	Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele	Allel	Allele	Allele		Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_017	Allel	Allele	Allele	Allele					Allel	Allele	Allele					
LuSv_MHCIe3_SR_020	Allel	Allele	Allele	Allele					Allel	Allele	Allele					
LuSv_MHCIe3_SR_001	Allel	Allele			Allel	Allele			Allel	Allele			Allel	Allele		
LuSv_MHCIe3_SR_031																
LuSv_MHCIe3_SR_015					Allel	Allele	Allele	Allele					Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_019					Allel	Allele	Allele	Allele					Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_004					Allel	Allele	Allele	Allele					Allel	Allele	Allele	Allele
LuSv_MHCIe3_SR_007	Allel	Allele	Allele	Allele	Allel	Allele	Allele	Allele								
LuSv_MHCIe3_SR_003																
LuSv_MHCIe3_SR_027																
LuSv_MHCIe3_SR_012																
LuSv_MHCIe3_SR_021																
LuSv_MHCIe3_SR_013																

Appendix S8: Sequence logo, MHCIe3

Appendix S9: Network MHCIe3



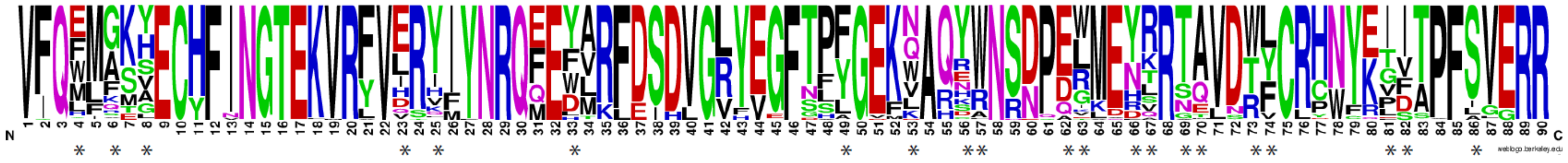
Appendix S10: Genotypes, MHCIIβe2

8 bluethroat individuals from Heimdalen, Norway

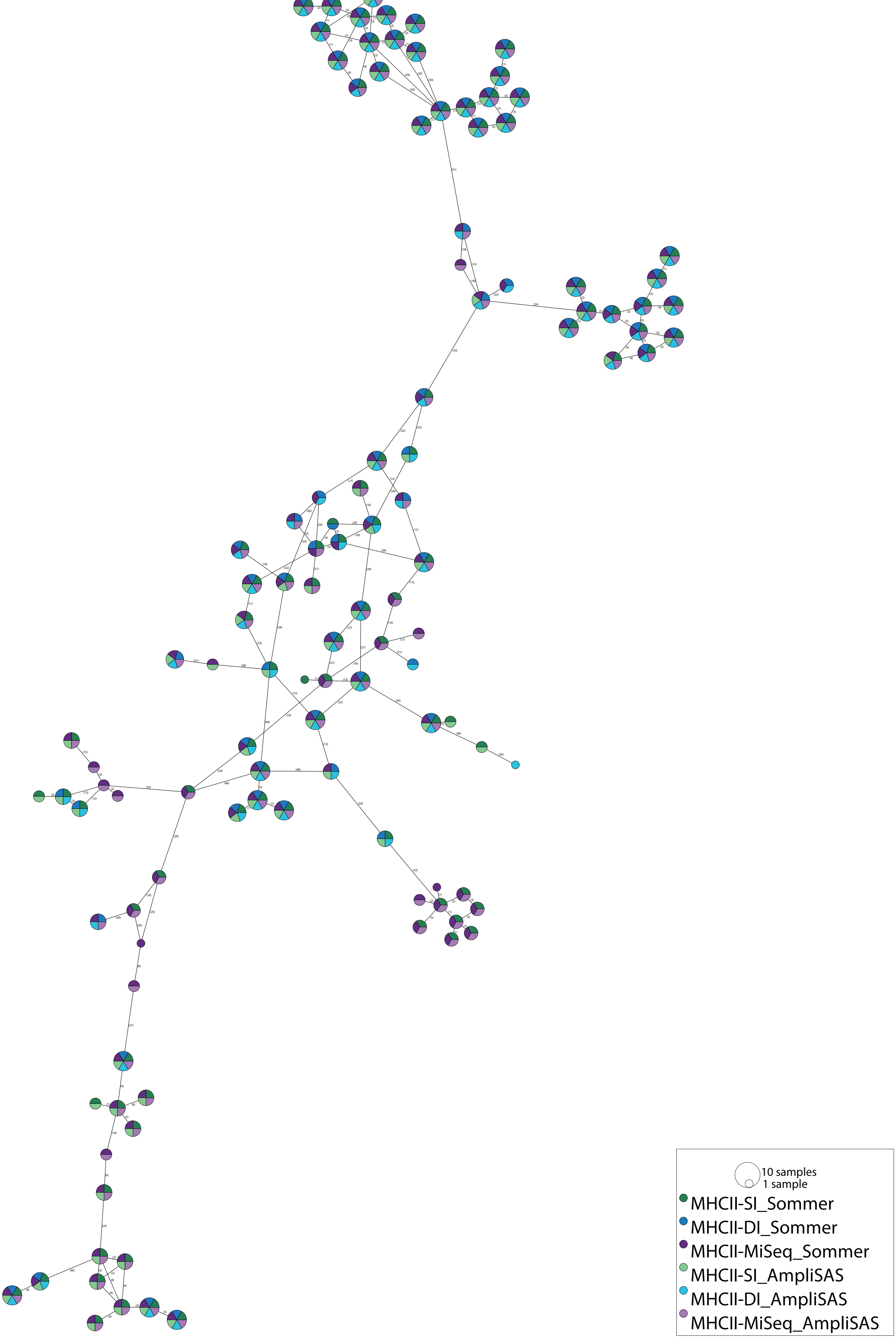
Individual	69392							69393							69390							69391						
Family role	Mother, family A							Father, family A							Offspring, family A							Offspring, family A						
Genotyping pipeline		SommerPipeline			AmpliSAS				SommerPipeline			AmpliSAS				SommerPipeline			AmpliSAS				SommerPipeline			AmpliSAS		
Primer strategy	Combined	MHCII-SI	MHCII-DI	MiSeq	MHCII-SI	MHCII-DI	MiSeq	Combined	MHCII-SI	MHCII-DI	MiSeq	MHCII-SI	MHCII-DI	MiSeq	Combined	MHCII-SI	MHCII-DI	MiSeq	MHCII-SI	MHCII-DI	MiSeq	Combined	MHCII-SI	MHCII-DI	MiSeq	MHCII-SI	MHCII-DI	MiSeq
Alleles found in all strategies	19							10							18							12						
"Reference genotype"	54							35							56							35						
No of alleles		47	31	41	35	30	29		31	17	26	22	17	22		45	29	40	39	30	31		31	20	23	20	21	21
No of pedigree errors																0	0	one failed rep	3	1	one failed rep	3	1	2	0	1	5	4
AlleleName_LuSv_SR	69392	Column4	Column5	69392_3	Column1	Column2	Column3	69393	Column9	Column10	69393_6	Column6	Column7	Column8	69390	69390_4	69390_5	69390_6	69390_	69390_2	69390_3	69391	69391_4	69391_5	69391_6	69391_	69391_2	69391_3
LuSv_MHCIIe2_SR_001	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_009																												
LuSv_MHCIIe2_SR_010																						Allele	Allele					
LuSv_MHCIIe2_SR_013																												
LuSv_MHCIIe2_SR_015	Allele	Allele		Allele			Allele															Allele	Allele		Allele			Allele
LuSv_MHCIIe2_SR_016	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_017								Allele	Allele	Allele	Allele		Allele									Allele	Allele	Allele			Allele	
LuSv_MHCIIe2_SR_018	Allele	Allele		Allele	Allele																	Allele	Allele		Allele	Allele		Allele
LuSv_MHCIIe2_SR_019																												
LuSv_MHCIIe2_SR_020	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele					
LuSv_MHCIIe2_SR_021																												
LuSv_MHCIIe2_SR_022	Allele	Allele			Allele			Allele	Allele			Allele			Allele	Allele			Allele			Allele	Allele			Allele		
LuSv_MHCIIe2_SR_023	Allele	Allele	Allele	Allele	Allele																	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_024	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_025	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_026																						Allele		Allele			Allele	
LuSv_MHCIIe2_SR_027	Allele	Allele		Allele			Allele								Allele	Allele		Allele			Allele							
LuSv_MHCIIe2_SR_028															Allele				Allele			Allele				Allele		
LuSv_MHCIIe2_SR_029	Allele	Allele	Allele	Allele	Allele	Allele									Allele	Allele	Allele	Allele	Allele	Allele								
LuSv_MHCIIe2_SR_030																												
LuSv_MHCIIe2_SR_031																												
LuSv_MHCIIe2_SR_032																												
LuSv_MHCIIe2_SR_033	Allele	Allele	Allele	Allele	Allele	Allele									Allele	Allele	Allele	Allele	Allele	Allele								
LuSv_MHCIIe2_SR_034	Allele		Allele	Allele	Allele	Allele									Allele		Allele	Allele	Allele									
LuSv_MHCIIe2_SR_035								Allele	Allele	Allele		Allele	Allele		Allele	Allele				Allele								
LuSv_MHCIIe2_SR_036	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele					
LuSv_MHCIIe2_SR_037	Allele	Allele		Allele				Allele	Allele		Allele			Allele	Allele			Allele			Allele	Allele	Allele		Allele			
LuSv_MHCIIe2_SR_038	Allele	Allele			Allele																	Allele	Allele	Allele		Allele	Allele	
LuSv_MHCIIe2_SR_039	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele			Allele			Allele	Allele						Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_040																												
LuSv_MHCIIe2_SR_041								Allele	Allele			Allele			Allele	Allele			Allele									
LuSv_MHCIIe2_SR_042	Allele	Allele	Allele		Allele																	Allele	Allele	Allele		Allele	Allele	
LuSv_MHCIIe2_SR_043	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_044	Allele	Allele		Allele			Allele								Allele	Allele		Allele			Allele							
LuSv_MHCIIe2_SR_045	Allele	Allele			Allele										Allele	Allele			Allele									
LuSv_MHCIIe2_SR_046	Allele	Allele	Allele	Allele		Allele																Allele	Allele	Allele	Allele		Allele	Allele
LuSv_MHCIIe2_SR_047																												
LuSv_MHCIIe2_SR_048	Allele			Allele											Allele			Allele	Allele									
LuSv_MHCIIe2_SR_049																												
LuSv_MHCIIe2_SR_050																												
LuSv_MHCIIe2_SR_051																												
LuSv_MHCIIe2_SR_052	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_053																												
LuSv_MHCIIe2_SR_054																												
LuSv_MHCIIe2_SR_055	Allele	Allele	Allele	Allele		Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele
LuSv_MHCIIe2_SR_056								Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_057	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_058																												
LuSv_MHCIIe2_SR_059	Allele	Allele						Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_060																												
LuSv_MHCIIe2_SR_061																												
LuSv_MHCIIe2_SR_062								Allele	Allele		Allele			Allele	Allele	Allele		Allele			Allele							
LuSv_MHCIIe2_SR_063								Allele	Allele		Allele	Allele		Allele	Allele	Allele		Allele										
LuSv_MHCIIe2_SR_064	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_065																												
LuSv_MHCIIe2_SR_066																												
LuSv_MHCIIe2_SR_067	Allele	Allele						Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_068	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_069																												
LuSv_MHCIIe2_SR_070								Allele	Allele		Allele	Allele		Allele	Allele	Allele		Allele	Allele		Allele							
LuSv_MHCIIe2_SR_071	Allele		Allele	Allele		Allele	Allele								Allele		Allele	Allele		Allele	Allele							
LuSv_MHCIIe2_SR_072	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_073																												
LuSv_MHCIIe2_SR_074	Allele	Allele	Allele	Allele	Allele	Allele	Allele															Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_075																												
LuSv_MHCIIe2_SR_076	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_077																												
LuSv_MHCIIe2_SR_078								Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_079																												
LuSv_MHCIIe2_SR_080																												
LuSv_MHCIIe2_SR_081	Allele	Allele	Allele		Allele	Allele									Allele	Allele	Allele		Allele									
LuSv_MHCIIe2_SR_082	Allele	Allele	Allele	Allele	Allele	Allele	Allele															Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_083																												
LuSv_MHCIIe2_SR_084								Allele		Allele	Allele		Allele															
LuSv_MHCIIe2_SR_085	Allele	Allele			Allele			Allele	Allele			Allele			Allele	Allele			Allele			Allele	Allele			Allele		
LuSv_MHCIIe2_SR_086								Allele			Allele			Allele								Allele			Allele			Allele
LuSv_MHCIIe2_SR_087																												
LuSv_MHCIIe2_SR_088	Allele	Allele						Allele	Allele						Allele	Allele			Allele									
LuSv_MHCIIe2_SR_089																												
LuSv_MHCIIe2_SR_090																												
LuSv_MHCIIe2_SR_091	Allele					Allele									Allele					Allele								
LuSv_MHCIIe2_SR_092	Allele	Allele	Allele		Allele	Allele									Allele	Allele	Allele		Allele	Allele								
LuSv_MHCIIe2_SR_093	Allele		Allele	Allele		Allele									Allele		Allele	Allele		Allele								
LuSv_MHCIIe2_SR_094	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_095																												
LuSv_MHCIIe2_SR_096																												
LuSv_MHCIIe2_SR_097																												
LuSv_MHCIIe2_SR_098	Allele	Allele		Allele			Allele	Allele	Allele		Allele			Allele	Allele	Allele		Allele			Allele	Allele	Allele		Allele			Allele
LuSv_MHCIIe2_SR_099								Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_100	Allele	Allele	Allele	Allele		Allele	Allele								Allele	Allele	Allele	Allele		Allele	Allele							
LuSv_MHCIIe2_SR_101																												
LuSv_MHCIIe2_SR_102																												
LuSv_MHCIIe2_SR_103	Allele	Allele		Allele	Allele										Allele	Allele		Allele	Allele									
LuSv_MHCIIe2_SR_104	Allele	Allele		Allele			Allele															Allele	Allele		Allele			Allele
LuSv_MHCIIe2_SR_105																												
LuSv_MHCIIe2_SR_106								Allele	Allele		Allele			Allele														
LuSv_MHCIIe2_SR_107																												
LuSv_MHCIIe2_SR_108								Allele	Allele		Allele	Allele		Allele	Allele	Allele		Allele	Allele		Allele							
LuSv_MHCIIe2_SR_109								Allele	Allele	Allele	Allele		Allele	Allele								Allele	Allele	Allele	Allele		Allele	Allele
LuSv_MHCIIe2_SR_110																												
LuSv_MHCIIe2_SR_111	Allele	Allele						Allele	Allele						Allele	Allele						Allele	Allele					
LuSv_MHCIIe2_SR_112	Allele	Allele		Allele	Allele																	Allele	Allele		Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_113																												
LuSv_MHCIIe2_SR_114																												
LuSv_MHCIIe2_SR_115																												
LuSv_MHCIIe2_SR_116								Allele	Allele		Allele	Allele		Allele	Allele	Allele			Allele									
LuSv_MHCIIe2_SR_117	Allele			Allele											Allele			Allele										
LuSv_MHCIIe2_SR_118																												
LuSv_MHCIIe2_SR_119								Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele	Allele	Allele		Allele								
LuSv_MHCIIe2_SR_120																												
LuSv_MHCIIe2_SR_121	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_122	Allele	Allele		Allele	Allele		Allele								Allele	Allele		Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_123	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_124								Allele	Allele	Allele		Allele	Allele									Allele	Allele	Allele		Allele	Allele	
LuSv_MHCIIe2_SR_125																												
LuSv_MHCIIe2_SR_126	Allele	Allele			Allele										Allele				Allele									
LuSv_MHCIIe2_SR_127	Allele			Allele			Allele	Allele			Allele			Allele	Allele			Allele			Allele	Allele			Allele			
LuSv_MHCIIe2_SR_128								Allele			Allele																	

Individual	69524							69525							69526							69528						
Family role	Mother, family B							Offspring, family B							Offspring, family B							Father, family B						
Genotyping pipeline		SommerPipeline			AmpliSAS				SommerPipeline			AmpliSAS				SommerPipeline			AmpliSAS				SommerPipeline			AmpliSAS		
Primer strategy	Combined	MHCII-SI	MHCII-DI	MiSeq	MHCII-SI	MHCII-DI	MiSeq	Combined	MHCII-SI	MHCII-DI	MiSeq	MHCII-SI	MHCII-DI	MiSeq	Combined	MHCII-SI	MHCII-DI	MiSeq	MHCII-SI	MHCII-DI	MiSeq	Combined	MHCII-SI	MHCII-DI	MiSeq	MHCII-SI	MHCII-DI	MiSeq
Alleles found in all strategies	9							11							15							15						
"Reference genotype"	43							40							43							42						
No of alleles		31	19	30	29	16	26		29	21	30	20	20	28		37	23	36	31	21	31		34	24	35	26	21	34
No of pedigree errors								2	0	0	2	0	2	3	1	2	2	1	3	2	3							
AlleleName_LuSv_SR	**69524**	69524_5	69524_4	**69524_6**	69524_	69524_2	**69524_3**	**69525**	69525_4	69525_5	**69525_6**	69525_	69525_2	**69525_3**	**69526**	69526_4	69526_5	**69526_6**	69526_	69526_2	**69526_3**	**69528**	69528_4	69528_5	**69528_6**	69528_	69528_2	**69528_3**
LuSv_MHCIIe2_SR_001	Allele	Allele		Allele	Allele			Allele	Allele		Allele	Allele		Allele														
LuSv_MHCIIe2_SR_009	Allele	Allele	Allele					Allele	Allele	Allele																		
LuSv_MHCIIe2_SR_010																												
LuSv_MHCIIe2_SR_013	Allele			Allele				Allele			Allele			Allele														
LuSv_MHCIIe2_SR_015								Allele			Allele																	
LuSv_MHCIIe2_SR_016	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_017																												
LuSv_MHCIIe2_SR_018																												
LuSv_MHCIIe2_SR_019															Allele	Allele		Allele	Allele		Allele	Allele	Allele		Allele	Allele		Allele
LuSv_MHCIIe2_SR_020								Allele	Allele													Allele	Allele					
LuSv_MHCIIe2_SR_021	Allele			Allele			Allele								Allele			Allele										
LuSv_MHCIIe2_SR_022	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele			Allele		
LuSv_MHCIIe2_SR_023																												
LuSv_MHCIIe2_SR_024																												
LuSv_MHCIIe2_SR_025								Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_026																												
LuSv_MHCIIe2_SR_027																												
LuSv_MHCIIe2_SR_028								Allele	Allele						Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_029																												
LuSv_MHCIIe2_SR_030	Allele	Allele		Allele			Allele	Allele	Allele		Allele			Allele														
LuSv_MHCIIe2_SR_031	Allele	Allele	Allele	Allele		Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_032	Allele	Allele		Allele	Allele		Allele								Allele	Allele		Allele	Allele		Allele							
LuSv_MHCIIe2_SR_033																												
LuSv_MHCIIe2_SR_034																												
LuSv_MHCIIe2_SR_035																												
LuSv_MHCIIe2_SR_036	Allele				Allele			Allele				Allele										Allele				Allele		
LuSv_MHCIIe2_SR_037								Allele			Allele			Allele														
LuSv_MHCIIe2_SR_038																												
LuSv_MHCIIe2_SR_039	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele	Allele		Allele		
LuSv_MHCIIe2_SR_040								Allele		Allele	Allele		Allele	Allele								Allele		Allele	Allele			Allele
LuSv_MHCIIe2_SR_041																												
LuSv_MHCIIe2_SR_042																												
LuSv_MHCIIe2_SR_043	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_044																												
LuSv_MHCIIe2_SR_045																												
LuSv_MHCIIe2_SR_046																												
LuSv_MHCIIe2_SR_047															Allele	Allele	Allele	Allele				Allele	Allele	Allele	Allele			Allele
LuSv_MHCIIe2_SR_048																												
LuSv_MHCIIe2_SR_049	Allele	Allele		Allele			Allele								Allele	Allele		Allele			Allele	Allele	Allele		Allele			Allele
LuSv_MHCIIe2_SR_050															Allele	Allele		Allele	Allele		Allele	Allele	Allele		Allele	Allele		Allele
LuSv_MHCIIe2_SR_051															Allele			Allele			Allele	Allele			Allele			Allele
LuSv_MHCIIe2_SR_052	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele						Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_053															Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_054	Allele		Allele	Allele		Allele	Allele	Allele		Allele	Allele		Allele	Allele														
LuSv_MHCIIe2_SR_055	Allele	Allele													Allele	Allele												
LuSv_MHCIIe2_SR_056																												
LuSv_MHCIIe2_SR_057	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_058															Allele	Allele		Allele	Allele		Allele	Allele	Allele		Allele			Allele
LuSv_MHCIIe2_SR_059	Allele				Allele										Allele	Allele			Allele			Allele	Allele			Allele		
LuSv_MHCIIe2_SR_060	Allele	Allele		Allele	Allele		Allele	Allele	Allele		Allele	Allele		Allele														
LuSv_MHCIIe2_SR_061															Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	
LuSv_MHCIIe2_SR_062																												
LuSv_MHCIIe2_SR_063																												
LuSv_MHCIIe2_SR_064	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_065	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_066															Allele			Allele			Allele							
LuSv_MHCIIe2_SR_067	Allele				Allele																							
LuSv_MHCIIe2_SR_068								Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_069	Allele	Allele	Allele	Allele	Allele		Allele								Allele	Allele	Allele	Allele	Allele		Allele							
LuSv_MHCIIe2_SR_070																												
LuSv_MHCIIe2_SR_071																												
LuSv_MHCIIe2_SR_072	Allele				Allele																	Allele				Allele		
LuSv_MHCIIe2_SR_073	Allele	Allele	Allele	Allele		Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele														
LuSv_MHCIIe2_SR_074																												
LuSv_MHCIIe2_SR_075	Allele	Allele		Allele			Allele								Allele	Allele		Allele			Allele							
LuSv_MHCIIe2_SR_076															Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_077															Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele
LuSv_MHCIIe2_SR_078																												
LuSv_MHCIIe2_SR_079								Allele			Allele			Allele								Allele			Allele			Allele
LuSv_MHCIIe2_SR_080	Allele			Allele			Allele	Allele			Allele																	
LuSv_MHCIIe2_SR_081	Allele	Allele	Allele	Allele	Allele	Allele									Allele	Allele	Allele	Allele	Allele	Allele								
LuSv_MHCIIe2_SR_082																												
LuSv_MHCIIe2_SR_083	Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_084																												
LuSv_MHCIIe2_SR_085	Allele	Allele			Allele			Allele	Allele			Allele			Allele	Allele			Allele			Allele	Allele			Allele		
LuSv_MHCIIe2_SR_086																												
LuSv_MHCIIe2_SR_087															Allele	Allele		Allele			Allele	Allele	Allele		Allele			Allele
LuSv_MHCIIe2_SR_088	Allele	Allele						Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_089															Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_090															Allele	Allele	Allele	Allele		Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele
LuSv_MHCIIe2_SR_091																												
LuSv_MHCIIe2_SR_092																												
LuSv_MHCIIe2_SR_093																												
LuSv_MHCIIe2_SR_094																												
LuSv_MHCIIe2_SR_095								Allele	Allele	Allele	Allele		Allele	Allele								Allele	Allele	Allele	Allele		Allele	Allele
LuSv_MHCIIe2_SR_096								Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_097	Allele		Allele	Allele		Allele	Allele	Allele		Allele			Allele															
LuSv_MHCIIe2_SR_098								Allele			Allele			Allele								Allele	Allele		Allele			Allele
LuSv_MHCIIe2_SR_099																												
LuSv_MHCIIe2_SR_100																												
LuSv_MHCIIe2_SR_101	Allele		Allele	Allele	Allele	Allele	Allele	Allele		Allele	Allele		Allele	Allele														
LuSv_MHCIIe2_SR_102	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele														
LuSv_MHCIIe2_SR_103	Allele	Allele		Allele	Allele		Allele	Allele	Allele		Allele	Allele		Allele														
LuSv_MHCIIe2_SR_104																												
LuSv_MHCIIe2_SR_105	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele														
LuSv_MHCIIe2_SR_106															Allele	Allele		Allele			Allele	Allele			Allele			Allele
LuSv_MHCIIe2_SR_107															Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele			Allele			Allele
LuSv_MHCIIe2_SR_108																												
LuSv_MHCIIe2_SR_109																												
LuSv_MHCIIe2_SR_110	Allele	Allele	Allele	Allele	Allele		Allele								Allele		Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_111	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_112																												
LuSv_MHCIIe2_SR_113								Allele	Allele	Allele	Allele	Allele	Allele	Allele								Allele	Allele	Allele	Allele	Allele		Allele
LuSv_MHCIIe2_SR_114															Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_115								Allele	Allele	Allele	Allele		Allele	Allele								Allele	Allele	Allele	Allele		Allele	Allele
LuSv_MHCIIe2_SR_116																												
LuSv_MHCIIe2_SR_117																												
LuSv_MHCIIe2_SR_118	Allele		Allele	Allele	Allele	Allele									Allele		Allele	Allele	Allele	Allele	Allele							
LuSv_MHCIIe2_SR_119																												
LuSv_MHCIIe2_SR_120	Allele	Allele		Allele			Allele	Allele	Allele		Allele			Allele														
LuSv_MHCIIe2_SR_121	Allele				Allele			Allele	Allele	Allele	Allele		Allele	Allele								Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_122																												
LuSv_MHCIIe2_SR_123	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele			Allele			Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
LuSv_MHCIIe2_SR_124																												
LuSv_MHCIIe2_SR_125	Allele	Allele		Allele	Allele		Allele								Allele	Allele		Allele	Allele		Allele							
LuSv_MHCIIe2_SR_126																												
LuSv_MHCIIe2_SR_127															Allele			Allele			Allele	Allele			Allele			Allele
LuSv_MHCIIe2_SR_128																												

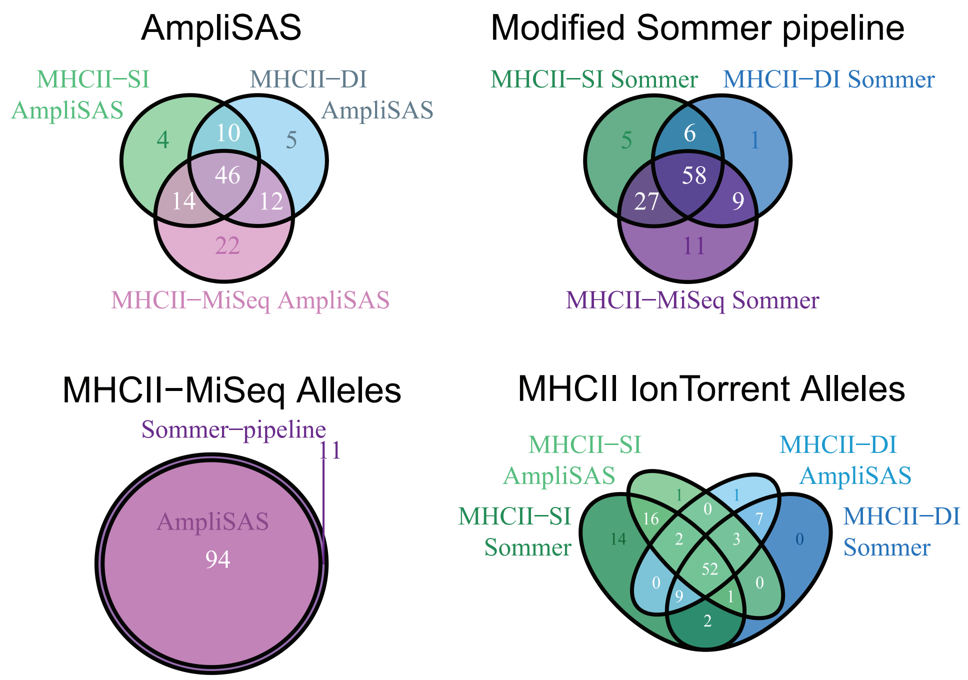
Appendix S11: Sequence logo, MHCIIβe2



Appendix S12: Network MHCIIβe2



Appendix S13: Venn diagram MHCIIβe2



II

Photo: Bjørn Aksel Bjerke

Received: 5 July 2019 | Revised: 30 September 2019 | Accepted: 14 October 2019
DOI: 10.1111/mec.15273

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Extra-pair mating in a passerine bird with highly duplicated major histocompatibility complex class II: Preference for the golden mean

Silje L. Rekdal | Jarl Andreas Anmarkrud | Jan T. Lifjeld | Arild Johnsen

Natural History Museum, University of Oslo, Oslo, Norway

Correspondence
Silje L. Rekdal, Natural History Museum, Blindern, Oslo, Norway.
Email: silje.rekdal@gmail.com

Funding information
Norwegian Research Council, Grant/Award Number: 107585/V40 and 146984/432; Nansen Endowment, Grant/Award Number: 147/98; Natural History Museum, University of Oslo, Norway

Abstract

Genes of the major histocompatibility complex (MHC) are essential in vertebrate adaptive immunity, and they are highly diverse and duplicated in many lineages. While it is widely established that pathogen-mediated selection maintains MHC diversity through balancing selection, the role of mate choice in shaping MHC diversity is debated. Here, we investigate female mating preferences for MHC class II (MHCII) in the bluethroat (*Luscinia svecica*), a passerine bird with high levels of extra-pair paternity and extremely duplicated MHCII. We genotyped family samples with mixed brood paternity and categorized their MHCII alleles according to their functional properties in peptide binding. Our results strongly indicate that females select extra-pair males in a nonrandom, self-matching manner that provides offspring with an allelic repertoire size closer to the population mean, as compared to offspring sired by the social male. This is consistent with a compatible genes model for extra-pair mate choice where the optimal allelic diversity is intermediate, not maximal. This golden mean presumably reflects a trade-off between maximizing pathogen recognition benefits and minimizing autoimmunity costs. Our study exemplifies how mate choice can reduce the population variance in individual MHC diversity and exert strong stabilizing selection on the trait. It also supports the hypothesis that extra-pair mating is adaptive through altered genetic constitution in offspring.

KEYWORDS
bluethroat, compatibility, extra-pair mating, major histocompatibility complex, mate choice, optimal MHC

1 | INTRODUCTION

Genes of the major histocompatibility complex (MHC) constitute an important part of the adaptive immune system in vertebrates. They code for proteins that present intracellular (MHC class I; MHCI) and extracellular (MHC class II; MHCII) pathogen-derived antigens to T-cells and hence trigger an immune response against the specific pathogens (Janeway, Travers, Walport, & Shlomchik, 2001). The arms race between hosts and parasites contributes to the maintenance of extensive polymorphism through balancing pathogen-mediated selection (Spurgin & Richardson, 2010) via heterozygote overdominance (Doherty & Zinkernagel, 1975; Hughes & Nei, 1988, 1989), negative frequency-dependent selection (Bodmer, 1972; Slade & McCallum, 1992; Snell, 1968; Takahata & Nei, 1990) and fluctuating selection (Hamilton & Zuk, 1982; Hedrick, 2002; Hedrick, Thomson, & Klitz, 1987). In addition to heterozygosity, gene duplications can expand intraindividual MHC repertoire and increase the number of pathogens that can be combatted (Nei, Gu, & Sitnikova, 1997).

The number of intraindividual MHC loci varies among species (e.g., Minias, Pikus, Whittingham, & Dunn, 2018). Duplications of MHC loci are presumably selected for when hosts are exposed to a broad array of pathogens (Westerdahl, Wittzell, & von Schantz, 2000). However, there are also costs of having a large number of MHC alleles; the risk of autoimmune diseases increases and the immune system could be less efficient due to negative selection of T-cells in the thymus (Lenz et al., 2015; Migalska, Sebastian, & Radwan, 2019; Nowak, Tarczy-Hornoch, & Austyn, 1992; Vidović & Matzinger, 1988, but see Borghans, Noest, & De Boer, 2003). This will lead to a trade-off in the number of intraindividual MHC alleles, and the optimal number could thus be expected to be intermediate rather than maximal (Kalbe et al., 2009; Wegner, Kalbe, Kurtz, Reusch, & Milinski, 2003; Woelfing, Traulsen, Milinski, & Boehm, 2009). It is conceivable that selection for an intermediate number of MHC alleles, i.e., the golden mean (Woelfing et al., 2009), could be especially pronounced in species possessing many MHC loci.

Whenever MHC affects fitness, females would be selected to choose a mate that will give rise to an optimal MHC diversity in the offspring (Milinski, 2006; Woelfing et al., 2009). Accordingly, genes of the MHC have been suggested to be candidate genes underlying female mate choice (Edwards & Hedrick, 1998; Penn & Potts, 1999; Yamazaki et al., 1976), but decades of studies of MHC and mate choice have rendered equivocal results (Kamiya, O'Dwyer, Westerdahl, Senior, & Nakagawa, 2014; Piertney & Oliver, 2006). If a female is capable of assessing her own MHC, she should choose a mate with a compatible genotype so that the diversity in the offspring will be optimal (Penn & Potts, 1999; Trivers, 1972). A preference for maximal MHC-dissimilar mates has been found in many vertebrates (e.g., Freeman-Gallant, Meguerdichian, Wheelwright, & Sollecito, 2003; Landry, Garant, Duchesne, & Bernatchez, 2001; Olsson et al., 2003; Strandh et al., 2012; Wedekind, Seebeck, Bettens, & Paepke, 1995; Yamazaki et al., 1976), while other studies have indicated choice for mates with intermediate dissimilarity (e.g., Baratti et al., 2012; Bonneaud, Chastel, Federici, Westerdahl, & Sorci, 2006; Eizaguirre, Yeates, Lenz, Kalbe, & Milinski, 2009; Forsberg, Dannewitz, Petersson, & Grahn, 2007), in line with the theoretical framework of an intermediate optimum. On the other hand, if females are not capable of self-referencing, they might choose mates with an optimal MHC diversity. Choice of the most MHC-diverse males has also been demonstrated in several taxa (e.g., fish [Reusch, Häberli, Aeschlimann, & Milinski, 2001], mammals [Ditchkoff, Lochmiller, Masters, Hoofer, & Bussche, 2001; Winternitz, Abbate, Huchard, Havlíček, & Garamszegi, 2017] and birds [Bonneaud et al., 2006; Dunn, Bollmer, Freeman-Gallant, & Whittingham, 2013; Richardson, Komdeur, Burke, & von Schantz, 2005; Whittingham, Freeman-Gallant, Taff, & Dunn, 2015]), while other studies have supported selection for males with an intermediate MHC diversity (e.g., Jäger et al., 2007; Slade, Watson, & MacDougall-Shackleton, 2017). In contrast, Dearborn et al. (2016) suggested that the benefits of MHC-based mate choice will be reduced in species with duplicated and diverged MHC loci, because diverse multilocus genotypes will then be inherited also under random mating. This could possibly explain the lack of MHC-based mate choice found by several studies (e.g., Paterson & Pemberton, 1997; Sepil et al., 2015; Westerdahl, 2004). Extending this argument further, MHC-based mate choice should be less pronounced in species with a high number of MHC loci.

Studying species that are socially monogamous but exhibit extra-pair paternity offers an opportunity to gain insights into the genetic basis of mate choice. While the social male might be chosen for his territory quality and parental abilities, the extra-pair male usually contributes only sperm and might be chosen for genetic benefits (Mays & Hill, 2004). Optimization of offspring MHC might confer such benefits, due to the importance of MHC in the defence against fast-evolving parasites (Milinski, 2006), but whether this holds true for species with extreme levels of MHC-diversity is not known.

Passerine birds are generally characterized by polygenic and polymorphic MHC (Westerdahl, 2007). In this study, we investigated the significance of MHC-based mate choice in a passerine species with highly duplicated MHC. The bluethroat (*Luscinia svecica*, Linnaeus, 1758) provides an excellent study system as it is among the bird species with the highest intraindividual MHCII diversity known to date (minimum 28 loci; Rekdal, Anmarkrud, Johnsen, & Lifjeld, 2018), and has an extensive extra-pair mating system (i.e., about 50% of the nests have extra-pair offspring; Johnsen & Lifjeld, 2003). Intriguingly, immunogenetic benefits of extra-pair copulations have indeed been suggested for this species: Johnsen, Andersen, Sunding, and Lifjeld (2000) and Fossøy, Johnsen, and Lifjeld (2008) found a higher immune response in extra-pair offspring than in both their maternal and paternal half-siblings. This suggests a preference for compatible genes in extra-pair mate choice in the bluethroat, which implies variable preferences that depend on the chooser's own genotype. In a previous study of this species, there were no correlations between male morphological traits within natural phenotypic variation and male success of extra-pair fertilisations (Johnsen, Lifjeld, Andersson, Örnborg, & Amundsen, 2001), which is consistent with a lack of directional selection on male secondary sexual traits through female choice of extra-pair males.

We based our analyses on bluethroat nests with known mixed paternities and identified genetic sires in order to be able to compare female choice of social males and extra-pair males. If females chose extra-pair males based on MHC compatibility, we predicted a difference in MHC diversity in the combined genotype of the female and her social mate and that of the female and her extra-pair male. Alternatively, if females based their choice on the male MHC genotypes alone, irrespective of their own genotype, we expected to find a difference between the intraindividual MHC diversity of the social male and the extra-pair male. Female MHC-based choice of extra-pair males should consequently lead to a difference in MHC diversity between within-pair and extra-pair offspring. If there was selection for maximized diversity, we expected a higher diversity in extra-pair than in within-pair offspring. Conversely, if there was selection for intermediate diversity, we expected MHC diversity to be more concentrated around an intermediate optimum (i.e., lower variance; Forsberg et al., 2007; Lenz, Eizaguirre, Scharsack, Kalbe, & Milinski, 2009) in extra-pair offspring.

2 | MATERIALS AND METHODS

2.1 | Study population and data collection

The present study is based on part of the same data set as used in Johnsen et al. (2000) and Fossøy et al. (2008). Blood samples were collected from adult and nestling bluethroats from a wild population in Øvre Heimdalen, Øystre Slidre, Norway (61°25′N, 8°52′E) during the spring and summer of 1998 and 1999. Because about 50% of all bluethroat nests in the study population contain one or more extra-pair young (EPY), and about 26% of all offspring are EPY (Johnsen & Lifjeld, 2003), the genetic parentage of the chicks were decided based on microsatellites in the previously published studies (Fossøy et al., 2008; Johnsen et al., 2000). In this study, we included 279 individuals from 38 complete families in which both females, EPY, within-pair young (WPY), social males (WPM) and extra-pair males (EPM) were known and sampled. DNA was extracted using E-Z 96 Blood DNA Kit (Omega Bio-Tek Inc. [D1199-01]), following the manufacturer's protocol.

2.2 | Sequencing and allele calling of MHCIIβe2

All DNA samples were amplified in duplicates using the primers MHCIIFihy-E2CF and MHCIIFihy-E2CR (Canal, Alcaide, Anmarkrud, & Potti, 2010) and the sample indexing setup described by Fadrosh et al. (2014). Details regarding PCR conditions and thermal profile are presented elsewhere (Rekdal et al., 2018). The amplicons were sequenced on an Illumina MiSeq instrument using v3 chemistry. The workflow used to call the MHCIIβe2 alleles resembles closely the pipeline outlined by Rekdal et al. (2018), which is based upon the allele identification methodology published by Sommer, Courtiol, and Mazzoni (2013). The use of replicates and family information facilitated allele calling. Several measures were taken to avoid artefacts. In short, this included reducing the number of PCR cycles to 25 to minimize artefact formation (Lenz & Becker, 2008), as well as implementing a strict filtering scheme in the allele calling process. For details, see Appendix S1.

We successfully genotyped 24 females and 35 males sampled in 1998, 12 females and 21 males sampled in 1999 (one female and three males were recaptures from 1998), as well as 98 WPY and 86 EPY in total from the 38 nests from both years. Because two female samples failed, we had only 36 complete trios (female, social male and extra-pair male) left for analysis, but 38 male duos (social and extra-pair male). Five offspring samples also failed during sequencing, leaving no WPY genotyped for one nest and no EPY genotyped for another. Thus, only 36 nests contained both WPY and EPY and were used for paired comparisons of the two groups. Several males sired both WPY and EPY in this data set. Every EPM also sired WPY in their own nest, although not all of these nests were included in this study. All pairs (i.e., combinations of male and female identity) were unique.

2.3 | Establishing genotypes for PSS alleles and supertypes

In order to consider the functional aspects of the MHC alleles, we employed the program CodeML in the package PAML (Yang, 2007) to identify sites under positive selection (positively selected sites; PSS). These are sites that probably are under pathogen-mediated selection, and thus presumably are important in antigen binding and hence the function of MHCII (Hughes & Nei, 1989; Sepil, Moghadam, Huchard, & Sheldon, 2012; Yang & Swanson, 2002). CodeML uses a codon substitution model on the sequence phylogeny to accomplish a likelihood ratio test (LRT), comparing a model with no positive selection (M7: $dN/dS < 1$) with a model that allow positive selection at amino acid sites (M8: $dN/dS > 1$). As the M8 model fitted the data significantly better than M7 (see Appendix S2), a Bayes Empirical Bayes-procedure (BEB) was used to identify sites under positive selection ($p > 95\%$) through a maximum likelihood framework. Of the 12 residues identified as PSS by CodeML, eight have also been described as antigen binding residues in bluethroats (Gohli et al., 2013, which again is based on the PBR of human MHCII by Tong et al., 2006) and in other passerines (Balakrishnan et al., 2010). These eight residues were thus selected as the basis for PSS sequences in this study: 4, 6, 8, 23, 25, 52, 55 and 66.

The unique PSS sequences were further subdivided into supertypes based on the physiochemical properties of the amino acid residues through *z*-descriptors (Sandberg, Eriksson, Jonsson, Sjöström, & Wold, 1998), aiming to group the PSS sequences with similar antigen binding properties (Doytchinova & Flower, 2005; Sepil et al., 2012). The R package adegenet (Jombart, 2008) was employed to infer clusters (i.e., supertypes) by *k*-means clustering (Doytchinova & Flower, 2005). There were 20 supertypes inferred from the PSS sequences. Details are given in Appendix S2. The number of unique PSS sequences per supertype ranged from four (cluster 5) to 25 (cluster 8), with an average of 15.6.

We designated genotypes for PSS alleles and supertypes for each individual based on their nucleotide genotype (Appendices S2 and S5), and used the PSS and supertype genotypes in all downstream analyses. We also obtained the number and identity of unique PSS alleles and supertypes within a pair, for all established pairs (social pairs [henceforth WPM-F], and extra-pair partners [EPM-F]).

2.4 | Statistical analyses

2.4.1 | Male diversity

To test for female choice of EPM for maximum male diversity, we computed the number of unique PSS alleles, sum of the amino acid distance between all pairs of unique PSS alleles, average amino acid distance between all pairs of unique PSS alleles and the number of supertypes for the individual males. Paired *t* tests were run to evaluate if there were any differences in the mean values of the parameters for WPM and EPM.

If females choose males with an intermediate number of alleles as EPM, we expect that the observed values will be more concentrated around an optimum in EPM than in WPM. We thus tested if the variances in the number of unique PSS alleles and supertypes differed between WPM and EPM using Levene's test (Brown-Forsythe type; Brown & Forsythe, 1974) for equality of variances, in the R package car (Fox & Weisberg, 2018). The intraindividual number of PSS alleles was further examined by paired *t* test, testing the distance to yearly population mean for WPM and EPM.

We built linear models to test if there were any significant correlations between the number of PSS alleles in the females and the males (WPM/EPM), in all observed pairs.

2.4.2 | Compatibility

If the females choose EPM based on her own genotype in order to maximize the MHCII diversity in her offspring, only the unique, nonshared alleles of the pair found exclusively within the males would be relevant for her choice. For every observed pair, we thus established these parameters: the number of nonshared PSS alleles only found within the male, sum of the amino acid distance between all pairs of nonshared PSS alleles only found within the male, average amino acid distance between all pairs of nonshared PSS alleles only found within the male and the number of nonshared supertypes only found within the male. Paired *t* tests were conducted in order to test if there were any differences in the abovementioned parameters between WPM-F and EPM-F within nests. We also employed Welch's unequal variances *t* test to further explore the differences between WPM-F and EPM-F in the total number of unique PSS alleles found within a pair.

As for WPM and EPM, we employed Levene's test to test the equality of variances in the observed number of PSS alleles and supertypes in WPM-F and EPM-F. If females choose males who render an intermediate number of alleles in the pair as EPMs, we correspondingly expect a smaller variance in EPM-F than in WPM-F. The difference between WPM-F and EPM-F in their distance to yearly population mean number of PSS alleles in the pair was also tested using a paired *t* test as well as Welch's unequal variances *t* test. Further, we ran simulations to test whether the observed numbers of PSS alleles within mating pairs differed from a random model. For each run in the simulations, we paired each female with a random male sampled in the same year, calculated the number of unique PSS alleles for the pair and listed its deviation (absolute value) from the overall population mean. We then calculated and recorded the mean deviation across all 36 females in our sample. This procedure was iterated 10,000 times, which yielded a distribution of 10,000 means, to which the observed means for social pairs and extra-pair partners could be compared.

2.4.3 | Offspring

Mate choice can also be tracked in the genotypes of offspring. If there is selection for females to choose EPM that will maximize MHCII diversity in the offspring, we expect EPY to have a higher number of PSS alleles or supertypes than WPY. We thus performed paired *t* tests on the number of PSS alleles and the number of supertypes in WPY and EPY, paired within nests. We also used the R package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) to build a linear mixed model of the correlation between the number of PSS alleles (response) and the status of the offspring (WPY/EPY; fixed effect), with dam and sire identity as random factors with random intercepts. We further established a corresponding linear mixed model to test if the distance to the population average number of PSS alleles was different in WPY and EPY. For the mixed models, we square root transformed the response variables (i.e., the number of PSS alleles and the distance to population average in absolute numbers of PSS alleles) to attain normality. We also performed a paired *t* test to test for a difference in the distance to the population average number of PSS alleles between WPY and EPY within nests. Similarly as for the adults, we carried out tests for equality of variances in the intraindividual number of PSS alleles (*F*-ratio test; see below) and supertypes (Levene's test) between WPY and EPY. Given the results we obtained from the adults, we had an expectation of EPY being closer to the population mean in the intraindividual number of PSS alleles than WPY. We thus applied one-sided tests testing this hypothesis (valid for two tests: the paired *t* test testing if EPY had a smaller distance to the population average number of PSS alleles than WPY, and a one-sided F-ratio test testing if EPY had a smaller variance in the number of PSS alleles than WPY). We square root transformed the intraindividual numbers of PSS alleles for the latter test, as it is highly sensitive to deviation from normality. For the variance tests and paired *t* tests in offspring, one WPY and one EPY were chosen randomly from each nest and used in the analyses, in order to avoid pseudoreplication. The analyses were repeated 10,000 times, and the resulting mean *t*- and *F*-values were used for calculation of *p*-values.

2.4.4 | For all analyses

Only the male with the highest number of offspring in the respective nest was included when there were two EPMs genotyped for one nest (relevant in four nests in this data set). All the statistical analyses were conducted in R (R Core Team, 2016, see Rekdal, Anmarkrud, Lifjeld, & Johnsen, 2019 for scripts and input data), while the amino acid distances were calculated by MEGA7 (Kumar, Stecher, & Tamura, 2016). The distance to the population average number of PSS alleles was square root transformed to endeavor normality in all tests using this variable. Visual inspection of QQ-plots and Shapiro–Wilk tests revealed normality for all tests (data not shown), with exception of slight deviation from normality in the Welch's *t* test between WPM-F and EPM-F in their distance

TABLE 1 The number of unique MHC class II variants at the different sequence levels, found across all individuals (adults and offspring) and within individuals (given as mean values and range)

Sequence level	Number of unique variants across all individuals	Alleles per individual	Alleles per adult	Alleles per offspring
Nucleotide	1,176	35.2 (16–58)	37.7 (16–58)	34.1 (19–52)
Translated	890	31.3 (16–47)	33.4 (16–47)	30.3 (18–45)
PSS	311	18.6 (7–32)	19.8 (11–32)	18.0 (7–32)
Supertypes	20	12.0 (5–18)	12.4 (8–18)	11.7 (5–18)

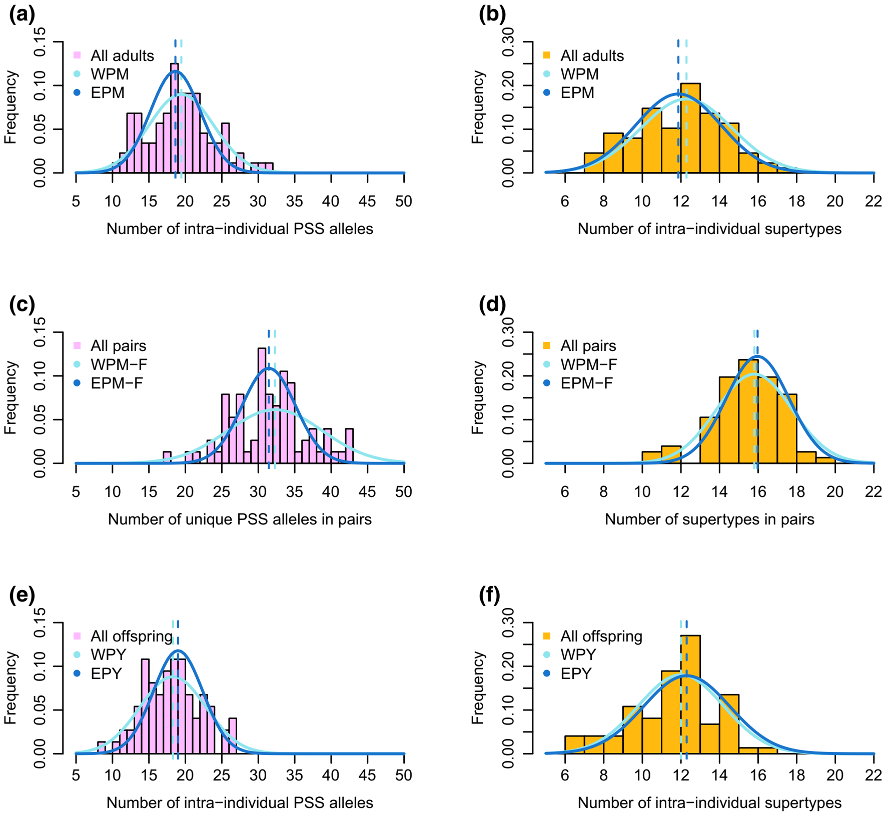


FIGURE 1 Frequency plots of the number of unique MHC class II PSS alleles (positively selected sites; left) and supertypes (right) found within all individual adults (top panel [a, b]), all observed partners (middle panel [c, d]) and offspring (bottom panel [e, f]). The observed values for social and extra-pair males (WPM/EPM [a, b]), social and extra-pair partners (WPM-F/EPM-F [c, d]) and within-pair and extra-pair young (WPY/EPY [e, f]) are visualized as normalized curves in light (WPM, WPM-F, WPY) and dark (EPM, EPM-F, EPY) blue. The mean of the normalized curves are given as dashed lines and coloured correspondingly

to yearly population mean number of PSS alleles in the pair. We approximated the optimal intermediate number of PSS alleles as the mean intraindividual number of alleles across all adults sampled within a year (1998:19.9 PSS alleles per individual, 1999:19.8 PSS alleles per individual), and the mean number of unique alleles within a pair over all possible pairs within a year (1998:32.0 PSS alleles per pair, 1999:32.7 PSS alleles per pair). This agrees with work on sticklebacks (*Gasterosteus aculeatus*) showing that the estimated optimum number of alleles is close to population average (Aeschlimann, Häberli, Reusch, Boehm, & Milinski, 2003). All significant tests remained significant after controlling for multiple testing using false discovery rate (Benjamini & Hochberg, 1995; $Q = 0.1$).

3 | RESULTS

3.1 | High number of MHCII alleles within individuals

The 1,176 nucleotide MHCII sequences translated into 890 unique amino acid sequences. When considering only the eight amino acid residues recognized as PSS, the sequences grouped into 311 unique PSS alleles, which further were divided into 20 supertypes (see Table 1 and Appendix S3). Across all adults sampled both years, the mean intraindividual number of nucleotide alleles, PSS alleles and supertypes were 37.7 (*SD* = 8.28), 19.8 (*SD* = 4.60) and 12.4 (*SD* = 2.32), respectively. The frequencies of each number of unique PSS alleles and supertypes within individual adults, pairs and offspring are visualized in Figure 1.

3.2 | Do females choose EPM based on male MHCII diversity?

There were no indications of females choosing males with maximized diversity as EPM, as there were no significant differences in the mean parameter values between WPM and EPM (paired *t* tests; see Appendix S4). EPMs did also not have a number of PSS alleles that was closer to the population mean than that of WPMs (paired *t* tests; $t_{37} = 1.34$, $p = .19$). Correspondingly, the variances in the individual number of PSS alleles and supertypes within WPM and EPM were not significantly different (Levene's test for equality of variances; $F_{1,74} = 2.93$, $p = .091$ [PSS alleles; Figure 1a] and $F_{1,74} = 0.026$, $p = .87$ [supertypes; Figure 1b]).

The numbers of PSS alleles found within the females and EPMs were negatively correlated (linear models; $R^2_{\text{adj}} = 0.16$, $F_{1,34} = 7.51$, $p = .0097$), but were uncorrelated between females and WPMs ($R^2_{\text{adj}} = 0.012$, $F_{1,34} = 1.42$, $p = .24$). The corresponding regression slopes for EPM and WPM were significantly different ($F_{1,68} = 6.28$, $p = .015$, see Figure 2). In other words, this suggests that females with few PSS alleles tended to choose EPM with many PSS alleles and vice versa, but this pattern was not found for female choice of WPM.

3.3 | Do females choose EPM based on compatibility at MHCII?

Females did not consistently choose a more dissimilar extra-pair male than their social male (paired *t* tests; see Appendix S4). There was further no significant difference in the average number of unique PSS alleles in the pair between WPM-F pairs and EPM-F pairs (Welch's *t* test; $t_{55.3} = 0.69$, $p = .49$, see Figure 3a). However, the number of unique PSS alleles in EPM-F pairs showed less variance than in WPM-F pairs (Levene's test; $F_{1,70} = 9.54$, $p = .0029$; Figure 1c), but not so for the number of supertypes ($F_{1,70} = 0.99$, $p = .32$; Figure 1d). EPM-F pairs were also significantly closer to the population mean number of unique PSS alleles within pairs than WPM-F pairs (paired *t* test; $t_{35} = 3.05$, $p = .0043$, effect size = 0.62, Welch's *t* test; $t_{64.4} = 2.65$, $p = .010$, see Figure 3b).

Assuming random mating, we calculated 10,000 averages of the number of unique PSS alleles within pairs across all females, and their deviations from the population mean. Figure 4 shows the distribution of these deviations. We used this distribution to test if the observed average number of unique alleles in EPM-F and WPM-F deviated significantly from a random model. We found that only 33 of the 10,000 simulated means had a lower value than the observed EPM-F (exact test: $p = .0068$, Figure 4), while the observed mean value for WPM-F was ranked as the 9091st observation when sorted ascendingly (exact test: $p = .18$).

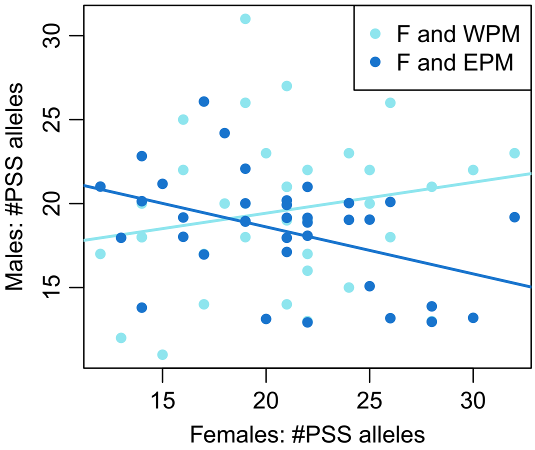


FIGURE 2 The number of unique MHC class II PSS (positively selected sites) alleles genotyped within each individual bluethroat in all observed pairs, divided in females and their social males (F and WPM; light blue), and females and their extra-pair males (F and EPM; dark blue)

3.4 | Offspring

In line with the results from the adults, we found no support for female choice of EPM for maximized diversity in the offspring data: there were no differences in mean number of alleles between WPY and EPY, neither in the paired *t* tests (PSS alleles: $t_{35} = 0.12$, $p = .91$, supertypes: $t_{35} = -0.45$, $p = .66$) nor in the linear mixed model for PSS alleles ($t_{54.7} = -0.26$, $p = .80$, see Figure 3c). Furthermore, in line with the above results for mate choice, we found that EPY had a number of PSS alleles that was significantly closer to the population average than that of WPY (linear mixed model: estimate = 0.31, $t_{49.6} = 2.68$, $p = .0099$, see Figure 3d, paired *t* test: $t_{35} = -2.01$, $p = .026$). Similarly, there was a tendency for EPY to have a smaller variance in the individual number of PSS alleles than WPY (F ratio test; $F_{36,35} = 1.67$, $p = .066$, see Figure 1e), but not in the number of supertypes (Levene's test; $F_{1,71} = 0.39$, $p = .54$, see Figure 1f).

4 | DISCUSSION

We have shown here that extra-pair mating in the bluethroat was nonrandom with respect to MHCII alleles. The number of unique, functional MHCII alleles among extra-pair parents was consistently closer to the population mean, with a significantly reduced variance, than that expected from a random model of pair combinations or that observed among social parents. Consequently, we found that extra-pair offspring received an allelic repertoire closer to the population mean than was the case for within-pair offspring. Our study therefore supports the hypothesis that females engage in extra-pair mating for genetic benefits and suggests that MHCII genes play a significant role in their mating preferences. Our results imply that females may be able to differentiate among alleles according to sequence variation at positively selected nucleotides in the peptide-binding region of the molecule, and not according to other physical properties assumed in the categorization of alleles into supertypes.

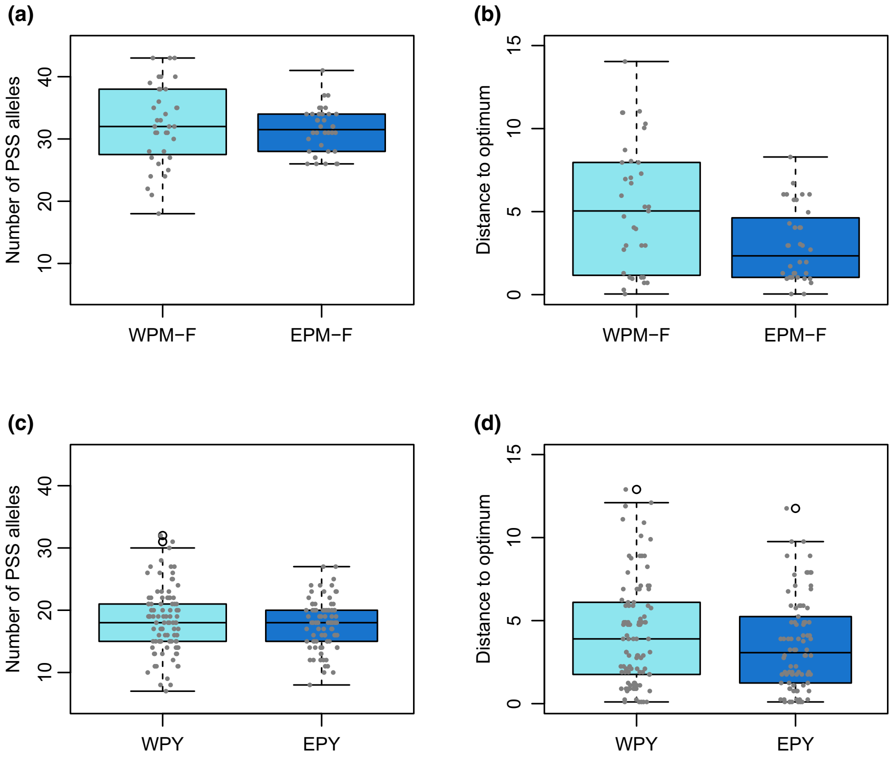


FIGURE 3 Boxplots of the observed values (grey, jittered dots) of the number of unique MHC class II PSS alleles (positively selected sites; a, c) and the absolute distance to the optimum (given in number of unique PSS alleles [nontransformed data]; b, d) within each pair in the data set, divided in social pairs and extra-pair partners (WPM-F/EPM-F [a, b]) and within each offspring, divided in within-pair young and extra-pair young (WPY/EPY [c, d])

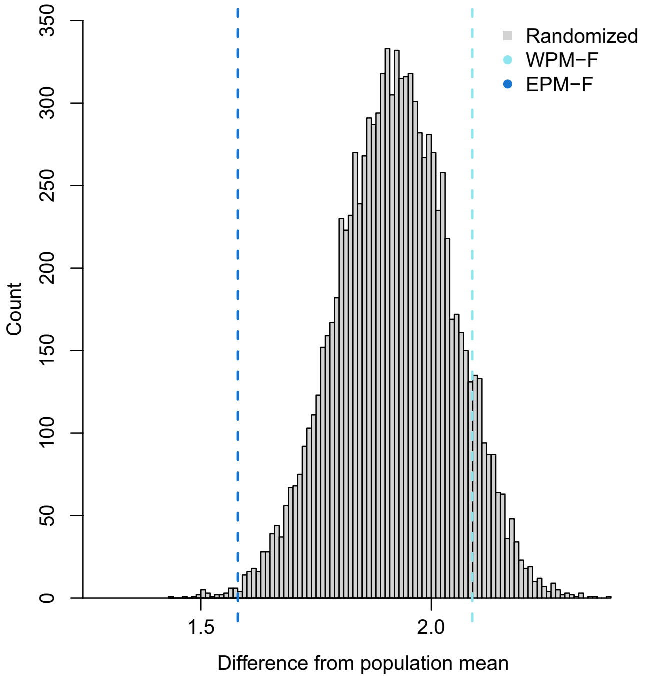


FIGURE 4 Histogram showing 10,000 simulated mean values of the distance to the yearly population mean (given in number of unique MHC class II PSS [positively selected sites] alleles within a pair, square root transformed), allowing each bluethroat female to mate with a random male within the data set. The observed mean values for social pairs (WPM-F) and extra-pair partners (EPM-F) are given as coloured, dashed lines (light blue and dark blue, respectively)

4.1 | MHC-based mate choice realized through extra-pair copulations

Dearborn et al. (2016) suggested that MHC-based mate choice may be superfluous in species with duplicated and diverged MHC loci, since offspring will inherit a diverse MHC genotype irrespective of mate choice. In contrast, we found substantial support for MHCII-based mate choice in a species possessing extensive MHCII duplications. Although we did not find any indications of disassortative mating with respect to MHCII, our results suggest that females choose extra-pair males that will render an intermediate number of functional MHCII alleles in the pair, leading to an intermediate, presumably optimal MHC diversity in extra-pair offspring.

A difference between within-pair and extra-pair units is expected if females choose social males for other than pure genetic benefits, while the extra-pair males might be chosen on the basis of their genes, as they probably only contribute sperm (Johnsen et al., 2001; Trivers, 1972). In the context of MHC, few studies on passerines have tested such differences, and among those which have, the results are mixed. Most studies did not find any support for increased compatibility between the MHC genotypes of the female and extra-pair male, compared to the female and the social male (Bollmer, Dunn, Freeman-Gallant, & Whittingham, 2012; Promerová et al., 2011; Richardson et al., 2005). Yet, Winternitz et al. (2015) found that in the scarlet rosefinch (*Carpodacus erythrinus*), the variance in intraindividual number of MHC alleles was lower in extra-pair offspring than within-pair offspring, which is consistent with our results. They did, however, not report any differences in MHC compatibility between social and extra-pair partners.

Other studies have found that paternity loss from social to extra-pair males could be negatively associated with MHC diversity, based on either the social male's MHC diversity (Promerová et al., 2011; Richardson et al., 2005) or MHC similarity in the social pair (Freeman-Gallant et al., 2003) – a pattern also found in a primate with high levels of extra-pair paternity (Schwensow, Fietz, Dausmann, & Sommer, 2008). While we have only included pairs with confirmed cuckoldry in this study, and hence do not know the MHCII compatibility in pairs with pure WPY broods, we cannot test predictions concerning overall paternity loss in relation to partner compatibility in MHCII genotypes with our data.

4.2 | Intermediate, not maximized MHCII diversity

Our results point to selection for an intermediate optimum number of MHCII alleles, rather than maximized MHCII diversity in the bluethroat. This is in line with the theoretical framework of a trade-off between recognizing a broad array of pathogens, and increased depletion of circulating T-cells following negative selection in the thymus and risk of autoimmune diseases with increased number of MHC alleles (e.g., Gough & Simmonds, 2007; Nowak et al., 1992; Woelfing et al., 2009). What level of intraindividual MHC diversity that constitutes the optimum might vary among species due to ecological differences, e.g., according to the pathogen load experienced (Minias et al., 2018; O'Connor, Cornwallis, Hasselquist, Nilsson, & Westerdahl, 2018; Westerdahl et al., 2000). This implies that the more pathogens a species is exposed to, the stronger the selective force for increased diversity will be, driving the optimum towards a higher diversity. Because the bluethroat is migratory, insectivorous and fairly promiscuous, it probably encounters a multitude of pathogens. While this could explain the large number of MHCII loci in the species (Anmarkrud, Johnsen, Bachmann, & Lifjeld, 2010; Gohli et al., 2013; Rekdal et al., 2018), it is, however, important to emphasize that we did not find support for selection for maximized diversity, but rather indications of stabilizing selection for a relatively high intermediate number of MHCII alleles. Still, stabilizing selection can lead to an increase in individual allelic diversity over evolutionary time, through a moving intermediate optimum.

Fossøy et al. (2008) compared microsatellite multilocus heterozygosity between within-pair and extra-pair units in a bluethroat data set including the data used in this study. They found that females were less genetically similar to the extra-pair male than the within-pair male, which presumably explained their results of extra-pair young being more heterozygous than their maternal within-pair half-siblings. Our results do, however, exhibit a different pattern; instead of increased MHCII diversity, we found that extra-pair partners and extra-pair young had a number of unique PSS alleles that were more concentrated around the population mean than their within-pair counterpart. This deviates from what we would expect based on microsatellites from Fossøy et al. (2008), and indicates that what we observed on MHCII was not due to genome-wide effects.

We restricted our analyses to MHCII. The bluethroat has relatively few MHCI loci (i.e., four; O'Connor, Strandh, Hasselquist, Nilsson, & Westerdahl, 2016; Rekdal et al., 2018), which might be due to less exposure to intra- than extracellular pathogens (Minias et al., 2018) or some compensatory immunological mechanism (e.g., Gangoso et al., 2012; Star et al., 2011). Recent studies have identified a link between MHCII composition and individual odor in birds, possibly mediated through microbial communities and uropygial gland secretions (Leclaire et al., 2019, 2014; Leclaire, Strandh, Mardon, Westerdahl, & Bonadonna, 2017; Slade et al., 2016; Strandh et al., 2012). As there is growing evidence that birds are able to use olfaction in MHC-based mate choice, also in a self-referencing manner (reviewed by Caro, Balthazart, & Bonadonna, 2015), MHCII is a prominent candidate for such a mate choice mechanism.

4.3 | Compatibility, not male diversity

Unlike selection for maximized or intermediate diversity in the male, in which the same males will be the best choice for all females, selection for compatibility implies that the best choice a female can make is dependent on her own genotype (Brown, 1997). One suggested approach for MHC-based mate choice is allele counting, in which females assess the number of MHC alleles in males, and choose mates accordingly (Aeschlimann et al., 2003; Reusch et al., 2001). A trend of mating-up preference by allele counting was supported by Griggio, Biard, Penn, and Hoi (2011), who found that female house sparrows (*Passer domesticus*) with a low number of MHC alleles preferred high diversity males. We found the same tendency in this study; females with a low number of MHCII alleles had males with a high number of alleles as extra-pair males, and those with many alleles had males with fewer alleles as extra-pair males (Figure 2).

Importantly, all our tests on the adult bluethroats rendered significant results only when considering MHCII diversity in the pairs combined, and not in the male alone. These results are consistent with the compatibility framework (Brown, 1997; Trivers, 1972), where females choose mates based on their own genotype in order to produce offspring with an optimal MHC diversity (Penn & Potts, 1999). Indeed, Johnsen et al. (2000) and Fossøy et al. (2008) demonstrated an increased immunocompentence in extra-pair offspring as compared to both their maternal and paternal half-siblings. Taken together, these studies offer extensive support for the compatibility hypothesis, realized through extra-pair mating in the bluethroat.

4.4 | Positive selected sites, not supertypes

The significant results in this study originated from analyses on PSS alleles, and not supertypes. The rationale for further grouping the PSS sequences into supertypes was to focus on a possible higher unit of selection, due to the overlap in binding repertoires by different MHC alleles (Matsumura, Fremont, Peterson, & Wilson, 1992; Sepil et al., 2012; Sette et al., 1989; Sette & Sidney, 1998; Trachtenberg et al., 2003). However, our results do not indicate any female discrimination of MHCII alleles at the level of supertypes. One conceivable explanation for the disparate results could be due to information loss in the grouping of PSS alleles into supertypes. Another possibility could be spurious inference of irrelevant properties of the supertypes. Supertypes might be functionally important in pathogen recognition, but it is possible that females can only discriminated MHCII alleles from information encoded in their nucleotide sequences.

5 | CONCLUSION

In conclusion, this study provides substantial support for extra-pair mating preferences associated with MHCII diversity in a passerine species with highly polymorphic and duplicated MHCII. The results are in agreement with a preference for a golden mean where an intermediate number of alleles in the individual is optimal, given an assumed trade-off between maximizing the range of pathogens that can be combatted and minimizing autoimmunity costs associated with too many alleles. We note, however, that we currently lack fitness data to verify that individuals with intermediate level of PSS alleles have higher survival than those at the more extreme ends of the allele number distribution. Our study provides additional empirical support for the hypothesis that females engage in extra-pair mating nonrandomly. It further suggests that this behaviour is associated with the genetic constitution of the immune system and the survival prospects of offspring under strong pathogen-mediated selection pressures.

ACKNOWLEDGEMENTS

We wish to thank B. A. Bjerke, C. K. Aas, V. Andersen and C. Sunding for help in the field. Thanks are also due to F. Fossøy, G. Bjørnstad and L. Thorbek for assistance in the laboratory. The Illumina MiSeq sequencing was performed at the Norwegian Sequencing Center. The field work was carried out in compliance with the national standards for animal research at the time, which required a permit for mist-netting and ringing birds, but no additional permit for blood sampling. Five anonymous reviewers improved the manuscript. This study was supported by several grants from the Norwegian Research Council (grant numbers: 107585/V40 and 146984/432), as well as a grant from the Nansen Endowment (grant number: 147/98) and a PhD fellowship from Natural History Museum, University of Oslo (to SLR).

AUTHOR CONTRIBUTIONS

S.L.R., J.A.A., J.T.L., and A.J. designed the study. J.A.A. carried out the laboratory work. S.L.R. analyzed the data and drafted the manuscript which was revised by all authors.

ORCID

Silje L. Rekdal https://orcid.org/0000-0003-4593-1709
Jarl Andreas Anmarkrud https://orcid.org/0000-0002-7496-4430
Jan T. Lifjeld https://orcid.org/0000-0002-9172-9985
Arild Johnsen https://orcid.org/0000-0003-4864-6284

DATA AVAILABILITY STATEMENT

MHCIIβe2 nucleotide alleles are available in GenBank (accession numbers MN332585–MN333760). Raw sequence data are available through the NCBI Sequence Read Archive (BioProject accession number PRJNA560776). A bash script with the codes for making the jMHC input file from the raw data, as well as a R script for the statistical analyses and associated input files are deposited in Dryad (https://doi.org/10.5061/dryad.93tf68k). Individual genotypes, number of PSS alleles within pairs, nest affiliations, grouping of nucleotides into PSS alleles and supertypes, as well as nucleotide and PSS sequences can be extracted from Appendix S5. Barcodes and primer sequences are given in Appendix S6. The corresponding GenBank accession number to each nucleotide sequence allele used in this study is given in Appendix S7. Information on sampling and identification for each individual in the data set can be retrieved via the online Collection Explorer (http://nhmo-birds.collectionexplorer.org/accession.aspx), using the accession numbers found in Appendix S5.

REFERENCES

Aeschlimann, P., Häberli, M., Reusch, T., Boehm, T., & Milinski, M. (2003). Female sticklebacks *Gasterosteus aculeatus* use self-reference to optimize MHC allele number during mate selection. *Behavioral Ecology and Sociobiology*, *54*, 119–126. https://doi.org/10.1007/S00265-003-0611-6

Anmarkrud, J., Johnsen, A., Bachmann, L., & Lifjeld, J. (2010). Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes. *Journal of Evolutionary Biology*, *23*, 1206–1217. https://doi.org/10.1111/j.1420-9101.2010.01999.x

Balakrishnan, C. N., Ekblom, R., Völker, M., Westerdahl, H., Godinez, R., Kotkiewicz, H., … Edwards, S. V. (2010). Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biology*, *8*(29). https://doi.org/10.1186/1741-7007-8-29

Baratti, M., Dessì-fulgheri, F., Ambrosini, R., Bonisoli-alquati, A., Caprioli, M., Goti, E., … Saino, N. (2012). MHC genotype predicts mate choice in the ring-necked pheasant *Phasianus colchicus*. *Journal of Evolutionary Biology*, *25*, 1531–1542. https://doi.org/10.1111/j.1420-9101.2012.02534.x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bodmer, W. F. (1972). Evolutionary significance of the HL-A system. *Nature*, *237*, 139–183. https://doi.org/10.1038/237139a0

Bollmer, J. L., Dunn, P. O., Freeman-Gallant, C. R., & Whittingham, L. A. (2012). Social and extra-pair mating in relation to major histocompatibility complex variation in common yellowthroats. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*, 4778–4785. https://doi.org/10.1098/rspb.2012.1885

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. https://doi.org/10.1098/rspb.2005.3325

Borghans, J. A., Noest, A. J., & De Boer, R. J. (2003). Thymic selection does not limit the individual MHC diversity. *European Journal of Immunology*, *33*, 3353–3358. https://doi.org/10.1002/eji.200324365

Brown, J. L. (1997). A theory of mate choice based on heterozygosity. *Behavioral Ecology*, *8*, 60–65. https://doi.org/10.1093/beheco/8.1.60

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*, 364–367. https://doi.org/10.2307/2285659

Canal, D., Alcaide, M., Anmarkrud, J. A., & Potti, J. (2010). Towards the simplification of MHC typing protocols: Targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*. *BMC Research Notes*, *3*, 236. https://doi.org/10.1186/1756-0500-3-236

Caro, S. P., Balthazart, J., & Bonadonna, F. (2015). The perfume of reproduction in birds: Chemosignaling in avian social life. *Hormones and Behavior*, *68*, 25–42. https://doi.org/10.1016/j.yhbeh.2014.06.001

Dearborn, D. C., Gager, A. B., McArthur, A. G., Gilmour, M. E., Mandzhukova, E., & Mauck, R. A. (2016). Gene duplication and divergence produce divergent MHC genotypes without disassortative mating. *Molecular Ecology*, *25*, 4355–4367. https://doi.org/10.1111/mec.13747

Ditchkoff, S. S., Lochmiller, R. L., Masters, R. E., Hoofer, S. R., & Bussche, R. A. V. D. (2001). Major-histocompatibility-complex-associated variation in secondary sexual traits of white-tailed deer (*Odocoileus virginianus*): Evidence for good-genes advertisement. *Evolution*, *55*, 616–625. https://doi.org/10.1111/j.0014-3820.2001.tb00794.x

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

Doytchinova, I. A., & Flower, D. R. (2005). In silico identification of supertypes for class II MHCs. *The Journal of Immunology*, *174*, 7085–7095. https://doi.org/10.4049/jimmunol.174.11.7085

Dunn, P. O., Bollmer, J. L., Freeman-Gallant, C. R., & Whittingham, L. A. (2013). MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats. *Evolution*, *67*, 679–687. https://doi.org/10.1111/j.1558-5646.2012.01799.x

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. https://doi.org/10.1016/S0169-5347(98)01416-5

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. https://doi.org/10.1111/j.1365-294X.2009.04243.x

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, *2*, 6. https://doi.org/10.1186/2049-2618-2-6

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. https://doi.org/10.1111/j.1420-9101.2007.01380.x

Fossøy, F., Johnsen, A., & Lifjeld, J. T. (2008). Multiple genetic benefits of female promiscuity in a socially monogamous passerine. *Evolution*, *62*, 145–156. https://doi.org/10.1111/j.1558-5646.2007.00284.x

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression* (3rd ed.). Thousands Oaks, CA: Sage Publications, Inc.

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. https://doi.org/10.1046/j.1365-294X.2003.01968.x

Gangoso, L., Alcaide, M., Grande, J. M., Muñoz, J., Talbot, S. L., Sonsthagen, S. A., … Figuerola, J. (2012). Colonizing the world in spite of reduced MHC variation. *Journal of Evolutionary Biology*, *25*, 1438–1447. https://doi.org/10.1111/j.1420-9101.2012.02529.x

Gohli, J., Anmarkrud, J. A., Johnsen, A., Kleven, O., Borge, T., & Lifjeld, J. T. (2013). Female promiscuity is positively associated with neutral and selected genetic diversity in passerine birds. *Evolution*, *67*, 1406–1419. https://doi.org/10.1111/evo.12045

Gough, S. C. L., & Simmonds, M. J. (2007). The HLA region and autoimmune disease: Associations and mechanisms of action. *Current Genomics*, *8*, 453–465. https://doi.org/10.2174/138920207783591690

Griggio, M., Biard, C., Penn, D. J., & Hoi, H. (2011). Female house sparrows "count on" male genes: Experimental evidence for MHC-dependent mate preference in birds. *BMC Evolutionary Biology*, *11*, 44. https://doi.org/10.1186/1471-2148-11-44

Hamilton, W. D., & Zuk, M. (1982). Heritable true fitness and bright birds: A role for parasites? *Science*, *218*, 384–387. https://doi.org/10.1126/science.7123238

Hedrick, P. W. (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution*, *56*, 1902–1908. https://doi.org/10.1111/j.0014-3820.2002.tb00116.x

Hedrick, P. W., Thomson, G., & Klitz, W. (1987). Evolutionary genetics and HLA: Another classic example. *Biological Journal of the Linnean Society*, *31*, 311–331. https://doi.org/10.1111/j.1095-8312.1987.tb01996.x

Hughes, A. L., & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, *335*, 167–170. https://doi.org/10.1038/335167a0

Hughes, A. L., & Nei, M. (1989). Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the USA*, *86*, 958–962. https://doi.org/10.1073/pnas.86.3.958

Jäger, I., Eizaguirre, C., Griffiths, S., Kalbe, M., Krobbach, C., Reusch, T., … Milinski, M. (2007). Individual MHC class I and MHC class IIB diversities are associated with male and female reproductive traits in the three-spined stickleback. *Journal of Evolutionary Biology*, *20*, 2005–2015. https://doi.org/10.1111/j.1420-9101.2007.01366.x

Janeway, C. A., Travers, P., Walport, M., & Shlomchik, M. J. (2001). Antigen presentation to T lymphocytes. In C. A. Janeway, P. Travers, M. Walport, & M. J. Shlomchik (Eds.), *Immunobiology: The immune system in health and disease* (5th ed.). New York, NY: Garland Science.

Johnsen, A., Andersen, V., Sunding, C., & Lifjeld, J. T. (2000). Female bluethroats enhance offspring immunocompetence through extra-pair copulations. *Nature*, *406*, 296–299. https://doi.org/10.1038/35018556

Johnsen, A., & Lifjeld, J. T. (2003). Ecological constraints on extra-pair paternity in the bluethroat. *Oecologia*, *136*, 476–483. https://doi.org/10.1007/s00442-003-1286-4

Johnsen, A., Lifjeld, J., Andersson, S., Örnborg, J., & Amundsen, T. (2001). Male characteristics and fertilisation success in bluethroats. *Behaviour*, *138*, 1371–1390. https://doi.org/10.1163/156853901317367645

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kalbe, M., Eizaguirre, C., Dankert, I., Reusch, T. B., Sommerfeld, R. D., Wegner, K. M., & Milinski, M. (2009). Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 925–934. https://doi.org/10.1098/rspb.2008.1466

Kamiya, T., O'Dwyer, K., Westerdahl, H., Senior, A., & Nakagawa, S. (2014). A quantitative review of MHC-based mating preference: The role of diversity and dissimilarity. *Molecular Ecology*, *23*, 5151–5163. https://doi.org/10.1111/mec.12934

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*, 1870–1874. https://doi.org/10.1093/molbev/msw054

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

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 of London. Series B: Biological Sciences*, *268*, 1279–1285. https://doi.org/10.1098/rspb.2001.1659

Leclaire, S., Strandh, M., Dell'Ariccia, G., Gabirot, M., Westerdahl, H., & Bonadonna, F. (2019). Plumage microbiota covaries with the major histocompatibility complex in blue petrels. *Molecular Ecology*, *28*, 833–846. https://doi.org/10.1111/mec.14993

Leclaire, S., Strandh, M., Mardon, J., Westerdahl, H., & Bonadonna, F. (2017). Odour-based discrimination of similarity at the major histocompatibility complex in birds. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *284*, 20162466. https://doi.org/10.1098/rspb.2016.2466

Leclaire, S., van Dongen, W. F. D., Voccia, S., Merkling, T., Ducamp, C., Hatch, S. A., ... Wagner, R. H. (2014). Preen secretions encode information on MHC similarity in certain sex-dyads in a monogamous seabird. *Scientific Reports*, *4*, 6920. https://doi.org/10.1038/srep06920

Lenz, T. L., & Becker, S. (2008). Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci—Implications for evolutionary analysis. *Gene*, *427*, 117–123. https://doi.org/10.1016/j.gene.2008.09.013

Lenz, T. L., Deutsch, A. J., Han, B., Hu, X., Okada, Y., Eyre, S., ... Raychaudhuri, S. (2015). Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. *Nature Genetics*, *47*, 1085. https://doi.org/10.1038/ng.3379

Lenz, T. L., Eizaguirre, C., Scharsack, J. P., Kalbe, M., & Milinski, M. (2009). Disentangling the role of MHC-dependent 'good genes' and 'compatible genes' in mate-choice decisions of three-spined sticklebacks *Gasterosteus aculeatus* under semi-natural conditions. *Journal of Fish Biology*, *75*, 2122–2142. https://doi.org/10.1111/j.1095-8649.2009.02410.x

Matsumura, M., Fremont, D. H., Peterson, P. A., & Wilson, I. A. (1992). Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science*, *257*, 927–934. https://doi.org/10.1126/science.1323878

Mays, J. H. L., & Hill, G. E. (2004). Choosing mates: Good genes versus genes that are a good fit. *Trends in Ecology & Evolution*, *19*, 554–559. https://doi.org/10.1016/j.tree.2004.07.018

Migalska, M., Sebastian, A., & Radwan, J. (2019). Major histocompatibility complex class I diversity limits the repertoire of T cell receptors. *Proceedings of the National Academy of Sciences of the USA*, *116*, 5021–5026. https://doi.org/10.1073/pnas.1807864116

Milinski, M. (2006). The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 159–186. https://doi.org/10.1146/annurev.ecolsys.37.091305.110242

Minias, P., Pikus, E., Whittingham, L. A., & Dunn, P. O. (2018). Evolution of copy number at the MHC varies across the avian tree of life. *Genome Biology and Evolution*, *11*, 17–28. https://doi.org/10.1093/gbe/evy253

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 USA*, *94*, 7799–7806. https://doi.org/10.1073/pnas.94.15.7799

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 USA*, *89*, 10896–10899. https://doi.org/10.1073/pnas.89.22.10896

O'Connor, E. A., Cornwallis, C. K., Hasselquist, D., Nilsson, J.-Å., & Westerdahl, H. (2018). The evolution of immunity in relation to colonization and migration. *Nature Ecology & Evolution*, *2*, 841. https://doi.org/10.1038/s41559-018-0509-3

O'Connor, E. A., Strandh, M., Hasselquist, D., Nilsson, J.-Å., & Westerdahl, H. (2016). The evolution of highly variable immunity genes across a passerine bird radiation. *Molecular Ecology*, *25*, 977–989. https://doi.org/10.1111/mec.13530

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 of London. Series B: Biological Sciences*, *270*(Suppl), S254–S256. https://doi.org/10.1098/rsbl.2003.0079

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 of London. Series B: Biological Sciences*, *264*, 1813–1819. https://doi.org/10.1098/rspb.1997.0250

Penn, D. J., & Potts, W. K. (1999). The evolution of mating preferences and major histocompatibility complex genes. *The American Naturalist*, *153*, 145–164. https://doi.org/10.1086/303166

Piertney, S., & Oliver, M. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, *96*, 7–21. https://doi.org/10.1038/sj.hdy.6800724

Promerová, M., Vinkler, M., Bryja, J., Poláková, R., Schnitzer, J., Munclinger, P., & Albrecht, T. (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. https://doi.org/10.1111/j.1600-048X.2010.05221.x

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rekdal, S. L., Anmarkrud, J. A., Johnsen, A., & Lifjeld, J. T. (2018). Genotyping strategy matters when analyzing hypervariable major histocompatibility complex-Experience from a passerine bird. *Ecology and Evolution*, *8*, 1680–1692. https://doi.org/10.1002/ece3.3757

Rekdal, S. L., Anmarkrud, J. A., Lifjeld, J. T., & Johnsen, A. (2019). Data from: Extra-pair mating in a passerine bird with highly duplicated major histocompatibility complex class II: Preference for the golden mean. Dryad Digital Repository, https://doi.org/10.5061/dryad.93tf68k

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. https://doi.org/10.1038/35104547

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. https://doi.org/10.1098/rspb.2004.3028

Sandberg, M., Eriksson, L., Jonsson, J., Sjöström, M., & Wold, S. (1998). New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *Journal of Medicinal Chemistry*, *41*, 2481–2491. https://doi.org/10.1021/jm9700575

Schwensow, N., Fietz, J., Dausmann, K., & Sommer, S. (2008). MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate. *Evolutionary Ecology*, *22*, 617–636. https://doi.org/10.1007/s10682-007-9186-4

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. https://doi.org/10.1186/1471-2148-12-68

Sepil, I., Radersma, R., Santure, A. W., De Cauwer, I., Slate, J., & Sheldon, B. C. (2015). No evidence for MHC class I-based disassortative mating in a wild population of great tits. *Journal of Evolutionary Biology*, *28*, 642–654. https://doi.org/10.1111/jeb.12600

Sette, A., Buus, S., Appella, E., Smith, J. A., Chesnut, R., Miles, C., … Grey, H. M. (1989). Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proceedings of the National Academy of Sciences of the USA*, *86*, 3296–3300. https://doi.org/10.1073/pnas.86.9.3296

Sette, A., & Sidney, J. (1998). HLA supertypes and supermotifs: A functional perspective on HLA polymorphism. *Current Opinion in Immunology*, *10*, 478–482. https://doi.org/10.1016/S0952-7915(98)80124-6

Slade, J., Watson, M., Kelly, T., Gloor, G., Bernards, M., & MacDougall-Shackleton, E. (2016). Chemical composition of preen wax reflects major histocompatibility complex similarity in songbirds. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *283*, 20161966. https://doi.org/10.1098/rspb.2016.1966

Slade, J., Watson, M., & MacDougall-Shackleton, E. (2017). Birdsong signals individual diversity at the major histocompatibility complex. *Biology Letters*, *13*, 20170430. https://doi.org/10.1098/rsbl.2017.0430

Slade, R. W., & McCallum, H. I. (1992). Overdominant vs. frequency-dependent selection at MHC loci. *Genetics*, *132*, 861.

Snell, G. (1968). The H-2 locus of the mouse: Observations and speculations concerning its comparative genetics and its polymorphism. *Folia Biologica*, *14*, 335–358.

Sommer, S., Courtiol, A., & Mazzoni, C. J. (2013). MHC genotyping of non-model organisms using next-generation sequencing: A new methodology to deal with artefacts and allelic dropout. *BMC Genomics*, *14*, 542. https://doi.org/10.1186/1471-2164-14-542

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. https://doi.org/10.1098/rspb.2009.2084

Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T. F., … Jakobsen, K. S. (2011). The genome sequence of Atlantic cod reveals a unique immune system. *Nature*, *477*, 207. https://doi.org/10.1038/nature10342

Strandh, M., Westerdahl, H., Pontarp, M., Canbäck, B., Dubois, M.-P., Miquel, C., … Bonadonna, F. (2012). Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 4457–4463. https://doi.org/10.1098/rspb.2012.1562

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

Tong, J. C., Bramson, J., Kanduc, D., Chow, S., Sinha, A. A., & Ranganathan, S. (2006). Modeling the bound conformation of Pemphigus Vulgaris-associated peptides to MHC class II DR and DQ alleles. *Immunome Research*, *2*(1). https://doi.org/10.1186/1745-7580-2-1

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T. B., Hraber, P. T., Hayes, E., … Wolinsky, S. (2003). Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, *9*, 928. https://doi.org/10.1038/nm893

Trivers, R. (1972). Parental investment and sexual selection. In B. G. Campbell (Ed.), *Sexual selection and the descent of man* (1st ed., pp. 1871–1971). Chicago, IL: Aldine Publishing Company.

Vidović, D., & Matzinger, P. (1988). Unresponsiveness to a foreign antigen can be caused by self-tolerance. *Nature*, *336*, 222. https://doi.org/10.1038/336222a0

Wedekind, C., Seebeck, T., Bettens, F., & Paepke, A. J. (1995). MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *260*, 245–249. https://doi.org/10.1098/rspb.1995.0087

Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B., & Milinski, M. (2003). Parasite selection for immunogenetic optimality. *Science*, *301*, 1343. https://doi.org/10.1126/science.1088293

Westerdahl, H. (2004). No evidence of an MHC-based female mating preference in great reed warblers. *Molecular Ecology*, *13*, 2465–2470. https://doi.org/10.1111/j.1365-294X.2004.02238.x

Westerdahl, H. (2007). Passerine MHC: Genetic variation and disease resistance in the wild. *Journal of Ornithology*, *148*, 469–477. https://doi.org/10.1007/s10336-007-0230-5

Westerdahl, H., Wittzell, H., & von Schantz, T. (2000). Mhc diversity in two passerine birds: No evidence for a minimal essential Mhc. *Immunogenetics*, *52*, 92–100. https://doi.org/10.1007/s002510000256

Whittingham, L. A., Freeman-Gallant, C. R., Taff, C. C., & Dunn, P. O. (2015). Different ornaments signal male health and MHC variation in two populations of a warbler. *Molecular Ecology*, *24*, 1584–1595. https://doi.org/10.1111/mec.13130

Winternitz, J., Abbate, J., Huchard, E., Havlíček, J., & Garamszegi, L. (2017). Patterns of MHC-dependent mate selection in humans and nonhuman primates: A meta-analysis. *Molecular Ecology*, *26*, 668–688. https://doi.org/10.1111/mec.13920

Winternitz, J. C., Promerova, M., Polakova, R., Vinker, M., Schnitzer, J., Munclinger, P., … Albrecht, T. (2015). Effects of heterozygosity and MHC diversity on patterns of extra-pair paternity in the socially monogamous scarlet rosefinch. *Behavioral Ecology and Sociobiology*, *69*, 459–469. https://doi.org/10.1007/s00265-014-1858-9

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. https://doi.org/10.1098/rstb.2008.0174

Yamazaki, K., Boyse, E., Mike, V., Thaler, H., Mathieson, B., Abbott, J., … Thomas, L. (1976). Control of mating preferences in mice by genes in the major histocompatibility complex. *The Journal of Experimental Medicine*, *144*, 1324–1335. https://doi.org/10.1084/jem.144.5.1324

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, Z., & Swanson, W. J. (2002). Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Molecular Biology and Evolution*, *19*, 49–57. https://doi.org/10.1093/oxfordjournals.molbev.a003981

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Rekdal SL, Anmarkrud JA, Lifjeld JT, Johnsen A. Extra-pair mating in a passerine bird with highly duplicated major histocompatibility complex class II: Preference for the golden mean. *Mol Ecol*. 2019;00:1–12. https://doi.org/10.1111/mec.15273

MOLECULAR ECOLOGY

Appendices for:

Extra-pair mating in a passerine bird with highly duplicated major histocompatibility complex class II: Preference for the golden mean

Silje L. Rekdal, Jarl Andreas Anmarkrud, Jan T. Lifjeld, Arild Johnsen

Table of Contents:

S1: Allele calling of MHCIIβe2

The allele calling of bluethroat MHCIIβe2 are closely following Rekdal, Anmarkrud, Johnsen, and Lifjeld (2018), which on their part is based on the methodology published by Sommer, Courtiol, and Mazzoni (2013). The 278 samples were sequenced in duplicates on Illumina MiSeq together with 20 amplicons not included in this study. Individual barcodes and primer set up are given in Appendix S6. After sequencing, the raw paired MiSeq reads were merged in FLASH v1.2.11 (Magoč & Salzberg, 2011). Fastx toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/) was used to filter out reads of poor quality, i.e. reads where >5% of the nucleotides had a quality of <20 on the Phred scale. The heterogenic spacer motifs were subsequently removed by replacing all complete barcode-spacer-primer sequences with the corresponding barcode-primer sequences with UNIX commands. Next, these reads were clustered into variants and assigned to amplicons based on the barcodes using jMHC v1.6.1624 (Stuglik, Radwan, & Babik, 2011). Only variants that had >2 reads over all amplicons were outputted from the software. The jMHC results were imported into an Excel spreadsheet, in which we discarded variants with <3 reads in any amplicon. Only amplicons with >500 reads in total after this step were considered further for genotyping purposes. Next, all variants with an intra-amplicon frequency of >0.2% were marked as putative alleles for the respective amplicons, based on the rationale from Rekdal et al. (2018). All unique putative alleles were then aligned to previously published sequences (GenBank accession number HQ539575-HQ539614; Gohli et al., 2013) in MEGA7 (Kumar, Stecher, & Tamura, 2016), using the ClustalW algorithm (Thompson, Higgins, & Gibson, 1994). The variants were trimmed to the length of the HQ-sequences (267±3 bp), and variants with shift in reading-frame, stop-codons or that were lacking the conserved residues Cys10 and Cys74 (Gohli et al., 2013) were discarded. The variants now identical due to trimming were collapsed using fastx_collapser in the fastx toolkit v0.0.14, and their read depths were combined.

For each amplicon, the most frequent, remaining variant was called as allele, based on the assumption that true alleles will amplify to a larger extent than artefacts (Babik, Taberlet, Ejsmond, & Radwan, 2009; Lighten, Oosterhout, & Bentzen, 2014). Chimeras were detected within each amplicon by the uchime_denovo command in USEARCH v7.0.1090, in the software UCHIME (Edgar, Haas, Clemente, Quince, & Knight, 2011). The distance in terms of number of different nucleotides to the most similar, more frequent variant was calculated for every putative allele in every amplicon, by MEGA7. The non-chimeric putative alleles were hence given the status ">1bp" or "=1bp" based on these results.

A chimeric variant was considered an artefact for the respective individual if it was designated as a chimera also in the replicate, or if it was not present as a non-chimeric variant in the replicate. Likewise, the "=1bp"-variants were called as artefacts if not present in the replicate, while the ">1bp"-variants were considered an artefact if not present in neither the replicate nor any amplicons of the genetic parents or offspring.

After identifying artefacts within each family, the putative alleles were called as alleles if not scored as artefacts as stated above (i.e., a putative allele was scored as an allele if present in the replicate. ">1bp"-variants were however also called as alleles if present in any amplicons of the genetic parents or offspring, although not present in the replicate). For the individuals with only one successful amplicon (total number of reads >500), the putative alleles were called as alleles if also called as allele in any of the genetic parents or offspring. Additionally, the variants that were called as alleles for offspring but not for any of their parents were further processed manually. If these variants were scored as chimeras or were present with reads although not above the cut-off threshold in any parent, the alleles were however also scored for them. Of the 1886 nucleotide variants initially aligned in MEGA, 1176 unique variants survived further filtering and were considered an allele in at least one individual.

The MiSeq sequencing yielded 11 655 044 paired reads, and of these, 81.9% were retained after merging and quality filtering. On average, 10 870 reads were assigned to each amplicon after removing variants that had <3 reads in any amplicon (range 80 - 29 830 reads per amplicon). Thirty of the 556 amplicons had <500 reads in total assigned to them, and were hence disregarded. Both amplicons were disregarded for six individuals; one female, three extra-pair offspring and two within-pair offspring. We did not retrieve any DNA from one female sample, leaving 36 complete trios with the female and her social and extra-pair male, as well as 38 duos with social and extra-pair males, for the analyses.

A possible bias in the number of alleles called arising from the number of offspring was explored using the males that were solely social or extra-pair partner in the dataset (linear model: `lm(NoOffspring ~ NucleotideAlleles*Status)`). A two-way analysis of variance (ANOVA) showed no correlations between the number of alleles called and the number of offspring, nor for the interaction with status (p>>0.05, data not shown).

After the allele calling, a phylogenetic maximum likelihood tree was constructed on the nucleotide sequences using MEGA7, and visualized with iTOL (Letunic & Bork, 2016; see Appendix S3). The same program was used to translate the sequences by the standard genetic code.

S2: Details on establishment of MHCII PSS-alleles and supertypes

The software CodeML in the PAML package (Yang, 2007) was used to identify sites under positive selection (positively selected sites; PSS). As suggested by Wong, Yang, Goldman, and Nielsen (2004), the program was run multiple times; each with a different initial value of the parameters ω (dN/dS: 0.4, 1.5 and 3) and κ (transition/transversion rate: 2, 1.5 and 1). We initially restricted the analysis to the ten nucleotide sequences that were scored as alleles in the highest number of (presumably unrelated) adults. For a more comprehensive analysis, the 96 most common alleles were used as input and similarly run three times. This cut-off was set to include the number of alleles closest to 100 that were scored as alleles in a given number of individual adults or more (the 96 most common alleles were scored in at least seven adults, the 112 most common alleles were scored in at least six adults). Due to technical limitations we could not use all 1176 nucleotide sequences as input. Because so few positively selected sites were outputted from the runs with the 96 most common alleles (i.e. three PSS; see table S1), we additionally divided this input file into four, to avoid missing out on other PSS with a weaker signal.

The model that allowed positive selection (M8), fitted the data significantly better than the model without selection (M7; see table S1). The *P*-values were given by twice the log likelihood difference between M7 and M8 compared against a χ^2 distribution with two degrees of freedom (Yang, 2007). The same positively selected sites (PSS) were obtained in the different runs with different initial values of ω and κ, using the same input file. To constitute the amino acid residues in the PSS sequences, we selected those that both were identified as PSS by at least one CodeML run and that have previously been described as antigen binding residues in the bluethroat (Gohli et al., 2013, based upon PBR from human MHC DR and DQ from Tong et al., 2006): 4, 6, 8, 23, 25, 52, 55 and 66. These residues are all found to be under positive selection in other passerines (Balakrishnan et al., 2010). Additionally, exploration of the nucleotide sequences in FUBAR (Murrell et al., 2013) confirmed that these residues were under positive selection (running the 1176 sequences in bathces of 400+400+379 sequences on www.datamonkey.org/fubar; data not shown). However, running all 1176 in FUBAR through HyPhy v2.5.0 (Kosakovsky Pond, Frost, & Muse, 2005) locally did not identify all the PSS used in this study as positively selected.

To group the PSS sequences with similar antigen binding properties, we divided the PSS sequences into so-called supertypes (Doytchinova & Flower, 2005; Sepil, Moghadam, Huchard, & Sheldon, 2012). The PSS sequences were first listed in a spreadsheet, and each residue was given five values, according to the *z*-descriptors of the respective amino acid in Sandberg, Eriksson, Jonsson, Sjöström, and Wold (1998): z_1: lipophilicity/hydrophilicity, z_2:steric bulk/polarizability, z_3: polarity, z_4 and z_5: electronic effects. The *z*-descriptors are themselves based on a PCA established by Sandberg et al. (1998), in which 26 physiochemical variables for each amino acid were incorporated. The PSS sequences, each now characterized by 40 consecutive values, were then subjected to *k*-means clustering (Doytchinova & Flower, 2005), using the R package adegenet (Jombart, 2008). This program uses discriminant analysis of principal components (DAPC) to infer clusters, pursuing to maximize among-cluster variation and minimize within-cluster variation (Jombart, Devillard, & Balloux, 2010). *K*-means clustering is run iteratively with an increasing number of clusters (*k*), and an associated likelihood is calculated given the data. The optimal number of supertypes derived in the dataset was determined by the lowest number of clusters in adegenet, after which the associated Bayesian information criterion-values (BIC) changed only marginally (Jombart et al., 2010). Based on the BIC-values, we inferred 20 supertypes among the 311 PSS sequences (see figure S1 and figure S2).

The individual genotypes on PSS level were obtained by collapsing the nucleotide alleles into the corresponding PSS alleles across all individuals using an Excel sheet. Similarly, the supertype genotypes were established by grouping the PSS alleles into the inferred supertypes across all individuals.

Table S1: Results from the CodeML runs, when identifying the positively selected sites (PSS) among the MHCIIβe2 nucleotide sequences in 88 bluethroat individuals. The maximum log likelihood (ML) of the two models M7 (no positive selection allowed on any sites) and M8 (positive selection allowed) are given. The *P*-value is twice the log likelihood difference between the two compared models (M7 and M8), compared against a χ^2 distribution with two degrees of freedom. PSS designates the inferred positively selected sites, given by their position in the translated sequence. Only the 96 most common alleles (full set and subdivided into four subsets: A, B, C and D in the table) and the ten most common alleles were included in the different runs. When using the four subsets of the 96 most common alleles and the ten most common alleles as input files, different initial values of ω and κ gave the same maximal log likelihoods, *P*-value and PSS. These numbers are hence just given for the 96-allele dataset. The identified PSS previously described as antigen binding residues in the bluethroat (Gohli et al., 2013) are marked in bold type.

	The 96 most common alleles			**Subsets of the 96 most common alleles**				**The 10 most common alleles**
				A	**B**	**C**	**D**	
	ω=0.4	**ω=1.5**	**ω=3**					
	κ=2	**κ=1.5**	**κ=1**					
M7: maximum log likelihood	-3616	-3616	-3615	-1488	-1649	-1778	-2134	-1127
M8: maximum log likelihood	-3581	-3582	-3579	-1479	-1638	-1765	-2112	-1118
***P*-value**	5.3 e-16	9.4 e-16	2.6 e-16	0.00013	1.7 e-05	3.5 e-06	3.0 e-10	0.00013
PSS	**6**	**6**	**6**	**4**	**4**	**6**	**4**	**23**
	66	**66**	**66**	**23**	**6**		**6**	**25**
	68	68	68	**52**	**23**		**8**	47
				55	**52**		48	**52**
				66	**66**		**66**	**55**
				68	80			**66**
				80				

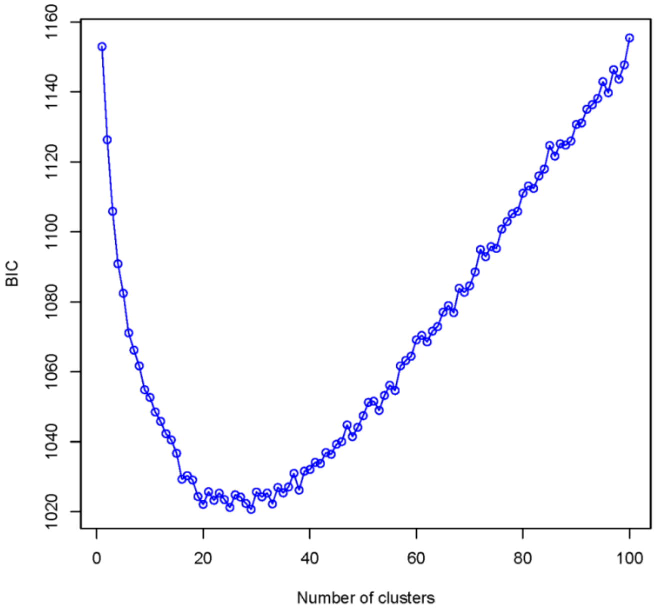


Figure S1: The Bayesian information criterion (BIC) obtained for each *k* (number of supertypes, *i.e.* clusters) when the 311 unique PSS (positively selected sites) MHC class II sequences were run using the package adegenet (Jombart, 2008) in R.

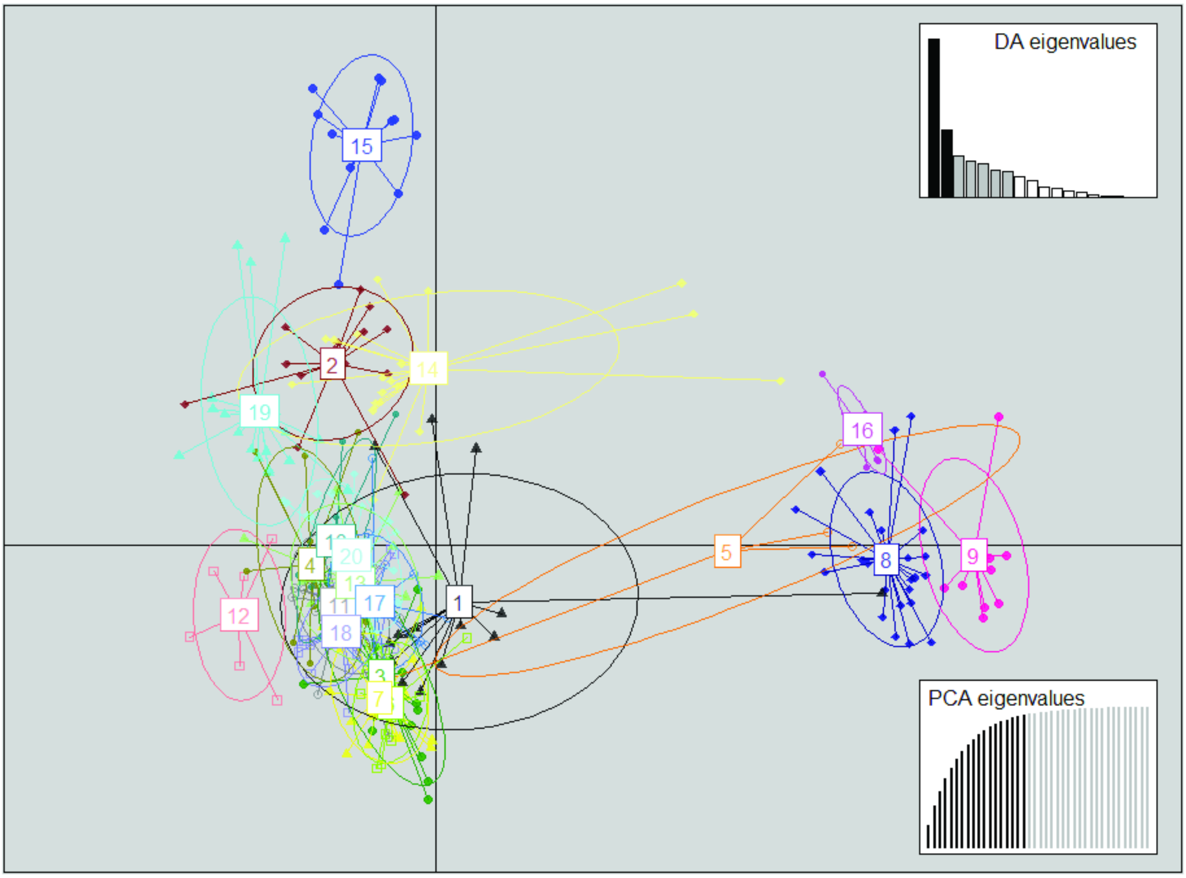


Figure S2: A scatterplot of the 311 PSS (positively selected sites) sequences (points) clustered into 20 supertypes (ellipses) and colored correspondingly. The plot is visualized on the two first principal components (PC), when analyzed with discriminant analysis of principal component (DAPC) in the R package adegenet (Jombart, 2008). The inserts represent the eigenvalues of the PCs in the discriminant analysis (top right) and the PCA used to transform the data to uncorrelated variables (bottom right). When visualizing the relationships among the supertypes, the 18 first PCs (from the PCA used in the transformation of the data into uncorrelated variables: black bars in bottom right inserted figure), as well as seven discriminant functions (DA: black and grey bars in the top right inserted figure), were retained in the analysis, which conserved 94.4% of the variation.

S3: Phylogeny of the MHCII PSS-sequences

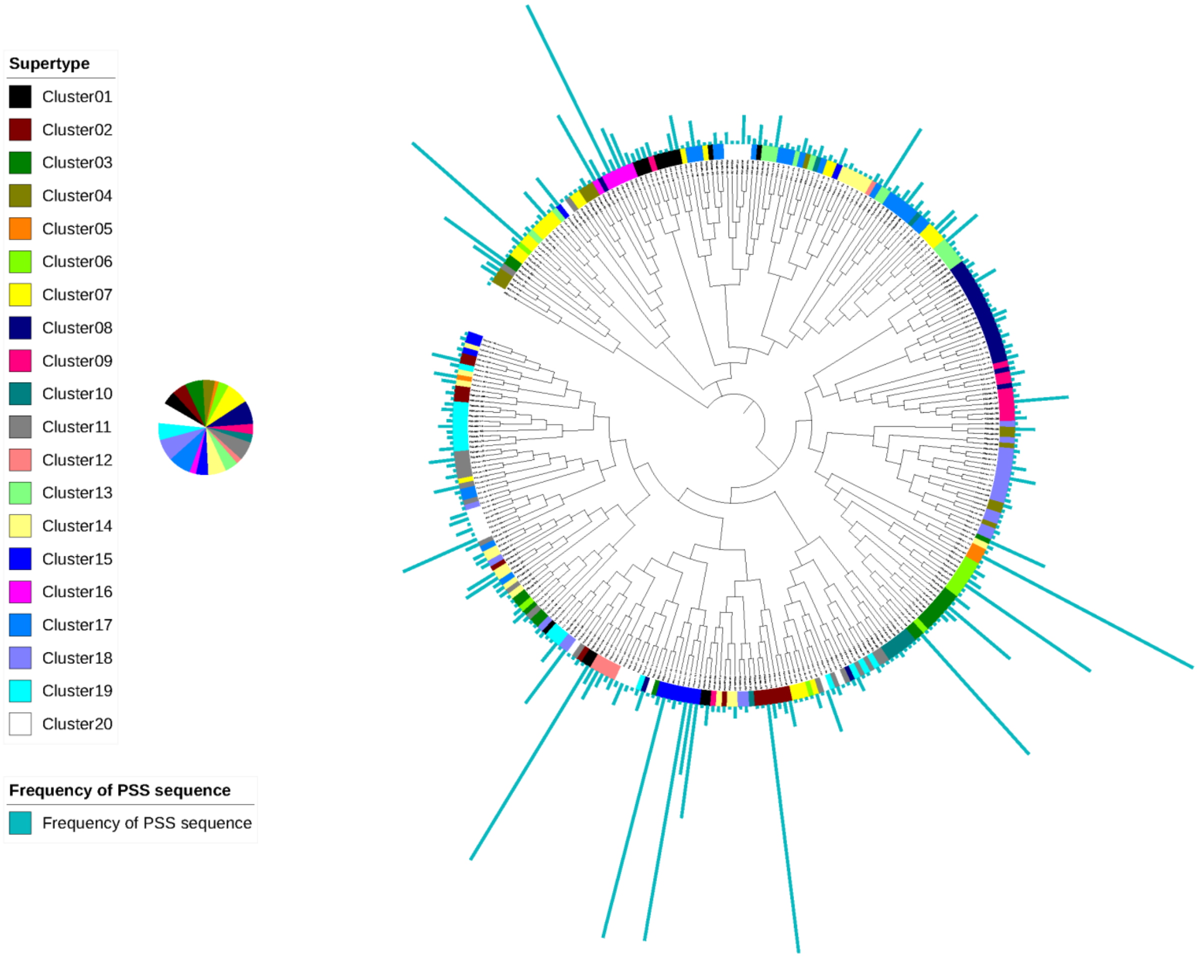


Figure S3: A phylogeny of the 311 PSS (positively selected sites) MCH class II sequences, generated with iTOL (Letunic & Bork, 2016). The colors correspond to the supertype to which each PSS sequence belongs, while the frequency bars indicate the number of (presumably unrelated) adults the PSS sequence is genotyped in. The pie chart illustrates the number of PSS sequences each supertype is consisting of.

S4: Results from paired t-tests, maximized diversity (adults)

Table S2: Results obtained from paired t-test, when testing for female choice of extra-pair males (EPM) for maximized MHC class II diversity across different parameters. If females choose EPM based on maximized diversity in the male genotypes alone, we would expect a significant difference between the social male (WPM) and EPM for the male diversity parameters. If instead females are able to self-reference and choose EPM based on a maximized diversity within the pair, we would expect a difference in the parameters between the social pairs (WPM-F) and extra-pair partners (EPM-F). The test units (WPM and EPM, and WPM-F and EPM-F) were paired within nest. The *P*-values are two-tailed. Parameter explanations: **MPSS**: The number of unique PSS alleles, found within the individual males, **MPSSSum**: Sum of the amino acid distance between all pairs of unique PSS alleles, found within the individual males, **MPSSAve**: Average amino acid distance between all pairs of unique PSS alleles, found within the individual males, **MST**: The number of supertypes, found within the individual males, $\mathbf{PSS_{Non\text{-}sharedMale}}$: The number of non-shared PSS alleles within the pairs (only found within the male), $\mathbf{PSSSum_{Non\text{-}sharedMale}}$: Sum of the amino acid distance between all pairs of non-shared PSS alleles within the pair (only found within the male), $\mathbf{PSSAve_{Non\text{-}sharedMale}}$: Average amino acid distance between all pairs of non-shared PSS alleles within the pair (only found within the male), $\mathbf{ST_{Non\text{-}sharedMale}}$: The number of non-shared supertypes within the pair (only found within the male).

Test unit	**Parameter**	***t***	**df**	**95% confidence interval**	***P*-value**
Male diversity (WPM and EPM)					
	MPSS	0.94	37	-0.95 - 2.58	0.36
	MPSSSum	1.15	37	-92.1 - 333	0.26
	MPSSAve	-0.057	37	-0.083 - 0.079	0.96
	MST	0.86	37	-0.57 - 1.41	0.39
Compatibility (WPM-F and EPM-F)					
	$PSS_{Non\text{-}sharedMale}$	1.07	35	-0.78 - 2.50	0.29
	$PSSSum_{Non\text{-}sharedMale}$	1.41	35	-35.1 - 194	0.17
	$PSSAve_{Non\text{-}sharedMale}$	-0.61	35	-0.18 - 0.098	0.55
	$ST_{Non\text{-}sharedMale}$	-0.55	35	-0.79 - 0.45	0.59

References

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. doi: 10.1111/j.1755-0998.2009.02622.x

Doytchinova, I. A., & Flower, D. R. (2005). In silico identification of supertypes for class II MHCs. *The Journal of Immunology, 174*, 7085-7095. doi: 10.4049/jimmunol.174.11.7085

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics, 27*, 2194-2200. doi: 10.1093/bioinformatics/btr381

Gohli, J., Anmarkrud, J. A., Johnsen, A., Kleven, O., Borge, T., & Lifjeld, J. T. (2013). Female promiscuity is positively associated with neutral and selected genetic diversity in passerine birds. *Evolution, 67*, 1406-1419. doi: 10.1111/evo.12045

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics, 24*, 1403-1405. doi: 10.1093/bioinformatics/btn129

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. doi: 10.1186/1471-2156-11-94

Kosakovsky Pond, S. L., Frost, S., & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics, 21*, 676-679. doi: 10.1093/bioinformatics/bti079

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution, 33*, 1870-1874. doi: 10.1093/molbev/msw054

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research, 44*, W242-W245. doi: 10.1093/nar/gkw290

Lighten, J., Oosterhout, C., & Bentzen, P. (2014). Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology, 23*, 3957-3972. doi: 10.1111/mec.12843

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics, 27*, 2957-2963. doi: 10.1093/bioinformatics/btr507

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution, 30*, 1196-1205. doi: 10.1093/molbev/mst030

Rekdal, S. L., Anmarkrud, J. A., Johnsen, A., & Lifjeld, J. T. (2018). Genotyping strategy matters when analyzing hypervariable major histocompatibility complex ‐ Experience from a passerine bird. *Ecology and Evolution, 8*, 1680-1692. doi: 10.1002/ece3.3757

Sandberg, M., Eriksson, L., Jonsson, J., Sjöström, M., & Wold, S. (1998). New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *Journal of Medicinal Chemistry, 41*, 2481-2491. doi: 10.1021/jm9700575

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. doi: 10.1186/1471-2148-12-68

Sommer, S., Courtiol, A., & Mazzoni, C. J. (2013). MHC genotyping of non-model organisms using next-generation sequencing: A new methodology to deal with artefacts and allelic dropout. *BMC Genomics, 14*, 542. doi: 10.1186/1471-2164-14-542

Stuglik, M. T., Radwan, J., & Babik, W. (2011). jMHC: Software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Molecular Ecology Resources, 11*, 739-742. doi: 10.1111/j.1755-0998.2011.02997.x

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research, 22*, 4673-4680. doi: 10.1093/nar/22.22.4673

Wong, W. S., Yang, Z., Goldman, N., & Nielsen, R. (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics, 168*, 1041-1051. doi: 10.1534/genetics.104.031153

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution, 24*, 1586-1591. doi: 10.1093/molbev/msm088

Supplemental information for the article: *Extra-pair mating in a passerine bird with highly duplicated major histocompatibility complex class II: Preference for the golden mean*

Silje L. Rekdal [1], Jarl Andreas Anmarkrud [1], Jan T. Lifjeld [1], Arild Johnsen [1]

[1] Natural History Museum, University of Oslo, Oslo, Norway

Table S3-IV: Individual genotypes of nucleotide alleles, alleles based on positively selected sites (PSS alleles) and supertypes. Each individual is only listed once. Note that individuals 2259, 2513, 1503 and 2914 aresampled in both sampling years. The number of PSS alleles within the observed pairs (social (WPM-F) and extra-pair (EPM-F) partners) are listed to the right of the table. Here, if the nest is cuckolded by more than one male, only the male with the highest number of offspring is included (valid for four nests: 1998_19, 1999_19, 1999_20, 1999_57).

Individual (accession number)	Role	Role (abbreviated)	Nest	#Nucleotide alleles Individuals	#PSS alleles Individuals	#Supertypes Individuals	Mulitple times in dataset?
1337	M	M	Multiple	35	13	9	Yes
1592	M	M	Multiple	16	13	9	Yes
2602	M	M	Multiple	36	14	9	Yes
2904	M	M	Multiple	40	19	11	Yes
3083	M	M	Multiple	34	17	12	Yes
3171	M	M	Multiple	39	22	14	Yes
3428	M	M	Multiple	33	20	12	yes
3882	M	M	Multiple	40	27	15	Yes
4899	M	M	Multiple	29	18	9	Yes
5019	M	M	Multiple	25	14	10	Yes
5508	M	M	Multiple	48	17	11	Yes
6701	M	M	Multiple	44	23	18	Yes
6779	M	M	Multiple	30	18	12	Yes
6890	M	M	Multiple	28	19	13	Yes
7412	M	M	Multiple	30	19	13	Yes
8300	M	M	Multiple	33	15	9	Yes
1503	M	M	Multiple	58	23	15	Yes (both years)
2259	M	M	Multiple	31	20	13	Yes (both years)
2513	M	M	Multiple	45	21	13	Yes (both years)
2914	F	F	Multiple	39	21	14	Yes (both years)
1137	M	M	1998_1	28	21	13	
1148	F	F	1998_1	39	21	14	
1159	Pullus 1	WPY	1998_1	38	19	12	
1170	Pullus 2	EPY	1998_1	27	15	11	
1181	Pullus 3	WPY	1998_1	27	19	12	
1192	Pullus 4	EPY	1998_1	37	16	10	
1203	Pullus 5	EPY	1998_1	35	16	13	
1214	Pullus 6	WPY	1998_1	29	22	15	
1248	M	M	1998_3	36	22	13	
1259	F	F	1998_3	51	30	16	
1270	Pullus 1	WPY	1998_3	40	27	17	
1281	Pullus 2	EPY	1998_3	43	18	10	
1292	Pullus 3	WPY	1998_3	41	19	11	
1303	Pullus 4	WPY	1998_3	45	22	11	
1314	Pullus 5	WPY	1998_3	46	30	18	
1348	F	F	1998_5	39	28	14	
1359	Pullus 1	EPY	1998_5	23	17	12	
1370	Pullus 2	EPY	1998_5	19	15	10	
1381	Pullus 3	WPY	1998_5	29	14	11	
1392	Pullus 4	EPY	1998_5	33	24	13	
1403	Pullus 5	WPY	1998_5	30	15	12	
1514	F	F	1998_7	40	20	13	
1525	Pullus 1	WPY	1998_7	37	16	13	
1536	Pullus 2	EPY	1998_7	26	15	12	
1547	Pullus 3	EPY	1998_7	25	14	11	
1559	Pullus 4	EPY	1998_7	24	18	10	
1570	Pullus 5	EPY	1998_7	28	16	9	
1581	Pullus 6	WPY	1998_7	38	18	13	
1768	M	M	1998_10	24	11	10	
1780	F	F	1998_10	38	15	10	
1790	Pullus 1	EPY	1998_10	37	14	11	
1801	Pullus 2	EPY	1998_10	38	17	12	
1812	Pullus 3	WPY	1998_10	28	14	10	
1823	M	M	1998_11	38	21	16	
2169	M	M	1998_15	40	26	15	
2180	F	F	1998_15	39	26	16	
2191	Pullus 1	WPY	1998_15	44	31	16	
2202	Pullus 2	EPY	1998_15	39	24	16	
2215	Pullus 3	WPY	1998_15	29	15	9	
2226	Pullus 4	WPY	1998_15	30	19	12	
2237	Pullus 5	WPY	1998_15	27	13	8	
2248	Pullus 6	WPY	1998_15	45	32	17	
2358	M	M	1998_17	48	31	17	
2369	F	F	1998_17	35	19	12	
2380	Pullus 1	WPY	1998_17	43	26	15	
2391	Pullus 2	WPY	1998_17	41	27	15	
2402	Pullus 3	EPY	1998_17	29	18	15	
2413	Pullus 4	EPY	1998_17	32	17	11	
2424	Pullus 5	WPY	1998_17	42	27	15	
2524	F	F	1998_19	45	28	16	
2535	Pullus 1	EPY	1998_19	40	20	10	
2547	Pullus 2	EPY	1998_19	38	21	13	
2558	Pullus 3	EPY	1998_19	35	16	10	
2569	Pullus 4	EPY	1998_19	35	18	12	
2580	Pullus 5	WPY	1998_19	41	24	13	
2613	F	F	1998_20	34	21	15	
2619	Pullus 1	EPY	1998_20	35	20	13	
2629	Pullus 2	WPY	1998_20	34	13	8	

Nest	#PSS-alleles WPM-F	#PSS-alleles EPM_F
1998_1	35	31
1998_3	43	34
1998_5	35	35
1998_7	31	28
1998_10	21	29
1998_15	43	37
1998_17	40	32
1998_19	40	33
1998_20	28	31
1998_24	31	31
1998_27	25	35
1998_28	36	34
1998_31	38	31
1998_32	27	26
1998_34	31	35
1998_36	33	32
1998_41	18	26
1998_42	39	33
1998_47	31	28
1998_56	32	26
1998_62	24	28
1998_64	35	31
1998_69	22	26
1998_70	27	33
1999_2	34	27
1999_12	32	34
1999_17	38	34
1999_19	43	41
1999_20	30	30
1999_31	32	34
1999_35	40	37
1999_44	26	26
1999_46	28	31
1998_16	38	34
1998_19	33	32
1999_57	24	27

Individual (accession number)	Role	Role (abbreviated)	Nest	#Nucleotide alleles Individuals	#PSS alleles Individuals	#Supertypes Individuals	Mulitple times in dataset?
2640	Pullus 3	WPY	1998_20	31	11	7	
2652	Pullus 4	WPY	1998_20	32	15	13	
2663	M	M	1998_21	51	26	17	
2751	M	M	1998_22	33	19	11	
2924	Pullus 1	WPY	1998_24	30	20	13	
2935	Pullus 2	WPY	1998_24	42	16	10	
2946	Pullus 3	WPY	1998_24	27	20	13	
2957	Pullus 4	WPY	1998_24	41	13	8	
2968	Pullus 5	WPY	1998_24	42	17	10	
2980	Pullus 6	EPY	1998_24	35	20	14	
3094	F	F	1998_27	36	17	11	
3105	Pullus 1	EPY	1998_27	49	27	16	
3116	Pullus 2	EPY	1998_27	28	14	11	
3127	Pullus 3	WPY	1998_27	36	15	10	
3138	Pullus 4	EPY	1998_27	28	11	9	
3149	Pullus 5	EPY	1998_27	29	13	8	
3160	Pullus 6	EPY	1998_27	27	12	7	
3182	F	F	1998_28	36	25	13	
3194	Pullus 1	WPY	1998_28	30	21	12	
3205	Pullus 2	EPY	1998_28	23	18	12	
3216	Pullus 3	WPY	1998_28	33	21	12	
3227	Pullus 4	EPY	1998_28	26	20	13	
3238	Pullus 5	WPY	1998_28	31	20	12	
3271	M	M	1998_31	48	26	13	
3282	F	F	1998_31	36	19	11	
3293	Pullus 1	EPY	1998_31	27	12	11	
3306	Pullus 2	WPY	1998_31	46	26	13	
3317	Pullus 3	EPY	1998_31	27	12	8	
3328	Pullus 4	EPY	1998_31	27	15	12	
3350	M	M	1998_32	37	20	11	
3361	F	F	1998_32	31	14	10	
3372	Pullus 1	WPY	1998_32	22	18	10	
3383	Pullus 2	EPY	1998_32	40	19	13	
3439	M	M	1998_34	39	15	9	
3450	F	F	1998_34	45	24	15	
3461	Pullus 1	WPY	1998_34	39	17	11	
3472	Pullus 2	WPY	1998_34	21	13	10	
3483	Pullus 3	WPY	1998_34	36	16	11	
3494	Pullus 4	WPY	1998_34	35	21	14	
3505	Pullus 5	EPY	1998_34	27	19	12	
3516	Pullus 6	WPY	1998_34	44	20	13	
3536	M	M	1998_36	38	19	12	
3547	F	F	1998_36	40	22	14	
3558	Pullus 1	WPY	1998_36	38	25	12	
3569	Pullus 2	WPY	1998_36	32	16	12	
3580	Pullus 3	EPY	1998_36	44	24	15	
3590	Pullus 4	WPY	1998_36	30	19	12	
3612	M	M	1998_37	52	21	14	
3793	M	M	1998_41	33	12	8	
3804	F	F	1998_41	33	13	8	
3815	Pullus 1	WPY	1998_41	35	11	6	
3826	Pullus 2	WPY	1998_41	34	8	5	
3837	Pullus 3	WPY	1998_41	34	8	5	
3849	Pullus 4	WPY	1998_41	35	12	8	
3860	Pullus 5	EPY	1998_41	28	16	9	
3871	Pullus 6	EPY	1998_41	26	12	9	
3893	F	F	1998_42	46	21	14	
3904	Pullus 1	WPY	1998_42	41	22	13	
3915	Pullus 2	EPY	1998_42	43	18	10	
3926	M	M	1998_43	41	20	13	
3947	Pullus 1	WPY	1998_43	25	16	11	
3959	Pullus 2	WPY	1998_43	32	18	12	
3970	Pullus 3	WPY	1998_43	28	14	10	
3980	Pullus 4	EPY	1998_43	38	20	13	
3991	Pullus 5	EPY	1998_43	30	21	15	
4002	Pullus 6	WPY	1998_43	37	20	12	
4113	M	M	1998_46	38	13	9	
4202	M	M	1998_47	44	16	11	
4213	F	F	1998_47	28	22	13	
4224	Pullus 1	EPY	1998_47	26	16	12	
4235	Pullus 2	EPY	1998_47	26	20	11	
4246	Pullus 3	WPY	1998_47	40	22	12	
4257	Pullus 4	EPY	1998_47	26	20	11	
4268	Pullus 5	WPY	1998_47	45	23	12	
4555	M	M	1998_54	51	24	13	
4710	M	M	1998_56	34	22	14	
4721	F	F	1998_56	33	16	9	
4733	Pullus 1	WPY	1998_56	22	10	8	
4744	Pullus 2	EPY	1998_56	27	14	10	
4755	Pullus 3	EPY	1998_56	30	15	9	
4766	Pullus 4	EPY	1998_56	25	11	8	
4777	Pullus 5	WPY	1998_56	31	11	9	
4788	Pullus 6	WPY	1998_56	41	23	13	
4799	Pullus 7	WPY	1998_56	36	21	13	
4921	M	M	1998_60	37	20	15	
5030	F	F	1998_62	27	17	11	

Individual (accession number)	Role	Role (abbreviated)	Nest	#Nucleotide alleles Individuals	#PSS alleles Individuals	#Supertypes Individuals	Mulitple times in dataset?
5041	Pullus 1	WPY	1998_62	21	18	12	
5052	Pullus 2	WPY	1998_62	27	13	10	
5064	Pullus 3	WPY	1998_62	34	14	9	
5075	Pullus 4	WPY	1998_62	23	17	12	
5175	F	F	1998_64	40	22	15	
5186	Pullus 1	WPY	1998_64	41	23	15	
5197	Pullus 2	WPY	1998_64	25	15	11	
5208	Pullus 3	WPY	1998_64	29	19	15	
5219	Pullus 4	EPY	1998_64	40	14	9	
5230	Pullus 5	WPY	1998_64	37	22	15	
5241	Pullus 6	WPY	1998_64	31	17	13	
5519	F	F	1998_69	26	12	8	
5530	Pullus 1	WPY	1998_69	34	11	8	
5540	Pullus 2	WPY	1998_69	29	7	5	
5551	Pullus 3	EPY	1998_69	29	17	11	
5562	Pullus 4	WPY	1998_69	40	9	7	
5573	Pullus 5	WPY	1998_69	34	18	11	
5584	Pullus 6	WPY	1998_69	46	15	10	
5593	M	M	1998_70	48	20	11	
5604	F	F	1998_70	33	18	11	
5616	Pullus 1	WPY	1998_70	29	15	11	
5627	Pullus 2	WPY	1998_70	33	19	12	
5638	Pullus 3	EPY	1998_70	43	27	14	
6090	M	M	1999_2	31	25	15	
6101	F	F	1999_2	50	16	13	
6112	Pullus 1	EPY	1999_2	38	18	13	
6123	Pullus 2	EPY	1999_2	39	8	7	
6134	Pullus 3	WPY	1999_2	31	15	11	
6145	Pullus 4	EPY	1999_2	40	21	14	
6156	Pullus 5	EPY	1999_2	37	10	9	
6168	Pullus 6	WPY	1999_2	35	15	13	
6201	M	M	1999_6	40	19	12	
6256	M	M	1999_7	43	19	13	
6345	M	M	1999_12	34	17	11	
6356	F	F	1999_12	26	22	13	
6367	Pullus 1	EPY	1999_12	32	18	13	
6378	Pullus 2	EPY	1999_12	31	18	13	
6390	Pullus 3	EPY	1999_12	34	20	12	
6401	Pullus 4	WPY	1999_12	25	17	11	
6412	Pullus 5	EPY	1999_12	28	14	10	
6423	Pullus 6	EPY	1999_12	25	19	11	
6534	M	M	1999_16	28	13	10	
6613	M	M	1999_17	35	18	13	
6624	F	F	1999_17	40	26	15	
6635	Pullus 1	WPY	1999_17	37	18	15	
6646	Pullus 2	WPY	1999_17	36	25	16	
6657	Pullus 3	WPY	1999_17	30	21	13	
6668	Pullus 4	EPY	1999_17	31	20	14	
6679	Pullus 5	EPY	1999_17	30	22	14	
6712	F	F	1999_19	43	32	15	
6724	Pullus 1	WPY	1999_19	40	28	17	
6735	Pullus 2	EPY	1999_19	25	20	13	
6746	Pullus 3	EPY	1999_19	32	23	15	
6757	Pullus 4	EPY	1999_19	29	18	13	
6768	Pullus 5	WPY	1999_19	52	26	15	
6790	F	F	1999_20	30	19	12	
6812	Pullus 2	EPY	1999_20	22	16	13	
6823	Pullus 3	EPY	1999_20	36	23	15	
7023	0	WPY	1999_20	27	14	10	
7279	M	M	1999_31	35	20	13	
7290	F	F	1999_31	57	19	14	
7301	Pullus 1	WPY	1999_31	36	20	13	
7312	Pullus 2	EPY	1999_31	39	23	15	
7323	Pullus 3	WPY	1999_31	44	21	13	
7334	Pullus 4	WPY	1999_31	41	15	12	
7345	Pullus 5	EPY	1999_31	46	22	14	
7356	Pullus 6	WPY	1999_31	40	20	13	
7523	M	M	1999_35	58	23	14	
7534	F	F	1999_35	42	24	12	
7545	Pullus 1	EPY	1999_35	33	19	12	
7556	Pullus 2	WPY	1999_35	43	22	15	
7567	Pullus 3	WPY	1999_35	39	22	12	
7577	Pullus 4	WPY	1999_35	48	19	14	
8000	M	M	1999_43	19	14	10	
8011	M	M	1999_44	35	18	11	
8022	F	F	1999_44	43	14	10	
8033	Pullus 1	EPY	1999_44	40	21	13	
8044	Pullus 2	WPY	1999_44	36	17	10	
8200	M	M	1999_46	22	13	8	
8211	F	F	1999_46	47	22	13	
8222	Pullus 1	EPY	1999_46	33	18	15	
8233	Pullus 2	EPY	1999_46	44	15	12	
8244	Pullus 3	EPY	1999_46	32	18	12	
8322	Pullus 1	WPY	1999_47	32	21	11	
8333	Pullus 2	EPY	1999_47	30	15	8	
8344	Pullus 3	WPY	1999_47	34	11	8	

Individual (accession number)	Role	Role (abbreviated)	Nest	#Nucleotide alleles Individuals	#PSS alleles Individuals	#Supertypes Individuals	Mulitple times in dataset?
8355	Pullus 4	EPY	1999_47	28	13	7	
8366	Pullus 5	EPY	1999_47	21	12	7	
8375	Pullus 6	EPY	1999_47	38	21	12	
8485	F	F	1999_49	44	25	14	
8508	Pullus 2	EPY	1999_49	40	20	14	
8519	Pullus 3	WPY	1999_49	31	19	13	
8552	Pullus 1	EPY	1999_50	36	17	11	
8563	Pullus 2	WPY	1999_50	41	18	12	
8574	Pullus 3	EPY	1999_50	30	19	14	
8585	Pullus 4	EPY	1999_50	39	23	15	
8596	Pullus 5	WPY	1999_50	51	22	15	
8608	Pullus 6	EPY	1999_50	37	20	15	
8897	M	M	1999_56	38	23	12	
8908	M	M	1999_57	40	18	14	
8919	F	F	1999_57	34	14	11	
8930	Pullus 1	EPY	1999_57	45	25	13	
8942	Pullus 2	EPY	1999_57	35	14	10	
8953	Pullus 3	EPY	1999_57	42	18	13	
8964	Pullus 4	WPY	1999_57	31	15	13	
8975	Pullus 5	EPY	1999_57	42	19	13	
8986	Pullus 6	WPY	1999_57	33	15	12	
8997	Pullus 7	EPY	1999_57	31	10	8	