LUND UNIVERSITY

MHC diversity and expression in a songbird

Temporal and transgenerational variation

Mellinger, Samantha

2023

Document Version:
Other version

Link to publication

Citation for published version (APA):
Mellinger, S. (2023). *MHC diversity and expression in a songbird: Temporal and transgenerational variation*. [Doctoral Thesis (compilation), Department of Biology, Faculty of Science]. Lund University.

Total number of authors:
1

Creative Commons License:
Unspecified

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

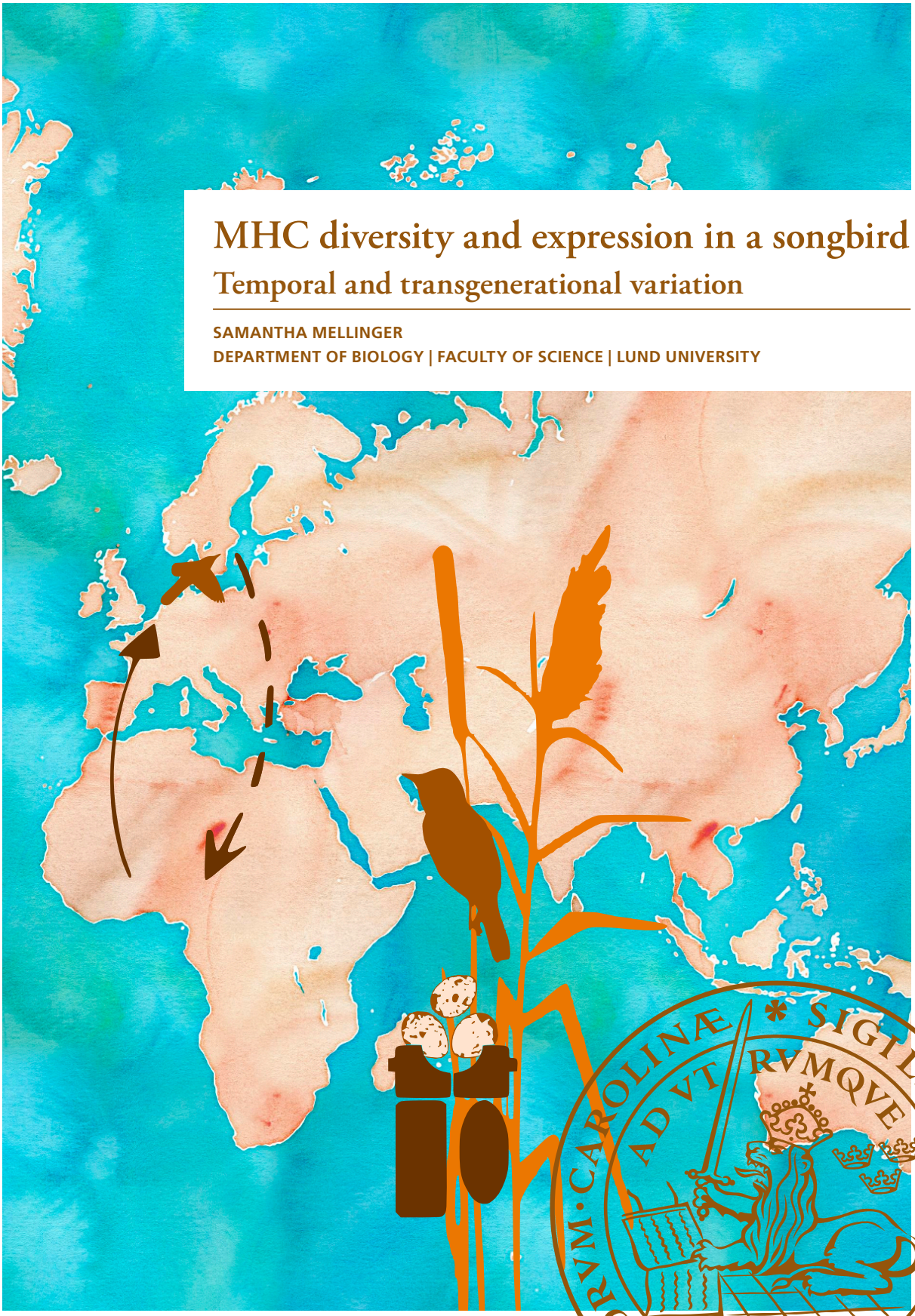
LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

MHC diversity and expression in a songbird

Temporal and transgenerational variation

SAMANTHA MELLINGER
DEPARTMENT OF BIOLOGY | FACULTY OF SCIENCE | LUND UNIVERSITY



MHC diversity and expression in a songbird

Temporal and transgenerational variation

Samantha Mellinger



LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Science at Lund University to be publicly defended on 29th of September at 09:30 in the Blue Hall, Ecology Building, Sölvegatan 37, Lund, Sweden.

Faculty opponent
Prof. Dr. Tobias L. Lenz
Evolutionary Immunogenomics group, Department of Biology
Universität Hamburg, Germany

Organization: LUND UNIVERSITY

Document name: DOCTORAL DISSERTATION **Date of issue:** 2023-09-29

Author(s): Samantha Mellinger **Sponsoring organization: NA**

Title and subtitle: MHC diversity and expression in a songbird: temporal and transgenerational variation.

Abstract: The immune system is of central importance to clear-off pathogens in order to keep an organism healthy. In vertebrates, the Major Histocompatibility Complex (MHC) is a genomic region that holds key genes involved in adaptive immunity and in particular in the antigen presenting process. The presentation relies on MHC class I (MHC-I) and MHC class II (MHC-II) molecules that are encoded by MHC-I and MHC-II genes, respectively. In passerine birds (songbirds; order Passeriformes), high MHC-I and MHC-IIB diversity (*i.e.,* number of MHC alleles in an individual) has been reported compared to other avian orders. Each MHC molecule can present a limited number of antigens, therefore high MHC diversity could potentially be an advantage for presenting various antigens from different pathogens. My aim was to gain a deeper understanding on what mechanisms have led to high MHC diversity in passerines, how much of this MHC diversity is expressed and whether it varies in time and across generation. In this thesis, I mainly focused on a particular passerine, the great reed warbler (GRW; *Acrocephalus arundinaceus*) for which high MHC diversity has been previously described. In chapter I, the MHC genomic region was characterized in the GRW and also in three additional passerine species from which long-read genome assemblies were available. The MHC region is extended in Passeriformes and the MHC diversity in the GRW has expanded through repeated and multiple gene duplication events of single and trios of MHC genes; these MHC paralogs are likely to have evolved under the birth-and-death model process. In chapter II, I wanted to assess the correctness of the estimated haploid MHC diversity found in the GRW genome and to compare the resolution between the estimated diploid MHC diversity found with long- and short-read sequencing techniques. The haplotype resolution was improved using short-read amplicons and I found that two MHC-IIB scaffolds holding tandemly duplicated genes were most likely complementary haplotypes. In the last two chapters, I wanted to investigate the functional relevance of the MHC-I diversity in the GRW by looking at gene expression in individuals from our long-term monitored study population breeding at Lake Kvismaren in Sweden. I found that 71% of the MHC-I diversity is expressed in the GRW but the different MHC-I alleles are expressed at different levels with generally only one to two alleles being highly expressed in each individual. Therefore, I investigated temporal variation in MHC-I relative expression (chapter III) in individuals sampled both as nestlings and as breeding adults at our study site. I found that MHC-I relative expression in blood is highly stable throughout an individual lifetime. In chapter IV, I investigated transgenerational variation in MHC-I relative expression in 17 families. MHC-I relative expression in haplotypes in offspring is to a large extent genetically inherited from their parents and does not change with offspring age. I found a stronger relationship between mothers and offspring suggesting a maternal effect, but I have not found any epistasis effect. To conclude, a high proportion of the MHC diversity is expressed in the great reed warbler but only a small fraction is highly expressed. Additionally, MHC-I expression does not vary over time and is to a large extent inherited.

Key words: MHC diversity, MHC-I gene expression, songbirds, genome assembly, amplicon HTS, temporal variation, family data, great reed warbler (*Acrocephalus arundinaceus*).

Classification system and/or index terms (if any) Supplementary bibliographical information

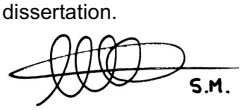
Language: English **ISSN and key title**:

ISBN: Printed version: 978-91-8039-766-7

ISBN: Electronic version: 978-91-8039-765-0

Recipient's notes **Number of pages**: 190

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature  Date 2023-09-29

MHC diversity and expression in a songbird

Temporal and transgenerational variation

Samantha Mellinger



LUND
UNIVERSITY

Coverphoto created by Tamara Emmenegger
Copyright pp 1-61 Samantha Mellinger

Paper 1 © John Wiley & Sons Ltd.
Paper 2 © PeerJ (O'Reilly and SAGE).
Paper 3 © by the Authors (Manuscript unpublished).
Paper 4 © by the Authors (Manuscript unpublished).

Faculty of Science
Department of Biology

ISBN (print): 978-91-8039-766-7
ISBN (electronic): 978-91-8039-765-0

Printed in Sweden by Media-Tryck, Lund University
Lund 2023



Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se
MADE IN SWEDEN

"It's no use going back to yesterday, because I was a different person then" *(Alice in Wonderland - Lewis Carroll)*

Table of Contents

List of papers 8
Author's contribution to the papers 9
Abstract 10
Popular science summary 12
Populärvetenskaplig sammanfattning 15

Background 19
The vertebrate immune system 19
The innate immune system 19
The adaptive immune system 20
Antigen presentation by MHC molecules 20
The Major Histocompatibility complex (MHC) 22
The MHC region 22
Evolution of the MHC multigene family 24
Forces maintaining MHC diversity and polymorphism 25
MHC diversity and gene expression in birds 26
Avian MHC diversity 26
MHC-I gene expression 28

Aims of the thesis 31

Methodology 32
Study species and study area 32
The great reed warbler genome assembly and the characterized MHC region 33
Assessing MHC diversity and MHC-I expression 34

Results and Discussion 37
Chapter I. The genomic architecture of the passerine MHC region: High repeat content and contrasting evolutionary histories of single copy and tandemly duplicated MHC genes 37
Chapter II. Improved haplotype resolution of highly duplicated MHC genes in a long-read genome assembly using MiSeq amplicons 40

Chapter III. Temporal variation and age-associated changes in MHC-I gene expression in great reed warblers. 43
Chapter IV. Inheritance of MHC-I gene expression in great reed warblers 46

Conclusion 49
Future directions and preliminary results 50

References 52

Acknowledgments – Thank you! 59



Picture 1. Silhouette of a great reed warbler singing.

List of papers

I. Westerdahl H., **Mellinger S.,** Sigeman H., Kutschera V.E., Proux-Wéra E., Lundberg M., Weissensteiner M., Churcher A., Bunikis I., Hansson B., Wolf J.B.W. and Strandh M. (2022). The genomic architecture of the passerine MHC region: High repeat content and contrasting evolutionary histories of single copy and tandemly duplicated MHC genes. *Molecular Ecology Resources,* 22, pp 2379-2395, doi: 10.1111/1755-0998.13614.

II. **Mellinger S.,** Stervander M., Lundberg M., Drews A. and Westerdahl H. (2023). Improved haplotype resolution of highly duplicated MHC genes in a long-read genome assembly using MiSeq amplicons. *PeerJ*, 11(5): e15480, doi: 10.7717/peerj.15480.

III. **Mellinger S.,** Råberg L., Hasselquist D. and Westerdahl H. Temporal variation and age-associated changes in MHC-I gene expression in great reed warblers. Manuscript.

IV. **Mellinger S.,** Råberg L., Hasselquist D. and Westerdahl H. Inheritance of MHC-I gene expression in great reed warblers. Manuscript.

In addition, I am a co-author for the following papers published during my PhD studies. These papers are **NOT** included in the thesis.

V. Neto J.M., **Mellinger S.,** Halupka L., Marzal A., Zehtindjiev P. and Westerdahl H. (2020). Seasonal dynamics of haemosporidian (Apicomplexa, Haemosporida) parasites in house sparrows *Passer domesticus* at four European sites: comparison between lineages and the importance of screening methods. *International Journal of Parasitology,* 50, pp 523-532, doi: 10.1016/j.ijpara.2020.03.008.

VI. Eltschkner S., **Mellinger S.,** Buus S., Nielsen M., Paulsson K.M., Lindkvist-Petersson K. and Westerdahl H. (2023). Extraordinary peptide-binding mode of a songbird MHC class-I molecule suggests mechanism to counter pathogen immune evasion. *Frontiers in Immunology*, 14: 1209059, doi: 10.3389/fimmu.2023.1209059.

VII. Arct A., Drobniak S.M., **Mellinger S.,** Martyka R., Gustafsson L. and Cichon M. (2022). Extra-pair paternity in Blue Tits (*Cyanistes caeruleus*) depends on the combination of social partners' age. *Ibis,* 164, pp 388-395, doi: 10.1111/ibi.13022.

Author's contribution to the papers

I. H.W. and M.S. conceptualized the study. M.S. and S.M. contributed to the methodology. H.S. and I.B. were involved in the bioinformatic assembly. S.M., V.E.K., E.P-W., A.C. and M.L. were involved in the bioinformatic annotation and/or analysis. H.W. performed the formal analysis. H.W. and M.S. wrote the original manuscript. All authors revised and approved the final manuscript.

II. S.M. and H.W. designed the study. S.M., M.S. and A.D. performed the laboratory work. S.M. and M.L. analysed the data. S.M. and H.W. conceptualized the figures and tables. S.M. prepared the figures and tables with inputs from all authors. S.M. drafted the manuscript with inputs from H.W. All authors revised and approved the final manuscript.

III. S.M. and H.W. designed the study. D.H. was responsible for the fieldwork and collection of all samples at Lake Kvismaren. S.M. performed the laboratory work. S.M., L.R. and H.W. analysed the data and conceptualized the figures and tables. S.M. prepared the figures and tables with inputs from all authors. S.M. drafted the manuscript with inputs from all authors.

IV. S.M. and H.W. designed the study. D.H. was responsible for the fieldwork and collection of all samples at Lake Kvismaren. S.M. performed the laboratory work. S.M., L.R. and H.W. analysed the data and conceptualized the figures and tables. S.M. prepared the figures and tables with inputs from all authors. S.M. drafted the manuscript with inputs from all authors.

Abstract

The immune system is of central importance to clear-off pathogens in order to keep an organism healthy. In vertebrates, the Major Histocompatibility Complex (MHC) is a genomic region that holds key genes involved in adaptive immunity and in particular in the antigen presenting process. The presentation relies on MHC class I (MHC-I) and MHC class II (MHC-II) molecules that are encoded by MHC-I and MHC-II genes, respectively. In passerine birds (songbirds; order Passeriformes), high MHC-I and MHC-IIB diversity (*i.e.,* number of MHC alleles in an individual) has been reported compared to other avian orders. Each MHC molecule can present a limited number of antigens, therefore high MHC diversity could potentially be an advantage for presenting various antigens from different pathogens. My aim was to gain a deeper understanding on what mechanisms have led to high MHC diversity in passerines, how much of this MHC diversity is expressed and whether it varies in time and across generation. In this thesis, I mainly focused on a particular passerine, the great reed warbler (GRW; *Acrocephalus arundinaceus*) for which high MHC diversity has been previously described.

In chapter I, the MHC genomic region was characterized in the GRW and also in three additional passerine species from which long-read genome assemblies were available. The key findings were that the MHC region is extended in Passeriformes and the MHC diversity in the GRW has expanded through repeated and multiple gene duplication events of single and trios of MHC genes; these MHC paralogs are likely to have evolved, and still evolve, under the birth-and-death model process. In chapter II, I wanted to assess the correctness of the estimated haploid MHC diversity found in the GRW genome and to compare the resolution between the estimated diploid MHC diversity found with long- and short-read sequencing techniques. The haplotype resolution was improved using short-read amplicons and I found that two MHC-IIB scaffolds holding tandemly duplicated genes were most likely complementary haplotypes. MHC diversity was equally well estimated when using long- and short-read sequencing but short-read amplicons give limited information on the putative functional MHC diversity. In the last two chapters, I wanted to investigate the functional relevance of the MHC-I diversity in the GRW by looking at gene expression in individuals from our long-term monitored study population breeding at Lake Kvismaren in Sweden. I found that 71% of the MHC-I diversity is expressed in the GRW but the different MHC-I alleles are expressed at different levels (*i.e.,* different relative expression) with generally one to two alleles being highly expressed and the other alleles being moderately or lowly expressed in each individual. In chapter III, I investigated temporal variation in MHC-I relative expression in individuals sampled both as nestlings and as breeding adults at our study site. I found that MHC-I relative expression in blood is highly stable throughout an individual's lifetime. In chapter IV, I investigated transgenerational variation in MHC-I relative expression in 17 families. Haplotypes were determined

based on the segregation of parental alleles in offspring. MHC-I relative expression in haplotypes in offspring is to a large extent genetically inherited from their parents and does not change with offspring age. I found a stronger relationship between mothers and offspring suggesting a maternal effect, but I have not found any epistasis effect.

To conclude, I found that a high proportion of the MHC diversity was expressed in blood in the great reed warbler but only a small fraction was highly expressed. Also, MHC-I gene expression does not vary over time and is to a large extent inherited.

Popular science summary

The survival of an organism depends greatly on the ability of its immune system to detect and fight-off pathogens entering its body. To do so, the immune system must be able to distinguish between components from the organism (*i.e.,* self) and components of foreign origin like pathogens (*i.e.,* non-self). An immune response can be either fast but with low specificity, which relies on the innate immunity, or slow but with high specificity, which relies on the adaptive immunity. The later matures over lifetime depending on which pathogens an individual has encountered and can provide long-term immunity against specific pathogens.

The Major Histocompatibility Complex (MHC) molecules play an essential role in adaptive immunity because they help the immune cells (T-lymphocytes) to identify pathogen-infected cells. MHC molecules present small proteins, so called peptides, at the surface of nucleated cells. MHC molecules can present both self-peptides *(i.e.,* from the own body) and non-self-peptides (*i.e.,* antigens) and each "MHC molecule-peptide" complex interacts with receptors on the surface of T-lymphocytes. If a T-lymphocyte receptor recognizes an antigen bound to an MHC molecule, the T-lymphocyte will trigger an adaptive response leading, for example, to the destruction of the infected cell. An appropriate immune response must only be triggered when an MHC molecule presents an antigen (non-self-peptide). If the antigen was synthetized inside the cell (intra-cellular), for example a cell infected by a virus, MHC class I (MHC-I) molecules present the antigen. If the antigen is extra-cellular, for example a microbe ingested by an antigen presenting cell, MHC class II (MHC-II) molecules present the antigen.

MHC molecules are encoded by MHC genes and the number of MHC genes vary considerably between different species. For example, there are three MHC-I genes in humans and two MHC-I genes in the chicken (*Gallus gallus*), while the number of MHC genes is much higher in songbirds. Illustrated by the willow warbler (*Phylloscopus trochilus*) were a minimum of 19 MHC-I genes was estimated. Having a high number of MHC genes could be an advantage especially because each MHC molecule only can present a limited number of antigens. Therefore, the more MHC genes an individual have, the more antigens it can potentially present. Does this mean that songbirds potentially have a "super-immune system"?

I wanted to understand why songbirds have so many MHC genes and whether all these MHC genes have functional relevance. In other words, are all MHC genes encoding MHC molecules with an important function in adaptive immunity? One way to investigate functional relevance is to look at gene expression. My work mainly focused on a wild, long-distance migratory songbird, the great reed warbler (*Acrocephalus arundinaceus*). It is a medium-sized songbird that weight about 30 grams. Individuals used in this thesis are part of a long-term monitored study

population breeding at lake Kvismaren in south central Sweden and wintering in sub-Saharan Africa.

To start, I wanted to know how many MHC genes great reed warblers have. To do so, I used the available great reed warbler genome assembly, which is based on a single individual, and for which the MHC region has been partly characterized. According to this genome the great reed warbler has 15 MHC-I and 56 MHC-IIB genes that seem to be functional. These gene copies were most likely generated by multiple and possibly repeated gene duplication events. Even more fascinating, I found evidence that "trios" (three genes in tandem, close together) of MHC-I and "trios" MHC-IIB genes have been duplicated. Gene duplications of several genes together might explain the high number of MHC gene copies observed in great reed warblers.

It is challenging to assemble the MHC region with the high number of MHC gene copies in the great reed warbler, because some parts of these MHC genes are highly similar. Estimating the number of MHC gene copies in the great reed warbler genome is also challenging, because other parts of the different MHC gene copies are highly dissimilar, thus the number of MHC genes can be wrongly estimated. Therefore, I used the DNA sequence information on MHC genes obtained from the great reed warbler genome together with sequence information obtained with a different sequencing technique, so called short-read amplicon sequencing, widely used in MHC studies in non-model organisms. As a result, I found that the number of MHC-IIB genes was overestimated in the great reed warbler genome and should only comprise 48 MHC-IIB genes, and not 56 MHC-IIB genes as previously thought.

Thereafter, I investigated the functional relevance of these MHC genes and focused in particular on MHC-I gene expression. Like humans, birds are diploid organisms, thus can have one or two variants (*i.e.,* alleles) for a single gene and from now on, all the MHC alleles found in an individual are referred as 'MHC diversity'. I found that on average 71% of the MHC-I diversity in the genome is expressed in the blood in great reed warblers. However, the level of expression between MHC alleles differs with one to two MHC-I alleles being highly expressed and the rest being moderately or lowly expressed. How come that so few MHC-I alleles are highly expressed? Perhaps different MHC alleles are highly expressed at different time points?

Therefore, I further investigated whether there were temporal (from nestling to adult stage) and transgenerational (from parents to offspring) variation in MHC-I gene expression in blood. The great reed warbler is a suitable study species because of its high philopatry. In fact, a large proportion of the individuals in our study population return to their breeding site after wintering in Africa. Thus, "family pedigree" of each breeding pair are recorded every year and individuals are monitored over several breeding seasons. These records are particularly unique when working with

wild songbird. I found that MHC-I gene expression is temporally stable within individual from the nestling stage to the adult stage and onwards. The same one or two MHC-I alleles were highly expressed independent of age. I also found that MHC-I gene expression is to a large extent inherited from both parents, hence overall the MHC-I expression pattern is highly stable across generations. Although, MHC-I gene expression in the offspring tend to resemble the MHC-I gene expression of their mothers (*i.e.,* maternal effect) more than that of their fathers. I found that MHC-I gene expression is genetically set and remain stable throughout an individual's lifetime.

Taken together, my results highlight the importance of further investigating the functional relevance of MHC genes in wild animals. Indeed, infectious diseases can constitute a major threat for wildlife and it remains essential to understand the adaptive advantage of high MHC diversity in songbirds.

Populärvetenskaplig sammanfattning

En individs överlevnad beror till stor del på hur bra immunförsvaret upptäcker och oskadliggör patogener som lyckas ta sig in i kroppen. För att lyckas med detta måste immunförsvaret skilja på kroppsegena (själv) och främmande ämnen (patogener). En immunreaktion kan antingen vara snabb och med låg specificitet, det nedärvda immunförsvaret, eller långsam och med hög specificitet, det adaptiva immunförsvaret. Det senare ändras under livets gång och speglar vilka patogener en individ har varit utsatt för och kan ge livslång immunitet mot specifika patogener.

Molekylerna från det så kallade histokompatibilitetskomplexet (Major Histocompatibility Complex, MHC) har en viktig roll i det adaptiva immunförsvaret och hjälper immunceller (T-lymfocyter) att identifiera infekterade celler. MHC-molekylerna presenterar små proteiner, så kallade peptider, på cellytan. MHC-molekylerna kan presentera både själv-peptider, peptider från egna kroppen, och icke-själv-peptider, peptider från patogener (antigen), och varje 'MHC-molekyl-antigen-komplex' interagerar med receptorer på cellytan av T-lymfocyter. Om T-lymfocyt-receptorn binder till antigenet som presenteras av en MHC molekyl kommer T-lymfocyten att initiera en adaptiv immunrespons som resulterar i att cellen oskadliggörs. Denna adaptiva immunrespons initieras endast när antigenet som presenteras av MHC molekylen är 'icke-själv'. Om antigenet har syntetiserats inne i cellen, från till exempel en virusinfekterad cell, så kommer antigenet att presenteras på en MHC klass I molekyl (MHC-I). Om antigenet är extra-cellulärt, från till exempel en mikrob som har tagits upp av en antigen presenterande cell, så kommer antigenet att presenteras av en MHC klass II molekyl (MHC-II).

MHC-molekylerna kodas av MHC-generna och antalet MHC gener per individ varierar avsevärt mellan olika arter. Det finns till exempel tre MHC-I gener hos människor och två MHC-I gener hos höns (*Gallus gallus*), medan antalet MHC-gener är mycket högre hos sångfåglar; lövsångaren (*Phylloscopus trochilus*) har minst 19 olika MHC-I gener. Det kan vara en fördel att ha många MHC-gener eftersom varje MHC-molekyl endast kan presentera ett begränsat antal antigen. Därför är det möjligt att ju fler MHC-gener en individ har desto fler antigener kan den potentiellt presentera. Betyder detta att sångfåglar har ett "super-immunförsvar"?

Jag ville förstå vilka mekanismer som ligger bakom det stora antalet MHC-gener hos sångfåglar och ifall alla dessa gener är funktionella. Med andra ord, kodar alla MHC-gener för MHC-molekyler med en viktig funktion i det adaptiva immunförsvaret? Ett sätt att undersöka funktion är att ta reda på ifall generna uttrycks. Mitt avhandlingsarbete har fram för allt fokuserat på en lång-flyttande sångfågel, trastsångaren (*Acrocephalus arundinaceus*). Trastsångaren är en mellanstor sångfågel som väger 30 gram. Trastsångarindividerna i den här avhandlingen kommer från en långtidsstudie av en population vid sjön Kvismaren i

mellansverige och fåglarna som häckar i vår studiepopulation övervintrar i Afrika söder om Sahara.

Först ville jag ta reda på hur många MHC-gener trastsångaren har och därför använde jag trastsångargenomet vars innehåll är baserat på en enda individ, och i detta genom är MHC-regionen delvis karaktäriserad. Enligt genomet har trastsångare 15 MHC-I gener och 56 MHC-II gener och dessa gener verkar vara funktionella. Gen-kopiorna har troligtvis skapats av många, och möjligtvis upprepade, genduplikationer. Ännu mer fascinerande är att det finns bevis för att ”trios” (tre gener i rad, tätt ihop) av MHC-I och ”trios” av MHC-IIB gener har duplicerats. Genduplicering av många gener samtidigt kanske kan förklara det stora genantalet vi fann hos trastsångare.

Det är utmanande att bygga ihop MHC-regionen i trastsångargenomet eftersom vissa delar av MHC-generna är mycket lika varandra när man jämför de olika genkopiorna. Vidare är det utmanande att estimera hur många MHC-gener det faktiskt finns i trastsångargenomet, eftersom andra delar av MHC generna är mycket olika varandra mellan de olika genkopiorna och därmed kan antalet MHC-gener i genomet överskattas. Därför använde jag DNA-sekvensinformation från både trastsångargenomet och DNA-sekvensinformation baserat på en ytterligare sekvenseringsmetod, så kallad kort-sekvens amplikon-sekvensering, en metod som ofta används för MHC studier av icke-modell-organismer. Jag fann att antalet MHC-IIB gener i trastsångargenomet hade överskattats och att det endast fanns 48 MHC-IIB gener och inte 58 MHC-IIB gener som man tidigare trott.

Därefter undersökte jag den funktionella relevansen av dessa MHC-gener och fokuserade speciellt på genuttryck. Precis som människor är fåglar diploida organismer, och de kan därför ha en eller två olika genvarianter (alleler) för varje gen och det totala antalet MHC-alleler per individ kallar jag från och med nu ’MHC-diversitet’. Jag fann att i medeltal uttrycks 71% av MHC-I-diversiteten som finns i genomet i blodet. Men graden av genuttryck skiljer sig mellan MHC-allelerna och endast 1-2 gener är höguttryckta medan resten är medelhögt eller lågt uttryckta. Varför är så få MHC-alleler högt uttryckta? Kanske olika MHC alleler är högt uttryckta vid olika tidpunkter?

Därför gick jag vidare och undersökte ifall genuttrycket av MHC-I i blod skiljer sig över tid (från ung, dvs bo-unge, till adult individ) och över generationsgränser (mellan föräldrar och avkomma). Trastsångaren är en bra studieorganism eftersom den är ortstrogen. En stor andel av individerna i vår studiepopulation återvänder till sitt häckningsområde efter att ha övervintrat i Afrika. Familjesläktskapen av varje häckande par noteras varje år och individerna studeras över flera häckningssäsonger. Dessa observationer är speciellt unika när man studerar vilda fåglar. Jag fann att uttrycket av MHC-I generna var stabilt över tiden från ung till adult individ och även vidare genom livet som adult individ. Samma två MHC-I alleler var höguttryckta oberoende av ålder. Jag fann också att MHC-I genuttrycket

till stor del nedärvdes från föräldrarna. MHC-I genuttrycket var alltså mycket stabilt över generationsgränser. Men, MHC-I genuttrycket i avkomman tenderade att likna gen-uttrycket i sin mamma mer än i sin pappa. Jag fann att MHC-I genuttrycket bestäms genetiskt och är stabilt under en individs livstid.

För att sammanfatta, mina resultat visar att det är viktigt att undersöka funktionen av MHC generna i vilda djur. Infektionssjukdomar kan utgöra ett stort hot mot vilda djur och det är viktigt att förstå de adaptiva fördelarna med den höga MHC diversitet som finns hos sångfåglar.

Background

The major histocompatibility complex (MHC) genes are highly diverse and polymorphic and the maintenance of this high genetic diversity has gained a lot of attention from evolutionary ecologists and biologists (Sommer, 2005). MHC genes are involved in antigen presentation, a key feature of adaptive immunity, and are subject to pathogen-driven selection which is thought to maintain high MHC polymorphism (Radwan *et al.* 2020). However, MHC genes are also involved in other processes such as autoimmunity, maternal-foetal interactions and mate choice. Therefore, MHC genes represent some of the most fascinating genes to study, past and present, evolutionary and adaptive processes.

The vertebrate immune system

The main function of the immune system is to protect against and fight off pathogens that cause infection and illness in organisms (Abbas *et al.* 2020). Physical and chemical barriers (*e.g.,* the skin or mucous secretions) trap infectious agents and prevent them from entering tissues. Pathogens able to cross those barriers will then face two different but complementary immune defence systems: the innate and the adaptive immunity (Riera Romo *et al.* 2016; Marshall *et al.* 2018). Both rely on a key mechanism which is to differentiate between self (*i.e.,* the organism) and non-self (*i.e.,* pathogens or foreign elements) (Chaplin, 2010).

The innate immune system

Innate immunity is the first line of an organism's immune defence against pathogens such as parasites, bacteria and viruses (Abbas *et al.* 2020). Pathogen recognition is possible through the recognition of conserved molecular structures shared by pathogens (pathogen-associated molecular patterns) but that are not present in the host. Thus, the innate immune system has low specificity but triggers a fast immune response (Chaplin, 2010). Recognition mechanisms involved in innate immunity can also stimulate an appropriate adaptive immune response (Iwasaki & Medzhitov, 2015; Marshall *et al.* 2018).

The adaptive immune system

Adaptive immunity is the second line of defence. It arose in jawed vertebrates to detect and fight off pathogens with high specificity (Cooper & Alder, 2006). It relies on the ability of specialized immune cells (lymphocytes) to discriminate between self (host) and non-self (*e.g.,* pathogen) molecules, also called antigens (Abbas *et al.* 2020). The adaptive immune response is delayed but can give long-term protection, called immunological memory. Therefore, if the same pathogen is encountered again, the immune response will be much faster and also more efficient (Chaplin, 2010; Abbas *et al.* 2020).

There are two types of adaptive immune responses: humoral immunity involves B-lymphocytes that secrete antibodies which bind directly to structures of extracellular pathogens and cell-mediated immunity which involves cytotoxic T-lymphocytes that eliminates intracellular pathogens (Chaplin, 2010, Abbas *et al.* 2020). B-lymphocytes can be activated with or without T-lymphocytes help (Abbas *et al.* 2020). Thus, both adaptive immune responses rely on antigen presentation on the surface of nucleated host cells by the Major Histocompatibility Complex (MHC) molecules (Figure 1).

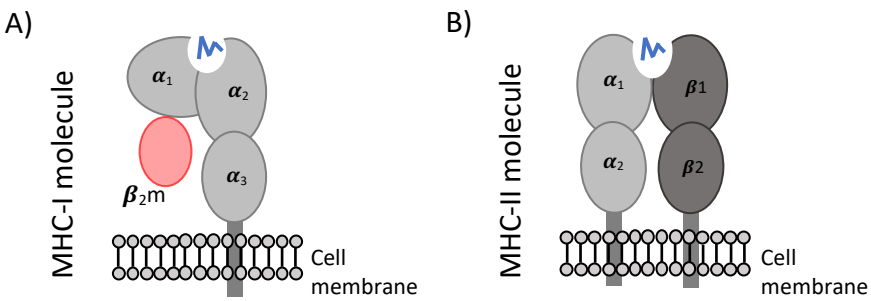


Figure 1. Schematic representation of an A) MHC-I and an B) MHC-II molecule presenting peptides (blue zigzag) at the cell surface.
The cell membrane separates the inside (below) and the outside (above) of the cell. The "MHC molecule-peptide" complex is anchored to the cell membrane and is displayed outside the cell. A) The α_1, α_2 and α_3 domains are indicated along the MHC-I molecule alpha chain (light grey). B) The MHC-II molecule is composed of the α chain (light grey; for which the α_1 and α_2 domains are indicated) and the beta chain (dark grey; for which the ß $_1$ and ß $_2$ domains are indicated).

Antigen presentation by MHC molecules

MHC molecules are key components of antigen presentation and they enable appropriate adaptive immune responses. Displayed on the surface of nucleated host cells, they present peptides from the host (self) or from pathogens (non-self) in their peptide binding pocket (Neefjes *et al.* 2011; Rock *et al.* 2016). Those "MHC molecule-peptide" complexes interact with T-lymphocyte receptors (TCR) situated on T-lymphocytes' surface. If T-lymphocytes recognize a non-self peptide (*i.e.,* an

antigen), it will trigger an immune response towards the infected cell (Abbas *et al.* 2020). Each MHC molecule can present a limited number of peptides. We can distinguish two classes of MHC molecules, MHC class I (MHC-I) and MHC class II (MHC-II), which are similar in protein structure but are encoded by different MHC genes and interact with two different T-lymphocyte types (Figure 1).

MHC class I (MHC-I) molecules

MHC-I molecules are expressed on all nucleated cells and they present peptides from intra-cellular pathogens to cytotoxic T-lymphocytes (CD8+ T-lymphocytes). They are heterodimers composed of a polymorphic heavy chain (the alpha chain including α_1, α_2 and α_3 domains, encoded by an MHC class I gene) and an associated monomorphic light chain, the beta-2-microglobulin (ß2m) (Figure 1A, Figure 2, Rock *et al.* 2016; Abbas *et al.* 2020).

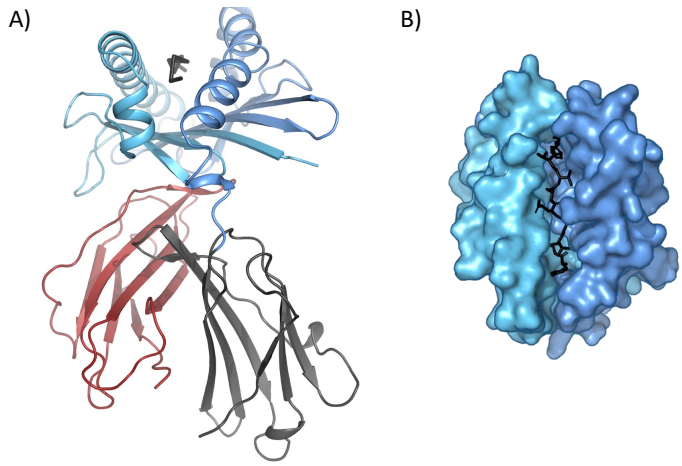


Figure 2. Structure of a great reed warbler "MHC-I molecule-peptide" complex (ribbon representation) and 3-D conformation of the peptide binding pocket holding the peptide.
A) Quaternary structure of a MHC-I molecule with $\boldsymbol{\alpha}_1$ (light blue), $\boldsymbol{\alpha}_2$ (dark blue) and $\boldsymbol{\alpha}_3$ (dark grey) domains and the associated ß2m light chain (red). The MHC-I molecule holds a peptide (black) in the peptide binding pocket composed of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ domains. B) 3-D conformation of a peptide (black stick) bound to the MHC-I peptide binding pocket (top view) composed of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ domains. Figures made by Sandra Eltschkner.

MHC class II (MHC-II) molecules

MHC-II molecules are mostly expressed on specialized antigen presenting cells (APCs) such as dendritic cells, macrophages and B-cells. They present peptides from phagocytosed pathogens to helper T-lymphocytes (CD4+ T-lymphocytes) (Rock *et al.* 2016; Abbas *et al.* 2020). They are heterodimers composed of two chains which are encoded by two different MHC-II genes: the MHC-II alpha (MHC-IIA) gene encodes the α chain and the MHC-II beta (MHC-IIB) gene encodes the ß chain (Figure 1B).

The Major Histocompatibility complex (MHC)

The MHC region

In humans, the MHC is a single large genomic region (4–7 Mbp) that comprises a set of linked loci encoding MHC molecules and many other immune-related molecules (Horton *et al.* 2004). MHC genes are a multigene family which is present in all jawed vertebrates and shares similar functions among species (Nei & Rooney, 2005). Characterization of the MHC region(s) and the MHC diversity in non-model organisms is an essential step toward comparative analyses of the MHC region organization between different vertebrate species (Kulski *et al.* 2002; Kelley *et al.* 2005) and to reconstruct the macroevolutionary history of the MHC genes in particular.

The mammalian MHC

The human MHC is called the Human Leukocyte Antigen (HLA) and is located on chromosome 6. The class I region includes three classical MHC genes (HLA-A, B and C) and three non-classical MHC genes (HLA-E, F and G). The class II region also includes three classical MHC loci (HLA-DP, -DQ and -DR), but only two non-classical MHC loci (HLA-DM and -DO) (Horton *et al.* 2004). Classical and non-classical genes are distinguished by their level of polymorphism, expression patterns and functions (Rodgers & Cook, 2005; Allen & Hogan, 2013), where classical genes have high polymorphism and high expression, whereas non-classical genes have low polymorphism and are often lowly expressed.

In other mammalian species, MHC has been relocated and differ in size (Kelley *et al.* 2005; Shiina *et al.* 2017), though certain characteristics have been kept rather conserved: several regions with similar function, including the class I and class II regions (Trowsdale, 1995; Kumánovics *et al.* 2003). However, the number of MHC gene copies within a locus (genes) and the number of MHC loci differ between species. For example, cats (*Felis catus*) have lost loci homologous to the HLA-DQ locus in humans (Yuhki *et al.* 2003) whereas the rhesus macaque (*Macaca mulatta*) gained nine MHC-B genes homologous to the HLA-B locus in humans (Kelley *et al.* 2005; Heijmans *et al.* 2020).

The avian MHC

The avian MHC was first described in the chicken (*Gallus gallus*). It is located on chromosome 16 and includes two genetically unlinked loci, classical MHC genes (the B complex) and non-classical MHC genes (the Rfp-Y region) (Kaufman *et al.* 1995; Kaufman *et al.* 1999).

The core MHC region in chicken is called the B complex, also called the “minimal essential MHC”. The BF region corresponds to the mammalian class I region and

the BL region corresponds to the mammalian class II region. The chicken B-complex is condensed both in size (~92 kb) and gene number compared to the human MHC (Kaufman *et al.* 1999; Kelley *et al.* 2005; Shiina *et al.* 2006). The BF region comprises two classical MHC-I genes (B-F1 and B-F2) and the BL comprises two classical MHC-IIB (B-LB1 and B-LB2) genes (Figure 3). There is only a single MHC-IIA (B-LA) gene in the chicken and it is located away from the core MHC region while in mammals MHC-IIA and MHC-IIB genes are found in pairs. Another striking difference in the genomic organization between chicken and humans is the genes encoding transporters associated with antigen processing (TAP genes, involved in the antigen loading within the cell). The two MHC class I genes in chicken are neighbouring the TAP genes in the class I region (Figure 3) while these are situated in the class II region in mammals (Kaufman *et al.* 1995; Kaufman *et al.* 1999; Hess & Edwards, 2002; Lankat-Buttgereit & Tampé, 2002; Walker *et al.* 2005).

So far, the core MHC region has been characterized in several bird species among different avian orders, including Galliformes (turkey (*M. gallopavo*), Chaves *et al.* 2009; golden pheasant (*C. pictus*), Ye *et al.* 2012; black grouse (*T. tetrix*), Wang *et al.* 2012; greater prairie chicken (*T. cupido*) Eimes *et al.* 2013; Japanese quail (*C. japonica*), Shiina *et al.* 2004; Shiina *et al.* 2006), Pelecaniformes (crested ibis (*N. nippon*), Chen *et al.* 2015) and Ciconiiformes (oriental stork (*C. boyciana*), Tsuji *et al.* 2017) (Figure 3). Note that the MHC core region has been partially described as well in Anseriformes with the mallard duck (*A. platyrhynchos*, Moon *et al.* 2005) and in Passeriformes with the zebra finch (*T. guttata*, Balakrishnan *et al.* 2010; Ekblom *et al.* 2011).

In Galliformes and Ciconiiformes, the core MHC region is compact but in Pelecaniformes several MHC-I genes are found in an extended class I region possibly separated from the core MHC region. The gene content and gene synteny differ slightly between avian orders and is outlined below for TAP, MHC-I and MHC-II genes.

i) In Galliformes and Ciconiiformes TAP genes and classical MHC-I genes are situated next to each other but in Pelecaniformes, several MHC-I genes were also found in the extended class I region as mentioned above (Figure 3). The partial characterization of MHC in Anseriformes has shown that TAP and MHC-I genes are also situated next to each other. However, in the zebra finch (Passeriformes), TAP genes are not closely situated to MHC-I genes (Balakrishnan *et al.* 2010).

ii) MHC-IIA and MHC-IIB are found in “dyads” in Pelecaniformes and in Ciconiiformes (Figure 3) whereas in Galliformes the single MHC-IIA gene is located away from the MHC-IIB genes.

iii) The number of MHC-I and MHC-IIB genes in the core MHC region differs between avian orders even between closely related (Figure 3). For example, the

Japanese quail has seven MHC-I genes (of these, only four MHC-I genes are functional) whereas the chicken has only two MHC-I genes.

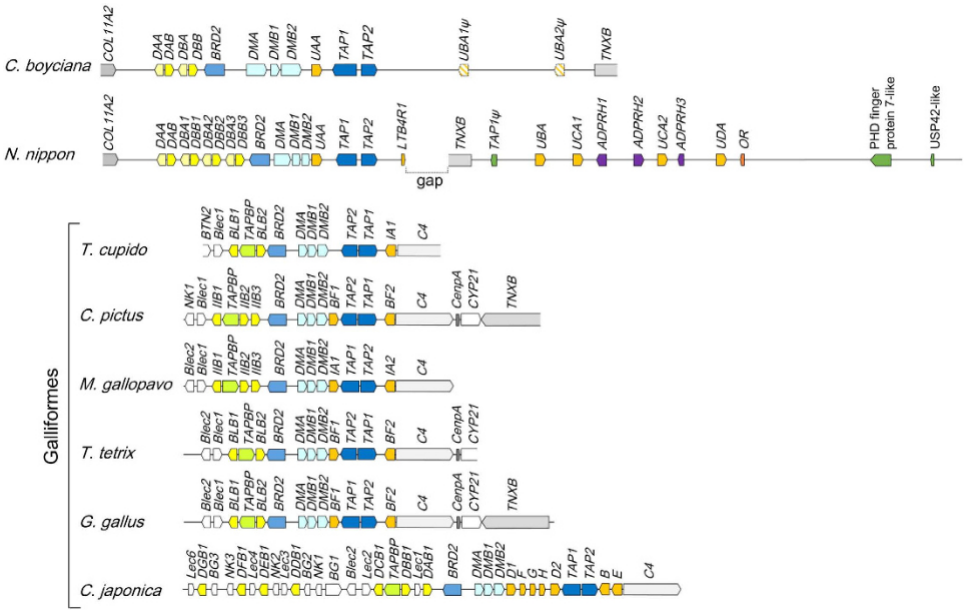


Figure 3. Schematic representation of the core MHC region organization in different avian orders. The MHC core region have been characterized in six species belonging to the Galliformes order and in two other species belonging to the Pelecaniformes (*N. nippon*) and to the Ciconiiformes (*C. boyciana*) orders. The core MHC region comprises a set of linked loci, MHC-I genes (orange), MHC-IIA and MHC-IIB genes (yellow) and TAP genes (dark blue). Figure taken from Tsuji *et al.* (2017).

Evolution of the MHC multigene family

The number of MHC gene copies per individual (*i.e.,* MHC diversity) varies between species and in some cases the MHC genes have expanded leading to many paralogs whereas in other species they have not expanded or even have been lost (Burri *et al.* 2010). Comparing the MHC region organization between species in combination with phylogenetic analysis of MHC genes is key in understanding the evolution of MHC genes (Trowsdale, 1995; Kelley *et al.* 2005, Kaufman, 2018). The birth and death hypothesis has been used to describe the evolutionary history of multigene families like the immunoglobulin genes (Ohta & Nei, 1994) and the MHC genes (Nei *et al.* 1997). This hypothesis proposes that new MHC genes arise by gene duplication, a phenomenon suggested to happen through ectopic recombination at meiosis (Ohta, 2000). Thereafter, new MHC genes are either maintained as functional genes ("birth") or lost ("death", *e.g.,* becoming pseudogenes because of deleterious mutations) (Nei *et al.* 1997; Nei & Rooney, 2005) (Figure 4A). This hypothesis has been supported to explain the evolutionary history of the avian class IIB genes. The two lineages, DAB1 and DAB2, emerged

by duplication before the avian radiation and independent losses of either DAB1 or DAB2 in some avian lineages occurred independently (Burri *et al.* 2010; Goebel *et al.* 2017) (Figure 4B).

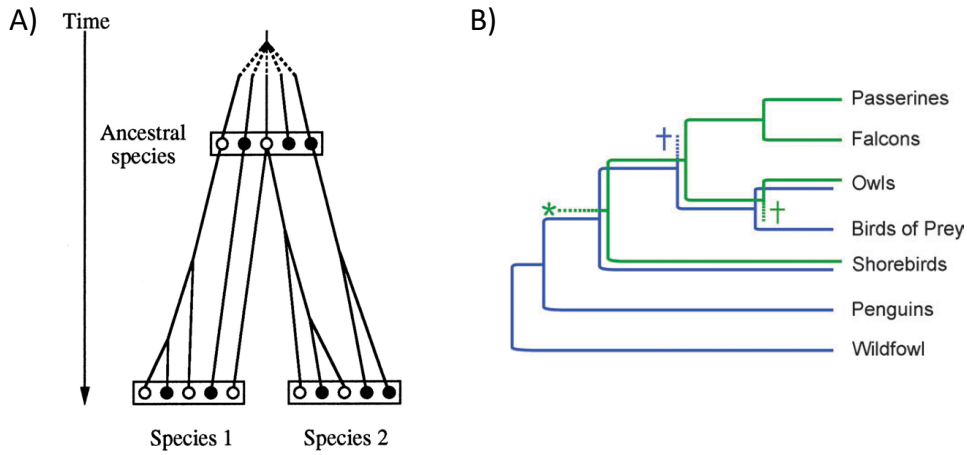


Figure 4. The birth and death model and the proposed evolutionary history of avian MHCIIB. A) Schematic representation of gene duplication events resulting in several paralogs (circles) in the ancestral species with either gene birth (white circles) or gene death (black circles). Figure taken from Nei *et al.* (1997). B) Hypothetical evolutionary history of avian MHCIIb exon 3 (Burri *et al.* 2010). Avian MHC-DAB1 (green) and MHC-DAB2 (blue) lineages are the result of a single gene duplication event (*). Independent losses of one of the paralogs (a cross) have happened twice in the bird phylogeny and the pattern agrees with the birth and death model. Figure taken from Burri *et al.* (2010).

However, assessing the evolutionary relationship of MHC genes between bird species is challenging (Goebel *et al.* 2017), particularly when paralogous genes are more similar within species than orthologous genes between species. This can happen when gene duplication precedes the speciation event whereafter inter-locus genetic exchange, also called gene conversion, makes the paralogs highly similar within species (Li, 1997; Nei & Rooney, 2005, Spurgin *et al.* 2011). This phenomenon leads to concerted evolution and has been shown for the MHC class IIB genes (DAB1 and DAB2) in the ring-necked pheasant (*Phasianus colchicus*). The DAB1 and DAB2 alleles clustered according to the species they belonged to even though the two genes arose through duplication before the two species split (Wittzell *et al.* 1999).

Forces maintaining MHC diversity and polymorphism

There are several non-exclusive mechanisms which can result in balancing selection on MHC genes which maintains genetic diversity. One of the mechanisms is

heterozygote advantage where heterozygous individuals will be able to recognize more pathogens than homozygous individuals and therefore have higher fitness (Doherty & Zinkernagel, 1975; Hughes & Nei, 1988; Radwan *et al.* 2020). This process selects heterozygotes which results in even MHC allele frequencies in the population. Another mechanism is negative frequency-dependent selection where novel rare alleles will provide greater resistance against pathogen and will increase in frequency in a population whereas common alleles will confer lower resistance and therefore decrease in frequency in the population (Takahata & Nei, 1990; Spurgin & Richardson, 2010). The divergent allele advantage hypothesis states that highly divergent alleles at a locus will encode molecules able to present a broader range of peptides from different pathogens (Wakeland *et al.* 1990; Lenz, 2011). Finally, MHC alleles can also be selected depending on the type and abundance of pathogens encountered that fluctuate in time and space (Hill, 1991; Hedrick, 2002; Westerdahl *et al.* 2004a; Spurgin & Richardson, 2010). This process is referred as fluctuating selection and is driven by biotic and abiotic variation.

Although MHC polymorphism is thought to be mainly promoted and maintained by selection from pathogens, additional mechanisms have also been suggested to maintain diversity such as mate choice (Zelano & Edwards, 2002; Piertney & Olivier, 2005; Milinski, 2006).

MHC diversity and gene expression in birds

Avian MHC diversity

Passerines (songbirds) are a particularly interesting avian model system. They harbor, in general, a greater MHC diversity than non-passerines (Minias *et al.* 2018; O'Connor *et al.* 2019), although it varies a lot between species. Several studies in songbirds have reported a huge diversity of MHC-I alleles, especially in birds of the genera *Sylvioida* and *Vireo* ((Figure 5; Minias *et al.* 2018). For example, up to 65 MHC-I alleles per individual have been found in the sedge warbler *Acrocephalus schoenobaenus* (Biedrzycka *et al.* 2017; O'Connor *et al.* 2019) (Figure 5) and up to 26 MHC-I alleles per individual have been found in the great reed warbler *Acrocephalus arundinaceus* (Roved *et al.* 2022) The same observations have been made for MHC-IIB in songbirds, for example, more than 40 alleles have been found in a single common yellowthroat individual (*Geothlypis trichas*) (Bollmer *et al.* 2010; O'Connor *et al.* 2019).

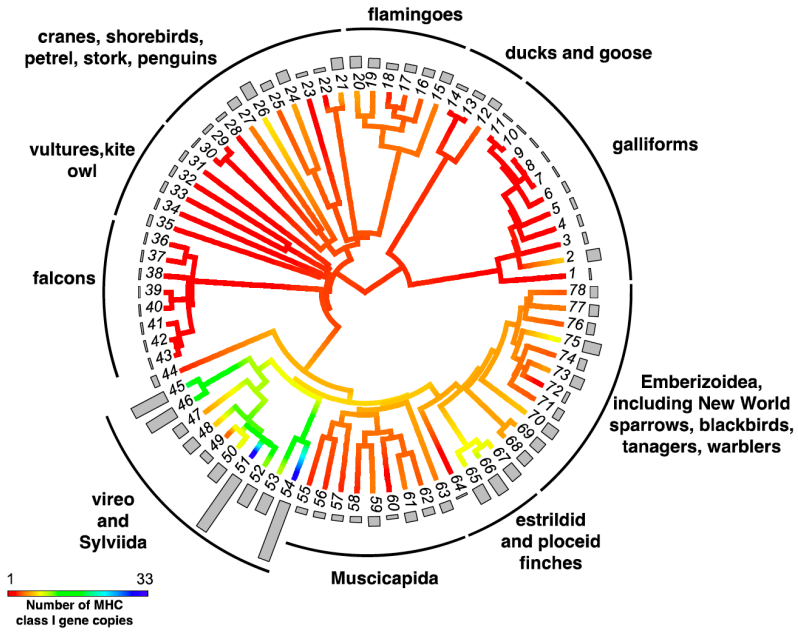


Figure 5. Estimates of the number of MHC-I gene copies along the branches of the avian phylogenetic trees.
Bars indicated the estimated number of MHC-I gene copies. Each number represent a species with for example the sedge warbler corresponding to the 51st branch. Genera belonging to the Passeriformes order include the Emberizoidae, the estrildid and ploceid finches, the Muscicapida and the vireo and Sylviida. Figure from O'Connor *et al.* (2019) modified from Minias *et al.* (2018).

Most studies investigating MHC diversity target MHC-I exon 3 and MHC-IIB exon 2 sequences which are thought to be the most variable gene regions, encoding parts of the peptide-binding pocket that interact with peptides. The genotyping of the MHC-I and IIB genes is performed by Polymerase Chain Reaction (PCR) that aims to amplify the target exons in all available MHC gene copies in each individual, where after these amplicons are high-throughput sequenced (Babik, 2010; O'Connor *et al.* 2016). However, in recent years, DNA sequencing technologies have become more widely available and long read sequenced avian genomes have also become available. Therefore, estimates of MHC diversity can be performed by blast searching MHC alleles in long-read genomes. He *et al.* (2021) have identified up to 185 MHC-IIB gene copies in the golden-collared manakin (*Manacus vitellinus*, Pipridae, Suboscine), a diversity far more impressive than previously found in other passerine species. Therefore, the combination of short-read sequencing (targeting MHC class I exon 3 and MHC class IIB exon 2) and long-read sequencing (to obtain high-quality genome sequencing) is an important step to improve the estimation of MHC diversity in passerines (O'Connor *et al.* 2019; He *et al.* 2021).

MHC-I gene expression

The expressed MHC diversity

The expressed MHC diversity (*i.e.,* the number of MHC genes/alleles expressed compared to the genomic MHC diversity) varies between species of different avian orders. Among Galliformes, the two classical MHC-I genes were expressed in the chicken (Kaufman *et al.* 1999). Likewise, in the quail (Shiina *et al.* 2004), two classical MHC-I genes were expressed but also two putatively non-classical MHC-I genes. In the mallard duck (Anseriformes), two out of three functional classical MHC-I genes have been found expressed (Moon *et al.* 2005) and in the crested ibis (Pelacaniformes) one major classical and three minor classical MHC-I genes were expressed (Chen *et al.* 2015). In the red knot (*C. canatus;* Charadriiformes), it is nine out of 11 MHC-I alleles that were expressed (found in complementary DNA (cDNA), Buehler *et al.* 2013).

A recent study comparing the expressed MHC diversity between 13 passerine species have shown that all passerine species express fewer MHC-I alleles (in cDNA) than found at the genomic level (gDNA), hence only a proportion of the alleles are expressed (Figure 6; O'Connor & Westerdahl, 2021). Different proportions of MHC-I alleles were expressed in different species (Figure 6) but the proportion of expressed MHC-I genes within a species remains relatively similar within avian families (O'Connor & Westerdahl, 2021).

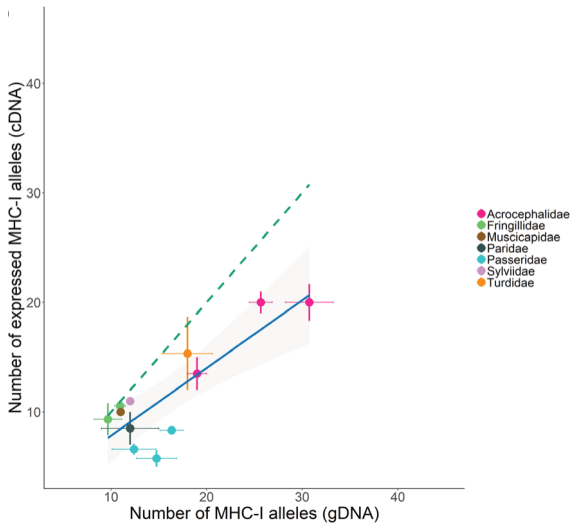


Figure 6. Empirical relationships between genomic and expressed MHC diversity across 13 songbird species from seven avian families.
Each dots represents a songbird species and shows the number of MHC-I alleles detected in genomic DNA (gDNA) and the number that are expressed in complementary DNA (cDNA). Each color represent an avian family and the blue line indicates the relationship between genomic and expressed MHC diversity across all songbird species.In comparison, the green dotted line represent a scenario in which all MHC-I alleles are expressed. Figure taken from O'Connor and Westerdahl (2021).

Among the *Fringillidae* (Figure 6), 97% of the genomic MHC-I alleles have been found expressed in the European greenfinch (*Carduelis chloris;* O'Connor *et al.* 2021) and in the Eurasian siskin (*Spinus spinus*; Drews & Westerdahl, 2019).

MHC genes/alleles are expressed at different degrees

Interestingly, classical MHC-I genes vary in degree of expression (*i.e.,* difference in total expression between MHC genes/alleles within an individual). In the chicken, the degree of expression differs between the two gene copies for MHC-I (Kaufman *et al.* 1999) and only one gene copy, the "major" gene, is highly expressed, whereas the other gene copy, the "minor" gene, is weakly expressed (Kaufman *et al.* 1995). In the red knot (*C. canatus;* Charadriiformes), five out of 11 MHC-I alleles were highly expressed (Buehler *et al.* 2013). So far, MHC-I expression in passerines has mainly been studied by amplicon sequencing and the relative expression of the different alleles has been measured as the read depth of the transcribed cDNA alleles. In two species belonging to the *Passer* genus, the house sparrow (*P. domesticus*) and the tree sparrow (*P. montanus*), one or two classical MHC-I alleles per individuals were highly expressed (out of a maximum of four classical expressed alleles) while the remaining alleles were lowly expressed (Drews *et al.* 2017). This expression pattern contrasts with the results found in the siskins (*S. spinus*) for which on average three out of seven classical alleles are highly expressed with a maximum of five out five classical alleles found highly expressed (Drews & Westerdahl, 2019).

MHC expression varies between tissues

Few studies have reported differences in degree of MHC-I expression between different tissues. In the chicken only the "major gene" has been found highly expressed in erythrocytes and spleen cells (Kaufman *et al.* 1995). In the mallard duck, one MHC-I gene was highly expressed and the expression of this gene was higher in spleen and liver than in the heart, kidney and testis (Moon *et al.* 2005). In the quail, the two classical MHC-I genes have been found highly and ubiquitously expressed (Shiina *et al.* 2004) whereas the two other MHC-I gene have been found expressed in a tissue-specific manner. In the crested ibis, one out of the four expressed MHC-I genes were highly expressed in nine different tissues, whereas the other three MHC-I genes were highly expressed in four or five tissues each (Chen *et al.* 2015).

When I started my PhD studies there were very limited information available on the genomic organization of the MHC in birds of the order Passeriformes. This information is essential to further investigate and understand why passerines have such high MHC diversity. Moreover, very little is known about the expression of the different MHC genes in passerines and even less is known to what extent the MHC expression pattern varies temporally and whether the MHC expression pattern changes between tissues.

Aims of the thesis

The aims of this thesis are to gain a deeper understanding of the high MHC diversity found in passerine birds and relate MHC diversity to functional relevance with a main focus on a particular songbird, the great reed warbler. My aim is to understand how MHC diversity has expanded in passerines, how to best estimate MHC diversity in passerines when using sequencing methods providing different levels of resolution, and finally to investigate temporal and transgenerational variation in MHC-I gene expression in blood in great reed warblers.

Chapter I aims to characterize the MHC region and the MHC multigene family in the great reed warbler genome, but also in three other passerines. The inferred duplication history of MHC genes in the great reed warbler and the phylogenetic reconstruction of full-length MHC genes in the four passerine species provided new insights into the evolutionary history of the MHC multigene family in the order Passeriformes. In addition, the MHC organization was compared between the four passerine species to describe structural differences and genomic rearrangements in Passeriformes compared with other avian orders.

Chapter II aims to verify the MHC diversity and the haplotype resolution described in the great reed warbler long-read genome assembly, which is based on a single individual. To do so, short-read amplicon sequencing data, from both the genome individual and its parents, were used and then compared with the full (Falcon-2017) and haploid (Purge Haplotigs) long-read genome assemblies. Both quantitative and qualitative estimates of MHC diversity were compared between the genotyping methods to assess MHC diversity with different resolution.

Chapter III aims to investigate temporal (*i.e.,* between year) variation in expressed MHC diversity and also the relative expression of each MHC-I allele within great reed warbler individuals. Both life stages and age-related changes in MHC-I relative expression were analysed.

Chapter IV aims to assess whether the relative MHC-I expression is inherited in great reed warblers. Family data were used to separate MHC-I alleles into parental haplotypes and the relative MHC-I expression at the haplotype level was determined in parents and offspring.

Methodology

Study species and study area

The great reed warbler (*Acrocephalus arundinaceus*) is a long-distance migratory songbird (Aves: Passeriformes, Figure 7). The breeding area of this species range from western Europe to western Asia and the wintering area is in sub-Saharan Africa (Kolecek *et al.* 2016). It is a medium-sized songbird that weight about 30 g and has an average life-span of 2.5 years (Bench *et al.* 1998). It has a facultatively polygynous social mating system (Bensch, 1996) and a male can bond with several females simultaneously and have several broods in the same breeding season (May to late July). At each breeding event, one to six fledglings can be raised (Bensch *et al.* 1998).



Figure 7. Great reed warbler at two different life-stages at lake Kvismaren: adult individual during the breeding season (left picture) and five days old nestlings (right picture). Pictures taken by David Gómez Blanco.

Our study population is located at lake Kvismaren (59°10'N, 15°25'S) in southern Central Sweden. The study area is a mosaic of small lakes within a 6 km^2 area

(Gómez-Blanco, 2023) that provide the optimal wetland habitat with a mosaic of reed and water providing reed edges where great reed warblers like to breed. The great reed warbler is particularly relevant as study species because it has high levels of natal and breeding-area philopatry (Bensch & Hasselquist, 1991; Hansson *et al.* 2002a; Mátrai *et al.* 2012). In our study population, 60-80% of the surviving juveniles of both sexes come back on their birth sites the following years (Hansson *et al.* 2002b; Hansson *et al*. 2003) and up to 90% of the surviving adults come back to breed in the same or neighbouring locations (~3 kms perimeter). Therefore, individuals can be followed-up between years when they return to the study area during the breeding season.

Our study population has been monitored for over 40 years and information such as arrivals date, male territory, song repertoire, reproductive success, dispersal pattern and morphological measurements have been collected (Bensch & Hasselquist 1991; Hasselquist *et al.* 1996; Bensch *et al.* 1998; Hasselquist, 1998; Hansson *et al.* 2002b; Tarka *et al.* 2010; Tarka *et al.* 2015). In addition, blood samples have been collected to perform DNA analysis such as for example molecular sexing, microsatellite analysis, detection of blood parasites, MHC genotyping and telomere measurements (Hasselquist *et al.* 1996; Westerdahl *et al.* 2000; Westerdahl *et al.* 2004b; Bensch *et al.* 2007; Asghar *et al.* 2015; Roved *et al.* 2018; Gómez-Blanco, 2023). Since 2015, blood samples have been also collected to perform RNA analysis to investigate gene expression.

The great reed warbler genome assembly and the characterized MHC region

In this thesis, I had the opportunity to use the characterized MHC region in the genome assembly available from a single female great reed warbler individual. The *de novo* genome assembly has been performed using a combination of long-read sequencing (PacBio), linked-read and short-read sequencing (Illumina) as well as optical mapping (Bionano) as reported in Sigeman *et al.* (2021). The assembly of all DNA sequencing data has been performed with the FALCON assembler (Chin *et al.* 2016). It resulted in 3,013 scaffolds (*i.e.,* long stretches of different length of DNA sequences) for a total size of 1.2 Gb and an N50 of 21.4 Mb (half of the assembly is covered with scaffolds that are ≥ 21.4 Mb long). To produce an improved version of the haploid genome assembly, the Purge Haplotigs pipeline has been used (Roach *et al.* 2018) to remove sequences that were possibly alternative haplotypes of another scaffold. The MHC region in the great reed warbler has been characterized by performing BLAST searches with MHC-associated genes annotated in the chicken and in the crested ibis (more details in Chapter I: Westerdahl *et al.* 2022). Briefly, full-length MHC alleles have been identified by

blasting each exon of MHC-I and MHC-IIB cDNA Sanger sequences from the same female individual for which the genome assembly is available (Westerdahl *et al.* 1999; Westerdahl *et al.* 2000).

Assessing MHC diversity and MHC-I expression

Most avian MHC studies nowadays use high-throughput sequencing (HTS) of short amplicons amplified by Polymerase Chain Reaction (PCR) (*i.e.,* simultaneous amplification of all MHC alleles by PCR). Such methods efficiently produce a great amount of amplicons from multiple loci being sequenced in parallel from many individuals (O'Connor *et al.* 2016; Drews *et al.* 2017; Biedrzycka *et al.* 2017; O'Connor *et al.* 2019; Roved *et al.* 2022).

In this thesis, I assessed MHC diversity and MHC-I expression using genomic DNA (gDNA) and RNA extracted from blood samples coupled with HTS. Note that MHC-I molecules are expressed on all nucleated cells and are therefore likely to be present on erythrocytes in birds (as birds, unlike for example mammals, have nucleated erythrocytes). Hence, MHC expression can be obtained by sampling a small amount of blood and does not require that birds are sacrificed.

To measure MHC diversity, I amplified fragments of MHC-I exon 3 (Figure 8A) and MHC-IIB exon 2 from gDNA. As mentioned, these exons are the most variable MHC gene regions. The MHC diversity is the total number of different MHC alleles found in an individual. To measure MHC-I expression, I used mRNA that was reversely transcribed to complementary DNA (cDNA). As described above for MHC diversity, I amplified fragments of MHC-I exon 3 (Figure 8B) from cDNA. Investigating MHC-I expression using HTS will give two informative measures for each individual: expressed MHC-I diversity (*i.e.*, total number of alleles expressed) and relative MHC-I expression (the proportion of reads obtained from each expressed MHC-I allele).

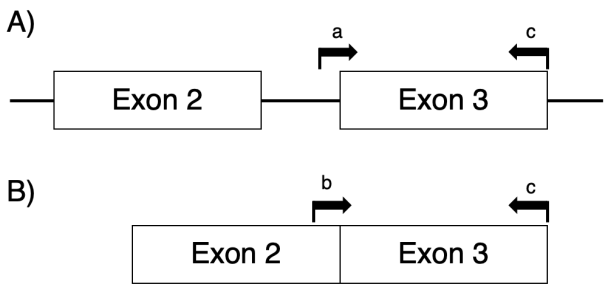


Figure 8. Targeted MHC-I exon 3 region to be PCR amplified in gDNA and in cDNA.
Primers pairs are indicated with arrows. A) In gDNA, both exons and introns are present. The forward primer a and reverse primer c were used to amplify a fragment of MHC-I exon 3. B) In cDNA, only exons are kept. The forward primer b and reverse primer c were used to amplify a fragment of MHC-I exon 3.

Although amplicon HTS gives results with high read-depth using this approach also come with some limitations such as possible primer-biased amplification of some alleles, short fragment size (up to 300 bp) and generation of artefactual alleles (Babik, 2010, O'Connor *et al.* 2019). Therefore, it is good to keep in mind that MHC genotyping using HTS remains an estimate of the functional diversity in an individual.

Results and Discussion

Chapter I. The genomic architecture of the passerine MHC region: High repeat content and contrasting evolutionary histories of single copy and tandemly duplicated MHC genes

The MHC in most Galliformes species comprises a core MHC region that is compact, conserved and with rather few MHC genes (Kaufman *et al.* 1999; Wang *et al.* 2012). However, the architecture of the MHC region, the size, the gene content and gene synteny differ slightly in other avian orders (Chen *et al.* 2015; Tsuji *et al.* 2017). In passerines (Passeriformes), the MHC region has been partly characterized though yet with limited resolution (Balakrishnan *et al.* 2010; Ekblom *et al.* 2011). Also, there is limited knowledge from genome assemblies on the evolutionary history of MHC genes that have led to the high diversity among Passeriformes such as in the great reed warbler, a passerine for which high MHC diversity has been described (Westerdahl *et al.* 1999; Westerdahl *et al.* 2004b; Roved *et al.* 2022). In the present study, the MHC region has been characterized in a single great reed warbler individual and in three other passerine species where long-read genomes were available, the jackdaw *Coloeus manedula*, the hooded crow *Corvus cornix* and the zebra finch.

In total, 15 MHC-I, one MHC-IIA and 56 MHC-IIB full-length genes in open reading frame, *i.e.,* putatively functional genes, were found in the great reed warbler individual. The MHC-I and MHC-IIB genes were found in two genomic arrangements: as single gene copies separated by long intergenic distances and placed among non-MHC genes or as tandemly duplicated gene copies with short intergenic distances (Figure 9).

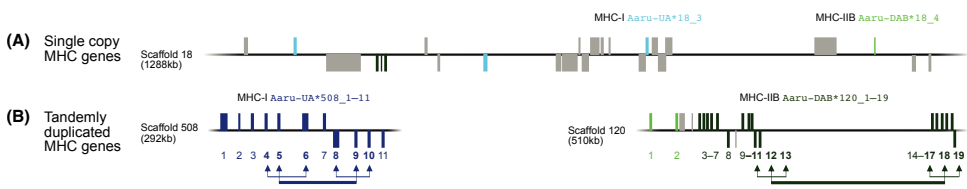


Figure 9. The great reed warbler MHC genes are found in two genomic arrangements.
A) Single and B) tandemly duplicated gene copies are shown among three MHC scaffolds in the great reed warbler drawn to scale: Aaru_Scaffold 18 (1,288 kb), Aaru_Scaffold 508 (292 kb) and Aaru_Scaffold 120 (510 kb). Aaru_Scaffold 18 (holds the core MHC region) has three single MHC-I genes (light blue boxes) and a one single MHC-IIB (light green boxes) gene. Aaru_Scaffold 508 holds 11 MHC-I tandemly duplicated genes (dark blue boxes) and Aaru_Scaffold 120 holds 17 MHC-IIB tandemly duplicated genes (dark green boxes). Trios of MHC-I and MHC-IIB genes have been duplicated and are shown with arrows on scaffold Aaru_Scaffold 508 and Aaru_Scaffold 120.

The high MHC diversity in the great reed warbler most likely arose through multiple and repeated gene duplication events of either single or many MHC genes with trios of tandemly organized MHC-I and MHC-IIB genes that have been duplicated as blocks (Figure 9). Most MHC-I and MHC-IIB genes were found to cluster within species (monophyly) meaning that the MHC genes have duplicated after speciation events. However, one single MHC-I gene was found to be shared within clades for the jackdaw and the hooded crow and for the great reed warbler and the zebra finch (Figure 10A). A single MHC-IIB gene was found to be a putatively orthologous gene among the species studied as it formed a separate cluster shared among three of the four Passeriformes species, whereas all other MHC-IIB genes are strictly monophyletic within species (Figure 10B).

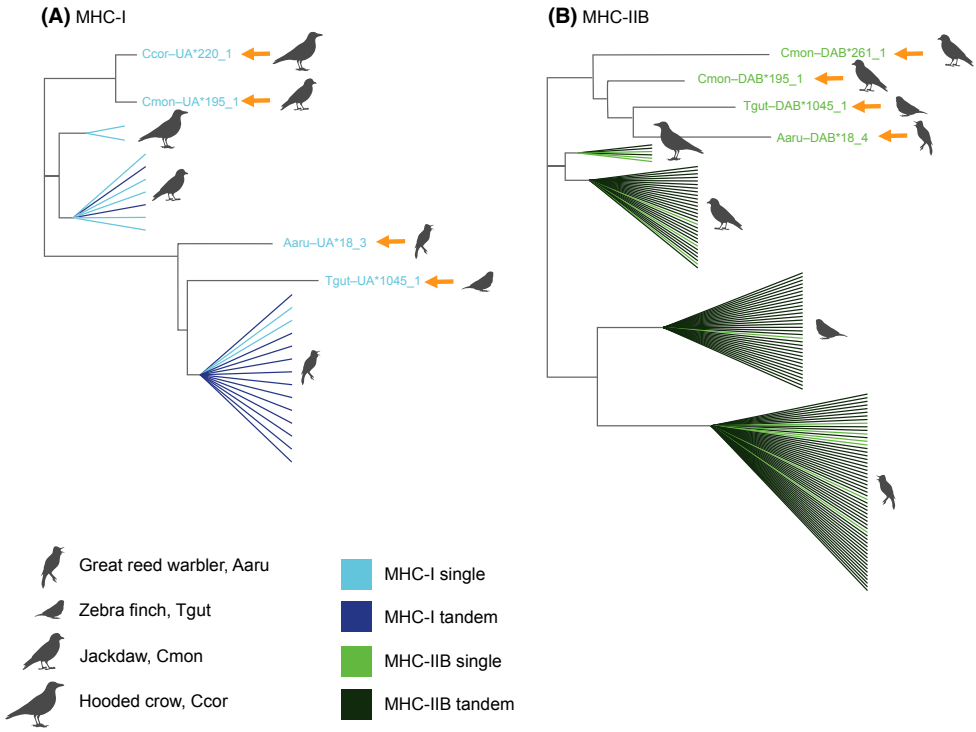


Figure 10. Phylogenetic reconstructions of full-length MHC-I (A) and MHC-IIB (B) single and duplicated genes in open-reading frame in four passerine species.
Most MHC genes are monophyletic within species (collapsed in the trees) but one single copy of MHC-I and MHC-IIB are are found to cluster and are putative orthologs. Both putative MHC-I and MHC-IIB orthologs, indicated by orange arrows, are found in an order that is shared among species in the core MHC region, which is situated on Aaru_Scaffold 18 in the great reed warbler. (A) Phylogenetic reconstruction of MHC-I genes, called in a species-specific manner: great reed warbler (Aaru-UA*), zebra finch (Tgut-UA*), jackdaw (Cmon-UA*) and hooded crow (Ccor-UA*). Single MHC-I copy genes are indicated in light blue and tandemly duplicated MHC-I genes are indicated in dark blue. (B) Phylogenetic reconstruction of MHC-IIB genes, called in a species-specific manner: in the great reed warbler (Aaru-DAB*), zebra finch (Tgut-DAB*), jackdaw (Cmon-DAB*)and hooded crow (Ccor-DAB*). Single MHC-IIB copy genes are indicated in light green and tandemly duplicated MHC-IIB genes are indicated in dark green.

The close association between TAP and classical MHC-I gene seems to have been lost in Passeriformes (Figure 11). Instead, putatively functional MHC-I genes are found on other scaffolds, noteworthily as tandemly duplicated copies in the great reed warbler (Figure 11). The number of MHC-I and MHC-IIB genes varied greatly between the four Passeriformes species with the highest diversity found in the great reed warbler, while only a single MHC-I gene was found in the zebra finch. A single MHC-IIA gene was found close to tandemly duplicated MHC-IIB genes in Passeriformes whereas in Galliformes, the single MHC-IIA gene is located away from the MHC-IIB genes and in Pelecaniformes and in Ciconiiformes, MHC-IIA and MHC-IIB are found in "dyads".

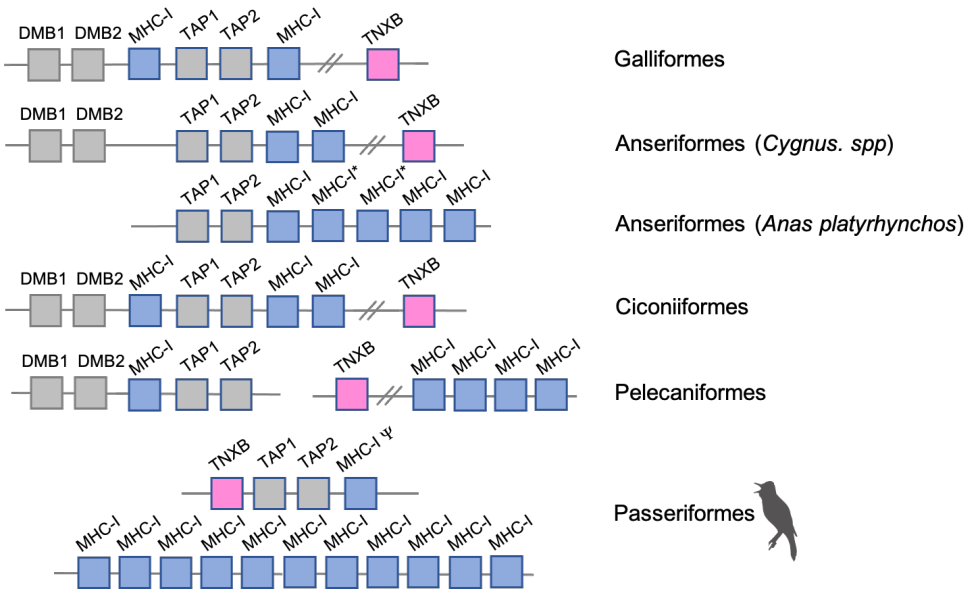


Figure 11. Simplified structural differences in the MHC-I genomic region between five avian orders.

The gene order and organization of the MHC-I (blue) genes differ among the bird orders. The passerines, with data from the great reed warbler, seems to have lost their functional MHC-I genes from this region. The gene order and organization of DMB1, DMB2, TAP1 and TAP2 (grey colour) in this region are shared among birds of the orders Galliformes, Ciconiiformes and Pelecaniformes; whereas the gene TNXB (pink) has been rearranged and is located next to TAP1 and TAP2 genes in Passeriformes. *Three MHC-I genes are not expressed in Anseriformes. Pseudogenes are indicated with ψ. Simplified comparative genomic maps, not drawn to scale, double bars indicates where other genes have been found without detailed information. For more information, refer to the following: Kaufman *et al.* 1999; Chaves *et al.* 2009; Wang *et al.* 2012; Ye *et al.* 2012; Eimes *et al.* 2013 (Galliformes), Karawita *et al.* 2023 (Anseriformes, *Cygnus spp.*), Moon *et al.* 2005 (Anseriformes, *Anas platyrhynchos*), Tsuji *et al.* 2017 (Ciconiiformes), Chen *et al.* 2015 (Pelacaniformes) and Chapter I (Passeriformes).

Chapter II. Improved haplotype resolution of highly duplicated MHC genes in a long-read genome assembly using MiSeq amplicons

The MHC region in the great reed warbler has been characterized in a single individual based on a long-read genome assembly (Sigeman *et al.* 2021; Chapter I: Westerdahl *et al.* 2022). The MHC-I and -IIB genes were highly duplicated and mainly monophyletic, which means that the MHC alleles (*i.e.,* full-length genes, an open reading frame) were often more similar within than between species (Chapter I: Westerdahl *et al.* 2022). High MHC sequence dissimilarity between genes make it difficult to produce a haploid representation of the MHC region. Therefore, we used amplicon short-read sequencing to amplify MHC-I exon 3 and MHC-IIB exon 2 sequences in the genome individual and its parents and mapped these amplicons against two different genome assemblies: "Falcon-2017" and "Purge Haplotigs". The Falcon-2017 assembly initially separate haplotypes but mostly represent MHC diversity under diploidy. An additional pipeline has been used to produce the Purge Haplotigs assembly which further separate haplotypes wrongly assigned in the Falcon-2017 assembly to get a better haploid representation. We were able to investigate i) the MHC diversity in Falcon-2017 and ii) the segregation of full-length alleles in Purge Haplotigs based on both amplicon short-read sequencing and long-read sequencing.

The Falcon-2017 assembly and short-read amplicon sequencing gave similar estimates of MHC-I and MHC-IIB diversity under diploidy, although for different reasons. Firstly, because the techniques have different resolution (Figure 12). In fact, only a small proportion of the amplicon alleles (55% of the MHC-I alleles and 61% of the MHC-IIB alleles) are necessary to detect most of the MHC-I and MHC-IIB diversity found in the Falcon-2017 genome assembly (Figure 12). This is because the amplicon alleles, MHC-I exon 3 and MHC-IIB exon 2 sequences, only represent a small portion of a full-length annotated gene, and this particular gene region can be shared between several MHC gene copies in the genome assembly. Secondly, a considerable proportion of the amplicon alleles (45% of the MHC-I alleles and 39% of the MHC-IIB alleles) are only found as amplicon alleles and are not found in the genome assembly. This observation suggests either that the genome assembly is incomplete and/or that a certain proportion of these amplicon alleles represents fragments of genes and not full-length genes.

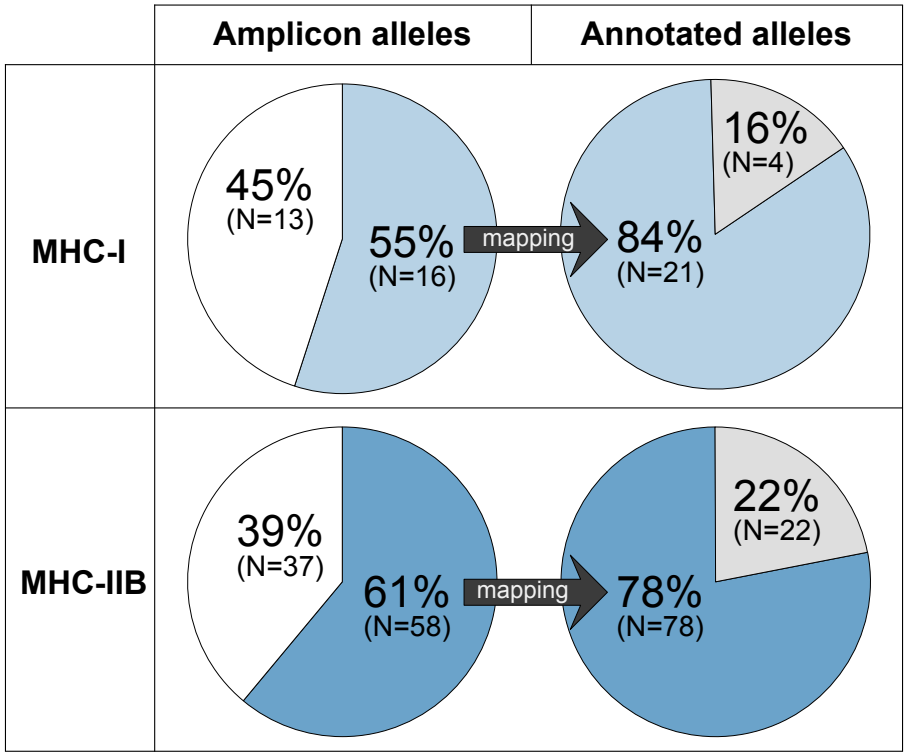


Figure 12. A moderate proportion of the MHC-I and MHC-IIB amplicon alleles (HTS) detects a considerable proportion of the annotated alleles (Falcon-2017 assembly) (Chapter II).
In total, there were 29 MHC-I and 95 MHC-IIB amplicon alleles and 25 MHC-I and 100 MHC-IIB full-length annotated alleles in Falcon-2017. Amplicon alleles mapping to annotated alleles are indicated in blue (MHC-I: upper panel, light blue; MHC-IIB: lower panel, dark blue), unmapped amplicon alleles are indicated in white, and annotated alleles with no matching amplicon alleles are indicated in grey.

The Purge Haplotigs assembly represents a haploid version of the genome and using amplicon sequencing data from the genome individual and her parents I was able to identify the parental origin of each annotated allele in every scaffold holding tandemly duplicated MHC-I and MHC-IIB genes (Figure 13).

For MHC-I, scaffold 508 (Aaru-UA*508) contains seven MHC-I genes of paternal origin and this scaffold is likely to represent a paternal haplotype. For MHC-IIB, I found that scaffold 357 (Acar-DAB*357) was putatively of paternal origin, scaffold 554 and 120 (Acar-DAB*554 and *120) was fully or partly of maternal origin, while scaffolds 45 and 301 (Acar-DAB*45 and *301) were of mixed parental origin (Figure 13). Of the MHC-IIB scaffolds, two were potential complementary haplotypes as they had the same number of genes and were from different parental origin with one maternal scaffold (Acar-DAB*554) and one paternal scaffold (Acar-DAB*357). To verify this, I compared the sequence similarity between genes among all scaffolds with tandemly duplicated MHC-IIB genes. I found that the

mean nucleotide p-distance between genes in the two scaffolds Aaru-DAB*554 and Aaru-DAB*357 was the lowest compared to other scaffold comparisons. Therefore, the Purge Haplotigs genome assembly (haploid representation) slightly overestimated the MHC-IIB diversity because it has not considered Aaru-DAB*554 and Aaru-DAB*357 as complementary haplotypes but instead as two different scaffolds (*i.e.,* two different genomic regions).

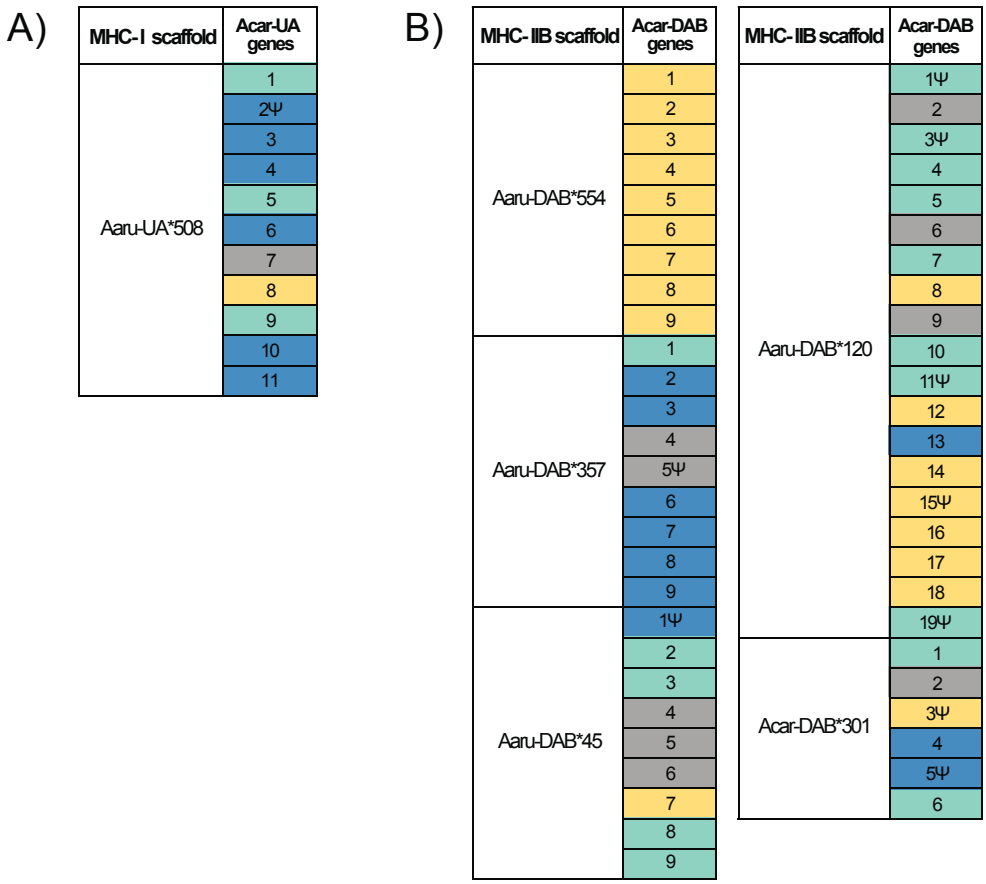


Figure 13. Inferred parental origin for annotated MHC genes, in MHC-I and MHC-IIB scaffolds holding tandemly duplicated genes in the Purge Haplotigs assembly, using the MHC amplicon allele information.

MHC scaffolds are indicated as "Aaru-UA*" for MHC-I genes (A) and "Aaru-DAB*" for MHC-IIB genes (B). Gene copies are indicated as "Acar-UA" and "Acar-DAB" and are named after their position on annotated primary scaffolds in the Purge Haplotigs assembly. Amplicon alleles were separated into three categories based on their inheritance in the focal individual: paternal alleles (blue), maternal alleles (yellow) and unresolved alleles (turquoise). Non-functional genes are indicated with the symbol ψ. Annotated MHC genes with no matching or no assigned amplicon alleles are in grey.

Chapter III. Temporal variation and age-associated changes in MHC-I gene expression in great reed warblers.

The MHC diversity (*i.e.,* total number of MHC genes or alleles per individual) has expanded in avian species belonging to Passeriformes (O'Connor *et al.* 2016; Minias *et al.* 2018; Chapter I: Westerdahl *et al.* 2022). A recent study has shown that the expressed MHC-I diversity also varies between species (O'Connor & Westerdahl, 2021). In addition, MHC alleles vary in how much they are expressed, *i.e.,* how much each allele is expressed relative to the other alleles (relative degree of expression*,* measured as the proportion of reads represented by an allele compared to the total number of reads using amplicon sequencing) (Drews *et al.* 2017; Drews & Westerdahl, 2019). However, little is known about how stable the MHC-I relative gene expression remains over time. We set out to investigate temporal variation in MHC-I expression (in the blood transcriptome) in great reed warbler, a migratory songbird that shows high level of natal and breeding-area philopatry every year. Therefore, I could study life-stage and age-related changes in MHC-I expression.

The great reed warbler express 71% of the MHC-I alleles found in genomic DNA (gDNA), which corresponds to an average expression of 14.3 alleles out of 20.1 gDNA alleles. The relative expression of each MHC-I allele varied within individuals and alleles could be defined as being highly, moderately or lowly expressed. A general pattern across all individuals (Figure 14) was that only one or two MHC-I alleles were highly expressed, zero to seven MHC-I were moderately expressed whereas the vast majority (5-15) of the alleles were lowly expressed.

The relative expression of each MHC-I allele was highly consistent between the sampling events of each individual (Figure 14). I also looked at gene expression using ranking, the most highly expressed allele was given the rank 1 etc… I observed a slight shuffle of the ranking among moderately and lowly expressed alleles in our analyses (Figure 15). The relative expression based on ranking was lower between life stages, *i.e.,* from nestling stage (8-10 days old at time t_0) to recruited adults (≥1 year old at time t_1) than within life-stage (from time t_1 to time t_2). Thus, individuals were more similar between years at adulthood (Figure 15).

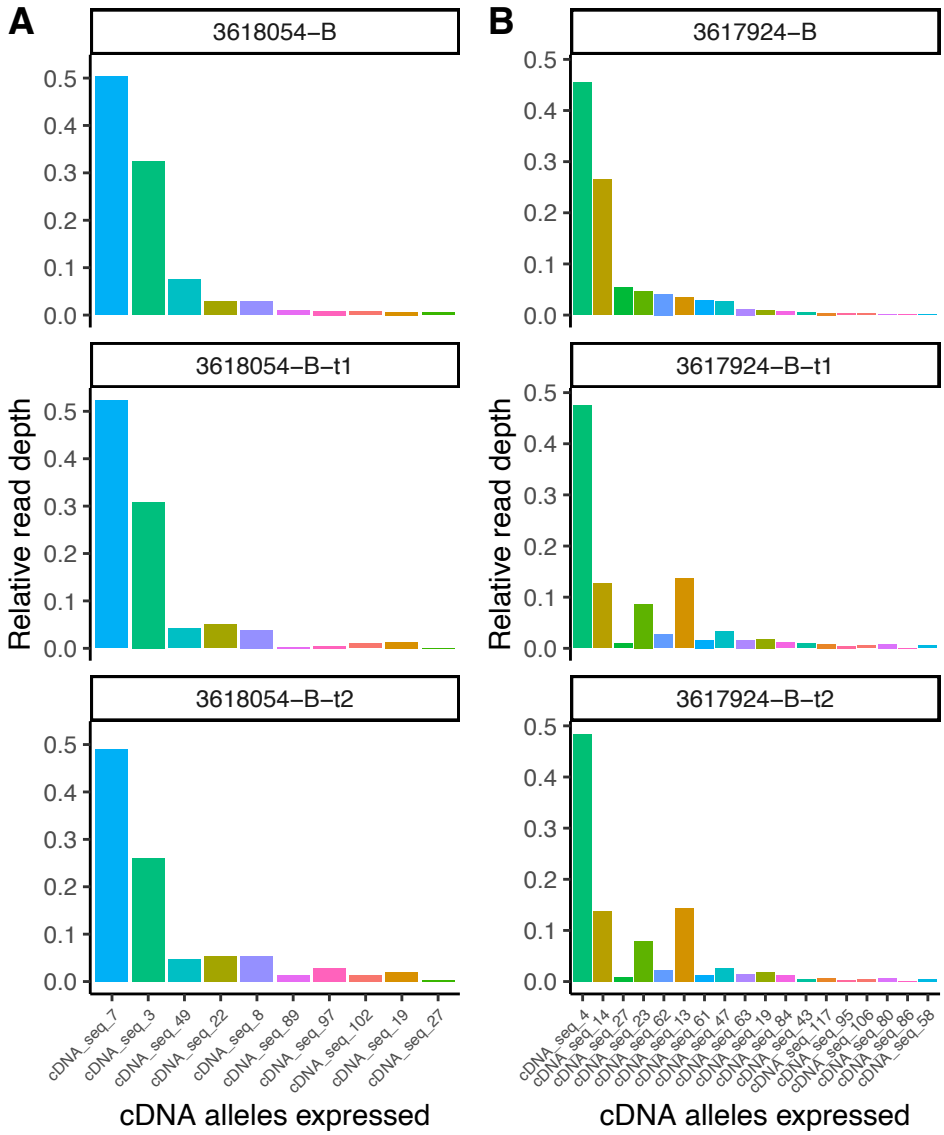


Figure 14. Temporal variation of relative expression of MHC-I alleles (cDNA) in two representative great reed warbler individuals.

Examples of two individuals (out of in total nine individuals investigated) that were caught as nestlings (8-10 days old) at the first sampling event (time t_0, top row) and then recaptured as adult the following years, both at the second sampling event (time t_1, middle row) and at the third sampling event (time t_2, bottow row). Each panel, from right to left, represents the relative MHC-I expression patterns for each of the two individuals, (A) individual "3618054", (B) individual "3617924". Each bar represents a single expressed allele (cDNA_seq_x) and all alleles are ordered according to the relative expression at time point time t_0. Expressed alleles are uniquely colored in each individual and colors are not shared between individuals.

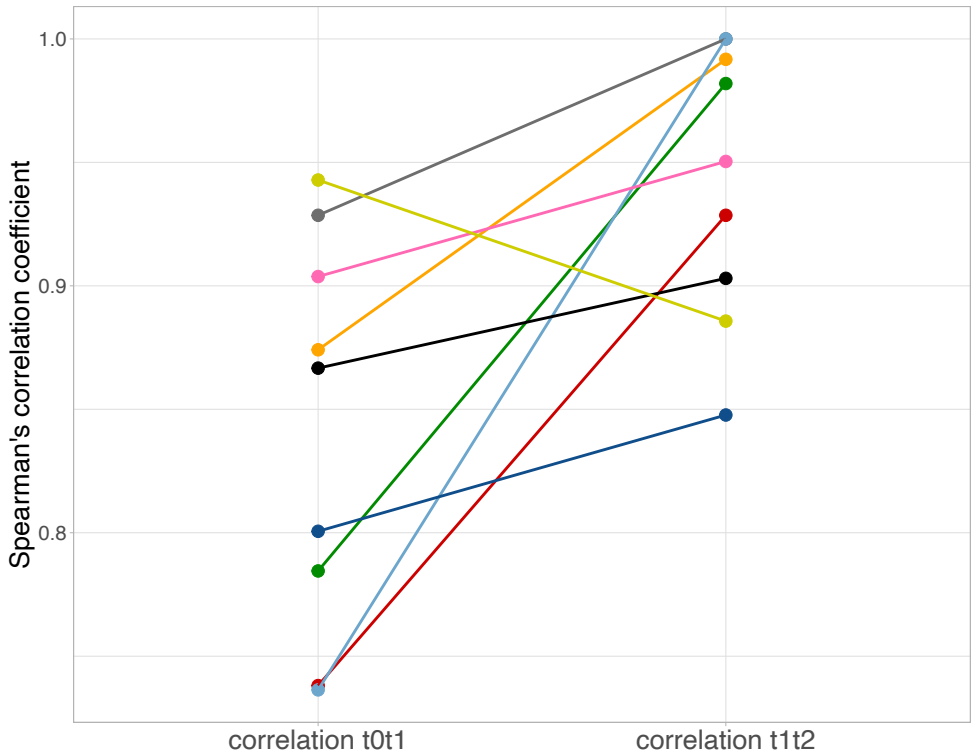


Figure 15. Spearman's ρ rank correlation for changes between life stages (correlation between time t_0 to time t_1) and between ages at adulthood (correlation between time t_1 to time t_2) in nine great reed warblers (Chapter III).

Each color represents distinct individuals for which correlation coefficients were plotted between life stages and between ages at adulthood. Individuals were more similar between years at adulthood than between the nestling and the 1-year of age life stages (Paired samples Wilcoxon Test, p = 0.027).

Chapter IV. Inheritance of MHC-I gene expression in great reed warblers

MHC genes in passerines are generally inherited as linked haplotypes (Westerdahl *et al.* 2004b, Karlsson & Westerdahl, 2013; Roved *et al.* 2022). Family data offers a great opportunity to look at gene expression at the haplotype level and investigate to what extent MHC-I gene expression is inherited. Here, I have selected 17 great reed warbler families and based on the segregation of alleles in offspring, I could group alleles into resolved maternal and paternal haplotypes (a minor proportion of alleles were unresolved; Figure 16). I have investigated whether MHC-I gene expression (*i.e.,* relative MHC-I expression) in haplotypes was inherited, by comparing the relative expression of alleles on a haplotype in parent and offspring. Moreover, I could test whether the genetic background (*i.e.,* the complementary haplotype) did affect the MHC-I gene expression by comparing expression of alleles on one common haplotype in siblings that had different alternative haplotypes.

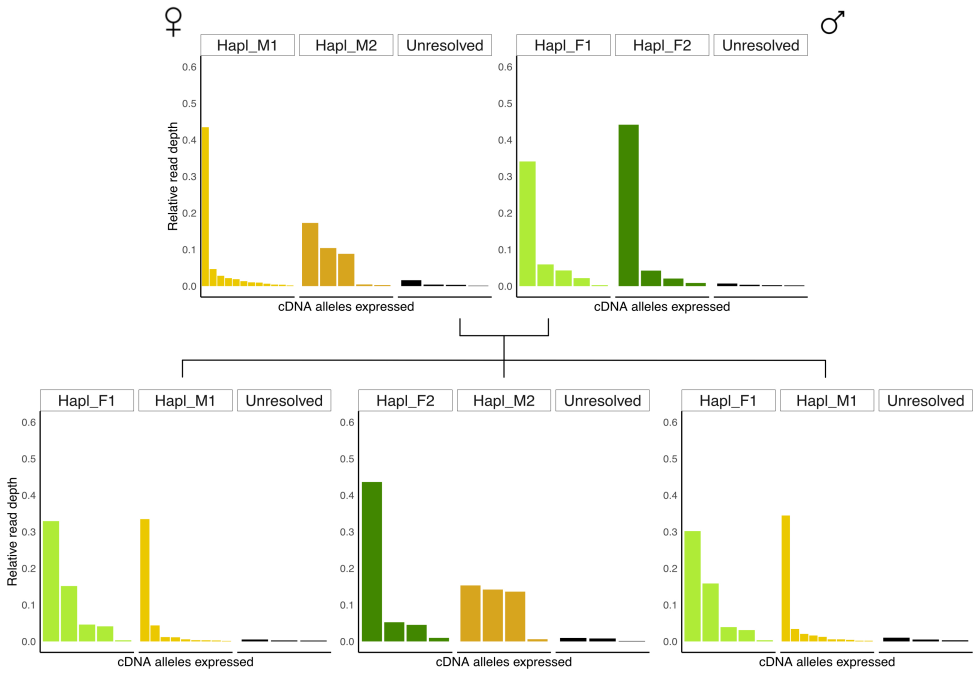


Figure 16. Relative MHC-I expression by haplotypes in one representative great reed warbler family (Chapter IV).
Great reed warbler parents on the top row and three offspring on the bottom row. Every parent has two resolved haplotypes: Hapl_M1 (Mother 1, yellow) and Hapl_M2 (Mother 2, orange) in mothers, and Hapl_F1 (Father 1, light green) and Hapl_F2 (Father 2, dark green) in fathers. Alleles that could not be assigned to resolved parental haplotypes were considered unresolved (black). Each allele is represented by a single bar.

The relative MHC-I gene expression of both maternal and paternal haplotypes in offspring resemble the MHC-I relative expression in their parents (Figure 17). However, I did observe a stronger relationship for maternal than paternal haplotypes. In other words, MHC-I gene expression in offspring tend to resemble more the MHC-I gene expression of their mothers than their fathers (Figure 17). The latter result was similarly found with offspring at the nestling stage (8-10 days old) and with offspring at the adult stage (≥1 year old). This pattern suggests a maternal effect on MHC-I expression that could be the result of genomic partial imprinting responsible for differential expression between maternal and paternal alleles.

Offspring can inherit either of two different haplotypes from each parent. I have selected siblings sharing at least one parental haplotype and then either sharing the same or a different complementary haplotype from the second parent (*i.e.,* same or different genetic background). Comparison of siblings with same or different parental haplotype combination within families enable us to look at the effect of different genetic background on the MHC-I gene expression of the shared haplotypes. Neither maternal nor paternal haplotypes were affected by the MHC-I genetic background (Figure 17).

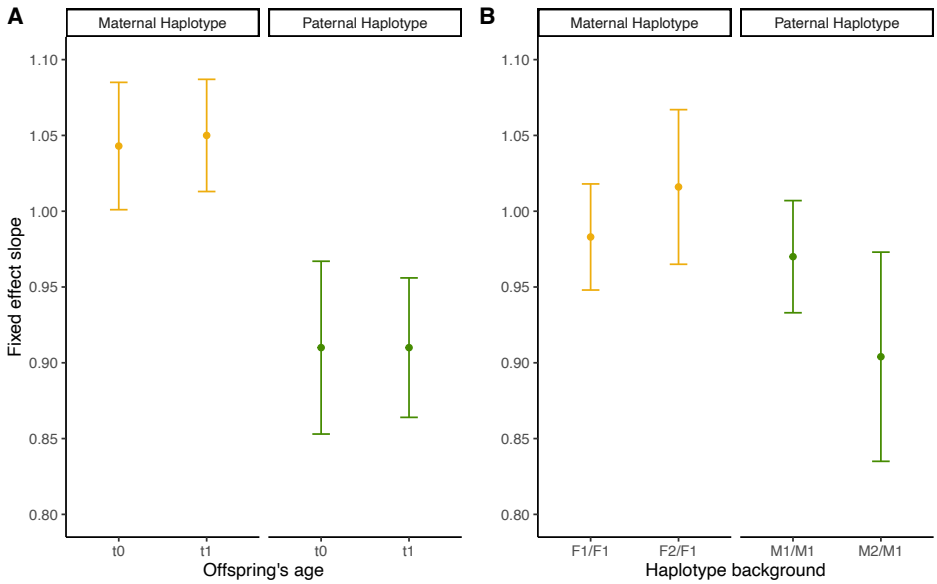


Figure 17. Fixed effect slopes for inheritance (A) and effect of MHC-I haplotype background (B) of relative MHC-I expression in great reed warbler offspring (Chapter IV).
A) Fixed effect slopes (± standard errors) report the strength of inheritance of relative MHC-I expression in maternal (yellow) and paternal (green) haplotypes in hatched (t_0) and recruited offspring (t_1). B) Fixed effect slopes (± standard errors) report the strength of the effect of haplotype background on the relative MHC-I expression in maternal (yellow) and paternal (green) haplotypes in hatched offspring. Maternal haplotypes were tested when offspring (pair of siblings in each family) had identical (F1/F1) or different (F2/F1) paternal haplotype background. Paternal haplotypes were tested when offspring (pair of siblings in each family) had identical (M1/M1) or different paternal haplotype background (M2/M1).

Conclusion

This thesis aimed to bring some insights in i) understanding what plausible mechanisms have generated such high MHC diversity in Passeriformes, ii) evaluating sequencing methods used to best estimate MHC diversity and iii) relating genomic MHC diversity to expressed MHC diversity in blood and investigate variation in gene expression in time and across generations.

The MHC region have gone through genomic rearrangements and has considerably expanded in Passeriformes compared with other avian orders. For example, the tightly linked genes found in the core MHC in Galliformes are rather distant in Passeriformes as they have been found in large scaffolds away from the core region. The presence of tandemly duplicated MHC paralogs in several rather large scaffolds shows the extended feature of the MHC in Passeriformes and can possibly explain how the rapid expansion of the MHC diversity happened. Mechanisms involved in maintenance and loss of MHC diversity *e.g.,* ectopic recombination and the birth-and-death model of evolution, align with our finding in the four passerines and indicates that the MHC region is dynamic and can also explain why MHC diversity vary a lot between species, also within Passeriformes.

The use of short-read amplicon sequencing and family data has proven to be efficient and complementary to assess MHC-I and MHC-IIB diversity quantitatively (*i.e.,* total number alleles per individual) and to verify the haplotype segregation in genome assemblies. However, short-read amplicon sequencing lacks resolution due to the fragment length being sequenced (the amplicon) and the amplicons only partially represent the true functional MHC diversity. Therefore, I encourage studies using short-read amplicon sequencing for estimating MHC diversity to be aware of the limits and to take them into account in their interpretation.

The expressed MHC-I diversity as well as the relative MHC-I expression are highly stable throughout time in great reed warblers. Also, I found that the relative MHC-I expression is to a large extent inherited from both parents, with no indication of epistasis. I did observe a maternal effect, though. Taken together, this indicates that the relative MHC-I expression is set at an early life-stage, probably during the first week in hatched offspring or possibly even at the embryogenic stage. The stable gene expression suggests that regulation of MHC-I expression is most likely happening at all MHC-I loci simultaneously and not at a particular locus.

MHC-I relative expression varies between alleles and every individual has one to two highly expressed alleles. As such, they are maybe subject to stronger pathogen-mediated selection than the lowly expressed MHC-I alleles. Therefore, I recommend future studies to focus on investigating the most highly expressed MHC-I alleles as they might give relevant adaptive advantage.

Future directions and preliminary results

The expressed MHC diversity in songbirds has mainly been analysed in blood samples and it is currently to a large extent unknown to what degree the expressed MHC diversity in blood is representative of other tissues. Therefore, I present some preliminary results on MHC-I relative expression in four different tissues (blood, intestine, liver and lung) in two great reed warbler individuals (Figure 18). The methods used to estimate the MHC-I diversity is the same as presented in chapters III and IV.

As previously described for MHC-I expression in blood (Chapter III and IV), the MHC-I alleles are expressed at different levels (highly, moderately and lowly expressed), and the level of expression seem to differ a little between the different tissues (Figure 18). The most highly expressed alleles (five alleles in Bird 1 and three alleles in Bird 2) in the blood remained the most highly expressed alleles also in the three other tissues. However, the level of expression of these alleles differed slightly between tissues (Figure 18). In one individual (Bird 2, Figure 1B), the most highly expressed allele in blood (orange, cDNA_seq_004) is not the most highly expressed allele in any of the other tissues. However, in the other individual (Bird 1, Figure 1A), the most highly expressed allele in blood (dark green, cDNA_seq_040) remains the most highly expressed allele in all the other tissues. These preliminary results show limited variation in MHC-I relative expression between tissues in Bird 1 but rather substantial variation in Bird 2, and therefore it is difficult to draw any general conclusions considering gene expression in different tissues in the great reed warbler at present. Nevertheless, my preliminary findings emphasize the need to further investigate the role of the most highly expressed MHC-I alleles in the great reed warbler and in songbirds in general, as mentioned above, but also to investigate the function of the large repertoire of lowly expressed alleles.

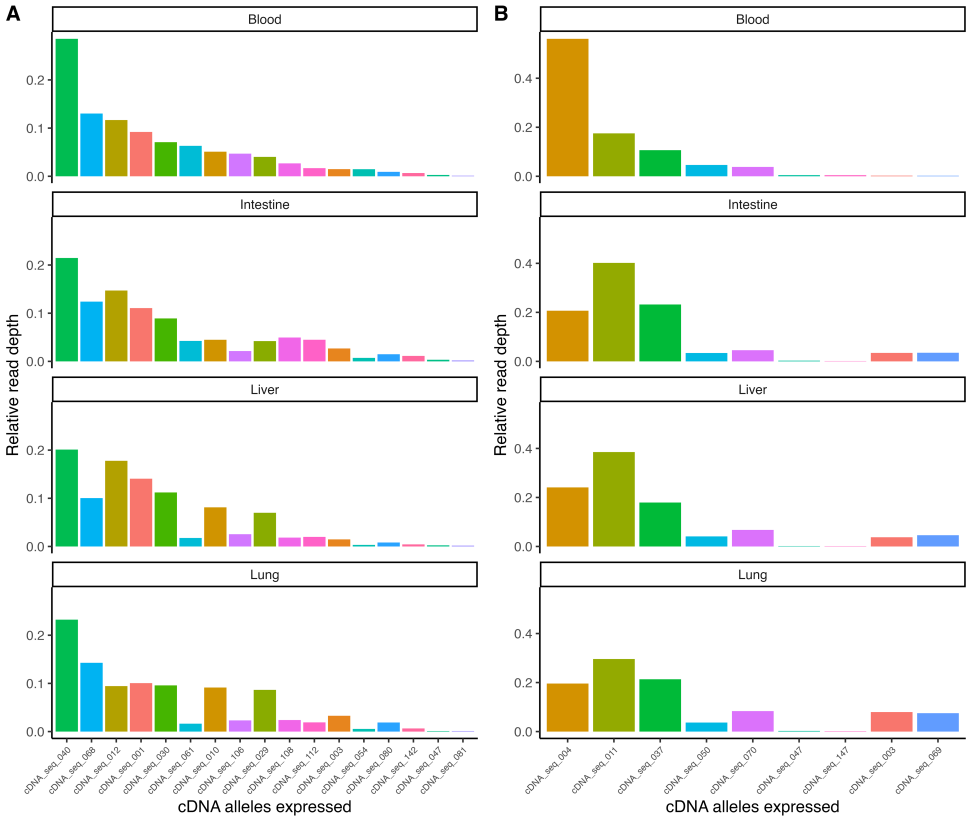


Figure 18. Relative expression of MHC-I alleles (cDNA) in four different tissues in two representative great reed warbler individuals.

Each panel, right (A) and left (B), represents the relative expression of MHC-I alleles for each individual. The sampled tissues were obtained from two adult birds kept in captivity as part of an experiment. The relative expression of MHC-I alleles for each tissue is represented from top to bottom: blood, intestine, liver and lung. Each bar represents a single expressed allele (cDNA_seq_x) within an individual and all alleles are ordered according to their relative expression found in blood. Expressed alleles are uniquely colored in each individual and colors are not shared between individuals.

References

Abbas A.K., A.H. Lichtman & Pillai S. (2020). Basic Immunology: Functions and disorders of the immune system. (6th Edition). Philadelphia, PA: Elsevier Saunders.

Allen R.L. & Hogan L. (2013). Non-Classical MHC Class I Molecules (MHC-Ib). eLS: (1-12), John Wiley & Sons, Ltd (Ed.). DOI: 10.1002/9780470015902.a0024246.

Asghar M., D. Hasselquist, B. Hansson, P. Zehtindjiev, H. Westerdahl & Bensch S. (2015). Hidden costs of infection: chronic malaria accelerates telomere degradation and senescence. *Science*, 347: 436-438. DOI: 10.1126/science.1261121.

Babik W. (2010). Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources*, 10(2): 237-251. DOI: 10.1111/j.1755-0998.2009.02788.x.

Balakrishnan C.N., R. Ekblom, M. Völker, H. Westerdahl, R. Godinez, H. Kotwiewicz, D.W. Burt, T. Graves, D.K. Griffin, W.C. Warren & Edwards S.V. (2010). Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biology*, 8: 29. DOI: 10.1186/1741-7007-8-29.

Bensch S., J. Waldenström, N. Jonzén, H. Westerdahl, B. Hansson, D. Sejberg & Hasselquist D. (2007). Temporal dynamics and diversity of avian malaria parasites in a single host species. *Journal of Animal Ecology,* 76(1): 112-22. DOI: 10.1111/j.1365-2656.2006.01176.x.

Bensch S, D. Hasselquist, B. Nielsen & Hansson B. (1998). Higher fitness for philopatric than for immigrant males in a semi-isolated population of great reed warblers *Evolution*, 52(3): 877–883. DOI: 10.1111/j.1558-5646.1998.tb03712.x.

Bensch, S. (1996). Female mating status and reproductive success in the great reed warbler: is there a potential cost of polygyny that requires compensation? *Journal of Animal Ecology*, 65: 283-296. DOI: 10.2307/5875

Bensch S., & Hasselquist D. (1991). Territory Infidelity in the Polygynous Great Reed Warbler *Acrocephalus arundinaceus*: The Effect of Variation in Territory Attractiveness. *Journal of Animal Ecology*, 60(3): 857–871. DOI: 10.2307/5418.

Biedrzycka A., E. O'Connor, A. Sebastian, M. Migalska, J. Radwan, T. Zając, W. Bielański, W. Solarz, A. Ćmiel & 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: 159. DOI: 10.1186/s12862-017-0997-9.

Bollmer, J. L., Dunn, P. O., Whittingham, L. A., & Wimpee, C. (2010). Extensive MHC class II B gene duplication in a passerine, the common Yellowthroat (*Geothlypis trichas*). *Journal of Heredity*, 101(4), 448-460. DOI: 10.1093/jhered/esq018.

Buehler D. M., Y.I. Verkuil, E.S. Tavares & Baker A. J. (2013). Characterization of MHC class I in a long-distance migrant shorebird suggests multiple transcribed genes and intergenic recombination. *Immunogenetics*, 65(3), 211-225. DOI: 10.1007/s00251-012-0669-2.

Burri R., N. Salamin, R.A. Studer, A. Roulin & Fumagalli L. (2010). Adaptive divergence of ancient gene duplicates in the avian MHC class II beta. *Molecular Biology and Evolution*, 27(10): 2360–2374. DOI: 10.1093/molbev/msq120.

Chaplin D.D. (2010). Overview of the immune response. *Journal of Allergy and Clinical Immunology*, 125(2), S3-23. DOI: 10.1016/j.jaci.2009.12.980.

Chaves L.D., S.B. Krueth & Reed K.M. (2009). Defining the turkey MHC: sequence and genes of the B locus. *The Journal of Immunology*, 183(10): 6530–6537. DOI: 10.4049/jimmunol.0901310.

Chen L.C., H. Lan, L. Sun, Y.L. Deng, K.Y. Tang & Wan Q.H. (2015). Genomic organization of the crested ibis MHC provides new insight into ancestral avian MHC structure. *Scientific Reports*, 5: 7963. DOI: 10.1038/srep07963.

Chin C-S., P. Peluso, F:J. Sedlazeck, M. Nattestad, G.T. Concepcion, A. Clum, C. Dunn, R. O'Malley, R. Figueroa-Balderas, A. Morales-Cruz, G.R. Cramer, M. Delledonne, C. Luo, J.R. Ecker, D. Cantu, D.R. Rank & Schatz M.C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, 13(12): 1050–1054. DOI: 10.1038/nmeth.4035.

Cooper M.D. & Alder M.N. (2006) The evolution of adaptive immune systems. *Cell*, 124(4): 815–822. DOI: 10.1016/j.cell.2006.02.001.

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

Drews A. & Westerdahl H. (2019). Not all birds have a single dominantly expressed MHC-I gene: Transcription suggests that siskins have many highly expressed MHC-I genes. *Scientific Reports*, 9(1):19506. DOI: 10.1038/s41598-019-55800-9.

Drews A., M. Strandh, L. Råberg & Westerdahl H. (2017). Expression and phylogenetic analyses reveal paralogous lineages of putatively classical and non-classical MHC-I genes in three sparrow species (Passer). *BMC Evolutionary Biology*, 17: 152 (2017). DOI: 10.1186/s12862-017-0970-7.

Eimes J.A., K.M. Reed, K.M. Mendoza, J.L. Bollmer, L.A. Whittingham, Z.W. Bateson & Dunn P.O. (2013). Greater prairie chickens have a compact MHC-B with a single class IA locus. *Immunogenetics*, 65(2): 133–144. DOI: 10.1007/s00251-012-0664-7.

Ekblom R., J. Stapley, A.D. Ball, T. Birkhead, T. Burke & Slate J. (2011). Genetic mapping of the major histocompatibility complex in the zebra finch (*Taeniopygia guttata*). *Immunogenetics*, 63(8): 523–530. DOI: 10.1007/s00251-011-0525-9.

Goebel J., M. Promerova, F. Bonadonna, K.D. McCoy, C. Serbielle, M. Strandh, G. Yannic, R. Burri & Fumagalli L. (2017). 100 million years of multigene family evolution: origin and evolution of the avian MHC class IIB. *BMC Genomics*, 18(1): 460. DOI: 10.1186/s12864-017-3839-7.

Gómez-Blanco D. (2023). Telomeres in ecology and evolution: Hypotheses and connections between telomere length, infections and life history trade-offs – Lund University.

Hansson B., S. Bensch & Hasselquist D. (2003). Heritability of dispersal in the great reed warbler. *Ecology Letters*, 6(4): 290-294. DOI: 10.1046/j.1461-0248.2003.00436.x.

(a)Hansson B., S. Bensch, D. Hasselquist & Nielsen B. (2002). Restricted dispersal in a long-distance migrant bird with patchy distribution, the great reed warbler. *Oecologia*, 130: 536–542. DOI: 10.1007/s00442-001-0831-2.

(b) Hansson B., S. Bensch & Hasselquist D. (2002). Predictors of Natal Dispersal in Great Reed Warblers: Results from Small and Large Census Areas. *Journal of Avian Biology*, 33(3): 311–314. DOI: 10.1034/j.1600-048X.2002.330314.x.

Hasselquist D. (1998). Polygyny in great reed warblers: a long-term study of factors contributing to male fitness. *Ecology*, 79: 2376–2390. DOI: 10.1890/0012-9658(1998)079[2376:PIGRWA]2.0.CO;2.

Hasselquist D., S. Bensch & von Schantz T. (1996). Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Nature* 381, 229–232. DOI: 10.1038/381229a0.

He K., P. Minias & Dunn P.O. (2021). Long-Read Genome Assemblies Reveal Extraordinary Variation in the Number and Structure of MHC Loci in Birds. *Genome Biology and Evolution*, 13(2). DOI: 10.1093/gbe/evaa270.

Hedrick, P. W. (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution*, 56(10), 1902-1908. DOI: 10.1111/j.0014-3820.2002.tb00116.x

Heijmans C.M.C., N.G. de Groot & Bontrop R.E. (2020). Comparative genetics of the major histocompatibility complex in humans and nonhuman primates. *International Journal of Immunogenetics*, 47: 243–260. DOI: 10.1111/iji.12490.

Hess C.M. & Edwards S.V. (2002). The evolution of the Major Histocompatibility Complex in Birds: Scaling up and taking a genomic approach to the major histo compatibilty complex (MHC) of birds reveals surprising departures from generalities found in mammals in both large-scale structure and the mechanisms shaping the evolution of the MHC. *BioScience*, 52(5): 423–431. DOI: 10.1641/0006-3568(2002)052[0423:TEOTMH]2.0.CO;2.

Hill A. V. S. (1991). HLA associations with malaria in Africa: some implications for MHC evolution. In Molecular evolution of the major histocompatibility complex (eds Klein J., Klein D.), pp: 403–419 Berlin, Germany: Springer.

Horton R., L. Wilming, V. Rand, R.C. Lovering, E.A. Bruford, V.K. Khodiyar, M.J. Lush, S. Povey, C.C. Talbot Jr, M.W. Wright, H.M. Wain, J. Trowsdale, A. Ziegler & Beck S. (2004). Gene map of the extended human MHC. *Nature Reviews Genetics*, 5(12): 889–899. DOI: 10.1038/nrg1489.

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

Iwasaki A. & Medzhitov R. (2015). Control of adaptive immunity by the innate immune system. *Nature Immunology,* 16: 343–353. DOI: 10.1038/ni.3123.

Karawita A.C., Y. Cheng, K.Y. Chew, A. Challagulla, R. Klaus, R.C. Mueller, M.Z.W. Tong, K.D. Hulme, H. Bielefeldt-Ohmann, L.E. Steele, M. Wu, J. Sng, E. Noye, T.J. Bruxner, G.G. Au, S. Lowther, J. Blommaert, A. Suh, A.J. McCauley *et al.* (2023). The swan genome and transcriptome, it is not all black and white. *Genome Biology*, 24(1): 13. DOI: 10.1186/s13059-022-02838-0.

Karlsson M. & Westerdahl H. (2013). Characteristics of MHC class I genes in house sparrows *Passer domesticus* as revealed by long cDNA transcripts and amplicon sequencing. *Journal of Molecular Evolution*, 77(1-2): 8-21. DOI: 10.1007/s00239-013-9575-y.

Kaufman J. (2018). Unfinished Business: Evolution of the MHC and the Adaptive Immune System of Jawed Vertebrates. *Annual Review of Immunology*. (36), 383-409. DOI: 10.1146/annurev-immunol-051116-052450.

Kaufman J., S. Milne, T.W. Göbel, B.A. Walker, J.P. Jacob, C. Auffray, R. Zoorob & Beck S. (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, 401(6756): 923–925. DOI: 10.1038/44856.

Kaufman J., H. Völk & Wallny H.J. (1995). A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunological Reviews*, 143: 63–88. DOI: 10.1111/j.1600-065x.1995.tb00670.x.

Kelley J., L. Walter & Trowsdale J. (2005). Comparative genomics of major histocompatibility complexes. *Immunogenetics*, 56(10): 683–695. DOI: 10.1007/s00251-004-0717-7.

Koleček J., P. Procházka, N. El-Arabany, M. Tarka, M. Ilieva, S. Hahn, M. Honza, J. de la Puente, A. Bermejo, A. Gürsoy, S. Bensch, P. Zehtindjiev, D. Hasselquist & Hansson B. (2016). Cross-continental migratory connectivity and spatiotemporal migratory patterns in the great reed warbler. *Journal of Avian Biology*, 47(6): 756-767. DOI: 10.1111/jav.00929.

Kulski J.K., T. Shiina, T. Anzai, S. Kohara & Inoko H. (2002). Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunological Reviews*, 190: 95–122. DOI: 10.1034/j.1600-065X.2002.19008.x.

Kumánovics A., T. Takada & Fischer Lindahl K. (2003). Genomic Organization of the Mammalian Mhc. *Annual Review of Immunology*, 21: 629-657. DOI: 10.1146/annurev.immunol.21.090501.080116.

Lankat-Buttgereit B. & Tampé R. (2002). The transporter associated with antigen processing: function and implications in human diseases. *Physiological Reviews*, 82(1): 187-204. DOI: 10.1152/physrev.00025.2001.

Lenz, T. L. (2011). Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution*, 65(8): 2380-2390. DOI: 10.1111/j.1558-5646.2011.01288.x.

Li W-H. (1997). Concerted evolution of multigene families. In: Li W-H (ed) Molecular evolution. Sinauer Associates, Sunderland, MA, p.309.

Marshall J.S., R. Warrington, W. Watson & Kim H.L. (2018). An introduction to immunology and immunopathology. *Allergy, Asthma & Clinical Immunology*, 14 (2): 49. DOI: 10.1186/s13223-018-0278-1.

Mátrai N., J. Gyurácz, M. Lenczl, G. Hoffmann, G. Bakonyi & Mátics R. (2012). Philopatry analysis of the great reed warbler (*Acrocephalus arundinaceus*) based on ringing data in Europe. *Biologia,* 67: 596–601. DOI: 10.2478/s11756-012-0043-8.

Milinski, M. (2006). The Major Histocompatibility Complex, Sexual Selection, and Mate Choice. *Annual Review of Ecology, Evolution, and Systematics*, 37(1): 159-186. DOI: 10.1146/annurev.ecolsys.37.091305.110242.

Minias P., E. Pikus, L.A. Whittingham & Dunn P.O. (2018). Evolution of Copy Number at the MHC Varies across the Avian Tree of Life. *Genome Biology and Evolution*, 11(1): 17-28. DOI: 10.1093/gbe/evy253.

Moon D.A., S.M. Veniamin, J.A. Parks-Dely & Magor K.E. (2005). The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed class I genes. *The Journal of Immunology*, 175(10): 6702–6712. DOI: 10.4049/jimmunol.175.10.6702.

Neefjes J., M.L.M. Jongsma, P. Paul & Bakke O. (2011). Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, 11(12): 823–836. DOI: 10.1038/nri3084.

Nei M. & Rooney A.P. (2005). Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, 39: 121–152. DOI: 10.1146/annurev.genet.39.073003.112240.

Nei M., X. Gu & Sitnikova T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *PNAS*, 94(15): 7799–7806. DOI: 10.1073/pnas.94.15.7799.

O'Connor E.A & Westerdahl H. (2021). Trade-offs in expressed major histocompatibility complex diversity seen on a macroevolutionary scale among songbirds, *Evolution*, 75(5): 1061–1069. DOI: 10.1111/evo.14207.

O'Connor E.A., H. Westerdahl, R. Burri & Edwards S.V. (2019). Avian MHC Evolution in the Era of Genomics: Phase 1.0. *Cells*, 8(10). DOI: 10.3390/cells8101152.

O'Connor E.A., M. Strandh, D. Hasselquist, J.Å. Nilsson & Westerdahl H. (2016). The evolution of highly variable immunity genes across a passerine bird radiation. *Molecular Ecology*, 25(4): 977-89. DOI: 10.1111/mec.13530.

Ohta T. (2000). Evolution of gene families. *Gene*. 259(1-2):45-52. DOI: 10.1016/s0378-1119(00)00428-5.

Ota T. & Nei M. (1994). Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Molecular Biology and Evolution*, 11(3), 469-482. DOI: 10.1093/oxfordjournals.molbev.a040127.

Piertney S. B. & Oliver M. K. (2005). The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96(1), 7-21. DOI: 10.1038/sj.hdy.6800724.

Radwan J., W. Babik, J. Kaufman, T.L. Lenz & Winternitz J. (2020). Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends in Genetics*, 36(4): 298–311. DOI: 10.1016/j.tig.2020.01.008.

Riera Romo M., D. Pérez-Martínez & Castillo Ferrer C. (2016). Innate immunity in vertebrates: an overview. *Immunology*, 148: 125–139. DOI: 10.1111/imm.12597.

Roach M.J., S.A. Schmidt & Borneman A.R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, 19(1): 460. DOI: 10.1186/s12859-018-2485-7.

Rock K.L, E. Reits & Neefjes J. (2016). Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends in Immunology*, 37(11): 724–737. DOI: 10.1016/j.it.2016.08.010.

Roved J., B. Hansson, M. Stervander, D. Hasselquist & Westerdahl H. (2022). MHCtools – an R package for MHC high-throughput sequencing data: Genotyping, haplotype and supertype inference, and downstream genetic analyses in non-model organisms. *Molecular Ecology Resources*, 22: 2775–2793. DOI: 10.1111/1755-0998.13645.

Roved, J., Hansson, B., Tarka, M., Hasselquist, D., & Westerdahl, H. (2018). Evidence for sexual conflict over major histocompatibility complex diversity in a wild songbird. *Proceedings: Biological Sciences*, 285(1884). DOI: 10.1098/rspb.2018.0841.

Shiina T., A. Blancher, H. Inoko & Kulski J.K. (2017). Comparative genomics of the human, macaque and mouse major histocompatibility complex. *Immunology*, 150(2): 127–138. DOI: 10.1111/imm.12624.

Shiina T., K. Hosomichi & Hanzawa K. (2006). Comparative genomics of the poultry major histocompatibility complex. *Animal Science Journal*, 77: 151–162. DOI: 10.1111/j.1740-0929.2006.00333.x.

Shiina T., S. Shimizu, K. Hosomichi, S. Kohara, S. Watanabe, K. Hanzawa, S. Beck, J.K. Kulski & Inoko H. (2004). Comparative genomic analysis of two avian (quail and chicken) MHC regions. *The Journal of Immunology*, 172(11): 6751–6763. DOI: 10.4049/jimmunol.172.11.6751.

Sigeman, H., M. Strandh, E. Proux-Wéra, V.E. Kutschera, S. Ponnikas, H. Zhang, M. Lundberg, L. Soler, I. Bunikis, M. Tarka, D. Hasselquist, B. Nystedt, H. Westerdahl & Hansson B. (2021). Avian Neo-Sex Chromosomes Reveal Dynamics of Recombination Suppression and W Degeneration. *Molecular Biology and Evolution*, 38(12): 5275–5291. DOI: 10.1093/molbev/msab277.

Sommer S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, 2: 16. DOI: 10.1186/1742-9994-2-16.

Spurgin L.G., C. van Oosterhout, J.C. Illera, S. Bridgett, K. Gharbi, B.C. Emerson & Richardson D.S. (2011). Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, 20: 5213–5225. DOI: 10.1111/j.1365-294X.2011.05367.x.

Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings: Biological Sciences*, 277(1684): 979-988. DOI: 10.1098/rspb.2009.2084.

Takahata, N., & Nei, M. (1990). Allelic Genealogy under overdominant and frequency-dependent selection and Polymorphism of Major Histocompatibility Complex loci. *Genetics*, 124, 967-978. DOI: 10.1093/genetics/124.4.967.

Tarka M., B. Hansson & Hasselquist D. (2015). Selection and evolutionary potential of spring arrival phenology in males and females of a migratory songbird. *Journal of Evolutionary Biology*, 28: 1024–1038. DOI: 10.1111/jeb.12638.

Tarka M., M. Åkesson, D. Beraldi, J. Hernández-Sánchez, D. Hasselquist, S. Bensch & Hansson B. (2010). A strong quantitative trait locus for wing length on chromosome 2 in a wild population of great reed warblers. *Proceedings: Biological Sciences*, 277(1692). DOI: 10.1098/rspb.2010.0033.

Trowsdale J. (1995). "Both man & bird & beast": comparative organization of MHC genes. *Immunogenetics*, 41: 1–17. DOI: 10.1007/BF00188427.

Tsuji H., Y. Taniguchi, S. Ishizuka, H. Matsuda, T. Yamada, K. Naito & Iwaisaki H. (2017). Structure and polymorphisms of the major histocompatibility complex in the Oriental stork, *Ciconia boyciana*. *Scientific Reports*, 7: 42864. DOI: 10.1038/srep42864.

Wakeland E. K., S. Boehme, , J. X . She, C.-C. Lu, R.A. McIndoe, I. Cheng & Potts W. K. (1990). Ancestral Polymorphisms of MHC Class II Genes: Divergent Allele Advantage. *Immunologic Research*, 9(2): 115-122. DOI: 10.1007/BF02918202.

Walker BA, van Hateren A, Milne S, Beck S, Kaufman J. Chicken TAP genes differ from their human orthologues in locus organisation, size, sequence features and polymorphism. Immunogenetics. 2005;57:232–247.

Wang B., R. Ekblom, T.M. Strand, S. Portela-Bens & Höglund J. (2012). Sequencing of the core MHC region of black grouse (*Tetrao tetrix*) and comparative genomics of the galliform MHC. *BMC Genomics*, 13: 553. DOI: 10.1186/1471-2164-13-553.

Westerdahl H., S. Mellinger, H. Sigeman, V.E. Kutschera, E. Proux-Wéra, M. Lundberg, M. Weissensteiner, A. Churcher, I. Bunikis, B. Hansson, J.B.W. Wolf & Strandh M. (2022). The genomic architecture of the passerine MHC region: High repeat content and contrasting evolutionary histories of single copy and tandemly duplicated MHC-genes. Molecular Ecology Resources, 00: 1-17. DOI: 10.1111/1755-0998.13614.

(a)Westerdahl H., B. Hansson, S. Bensch & Hasselquist D. (2004). Between-year variation of MHC allele frequencies in great reed warblers: selection or drift? *Journal of Evolutionary Biology*, 17(3): 485-492. DOI: 10.1111/j.1420-9101.2004.00711.x

(b)Westerdahl H., H. Wittzell, T. Schantz & Bensch S. (2004). MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. *Heredity*, 92: 534–542. DOI: 10.1038/sj.hdy.6800450

Westerdahl H., H. Wittzell & von Schantz T. (2000). Mhc diversity in two passerine birds: no evidence for a minimal essential Mhc. *Immunogenetics,* 52(1-2): 92-100. DOI: 10.1007/s002510000256

Westerdahl H., H. Wittzell & von Schantz T. (1999). Polymorphism and transcription of Mhc class I genes in a passerine bird, the great reed warbler. *Immunogenetics*, 49: 158-170. DOI: 10.1007/s002510050477.

Wittzell H., A. Bernot, C. Auffray & Zoorob R. (1999). Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. *Molecular Biology and Evolution*, 16(4): 479-90. DOI: 10.1093/oxfordjournals.molbev.a026130.

Ye Q., K. He, S.Y. Wu & Wan Q.H. (2012). Isolation of a 97-kb Minimal Essential MHC B Locus from a New Reverse-4D BAC Library of the Golden Pheasant. *PLOS ONE*, 7(3): e32154. DOI: 10.1371/journal.pone.0032154.

Yuhki N., T. Beck, R.M. Stephens, Y. Nishigaki, K. Newmann & O'Brien S.J. (2003). Comparative genome organization of human, murine, and feline MHC class II region. *Genome Research*, 13(6A): 1169–1179. DOI: 10.1101/gr.976103.

Zelano B & Edwards S. V. (2002). An Mhc Component to Kin recognition and Mate choice in Birds: Predictions, Progress and Prospects. *The American Naturalist*, 160: 225-237. DOI: 10.1086/342897.

Acknowledgments – Thank you!

I feel very lucky to have met so many different and amazing people since I arrived in Lund during a very cold day in January 2018 and throughout my PhD journey! I have learned so many things from all of you and I feel like I grown up so much. In my flat, I have a sign that says "Lifes truest happiness is found in the friendships we make along the way" and I can sincerely relate to that sign thanks to all of you.

First of all, I would to thank **Helena**. You have been an amazing supervisor during all these years. You gave me a chance to come for a Master internship, then you partly employed me and to finish, you offered me one of the most wonderful opportunity: to be a PhD student. You taught me so much. You never gave up on me even in the worst moments and I would have never finished this PhD if you would not have pushed me. Tack så mycket!

A big thank you to my co-supervisors. Your support and your valuable feedbacks were especially helpful. **Lars,** thank for your help for the statistics and for your clear explanations. It was always such a pleasure to talk to you. **Dennis**, thank you for letting me join the great reed warbler fieldwork and for your nice words when I was down.

Thank you, **Karin**, for being such a great department representative, you really followed me up carefully during the past years. You supported me a lot to find help and to make me think twice on my thoughts last year.

Anna, my friend, I would never be able to thank you enough for your guidance, your support, your patience in listening to me about the problems in the lab or about the things that happened in my personal life, your kindness, your pragmatism… Everything seemed and still seems easy with you by my side.

Jane, you have been such a great friend throughout the years. You have helped me so much. Not only in the lab but also everytime I needed to talk. You shared so much with me and I am the most grateful.

Thank you to all the former and present collaborators in the MHC group. **Julio**, it was always such a pleasure to discuss with you about science and birds and malaria and many more things. **Maria**, I really miss the time when we were sharing the office. **Sandra**, I feel very lucky for the time we spent together doing science and talking about life. **Nisha**, I am so grateful for the strong friendship we build. Soon, we will defend also your thesis and I know that you will be fantastic! Thank you

also to **Emily** (for helping to check my English), to **Hannah** (for dealing with the robot almost forever), to **Jacob**, to **Kelly**, to **Iris** and to **Jon.**

Thank you, **Mridula.** You were such a good friend. I always felt understood with you and I was never scared of talking freely about everything. You helped me so much to see clearer in my feelings.

Thank you, **Max**. I always enjoyed our interesting conversations at fikas or during lunches. I will always be amazed by your ability to simply explain to me complicated bioinformatic stuffs.

Thank you to all fellow PhD-student mates. We all have started almost at the same time and we have been through a lot together or not. I never felt alone. Thank you to **David** (I really loved chatting and collaborating with you and thank you for being nice with me during the field season), **Hongkai** (I will always have lollipops for you), **Elsie** (you are always so helpful and nice with me), **Kristaps** ("you're lack of intellectual capacities frightens me" and …) and **Victor**.

Thank you to all people I had the chance to spent time with in the past or still in the present, at work or outside work, during fikas, lunches or at the pub: **Chiara** (Ciao, bella!), **Violeta** (for your forever honesty that make me laugh so much everytime), **Sara** (the real italian coffee expert), **Zsófia** (I promise to always sing for you if you ask me), **Micaela** (we shared so much and I know we really understand each other), **Twinkle** (for all our conversations), **Tamara** (you were by my side until the very last day), **Erica** (my favorite Easter bunny, you are amazing), **Ann-Kathrin** (I will be always grateful for your advices when I needed them the most), **Linus** (Why the mammoths have disappeared?), **Angela and Théo** (everytime I enjoy our conversations), **Sofie** (always smiling), **Evelina, Alessia, Katie, Javier A., Mingyue, Kat**, **Kalle**, **Qinyang**, **Tianhao**, **Javier P., Robin, Agnes**, **Jocke** (for your true and deep kindness), **Maja** (thank you for your encouraging words), **Staffan B., Bengt H., Rae, Julian, Arif, Fredrik, Hanna S., Philip, Suvi, Luz, Vincenzo, Maria S.C., Pablo, Juan Pablo, Moritz, Elin, Susana, Johan N., Jessica A., Sara W., Carlos, Olivier** (for being my scientific mentor), **Hanna B., Martin S., Raphaël, Mikkel, Sissel, Aivars, Simon, Emma K., Yesbol, Damandeep and Erik S.**

Thank you to the MEEL group and the "second floor" in general.

Thank you to **Annika** (every time I had a question, you had an answer), **Ewa and Izabella** for your support.

Thank you to all the people I met in the "Aquarium". First and foremost, my beloved friend **Saranda** with whom I laugh so much the first months. Thank you to **Pallavi** for all the good time at fikas. Thank you to **Mariana** (I loved talking with you. You are the only one who have seen my little "Staying alive" dance, I really miss you my friend). Thank you to **Mélanie** (Count on me my friend to come and visit you in

France) and to **Farisia** (I am so lucky to know you). Thank you to **Vignesh**, **Jésus V., Jéssica, Daniel, Sachin, Alejandro, Paulius, Vojtech** and so many others…

To all my colleagues at the Ecology Building, thank you for making this place the most vibrant and inspiring. Thank you to **Tomas** (thanks to you I could sequence my amplicons in house!), to **Anna** (for all the sanger sequencing you did for me and for your help during the Molecular Ecology Course).

Thank you to the GRW fieldwork team (2019) especially **Gintaras**, **Laila**, **Lucia and Andrea.**

I am sorry in advance if I missed any names. I thank you all from the bottom of my heart.

I would like to thank another scientific mentor, **Helena P.**, who have helped me a lot and always gave me good advices. I would like thank my psychologist **Johan** for listening to me.

Thank you to my amazing flatmates **Marja** and **Deborah**. You are part of my family now and I am so grateful to have lived with you all these years. We shared so many good moments and when I was feeling down, I could always count on you. I love you girls. Thank you to **Anders** for all his good advices and for the laughs.

To finish, I would like to thank my family and friends in France. À mes chers parents, je n'aurais jamais pu accomplir tout ce chemin sans vous. À ma maman, pour toutes les fois où tu m'as réconforter au telephone et à la maison. À mon papa, pour toutes les fois où tu as dû me chercher à l'aéroport. Je vous dois tout et je vous aime de tout mon petit coeur.

List of papers

I. Westerdahl H., **Mellinger S.**, Sigeman H., Kutschera V.E., Proux-Wéra E., Lundberg M., Weissensteiner M., Churcher A., Bunikis I., Hansson B., Wolf J.B.W. and Strandh M. (2022). The genomic architecture of the passerine MHC region: High repeat content and contrasting evolutionary histories of single copy and tandemly duplicated MHC genes. *Molecular Ecology Resources*, 22, pp 2379-2395, doi: 10.1111/1755-0998.13614.

II. **Mellinger S.**, Stervander M., Lundberg M., Drews A. and Westerdahl H. (2023). Improved haplotype resolution of highly duplicated MHC genes in a long-read genome assembly using MiSeq amplicons. *PeerJ*, 11(5): e15480, doi: 10.7717/peerj.15480.

III. **Mellinger S.**, Råberg L., Hasselquist D. and Westerdahl H. Temporal variation and age-associated changes in MHC-I gene expression in great reed warblers. Manuscript.

IV. **Mellinger S.**, Råberg L., Hasselquist D. and Westerdahl H. Inheritance of MHC-I gene expression in great reed warblers. Manuscript.

Printed by Media-Tryck, Lund 2023 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Department of Biology
Faculty of Science

ISBN 978-91-8039-766-7