Year : 2017

MHC EVOLUTIONARY ECOLOGY IN EUROPEAN BARN OWLS [TYTO ALBA)

Gaigher Arnaud

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département d'écologie et évolution

MHC EVOLUTIONARY ECOLOGY IN EUROPEAN BARN OWLS (*TYTO ALBA*)

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Arnaud Gaigher

Master de l'université Pierre et Marie Curie (Paris VI)

Jury

Prof. Philippe Reymond, Président

Dr. Luca Fumagalli, Directeur de thèse

Prof. Alexandre Roulin, Co-directeur de thèse

Prof. Nicolas Salamin, expert

Prof. Tobias L. Lenz, expert

Lausanne 2017

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président· e	Monsieur	Prof.	Philippe	**Reymond**
Directeur· rice de thèse	Monsieur	Dr	Luca	**Fumagalli**
Co-directeur· rice	Monsieur	Prof.	Alexandre	**Roulin**
Experts· es	Monsieur	Prof.	Nicolas	**Salamin**
	Monsieur	Prof.	Tobias L.	**Lenz**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Arnaud Gaigher

Master de l' Université Pierre et Marie Curie - Paris VI, France

intitulée

MHC EVOLUTIONARY ECOLOGY IN EUROPEAN BARN OWLS *(TYTO ALBA)*

Lausanne, le 11 septembre 2017

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Philippe Reymond

Index

Remerciements

Je voudrais en premier lieu remercier mes deux superviseurs Luca Fumagalli et Alexandre Roulin pour m'avoir donné l'opportunité de réaliser cette thèse au sein du département d'écologie et évolution. Vous avez tous les deux des approches scientifiques différentes mais complémentaires, dont j'ai pu tirer profit au cours de ma thèse. Je vous remercie également pour l'autonomie que vous m'avez accordée durant cette thèse, me laissant développer certains projets comme je le souhaitais. Finalement encore merci pour votre enthousiasme à toute épreuve. En second lieu, je voudrais remercier Reto Burri… Je ne sais pas par où commencer, mais je te remercie chaleureusement pour tout le soutien que tu m'as apporté au cours de cette thèse. Bien que je t'ai « physiquement » très peu vu, ta présence, ton aura dans cette thèse est incontestable. Merci Reto pour ta supervision à distance, ton professionnalisme, ton esprit critique et ta disponibilité. Amitié…

Ensuite j'aimerais remercier tous les membres de mon jury de demi-thèse ainsi que de thèse, à savoir mes présidents Dario Diviani et Philippe Reymond, et mes experts, Jérôme Goudet, Nicolas Salamin et Tobias L. Lenz. Un grand merci à vous pour avoir pris de votre temps pour évaluer mes travaux de recherches. Thank you Tobias for having taken your time to review my thesis manuscript and give constructive feedback.

J'aimerais remercier mes collaborateurs de cette thèse, Luis San Jose Garcia, Sylvain Antoniazza, Pierre Taberlet et Walid Gharib. Merci à eux pour avoir significativement amélioré et facilité mes projets de recherches. Merci Sylvain pour m'avoir guidé lors de mon arrivée à Lausanne, puis de m'avoir soutenu au cours de ma thèse. Un grand merci à tous les collaborateurs qui de près ou de loin ont participé à l'échantillonnage, merci pour ce fabuleux travail sur le terrain.

Un grand merci à tous le département DEE, pour toutes les interactions scientifiques mais aussi toutes les interactions sociales à travers les (très nombreux) apéros. Je remercie aussi les filles qui m'ont donné pleins de conseils au labo et pour leurs bonnes humeurs, Céline St, Céline Si, Anne-Lyse et Nadège. Merci à toi Céline pour les petits moments sportifs au terrain de tennis. J'aimerais également remercier tous les copains/amis de l'UNIL avec qui j'ai passé des moments inoubliables au cours de mes 5 ans à Lausanne, bella Vera, Valouch, mon loulou Julien, Helenita, Esra, Jessica, Luisso, Paulochon, Popo, Ana, Robinouch, Kim et Guillaume. Merci aussi aux sportifs et amis, Yannis, Rudaï, Baudouin et Pascal. Un grand merci à Sophie, Kévin, Amandine, Loeiza, Christie et Laure pour votre amitié durant ma thèse. Big thank à mes colocs que j'adore Laeti et Gé. Je voudrais tout particulièrement remercier Luis, Vera et Valouch, mes amis, qui ont été d'un très grand support à la fois moral et scientifique pendant ma thèse et m'ont aidé à finaliser ce manuscrit. Enfin merci à toi Tchoukinette !!!

Mes derniers remerciements vont tout droit à ma famille. Pépère, Maman, Papa, ma sœurette, Aldja, Christian et Fafa merci pour votre soutien inconditionnel. Même si la distance nous a physiquement éloignés, vous avez au cours de mes années scolaires toujours été très présents pour moi et m'avez laissé voler de mes propres ailes. Donc si ce manuscrit de thèse voit le jour c'est grâce à vous. Aussi, un grand merci à mon père pour les superbes dessins de ce manuscrit.

Summary

Elucidating the mechanisms promoting and maintaining adaptive genetic diversity (i.e., variation at loci with a direct effect on fitness) is a central issue in evolutionary and conservation biology. Due to their essential role in pathogen resistance and extraordinary levels of polymorphism, genes of the major histocompatibility complex (MHC) constitute perfect candidates for studying adaptive genetic variability. However, despite the accumulation of empirical data, estimating the relative contribution of adaptive against non-adaptive processes that drive the MHC evolution in natural populations is still challenging, and consequently requires further investigation.

In this thesis, my general aim was to understand the pattern of MHC evolution by using the European populations of barn owl (*Tyto alba*) as model species. We adopted a sampling strategy including individuals from family data, for which fitness-related trait information from experimental data were available, as well as an extensive number of individuals from many localities in Europe. This sampling resulted in a total of 1'300 barn owls that were sequenced for both MHC class Iα exon3 (MHC-I) and MHC class IIβ exon 2 (MHC-IIB) using a high-throughput sequencing technology. Our strategy allowed us to address some of the important topics in the MHC field, i.e. MHC genotyping, molecular evolution of MHC genes, the link between MHC diversity and individual fitness-related traits, and the spatial pattern of MHC diversity.

I first characterized the diversity at the two MHC classes (MHC-I and MHC-IIB), and revealed only two duplicated genes in each class. All four genes displayed the characteristics of functional MHC genes, i.e. footprints of historical positive selection, high genetic diversity, and footprints of recombination and gene conversion. Additionally, I highlighted a remarkable discrepancy in the evolutionary dynamic between the two MHC classes, where alleles between MHC-I genes were extremely homogenized, while the ones at MHC-IIB genes were strongly divergent. These findings suggest that the different evolutionary histories at the two MHC classes may constrain or promote the evolution for an optimal MHC-mediated pathogen defence. Taking advantage of this detailed characterization, in a follow-up study, I investigated whether specific aspects of MHC diversity (i.e. high diversity or specific alleles) confer higher immunocompetence. Overall I found low support in this sense, but most importantly these results highlighted that if an effect exists, it is of small to moderate effect size, which is important for future studies in order to achieve enough power to detect small-effects associations. Finally, I took a global view and investigated the pattern of MHC diversity across Europe. I found that MHC diversity has been mainly driven by the species colonization history, even if certain localities show intriguing patterns that may allow me to think that selection has also contributed to shape MHC diversity. This study revealed the complexity of distinguishing the relative contribution of evolutionary forces shaping the spatial MHC diversity and requires pathogen community information to properly discern the forces in action.

The findings of this PhD, using a multilevel approach coupled with high throughput sequencing technologies, have permitted to bring new insights into the MHC evolutionary ecology, but have also opened new challenging perspectives, such as the importance of using genomic scans and sampling strategy efforts, in order to overcome the general difficulty of fully understanding the outcomes of host-pathogen interactions.

Résumé

Comprendre les mécanismes à l’origine et qui maintiennent la diversité génétique adaptative (c'est-à-dire la diversité aux gènes avec un effet direct sur la fitness) est un sujet central en biologie évolutive et biologie de la conservation. En raison de leur rôle essentiel dans la résistance aux agents pathogènes et de leurs niveaux exceptionnels de diversité, les gènes du complexe majeur d'histocompatibilité (CMH) constituent des candidats parfaits pour étudier la variabilité génétique adaptative. Cependant, malgré l'accumulation de données empiriques, estimer le rôle relatif de la sélection et des processus neutres qui façonne l'évolution du CMH dans les populations naturelles est encore difficile et, par conséquent nécessite des études plus approfondies.

Dans cette thèse, mon objectif général était de comprendre le modèle d'évolution de la diversité du CMH chez la chouette effraie (*Tyto alba*). Nous avons adopté une stratégie d'échantillonnage comprenant des individus provenant de données familiales, pour lesquelles des informations sur la fitness étaient disponibles, ainsi qu'un grand nombre d'individus d’Europe. Cet échantillonnage a abouti à un total de 1 300 chouettes qui ont été séquencés pour l’exon 3 du CMH classe I (MHC-I) et le l’exon 2 du CMH classe II (MHC-IIB) avec la technologie de séquençage à haut débit. Notre stratégie nous a permis d'aborder des sujets clés dans le domaine du CMH, à savoir le génotypage du CMH, l'évolution moléculaire des gènes du CMH, le lien entre la diversité au CMH et des traits liés à la fitness individuel, ainsi que la variation spatiale du CMH.

J'ai d'abord révélé que chaque classe du CMH a seulement deux gènes dupliqués. Les quatre gènes ont montré les caractéristiques habituelles de gènes fonctionnels, c'est-à-dire des empreintes de sélection positive, une grande diversité génétique et l’effet de la recombinaison et de conversion génique. J'ai mis en évidence un écart remarquable dans la dynamique évolutive entre les deux classes du CMH ; les allèles entre les gènes du MHC-I étaient extrêmement homogénéisés, tandis que ceux des gènes du MHC-IIB étaient fortement divergents. Ces résultats suggèrent que ces différences évolutives dans les deux classes du CMH peuvent contraindre ou favoriser l'évolution d'une défense optimale contre les agents pathogènes. Profitant de cette caractérisation détaillée, j'ai ensuite étudié si des aspects spécifiques de la diversité du CMH (c'est-à-dire une grande diversité ou des allèles spécifiques) confèrent une immunocompétence plus élevée. Dans l'ensemble, j'ai trouvé un faible soutien dans ce sens, mais surtout, ces résultats ont mis en évidence que si un effet existe, il est de taille faible à modérée. Ce résultat est important pour les études futures afin d'obtenir suffisamment de pouvoir statistique pour détecter les associations de petits effets. Enfin, j'ai pris une vision plus globale et j'ai étudié la diversité du CMH en Europe. J'ai constaté que la diversité du CMH a été principalement façonnée par l'histoire de colonisation de la chouette, même si certaines localités Européen montrent des modèles intrigants qui laissent à penser que la sélection a également pu contribuer à la diversité du CMH observé. Cette étude a révélé la complexité de distinguer la contribution relative des forces évolutives qui façonnent la diversité spatiale du CMH et nécessite des informations sur les communautés pathogènes pour discerner correctement les forces en action.

Grâce à une approche multiniveau couplée à des technologies de séquençage à haut débit, les résultats de ce doctorat ont permis d'apporter de nouvelles connaissances sur l'écologie évolutive du CMH, mais ont également ouvert de nouvelles perspectives pour les future études, telles que l'importance de l'utilisation d'analyses génomiques et des efforts de stratégie d'échantillonnage, afin d’améliorer notre compréhension des interactions hôte-pathogène.

GENERAL INTRODUCTION

Historical background

The relative role of alternative evolutionary processes, in underpinning the molecular and phenotypic variation in natural populations, is a central issue in evolutionary biology, and it has long been a topic of intense debate. In the early 60s, the development of new molecular techniques, such as the enzymatic electrophoresis, unraveled a high level of polymorphisms within species (Lewontin & Hubby 1966), which fueled the famous debate on whether this variation can be explained either by balancing selection or neutral processes. The works of Kimura (1968) and of Kingand Jukes (1969) were the first to formally challenge the preponderant, neo-Darwinist dogma about the major role of natural selection in shaping genetic variation, i.e. most polymorphisms are adaptive. This neutralist view, the so-called neutral theory of molecular evolution (reviewed in Nei 2005), states that the majority of genetic changes are neutral and result from the joint action of genetic drift and mutation, and it has been controversial for a long time. This vision does not exclude the effect of selection, but rather removes its exclusivity as the main mechanism shaping genetic diversity. Although there are still remnants of this debate (Lynch 2007; Hahn 2008; Nei *et al.* 2010), neutral theory has gained support due to the growing body of evidence showing that neutral processes can play an important role in shaping genetic variation, to a similar extent as selective processes. The neutral theory is now a major conceptual framework in evolutionary biology, and has become the null hypothesis against which the presence of selection can be tested (Kreitman 1996). Nowadays, interest has shifted to understand what regions of the genome evolve under what evolutionary processes, and, having discarded neutral forces, to fully characterize the mechanisms promoting and maintaining the adaptive genetic diversity.

From neutral to adaptive genetic variation

The characterization of neutral diversity has been used to cover various topics in evolutionary biology, such as for instance phylogenetic relationships among organisms, past demographic events (expansions or bottlenecks), population structure, dispersal patterns, paternity assessment, or species delimitation. Although extremely useful in the above-mentioned contexts, neutral diversity cannot provide information to understand how organisms are adapted to their environment, nor on the adaptive potential of populations to cope with environmental changes. Nevertheless, due to the difficulty to unravel which genes are or have been under selection, levels of neutral genetic variation have been used in the past as a proxy for the level of adaptive genetic diversity of a species or a population. Although this approach might be valid in certain cases and it has certainly allowed for substantial development in several areas, it assumes the existence of a strong correlation between neutral and adaptive genetic diversity. However, such correlation can be rather weak (Hedrick 2001; Reed & Frankham 2001; Aguilar *et al.* 2004) and, hence, only the direct study of genetic markers that have an impact on fitness will give accurate information to study the evolutionary processes that maintain adaptive genetic variation (van Tienderen *et al.* 2002).

The advent of high throughput sequencing technologies has permitted the emergence of wide- or even whole-genome analyses, and a considerable amount of genetic data, which has become the basis for identifying genes underlying genomic signatures of selection, and to study their evolutionary dynamics. Coupled with this, many sophisticated tools have been developed to detect selection across the genome (Vitti *et al.* 2013), and, as a consequence, there exists now a plethora of studies discerning those genomic regions evolving under the null assumption of neutrality, from those hosting adaptive evolutionary changes (Yi *et al.* 2010; Qiu *et al.* 2012; Zhan *et al.* 2013; Poelstra *et al.* 2014; Zhang *et al.* 2014; Malinsky *et al.* 2015). Genome-wide scans can constitute an interesting approach to identify regions under selection (Oleksyk *et al.* 2010), however, several researchers have concentrated their efforts on specific candidate genes as an alternative (Piertney & Webster 2010), which have resulted in important discoveries (for instance the lactose digestion capacity, Bersaglieri *et al.* 2004; Tishkoff *et al.* 2007). Candidate genes are genes that have been identified to influence a phenotypic trait in one organism, and are expected to influence a similar phenotypic trait in another organism (Fitzpatrick *et al.* 2005). Focusing on such an approach confers some advantages in comparison to a genome-wide scan approach, i.e. a very high number of individuals can be sequenced for a reduced cost, it can reveal the effect of a specific mutation, and finally, the structure and function of proteins for which genes are encoding are often well understood, which facilitates the identification of which pressure or pressures are causing selection. There are some examples of well-characterized functional candidate genes, such as the melanocortin-1-receptor (MC1R) gene involved in colour polymorphism

(Mundy 2005). Molecular changes at the MC1R have been shown to mediate adaptation via background matching of coat colouration (Nachman *et al.* 2003; Hoekstra *et al.* 2006). Similarly, genes of the major histocompatibility complex (MHC) are advocated among the best candidates to study adaptive genetic diversity (Bernatchez & Landry 2003). Indeed, due to their direct link with individual pathogen resistance, hence fitness, the MHC system has received considerable attention across different disciplines (evolutionary biology, molecular biology, conservation biology, medicine, epidemiology, etc.), and consequently, the evolution, the structure, the molecular pathway, and the function of MHC proteins are now well understood, even in non-model species (Kaufman *et al.* 1999; Jeffery & Bangham 2000; Hess & Edwards 2002; Bernatchez & Landry 2003; Jensen 2007). The number of studies using such candidate genes is still growing because they offer an effective tool for the interpretation of the role of selection in maintaining genetic diversity, and to understand what ecological processes drive these systems in natural populations.

The MHC: a suited system for studying adaptive genetic diversity

Structure and function

The MHC is a multigene family that plays an essential role in the vertebrate adaptive immune response in vertebrates. The MHC system includes two major subfamilies, MHC class I and class II, which encode for cell-surface proteins (Fig. 1) that, in turn, present antigen-peptides to the T-cell receptors to trigger an appropriate immune response. Although both MHC classes share many structural and functional characteristics, these molecules are involved in different pathways of antigen presentation. Both pathways are complex and have been described in several good reviews (e.g. Klein & Sato 2000; Jensen 2007; Neefjes *et al.* 2011; Blum *et al.* 2013). Briefly, MHC-I genes encode for MHC proteins located in all nucleated cells and that interact with peptides derived from proteins synthesized within the cell (endogenous origin, for instance viral proteins). These endogenous proteins are first digested into

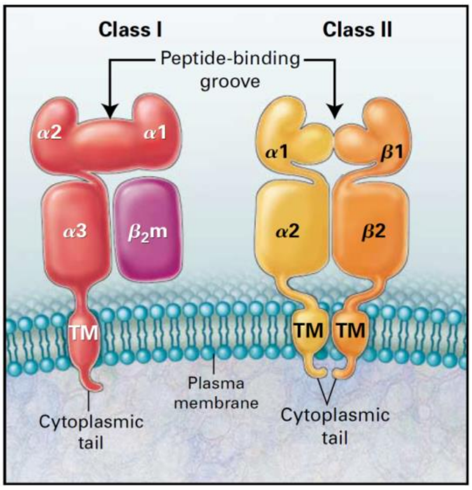


Figure 1: Structure of MHC class I and class II proteins. From Klein & Sato (2000).

smaller peptides by the proteasome and proteases in the cytosol. These peptides are then transported within the endoplasmic reticulum, where the peptide-MHC molecule complex can be formed. This complex is finally exposed to the cell surface for presentation to $CD8^+$ T-cells, which can identify and destroy infected cells to avoid the spread of the infection. On the other hand, the MHC class II proteins are expressed only on a subset of specific cells, called antigen presentation cells (including for instance dendritic cells and macrophages). MHC-II molecules bind peptides with an exogenous origin, which enter in the cell by processes of phagocytosis or endocytosis (for instance bacteria infecting a wound). Once the peptide-MHC molecule complex is formed and transported to the cell surface, the complex is presented to $CD4^+$ T cells. The recognition by $CD4^+$ T cells triggers the activation and regulation of various pathways of a protective immune response against extracellular pathogens, such as the proliferation of lymphocytes and the antibody production to neutralize pathogens (Janeway *et al.* 2001; Holling *et al.* 2004). However, the strict demarcation of these two pathways is not absolute and evidence indicates the possibility of cross-presentation: exogenous peptides can enter the MHC-I pathway, or inversely, endogenous peptides can enter the MHC-II pathway (Heath & Carbone 2001).

One of the most important regions of the MHC molecule is the groove or the so-called peptide-binding region (PBR, or peptide binding groove in Fig. 1), which recognizes and binds foreign peptides for presentation to the T-cells. MHC-I binds peptides of usually 8-11 amino acids, whereas MHC-II interacts with longer peptides of 14-20 amino acids (Jeffery & Bangham 2000). Due to the functional importance of the PBR, most studies directly target the genes encoding for these regions. For MHC-I, the two extracellular domains (α1 and α2) are involved in the PBR and are encoded by exon 2 and 3 from a single gene (Bjorkman *et al.* 1987) (Fig. 1). On the other hand, α1 and β1 domains involved in the PBR of the MHC-II are encoded by two different genes (Fig. 1). A given PBR binds to a specific subset of peptides which are not necessarily specific to a given pathogen. Intriguingly, the polymorphisms in MHC molecules are mostly located at sites involved in the PBR, which should allow for the recognition of a wider variety of pathogen-derived antigens. It is therefore not surprising that it was quickly postulated that the extraordinary level of polymorphisms at MHC genes is governed by selection imposed by pathogens (Doherty & Zinkernagel 1975; Apanius *et al.* 1997; Jeffery & Bangham 2000).

Mechanisms maintaining MHC diversity

MHC genes are among the most polymorphic genes in vertebrates (Gaudieri *et al.* 2000), with some of them exhibiting hundreds of alleles in human populations (Robinson *et al.* 2015). Due to their direct link with pathogen recognition and resistance, it has long been suggested that pathogens are the major selective agent in driving MHC evolution and somehow involved in maintaining MHC variation in natural populations (Apanius *et al.* 1997; Jeffery & Bangham 2000; Spurgin & Richardson 2010). This selection is referred as the pathogen-mediated selection (PMS), which can take the shape of one of three hypothesized balancing selection mechanisms: (i) the heterozygote advantage, (ii) the rare-allele advantage (or negative frequency-dependent selection), and (iii) the diversifying selection at a spatial or temporal scale (reviews in Sommer 2005; Piertney & Oliver 2006; Spurgin & Richardson 2010).

Under the heterozygote advantage hypothesis, heterozygous individuals will present a large number of different MHC molecules and are then able to screen a broader array of pathogen-peptides, which will consequently confer them a better resistance against pathogens and a selective advantage over homozygous individuals (Doherty & Zinkernagel 1975). This hypothesis was later extended to also take into account allelic diversity within genotype. Thus, it was proposed that some heterozygote genotypes which compile a higher allelic divergence can interact with a wider range of pathogen-peptides than heterozygotes with low allelic diversity (the so-called "divergent allele advantage", Wakeland *et al.* 1990; Lenz 2011). Under the rare-allele advantage hypothesis, the new or rare alleles provide a better pathogen resistance, as pathogens have not yet adapted to them during the evolutionary host-pathogen arms race (Takahata & Nei 1990). Despite that MHC allele frequencies are expected to constantly fluctuate in the population, this dynamic cyclic process can nevertheless result in the persistence of a large set of MHC alleles (Slade & McCallum 1992; Borghans *et al.* 2004). Finally, under the diversifying selection hypothesis, diversity at the MHC may be maintained by spatiotemporal variability of pathogens (Hedrick 2002). The diversifying selection may cause important shifts in the MHC allele frequency distribution for instance between populations, reflecting their adaptations to their own pathogen communities.

Although these models of selection are easy to define, pointing out their relative contribution in maintaining the levels of MHC polymorphisms remains challenging, especially as they are expected to act in concert and yield similar outcomes (Spurgin &

Richardson 2010). In addition, the complexity of the overall picture increases, since other mechanisms such as sexual selection through MHC-dependent mate choice, has been suggested to contribute to MHC polymorphism levels (reviews in Penn & Potts 1999; Milinski 2006; Kamiya *et al.* 2014), and can amplify the previously mentioned PMS models.

The MHC system has been attractive not only in linking MHC diversity with pathogen resistance, but also for its particular pattern of molecular evolution. One remarkable characteristic of the MHC system is its high levels of evolution by gene duplication. Usually, species exhibit several duplicated MHC genes, and their number and organization in the genome can drastically vary among organisms (Kelley *et al.* 2005), especially in birds (Hess & Edwards 2002). For instance, the simple and compact organization observed in Galliforms (chicken, Kaufman *et al.* 1999), contrasts with the occurrence of many functional paralogs and pseudogenes in Passeriforms (Westerdahl 2007; Promerová *et al.* 2009; Zagalska-Neubauer *et al.* 2010; Sepil *et al.* 2012; Biedrzycka *et al.* 2016; O'Connor *et al.* 2016). In addition, the high rates of recombination and gene conversion, that usually characterize the MHC system (Hess & Edwards 2002), can deeply contribute to shape and maintain the diversity at MHC genes (Spurgin *et al.* 2011). Consequently, studying the molecular evolution and genomic structure of such a multigene family is an important prerequisite to understand the significance of the MHC diversity in relation to pathogen resistance. This also entails certain practical disadvantages, given that the complex and dynamic evolution of the MHC system renders the access to the MHC diversity laborious from a methodological point of view. Hence, research effort in the latest years has focused in overcoming these difficulties.

MHC genotyping: 10 years of improvement

Traditional MHC genotyping methods[1] were connected to serious limitations in terms of time and money (Babik 2010). In addition, due to the high number of alleles and MHC loci, applying these methods was inadequate and prone to underestimating MHC diversity (Promerová *et al.* 2012). These limitations have now been overcome with the advent of high-throughput sequencing technology, which provides a powerful tool to better understand what are the processes driving MHC evolution in wild populations at a large

[1] DGGE, SSCP, RSCA or cloning/Sanger sequencing

scale. The major advantage of this technology is that it enables the sequencing of thousands of samples at several MHC loci in parallel, and in a cost-effective manner (Wegner 2009). Although the benefit of such tools is indisputable, the central challenge of generating millions of sequences remains in defining reliable individual MHC genotypes by distinguishing true alleles from the artefactual ones. In this sense, compared to other gene systems, genotyping MHC genes is particularly difficult due to their highly dynamic evolution (occurrence of many duplicated genes, high rates of recombination and gene conversion, and high levels of polymorphism; see previous section).

The first problem is the conception of primers that independently capture all the genes of interest. Indeed, the recent duplication events or high level of recombination *sensu lato* can result in highly similar sequences between duplicated MHC loci (Hess & Edwards 2002), and, consequently, the design of locus-specific primers is often impossible (Babik 2010; Burri *et al.* 2014). Generally, researchers are obliged to co-amplify multiple MHC genes simultaneously, which can in turn increase the complexity of the subsequent genotyping given that the distinction between true alleles and artefacts is more difficult. Evidently, the level of difficulty of MHC genotyping will be proportional to the system complexity. This is particularly true in passerines, which usually have many functional duplicated loci, for example up to 16 MHC-I loci in great tits (Sepil *et al.* 2012).

The second major issue is that errors generated during PCR and sequencing may be difficult to discriminate from true alleles. Indeed, true MHC alleles can differ by only one base pair, and some alleles can be a melting pot of other alleles via intra- or inter-locus recombination. Consequently, PCR and sequencing errors, including indels, substitutions or chimeras, can easily mimic true alleles. In addition, PCR may produce amplification biases across alleles (Sommer *et al.* 2013; Burri *et al.* 2014), resulting in over-representation of certain alleles compared to others, which can generate errors in assessing the level of zygosity. Therefore, errors and biases need to be carefully detected and discarded by applying strict filtering procedures to gain reliable genotypes. In this context, a considerable effort has been done to i) understand the source of errors and how to reduce them, ii) develop post-sequencing methodologies for MHC genotyping (for instance validation threshold method, clustering method or relative sequencing depth method), and iii) evaluate what methodologies perform better for the acquisition of reliable data (Lenz & Becker 2008; Babik *et al.* 2009; Galan *et al.* 2010; Meglecz *et al.* 2011; Oomen *et al.* 2013; Pavey *et al.* 2013; Sommer *et al.* 2013; Lighten *et al.* 2014a; Lighten *et* *al.* 2014b; Stutz & Bolnick 2014; Ferrandiz-Rovira *et al.* 2015; Biedrzycka *et al.* 2016; Gaigher *et al.* 2016; Grogan *et al.* 2016; Sebastian *et al.* 2016). The development of all these new methodologies highlights that defining a consensus genotyping method is difficult, but rather the methodology to apply should be system-dependent (i.e. depending on the sequencing strategy, the molecular complexity of the species, etc.).

To date, Illumina MiSeq technology is the method of choice and has proven to bring a great value for MHC genotyping due to its high sequencing depth (Herdegen *et al.* 2014; Lighten *et al.* 2014b; Gaigher *et al.* 2016). In the first place, the high sequencing coverage provides a higher confidence in genotyping of species with high number of alleles (Biedrzycka *et al.* 2016), but most importantly, such a coverage enables a quantitative assessment of the abundance of alleles within amplicons[2], which before was impossible (Oomen *et al.* 2013; Lighten *et al.* 2014b; Gaigher *et al.* 2016). Finally, coupled with high sequencing coverage, family-based studies can significantly contribute to accurate MHC genotyping and insights into the MHC architecture, but until now have been rarely used (but see Gaigher *et al.* 2016).

To conclude, an accurate and reliable genotyping procedure is essential to capture MHC diversity in all individuals, which can in turn reduce bias in subsequent analyses and improve our knowledge of the mechanisms at the origin of MHC variability. This can be more easily achieved in study species presenting a relatively simple MHC organization, such as in raptors (Alcaide *et al.* 2007; Alcaide *et al.* 2008; Burri *et al.* 2008b).

General aims of the PhD

As outlined in the introduction, the MHC multigene family constitutes a remarkable system to understand the mechanisms that promote and maintain adaptive genetic diversity. However two main points remain challenging: i) providing a detailed MHC characterization to address poorly studied aspects of MHC diversity (such as haplotypic organization) in an evolutionary context, and ii) disentangling the relative role of evolutionary forces involved in MHC variation. To tackle these issues, more studies need to be conducted using family data in new model species, with simple MHC organizations, and at a large spatial scale.

[2] In the present case, a DNA sequence produced during artificial amplification (PCR)

The general aim of my thesis was to characterize the MHC diversity in a new model species and to investigate the evolutionary processes influencing the MHC genetic variation at different scales. In this thesis, I will address some of the most important topics in the MHC area, i.e. the MHC genotyping and the molecular evolution of MHC genes, the link between MHC diversity and individual fitness-related traits, and the spatial pattern of MHC diversity. Furthermore, within these areas I will address issues that have received little attention to date.

In chapter 1 and 2, my objective was to characterize the molecular basis of MHC genes. Providing a reliable and detailed molecular characterization of MHC diversity is the prerequisite to address the role of MHC variation in natural populations. For this purpose, an extensive family data set coupled with high throughput sequencing was used. Specifically, in chapter 1, I investigated to what extent family data can bring insights into the genetic architecture of MHC variation, by revealing how alleles are distributed among MHC loci (i.e. haplotypes), the degree of linkage among duplicate MHC loci, and the number of MHC copies present in each individual. In chapter 2, under the hypothesis that higher diversity grants a better immunity, I tested whether genetic diversity combined across MHC duplicate loci was higher than expected from randomly associated alleles. The issues of these two first chapters are of considerable value to perform accurate analyses in an evolutionary ecology context, and to understand how selective pressure by pathogens may influence genomic changes. However, until now, they were poorly studied due to the difficulties in accessing haplotypic data.

In chapter 3, I investigated whether there is a link between MHC diversity and individual fitness. For this purpose, I tested whether the diversity at MHC genes, or specific alleles, influences the strength of immunocompetence related traits. Such associations with immunocompetence have been rarely addressed, although a link is expected, given that the MHC is an essential component of antigen recognition and lymphocyte activation. In addition, the absence of support in such associations in the majority of previous studies, can result from a general lack of power (small sample sizes), or from general small-effect-size associations, an issue that I touched in this chapter.

Finally, in chapter 4, I explored what are the relative contributions of demographic events and selective processes to shape the MHC variability at a large spatial scale. Populations having undergone colonization history are expected to harbor a reduced genetic diversity, while such genetic erosion can limit the adaptive potential of populations, especially at MHC genes. Therefore, I also wanted to investigate whether selection can counteract the loss of diversity in populations with colonisation history. With this aim, I described the MHC diversity through a large continental scale, and contrasted the MHC diversity and structure with those of neutral markers.

The barn owl: a model species

The barn owl (*Tyto alba*; Tytonidae family) is a nocturnal predator bird with a wide range of variation in colour and size, with over 25 suggested subspecies (Del Hoyo *et al.* 2014). It constitutes one of the rare species of birds with a worldwide distribution, inhabiting open habitats in temperate, subtropical and tropical zones, including many islands (Fig. 2). It exhibits several characteristics that make it a suited model to study MHC diversity. First, given its distribution around the world over various environments, the barn owl presumably is exposed to a variety of pathogens which can result in variation at the MHC diversity. Second, compared to many other bird species, raptors (Alcaide *et al.* 2007; Alcaide *et al.* 2009), including the barn owl, exhibit a simple MHC organization (Burri *et al.* 2008b). Indeed, this species possesses only two functional MHC class IIβ genes that can be amplified specifically, and thus are distinguishable (Burri *et al.* 2008b).

In our group, an intensive effort has been done to collect samples from all over Europe, with a particular focus on Switzerland. As a result, the phylogeography and pattern of colonization history in the European barn owl is well understood (Antoniazza *et al.* 2010; Antoniazza *et al.* 2014; Burri *et al.* 2016) (see Appendix) and supports a ring-like population structure around the Mediterranean. First, barn owls colonized North

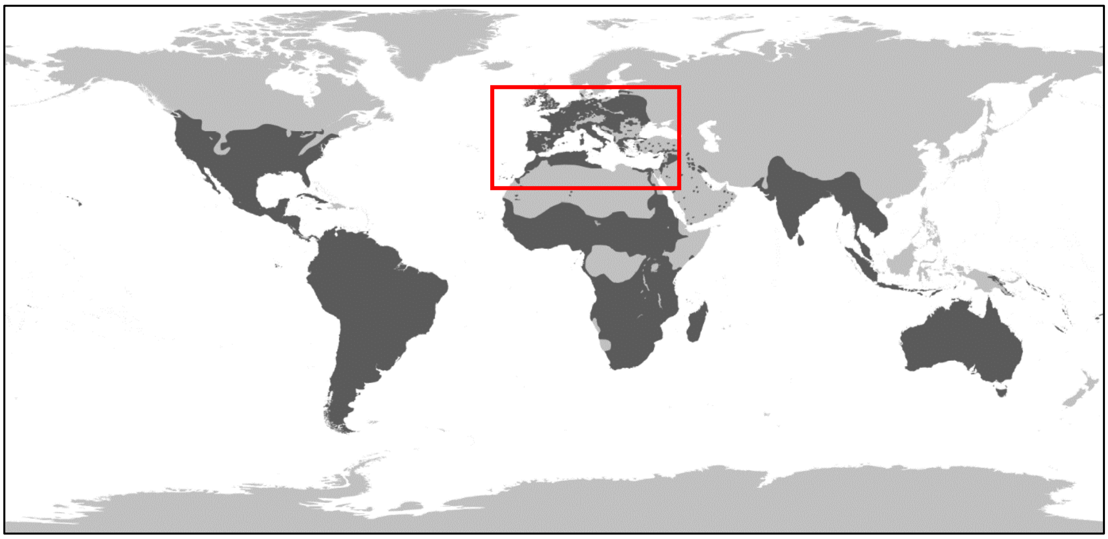


Figure 2: World distribution map of the barn owl. Red rectangle indicate the study area. From IUCN.

Africa and Iberia from the Middle East, and then Europe from the Iberian Peninsula after the last glaciation. In addition, this species colonized all islands within the Mediterranean Sea, as well as the Canary Islands in front of the Moroccan coast. Such colonization histories of Eastern Europe and Islands offer a unique opportunity to investigate how MHC diversity is affected in the face of such strong demographic events, and whether selection can counteract the loss of diversity, a topic that I will touch in Chapter 4.

At a smaller scale, in the western part of Switzerland, this species is well monitored since more than 20 years with around 300 currently installed nest boxes. Each year, in the period extending from April to July, nest boxes are controlled. When occupied by a family, breeding adults and their offspring are ringed, while morphological parameters, blood, and feather samples are collected. In addition, some immuno-ecological challenges were performed in this population, providing the possibility to test whether MHC diversity can influence the capacity to mount an efficient immune response in the barn owl (Chapter 3).

Therefore, all collected samples in Europe and Switzerland, plus the available molecular resources, including MHC class Iα, MHC class IIβ (Burri *et al.* 2008b; Gaigher *et al.* 2016) and neutral markers (Burri *et al.* 2008a), open highly promising perspectives for studying the above-mentioned aims of my thesis. To this end, a sampling strategy was designed encompassing the spatial scale, family data and fitness-related trait information from experimental data. The sampling comprises a total of around 1'300 barn owls (including around 400 throughout Europe and 900 from Switzerland). All these individuals were sequenced for both exon 3 of MHC class Iα genes and exon 2 of MHC class IIβ genes (i.e., where the PBR is located) with high-throughput sequencing technology.

References

Aguilar A, Roemer G, Debenham S*, et al.* (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 3490-3494.

Alcaide M, Edwards S, Cadahía L, Negro J (2009) MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. *Conservation Genetics*, **10**, 1349-1355.

Alcaide M, Edwards SV, Negro JJ (2007) Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. *Journal of Molecular Evolution*, **65**, 541-554.

Alcaide M, Edwards SV, Negro JJ, Serrano D, Tella JL (2008) Extensive polymorphism and geographical variation at a positively selected MHC class IIB gene of the lesser kestrel (*Falco naumanni*). *Molecular Ecology*, **17**, 2652-2665.

Antoniazza S, Burri R, Fumagalli L, Goudet J, Roulin A (2010) Local adaptation maintains clinal variation in melanin-based coloration of european barn owls (*Tyto alba*). *Evolution*, **64**, 1944-1954.

Antoniazza S, Kanitz R, Neuenschwander S*, et al.* (2014) Natural selection in a postglacial range expansion: the case of the colour cline in the European barn owl. *Molecular Ecology*, **23**, 5508-5523.

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, **17**, 179-224.

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

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

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

Bersaglieri T, Sabeti PC, Patterson N*, et al.* (2004) Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, **74**, 1111-1120.

Biedrzycka A, Sebastian A, Migalska M, Westerdahl H, Radwan J (2016) Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Molecular Ecology Resources*, In press.

Bjorkman PJ, Saper MA, Samraoui B*, et al.* (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, **329**, 512-518.

Blum JS, Wearsch PA, Cresswell P (2013) Pathways of Antigen Processing. *Annual Review of Immunology*, **31**, 443-473.

Borghans JAM, Beltman JB, De Boer RJ (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics*, **55**, 732-739.

Burri R, Antoniazza S, Gaigher A*, et al.* (2016) The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean. *Evolution*, **70**, 140-153.

Burri R, Antoniazza S, Siverio F*, et al.* (2008a) Isolation and characterization of 21 microsatellite markers in the barn owl (*Tyto alba*). *Molecular Ecology Resources*, **8**, 977-979.

Burri R, Niculita-Hirzel H, Roulin A, Fumagalli L (2008b) Isolation and characterization of major histocompatibility complex (MHC) class IIB genes in the Barn owl (Aves: *Tyto alba*). *Immunogenetics*, **60**, 543-550.

Burri R, Promerova M, Goebel J, Fumagalli L (2014) PCR-based isolation of multigene families: lessons from the avian MHC class IIB. *Molecular Ecology Resources*, **14**, 778-788.

Del Hoyo J, Collar NJ, Christie DA, Elliott A, Fishpool LDC (2014) *HBW and BirdLife International Illustrated Checklist of the Birds of the World* Lynx Edicions in association with BirdLife International, Barcelona, Spain and Cambridge, UK.

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

Ferrandiz-Rovira M, Bigot T, Allaine D, Callait-Cardinal MP, Cohas A (2015) Large-scale genotyping of highly polymorphic loci by next-generation sequencing: how to overcome the challenges to reliably genotype individuals? *Heredity*, **114**, 485-493.

Fitzpatrick MJ, Ben-Shahar Y, Smid HM*, et al.* (2005) Candidate genes for behavioural ecology. *Trends in Ecology & Evolution*, **20**, 96-104.

Gaigher A, Burri R, Gharib WH*, et al.* (2016) Family-assisted inference of the genetic architecture of major histocompatibility complex variation. *Molecular Ecology Resources*, **16**, 1353-1364.

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

Gaudieri S, Dawkins RL, Habara K, Kulski JK, Gojobori T (2000) SNP profile within the Human Major Histocompatibility Complex reveals an extreme and interrupted level of nucleotide diversity. *Genome Research*, **10**, 1579-1586.

Grogan KE, McGinnis GJ, Sauther ML, Cuozzo FP, Drea CM (2016) Next-generation genotyping of hypervariable loci in many individuals of a non-model species: technical and theoretical implications. *Bmc Genomics*, **17**, 204.

Hahn MW (2008) Toward a selection theory of molecular evolution. *Evolution*, **62**, 255-265.

Heath WR, Carbone FR (2001) Cross-presentation in viral immunity and self-tolerance. *Nature Reviews Immunology*, **1**, 126-134.

Hedrick PW (2001) Conservation genetics: where are we now? *Trends in Ecology & Evolution*, **16**, 629-636.

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

Herdegen M, Babik W, Radwan J (2014) Selective pressures on MHC class II genes in the guppy (*Poecilia reticulata*) as inferred by hierarchical analysis of population structure. *Journal of Evolutionary Biology*, **27**, 2347-2359.

Hess CM, Edwards SV (2002) The evolution of the major histocompatibility complex in birds. *Bioscience*, **52**, 423-431.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, **313**, 101-104.

Holling TM, Schooten E, van Den Elsen PJ (2004) Function and regulation of MHC class II molecules in T-lymphocytes: of mice and men. *Human Immunology*, **65**, 282-290.

Janeway CA, Travers P, Walport M, Shlomchik MJ (2001) *Immunobiology: the immune system in health and disease* New York: Garland Sience.

Jeffery KJM, Bangham CRM (2000) Do infectious diseases drive MHC diversity? *Microbes and Infection*, **2**, 1335-1341.

Jensen PE (2007) Recent advances in antigen processing and presentation. *Nature Immunology*, **8**, 1041-1048.

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

Kaufman J, Milne S, Gobel TWF, *et al.* (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, **401**, 923-925.

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

Kimura M (1968) Evolutionary rate at the molecular level. *Nature*, **217**, 624-626.

King JL, Jukes TH (1969) Non-Darwinian evolution. *Science*, **164**, 788-798.

Klein J, Sato A (2000) The HLA System. *New England Journal of Medicine*, **343**, 702-709.

Kreitman M (1996) The neutral theory is dead. Long live the neutral theory. *BioEssays*, **18**, 678-683.

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

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

Lewontin RC, Hubby JL (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, **54**, 595-609.

Lighten J, van Oosterhout C, Bentzen P (2014a) Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology*, **23**, 3957-3972.

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

Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 8597-8604.

Malinsky M, Challis RJ, Tyers AM*, et al.* (2015) Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, **350**, 1493-1498.

Meglecz E, Piry S, Desmarais E*, et al.* (2011) SESAME (SEquence Sorter & AMplicon Explorer): genotyping based on high-throughput multiplex amplicon sequencing. *Bioinformatics*, **27**, 277-278.

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

Mundy NI (2005) A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 1633-1640.

Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5268-5273.

Neefjes J, Jongsma MLM, Paul P, Bakke O (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, **11**, 823-836.

Nei M (2005) Selectionism and Neutralism in molecular evolution. *Molecular Biology and Evolution*, **22**, 2318-2342.

Nei M, Suzuki Y, Nozawa M (2010) The neutral theory of molecular evolution in the genomic era. *Annual Review of Genomics and Human Genetics*, **11**, 265-289.

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

Oleksyk TK, Smith MW, O'Brien SJ (2010) Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 185-205.

Oomen RA, Gillett RM, Kyle CJ (2013) Comparison of 454 pyrosequencing methods for characterizing the major histocompatibility complex of nonmodel species and the advantages of ultra deep coverage. *Molecular Ecology Resources*, **13**, 103-116.

Pavey SA, Sevellec M, Adam W*, et al.* (2013) Nonparallelism in MHCII diversity accompanies nonparallelism in pathogen infection of lake whitefish (*Coregonus clupeaformis*) species pairs as revealed by next-generation sequencing. *Molecular Ecology*, **22**, 3833-3849.

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

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

Piertney SB, Webster LMI (2010) Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach. *Genetica*, **138**, 419-432.

Poelstra JW, Vijay N, Bossu CM*, et al.* (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, **344**, 1410-1414.

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

Promerová M, Babik W, Bryja J*, et al.* (2012) Evaluation of two approaches to genotyping major histocompatibility complex class I in a passerine-CE-SSCP and 454 pyrosequencing. *Molecular Ecology Resources*, **12**, 285-292.

Qiu Q, Zhang G, Ma T*, et al.* (2012) The yak genome and adaptation to life at high altitude. *Nature Genetics*, **44**, 946-949.

Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution*, **55**, 1095-1103.

Robinson J, Halliwell JA, Hayhurst JD*, et al.* (2015) The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Research*, **43**, D423-D431.

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

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

Slade RW, McCallum HI (1992) Overdominant vs. frequency-dependent selection at MHC loci. *Genetics*, **132**, 861-864.

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

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

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

Spurgin LG, van Oosterhout C, Illera JC*, et al.* (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, **20**, 5213-5225.

Stutz WE, Bolnick DI (2014) Stepwise threshold clustering: a new method for genotyping MHC loci using next-generation sequencing technology. *PLoS One*, **9**, e100587.

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

Tishkoff SA, Reed FA, Ranciaro A*, et al.* (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, **39**, 31-40.

van Tienderen PH, de Haan AA, van der Linden CG, Vosman B (2002) Biodiversity assessment using markers for ecologically important traits. *Trends in Ecology & Evolution*, **17**, 577-582.

Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting natural selection in genomic data. *Annual Review of Genetics*, **47**, 97-120.

Wakeland E, Boehme S, She J*, et al.* (1990) Ancestral polymorphisms of MHC class II genes: divergent allele advantage. *Immunologic Research*, **9**, 115-122.

Wegner KM (2009) Massive parallel MHC genotyping: titanium that shines. *Molecular Ecology*, **18**, 1818-1820.

Westerdahl H (2007) Passerine MHC: genetic variation and disease resistance in the wild. *Journal of Ornithology*, **148**, S469-S477.

Yi X, Liang Y, Huerta-Sanchez E*, et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, **329**, 75-78.

Zagalska-Neubauer M, Babik W, Stuglik M*, et al.* (2010) 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *Bmc Evolutionary Biology*, **10**, 395.

Zhan X, Pan S, Wang J*, et al.* (2013) Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nature Genetics*, **45**, 563-566.

Zhang G, Li C, Li Q*, et al.* (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, **346**, 1311-1320.

Chapter 1

Family-assisted inference of the genetic architecture of major histocompatibility complex variation

A. Gaigher, R. Burri, W. H. Gharib, P. Taberlet, A. Roulin & L. Fumagalli

*This chapter is published in **Molecular Ecology Resources (2016) 16, 1353-1364**.*

MOLECULAR ECOLOGY
RESOURCES

Molecular Ecology Resources (2016) **16**, 1353–1364 doi: 10.1111/1755-0998.12537

Family-assisted inference of the genetic architecture of major histocompatibility complex variation

A. GAIGHER,* R. BURRI,† W. H. GHARIB,‡ P. TABERLET,§¶ A. ROULIN* and L. FUMAGALLI*

**Laboratory for Conservation Biology, Department of Ecology and Evolution, University of Lausanne, Biophore, Lausanne CH-1015, Switzerland, †Department of Evolutionary Biology, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden, ‡Interfaculty Bioinformatics Unit, University of Bern, CH-3012 Bern, Switzerland, §CNRS, Laboratoire d'Ecologie Alpine (LECA), 38000 Grenoble, France, ¶ Laboratoire d'Ecologie Alpine (LECA), University of Grenoble Alpes, 38000 Grenoble, France*

Abstract

With their direct link to individual fitness, genes of the major histocompatibility complex (MHC) are a popular system to study the evolution of adaptive genetic diversity. However, owing to the highly dynamic evolution of the MHC region, the isolation, characterization and genotyping of MHC genes remain a major challenge. While high-throughput sequencing technologies now provide unprecedented resolution of the high allelic diversity observed at the MHC, in many species, it remains unclear (i) how alleles are distributed among MHC loci, (ii) whether MHC loci are linked or segregate independently and (iii) how much copy number variation (CNV) can be observed for MHC genes in natural populations. Here, we show that the study of allele segregation patterns within families can provide significant insights in this context. We sequenced two MHC class I (MHC-I) loci in 1267 European barn owls (*Tyto alba*), including 590 offspring from 130 families using Illumina MiSeq technology. Coupled with a high per-individual sequencing coverage (~3000×), the study of allele segregation patterns within families provided information on three aspects of the architecture of MHC-I variation in barn owls: (i) extensive sharing of alleles among loci, (ii) strong linkage of MHC-I loci indicating tandem architecture and (iii) the presence of CNV in the barn owl MHC-I. We conclude that the additional information that can be gained from high-coverage amplicon sequencing by investigating allele segregation patterns in families not only helps improving the accuracy of MHC genotyping, but also contributes towards enhanced analyses in the context of MHC evolutionary ecology.

Keywords: adaptive genetic diversity, birds, copy number variation, gene duplication, immunogenetics, major histocompatibility complex

Received 1 December 2015; revision received 7 April 2016; accepted 18 April 2016

Introduction

Understanding the mechanisms that promote and maintain adaptive genetic diversity is a central issue in evolutionary and conservation biology. Due to their known direct link with individual fitness and their exceptional levels of diversity, genes of the major histocompatibility complex (MHC) are placed among the best candidates to study adaptive genetic diversity (Bernatchez & Landry 2003). Classical MHC genes encode cell surface proteins involved in the activation of the adaptive immune response by binding antigens derived from pathogens (Klein & Sato 2000). It has long been suggested that pathogens are the major selective force maintaining MHC diversity in natural populations (Apanius *et al.* 1997; Jeffery & Bangham 2000; Spurgin & Richardson 2010). Typical issues addressed in MHC evolutionary ecology are therefore concerned with the link between MHC diversity and disease resistance and fitness-related life-history traits, the spatial and temporal distribution of MHC diversity (reviewed in Sommer 2005; Piertney & Oliver 2006; Spurgin & Richardson 2010), and MHC-related mate choice (reviewed in Penn & Potts 1999; Milinski 2006; Kamiya *et al.* 2014).

Due to their high evolutionary dynamics—including gene duplication and loss, high rates of recombination and gene conversion (Hess & Edwards 2002; Nei & Rooney 2005) and long-term action of balancing selection (Klein *et al.* 1998)—the isolation, characterization and genotyping of MHC genes, which constitute first crucial steps in the study of MHC evolutionary ecology, often represent a challenging task (Babik 2010; Burri *et al.* 2014). In taxa like birds and fish, high rates of duplication,

Correspondence: Arnaud Gaigher, Fax: +4121-692-42-65; E-mail: arnaud.gaigher@unil.ch

recombination and gene conversion often render the specific amplification of single MHC loci impossible. Instead, often multiple loci are amplified simultaneously (Lenz *et al.* 2009; Michel *et al.* 2009; Zagalska-Neubauer *et al.* 2010; Promerová *et al.* 2012; Sepil *et al.* 2012), posing multiple challenges to the characterization of the architecture of MHC variation and to downstream analyses. The number of MHC loci present in a species in such cases has to be approximated from the number of alleles detected per individual, an approach that only works if alleles are not shared among distinct MHC loci. However, determining whether alleles segregate at multiple loci and estimating the extent to which MHC loci are linked are difficult. Therefore, evidence for copy number variation (CNV) of MHC genes has also been to date rather speculative for many species (e.g. Zagalska-Neubauer *et al.* 2010; Promerová *et al.* 2012; Sepil *et al.* 2012; Sommer *et al.* 2013). Ultimately, the incomplete information on (i) the sharing of alleles among loci, (ii) the linkage among coamplified loci and (iii) the number of MHC copies present in each individual implies uncertainties or biased genotypes and diversity estimates that are carried over to analyses relating MHC variation to individual fitness. Resolving these issues is therefore of central importance to MHC evolutionary ecology.

High-throughput sequencing technology has moved the field a large leap forward during the past decade (Babik *et al.* 2009; Wegner 2009; Galan *et al.* 2010), and the ultra-high coverage provided by Illumina's MiSeq technology has recently enabled significant improvements in terms of costs and genotyping accuracy (Herdegen *et al.* 2014; Lighten *et al.* 2014b). Importantly, the high coverage provided by MiSeq technology enables a quantitative assessment of the abundance of alleles within amplicons (Oomen *et al.* 2013; Lighten *et al.* 2014b) in the absence of amplification biases (but see Sommer *et al.* 2013; Burri *et al.* 2014).

Here, we suggest that together with family data, the high sequencing coverage can be taken full advantage of to help resolving the above-mentioned issues. We isolated two MHC-I loci in the European barn owl (*Tyto alba*) and sequenced 1306 individuals including extensive family data from Switzerland using Illumina MiSeq technology. We show how family data with high sequencing coverage provide insights into the architecture of MHC variation in barn owls, including the confident assessment of CNV, and to gain confidence in the accuracy of MHC genotypes.

Material and methods

Sampling and DNA extraction

Tissue samples (feathers, blood or muscle) were collected from 1306 barn owls from 14 populations throughout Europe, one population from the Middle East, one population on the eastern Canary Islands (Lanzarote and Fuerteventura) and three individuals of *T. (a.) pratincola* from the USA (Table S1, Supporting Information). These samples included 800 individuals from 132 putative families from a Swiss population collected between 1997 and 2003. Each family was composed of two parents and on average four (range 1–17) offspring. To ensure that family-based analyses were not affected by erroneous family relationships caused for instance by extra-pair paternity that can occur in barn owls (Henry *et al.* 2013), all individuals were genotyped at 10 microsatellite markers (multiplex sets 3 and 4 in Burri *et al.* (2016)) and parent–offspring relationships were verified using CERVUS (Kalinowski *et al.* 2007). Family relationships confirmed by positive LOD scores were considered confident.

DNA was extracted using a BioSprint 96 extraction robot and the BioSprint 96 DNA blood kit, or using the DNeasy blood and tissue kit following the manufacturer's instructions (Qiagen, Hilden, Germany).

MHC class I primer development, PCR amplification and cloning

In this study, we investigated exon 3 of MHC class Iα (MHC-I) genes, which together with exon 2 encodes for the highly polymorphic sequence encoding the peptide-binding region (PBR, Bjorkman *et al.* 1987; Strandh *et al.* 2011). To develop primers that amplify the exon 3 of MHC class I, we used primers MHCI-ex2F and MHCI-ex4R (Alcaide *et al.* 2009) designed to isolate genomic MHC class I sequences in birds. The resulting sequences were aligned to avian MHC-I sequences available on the GenBank database, and primers Tyal-MHCI-int2F (5′-GCT CCC CGG ACG TGT TTC AG-3′) and Tyal-MHCI-int3R (5′-GGC ACA CGC ACC CGC AGC AC-3′) were designed within the most conserved regions of the exon 3-flanking introns. These primers amplify a 304-bp fragment (excluding primers), including the full MHC class Iα exon 3 sequence and a small part of the introns. To obtain an estimate of the number of coamplified loci, we first cloned eight individuals. Cloning was performed as follows: in line with recommendations in Burri *et al.* (2014), we amplified MHC-I fragments on Applied Biosystems Veriti® thermocyclers in a final volume of 25 μL containing 1× buffer Gold, 2.0 mM MgCl2 Gold, 1× Q solution (Qiagen), 0.2 mM dNTPs, 0.25 μM each primer and 0.5 U AmpliTaq Gold (Applied Biosystems, Switzerland). PCR conditions included an initial denaturation step at 95 °C for 10 min, 30 cycles of denaturation at 95 °C for 30 s, annealing at 68 °C for 45 s and extension at 72 °C for 45 s. A final step at 72 °C for 7 min was used to complete primer extension. PCR products were checked on agarose gels and then purified using the MinElute PCR purification kit (Qiagen). PCR products were then cloned using the pGEM®-T Easy Vector System (Promega) protocol followed by transformation following the manufacturer's instructions, but using MAX Efficiency® DH5α™ Competent Cells (Invitrogen). Thirty-two colonies per individual were amplified using the SP6 and T7 vector primers with PCR conditions including an initial denaturation step at 95 °C for 5 min, 30 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for 30 s and a final elongation at 72 °C for 10 min. Twenty-four clones per individual with expected insert size were purified using QIAquick PCR purification kits (Qiagen) and sequenced in both directions by Microsynth (Balgach, Switzerland). Cloning showed that these primers amplify two MHC-I loci simultaneously.

Preparations for Illumina MiSeq sequencing

We used the Illumina MiSeq protocol to sequence the isolated MHC-I loci in 1477 samples (including replicates, see Table S1, Supporting Information). Primers Tyal-MHCI-int2F and Tyal-MHCI-int3R were modified by adding a random label composed by three nucleotides (NNN) and a barcode sequence to the 5′-end. The three random nucleotides (NNN) were added to increase the diversity, because the MiSeq technology requires nucleotide diversity to properly identify clusters on the flow cell. Barcodes consisted of eight base pairs differing by at least five base pairs. A total of 32 forward primers and 24 reverse primers with specific barcodes were designed, offering 768 unique combinations, similar to Galan *et al.* (2010). Due to the large number of individuals used in this study, we divided samples among two libraries (A and B) with the same combination of barcodes. PCR reactions were carried out following the protocol previously described, but with 35 cycles for samples with very low DNA concentration.

All PCR products were checked on 1.5% agarose gel. PCR purification was performed using the QIAquick or MinElute PCR purification kit (Qiagen) by pooling eight PCR products of similar amplification intensity (evaluated from agarose gels) per column. Before final pooling, DNA concentrations of purified PCR products were quantified using a Qubit® 2.0 fluorometer (Life Technologies). Prior to sequencing, Illumina adaptors tagged with a specific index for the two libraries were added to amplicons by ligation. Both libraries were simultaneously sequenced with a 250-bp paired-end MiSeq protocol (Illumina) in one full run. Reliability of the sequencing was evaluated by including 171 replicated samples (13%) from independent PCR in the same run and by cloning eight individuals.

Raw data processing

We applied five consecutive filtering steps to the raw data to remove the lowest quality sequences (steps 1 and 2, performed with the OBITOOLS software package (Boyer *et al.* 2016)), rare artefactual variants (steps 3 and 4) and samples with too low coverage (step 5). A glossary of terms used for data processing is provided in the Table S2 (Supporting Information). Step 1: forward and reverse reads were assembled based on at least 20-bp overlap. Step 2: only sequences with a maximum of two errors within the primer sequence and none within the barcodes were conserved. Then, identical sequences were grouped according to individual barcode combinations, and both libraries were concatenated. Step 3: because the mean coverage was very high (3229 sequences per sample), all variants with <100× coverage in whole data set were discarded. Step 4: all variants with the number of sequences per individual never exceeding 20 were removed. Step 5: due to low amplification intensity, 26 samples were covered by <100 sequences (mean = 12.8 sequences), could not be genotyped reliably and were therefore excluded. At this step, we also checked the barcode combinations' consistency with 59 unused specific barcode combinations, which permitted to evaluate the erroneous sequence-sample associations. We found on average five sequences per unused combination (range: 0–14 sequences), corresponding to a false combination rate of 0.16%. Our result suggests that the impact of these false combinations, which can be generated by artificial tag jumping (Schnell *et al.* 2015), will have an negligible impact on genotyping quality.

MHC genotyping

We first applied two published genotyping methods on all individuals: (i) the threshold procedure developed by Galan *et al.* (2010) and (ii) the degree of change (DOC) method proposed by Lighten *et al.* (2014b). To establish the reliability of these methods, we subsequently benchmarked them against the genotypes obtained from family data. For the first method, we applied Galan's T2 threshold, which should separate artefacts from true alleles based on the frequency of a variant within individuals (*Fij*). We opted for a threshold of 8% because allele frequency distributions showed a break at around 8%, which appeared to be the critical zone separating true alleles from artefacts (Fig. 1). All variants below this threshold were rejected as artefacts. For variants above this threshold, we further removed variants with a chimeric composition comprised of parental true alleles (chimera was identified visually as mixed sequences from two different alleles). The DOC method permits the identification and estimation of the number of alleles (*Ai*) per individual. This procedure is based on the break point in sequencing coverage between variants within each individual and avoids choosing a subjective threshold to separate true alleles from artefacts. In this procedure, variants are sorted top-down by coverage, followed by the calculation of the coverage break point (DOC statistic) around each variant. The variant with the highest DOC value (i.e. with the greatest break point) is assumed to be the last true allele (see Lighten *et al.* 2014b).

Because segregation patterns within families can be used to verify genotypes, in a second phase, we genotyped all individuals from the Swiss families. After excluding individuals incorrectly attributed to families (17 individuals, based on microsatellite analysis) and after filtering, in total, our family data cover 765 barn owls from 130 families (data corresponding to the full family data set). For the family-based genotyping, we first removed all variants with a relative coverage within an individual (*Fij*; Galan *et al.* 2010) lower than 0.03. The genotypes of the individuals from the same family were then defined both based on the concordance between parental and offspring variants, and by considering ratios of the coverage of variants found within individuals. Offspring variants with high coverage that were shared with at least one parent were categorized as true alleles. Variants detected only in an offspring but not in its parents were considered artefacts. More complex situations occurred when (i) a variant was detected in a parent but not in offspring, or, rarely, when (ii) both a parent and offspring shared the same low-coverage variant. In the first case, the variant does not necessarily constitute an artefact, because it may simply not have been transmitted. However, it was considered an artefact if it showed both low coverage and could be explained by chimera, indels or single substitution errors from parental true alleles of higher coverages. In the second case, the variant can be categorized either as a true allele with downward-biased amplification, or as an artefact. The

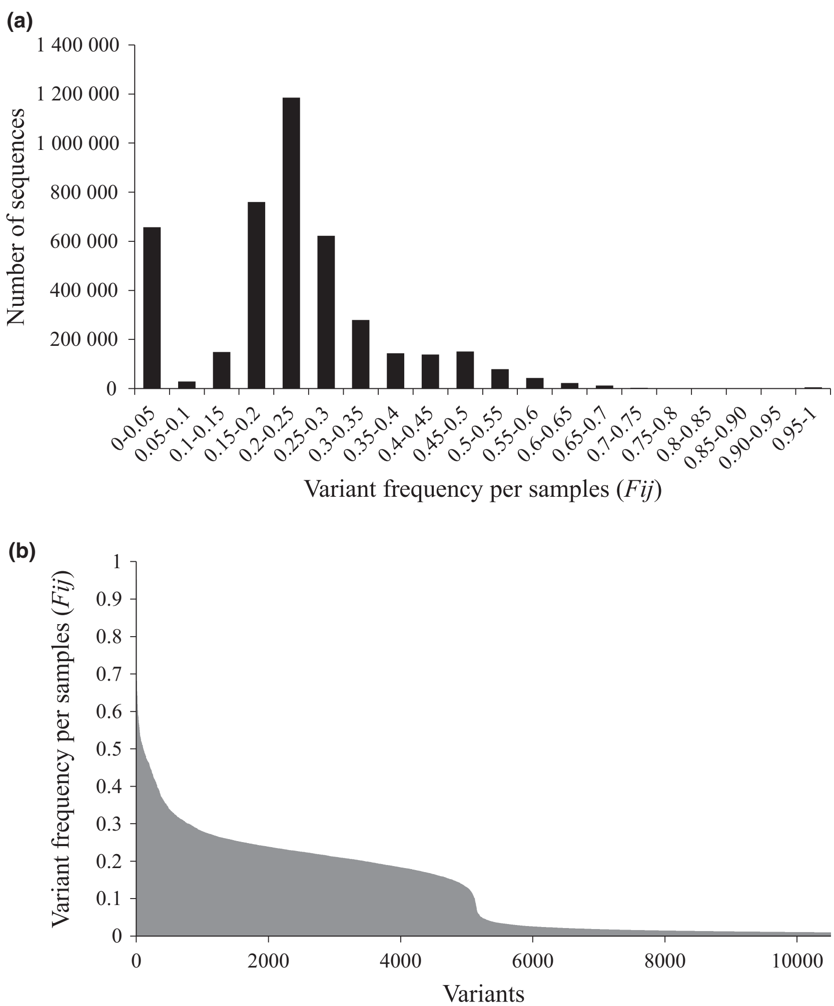


Fig. 1 Variant frequencies. (a) Histogram showing the distribution of the frequency of each variant j within each sample *i* (*Fij*). (b) Distribution of variants sorted by top-down *Fij*. For better readability, only variants occurring in samples with a frequency superior to 0.01 are shown.

variant was classified an artefact if a chimeric pattern or single substitution error from parental true alleles of higher coverages was identified for all individuals sharing the variant.

We then made use of the ratios of sequencing coverage observed for each allele within individuals to determine how many copies of each allele an individual carries. For two coamplified loci, we can expect the following ratios: (i) individuals with four different alleles should display about equal coverage for each allele (1:1:1:1). (ii) With three different alleles, one should be present at double the coverage of the other two (2:1:1). (iii) With two alleles, both should display the same coverage (2:2), unless an allele is shared between loci and present thrice, in which case a ratio of 3:1 can be observed. Finally, (iv) for double homozygotes, amplicons should be comprised of a single allele.

The amplification ratio of alleles can be significantly biased during PCR either by random process and/or biased amplification efficiency among alleles (Sommer *et al.* 2013; Burri *et al.* 2014). Prior to genotyping, we therefore evaluated the allele amplification efficiency following Sommer *et al.* (2013). Analyses were performed only with double heterozygotes to avoid biases related to alleles present in more than one copy within an individual. Analyses were performed using R with the script provided by Sommer *et al.* (2013). The standardized amplification efficiency was calibrated with Tyal-UA*01. Our results highlighted relatively weak variation in the allele amplification efficiency (Fig. S1, Supporting information), justifying the use of the above described coverage ratios to call genotypes. Segregation patterns within families were used to confirm genotypes and evaluate the ability of our procedure to estimate the number of allele copies for each individual.

Genetic architecture of MHC-I variation

Family data were used to estimate the linkage of the two loci based on the pattern of allelic segregation. Only large families were used, that is 40 families with a minimum of five offspring, comprising a total of 347 barn owls including 267 offspring (mean number of offspring per clutch ± SD: 6.7 ± 2.5). Because parents with two copies of a same allele (2:1:1) are uninformative regarding the occurrence of recombination, these 40 families comprised only ones in which both parents were double heterozygotes (1:1:1:1).

With free homologous recombination among loci, in offspring of two double heterozygote parents, we expect to observe a maximum of 16 different genotypes (Fig. 2a). If in contrary loci are tightly linked, only four different genotypic combinations should be observed, because alleles at the two linked loci are usually transmitted together (Fig. 2b). Following this logic, we inferred the frequency of recombinant gametes in our family data, which is indicative of the amount of linkage of the two loci.

In addition, we used the full family data set to evaluate the possibility of assigning alleles to one of the MHC-I loci and thus obtain complete genotypes for each locus ('complete genotypes' implying that both the exact number of alleles and the locus on which they occur are known). Complete genotypes can in principle be reconstructed using two alternative approaches that make different assumptions. (i) With homologous recombination among loci, recombinant gametes will occur. However, of the alleles located on the same locus within an individual, only one at a time will be transmitted to an offspring, and never both together (example Fig. 2a; for parent 1, alleles 01/02 and 03/04 are on the same locus, respectively). (ii) In case of strong linkage between loci, alleles can be attributed to loci under the assumption that no alleles are shared between loci. Under this assumption, we aligned all haplotypes derived from segregation patterns within families such that the same allele comes to lie always at the same locus (Fig. 3). For individuals with a 2:1:1 coverage pattern, the allele occurring twice is automatically considered to be located

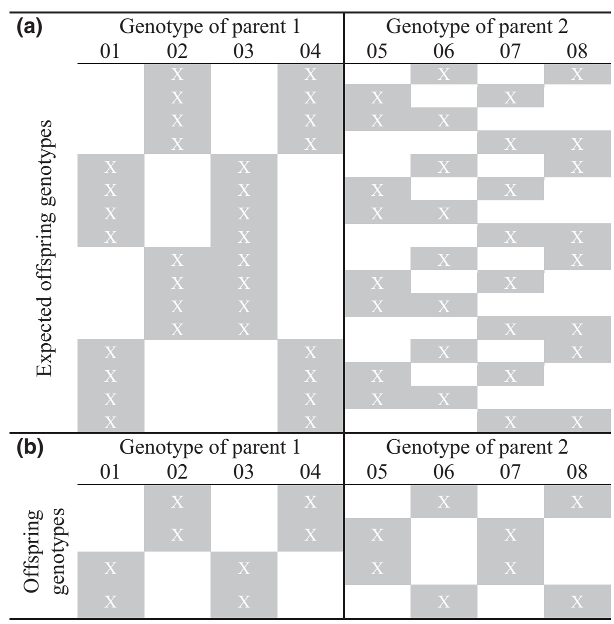


Fig. 2 General outline of linkage between loci using family data. Alleles of parents were named from 01 to 08. Grey and white shading indicate the presence and absence of alleles in offspring, respectively. (a) Family with expected offspring genotypes under free homologous recombination (no linkage), that is producing 16 different genotypes. (b) The same family as in (a) but with the expected pattern in the case of linkage among loci, that is only four observed genotypic combinations.

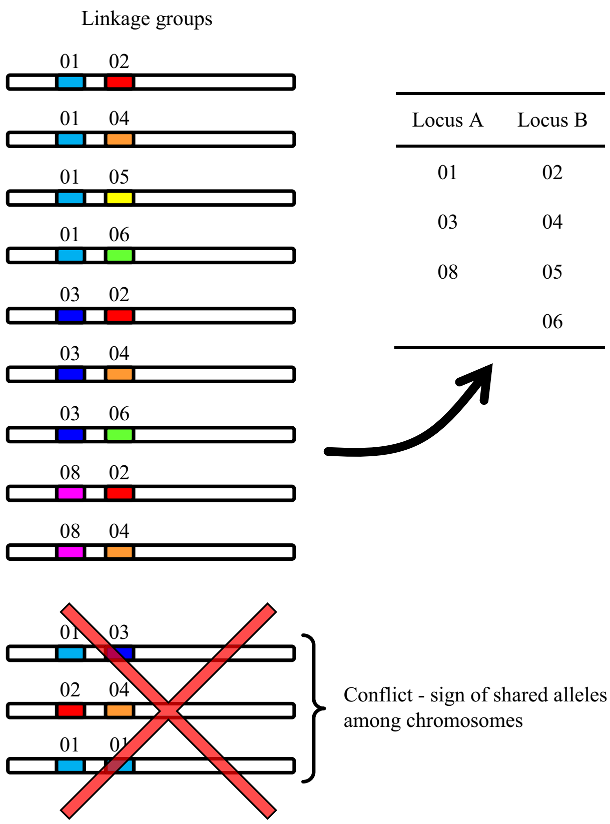


Fig. 3 Allelic linkage groups used to assign alleles to one of the MHC-I loci. To facilitate reading, all allele names (i.e. Tyal-UA*01) were reduced to the allele number, and each colour refers to a specific allele. The cross indicates linkage groups conflicting with the assignment of alleles to loci.

on one locus and the two other alleles are attributed to the second locus. Cases in which haplotypes cannot be aligned to others without one of the alleles being present at both loci provide an estimate of the minimum number of alleles shared between loci.

Sequence analysis

To verify that the studied MHC-I loci represent classical MHC-I genes, we performed a number of population genetic and phylogenetic analyses. Alignment of the identified alleles was performed using CLUSTALW (Thompson *et al.* 1994) implemented in MEGA 5 (Tamura *et al.* 2011). Nucleotide diversity (π) was estimated in DNASP (Librado & Rozas 2009) and compared among three data partitions: (i) the entire exon 3, (ii) codons of the PBR exclusively and (iii) codons of the non-PBR exclusively. PBR codons were defined from Human HLA and Chicken BF (Bjorkman *et al.* 1987; Wallny *et al.* 2006). To investigate the phylogenetic relationships among barn owl MHC-I alleles, we built a neighbor-net network using SPLITSTREE 4 (Huson & Bryant 2006). The presence of recombination was evaluated using the Φ_w test (Bruen *et al.* 2006) in SPLITSTREE 4, and the minimal number of historical recombination events, R_M (Hudson & Kaplan 1985), was inferred using the four-gamete test implemented in DNASP. To infer footprints of positive selection, we estimated maximum-likelihood site models using CODEML implemented in the PAML v4.7 (Yang 2007). Two likelihood ratio tests of positive selection were carried out comparing models M1a with M2a and models M7 with M8 (Yang *et al.* 2005). Models M1a and M7 are neutral, while models M2a and M8 allow for a proportion of sites under positive selection. The Bayes empirical Bayes (BEB) procedure was used to identify sites under positive selection. The phylogenic tree used to run CODEML was obtained from MRBAYES v3.2.3 (Ronquist & Huelsenbeck 2003) based on the GTR+Γ model. To test whether signals of selection were sensitive to tree topology, we used the best tree as input and then reran the CODEML analysis with nine other topologies randomly chosen from the posterior distribution of topologies. All analyses were focused on alleles from European populations.

Results

Illumina sequencing and MHC-I characterization

Considering all 1306 samples and 171 replicates, we obtained 12 563 792 reads before filtering. After filtering, 4 282 708 sequences were retained (Table S3, Supporting Information). The *Fij* distribution with a break at around 8%, that separates true alleles from artefacts, indicates that artefacts were 5.3 times less sequenced than true alleles (Fig. 1). Of the 1124 initially detected variants, 159 variants were attributed to true alleles; remaining variants predominantly represented chimeras, few variants with single substitution errors, or single-base pair frameshifts in a homopolymer region directly following the primer sequence. Of initially 1477 samples (including replicates), 1437 were genotyped with confidence based on a final average coverage of 2946 sequences per individual (SD: 1081) (Table S4, Supporting Information). The highest relative coverage of an artefact within an individual was 0.091, while the lowest of a true allele was 0.076. Among the 1267 individuals, 0.2%, 6.2%, 29% and 64.6% displayed one, two, three and four different alleles, respectively (Table S4, Supporting Information). Genotypic diversity was very high, with a total of 812 allelic compositions (Table S4, Supporting Information). All replicates displayed identical genotypes except in one case for which the individual was removed from subsequent analyses. Despite differences in sequencing depth between replicates (mean difference ± SD: 1048 ± 880 reads), the relative coverage between same alleles from two replicates was very close and differed on average by only 3% (SD: 2.5%), highlighting the good repeatability of our genotyping. Finally, cloning of eight individuals produced the same alleles identified with Illumina MiSeq data for these individuals, but it would need more clones to really confirm some underrepresented alleles.

Considering only the individuals across Europe for the entire exon 3, we detected 146 alleles that translated into 132 amino acid sequences (GenBank Accession nos KX189198–KX189343). Sequence analyses showed that the two MHC-I loci exhibited the typical characteristics of functional MHC-I genes, that is a high sequence diversity, evidence for positive selection localized in the PBR (Table S5, Fig. S2, Supporting information) and footprints of recombination sensu lato [R_M = 14 and significant Φ_w test (P < 0.01); Fig. S3, Supporting information].

Comparison and validation of MHC-I genotyping procedures

The 8% T2 threshold and the DOC methods yielded genotypes which were highly congruent with the ones obtained from the full family data set analysis, with more than 99% and 97% of identical genotypes, respectively. Accordingly, genotypes retrieved with the 8% T2 threshold and DOC methods were identical in 97% of the individuals. Manual checking of nonmatching genotypes obtained by the DOC method revealed that these were due to a high proportion of artefacts, or a very high coverage of one true allele compared to the others, inflating the number of false alleles or deflating the proportion of true alleles, respectively. For the threshold method, a 3% T2 threshold was more appropriate to avoid rejecting true alleles that occurred below a relative coverage of 8%. After taking this into account, both methods yielded identical genotypes with the family-based ones as well as for replicates. Consequently, we were very confident to apply both methods to samples outside the families.

Tandem architecture of the barn owl MHC-I and allele sharing among loci

We used the pattern of allele segregation in 40 large families (267 offspring and thus 534 gametes) to detect recombination events, and thereby indirectly infer the extent of linkage between loci. In line with expectations under strong linkage of the two loci, our analyses revealed almost exclusively four genotypes per family (i.e. 38 of 40 families) (Fig. 4a). Of 534 observed gametes, only two (0.37%) appeared to be parental recombinants (Fig. 4b and Fig. S4, Supporting information), suggesting that the level of recombination is very low between the two MHC-I loci and that they are situated in tandem.

Furthermore, from the full family data set, in more than 95% of cases, it was possible to place an allele exclusively on the same locus (as outlined in Fig. 3), indicating that each allele might be restricted to one of the two loci. However, the observation of the simultaneous transmission of two copies of the same allele in 15 individuals from five families demonstrates the sharing of at least two alleles among loci (Fig. 5b).

Because of both the low recombination level and allele sharing among loci, the family data did not enable us to associate alleles to particular loci.

CNV in the barn owl MHC-I

We capitalized on the combination of high sequencing coverage and allele segregation in families to determine and confirm the number of allele copies for each individual. Of 765 individuals (from the full family data set),

(a)	Genotype of parent 1				Genotype of parent 2			
	05	22	23	71	01	04	15	16
	(0.29)	(0.23)	(0.25)	(0.23)	(0.24)	(0.29)	(0.29)	(0.18)
Offspring genotypes		0.27		0.23		0.25		0.25
		0.22		0.30		0.25		0.23
		0.26		0.24	0.22		0.28	
		0.29		0.16	0.31		0.24	
		0.25		0.20	0.30		0.25	
	0.25		0.26		0.25		0.24	
	0.26		0.26		0.24		0.24	
	0.27		0.24		0.21		0.28	
	0.27		0.23		0.24		0.26	
	0.27		0.23		0.22		0.28	
	0.28		0.24		0.23		0.25	
	0.27		0.21			0.24		0.28

(b)	Genotype of parent 1				Genotype of parent 2			
	04	10	12	29	01	11	15	44
	(0.24)	(0.23)	(0.32)	(0.21)	(0.24)	(0.26)	(0.24)	(0.26)
Offspring genotypes	0.24			0.18		0.30		0.28
	0.26			0.17		0.30		0.27
	0.27			0.17		0.26		0.30
	0.26			0.17		0.31		0.26
	0.25			0.18		0.28		0.29
	0.26			0.19		0.27		0.28
		0.24	0.29			0.21		0.26
		0.20	0.28			0.25		0.27
		0.25	0.24			0.20		0.31
		0.25	0.25		0.23		0.27	
	0.26		0.26		0.22		0.26	

Fig. 4 Genetic linkage between loci estimated using family data. To facilitate the reading, all allele names (i.e. Tyal-UA*01) were reduced to the allele number. Grey and white shading indicate the presence and absence of alleles in offspring, respectively. Values below allele names and in shaded background indicate the relative coverage of the allele within the individual (the allele's read coverage divided by the sum of coverage over all alleles). These two examples illustrate cases in which both parents are double heterozygotes (1:1:1:1) without shared alleles between them. (a) Family with only the four truly observed genotypic combinations in 12 offspring. (b) Family composed of 11 offspring, in which recombination has been detected. Black backgrounds indicate recombinant haplotypes.

(a)

Offspring genotypes	Genotype of parent 1: 03 (0.55)	03	06 (0.24)	54 (0.21)	Genotype of parent 2: 14 (0.28)	16 (0.19)	34 (0.24)	59 (0.29)	Observed pattern
		0.28		0.25			0.21	0.26	1:1:1:1
		0.27		0.24			0.20	0.29	1:1:1:1
		0.24		0.30			0.22	0.24	1:1:1:1
		0.27		0.27	0.27	0.19			1:1:1:1
		0.28		0.28	0.25	0.19			1:1:1:1
		0.26		0.24	0.27	0.23			1:1:1:1
	0.32		0.26		0.22	0.20			1:1:1:1
	0.27		0.29				0.19	0.25	1:1:1:1

(b)

Offspring genotypes	Genotype of parent 1: 01 (0.47)	01	07 (0.29)	62 (0.24)	Genotype of parent 2: 02 (0.24)	03 (0.26)	21 (0.25)	28 (0.25)	Observed pattern
			0.30	0.23		0.25		0.22	1:1:1:1
			0.31	0.22		0.26		0.21	1:1:1:1
	0.53	Same allele				0.23		0.24	2:1:1
	0.57	Same allele			0.24		0.19		2:1:1
	0.60	Same allele			0.19		0.21		2:1:1

Fig. 5 Family-assisted inference of allele sharing among loci. Nomenclature and shading follow Fig. 4. The presence of multiple copies of single alleles was estimated from the ratios of sequencing coverage (for instance 2:1:1), both observed in parents and offspring. (a) Family without allele sharing (representation expected without homologous recombination). (b) Family with allele sharing within the same chromosome. Alleles 01 are transmitted together on the same chromosome and thus shared among loci.

499 were double heterozygotes with about equal relative coverage for all alleles (1:1:1:1), confirming the presence of two loci in these individuals (Table 1). Accordingly, in 130 individuals that displayed three alleles, one allele had approximately twice the coverage of the others (2:1:1) and thus likely is present in two copies in these individuals (Fig. 6, Table 1). Finally, 24 individuals exhibited two alleles at approximately the same coverage (2:2) (Table 1; Fig. S5, Supporting information). However, in 100 individuals with three alleles, we found an unexpected pattern with all alleles exhibiting approximately the same coverage (1:1:1), indicating the presence of only three MHC-I copies in these individuals (Table 1). This result is likely explained by CNV. One of the parents would have transmitted a chromosome with only a single copy. In this case, the 1:1:1 pattern observed in offspring should therefore always also be found in at least one of the parents. The analysis of coverage patterns in these parents confirmed this expectation in all instances (Fig. 7). Furthermore, we identified cases in which two alleles were present at a 2:1 (N = 10) and 1:1 (N = 2) (Table 1; Fig. S6, Supporting information).

Table 1 Relative allelic coverage for each genotypic pattern. For patterns with multiple copies of the same allele, the mean relative coverage for the allele with the highest coverage (i.e. 2:1:1 and 2:1) was calculated independently as well as for remaining alleles within the pattern (numbers in bold and highlighted)

Pattern	Number of samples	Mean relative allelic coverage	Standard deviation
1:1:1:1	499	0.25	±0.04
2:1:1	130	0.52	±0.05
2:**1:1**	130	0.24	±0.03
2:2	24	0.50	±0.06
1:1:1	100	0.33	±0.04
2:1	10	0.71	±0.05
2:**1**	10	0.29	±0.05
1:1	2	0.50	±0.07

Discussion

The present study illustrates how allele segregation patterns within families in combination with high-coverage sequencing can bring insights into several aspects of the architecture of MHC diversity and assist assessing the reliability of MHC genotyping. Our analyses provide evidence of a strong linkage of MHC-I loci, the sharing of alleles among loci and the occurrence of CNV in the barn owl. We in turn discuss the nature of information provided by family data as well how this information can be made use of in downstream evolutionary ecology analyses.

In the present study, family data have provided information on the architecture of MHC variation in mainly two ways. First, segregation patterns within families have proven to be of great confirmatory value for the validation of CNV and genotypes. Odd patterns, such as

	Genotype of parent 1				Genotype of parent 2				Observed pattern
	02 (0.27)	11 (0.26)	31 (0.23)	37 (0.24)	02 (0.25)	14 (0.29)	24 (0.23)	38 (0.23)	
Offspring genotypes		0.38	0.22		0.20		0.20		1:1:1:1
		0.25	0.21		0.27		0.27		1:1:1:1
	0.29			0.23		0.24		0.24	1:1:1:1
	0.51			0.26	Same allele		0.23		2:1:1
	0.58			0.20	Same allele		0.22		2:1:1
	0.58			0.20	Same allele		0.22		2:1:1

Fig. 6 Family-assisted validation of the presence of two copies of an allele. Nomenclature and shading follow Fig. 4. This example illustrates case in which both parents are double heterozygotes (1:1:1:1), but share one allele between them (allele 02). This results in the presence of twice the same allele in offspring (i.e. allele 02), with a sequencing coverage ratio of 2:1:1.

	Genotype of parent 1				Genotype of parent 2				Observed pattern
	03 (0.39)	05 (0.27)	30 (0.34)	-	01 (0.22)	03 (0.28)	06 (0.25)	07 (0.25)	
Offspring genotypes	0.26		0.28		0.20			0.26	1:1:1:1
	0.50		0.25			Same allele	0.25		2:1:1
	0.50		0.22			Same allele	0.28		2:1:1
		0.33				0.34	0.33		1:1:1
		0.34				0.30	0.36		1:1:1

Fig. 7 Example of family showing copy number variation (CNV) pattern. Nomenclature and shading follow Fig. 4. Parent 1 lost one gene copy (1:1:1) in the chromosome carrying allele 05. A relative coverage pattern of 1:1:1 is observed in the two offspring that receive the allele 05. Both parents share one allele between them (allele 03), and this results in the presence of twice the same allele in offspring (allele 03), with a sequencing coverage ratio of 2:1:1.

allele coverage ratios of 1:1:1 linked to CNV, may be artefacts of unequal amplification or sequencing ratios among alleles. For these reasons, previous studies that already illustrated the value of high-coverage sequencing to study aspects of MHC variation that we touch upon here (Oomen *et al.* 2013; Lighten *et al.* 2014b) raised issues concerning the confidence with which conclusions can be drawn about for instance the occurrence of CNV. However, rules of inheritance provide clear predictions of how such patterns should segregate if they were real, and cross-checking among generations provides great confidence in conclusions drawn from such observations. Therefore, the approach based on family data that we put forward adds valuable information to the study of the architecture of MHC variation. Here, it for instance greatly assisted verifying the hypothesis of CNV by confirming that odd allele coverage ratios indicative of CNV (i.e. 1:1:1) segregate across generations. Also, our results show that family data are of great value to validate genotyping procedures, because the segregation of true alleles helps separating them from artefacts.

Second, we made use of direct information that can be drawn from family data to estimate linkage of loci, and allele sharing among loci. Although MHC genes are often found in tight linkage, the genomic architecture vastly varies among organisms (Kelley *et al.* 2005). In birds, knowledge on MHC architecture is still limited. In conjunction with the presence of only two MHC class IIB loci (Burri *et al.* 2008), the present results suggest an MHC organization in barn owls such as observed in chicken (Kaufman *et al.* 1999), which is rather simple compared to the one from passerines such as zebra finch (Balakrishnan *et al.* 2010). In nonmodel species that exhibit several simultaneously coamplified MHC loci, such as is often the case for MHC-I (for instance Alcaide *et al.* 2009; Promerová *et al.* 2012; Sepil *et al.* 2012; Jones *et al.* 2014), evaluating the strength of linkage among loci represents a challenging task. However, with the help of family data, linkage information can be established by studying the segregation of alleles across generations and reconstructing offspring haplotypes from parental genotypes. Using this approach, we were able to show that the linkage between the two barn owl loci is strong, implying that they are located and transmitted in tandem. Moreover, our analyses demonstrated the sharing of the same allele among loci, a pattern that is often suspected but rarely demonstrated (Wittzell *et al.* 1999) and prevents the establishment of complete genotype information. Oomen *et al.* (2013) recently addressed the same issue based on the assumption that in the case of allele sharing among loci, individuals with three copies of shared alleles (and thus a relative coverage of 3:1) should be observed. Given that no such cases were detected, this study assigned alleles to loci. However, even in our data set that comprises many hundreds of individuals, despite confirmation of allele sharing, no 3:1 ratios were detected, implying that allele sharing cannot be excluded based on the lack of observation of expected 3:1 ratios alone.

Establishing the architecture of MHC variation is not only of interest in its own right, but also impacts the way in which MHC variation is analysed in an evolutionary ecology context. Due to often still restricted data —and in some cases likely also missed cases of CNV— estimating the correct number of occurrences of each MHC allele within an individual remains a challenge. In the context of population genetic studies that aim at inferring the spatial structure of MHC variation, this will inevitably lead to biased allele frequency estimates. Allele frequencies are prone to be overestimated for rare alleles and underestimated for common alleles (Ekblom *et al.* 2007). This likely results in an overestimation of population structure and may therefore bias conclusions towards the detection of MHC-related local adaptation. Perhaps, more importantly, missing CNVs and variation in allelic copy number within individuals may have important consequences for the detection of MHC-linked fitness effects. In the present study, we identified individuals with the same allelic composition, but which differ in the number of copies at which one of the alleles is present, relative coverages being 2:1:1 and 1:1:1, respectively. Both theoretical considerations and empirical evidence suggest that these individuals might have different fitness, for instance, due to allele dosage effects. Individuals with four alleles may express higher concentrations of MHC-I receptors at the cell surface and hence more efficiently detect antigens. This is akin to effects found for instance in stickleback fish, where individuals with lower MHC diversity appear to compensate with a higher expression level (Wegner *et al.* 2006). Also, the massive duplication of the MHC-I family going along with the complete loss of MHC-IIB in cod may suggest that the lack of adequate MHC variation might be compensated by high MHC expression (Star *et al.* 2011). Although CNV is well understood in a theoretical framework and increasingly shown to evolve under selection (reviewed in Schrider & Hahn 2010), it has rarely been studied in the context of MHC evolutionary ecology (Lighten *et al.* 2014a). The family-assisted detection of CNV and variation in within-individual allele copy number outlined in the present study therefore contribute towards future investigations of the role of CNV in MHC-related disease susceptibility (Siddle *et al.* 2010).

Based on a system with a low number of MHC gene copies, the present study provides an illustration of the concepts of family-assisted inference of MHC architecture. In species with a higher copy number, these analyses may require significantly higher sequencing coverage, a problem that is offset by the ever improving yield of high-throughput sequencing methods. While the family-based genotype validation is straightforward to apply also in such species, analyses of recombination and allele sharing in these cases will depend to a greater extent on automated procedures for recombinant detection and analyses of shared alleles than required for species with a simple MHC architecture, such as the barn owl.

To conclude, the additional information that can be gained from high-coverage amplicon sequencing by investigating allele segregation patterns in families not only helps improving the accuracy of MHC genotyping, but contributes towards enhanced analyses in the context of MHC evolutionary ecology, including analyses relating MHC-I variation to individual fitness.

Acknowledgements

We thank all current and former members of Alexandre Roulin's group who participated in the sampling of the Swiss barn owl population. We thank Takeshi Kawakami for discussion and Marta Promerová and two reviewers for constructive comments on the manuscript. This study was supported by the Swiss National Science Foundation (no 31003A_138371 to LF and no 31003A_120517 to AR).

References

Alcaide M, Edwards S, Cadahía L, Negro J (2009) MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. *Conservation Genetics*, **10**, 1349–1355.

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, **17**, 179–224.

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

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

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

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

Bjorkman PJ, Saper MA, Samraoui B *et al.* (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, **329**, 512–518.

Boyer F, Mercier C, Bonin A *et al.* (2016) OBITOOLS: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, **16**, 176–182.

Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics*, **172**, 2665–2681.

Burri R, Niculita-Hirzel H, Roulin A, Fumagalli L (2008) Isolation and characterization of major histocompatibility complex (MHC) class II B genes in the Barn owl (Aves: *Tyto alba*). *Immunogenetics*, **60**, 543–550.

Burri R, Promerova M, Goebel J, Fumagalli L (2014) PCR-based isolation of multigene families: lessons from the avian MHC class IIB. *Molecular Ecology Resources*, **14**, 778–788.

Burri R, Antoniazza S, Gaigher A *et al.* (2016) The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean. *Evolution*, **70**, 140–153.

Ekblom R, Saether SA, Jacobsson P *et al.* (2007) Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology*, **16**, 1439–1451.

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

Henry I, Antoniazza S, Dubey S *et al.* (2013) Multiple paternity in polyandrous barn owls (*Tyto alba*). *PLoS One*, **8**, e80112.

Herdegen M, Babik W, Radwan J (2014) Selective pressures on MHC class II genes in the guppy (*Poecilia reticulata*) as inferred by hierarchical analysis of population structure. *Journal of Evolutionary Biology*, **27**, 2347–2359.

Hess CM, Edwards SV (2002) The evolution of the major histocompatibility complex in birds. *BioScience*, **52**, 423–431.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Jeffery KJM, Bangham CRM (2000) Do infectious diseases drive MHC diversity? *Microbes and Infection*, **2**, 1335–1341.

Jones MR, Cheviron ZA, Carling MD (2014) Variation in positively selected major histocompatibility complex class I loci in rufous-collared sparrows (*Zonotrichia capensis*). *Immunogenetics*, **66**, 693–704.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **16**, 1099–1106.

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

Kaufman J, Milne S, Gobel TWF *et al.* (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, **401**, 923–925.

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

Klein J, Sato A (2000) The HLA System. *New England Journal of Medicine*, **343**, 702–709.

Klein J, Sato A, Nagl S, O'Huigin C (1998) Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics*, **29**, 1–21.

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

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

Lighten J, van Oosterhout C, Bentzen P (2014a) Critical review of NGS analyses for *de novo* genotyping multigene families. *Molecular Ecology*, **23**, 3957–3972.

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

Michel C, Bernatchez L, Behrmann-Godel J (2009) Diversity and evolution of MHII P genes in a non-model percid species-The Eurasian perch (*Perca fluviatilis* L.). *Molecular Immunology*, **46**, 3399–3410.

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

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, **39**, 121–152.

Oomen RA, Gillett RM, Kyle CJ (2013) Comparison of 454 pyrosequencing methods for characterizing the major histocompatibility complex of nonmodel species and the advantages of ultra deep coverage. *Molecular Ecology Resources*, **13**, 103–116.

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

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

Promerová M, Babik W, Bryja J *et al.* (2012) Evaluation of two approaches to genotyping major histocompatibility complex class I in a passerine-CE-SSCP and 454 pyrosequencing. *Molecular Ecology Resources*, **12**, 285–292.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Schnell IB, Bohmann K, Gilbert MT (2015) Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, **15**, 1289–1303.

Schrider DR, Hahn MW (2010) Gene copy-number polymorphism in nature. *Proceedings of the Royal Society B-Biological Sciences*, **277**, 3213–3221.

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

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

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

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

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

Star B, Nederbragt AJ, Jentoft S *et al.* (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature*, **477**, 207–210.

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

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Wallny H-J, Avila D, Hunt LG *et al.* (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1434–1439.

Wegner KM (2009) Massive parallel MHC genotyping: titanium that shines. *Molecular Ecology*, **18**, 1818–1820.

Wegner KM, Kalbe M, Rauch G *et al.* (2006) Genetic variation in MHC class II expression and interactions with MHC sequence polymorphism in three-spined sticklebacks. *Molecular Ecology*, **15**, 1153–1164.

Wittzell H, Bernot A, Auffray C, Zoorob R (1999) Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. *Molecular Biology and Evolution*, **16**, 479–490.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**, 1107–1118.

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

R.B., A.R. and L.F. designed the study. A.G. performed molecular experiments and data analyses, with the help of P.T. and W.H.G. A.G. and R.B. wrote the manuscript with input of the others authors.

Data accessibility

DNA sequences: GenBank Accessions KX189198–KX189343. MHC-I genotypes and family information: Dryad database (doi: 10.5061/dryad.rm611).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Sampling.

Table S2 Definitions of terms used.

Table S3 Details of the filtering procedure.

Table S4 General statistic after filtering procedure (based on all populations).

Table S5 Genetic diversity at European barn owl MHC-I genes.

Fig. S1 Allele amplification efficiency following Sommer *et al*. (2013) calibrated with Tyal-UA*01.

Fig. S2 Diversity of MHC-I exon 3 in European barn owls.

Fig. S3 Phylogenetic network of barn owl MHC-I alleles.

Fig. S4 Genetic linkage between loci estimated using family data.

Fig. S5 Family-assisted validation of the presence of two copies of an allele.

Fig. S6 Examples of families showing CNV pattern.

Chapter 2

Lack of evidence for selection favouring MHC haplotypes that combine high functional diversity

A. Gaigher, A. Roulin, W. H. Gharib, P. Taberlet, R. Burri* & L. Fumagalli*

* Joint senior authors

*This chapter is in revision in **Heredity**.*

Abstract

High rates of gene duplication and the highest levels of functional allelic diversity in vertebrate genomes are the main hallmarks of the major histocompatibility complex (MHC), a multigene family with a primordial role in pathogen recognition. The usual tight linkage among MHC gene duplicates may provide an opportunity for the evolution of haplotypes that associate functionally divergent alleles and thus grant the transmission of optimal levels of diversity to coming generations. Even though such associations may be a crucial component of disease resistance, this hypothesis has been given little attention in wild populations. Here, we leveraged pedigree data from a barn owl (*Tyto alba*) population to characterize MHC haplotype structure across two MHC class I (MHC-I) and two MHC class IIB (MHC-IIB) duplicates, in order to test the hypothesis that haplotypes' genetic diversity is higher than expected from randomly associated alleles. Overall, we found limited evidence for shifts towards MHC haplotypes combining high diversity. Neither amino acid nor functional within-haplotype diversity were significantly higher than in random sets of haplotypes, regardless of MHC class. Our results therefore provide no evidence for selection towards high-diversity MHC haplotypes in barn owls. Rather, for MHC-IIB, fixed differences among loci may provide barn owls with already optimized functional diversity, while, in contrast, high rates of convergent evolution may constrain the evolution of high-diversity haplotypes at MHC-I. This suggests that at the MHC-IIB and MHC-I, respectively, different evolutionary dynamics may govern the evolution of within-haplotype diversity.

Introduction

Gene duplication is a major mechanism in the evolution of phenotypic complexity (Lynch & Conery 2000; Conant & Wolfe 2008), and has led to one of the most remarkable adaptations in vertebrates, the major histocompatibility complex (MHC). The MHC multigene family has a primordial role in pathogen resistance. Classical MHC class I (MHC-I) and class II (MHC-II) genes encode cell-surface proteins that present antigen-peptides derived from pathogens to T-lymphocytes, in order to trigger an adaptive immune response (Klein & Sato 2000). As a result of the host-pathogen arms race, MHC-I and MHC-II genes have evolved the highest genetic diversity known from any vertebrate genome region to date (Gaudieri *et al.* 2000; Bernatchez & Landry 2003; Piertney & Oliver 2006). This diversity entails not only the number of different alleles and the high degree of genetic divergence between them, but also the number of duplicated genes. MHC-I and -II diversity is typically distributed across multiple functional gene copies that are usually situated in tandem (Trowsdale & Parham 2004; Kelley *et al.* 2005).

Despite the growing amount of data on the characterization of MHC diversity and duplication history, the link between them, i.e. the combination of alleles of each duplicated MHC gene into haplotypes, has received little attention. Yet, diversity within haplotypes may deliver raw material that is selected at the ecological level. Until now, most of our knowledge about MHC haplotype structure is limited to human and poultry. In chicken, MHC-I variants appear to segregate together with co-adapted variants at strongly linked TAP genes that are fine-tuned with respect to their function of loading peptides on the MHC-I molecules (Walker *et al.* 2011). This coevolution is involved in MHC haplotype-related disease resistance, as for instance in economically important diseases such as Rous sarcoma virus or Marek's disease (Kaufman *et al.* 1999; Kaufman 2000; Wallny *et al.* 2006; Koch *et al.* 2007).

As MHC molecules are directly involved in the presentation of pathogen-peptides, MHC diversity should be optimized for a large number of different MHC molecules in individuals in order to fight a broader range of pathogens and thereby confers them with higher fitness (Doherty & Zinkernagel 1975; Bernatchez & Landry 2003; Sommer 2005; Spurgin & Richardson 2010). Individuals with highly divergent MHC alleles can interact with a wider range of pathogen-peptides than individuals with low allelic divergence (divergent allele advantage, Wakeland *et al.* 1990; Lenz 2011). Ample evidence has shown that high MHC diversity confers better pathogen resistance via heterozygote advantage or divergent allele advantage (for instance, Penn *et al.* 2002; Lenz *et al.* 2009; Oliver *et al.* 2009; Savage & Zamudio 2011), even if the optimum can be achieved by an intermediate level of MHC diversity due to the negative T-cell selection process (Nowak *et al.* 1992; Wegner *et al.* 2003). To optimize an individual's MHC diversity, mate choice for MHC-dissimilar partners may operate to increase the diversity in offspring. Alternatively, high-diversity haplotypes encompassing tightly linked MHC-I and/or MHC-II genes may ensure the transmission of a high amount of individual MHC diversity to progeny, even under random mating (Dearborn *et al.* 2016).

Under the latter hypothesis, high-diversity MHC haplotypes should be favoured by selection. In natural populations, selection for such haplotypes may be expressed in one of two ways. In the most extreme case, the diversity of observed haplotypes (i.e. the ones found in the population) exceeds the diversity levels expected for random subsets of the possible haplotypes (i.e. of all possible combinations of variants across duplicated genes, including haplotypes absent from the population); low-diversity haplotypes are purged from the population. More likely, however, high-diversity MHC haplotypes are found at higher frequencies than low-diversity haplotypes, and the average observed within-haplotype diversity should exceed the one expected under equal haplotype frequencies. These predictions should especially hold true for functional MHC diversity, i.e. the diversity observed at the residues of the peptide-binding region (PBR) involved in the detection of pathogen-derived peptides.

In most species, determining whether MHC haplotypes lock up higher than randomly expected diversity has been limited by the ability to reconstruct MHC haplotypes. Here we took advantage of extensive pedigree data to reconstruct MHC-I and MHC-IIB haplotypes, in order to investigate whether the tight linkage of MHC duplicates may favour haplotypes combining high diversity in a natural population of barn owl (*Tyto alba*), a species with two MHC-I and MHC-IIB duplicates (Burri *et al.* 2008; Gaigher *et al.* 2016). To this end, our main goals in the present study were to: (i) characterize evolutionary mechanisms that shape the MHC diversity; (ii) estimate the linkage between MHC loci; and (iii) compare the levels of genetic diversity across loci within observed haplotypes and expected haplotypes under random allelic combinations.

Material and Methods

Sampling and DNA extraction

We focused our study on a single population of barn owls breeding in nest boxes in western Switzerland. We collected blood and feather samples from adults and their offspring between 1997 and 2003 resulting in a total of 937 barn owls. These samples included 823 individuals from 140 families. Each family was formed of two parents and on average 4.5 (range 1-17) offspring.

DNA was extracted using the DNeasy blood and tissue kit following the manufacturer's instructions (Qiagen, Hilden, Germany). All individuals were genotyped at 10 microsatellite markers (multiplex sets 3 and 4 in Burri *et al.* 2016) to verify parent-offspring relationships using CERVUS (Kalinowski *et al.* 2007).

MHC sequencing and genotyping

We investigated exon 3 of MHC class Iα (MHC-I) genes and exon 2 of MHC class IIβ (MHC-IIB) genes, which encode for polymorphic sequences encoding the respective genes' peptide-binding regions. MHC-I primers were developed to specifically co-amplify the exon 3 of the two genes (see details in Gaigher *et al.* 2016). For specific amplification of both MHC-IIB genes (DAB1 and DAB2), we used forward primers Tyal-int1F and Tyal-DAB2-int1F together with the single reverse primer Tyal-int2R (Burri *et al.* 2008; Supplementary Methods).

Because each MHC class was sequenced at a different time period, and since the most updated technologies available at that time were used, libraries of the MHC-I and MHC-IIB genes were sequenced with the Illumina MiSeq technology and the 454 Titanium pyrosequencing protocol, respectively. All molecular protocols are described in Gaigher *et al.* (2016) and Burri *et al.* (2008), for MHC-I and MHC-IIB, respectively, and in the Supplementary Methods. In brief, all individuals were amplified for both MHC classes with individual barcoded primers. PCR products were quantified (either visually on agarose gels or using the QIAxcel screening system (Qiagen)), purified by pooling eight PCR products of similar amplification intensity per column, and finally pooled according to equimolar concentrations of purified PCR products. Library preparation and high-throughput sequencing were performed at Fasteris (Plan-les-Ouates, Switzerland).

The MHC-I data used in the current study were previously published, and details about the genotyping procedure can be found in Gaigher *et al.* (2016). Briefly, the Illumina approach used to sequence MHC-I yielded a very high coverage per individual (~3000x). To identify and estimate the number of MHC-I alleles per individual, we used the degree of change (DOC) (Lighten *et al.* 2014), that uses sequencing depth to distinguish true alleles from artifacts. Based on the pattern of allelic segregation within families, we have demonstrated that the DOC method provides accurate MHC genotyping (Gaigher *et al.* 2016). In addition, allelic segregation patterns, together with high per-individual sequencing coverage, revealed allele sharing among loci, as well as the presence of copy number variation (CNV) in the barn owl MHC-I (Gaigher *et al.* 2016).

The MHC-IIB data were generated for this study. The 454 technology used to sequence MHC-IIB loci resulted in an average coverage of 78 reads per individual. However, from these data a high proportion of artifacts was detected (mainly attributed to indels, but also including substitutions or chimera errors generated during PCR or sequencing). Consequently, in order to increase the coverage of true alleles to facilitate their identification we deployed a sequence similarity-based clustering approach to gather true alleles with all their potential artifacts, an approach in the same line of reasoning as Stutzand Bolnick (2014) and Sebastian *et al.* (2016). Our procedure relied on the three assumptions that: (i) in the whole dataset true alleles should be found at higher frequency than artifacts; (ii) artifacts should be highly similar to true alleles, differing only by 1 or 2 indels (especially in homopolymer regions) and/or substitutions; and (iii) artifacts have to co-occur with their true alleles within an individual. Generated clusters (i.e. the true allele plus its artifacts) were used to define MHC-IIB genotypes. Due to the independent amplification of both MHC-IIB loci, a maximum of two clusters per loci and per individual was expected. For details of the procedure see the Supplementary Methods. The MHC-IIB genotyping were judged reliable due to the correct matches in the pattern of allelic segregation within families. Furthermore, a subset of around 100 individuals were genotyped using the cloning/Sanger method, and showed congruent genotype results with the 454 sequencing.

Characterization of MHC-I and MHC-IIB

All identified alleles were designated according to standard nomenclature (Klein *et al.* 1990) and deposited in GenBank (XXXX-XXXX). Alignments of MHC-I and MHC-IIB alleles were performed separately using ClustalW (Thompson *et al.* 1994) implemented in MEGA5 (Tamura *et al.* 2011). For each MHC class, the average number of pairwise differences per base pair (π) was estimated in DnaSP (Librado & Rozas 2009), and Poisson corrected amino acid distances were obtained in MEGA5. These analyses were run on three data partitions: (i) the entire exon; (ii) codons of the peptide-binding region (PBR) exclusively; and (iii) codons of the non-PBR exclusively. PBR codons were defined from Human HLA and Chicken BF for MHC-I (Bjorkman *et al.* 1987; Wallny *et al.* 2006) and from Human HLA for MHC-IIB (Brown *et al.* 1993).

In order to investigate the phylogenetic relationships among MHC alleles, we built a molecular phylogeny for each MHC class separately, using MrBayes v3.2.3 (Ronquist & Huelsenbeck 2003) based on the GTR+Γ model, which was considered the best-fitting nucleotide substitution model by jModelTest (Darriba *et al.* 2012). Bayesian inference analyses were performed with two independent MCMC runs of 2×10^7 generations (three heated chains with a temperature of 0.15). Parameter values and tree topologies were sampled every 2000 generations. Posterior probabilities were calculated after removing the first 25% of the topologies as burn-in. Convergence was estimated using the average standard deviation of split frequencies between runs, the estimated sample size (ESS) and the potential scale reduction factor (PSRF) using MrBayes and Tracer v1.6 (Rambaut *et al.* 2014).

Recombination events were inferred using multiple methods implemented in RDP4, including RDP (Martin & Rybicki 2000), MaxChi (Smith 1992), and Chimerae (Posada & Crandall 2001). All default parameters were applied with a highest acceptable P-value of 0.05 and Bonferroni correction for multiple comparisons. In addition, we performed the Φw test (Bruen *et al.* 2006) in SplitsTree 4 (Huson & Bryant 2006), and estimated the minimal number of historical recombination events (Rm, Hudson & Kaplan 1985) using the four-gamete test in DnaSP. Finally, gene conversion events were tracked using Geneconv 1.81 (Sawyer 1999) with 10 000 permutations.

In order to investigate footprints of positive selection, we estimated maximum likelihood site-models using CodeML implemented in PAML v4.7 (Yang 2007). These analyses were performed independently for each MHC gene using the identified alleles as input. Two likelihood ratio tests of positive selection as proposed by Yang *et al.* (2005) were carried out comparing models M1a with M2a and models M7 with M8. Models M1a and M7 are neutral, while models M2a and M8 allow for a proportion of sites to evolve under positive selection. Likelihood ratio test statistics (i.e. $2^*(\ln L_b - \ln L_a)$) were compared to the χ^2 distribution with two degrees of freedom. When the best-fit model was M2a or M8, sites under positive selection were determined through the Bayes empirical Bayes (BEB) approach. Input tree files used to run CodeML were generated from MrBayes under the GTR+Γ model. In order to ensure that signals of selection were not sensitive to tree topology, we used the best tree as input, and then re-performed the CodeML analysis with nine other topologies randomly chosen from the posterior distribution of topologies.

Genetic architecture

The MHC haplotype reconstruction for each individual was performed based on the allelic segregation within families. From the resulting haplotypes, we investigated linkage among MHC loci. We estimated linkage between: (i) the two MHC-I loci; (ii) the two MHC-IIB loci; and (iii) MHC classes. Because homozygote parents are uninformative regarding the occurrence of recombination, our linkage estimation was based only on heterozygous parents that transmitted a minimum of five gametes. Because a given parent can be heterozygous at one MHC class but homozygous at the other, the number of parents to assess linkage differed between analyses involving MHC-I (103 parents, 804 gametes), MHC-IIB (57 parents, 409 gametes), and both classes (76 parents, 535 gametes).

Recombinant gametes were inferred from the rationale provided in Gaigher *et al.* (2016). From a fully heterozygous parent, a maximum of 16 different haplotypes are expected to be transmitted to offspring in case of free homologous recombination among all loci. If, in contrast, all MHC loci are linked, only two different haplotypic combinations should be observed in offspring; in this case, alleles at the four linked loci are generally transmitted together. Following this rationale, and assuming that allelic combinations resulted from a minimum number of recombination events, we deduced the frequency of recombinant gametes in our family data, which is indicative of the amount of linkage of the four loci.

Haplotype characterization

Firstly, we estimated the diversity combined within barn owl MHC-I and MHC-IIB haplotypes using three different genetic distances: (i) the nucleotide sequence-based p-distance; (ii) amino acid sequence-based p-distance; and (iii) amino acid functional distance. Nucleotide and amino acid distances between MHC alleles were calculated using MEGA5. Functional distances were measured as reported by Agbali *et al.* (2010) and

Dearborn *et al.* (2016). Briefly, the 20 amino acids were described as numerical measures according to five physicochemical properties (Sandberg *et al.* 1998), which were used to calculate a Euclidean distance between each pair of amino acids. The functional distance between alleles for MHC-I and MHC-IIB loci was estimated as the mean of Euclidean distances. Then, to test whether the diversity combined within MHC-I and MHC-IIB haplotypes was higher than expected, we performed two tests: test 1 investigated whether the haplotypes observed in the population combined more diversity than a random set of the same number of haplotypes sampled from all possible haplotypes. We then investigated whether haplotypes that combine high diversity are present at elevated frequencies in the population relative to a random combination of alleles, such as expected if selection favoured haplotypes combining higher than average diversity. To this end, in test 2 we tested whether the diversity observed with the population haplotype frequency distribution was higher than the one expected with a random combination of alleles, while considering the two loci's allele frequency distributions. Haplotype frequency used for test 2 was obtained from the different haplotypes that adults transmitted to offspring. For these two tests, 10^5 randomizations were run. These tests were performed independently on each MHC class and on three sequence partitions, namely the entire exon sequences, codons situated in the PBR, and codons inferred to be under positive selection. All statistical tests were performed in R 3.1.3 (R Core Team 2014).

Results

MHC-I and MHC-IIB characterization

Out of 937 individuals, 96%, 79% and 83% were successfully genotyped for MHC-I, MHC-IIB DAB1 and MHC-IIB DAB2, respectively. The remaining individuals could not be genotyped mainly due to low coverage. A total of 69 MHC-I alleles, 25 MHC-IIB DAB1 alleles and 17 MHC-IIB DAB2 alleles were identified (Figure 1). None showed evidence of non-functionality, such as frameshift mutations or stop codons. All nucleotide sequences translated into unique amino acid sequences for MHC-IIB, and only four were synonymous for MHC-I. Sequence analyses revealed that both MHC-I and MHC-IIB loci exhibited the classical characteristics of functional MHC genes: (i) high genetic diversity mainly located in the peptide-binding regions (Figure 1; Supplementary Table S1); (ii) evidence of positive selection (Figure 1); and (iii) footprints of recombination and gene conversion (Supplementary Table S2). DAB1 displayed a higher diversity than DAB2 (π: DAB1, 0.071; DAB2, 0.053; Supplementary Table S1) with a different amino acid composition (Figure 1). Our population covers a large variation of allele frequencies from very common to very rare alleles (the frequency of the most common alleles for MHC-I,

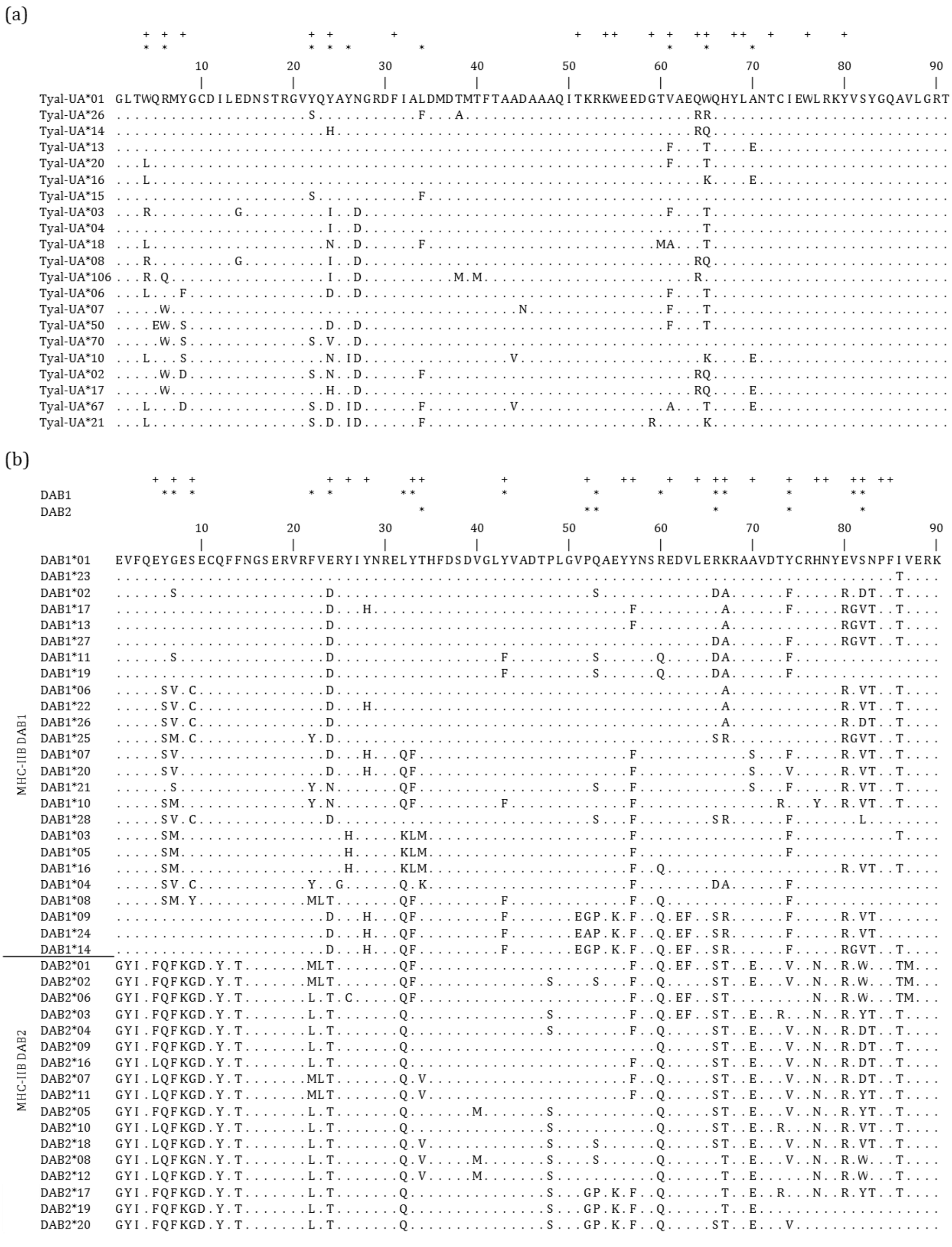


Figure 1: Amino acid sequences of MHC-I (a) and MHC-IIB alleles (b) of barn owl. Only the most divergent alleles are shown to illustrate the diversity in MHC-I, whereas all alleles are presented for MHC-IIB. +, residues associated to the PBR. *, sites identified to have evolved under positive selection according the M8 model (p > 95%) %) (based on the Bayes empirical Bayes approach). The evidence for positive selection at residues 5 and 63 of DAB2 was very sensitive to tree topology and consequently not considered as robust. The same applied to residue 8 of MHC-I.

MHC-IIB DAB1 and MHC-IIB DAB2 genes were 0.12, 0.26 and 0.50 respectively; Supplementary Figure S1).

In line with the monophyly of MHC-IIB loci in the phylogenetic tree (Supplementary Figure S2), MHC-IIB exon 2 is highly divergent between both loci (mean amino acid p-distance between loci, within DAB1, and within DAB2, respectively: 0.292, 0.138, 0.099) (Figure 2; Supplementary Figure S3). In contrast, the MHC-I tree exhibited a polytomic topology indicative of reticulate evolution of alleles not only within but also between the two loci (Supplementary Figure S2 and Figure S3). The MHC-I pairwise genetic distances revealed a unimodal distribution, with a mean amino acid p-distance of 0.075 (Figure 2). Although assigning alleles to loci based on the MHC-I tree was

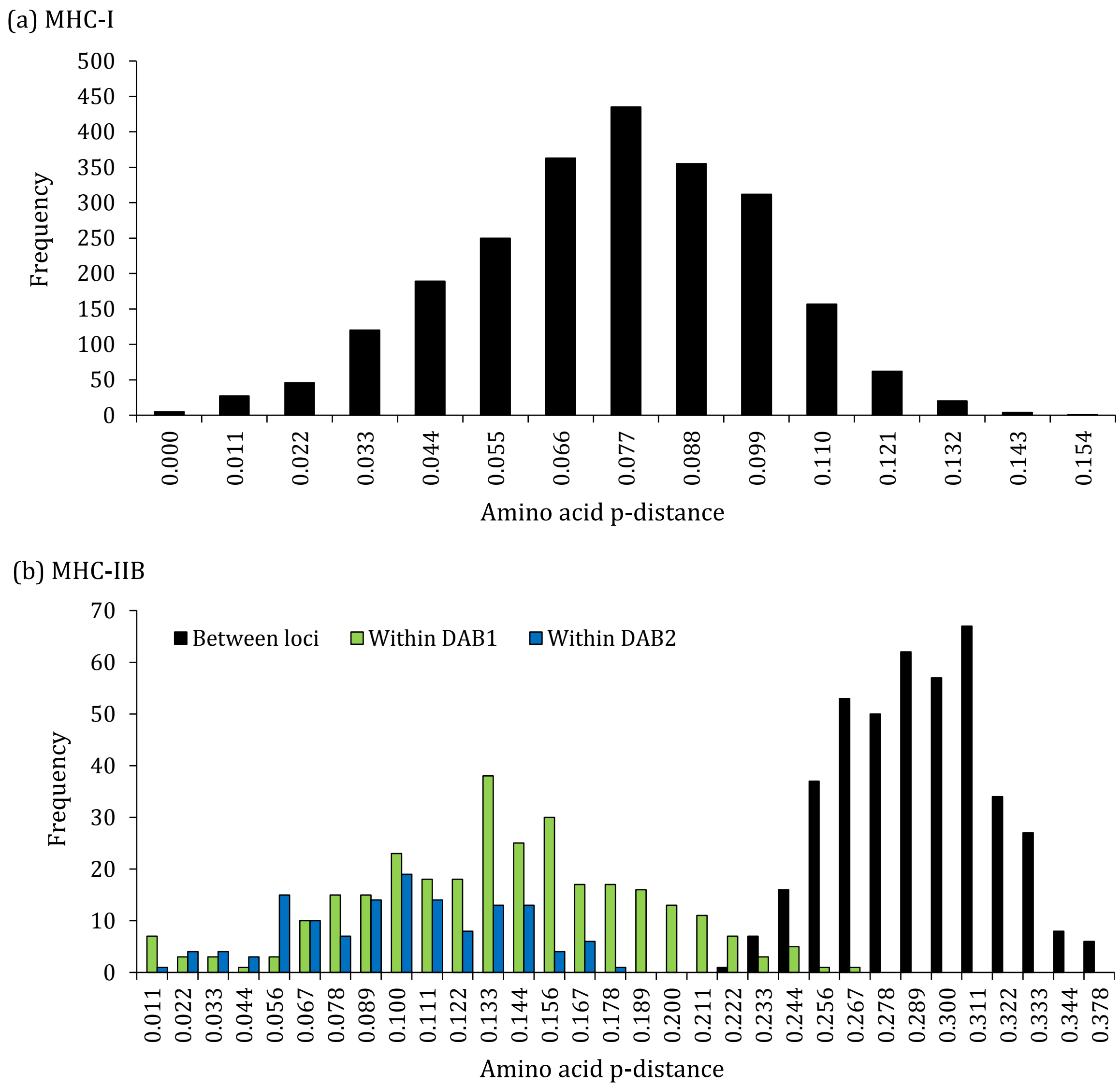


Figure 2: Amino acid p-distance frequencies of MHC-I (a) and MHC-IIB (b).

impossible, this could be achieved based on family data. Indeed, given that we observe a set of alleles combining only with another specific set of alleles, we were able to attribute alleles to loci (Supplementary Figure S4). However, this analysis reveals allele sharing among loci, for instance, Tyal-UA*01 allele occurred on the two MHC-I loci within the same haplotype (Supplementary Figure S4) (Gaigher *et al.* 2016).

Linkage within and between MHC classes

We inferred MHC-I/MHC-IIB haplotypes in offspring based on the pattern of allele segregation within families, and tracked recombination events to estimate linkage among MHC loci. In line with expectations of tight linkage between MHC loci, our analyses revealed that for both classes each parent almost exclusively transmitted two different haplotypes to offspring (Figure 3a). Within 409 analysed gametes, no recombination

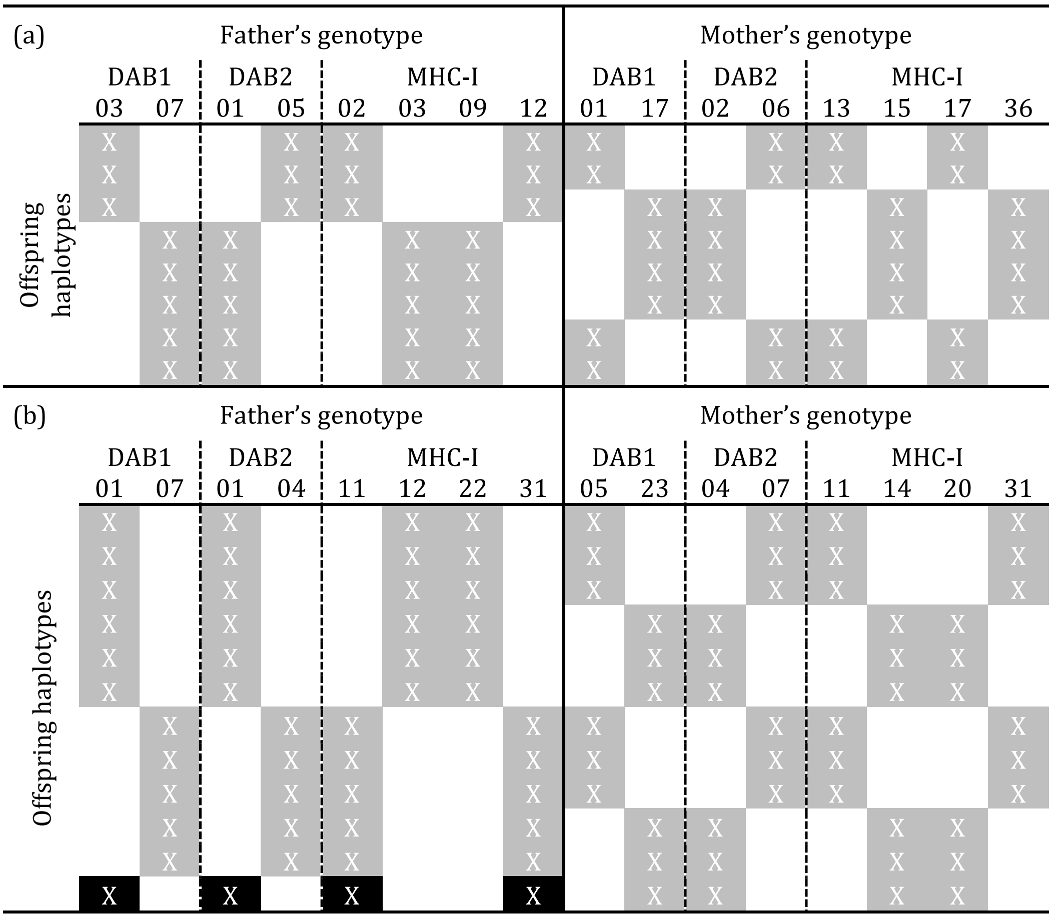


Figure 3: Genetic linkage between MHC loci. To facilitate the reading, all allele names were reduced to the allele number. Crosses indicate the presence of alleles in offspring. These two examples illustrate families in which both parents are heterozygotes for all loci. (a) Family with only four observed haplotypes in 8 offspring. For instance, parent 1 transmitted only two different haplotypes to offspring: DAB1*03/DAB2*05/MHC-I*02/MHC-I*12 and DAB1*07/DAB2*01/MHC-I*03/MHC-I*09. (b) Family composed of 12 offspring, in which recombination has been detected between MHC classes. Black backgrounds indicate recombinant haplotypes.

event was detected between MHC-IIB loci, and for MHC-I out of 804 gametes only three showed evidence for recombination between loci. In contrast, between MHC classes eight recombination events were detected within 535 gametes (Figure 3b). In addition, nine other recombinant gametes were detected; however due to homozygosity of parents for one locus, recombination events were impossible to locate (i.e. between MHC classes or between loci of the same class). In total, we found evidence for 20 recombination events, implying that MHC loci are linked (lower than 3 cM), but with a stronger linkage within than between MHC classes, and with a stronger linkage between MHC-IIB loci than between MHC-I loci. As may be expected from the latter result, the most common MHC-I alleles are found in haplotypes in combination with many different alleles (for instance Tyal-UA*01, *02, and *03 combine with 13, 12, 12 different alleles, respectively), whereas the most common MHC-IIB DAB1 alleles group with exclusively one or a few DAB2 alleles (Tyal-DAB1*01, Tyal-DAB1*10, and Tyal-DAB1*05 combine with two, two and one DAB2 alleles) (Supplementary Figure S4). This last point was supported by the strong linkage between MHC-IIB loci estimated from the likelihood ratio test ($p < 0.001$).

Haplotype characterization

A total of 111 MHC-I and 40 MHC-IIB different haplotypes were observed (Supplementary Figure S4). Across MHC classes, 210 different haplotypes were identified. Our data highlighted that only 11% and 9% of all possible allelic combinations were realized for MHC-I and MHC-IIB, respectively. In addition, our population compiles a wide variation of haplotype frequencies from common to rare haplotypes (Supplementary Figure S4), with important amino acid divergence between alleles (Figure 1 and Figure 2). Consequently, we took advantage of our data to first test whether the diversity combined within the MHC-I and MHC-IIB haplotypes that are observed in the population was higher than expected under a random set of all possible haplotypes. We found no support in this direction. Neither nucleotide, amino acid nor functional within-haplotype diversity in the population were significantly higher than in random sets of haplotypes, regardless of the MHC class (Test 1, Table 1). Then, we tested whether MHC haplotypes with higher frequencies combine the highest diversity, relative to an expected haplotype frequency distribution (Test 2, Table 1). The most common MHC-IIB DAB2 allele (MHC-IIB DAB2*01) displays on average the highest amino acid distance with DAB1 alleles (mean amino acid p-distance: 0.311); hence we performed the second test considering allele frequencies in the expected distribution, in order to account for processes unrelated to selection (Test 2, Table 1). We found low support in this direction, with only a significant shift for high diversity at the nucleotide level in the PBR and PSS data, as well as at the amino acid level in the PSS (Test 2, Table 1). Overall, an inverse trend was observed for MHC-I haplotypes; i.e. observed haplotypes appear to have lower diversity compared to random expectations (Table 1).

Table 1: Mean within-haplotype diversity for MHC-I and MHC-IIB in Swiss barn owls. Test 1 investigates whether the diversity combined within haplotypes observed in the population is higher than a random set of the same number of haplotypes sampled from all possible haplotypes. Tests 2 investigates whether the diversity combined within the observed haplotype frequency distribution is higher than within a random combination of alleles, while considering the allele frequency distribution.

		MHC-I						MHC-IIB					
		Test 1			Test 2			Test 1			Test 2		
		Exp.	Obs.	P	Exp.	Obs.	P	Exp.	Obs.	P	Exp.	Obs.	P
Nucleotide p-distance	Entire	0.034	0.030	1.000	0.030	0.029	0.813	0.174	0.176	0.183	0.178	0.178	0.061
	PBR	0.152	0.133	0.999	0.136	0.134	0.710	0.317	0.319	0.341	0.314	0.316	0.002
	PSS	0.224	0.195	0.999	0.195	0.188	0.918	0.455	0.463	0.135	0.446	0.450	0.000
Amino acid p-distance	Entire	0.069	0.061	0.999	0.060	0.060	0.520	0.292	0.295	0.259	0.303	0.304	0.138
	PBR	0.316	0.281	0.996	0.283	0.289	0.216	0.491	0.495	0.316	0.496	0.498	0.092
	PSS	0.483	0.423	1.000	0.414	0.413	0.510	0.749	0.762	0.178	0.751	0.755	0.030
Amino acid functional distance	Entire	0.395	0.351	0.998	0.355	0.355	0.510	1.495	1.496	0.480	1.537	1.533	0.937
	PBR	2.059	1.859	0.989	1.927	1.945	0.362	2.594	2.603	0.420	2.667	2.655	0.981
	PSS	3.039	2.708	0.997	2.754	2.734	0.614	3.645	3.656	0.435	3.710	3.685	0.991

Exp., mean expected diversity; Obs., mean observed diversity; P, p-value, Entire: entire sequence; PBR, peptide-binding region; PSS, positively selected site.

Discussion

In the present study we investigated whether tightly linked MHC genes may favour the evolution of haplotypes that associate functionally divergent alleles and thus grant the transmission of a high amount of MHC diversity to progeny. First, our analyses revealed contrasted evolutionary dynamics between MHC classes. While the two MHC-I loci are indistinguishable due to their high sequence similarity, the two MHC-IIB loci are strongly divergent. Second, our results showed that all MHC loci are linked, with as expected a stronger linkage within than between MHC classes. Finally, when we tested whether the diversity combined within MHC-I and MHC-IIB haplotypes is higher than expected from randomly associated alleles, overall, our tests supported no evidence for shifts towards high MHC within-haplotype diversity at the amino acid sequence level in our population. As our dataset provided a good representation of the barn owl haplotype diversity in the study population, sample size is unlikely to explain the lack of evidence for evolution towards high-diversity haplotypes. Given likely biological meaning of our finding, we therefore discuss how the evolution of high-diversity haplotypes in our population may be constrained by the molecular evolution of MHC genes.

At the ultimate functional level, that is, at the cell surface, the only thing that may matter is the diversity of different MHC molecules, regardless of how it was inherited. Still, high-diversity haplotypes may be favoured because they assure inheritance of a minimal level of MHC diversity. A previous study in the MHC-DRB of wild baboons suggested that selection favours haplotypes combining different sets of DRB supertypes (i.e. clusters of alleles based on their similar amino acid physicochemical properties), leading to an overall high diversity over multiple loci in individuals (Huchard *et al.* 2008). In contrast, here we found only low support for high within-haplotype diversity. Explanations for this finding may be fundamentally different between the two MHC classes.

For MHC-I, the evolution of high-diversity haplotypes may be constrained by high rates of recombination and gene conversion. These processes have previously been documented to shape MHC diversity especially in birds (Hess & Edwards 2002; Go *et al.* 2003; Spurgin *et al.* 2011; Promerová *et al.* 2013; Goebel *et al.* 2017). In line with this, we previously demonstrated allele sharing among barn owl MHC-I loci, as well as copy number variation (Gaigher *et al.* 2016), both of which decrease the level of divergence between loci. Barn owl MHC-I diversity therefore tends towards a homogenization across both loci, suggesting high rates of gene conversion. Our results now even suggest that observed haplotypes combine lower diversity compared to random expectations. Whether this is promoted by selection or reflects evolutionary constraints imposed by local recombination rates remains to be addressed.

In contrast, the highly divergent evolutionary history between the two MHC-IIB loci may inherently have promoted the evolution toward high-diversity haplotypes. Here, it is important to note that, had we randomized alleles between rather than within loci, haplotypes would be significantly more diverse than by chance: the two barn owl MHC-IIB loci exhibit fixed differences in the amino acid sequence, especially within the PBR in

5' of the sequence (Burri *et al.* 2008). Moreover, as the fixed differences between loci translate into different MHC supertypes, haplotypes are bound to consistently combine different supertypes; a result similar to what has been reported from wild baboons (Huchard *et al.* 2008). These fixed differences generate much higher allelic diversity between than within the MHC-IIB loci. Consequently, as the two loci are already divergent, an even higher level of divergence may be not be of additional advantage, as the fixed differences between duplicates may already ensure the transmission of a sufficient amount of diversity to the next generation.

In the MHC context, the evolution of high-diversity haplotypes may be promoted by the very same mechanism restricting the co-segregation of co-adapted alleles, i.e. recombination. Recombination (*sensu lato*) is essential to generate new allelic combinations at the MHC, which may offer an adaptive potential against pathogens. When selection is strong enough, new high-diversity combinations of alleles can be locked, and increase in frequency in the population. At the same time however, if recombination rates are high enough to recombine divergent alleles into beneficial high-diversity haplotypes, it may be equally likely to break up such advantageous combinations. Our results therefore may suggest that in a system involved in defence against pathogens, such as the MHC, considerable flexibility – and hence recombination – may be required to parallel the dynamics of pathogens in time and space (Milinski 2006), and that the advantages of recombination surmount those of suppressed recombination to maintain high-diversity haplotypes.

To conclude, selection for high-diversity MHC haplotypes may be weak in the studied population. The high level of MHC diversity in the barn owl may not require selection for diverse haplotypes. Nevertheless, in systems with less diversity, there might be an effect of haplotypes, a hypothesis that should be tested in the future. In addition, we focused only on the hypothesis that high divergence confers better resistance, and it remains to be tested whether selection can favour the maintenance of specific and optimal interactions, such as proposed for the chicken MHC-I and TAP genes (Kaufman *et al.* 1999). In this context, studies correlating specific haplotypes to fitness-related traits can offer new insight into gene interaction (Gregersen *et al.* 2006; Sin *et al.* 2014).

References

Agbali M, Reichard M, Bryjová A, Bryja J, Smith C (2010) Mate choice for nonadditive genetic benefits correlate with MHC dissimilarity in the rose bitterling (*Rhodeus ocellatus*). *Evolution*, **64**, 1683-1696.

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

Bjorkman PJ, Saper MA, Samraoui B*, et al.* (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, **329**, 512-518.

Brown JH, Jardetzky TS, Gorga JC*, et al.* (1993) Three-dimensional structure of the human class-II histocompatibility antigen HLA-DR1. *Nature*, **364**, 33-39.

Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics*, **172**, 2665-2681.

Burri R, Antoniazza S, Gaigher A*, et al.* (2016) The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean. *Evolution*, **70**, 140-153.

Burri R, Niculita-Hirzel H, Roulin A, Fumagalli L (2008) Isolation and characterization of major histocompatibility complex (MHC) class IIB genes in the Barn owl (Aves: *Tyto alba*). *Immunogenetics*, **60**, 543-550.

Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics*, **9**, 938-950.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772-772.

Dearborn DC, Gager AB, McArthur AG*, et al.* (2016) Gene duplication and divergence produce divergent MHC genotypes without disassortative mating. *Molecular Ecology*, **25**, 4355-4367.

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

Gaigher A, Burri R, Gharib WH*, et al.* (2016) Family-assisted inference of the genetic architecture of major histocompatibility complex variation. *Molecular Ecology Resources*, **16**, 1353-1364.

Gaudieri S, Dawkins RL, Habara K, Kulski JK, Gojobori T (2000) SNP profile within the Human Major Histocompatibility Complex reveals an extreme and interrupted level of nucleotide diversity. *Genome Research*, **10**, 1579-1586.

Go Y, Satta Y, Kawamoto Y*, et al.* (2003) Frequent segmental sequence exchanges and rapid gene duplication characterize the MHC class I genes in lemurs. *Immunogenetics*, **55**, 450-461.

Goebel J, Promerová M, Bonadonna F*, et al.* (2017) 100 million years of multigene family evolution: origin and evolution of the avian MHC class IIB. *BMC Genomics*, **18**, 460.

Gregersen JW, Kranc KR, Ke X*, et al.* (2006) Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature*, **443**, 574-577.

Hess CM, Edwards SV (2002) The evolution of the major histocompatibility complex in birds. *Bioscience*, **52**, 423-431.

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

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147-164.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254-267.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **16**, 1099-1106.

Kaufman J (2000) The simple chicken major histocompatibility complex: life and death in the face of pathogens and vaccines. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **355**, 1077-1084.

Kaufman J, Milne S, Gobel TWF, *et al.* (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, **401**, 923-925.

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

Klein J, Bontrop RE, Dawkins RL, *et al.* (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics*, **31**, 217-219.

Klein J, Sato A (2000) The HLA System. *New England Journal of Medicine*, **343**, 702-709.

Koch M, Camp S, Collen T, *et al.* (2007) Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. *Immunity*, **27**, 885-899.

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

Lenz TL, Wells K, Pfeiffer M, Sommer S (2009) Diverse MHC IIB allele repertoire increases parasite resistance and body condition in the Long-tailed giant rat (*Leopoldamys sabanus*). *BMC Evolutionary Biology*, **9**, 269.

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

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

Lynch M, Conery JS (2000) The Evolutionary Fate and Consequences of Duplicate Genes. *Science*, **290**, 1151-1155.

Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, **16**, 562-563.

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

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

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

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

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

Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 13757-13762.

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

R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria. URL http://www.R-project.org/.

Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer v1.6, available from http://beast.bio.ed.ac.uk/Tracer.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.

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

Savage AE, Zamudio KR (2011) MHC genotypes associate with resistance to a frog-killing fungus. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 16705-16710.

Sawyer S (1999) GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis. Available at http://www.math.wustl.edu/~sawyer/geneconv/.

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

Sin YW, Annavi G, Dugdale HL, *et al.* (2014) Pathogen burden, co-infection and major histocompatibility complex variability in the European badger (*Meles meles*). *Molecular Ecology*, **23**, 5072-5088.

Smith J (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**, 126-129.

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

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

Spurgin LG, van Oosterhout C, Illera JC, *et al.* (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, **20**, 5213-5225.

Stutz WE, Bolnick DI (2014) Stepwise threshold clustering: a new method for genotyping MHC loci using next-generation sequencing technology. *PLoS One*, **9**, e100587.

Tamura K, Peterson D, Peterson N, *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731-2739.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673-4680.

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

Wakeland E, Boehme S, She J, *et al.* (1990) Ancestral polymorphisms of MHC class II genes: divergent allele advantage. *Immunologic Research*, **9**, 115-122.

Walker BA, Hunt LG, Sowa AK, *et al.* (2011) The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 8396-8401.

Wallny H-J, Avila D, Hunt LG, *et al.* (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1434-1439.

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

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

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**, 1107-1118.

Supplementary material

Supplementary materials include the following documents:

- Supporting Methods
- Supporting Results
- Supporting Data:

Table S1: Genetic diversity at barn owl MHC-I and MHC-IIB genes.

Table S2: Results of recombination analysis.

Figure S1: Allele frequencies at MHC-I, MHC-IIB DAB1 and MHC-IIB DAB2 genes.

Figure S2: Bayesian phylogenetic trees of (a) MHC-I exon 3 and (b) MHC-IIB exon 2.

Figure S3: Neighbor-net networks of (a) MHC-I exon 3 and (b) MHC-IIB exon 2 alleles.

Figure S4: MHC-I (a) and MHC-IIB (b) haplotype combinations in Swiss barn owls.

Supporting Methods

MHC-IIB: primer development, PCR amplification and 454 sequencing preparation

Barn owl's MHC class IIβ (MHC-IIB) genes DAB1 and DAB2 were previously isolated and characterized (Burri *et al.* 2008). Forward primers Tyal-int1F (Burri *et al.* 2008) and Tyal-DAB2-int1F (5'-CTCCCCGTGTCTGCCTGTGC-3') are situated in the region of intron 1 divergent between DAB1 and DAB2 and together with the single reverse primer Tyal-int2R (5'-GACGCGCTGCCACGCACTC-3') allow for the specific amplification of the species' polymorphic exon 2.

PCR reactions were carried out on Biometra T3000 thermocycler in a final volume of 25 µl containing approximately 10ng DNA, 1× buffer Gold, 2.0 mM MgCl2, 1x Q solution (Qiagen), 0.2 mM dNTPs, 0.25 µM each primer, and 1 U AmpliTaq Gold (Applied Biosystems, Switzerland). PCR conditions included an initial denaturation step at 95°C for 10 min, 30 cycles of denaturation at 95°C for 30 sec, primer annealing at 60°C (DAB1 and DAB2) for 45 sec, and primer extension at 72°C for 45 sec. A final step at 72°C for 7 min was used to complete primer extension.

We chose the 454 pyrosequencing protocol (Roche, Basel, Switzerland) to sequence efficiently MHC-IIB genes for the barn owl. One full run divided in eight different regions within the PicoTiterPlate was used for the parallel sequencing of the Swiss population. In order to identify individuals within region, PCR primers were tagged in 5' using 128 different tags of seven bp.

PCR purification prior to sequencing was carried out using the QIAquick PCR purification kit by pooling eight PCR products of similar amplification strength simultaneously on a single column. In order to equilibrate DNA volumes among PCRs, DNA concentrations of target amplicons were previously quantified either visually on agarose gels or using the QIAxcel screening kit (Qiagen) on an eGene HDA-GT12™ machine. Prior to multiplexing of PCR products per sequencing region, DNA concentrations of purified PCR products were quantified using a Nanodrop ND-1000 spectrophotometer, in order to multiplex equal DNA volumes per individual and locus.

MHC-IIB: raw data processing and genotyping

Contrasting with most studies, here we took advantage of (i) the independent amplification of both MHC-IIB loci, with the expectation of a maximum of two alleles per sample, and (ii) the previously known allelic characterization (Burri *et al.* 2008). The 454 technology used to sequence MHC-IIB loci resulted in an average coverage of 78 sequences per-amplicon. However, preliminary view of the data showed that many samples have a low sequence count and/or high proportion of artifacts. Consequently, we deployed a method to cluster true alleles with their potential artifacts. As a result, the sequence number of true alleles increases, facilitating their identification.

The processing phase is composed of three main steps. During these steps we tried to keep the same line of reasoning of previous studies using a high-throughput sequencing approach (Galan *et al.* 2010; Sommer *et al.* 2013; Lighten *et al.* 2014; Stutz & Bolnick 2014), but adapted for our data. Whereas the two first steps are common to the MHC genotyping area, the last step aims to generate clusters encompassing true alleles with their artifacts. All these steps are described below and illustrated by a flow chart (Figure A).

The first processing phase intends to conserve the best quality sequences from the 454 raw data. Data were filtered to keep only sequences with a maximum of two errors within primers, and none within tags. Then, sequences longer than 200 bp and with a count greater than one were retained (=singletons removed). Finally, identical sequences were grouped according to individual barcodes.

The second processing phase aims to reduce the number of variants on the whole data set by removing rare artifactual variants and samples with low coverage. Each variant with a maximum sequence count per individual lower than three were discarded. Due to low amplification intensity, samples covered by less than 10 sequences were excluded from the analyses.

The third processing phase aims to cluster true alleles with their potential artifacts. Before starting the procedure, data were organized as follows: (i) a table containing the sequence count for each variant and sample (variants in rows and samples in columns) and (ii) variants were sorted according to sequence count (from largest to smallest). The clustering procedure uses a top-down variant similarity comparison at two different scales, first at a whole dataset scale, and then at the individual scale to group artifacts with their true alleles. The procedure is based on three assumptions: (i) at the whole dataset, true alleles should be found at higher frequency than their own artifacts, (ii) artifacts should be similar to true alleles, differentiating only by 1 or 2 indels (especially in homopolymer regions) and/or substitutions, and (iii) artifacts have to always co-occur with their own true alleles in the individual amplicon (for artifacts arising during PCR).

First, variants sorted according to their sequence count were aligned using MAFFT (Katoh & Standley 2013). Paired comparisons were carried out by a top-down analysis (Figure B). When the top variant did not possess ambiguous sites (*i.e.* "n") and had the expected length of 279 and 270 bp respectively for DAB1 and DAB2, we performed a comparison with the second variant of the list. If only one or two indels of divergence were observed between both sequences, the second variant was accepted as an artifact of the top variant. All the information related to the second variant (*i.e.* sequence count occurring in each individual) were added to the top one. When the first variant was compared with all other variants, we moved to the next one, and re-performed the procedure. At the end, we obtained a reduced table, in which artifactual variant sequence counts were added to their cluster variant. During the procedure we control that, (i) the top variants receiving new sequences should occur within samples, and (ii) each variant should not be clustered in several cluster variants.

Second, we performed a procedure similar to the previous one, but here at the individual scale. Variants were aligned independently within each individual using MAFFT. Then, paired comparisons were carried out based on the following scoring method within each individual:

<u>Score:</u>

Substitution in non-conserved region: -12

Substitution in conserved region: -4

Indel: -1

Match: 0

Threshold: -11

Delimitations between conserved and non-conserved regions across the nucleotide sequence have been based on the peptide-binding region (PBR) sites and previous knowledge on polymorphic sites. If comparison between variants resulted in a score lower than -12, then the second variant was added to the top cluster variant. When the top variant was compared with all variants, we moved to the second variant, and re-performed the procedure. After the top-down procedure was done for each sample independently, we checked that each variant was not found in several cluster variants. Then, we generated a file gathering all individuals with their cluster variants (which are potential true alleles). Finally, artifactual clusters resulting from chimeras or substitution errors were discarded based on cluster features (*i.e.* low number of sequences in this cluster variant and the number of individuals that possess this cluster variant), and on a meticulous observation of the nucleotide sequence. Retained clusters were used to define MHC-IIB genotypes.

Reliability of the MHC-IIB DAB1 and DAB2 genotyping was evaluated with segregation patterns within families. Concretely, after the genotyping we checked that sequences attributed as true allele in offspring were also found and attributed as true allele in one of the parents. Over our 140 families, we found almost 100% matches. In addition, around 100 individuals were also genotyped using cloning/Sanger method, and showed congruent results with the 454 sequencing.

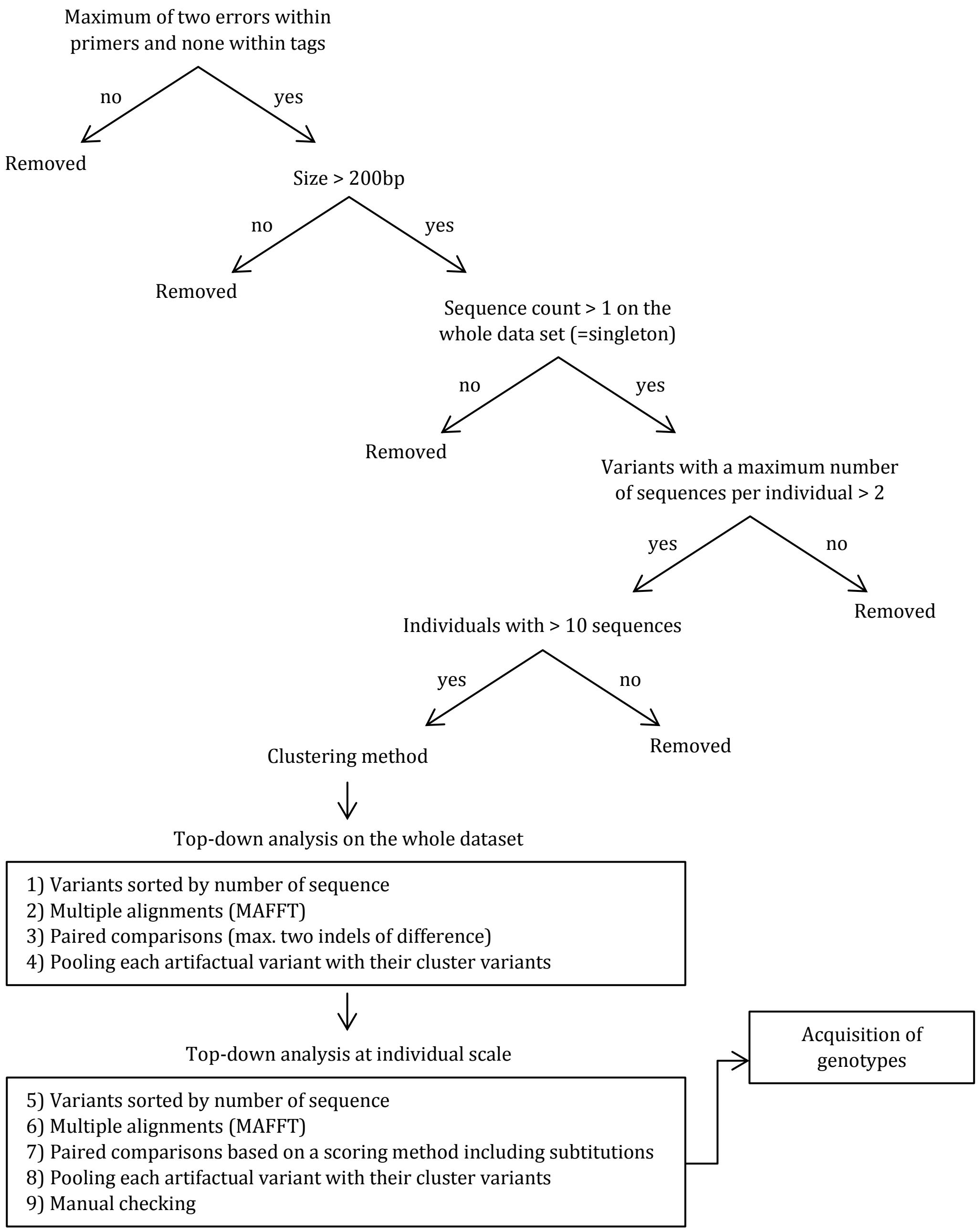


Figure A: Flow chart of the MHC-IIB genotyping procedure.

Figure B: Illustration of the top-down analysis at the whole dataset (MHC-IIB DAB1 example). Data were sorted according to the sequence count. Variant 0001 represent the top variant (*i.e.* with the maximum sequence count). The only difference between 0001 and 0007 variants is one indel, consequently sequence count at each sample of the variant 0007 are pooled with data of the variant 0001.

Variant name	Sequence count	Sample count	Max. sample count	Sequence length	Nucleotidic sequence	Sample 833858	Sample 871106	...
Var_0001	9970	560	186	279	caaacagaggttttccagg…	0	39	...
Var_0002	6290	316	169	279	caaacagaggttttccagg…	44	0	...
Var_0003	5282	288	118	279	caaacagaggttttccagg…	46	0	...
Var_0004	3332	195	112	279	caaacagaggttttccagg…	0	25	...
Var_0005	3206	169	80	279	caaacagaggttttccagg…	0	0	...
Var_0006	2763	189	65	279	caaacagaggttttccagg…	0	0	...
Var_0007	2583	272	88	278	caaacagaggttttccagg…	0	1	...
...	...	...	...	...	...	...	...	...

Supporting Results

MHC-I and MHC-IIB characterization

A total of 69 MHC-I alleles, 25 MHC-IIB DAB1 alleles and 17 MHC-IIB DAB2 alleles were identified (Figure 1). None of these sequences showed evidence of non-functionality, such as frameshift mutations or stop codons. All nucleotide alleles translated into unique amino acid sequences for MHC-IIB, only four were synonymous for MHC-I. Comparison between MHC classes revealed more alleles for MHC-I, however the genetic sequence diversity was higher by a factor of three for MHC-IIB, as illustrated by the mean pairwise differences, the nucleotide diversity (π) and amino-acid distances (MHC-I, 10.14, 0.036, and 0.078; MHC-IIB, 32.04, 0.119, and 0.240; Table S1). For both MHC classes, the diversity was concentrated within the respective PBR, where it was 10 and three times higher than at non-PBR for MHC-I and MHC-IIB, respectively (Table S1, Figure 1). Between MHC-IIB loci, DAB1 displayed a higher diversity compared to DAB2 (π: DAB1, 0.071; DAB2, 0.053; Table S1).

Positive selection and recombination (*sensu lato*) were shown to play an important role in shaping MHC diversity at both classes. Accordingly, branch-site tests of positive selection revealed that M2a and M8 were best-fit models for both MHC classes. For both models, nine positively selected sites (PSS) were found at MHC-I (Figure 1). For MHC-IIB combined, a total of 9 sites was identified as PSS. More than double the number of PSS were identified in MHC-IIB DAB1 compared to DAB2 in both models. In total 15 and 6 PSS were detected with M8 for DAB1 and DAB2 respectively, whereas based on M2a, 12 and 4 sites were conserved as PSS (Figure 1). In more than 70 percent of cases, sites detected under positive selection were located within the PBR (Figure 1). Finally, based on a set of methods we detected evidence for recombination (*sensu lato*) for all MHC loci (Table S2). Although, statistical analysis performed with RDP4 or Geneconv failed to detected recombination events in MHC-I, some of the applied tests are known to have limited power when gene conversion is too frequent (Mansai & Innan 2010).

Supporting Data

Table S1: Genetic diversity at barn owl MHC-I and MHC-IIB genes.

	Number of alleles	Number of sites	Number of codons	S	k	π (S.D.)	AA distance (S.E.)
MHC-I							
All	69	276	91	39	10.14	0.036 (0.001)	0.075 (0.019)
PBR	49	39	13	17	6.32	0.162 (0.006)	0.336 (0.079)
Non-PBR	43	237	78	22	3.82	0.016 (0.001)	0.031 (0.011)
MHC-IIB combined							
All	42	270	90	86	32.04	0.119 (0.004)	0.208 (0.026)
PBR	41	72	24	41	16.54	0.230 (0.005)	0.378 (0.055)
Non-PBR	33	198	66	45	15.49	0.078 (0.004)	0.146 (0.026)
MHC-IIB DAB1							
All	25	270	90	57	19.08	0.071 (0.004)	0.138 (0.023)
PBR	25	72	24	32	11.74	0.163 (0.007)	0.309 (0.051)
Non-PBR	19	198	66	25	7.34	0.033 (0.004)	0.075 (0.019)
MHC-IIB DAB2							
All	17	270	90	41	14.07	0.052 (0.005)	0.099 (0.018)
PBR	16	72	24	19	7.56	0.105 (0.008)	0.175 (0.048)
Non-PBR	14	198	66	22	6.51	0.033 (0.004)	0.072 (0.018)

S, number of polymorphic sites; ps, proportion of segregating sites; k, average number of nucleotide differences; π, average number of pairwise differences per base pair; AA distance, amino acid pairwise distance.

Table S2: Results of recombination analysis. Values indicate the number of detected recombination event estimated by several methods.

	Rm	Φw test	Geneconv	MaxChi	Chimerae	RDP
MHC-I	11	$p < 10^{-3}$	0	0	0	0
MHC-IIB combined	14	$p < 10^{-3}$	24	3	2	1
MHC-IIB DAB1	14	$p < 10^{-3}$	6	2	2	1
MHC-IIB DAB2	9	$p = 0.437$	1	1	0	0

Figure S1: Allele frequencies at MHC-I, MHC-IIB DAB1 and MHC-IIB DAB2 genes.

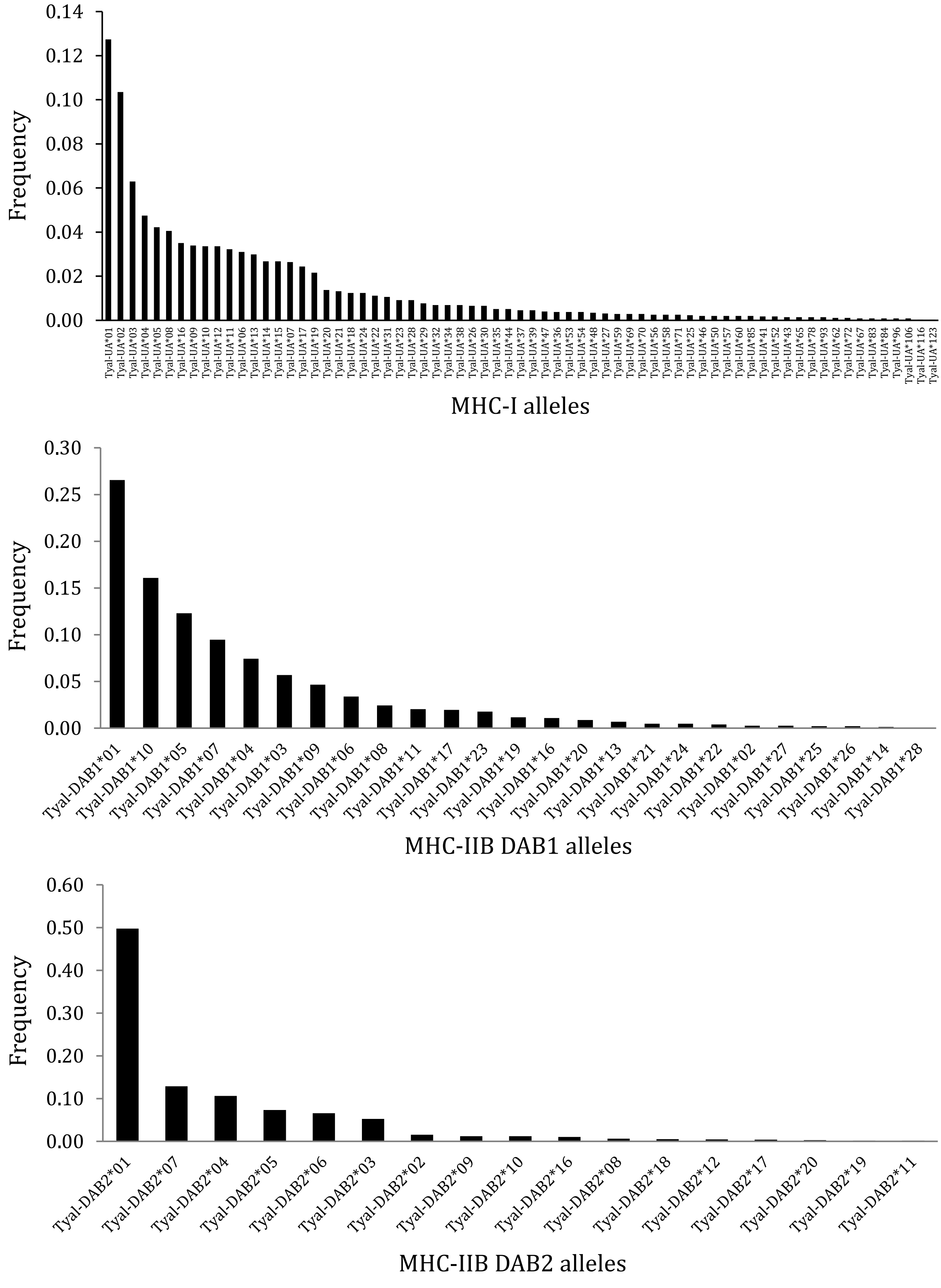
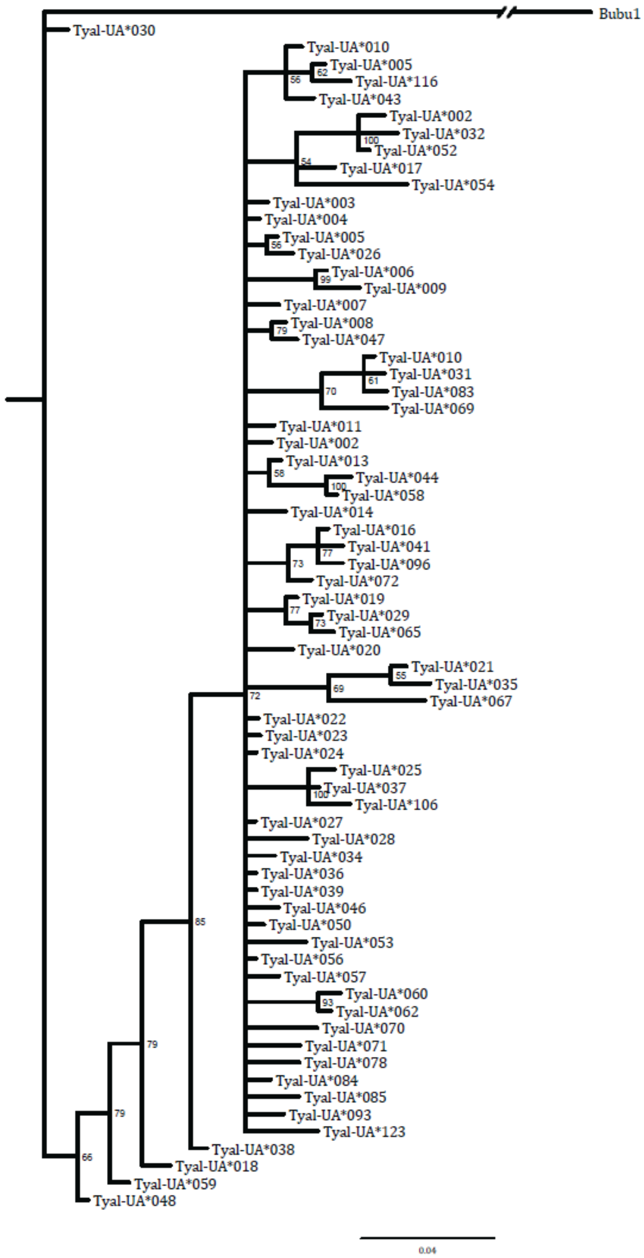


Figure S2: Bayesian phylogenetic trees of (a) MHC-I exon 3 and (b) MHC-IIB exon 2. Trees have been rooted with *Bubo bubo* sequences. Node values represent Bayesian posterior probability. Green: MHC-IIB DAB1, Blue: MHC-IIB DAB2.

(a)



(b)

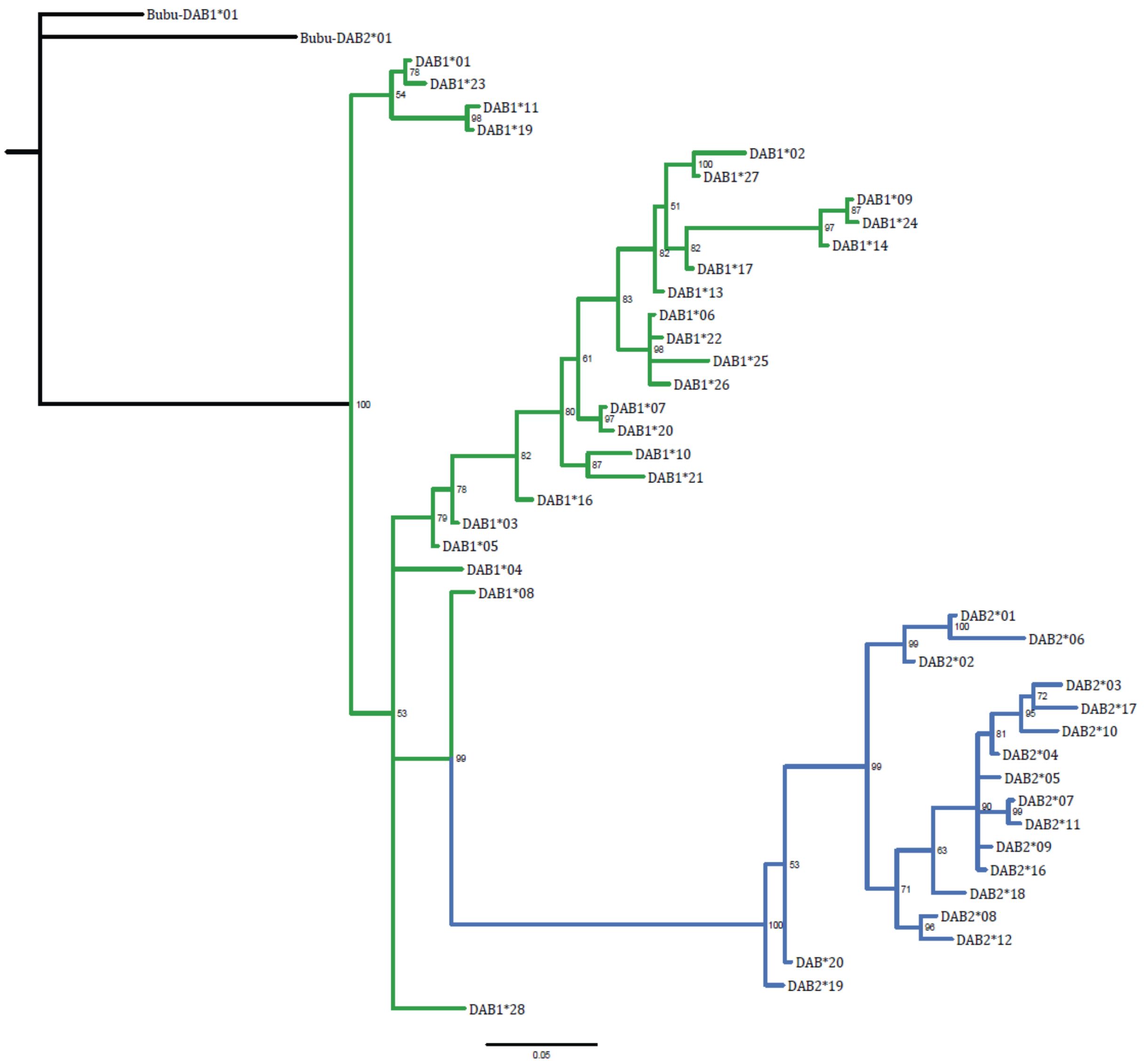
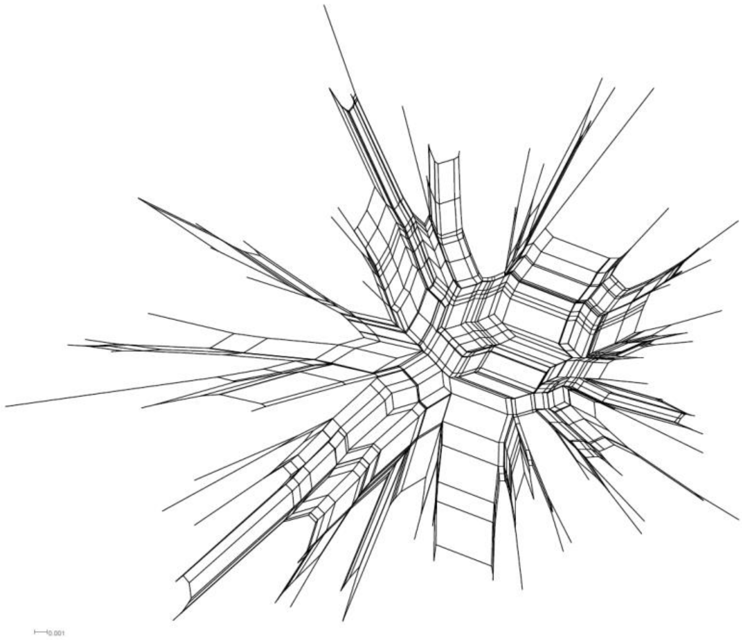

Bubu-DAB1*01
Bubu-DAB2*01
DAB1*01
DAB1*23
DAB1*11
DAB1*19
DAB1*02
DAB1*27
DAB1*09
DAB1*24
DAB1*14
DAB1*17
DAB1*13
DAB1*06
DAB1*22
DAB1*25
DAB1*26
DAB1*07
DAB1*20
DAB1*10
DAB1*21
DAB1*16
DAB1*03
DAB1*05
DAB1*04
DAB1*08
DAB2*01
DAB2*06
DAB2*02
DAB2*03
DAB2*17
DAB2*10
DAB2*04
DAB2*05
DAB2*07
DAB2*11
DAB2*09
DAB2*16
DAB2*18
DAB2*08
DAB2*12
DAB*20
DAB2*19
DAB1*28
0.05

Figure S3: Neighbor-net networks of (a) MHC-I exon 3 and (b) MHC-IIB exon 2 alleles. Networks were built with uncorrected p-distances using Splitstree 4 (Huson & Bryant 2006).

(a)



(b)

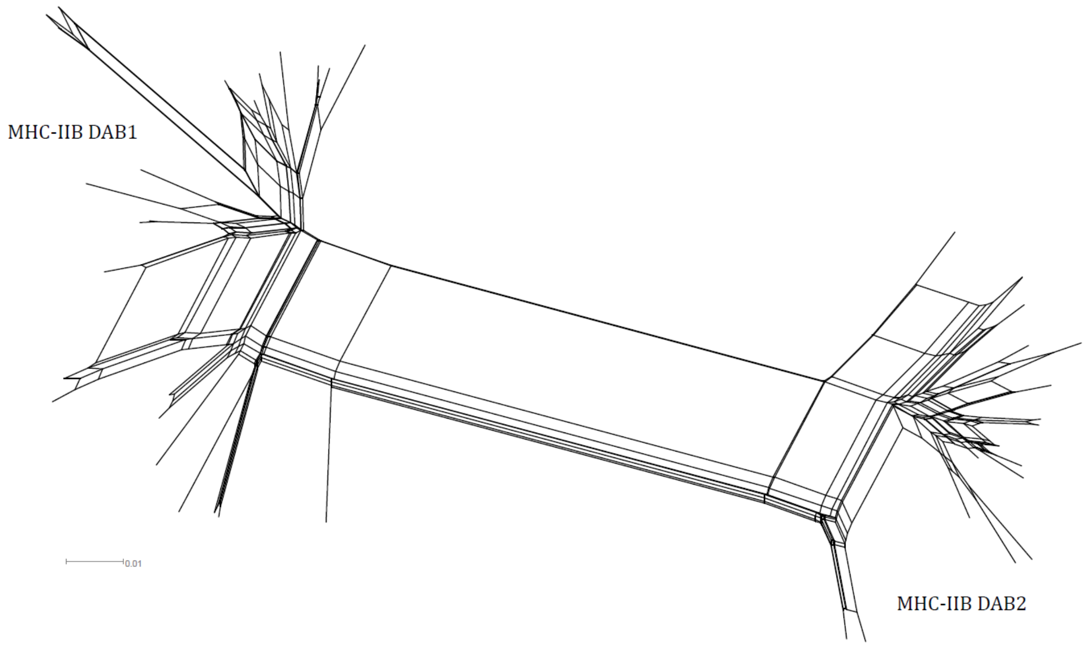
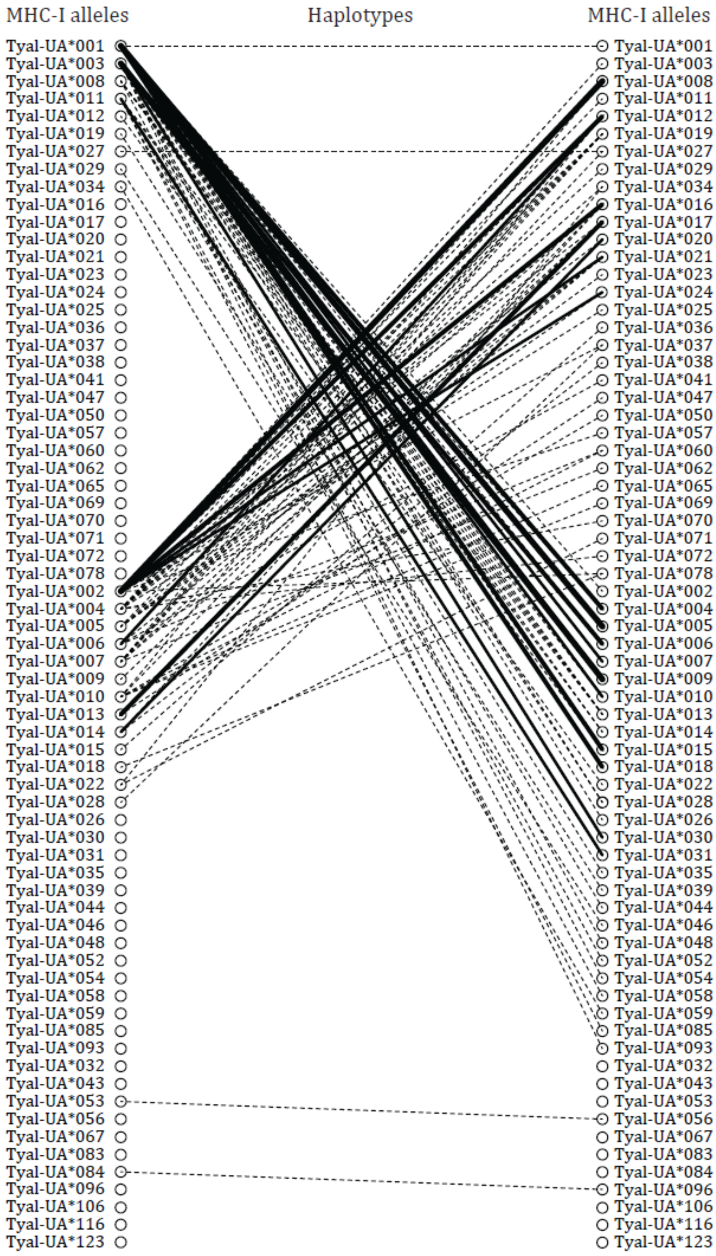
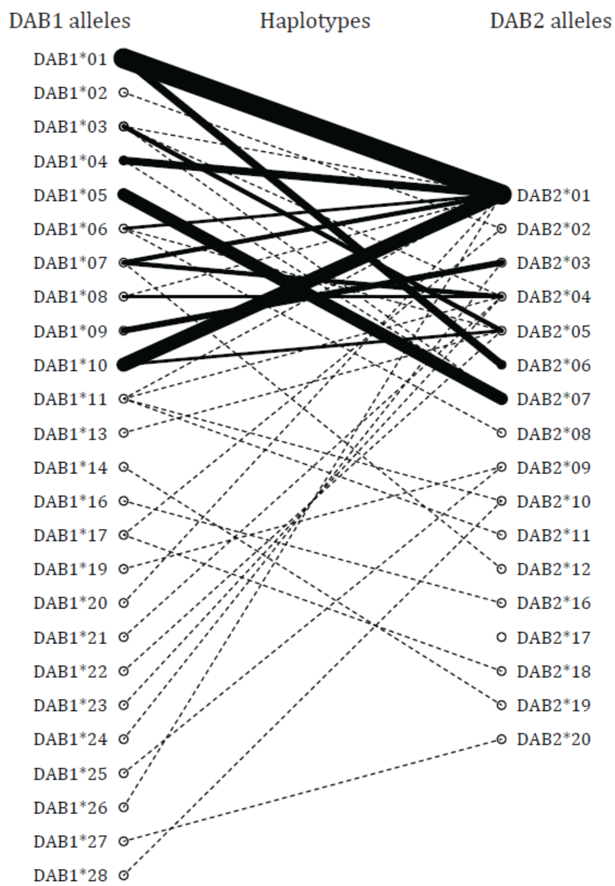


Figure S4: MHC-I (a) and MHC-IIB (b) haplotype combinations in Swiss barn owls. Right and left columns represent alleles. Each line is an allelic combination between two loci. The lines' weight is proportional to the occurrence of the haplotype in the population. Dashed lines are rare haplotypes. (a) MHC-I haplotype combinations. Due to the co-amplification of both MHC-I loci, and the unknown origin of alleles, each column is composed of the same alleles. However, this figure show that the same panel of alleles combine to another different set of alleles, resulting to possible assignation of alleles to loci. (b) MHC-IIB haplotype combinations. Right and left columns represent MHC-IIB DAB1 and DAB2 alleles respectively.

(a)



(b)



Supplementary material references

Burri R, Niculita-Hirzel H, Roulin A, Fumagalli L (2008) Isolation and characterization of major histocompatibility complex (MHC) class II B genes in the Barn owl (Aves: *Tyto alba*). *Immunogenetics*, **60**, 543-550.

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

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254-267.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772-780.

Mansai SP, Innan H (2010) The power of the methods for detecting interlocus gene conversion. *Genetics*, **184**, 517-527.

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

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

Stutz WE, Bolnick DI (2014) Stepwise threshold clustering: A new method for genotyping MHC loci using next generation sequencing technology. *PLoS One* **9**, e100587.

Chapter 3

Absence of strong effect-size associations between multiple immunocompetence proxies and the MHC system in barn owls

A. Gaigher, L.M. San-Jose, R. Burri, A. Roulin* & L. Fumagalli*

* Joint senior authors

Abstract

The ability to develop an effective immune response against pathogens can drastically vary between individuals, and disentangling the sources of such variation constitutes a major topic in eco-immunology and evolutionary biology. To date, the link between immunocompetence and the genetic background via the major histocompatibility complex (MHC) genes has been barely studied, even though it is a key component of immune system activation. The majority of these studies found no association between immunocompetence and MHC diversity or specific MHC alleles. The lack of association is usually perceived as the absence of a general biological association. However, lack of power owing to insufficient sample size and/or to the presence of small-effect-size associations between immunocompetence and the MHC system, may be affecting this type of studies. Here, we investigated whether the diversity at MHC-I and MHC-II classes influence eight distinct traits related to immune responses in the barn owl (*Tyto alba*). Overall, we found that MHC poorly explained the different immune-related traits. Consequently, we performed a set of power analyses to examine the power to detect associations of different effect sizes in the present study and previously published studies. We argued based on our findings, previous studies, and the genetic basis of the immune response, that if existing, small-to-moderate associations between the MHC system and immunocompetence are more likely to occur than large effect associations. Finally, larger sample sizes coupled with power analysis would be needed in studies linking the MHC diversity with fitness-related traits, in order to achieve a more robust understanding of such a link.

Introduction

Understanding the mechanisms that grant an effective immune response against pathogens is a central challenge in eco-immunology and evolutionary biology. The effectiveness of such response can drastically vary between organisms and often results from the simultaneous action of several factors including the history of infection, the physiological and environmental conditions, and the genetic background of an individual (Pedersen & Babayan 2011). To date, the genes of the major histocompatibility complex (MHC) have recurrently been used as candidate markers to study the link between genetic diversity and pathogens (Bernatchez & Landry 2003).

The MHC system, which includes two major gene subfamilies, MHC class I and II, plays a key role in the adaptive immune response, by encoding cell-surface proteins that present foreign antigen-peptides to the T-cells (Klein & Sato 2000; Jensen 2007). Currently, an important effort is being done to address the strength of the link between MHC diversity and disease resistance (see review in Sommer 2005), or, more generally, with fitness-related life-history traits, such as growth, reproduction and survival (Bonneaud *et al.* 2004; Worley *et al.* 2010; Barribeau *et al.* 2012; Radwan *et al.* 2012; Lenz *et al.* 2013; Sepil *et al.* 2013b). With this aim, several studies have investigated this link with the specific natural community of pathogens with whom the host can be confronted, and was supported in various organisms, including mammals (Oliver *et al.* 2009; Kloch *et al.* 2010; Sin *et al.* 2014; Osborne *et al.* 2015), fishes (Langefors *et al.* 2001; Wegner *et al.* 2003; Bolnick *et al.* 2014), amphibians (Savage & Zamudio 2011; Bataille *et al.* 2015), and birds (Westerdahl *et al.* 2012; Sepil *et al.* 2013a). However, rather than to focus on specific pathogens, a promising alternative of more general applicability to multiple species and scenarios would be to investigate the association between the MHC system with general parameters summarizing the general immunocompetence of an individual.

In the field of ecological immunology, immunocompetence refers to the capacity of an organism to mount an immune response against a specific antigen (Owens & Wilson 1999; but see Vinkler & Albrecht 2011). To assess immunocompetence, a wide range of techniques has been developed (Norris & Evans 2000) including "monitoring techniques" that provide general information about the health status of the host, and "challenging techniques" which activate the immune system by injecting an antigen to mimick an invasion by a pathogen (Norris & Evans 2000). The post-injection antibody-response can be followed, quantified and finally used as a measure of immunocompetence. Their simplicity allows researchers to study wild animals, and in this sense, methodologies such as the injection of sheep red blood cells (SRBC) or the phytohaemagglutinin (PHA) skin tests, have been widely used as challenging techniques to assess the immune capacity of animals, especially birds (Norris & Evans 2000; Owen *et al.* 2010). Such challenges trigger the activation of the innate, humoral and/or cell-mediated immunity, which involve key cell types and cellular components such as lymphocytes (Martin *et al.* 2006; Tella *et al.* 2008; Owen *et al.* 2010).

Most of the current knowledge of the link between MHC and immunocompetence comes from studies on fowl immunology (for instance Zhou & Lamont 2003; Shimizu *et al.* 2004; Sumners *et al.* 2012), and less information is available from wild species (Bonneaud *et al.* 2005; Charbonnel *et al.* 2010; Ekblom *et al.* 2013; Rodríguez *et al.* 2014). Since the MHC is an essential component of antigen recognition and lymphocyte activation in the humoral and cell-mediated immunity pathways (Janeway *et al.* 2001), a link between the MHC and immunocompetence is expected. Surprisingly however, contrasting results have been observed. Whereas some studies supported some associations between MHC and a better immunocompetence (Bonneaud *et al.* 2005; Charbonnel *et al.* 2010), others provided no evidence for an association between immunocompetence and MHC diversity, or specific MHC alleles (e.g.,Ekblom *et al.* 2013; Rodríguez *et al.* 2014). Such limited support may indicate the absence of a general biological association, although other factors cannot yet be discarded, such as a general lack of power owing to insufficient sample size of individuals to detect small effect-size. In addition, previous studies used only a few immunocompetence tests, and consequently considering only a few aspects of the immune system may oversimplify its complexity (Norris & Evans 2000; Roulin *et al.* 2007), which may reduce the likelihood to encounter significant associations between MHC and immunocompetence. Therefore, studies combining multiple proxies of immunocompetence coupled with statistical power analyses can provide a great insight into host-pathogen interactions.

In the present study, we investigated in the barn owl, *Tyto alba*, whether the diversity at both MHC families (MHC-I and MHC-II) is related to eight distinct traits related to the immune response, susceptibility to ectoparasites, and health status. Because the MHC poorly explained the different immune-related parameters, we explored the statistical power to detect associations of different effect sizes. If existing, the association between immunocompetence and the MHC system would be expected to be of little strength, and have gone undetected in our study, owing to little power to detect small-effect sizes. Our study highlights that in studies linking the MHC diversity with fitness-related traits, larger sample sizes and power analyses would be needed in order to achieve a more robust understanding of such a link.

Material and methods

Barn owl population and immunocompetence

The barn owl is a nocturnal, sedentary bird inhabiting open habitats of most continents expect Antarctica. Our study focused on a population breeding in nest-boxes in western Switzerland, which has been monitored for more than 20 years. Since 1996, adults and their offspring have been systemically caught during the breeding period from March to October. Once captured, individuals were ringed, and blood and feather samples collected.

The experiments were conducted in 1998 and 2002. They were parts of partial cross-fostering experiments, where approximately half of the hatchlings were exchanged between a pair of broods (Roulin *et al.* 2007). We investigated eight traits linked to (i) the immune response, (ii) the susceptibility to ectoparasites, and (iii) health status. Here, we provide the essential information of the manipulations, which are described in detail in Roulin *et al.* (2007).

(i) The humoral immunocompetence was challenged by vaccinating nestlings with sheep red blood cells (SRBC), human serum albumin (HSA) and tetanus toxoid (TT) as antigens. This assay elicits the antibody production against the SRBC, HAS and TT antigens. In 1998, 163 nestlings (age: in mean 36 days, range 14-45 days) from 38 broods were immuno-challenged by injecting in the neck 0.25 mL of mixed antigens (30 mg HSA and 8 IU TT) and, subsequently, by injecting 0.1 mL of 20% SRBC (10% v/v in phosphate-buffered saline, PBS, with 10mM phosphate, pH 7.4). To assess the concentration of antibodies, 100 µL of blood were sampled at five different times post-injection: before injection (day 0) and at 3, 8, 13 and 18 days post-injection. The antibody titers directed toward the three antigens were determined using an indirect haemagglutination assay as described in Roulin *et al.* (2000) and El-Lethey *et al.* (2003). As in Roulin *et al.* (2007), we calculated the logarithms of the antibody titers at days 3, 8, 13, and 18 post-injection, and used the mean values for the statistical analyses. The innate and cell-mediated immunocompetence were investigated using also the phytohaemagglutinin (PHA) skin test, which induces a cutaneous inflammatory reaction (Kennedy & Nager 2006; Martin *et al.* 2006; Tella *et al.* 2008). In 2002, 270 nestlings from 67 broods were injected in the wing with 0.1mg of PHA (dissolved in 0.02 mL of PBS). The "PHA-response" was referred to as the thickness difference of the wing web prior to the injection and 24 hours later.

(ii) The susceptibility to parasites was investigated using as a proxy the fecundity of the blood-sucking ectoparasite *Carnus haemapterus*. This fly constitutes the most abundant ectoparasite of nestling barn owls. Their pupae winter inside the nests and hatch in synchronization with the hatching of the owlets to feed on them (Roulin 1998). In 1998, 1458 gravid females were collected on 184 nestlings from 36 nests (the mean number of flies per nestling was 39) and stored in separate Eppendorf tubes at 37 °C (similar to nestling body temperature). After 24 hours, when all flies were dead, the eggs laid on the tubes' surfaces were counted. The mean number of eggs laid by the *C. hemapterus* sampled on each nestling was calculated and used as a response variable in the statistical analysis.

(iii) Health status was first evaluated by measuring the haematocrit and leucocytes blood concentration. These measurements were taken in 1998 on the same broods used to assay the humoral immunity. A capillary tube was used to collect the blood samples from the brachial vein. The samples were immediately centrifuged and the amount of plasma, leucocytes and erythrocytes were measured with a Vernier caliper to the nearest 0.1 mm, and expressed as a percentage of the total amount of blood in the capillary. Second, in 2002, the proportion of immunoglobulin proteins (Ig) in the serum was assessed in the same broods and at the same date as the PHA test described above. The serum was obtained after centrifuging each blood sample and the Ig proportion quantified by densitometric analysis (Roulin *et al.* 2007).

Sequencing, genotyping and explanatory variables related to MHC genes

DNA was extracted using the DNeasy blood and tissue kit following the manufacturer's instructions (Qiagen, Hilden, Germany). We targeted our study on the two exon 2 of MHC class IIβ (MHC-IIB DAB1 and DAB2) and the two exon 3 of MHC class Iα (MHC-I), which encode for the polymorphic regions involved in the peptide recognition (PBR). Both MHC-IIB DAB1 and DAB2 genes were independently amplified (Burri *et al.* 2008b; Chapter 2), whereas the two MHC-I genes were simultaneously co-amplified (Gaigher *et al.* 2016). Briefly, all individuals' MHC were amplified with individual barcode primers. All PCR products were checked/quantified on agarose gels, then purified, quantified (using the

Nanodrop ND-1000 spectrophotometer and the Qubit® 2.0 fluorometer), pooled and finally sequenced with the Illumina MiSeq technology (for MHC-I) and the 454 pyrosequencing protocol (for MHC-IIB). All details of protocols concerning primers development, PCR amplification and 454/Illumina sequencing preparations are described in detail in Burri *et al.* (2008b), Gaigher *et al.* (2016) and in the Chapter 2.

MHC genotyping was performed according to current standard methods for high throughput sequencing (Galan *et al.* 2010; Sommer *et al.* 2013; Lighten *et al.* 2014; Gaigher *et al.* 2016). Reliability of the sequencing was estimated with replicates and using the pattern of allelic segregation in family data. In total, we unambiguously genotyped 96%, 79% and 83% of the individuals for MHC-I, MHC-IIB DAB1 and MHC-IIB DAB2 respectively (Chapter 2). All details of the genotyping are described in Gaigher *et al.* (2016) and in the Chapter 2.

In the present study, we investigated whether diversity and/or specific alleles of the MHC system are linked to immunocompetence traits. The MHC-IIB diversity was calculated for each individual as the zygosity at each MHC-IIB genes. Due to the co-amplification of the two MHC-I loci, the MHC-I diversity was estimated as the number of different alleles per individual. In addition, the diversity was described as the functional allelic divergence. The mean amino acid functional divergence between alleles was calculated for each MHC class and each individual. The functional divergence was calculated as reported in the Chapter 2 using five amino acid physicochemical descriptors (Sandberg *et al.* 1998). The functional divergence was estimated on the entire exon and on codons of the PBR exclusively. PBR codons were identified from Human HLA and Chicken BF for MHC-I (Bjorkman *et al.* 1987; Wallny *et al.* 2006) and from Human HLA for MHC-IIB (Brown *et al.* 1993).

Second, the presence of specific alleles was used as predictors of immunocompetence. However, we faced a large number of alleles in our barn owl population, which can in turn limit the statistical power (69, 25, and 17 MHC-I, MHC-IIB DAB1, and DAB2 alleles, respectively; see also Chapter 2). To overcome this issue, we reduced the number of alleles by clustering them into supertypes on the basis of shared amino acid physicochemical properties of different alleles. This approach is now largely applied in MHC studies given that it is biologically more meaningful (Schwensow *et al.* 2007; Castillo *et al.* 2010; Sepil *et al.* 2012; Lillie *et al.* 2015; Sutton *et al.* 2015). The creation of supertypes was performed using the same amino acid descriptors reported in the previous paragraph (Sandberg *et al.* 1998), and using the approach in Sepil *et al.* (2012). This analysis resulted in 9, 13 and 10 supertypes for MHC-I, MHC-IIB DAB1 and DAB2, respectively. For the analyses, we used only the most common supertypes in the population (frequency higher than 10%) as explanatory variables (three supertypes for MHC-I: MHC-I*S3, MHC-I*S6, MHC-I*S7; three for DAB1: DAB1*S1, DAB1*S9, DAB1*S13; and four for DAB2: DAB2*S3, DAB2*S6, DAB2*S7, DAB2*S10).

Neutral diversity

To also account for a potential effect of genome-wide diversity rather than diversity in the MHC system, each individual was additionally genotyped for a panel of 10 microsatellite markers following the protocol described in Burri *et al.* (2008a). We calculated five different individual estimates of heterozygosity (proportion of heterozygous loci, standardized expected and observed heterozygosity, internal relatedness, homozygosity weighted by locus) using GenHet (Coulon 2010). However, because all estimates were highly correlated (all $r \geq 0.98$, all $P \leq 0.001$), we used only one (homozygosity weighted by locus) for our analyses (Aparicio *et al.* 2006).

Statistical analyses

To test for associations between MHC-related parameters and immunocompetence variables, we ran separate linear mixed models using in each model one of immunocompetence variables as dependent variable. All the mixed models included the random effect "nest of origin" and "nest of rearing" and the full models considered the fixed effects of several factors known to influence the immune response in the barn owl or in other species. (i) The brood size, which can negatively impact the immune function by increasing sibling competition. (ii) The rank in the within-brood hierarchy (i.e. hatching-rank), which can shape the individual infection load (Roulin *et al.* 2003). (iii) The age, which reflects the maturity of immune system. (iv) The body condition, which was calculated as the residuals of body mass on wing length. (v) The sex of individuals, given that sex-specific hormones can generate disparities in immune responses and parasite loads (Moore & Wilson 2002; Klein 2004). And finally, (vi) the season (as Julian date), which can affect the abundance and virulence of parasites.

In a first set of models, we added to the full models different proxies of MHC diversity as independent variables: the zygosity at DAB1 and DAB2 loci (0: homozygote, 1: heterozygote), as well as their interactions, the number of MHC-I alleles, the functional divergence at both MHC classes, and finally the homozygosity at neutral markers. In a second set of models, we added as independent variables several factors indicating the presence of the 10 specific MHC supertypes: MHC-I*S3, MHC-I*S6, MHC-I*S7, DAB1*S1, DAB1*S9, DAB1*S13, DAB2*S3, DAB2*S6, DAB2*S7, DAB2*S10 (variables were coded from 0 to 4, according to their occurrence). We then selected the model/s that better fitted the data by comparing the AICc (the Akaike Information Criterion corrected for sample size) of a set of competing models that considered all the possible combinations of the predictors up to a maximum of 7 parameters per model to avoid model overparametrization. Models with a difference of AICc equal or smaller than 2 to the model with the lowest AICc (ΔAICc ≤ 2) were retained as "best models". Robust estimates and the significance of the effects of the variables included in the set of "best models" were derived using the model averaging approach described in (Grueber *et al.* 2011), which takes into account uncertainty in model selection (i.e., several models being ranked as "best" models). All the analyses were conducted in R (R Core Team 2014), using the packages LME4 (Bates *et al.* 2015) and MuMIn (Bartoń 2016).

We ran *post-hoc* power analyses to assess the power of our study to detect the observed associations between immunocompetence variables and MHC-related parameters. For each immunocompetence variable, we re-ran similar linear mixed models as described above but including this time only one MHC-related parameter at a time. The parameter estimates from each model were then used to create 1000 simulated dependent variables, which were used to re-fit the linear mixed models. Power was then derived as the proportion of tests based on the simulated dependent variables that attained significance ($\alpha \leq 0.05$, two-tailed). In addition, we assessed the power of our study to detect associations of small, moderate, and large effect size (partial correlation coefficients, r, of 0.1, 0.3, and 0.5, respectively (Nakagawa & Cuthill 2007)). In addition to this 'conventional' values, we also assessed the power of our study to detect small-to-moderate effect sizes ($r = 0.2$) given that a preliminary analysis revealed that, for most of the variables in our study a relatively large power (0.8) is already achieved for effect sizes between $r = 0.1$ and $r = 0.3$. To assess the power for each value of effect size considered, we followed the same procedure as described above but the simulated data was generated each time according to one of the magnitudes of effect sizes listed above. For categorical predictors (e.g., specific MHC supertypes), we used Cohen's d values as a measure of the effect sizes, although these are reported as *r* values in the tables to facilitate comparisons across all the MHC parameters (Nakagawa & Cuthill 2007). Again, power for each effect size was finally derived as the proportion of tests based on the dependent variables simulated with a given effect size that attained significance ($\alpha \leq 0.05$, two tails).

Results

MHC diversity in relation to immunocompetence

The summary of sample size used in each analysis and for each immune parameter are shown in Table 1. We first tested whether the zygosity at MHC-IIB genes, number of MHC-I alleles and functional divergence at both classes predict any of the variables related to immune response, susceptibility to ectoparasites, and health status. For immune-response parameters, model selection based on AICc identified a large number of competing models with a ΔAICc < 2 (22 different models for SRBC, 13 for HSA, 14 for TT, 12 for PHA; see Tables S1-4). Although several MHC diversity parameters were included in the sets of best models for each immune-response parameter (Table S1-4), none of them attained significance when using averaged model estimates to take into account model selection uncertainty (Table 2) (MuMIn, Bartoń 2016). The *post-hoc* power analyses indicated that the power of our study to detect associations with effect sizes of 0.1, 0.2, 0.3, and 0.5 (*r* partial correlation coefficients) was, on average for all MHC diversity parameters and all immune response parameters, 0.20, 0.56, 0.82, and 0.94 (Table S5-8). Because a relative large power (>0.8) was attained for a moderate association (0.3, Nakagawa & Cuthill 2007), our study indicated that moderate-to-strong associations between MHC diversity and immune response parameters are unlikely to exist in the population.

Table 1: Summary of sample size in analyses for each immune parameter.

Immune parameter	Year	Sample size	
		MHC diversity	MHC supertypes
SRBC	1998	132	133
HSA	1998	132	133
TT	1998	132	133
PHA	2002	166	169
Ectoparasite fecundity	1998	122	123
Haematocrit	1998	129	130
Leucocyte	1998	129	130
Immunoglobulins	2002	152	155

Concerning the susceptibility to ectoparasites, model selection identified 10 models with a ΔAICc < 2 (Table S9). Based on averaged model estimates, none of the MHC diversity parameters predicted the susceptibility to ectoparasites (Table 2). As for immune-response parameters, the *post-hoc* power analyses revealed that a large power was achieved for a moderate association (for 0.1, 0.2, 0.3, and 0.5, the power was on average 0.18, 0.5, 0.78, 0.94, respectively; Table S10). Finally, for traits related to health status, a large number of models with a ΔAICc < 2 were selected (18 different models for haematocrit, 14 for leucocytes, and 16 for immunoglobulins proteins; Tables S11-13). Only one MHC diversity predictor was significantly associated with the haematocrit. The haematocrit was higher in individuals having fewer MHC-I alleles (Table 2). The power to detect an effect of this predictor with the current effect size (r = -0.177) was moderate: 0.681 (Table S14), and this effect needs to be interpreted with caution. Overall, and in the same line as before, a moderate association reached a large power (for 0.1, 0.2, 0.3, and 0.5; the power was on average 0.24, 0.65, 0.86, 0.95, respectively; Table S14-16).

Table 2: Averaged model estimates of MHC diversity with SRBC, HSA, TT, PHA, fecundity of the ectoparasite, amount of leucocytes, haematocrit, and immunoglobulins proteins. BC: body condition, BS: brood size, Rank: rank raised, HL: homozygoty (neutral), zygDAB1: zygozity DAB1, zygDAB2: zygozity DAB2, divMHCI: number of MHC-I alleles, FdivMHCI: functional divergence MHC-I, FdivMHCIIB: functional divergence MHC-IIB, zygDAB1:zygDAB2: interaction.

Immune parameter	Variables	Estimate	Adjusted SE	z value	Pr(>|z|)	
SRBC	**Age**	**0.096**	**0.020**	**4.888**	**0.000**	***
	zygDAB1	0.450	0.568	0.793	0.428	
	zygDAB2	0.519	0.661	0.785	0.432	
	zygDAB1:zygDAB2	-0.664	0.829	0.801	0.423	
	divMHCI	0.133	0.179	0.745	0.456	
	FdivMHCIIB	0.133	0.467	0.285	0.776	
	HL	-0.102	0.441	0.231	0.817	
	Rank	0.005	0.030	0.163	0.871	
	BS	-0.001	0.015	0.059	0.953	
	Date	0.000	0.001	0.087	0.931	
HSA	**Sex**	**-5.22E-02**	**2.52E-02**	**2.076**	**0.038**	*
	HL	1.57E-02	5.73E-02	0.273	0.785	
	FdivMHCIIB	-1.13E-02	4.38E-02	0.258	0.797	
	zygDAB2	-6.82E-04	6.17E-03	0.111	0.912	
	divMHCI	-5.48E-04	4.63E-03	0.118	0.906	
	FdivMHCI	-2.03E-04	3.60E-03	0.056	0.955	
	zygDAB1	5.84E-04	7.42E-03	0.079	0.937	
	BS	-4.44E-05	1.18E-03	0.038	0.970	
	Age	5.40E-06	1.60E-04	0.034	0.973	
TT	Rank	-1.84E-02	1.29E-02	1.423	0.155	
	Age	1.98E-03	3.47E-03	0.571	0.568	
	divMHCI	1.32E-04	2.24E-03	0.059	0.953	
	BC	-8.21E-07	2.77E-05	0.03	0.976	
	HL	2.64E-06	8.09E-04	0.003	0.997	
	Sex	1.78E-06	3.01E-04	0.006	0.995	
	FdivMHCIIB	-2.26E-06	6.00E-04	0.004	0.997	
	BS	5.04E-08	3.97E-05	0.001	0.999	
PHA	zygDAB2	-0.231	0.628	0.368	0.713	
	FdivMHCI	0.842	0.868	0.970	0.332	
	Date	**0.064**	**0.022**	**2.941**	**0.003**	**
	Rank	**0.714**	**0.330**	**2.165**	**0.030**	*
	BS	0.103	0.308	0.334	0.738	
	divMHCI	0.073	0.325	0.226	0.821	
	Age	0.008	0.052	0.156	0.876	
	BC	0.000	0.002	0.051	0.959	

Table 2 continued

Immune parameter	Variables	Estimate	Adjusted SE	z value	Pr(>|z|)	
Ectoparasite fecundity	zygDAB1	-8.600	4.652	1.849	0.065	
	zygDAB2	-5.248	4.990	1.052	0.293	
	Rank	**1.696**	**0.663**	**2.557**	**0.011**	*
	zygDAB1:zygDAB2	7.261	6.047	1.201	0.230	
	Sex	0.075	0.559	0.134	0.893	
	BS	0.071	0.410	0.174	0.862	
	FdivMHCIIB	-0.385	2.316	0.166	0.868	
	Date	0.002	0.020	0.088	0.930	
Haematocrit	HL	-0.948	1.999	0.474	0.635	
	divMHCI	**-1.127**	**0.451**	**2.502**	**0.012**	*
	Rank	-0.448	0.340	1.317	0.188	
	Sex	-0.435	0.563	0.774	0.439	
	FdivMHCIIB	0.343	1.063	0.323	0.747	
	Age	0.074	0.113	0.651	0.515	
	zygDAB2	0.025	0.174	0.145	0.885	
	Date	1.25E-04	0.003	0.038	0.969	
	BC	-1.09E-05	0.001	0.021	0.984	
Leucocytes	**Date**	**0.022**	**0.007**	**2.897**	**0.004**	**
	FdivMHCIIB	0.560	0.306	1.831	0.067	
	Sex	**0.203**	**0.088**	**2.320**	**0.020**	*
	FdivMHCI	0.037	0.078	0.473	0.636	
	BS	-0.032	0.071	0.454	0.650	
	HL	-0.019	0.125	0.152	0.879	
	divMHCI	0.003	0.019	0.130	0.896	
	Age	-0.001	0.007	0.124	0.902	
	Rank	-0.003	0.021	0.124	0.901	
	zygDAB1	0.000	0.011	0.041	0.967	
Immunoglobulins	FdivMHCIIB	-2.18E-02	1.35E-02	1.615	0.106	
	Date	**4.88E-04**	**9.16E-05**	**5.332**	**0.000**	***
	zygDAB1	5.57E-04	3.22E-03	0.173	0.863	
	Rank	-4.90E-06	1.15E-04	0.043	0.966	
	HL	4.74E-06	3.95E-04	0.012	0.990	
	zygDAB2	1.62E-06	1.25E-04	0.013	0.990	
	FdivMHCI	1.06E-06	8.71E-05	0.012	0.990	
	BS	-6.99E-08	1.80E-05	0.004	0.997	
	Age	8.36E-09	2.57E-06	0.003	0.997	
	BC	2.32E-10	1.92E-07	0.001	0.999	

Specific MHC supertypes related to immunocompetence

We tested whether the 10 most common MHC supertypes (MHC-I*S3, MHC-I*S6, MHC-I*S7, DAB1*S1, DAB1*S9, DAB1*S13, DAB2*S3, DAB2*S6, DAB2*S7, DAB2*S10) are significantly associated with the variables related to immune response, susceptibility to ectoparasites, and health status. For immune-response parameters, AICc-based model selection identified many models with a ΔAICc < 2 (5 models for SRBC, 12 for HSA, 17 for TT, 28 for PHA; see Tables S17-20). Based on averaged model estimates, only the MHC-I*S6 predictor was found to be significantly related with the SRBC response variable (Table 3). This predictor was found in all the models with ΔAIC < 2 (6 models; Table S17).

The antibody production against SRBC increased when individuals carried one copy of the MHC-I*S6 supertype compared to individuals without this supertype. The power analysis associated to the SRBC variable, revealed that the power to detect the observed effect in our data was high (effect size: r = 0.174; power: 0.94; Table S21). No significant difference was observed when comparing the antibody production against SRBC of individuals carrying two copies of the MHC-I*S6 supertype against individuals carrying one copy or against individuals carrying no copy of this supertype. Overall, power analyses for immune-response parameters revealed that a large power was achieved for a small-to-moderate association (for 0.1, 0.2, 0.3, and 0.5, the power was on average 0.53, 0.96, 0.99, 1, respectively; Table S21-24).

For the susceptibility to ectoparasites, model selection identified 14 models with a ΔAICc < 2, including the null model (Table S25). None of the MHC supertypes parameters predicted the susceptibility to ectoparasites, based on averaged model estimates (Table 3). The power associated to this response variable was achieved on average for a moderate association (for 0.1, 0.2, 0.3, and 0.5, the power was on average 0.44, 0.94, 0.99, 1, respectively; Table S26). Finally, for traits related to heath status, 19 different models with a ΔAICc < 2 were retained for haematocrit, 7 for leucocytes, and 12 for immunoglobulins proteins (Tables S27-29). Taking into account model selection uncertainty, we found that carrying the DAB1*S9 and DAB2*S7 supertypes was respectively negatively and positively related with the proportion of immunoglobulins proteins (Table 3). These two predictors were found in all the models with ΔAIC < 2 (13 models; Table S29). However, the current effect sizes for the supertypes DAB1*S9 and DAB2*S7 predictors were very low (-0.023 and 0.048), with a low power associated to these effects (0.079 and 0.254) (Table S32). Overall, a large power was reached on average for a small-to-moderate association (for 0.1, 0.2, 0.3, and 0.5, the power was on average 0.62, 0.98, 0.99, 1; Table S30-32).

Table 3: Averaged model estimates of specific MHC supertype with SRBC, HSA, TT, PHA, fecundity of the ectoparasite, amount of leucocytes, haematocrit, and immunoglobulins proteins.

Immune parameter	Variables	Estimate	Adjusted SE	z value	Pr(>|z|)	
SRBC	**MHC-I*S6-1**	**0.847**	**0.244**	**3.469**	**0.001**	***
	MHC-I*S6-2	0.292	0.262	1.114	0.265	
	Age	**0.085**	**0.020**	**4.335**	**0.000**	***
	DAB2*S6-1	0.120	0.244	0.490	0.624	
	DAB2*S6-2	-0.183	0.617	0.296	0.767	
	DAB2*S10-1	0.007	0.114	0.058	0.954	
	DAB2*S10-2	-0.192	0.492	0.390	0.697	
	Rank	0.003	0.025	0.132	0.895	
	Date	-9.29E-05	0.001	0.074	0.941	
HSA	**Sex**	**-5.46E-02**	**2.48E-02**	**2.201**	**0.028**	*
	DAB1*S13-1	1.19E-03	8.63E-03	0.138	0.890	
	BS	-1.15E-04	1.68E-03	0.068	0.946	
	Rank	-6.40E-05	1.05E-03	0.061	0.951	
	DAB1*S9-1	-3.29E-04	4.40E-03	0.075	0.940	
	DAB1*S9-2	7.24E-04	1.14E-02	0.063	0.950	
	Age	1.49E-05	2.72E-04	0.055	0.956	
	MHC-I*S3-1	-1.23E-04	3.10E-03	0.039	0.969	
	MHC-I*S3-2	-1.53E-04	4.06E-03	0.038	0.970	
	MHC-I*S3-3	-1.60E-04	7.08E-03	0.023	0.982	
TT	Rank	-2.17E-02	1.23E-02	1.764	0.078	
	Age	1.29E-03	2.94E-03	0.439	0.661	
	DAB2*S10-1	2.40E-05	4.12E-03	0.006	0.995	
	DAB2*S10-2	-2.43E-03	2.23E-02	0.109	0.913	
	BC	-5.91E-07	2.32E-05	0.025	0.980	
	Sex	8.76E-06	6.67E-04	0.013	0.990	
	DAB2*S7-1	-1.33E-08	2.96E-05	0.000	1.000	
	DAB2*S7-2	1.05E-07	1.51E-04	0.001	0.999	
	DAB1*S9-1	-4.32E-11	1.60E-06	0.000	1.000	
	DAB1*S9-2	2.94E-10	7.50E-06	0.000	1.000	
PHA	DAB2*S10-1	0.255	0.878	0.291	0.771	
	DAB2*S10-2	4.744	3.918	1.211	0.226	
	DAB2*S6-1	-0.548	1.095	0.500	0.617	
	DAB2*S6-2	3.745	2.920	1.283	0.200	
	MHC-I*S6-1	-1.696	1.312	1.293	0.196	
	MHC-I*S6-2	-1.421	1.448	0.982	0.326	
	MHC-I*S6-3	-1.756	1.873	0.937	0.349	
	Date	**0.082**	**0.023**	**3.503**	**0.000**	***
	Rank	**0.625**	**0.294**	**2.124**	**0.034**	*
	DAB1*S1-1	0.253	0.653	0.388	0.698	
	DAB1*S1-2	1.187	2.527	0.470	0.639	
	DAB2*S3-1	-0.458	1.020	0.449	0.654	
	DAB2*S3-2	-0.402	0.994	0.405	0.686	
	BS	0.042	0.205	0.206	0.836	
	Age	-4.97E-04	0.009	0.052	0.958	
	DAB1*S13-1	-0.002	0.071	0.031	0.975	
	DAB1*S13-2	-8.20E-05	0.114	0.001	0.999	
	BC	2.93E-06	3.41E-04	0.009	0.993	

Table 3 continued

Immune parameter	Variables	Estimate	Adjusted SE	z value	Pr(>|z|)	
Ectoparasite fecundity	MHC-I*S3-1	2.183	2.275	0.959	0.337	
	MHC-I*S3-2	-4.060	3.867	1.050	0.294	
	MHC-I*S3-3	-8.416	10.546	0.798	0.425	
	Rank	1.134	0.649	1.748	0.081	
	DAB2*S7-1	0.013	0.668	0.019	0.985	
	DAB2*S7-2	0.760	3.338	0.228	0.820	
	DAB1*S9-1	0.076	0.678	0.111	0.911	
	DAB1*S9-2	0.901	3.631	0.248	0.804	
	Date	0.002	0.022	0.101	0.920	
	Sex	0.004	0.130	0.029	0.977	
	BS	0.006	0.121	0.054	0.957	
	Age	-3.77E-04	0.014	0.028	0.978	
Haematocrit	DAB1*S9-1	0.765	0.671	1.141	0.254	
	DAB1*S9-2	-1.568	2.070	0.758	0.449	
	DAB2*S6-1	-0.993	0.946	1.049	0.294	
	Rank	-0.395	0.361	1.094	0.274	
	Sex	**-1.094**	**0.523**	**2.095**	**0.036**	*
	Age	0.106	0.132	0.807	0.420	
	BS	-0.072	0.241	0.299	0.765	
	Date	0.001	0.009	0.097	0.923	
	BC	-1.58E-06	2.11E-04	0.008	0.994	
Leucocytes	**Age**	**-0.055**	**0.021**	**2.604**	**0.009**	**
	Date	**0.018**	**0.007**	**2.644**	**0.008**	**
	Rank	**-0.157**	**0.058**	**2.702**	**0.007**	**
	Sex	**0.189**	**0.085**	**2.229**	**0.026**	*
	DAB2*S6-1	-0.046	0.101	0.453	0.651	
	DAB1*S1-1	-0.002	0.023	0.102	0.919	
	DAB1*S1-2	-0.01	0.058	0.175	0.861	
	DAB1*S9-1	0.003	0.028	0.121	0.904	
	DAB1*S9-2	-0.001	0.052	0.011	0.991	
	BS	-0.001	0.014	0.078	0.938	
	BC	-2.49E-06	1.11E-04	0.022	0.982	
Immunoglobulins	**DAB1*S9-1**	**-0.015**	**0.006**	**2.486**	**0.013**	*
	DAB2*S7-1	**0.018**	**0.007**	**2.634**	**0.008**	**
	Date	**4.66E-04**	**9.19E-05**	**5.070**	**0.000**	***
	Rank	-4.15E-06	1.04E-04	0.040	0.968	
	BS	-1.89E-06	9.19E-05	0.021	0.984	
	Age	3.20E-07	1.59E-05	0.020	0.984	
	DAB1*S13-1	2.06E-06	1.44E-04	0.014	0.989	
	DAB1*S13-2	4.26E-06	3.38E-04	0.013	0.990	
	DAB1*S1-1	-2.57E-07	4.91E-05	0.005	0.996	
	DAB1*S1-2	-1.99E-06	2.21E-04	0.009	0.993	
	BC	9.03E-09	1.19E-06	0.008	0.994	

Discussion

Although, the MHC is an essential component of immune pathways, its link with traits related to immunocompetence is poorly understood in wild species, with previous studies supporting the existence of this link in some species but not in others (Kurtz *et al.* 2004; Bonneaud *et al.* 2005; Charbonnel *et al.* 2010; Ekblom *et al.* 2013; Rodríguez *et al.* 2014;

Sin *et al.* 2016). In the present study, we examined whether the MHC diversity and specific MHC supertypes at both MHC classes (MHC-I and MHC-II) predict distinct traits related to immune responses. We found only punctual rather than general effects: a link between SRBC, haematocrit, and the proportion of immunoglobulin proteins with different MHC variables. However, and despite the screening of many immune parameters, MHC variables generally failed to predict immune-related parameters in our barn owl population. We revealed a good power to detect effects of moderate effect size (r = 0.2-0.3), suggesting that, if an association exists, it should be of small-effect size. The study of Roulin *et al.* (2007) previously investigated the effects of season, body condition, sex and age as determinants of immunocompetence in this population (i.e. on the same data set) and a detailed discussion of these effects can be found in that study. Consequently, we will focus our discussion on the aspects related to the MHC-system and its association with immunocompetence, which was not investigated in this species to date. We will first discuss our results in the light of previous findings, and then in relation to what magnitude of effects are expected to exist between the MHC-system and immunocompetence.

Usually, the effects of the MHC system are considered to be mediated by two different means. First, carrying a high MHC diversity in terms of number of alleles, or functional divergence between alleles, is expected to allow for the recognition of a broader range of foreign-derived antigens, which should in turn confer a more efficient immune response (Doherty & Zinkernagel 1975). Second, the immune response is expected to be enhanced when individuals possess certain specific MHC alleles conferring the recognition of a foreign antigen with high affinity, that will result in an efficient cascade of immune events (Hedrick 2002). In addition, such relation can also exist in bird-ectoparasite interactions (Owen *et al.* 2008). To date only a few studies in wild species investigated this and found support for a relationship between the MHC components and immunocompetence. Some studies showed a higher response of carrying a specific MHC allele or genotype when faced with the SRBC or PHA tests (for instance, in house sparrows, *Passer domesticus*, (Bonneaud *et al.* 2005), or in the montane water vole, *Arvicola scherman,* (Charbonnel *et al.* 2010). Less often, an association between MHC diversity and immune response has been observed (but see Kurtz *et al.* 2004). Although we found several significant associations, only one is associated to high power and hence trustworthy, and in line with previous studies, this significant association was found between particular MHC alleles rather than with MHC diversity indexes. Individuals that carried the MHC-I*S6 supertype significantly increased the antibody production against SRBC. The structure of this MHC supertype on the cell surface may have a higher affinity with SRBC antigens, compared to other supertypes, subsequently activating efficient immune response against SRBC antigens. Nevertheless, such a relationship is surprising as the cascade of reactions to ultimately produce a humoral immunity response, is thought to involve the MHC-II pathway. However, the demarcation of both MHC pathways does not seem absolute and the possibility of cross-presentation exists (Heath & Carbone 2001), even if these processes are not clear yet.

Studies investigating immunocompetence in relation to the MHC system have produced contrasting outcomes. For instance, traits related to innate immune responses were investigated in sticklebacks (*Gasterosteus aculeatus*) and European badgers (*Meles meles*), and yielded different results, whereas the level of MHC diversity is correlated in sticklebacks, no association was observed in badgers (Kurtz *et al.* 2004; Sin *et al.* 2016). A similar pattern exists when investigating the humoral or cell-mediated immunity through SRBC or PHA challenges. Some studies showed a correlation with MHC diversity (Bonneaud *et al.* 2005; Charbonnel *et al.* 2010), whereas others found no statistical evidence for an association (Ekblom *et al.* 2013; Rodríguez *et al.* 2014). Such disparity can reflect that the association between the MHC system and immunocompetence depends on species-specific factors, or ecological or evolutionary contexts yet-to-be unraveled. Nevertheless, methodological constraints cannot be discarded and the absence of evidence for a link between immunocompetence and MHC in some studies may be simply related to a lack of power. In line with this idea, we observed in our study that, if any association exists, it would be expected to be of little strength, given that our study had generally little power to detect associations of small effect size. The sample size used in this study is larger or of similar magnitude than in studies conducted in other species, suggesting that previous studies might have also had limited power to detect associations of at least small effect size. This problem was highlighted by Rodríguez *et al.* (2014) in their own work, and indeed, a *post-hoc* power analysis conducted by us revealed that, even the power of this study to detect large effects sizes (r = 0.5) was moderate (0.70) (power for r = 0.1 and 0.3 was 0.07 and 0.29). Similarly, the study of Ekblom *et al.* (2013), where no evidence for an association between MHC-IIB diversity and immune response was found, had little power to detect small-to-moderate effects (power for r = 0.1 and 0.3 was 0.11 and 0.58), although it did have substantial power to detect large effect size associations (power for $r = 0.5$ was 0.97). Although a specific analysis is still required (e.g., using meta-analytical tools), lack of power seems to be an important factor to explain disparity between studies investigating the link between the MHC system and immunocompetence and could be hindering our understanding of the general importance of this genetic system in mediating adaptation.

Despite having low power to detect small effect size associations, our study allows us to discard the existence of moderate-to-large associations between most of the MHC and immune-related parameters investigated in our barn owl population. Studies reporting significant associations showed small-to-moderate effects ($0.26 > r < 0.30$ in Bonneaud *et al.* (2005); $r = 0.17$ in Charbonnel *et al.* (2010); $r = 0.17$ in the present study, Table S21), although it should be noted that we could not calculate the effect size for several other studies, and the generality of associations between the MHC system and immunocompetence being of small-to-moderate effect, needs to be taken cautiously. Surprisingly, very few studies investigating the link between MHC and immunocompetence have provided a detailed theoretical explanation about how the MHC system influences general immunocompetence parameters, which precludes foreseeing the magnitude of the expected associations. We hypothesize that this link may be at best indirect, justifying small effect size associations, if any, given that immunocompetence may include the activation of many immune pathways and effectors in which MHC system is only a part. Other genes are expected to play a major role in the immunocompetence, such as Toll-like receptors (TLRs), which are molecules that act as a first line in the host defense against pathogens and involved the innate inflammatory responses, and additionally the adaptive immunity (Iwasaki & Medzhitov 2010). In addition, other molecules, such as cytokine, play a key contribution in immune signaling, and are involved in both innate and adaptive immunity (Janeway *et al.* 2001). Genes encoding TLRs or cytokines showed polymorphisms to some extent and have been previously shown to be involved in pathogen resistance and to evolve under pathogen-mediated selection (Hollegaard & Bidwell 2006; Downing *et al.* 2010; Alcaide & Edwards 2011; Bollmer *et al.* 2011; Quéméré *et al.* 2015). These groups of genes constitute perfect candidate genes in addition to common MHC-based eco-immunology studies, and can contribute to variation in pathogen resistance in wild species (Vinkler & Albrecht 2009).

In addition, an important effort has been done to understand for instance the effects of PHA on activation of the immune system, and have revealed a complex action of both innate and acquired immunity (Kennedy & Nager 2006; Martin *et al.* 2006; Tella *et al.* 2008). However, more recently a study suggested not to use this challenge as a direct proxy of T-cell responsiveness (Vinkler *et al.* 2010). All these studies highlight the complexity of the effect of PHA on immune pathways, and consequently make the interpretation behind MHC system on PHA difficult. This complexity can be generalized to not only to PHA, but all immuno-ecological challenges, which probably simultaneously involve multi-components of the immune system, making it difficult to highlight the effect of only one specific gene (i.e. MHC). Also, such links can be expected to be species-specific. Each species is confronted with a different set of pathogens, and consequently links may result from specific history of infection, as well as co-evolution between the host and their pathogens (Sironi *et al.* 2015).

If, in general, the association between the MHC system and immunocompetence is proven to be of small effect, we can ultimately wonder about the biological importance of such an association, to explain how individuals adapt to their environment for instance. We used conventional values of effect size as described by Cohen (1988) 30 years ago, based on observed effects in different behavior and psychological studies. Based on this, effects of 0.1, 0.3, and 0.5 are considered as small, medium, and large effect respectively, which may not necessarily be true in all biological contexts. In our opinion, it is necessary to challenge these conventional values in the specific context of the immune system and immunocompetence, given that small and moderate differences in immunocompetence can ultimately represent a strong effect in such context. Consequently, the compilation of the range of effect sizes reported in the literature investigating the immune system in relation to fitness related traits, such as survival, growth, reproduction, and pathogen load, would be required in the future in order to assess the importance of different magnitudes of effect size. This will finally bring a better, more biologically meaningful, understanding of the relationship between immunocompetence and underlying genetic factors, such as the MHC system.

To conclude, we showed in our study that, although we did find a few punctual associations between the MHC system and immunocompetence proxies, we did not find support for a general association between the two. We can discard that moderate-to-strong effect sizes exist between the MHC system and immunocompetence in our population, although low power to detect small effect sizes in our study, indicates that associations of such magnitude could still be possible. We argued that the lack of power of previous studies could explain why associations between the MHC system and immunocompetence seem to exist only in some animal species. We also argued, based on our findings, previous studies, and knowledge on the multiple genetic factors that influence the immune response, that if existing, small-to-moderate associations between the MHC system and immunocompetence are more likely to occur than large effect associations. More studies are certainly needed to confirm this, but this hypothesis should not be neglected in future studies, because in order to detect an effect, an important sampling effort will be needed, to achieve enough power to detect small-effect associations. Consequently, larger sample sizes coupled with power analysis would be of great utility in studies linking the MHC diversity with immunocompetence, and other fitness-related traits, in order to achieve a more robust understanding of this link. Finally, the accumulation of multiple studies over various taxa (including those reporting negative results) will permit to conduct meta-analysis, which will draw a more general conclusion in the relationship between MHC, immunocompetence and, ultimately, fitness.

References

Alcaide M, Edwards SV (2011) Molecular evolution of the Toll-like receptor multigene family in birds. *Molecular Biology and Evolution*, **28**, 1703-1715.

Aparicio JM, Ortego J, Cordero PJ (2006) What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology*, **15**, 4659-4665.

Barribeau SM, Villinger J, Waldman B (2012) Ecological immunogenetics of life-history traits in a model amphibian. *Biology Letters*, **8**, 405-407.

Bartoń K (2016) MuMIn: Multi-Model Inference. R package version 1.15.6. https://CRAN.R-project.org/package=MuMIn.

Bataille A, Cashins SD, Grogan L*, et al.* (2015) Susceptibility of amphibians to chytridiomycosis is associated with MHC class II conformation. *Proceedings of the Royal Society B: Biological Sciences*, **282**.

Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1-48.

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

Bjorkman PJ, Saper MA, Samraoui B*, et al.* (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, **329**, 512-518.

Bollmer JL, Ruder EA, Johnson JA, Eimes JA, Dunn PO (2011) Drift and selection influence geographic variation at immune loci of prairie-chickens. *Molecular Ecology*, **20**, 4695-4706.

Bolnick DI, Snowberg LK, Caporaso JG*, et al.* (2014) Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. *Molecular Ecology*, **23**, 4831-4845.

Bonneaud C, Mazuc J, Chastel O, Westerdahl H, Sorci G (2004) Terminal investment induced by immune challenge and fitness traits associated with major histocompatibility complex in the house sparrow. *Evolution*, **58**, 2823-2830.

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

Brown JH, Jardetzky TS, Gorga JC*, et al.* (1993) Three-dimensional structure of the human class-II histocompatibility antigen HLA-DR1. *Nature*, **364**, 33-39.

Burri R, Antoniazza S, Siverio F*, et al.* (2008a) Isolation and characterization of 21 microsatellite markers in the barn owl (*Tyto alba*). *Molecular Ecology Resources*, **8**, 977-979.

Burri R, Niculita-Hirzel H, Roulin A, Fumagalli L (2008b) Isolation and characterization of major histocompatibility complex (MHC) class IIB genes in the Barn owl (Aves: *Tyto alba*). *Immunogenetics*, **60**, 543-550.

Castillo S, Srithayakumar V, Meunier V, Kyle CJ (2010) Characterization of Major Histocompatibility Complex (MHC) DRB Exon 2 and DRA Exon 3 Fragments in a Primary Terrestrial Rabies Vector (*Procyon lotor*). *PLoS One*, **5**, e12066.

Charbonnel N, Bryja J, Galan M*, et al.* (2010) Negative relationships between cellular immune response, Mhc class II heterozygosity and secondary sexual trait in the montane water vole. *Evolutionary Applications*, **3**, 279-290.

Cohen J (1988) *Statistical power analysis for the behavioural sciences.* New Jersey: Lawrence Earlbaum Associates.

Coulon A (2010) Genhet: an easy-to-use R function to estimate individual heterozygosity. *Molecular Ecology Resources*, **10**, 167-169.

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

Downing T, Lloyd AT, O'Farrelly C, Bradley DG (2010) The differential evolutionary dynamics of avian cytokine and TLR gene classes. *The Journal of Immunology*, **184**, 6993-7000.

Ekblom R, Hasselquist D, Sæther SA*, et al.* (2013) Humoral immunocompetence in relation to condition, size, asymmetry and MHC class II variation in great snipe (*Gallinago media*) males. In: *ArXiv e-prints*.

El-Lethey H, Huber-Eicher B, Jungi TW (2003) Exploration of stress-induced immunosuppression in chickens reveals both stress-resistant and stress-susceptible antigen responses. *Veterinary Immunology and Immunopathology*, **95**, 91-101.

Gaigher A, Burri R, Gharib WH*, et al.* (2016) Family-assisted inference of the genetic architecture of major histocompatibility complex variation. *Molecular Ecology Resources*, **16**, 1353-1364.

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

Grueber CE, Nakagawa S, Laws RJ, Jamieson IG (2011) Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology*, **24**, 699-711.

Heath WR, Carbone FR (2001) Cross-presentation in viral immunity and self-tolerance. *Nature Reviews Immunology*, **1**, 126-134.

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

Hollegaard MV, Bidwell JL (2006) Cytokine gene polymorphism in human disease: on-line databases, Supplement 3. *Genes and Immunity*, **7**, 269-276.

Iwasaki A, Medzhitov R (2010) Regulation of adaptive immunity by the innate immune system. *Science*, **327**, 291-295.

Janeway CA, Travers P, Walport M, Shlomchik MJ (2001) *Immunobiology: the immune system in health and disease* New York: Garland Sience.

Jensen PE (2007) Recent advances in antigen processing and presentation. *Nature Immunology*, **8**, 1041-1048.

Kennedy MW, Nager RG (2006) The perils and prospects of using phytohaemagglutinin in evolutionary ecology. *Trends in Ecology & Evolution*, **21**, 653-655.

Klein J, Sato A (2000) The HLA System. *New England Journal of Medicine*, **343**, 702-709.

Klein SL (2004) Hormonal and immunological mechanisms mediating sex differences in parasite infection. *Parasite Immunology*, **26**, 247-264.

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

Kurtz J, Kalbe M, Aeschlimann PB*, et al.* (2004) Major histocompatibility complex diversity influences parasite resistance and innate immunity in sticklebacks. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 197-204.

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

Lenz TL, Mueller B, Trillmich F, Wolf JBW (2013) Divergent allele advantage at MHC-DRB through direct and maternal genotypic effects and its consequences for allele pool composition and mating. *Proceedings of the Royal Society B: Biological Sciences*, **280**.

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

Lillie M, Grueber CE, Sutton JT*, et al.* (2015) Selection on MHC class II supertypes in the New Zealand endemic Hochstetter's frog. *Bmc Evolutionary Biology*, **15**, 63.

Martin LB, Han P, Lewittes J*, et al.* (2006) Phytohemagglutinin-Induced Skin Swelling in Birds: Histological Support for a Classic Immunoecological Technique. *Functional Ecology*, **20**, 290-299.

Moore SL, Wilson K (2002) Parasites as a Viability Cost of Sexual Selection in Natural Populations of Mammals. *Science*, **297**, 2015-2018.

Nakagawa S, Cuthill IC (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, **82**, 591-605.

Norris K, Evans MR (2000) Ecological immunology: life history trade-offs and immune defense in birds. *Behavioral Ecology*, **11**, 19-26.

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

Osborne AJ, Pearson J, Negro SS*, et al.* (2015) Heterozygote advantage at MHC DRB may influence response to infectious disease epizootics. *Molecular Ecology*, **24**, 1419-1432.

Owen JP, Delany ME, Mullens BA (2008) MHC haplotype involvement in avian resistance to an ectoparasite. *Immunogenetics*, **60**, 621-631.

Owen JP, Nelson AC, Clayton DH (2010) Ecological immunology of bird-ectoparasite systems. *Trends in Parasitology*, **26**, 530-539.

Owens IPF, Wilson K (1999) Immunocompetence: a neglected life history trait or conspicuous red herring? *Trends in Ecology & Evolution*, **14**, 170-172.

Pedersen AB, Babayan SA (2011) Wild immunology. *Molecular Ecology*, **20**, 872-880.

Quéméré E, Galan M, Cosson J-F*, et al.* (2015) Immunogenetic heterogeneity in a widespread ungulate: the European roe deer (*Capreolus capreolus*). *Molecular Ecology*, **24**, 3873-3887.

R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria. URL http://www.R-project.org/.

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

Rodríguez A, Broggi J, Alcaide M, Negro JJ, Figuerola J (2014) Determinants and short-term physiological consequences of PHA immune response in lesser kestrel nestlings. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, **321**, 376-386.

Roulin A (1998) Cycle de reproduction et abondance du diptère parasite *Canus hemapterus* dans les nichées de chouettes effraies *Tyto alba*. *Alauda*, **66**, 265-272.

Roulin A, Brinkhof MWG, Bize P*, et al.* (2003) Which chick is tasty to parasites? The importance of host immunology vs. parasite life history. *Journal of Animal Ecology*, **72**, 75-81.

Roulin A, Christe P, Dijkstra C, Ducrest AL, Jungi TW (2007) Origin-related, environmental, sex, and age determinants of immunocompetence, susceptibility to ectoparasites, and disease symptoms in the barn owl. *Biological Journal of the Linnean Society*, **90**, 703-718.

Roulin A, Jungi TW, Pfister H, Dijkstra C (2000) Female barn owls (*Tyto alba*) advertise good genes. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 937-941.

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

Savage AE, Zamudio KR (2011) MHC genotypes associate with resistance to a frog-killing fungus. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 16705-16710.

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

Sepil I, Lachish S, Hinks AE, Sheldon BC (2013a) Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20130134.

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

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

Shimizu S, Shiina T, Hosomichi K*, et al.* (2004) MHC class IIB gene sequences and expression in quails (Coturnix japonica) selected for high and low antibody responses. *Immunogenetics*, **56**, 280-291.

Sin YW, Annavi G, Dugdale HL*, et al.* (2014) Pathogen burden, co-infection and major histocompatibility complex variability in the European badger (*Meles meles*). *Molecular Ecology*, **23**, 5072-5088.

Sin YW, Newman C, Dugdale HL*, et al.* (2016) No Compensatory Relationship between the Innate and Adaptive Immune System in Wild-Living European Badgers. *PLoS One*, **11**, e0163773.

Sironi M, Cagliani R, Forni D, Clerici M (2015) Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat Rev Genet*, **16**, 224-236.

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

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

Sumners LH, Cox CM, Kim S*, et al.* (2012) Immunological responses to Clostridium perfringens alpha-toxin in two genetically divergent lines of chickens as influenced by major histocompatibility complex genotype. *Poultry Science*, **91**, 592-603.

Sutton JT, Robertson BC, Jamieson IG (2015) MHC variation reflects the bottleneck histories of New Zealand passerines. *Molecular Ecology*, **24**, 362-373.

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

Vinkler M, Albrecht T (2009) The question waiting to be asked: Innate immunity receptors in the perspective of zoological research. *Folia Zoologica*, **58**, 15-28.

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

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

Wallny H-J, Avila D, Hunt LG*, et al.* (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1434-1439.

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

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

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

Zhou H, Lamont SJ (2003) Chicken MHC class I and II gene effects on antibody response kinetics in adult chickens. *Immunogenetics*, **55**, 133-140.

Supplementary material

Supplementary materials include a total of 32 tables. Tables show results of the relationship between MHC predictors and (i) immunocompetence, (ii) fecundity of the ectoparasite, and (iii) health status, as well as results of power analysis.

Table S1: Relationship between MHC diversity predictors and SRBC. AICc: AIC adjusted for small sample sizes; ΔAICc: difference in a AICc model with the AIC of the best model; wAICc: Akaike weight; ER: evidence ratio. BC: body condition, BS: brood size, Rank: rank raised, HL: homozygoty (neutral), zygDAB1: zygozity DAB1, zygDAB2: zygozity DAB2, divMHCI: number of MHC-I alleles, FdivMHCI: functional divergence MHC-I, FdivMHCIIB: functional divergence MHC-IIB, zygDAB1:zygDAB2: interaction.

Models	AICc	ΔAICc	wAICc	ER
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2	402.826	0.000	0.013	
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+Date	402.859	0.032	0.013	1.016
Age+divMHC-I	402.863	0.037	0.013	1.018
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+divMHC-I	403.040	0.213	0.012	1.112
Age+divMHC-I+Date	403.283	0.456	0.010	1.256
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+divMHC-I+Date	403.347	0.521	0.010	1.297
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+divMHC-I+Rank	403.468	0.641	0.010	1.378
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+Rank	403.607	0.781	0.009	1.478
Age+divMHC-I+Rank	403.898	1.071	0.008	1.709
Age+divMHC-I+FdivMHC-IIB	404.033	1.207	0.007	1.828
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+Date+Rank	404.154	1.327	0.007	1.942
Age	404.200	1.374	0.007	1.987
Age+Date	404.281	1.454	0.006	2.069
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+divMHC-I+Date+Rank	404.341	1.515	0.006	2.133
Age+zygDAB1+divMHC-I	404.461	1.635	0.006	2.265
Age+divMHC-I+FdivMHC-IIB+Date	404.557	1.731	0.006	2.376
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+FdivMHC-IIB+Date	404.724	1.897	0.005	2.582
Age+divMHC-I+Date+Rank	404.725	1.898	0.005	2.583
Age+divMHC-I+HL	404.767	1.941	0.005	2.639
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+BS	404.791	1.964	0.005	2.670
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+HL	404.803	1.977	0.005	2.687
Age+zygDAB1+zygDAB2+zygDAB1:zygDAB2+FdivMHC-IIB	404.809	1.982	0.005	2.694
Age+divMHC-I+FdivMHC-I	404.835	2.009	0.005	2.730

Table S2: Relationship between MHC diversity predictors and HSA. The description of the header and predictor abbreviations are found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Sex	-128.557	0.000	0.010	
Sex+divMHC-I	-127.701	0.856	0.007	1.535
Sex+FdivMHC-IIB	-127.629	0.928	0.006	1.590
Sex+HL	-127.546	1.011	0.006	1.658
Sex+FdivMHC-IIB+zygDAB1	-127.503	1.054	0.006	1.694
Sex+Age	-127.443	1.114	0.006	1.746
Sex+zygDAB2	-127.307	1.250	0.005	1.868
Sex+FdivMHC-IIB+zygDAB1+FdivMHC-I	-127.166	1.391	0.005	2.005
Sex+FdivMHC-IIB+zygDAB1+divMHC-I	-126.885	1.672	0.004	2.307
Sex+FdivMHC-IIB+divMHC-I	-126.775	1.782	0.004	2.438
Sex+FdivMHC-IIB+HL	-126.696	1.861	0.004	2.536
Sex+BS	-126.608	1.949	0.004	2.650
Sex+FdivMHC-I	-126.603	1.954	0.004	2.657
Sex+Rank	-126.541	2.016	0.004	2.740

Table S3: Relationship between MHC diversity predictors and TT. The description of the header and predictor abbreviations are found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Age+BC	-133.420	0.000	0.011	
Age+BC+Sex	-132.966	0.453	0.009	1.254
Age	-132.687	0.733	0.008	1.442
Rank	-132.190	1.230	0.006	1.849
BC+Rank	-132.099	1.321	0.006	1.936
Age+BC+Rank	-132.025	1.395	0.006	2.009
Age+BC+FdivMHC-IIB	-131.919	1.501	0.005	2.118
BC+Rank+Sex	-131.792	1.627	0.005	2.256
Age+BC+divMHC-I	-131.746	1.674	0.005	2.309
Age+BC+Rank+Sex	-131.704	1.716	0.005	2.358
Age+BC+BS	-131.660	1.759	0.005	2.410
Age+Rank	-131.593	1.827	0.004	2.493
Age+divMHC-I	-131.562	1.858	0.004	2.532
Age+BC+HL	-131.558	1.862	0.004	2.537
Age+BC+Sex+FdivMHC-IIB	-131.411	2.008	0.004	2.730

Table S4: Relationship between MHC diversity predictors and PHA. The description of the header and predictor abbreviations are found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Date+Rank+FdivMHC-I	1043.904	0.000	0.017	
Date+Rank	1044.237	0.333	0.015	1.181
Date+Rank+FdivMHC-I+BS	1044.939	1.035	0.010	1.678
Date+Rank+FdivMHC-I+Age	1045.226	1.323	0.009	1.937
Date+Rank+FdivMHC-I+BC	1045.311	1.407	0.009	2.021
Date+Rank+FdivMHC-I+zygDAB2	1045.419	1.515	0.008	2.133
Date+Rank+BS	1045.432	1.528	0.008	2.147
Date+Rank+BC	1045.499	1.595	0.008	2.220
Date+Rank+divMHC-I	1045.532	1.628	0.008	2.257
Date+Rank+zygDAB2	1045.754	1.850	0.007	2.522
Date+Rank+Age	1045.847	1.943	0.007	2.642
Date+Rank+FdivMHC-I+divMHC-I	1045.891	1.988	0.006	2.701
Date+Rank+FdivMHC-I+Sex	1046.088	2.184	0.006	2.980

Table S5: Power analysis results for the MHC diversity associated with SRBC. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
zygDAB1	0.063	0.192	0.106	0.189	0.516	0.846	0.999
zygDAB2	0.027	0.169	0.063	0.199	0.643	0.938	1
divMHC-I	0.159	0.156	0.522	0.25	0.712	0.973	1
FdivMHC-IIB	0.066	0.154	0.141	0.257	0.71	0.97	1
FdivMHC-I	0.009	0.156	0.049	0.254	0.729	0.972	1
HL	-0.045	0.160	0.096	0.254	0.688	0.958	1
zygDAB1:zygDAB2	-0.651	0.409	0.871	0.095	0.159	0.29	0.656

Table S6: Power analysis results for the MHC diversity associated with HSA. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
zygDAB1	0.022	0.218	0.05	0.156	0.436	0.77	0.995
zygDAB2	-0.085	0.189	0.15	0.188	0.549	0.881	0.998
divMHC-I	-0.103	0.177	0.201	0.191	0.586	0.905	1
FdivMHC-IIB	-0.095	0.173	0.196	0.211	0.628	0.916	1
FdivMHC-I	-0.036	0.175	0.066	0.207	0.604	0.912	1
HL	0.072	0.181	0.126	0.194	0.593	0.892	1
zygDAB1:zygDAB2	0.148	0.474	0.099	0.081	0.145	0.242	0.564

Table S7: Power analysis results for the MHC diversity associated with TT. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
zygDAB1	0.020	0.207	0.051	0.163	0.48	0.822	0.995
zygDAB2	0.033	0.172	0.061	0.172	0.618	0.926	1
divMHC-I	0.042	0.168	0.07	0.231	0.654	0.951	0.999
FdivMHC-IIB	-0.070	0.160	0.143	0.225	0.682	0.947	1
FdivMHC-I	-0.004	0.155	0.055	0.225	0.696	0.964	1
HL	0.050	0.164	0.094	0.231	0.649	0.953	1
zygDAB1:zygDAB2	0.225	0.434	0.167	0.086	0.145	0.255	0.601

Table S8: Power analysis results for the MHC diversity associated with PHA. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
zygDAB1	0.004	0.261	0.04	0.113	0.324	0.622	0.966
zygDAB2	-0.052	0.163	0.085	0.205	0.662	0.931	1
divMHC-I	0.075	0.147	0.182	0.278	0.742	0.968	1
FdivMHC-IIB	0.009	0.148	0.051	0.271	0.763	0.975	1
FdivMHC-I	0.127	0.151	0.376	0.272	0.721	0.974	1
HL	-0.011	0.150	0.053	0.255	0.731	0.972	1
zygDAB1:zygDAB2	-0.405	0.545	0.32	0.074	0.121	0.203	0.44

Table S9: Relationship between MHC diversity predictors and ectoparasite fecundity. The description of the header and predictor abbreviations are found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Rank+zygDAB1	943.481	0.000	0.012	
Rank	944.297	0.816	0.008	1.503
Rank+zygDAB1+Date	944.300	0.818	0.008	1.506
Rank+zygDAB1+BS	944.446	0.964	0.007	1.620
Rank+Date	944.729	1.248	0.006	1.866
Rank+FdivMHC-IIB	944.963	1.481	0.006	2.097
Rank+BS	945.027	1.545	0.006	2.165
Rank+zygDAB1+zygDAB2+zygDAB1:zygDAB2	945.233	1.752	0.005	2.401
Rank+zygDAB1+Sex	945.238	1.757	0.005	2.407
Rank+Date+FdivMHC-IIB	945.402	1.921	0.005	2.613
Rank+zygDAB1+BC	945.491	2.010	0.004	2.731

Table S10: Power analysis results for the MHC diversity associated with ectoparasite fecundity. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
zygDAB1	-0.186	0.221	0.378	0.166	0.44	0.77	0.991
zygDAB2	0.008	0.191	0.062	0.188	0.52	0.837	0.998
divMHC-I	-0.012	0.176	0.066	0.204	0.585	0.9	0.998
FdivMHC-IIB	-0.122	0.176	0.308	0.226	0.639	0.905	1
FdivMHC-I	-0.019	0.170	0.065	0.203	0.6	0.917	1
HL	-0.009	0.182	0.053	0.189	0.559	0.876	1
zygDAB1:zygDAB2	0.451	0.464	0.491	0.068	0.136	0.256	0.558

Table S11: Relationship between MHC diversity predictors and haematocrit. The description of the header and predictor abbreviations are found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
divMHC-I+Age	661.288	0.000	0.013	
divMHC-I+Age+Sex	661.379	0.092	0.013	1.047
divMHC-I+Rank	661.474	0.186	0.012	1.098
divMHC-I+Sex+Rank	661.703	0.415	0.011	1.231
divMHC-I+Age+Date	662.528	1.241	0.007	1.860
divMHC-I+Age+Rank	662.639	1.352	0.007	1.966
divMHC-I+Age+zygDAB2	662.772	1.485	0.006	2.101
divMHC-I+Age+Sex+Date	662.788	1.500	0.006	2.117
divMHC-I+Age+Sex+zygDAB2	662.810	1.523	0.006	2.141
divMHC-I+Age+Sex+FdivMHC-IIB	662.871	1.583	0.006	2.207
divMHC-I+Age+FdivMHC-IIB	662.907	1.619	0.006	2.247
divMHC-I+Age+Sex+Rank	662.948	1.661	0.006	2.294
divMHC-I+Age+BC	662.972	1.684	0.006	2.321
divMHC-I+Rank+HL	663.173	1.885	0.005	2.567
divMHC-I+Sex+Rank+HL	663.193	1.905	0.005	2.592
divMHC-I+Rank+BC	663.232	1.945	0.005	2.644
divMHC-I+Rank+FdivMHC-IIB	663.258	1.970	0.005	2.678
divMHC-I+Sex+HL	663.263	1.976	0.005	2.686
divMHC-I+Sex+Rank+FdivMHC-IIB	663.353	2.066	0.005	2.809

Table S12: Relationship between MHC diversity predictors and amount of leucocytes. The description of the header and predictor abbreviations are found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Date+Sex+Age+Rank+FdivMHC-IIB+FdivMHC-I+HL	226.221	0.000	0.014	
Date+Sex+Age+Rank+FdivMHC-IIB+FdivMHC-I	226.689	0.468	0.011	1.264
Date+Sex+Age+Rank+FdivMHC-IIB+HL	226.738	0.517	0.011	1.295
Date+Sex+FdivMHC-IIB+FdivMHC-I+BS	226.834	0.613	0.010	1.359
Date+Sex+Age+Rank+HL+zygDAB1	227.297	1.076	0.008	1.713
Date+Sex+Age+Rank+FdivMHC-IIB	227.324	1.103	0.008	1.736
Date+Sex+FdivMHC-IIB+FdivMHC-I	227.567	1.346	0.007	1.960
Date+Sex+Age+Rank+zygDAB1	227.629	1.408	0.007	2.022
Date+Sex+Age+Rank	227.687	1.466	0.007	2.081
Date+Sex+Age+Rank+HL	227.693	1.472	0.007	2.087
Date+Sex+FdivMHC-IIB+FdivMHC-I+BS+HL	227.807	1.586	0.006	2.210
Date+Sex+FdivMHC-IIB+BS	227.807	1.586	0.006	2.211
Date+Sex+FdivMHC-IIB+BS	227.902	1.681	0.006	2.317
Date+Sex+FdivMHC-IIB	228.185	1.964	0.005	2.670
Date+Sex+Age+Rank+FdivMHC-IIB+FdivMHC-I+BS	228.307	2.086	0.005	2.837

Table S13: Relationship between MHC diversity predictors and immunoglobulins proteins (Ig). The description of the header and predictor abbreviations are found in the caption of the table S1.

Model	AICc	ΔAICc	wAICc	ER
Date+FdivMHC-IIB+zygDAB1	-718.001	0.000	0.015	
Date+FdivMHC-IIB+zygDAB1+FdivMHC-I	-717.439	0.562	0.011	1.324
Date+FdivMHC-IIB+zygDAB1+zygDAB2	-717.215	0.786	0.010	1.482
Date+FdivMHC-IIB+zygDAB1+Rank	-717.150	0.851	0.010	1.530
Date+FdivMHC-IIB+zygDAB1+FdivMHC-I+zygDAB2	-716.876	1.125	0.008	1.755
Date+FdivMHC-IIB+zygDAB1+Age	-716.861	1.140	0.008	1.768
Date+FdivMHC-IIB+zygDAB1+BC	-716.656	1.345	0.008	1.959
Date+FdivMHC-IIB	-716.489	1.512	0.007	2.130
Date+FdivMHC-IIB+zygDAB1+Rank+FdivMHC-I	-716.472	1.529	0.007	2.148
Date+FdivMHC-IIB+zygDAB1+FdivMHC-I+Age	-716.416	1.585	0.007	2.209
Date+FdivMHC-IIB+Rank	-716.291	1.710	0.006	2.352
Date+FdivMHC-IIB+zygDAB1+Rank+zygDAB2	-716.184	1.817	0.006	2.481
Date+FdivMHC-IIB+zygDAB1+BS	-716.155	1.846	0.006	2.517
Date+FdivMHC-IIB+zygDAB1+HL	-716.136	1.865	0.006	2.541
Date+FdivMHC-IIB+zygDAB1+zygDAB2+Age	-716.044	1.957	0.006	2.661
Date+FdivMHC-IIB+zygDAB1+FdivMHC-I+BC	-716.028	1.973	0.006	2.682
Date+Rank	-715.981	2.019	0.005	2.745

Table S14: Power analysis results for the MHC diversity associated with haematocrit. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
zygDAB1	-0.004	0.177	0.056	0.212	0.626	0.913	1
zygDAB2	0.108	0.156	0.299	0.265	0.682	0.96	1
divMHC-I	-0.177	0.138	0.681	0.273	0.784	0.992	1
FdivMHC-IIB	0.084	0.137	0.235	0.307	0.803	0.993	1
FdivMHC-I	-0.088	0.134	0.247	0.307	0.829	0.989	1
HL	-0.021	0.148	0.066	0.28	0.765	0.984	1
zygDAB1:zygDAB2	-0.015	0.395	0.063	0.098	0.191	0.332	0.694

Table S15: Power analysis results for the MHC diversity associated with amount of leucocytes. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
zygDAB1	0.138	0.174	0.34	0.194	0.579	0.916	1
zygDAB2	0.009	0.145	0.066	0.262	0.733	0.975	1
divMHC-I	0.081	0.137	0.232	0.328	0.828	0.987	1
FdivMHC-IIB	0.114	0.133	0.373	0.298	0.821	0.995	1
FdivMHC-I	0.077	0.125	0.226	0.346	0.875	0.999	1
HL	-0.103	0.140	0.296	0.281	0.795	0.983	1
zygDAB1:zygDAB2	0.103	0.355	0.098	0.096	0.211	0.386	0.761

Table S16: Power analysis results for the MHC diversity associated with immunoglobulins proteins. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
zygDAB1	0.058	0.254	0.083	0.139	0.354	0.676	0.976
zygDAB2	0.020	0.165	0.06	0.212	0.652	0.932	1
divMHC-I	-0.047	0.148	0.1	0.255	0.734	0.96	1
FdivMHC-IIB	-0.115	0.147	0.287	0.224	0.733	0.974	1
FdivMHC-I	0.065	0.150	0.133	0.263	0.735	0.97	1
HL	0.064	0.147	0.137	0.261	0.737	0.971	1
zygDAB1:zygDAB2	-0.143	0.521	0.087	0.066	0.119	0.208	0.465

Table S17: Relationship between MHC supertype predictors and SRBC. The description of the header is found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
MHC-I*S6+Age	386.722	0.000	0.038	
MHC-I*S6+Age+Date	386.832	0.110	0.036	1.057
MHC-I*S6+Age+Rank	388.363	1.642	0.017	2.272
MHC-I*S6+Age+DAB2*S10	388.563	1.842	0.015	2.511
MHC-I*S6+Age+DAB2*S6	388.693	1.971	0.014	2.679
MHC-I*S6+Age+Date+Rank	388.741	2.019	0.014	2.745

Table S18: Relationship between MHC supertype predictors and HSA. The description of the header is found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Sex	-131.238	0.000	0.010	
Sex+Age	-130.868	0.370	0.008	1.203
Sex+MHC-I*S3	-130.680	0.558	0.008	1.322
Sex+DAB1*S9	-130.230	1.008	0.006	1.655
Sex+Age+MHC-I*S3	-130.137	1.101	0.006	1.734
DAB1*S13+Sex	-129.997	1.242	0.005	1.861
DAB1*S13+Sex+Age	-129.808	1.430	0.005	2.045
Sex+Rank	-129.760	1.478	0.005	2.094
Sex+Age+DAB1*S9	-129.716	1.523	0.005	2.141
Sex+BS	-129.673	1.565	0.005	2.187
Sex+DAB1*S9+MHC-I*S3	-129.534	1.704	0.004	2.345
Sex+MHC-I*S3+BS	-129.414	1.824	0.004	2.489
Sex+Age+Date	-129.227	2.011	0.004	2.733

Table S19: Relationship between MHC supertype predictors and TT. The description of the header is found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
BC+Sex+Age+DAB2*S10	-132.428	0.000	0.009	
BC+Age+DAB2*S10	-131.645	0.784	0.006	1.480
Age+DAB2*S10	-131.475	0.953	0.006	1.611
BC+Sex+Age+Rank+DAB2*S10	-131.223	1.206	0.005	1.827
Rank	-131.084	1.345	0.005	1.959
Sex+Age+DAB2*S10	-131.035	1.393	0.005	2.007
BC+Sex+Rank+DAB2*S10	-130.964	1.464	0.005	2.079
BC+Sex+Rank	-130.917	1.512	0.004	2.129
BC+Sex+Age+DAB2*S10+DAB2*S7	-130.875	1.554	0.004	2.174
BC+Sex+Age	-130.815	1.613	0.004	2.240
BC+Age	-130.801	1.627	0.004	2.256
BC+Sex+Rank+DAB2*S10+DAB2*S7	-130.786	1.642	0.004	2.273
BC+Rank	-130.726	1.702	0.004	2.342
Age	-130.567	1.861	0.004	2.536
BC+Sex+Age+DAB2*S10+DAB1*S9	-130.542	1.886	0.004	2.568
Rank+DAB2*S10	-130.532	1.896	0.004	2.581
BC+Sex+Rank+DAB2*S7	-130.459	1.969	0.004	2.677
Sex+Rank	-130.370	2.058	0.003	2.798

Table S20: Relationship between MHC supertype predictors and PHA. The description of the header is found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Date+Rank+DAB1*S1+DAB2*S3	1032.056	0.000	0.005	
Date+Rank	1032.341	0.285	0.004	1.153
Date+Rank+DAB2*S3	1032.387	0.331	0.004	1.180
Date+Rank+DAB2*S6	1032.835	0.780	0.003	1.477
Date+Rank+DAB2*S10	1032.910	0.855	0.003	1.533
Date+Rank+DAB1*S1	1032.970	0.915	0.003	1.580
Date+Rank+DAB1*S1+DAB2*S6	1033.095	1.039	0.003	1.682
Date+Rank+DAB1*S1+DAB2*S3+BS	1033.176	1.120	0.003	1.751
Date+Rank+DAB2*S10+MHC-I*S6	1033.234	1.178	0.003	1.803
Date+Rank+BS	1033.265	1.209	0.003	1.831
Date+Rank+DAB2*S3+BS	1033.276	1.220	0.003	1.840
Date+Rank+DAB1*S1+DAB2*S3+DAB2*S6	1033.336	1.281	0.003	1.897
Date+Rank+DAB2*S6+BS	1033.538	1.482	0.002	2.099
Date+Rank+DAB2*S3+DAB2*S6	1033.546	1.490	0.002	2.107
Date+Rank+MHC-I*S6	1033.581	1.525	0.002	2.144
Date+DAB1*S1+DAB2*S3+Age	1033.727	1.671	0.002	2.307
Date+Rank+DAB2*S6+DAB2*S10	1033.738	1.683	0.002	2.320
Date+DAB1*S1+DAB2*S3	1033.820	1.765	0.002	2.417
Date+Rank+Age	1033.821	1.765	0.002	2.417
Date+Rank+DAB2*S6+DAB2*S10+MHC-I*S6	1033.826	1.771	0.002	2.424
Date+Rank+DAB1*S13	1033.858	1.802	0.002	2.462
Date+Rank+DAB2*S10+BS	1033.914	1.858	0.002	2.532
Date+Rank+BC	1033.925	1.870	0.002	2.547
Date+Rank+DAB1*S1+DAB2*S3+BC	1033.937	1.882	0.002	2.562
Date+DAB1*S1+DAB2*S3+BS	1033.988	1.932	0.002	2.627
Date+Rank+DAB2*S6+MHC-I*S6	1033.994	1.939	0.002	2.636
Date+DAB1*S1+DAB2*S3+BS+Age	1034.006	1.951	0.002	2.652
Date+Rank+DAB2*S10+MHC-I*S6+BS	1034.010	1.954	0.002	2.657
Date+Rank+DAB1*S1+DAB2*S10	1034.060	2.004	0.002	2.724

Table S21: Power analysis results for specific MHC supertype associated with SRBC. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
DAB1*S1.1	0.019	0.089	0.078	0.576	0.992	1	1
DAB1*S1.2	0.023	0.107	0.073	0.454	0.954	1	1
DAB1*S9.1	-0.030	0.080	0.106	0.693	1	1	1
DAB1*S9.2	-0.010	0.098	0.048	0.525	0.978	1	1
DAB1*S13.1	0.022	0.079	0.083	0.694	0.997	1	1
DAB1*S13.2	-0.021	0.086	0.081	0.604	0.994	1	1
DAB2*S3.1	0.016	0.094	0.069	0.574	0.989	1	1
DAB2*S3.2	0.020	0.119	0.059	0.377	0.91	1	1
DAB2*S7.1	0.009	0.078	0.057	0.71	1	1	1
DAB2*S7.2	-0.008	0.092	0.052	0.565	0.985	1	1
DAB2*S10.1	0.019	0.082	0.08	0.651	0.994	1	1
DAB2*S10.2	-0.051	0.088	0.212	0.603	0.994	1	1
DAB2*S6.1	0.047	0.087	0.175	0.597	0.997	1	1
DAB2*S6.2	0.003	0.082	0.057	0.675	0.997	1	1
MHC-I*S6.1	0.174	0.097	0.937	0.546	0.979	1	1
MHC-I*S6.2	0.070	0.099	0.302	0.509	0.98	1	1
MHC-I*S6.3	0.044	0.122	0.109	0.355	0.89	1	1
MHC-I*S7.1	-0.025	0.089	0.101	0.599	0.99	1	1
MHC-I*S7.2	-0.075	0.106	0.288	0.474	0.951	1	1
MHC-I*S7.3	-0.013	0.114	0.06	0.416	0.935	1	1
MHC-I*S7.4	0.024	0.111	0.071	0.448	0.958	1	1
MHC-I*S3.1	-0.010	0.086	0.053	0.623	0.993	1	1
MHC-I*S3.2	0.037	0.100	0.14	0.488	0.969	1	1
MHC-I*S3.3	-0.074	0.096	0.339	0.522	0.985	1	1

Table S22: Power analysis results for specific MHC supertype associated with HSA. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
DAB1*S1.1	0.021	0.100	0.079	0.492	0.967	1	1
DAB1*S1.2	-0.041	0.121	0.112	0.372	0.913	1	1
DAB1*S9.1	-0.060	0.090	0.25	0.599	0.995	1	1
DAB1*S9.2	0.048	0.109	0.139	0.406	0.955	1	1
DAB1*S13.1	0.048	0.088	0.209	0.621	0.995	1	1
DAB1*S13.2	-0.128	0.094	0.75	0.523	0.99	1	1
DAB2*S3.1	-0.004	0.106	0.045	0.449	0.965	1	1
DAB2*S3.2	0.043	0.133	0.095	0.31	0.833	1	1
DAB2*S7.1	-0.031	0.089	0.108	0.587	0.996	1	1
DAB2*S7.2	0.048	0.102	0.15	0.462	0.971	1	1
DAB2*S10.1	0.001	0.094	0.054	0.528	0.983	1	1
DAB2*S10.2	-0.028	0.100	0.099	0.497	0.979	1	1
DAB2*S6.1	0.000	0.098	0.057	0.51	0.981	1	1
DAB2*S6.2	-0.001	0.094	0.057	0.55	0.991	1	1
MHC-I*S6.1	-0.029	0.112	0.086	0.417	0.933	1	1
MHC-I*S6.2	0.019	0.115	0.06	0.407	0.929	1	1
MHC-I*S6.3	-0.075	0.143	0.186	0.306	0.794	1	1
MHC-I*S7.1	-0.053	0.097	0.203	0.524	0.977	1	1
MHC-I*S7.2	0.035	0.116	0.099	0.376	0.922	1	1
MHC-I*S7.3	0.053	0.123	0.131	0.332	0.897	1	1
MHC-I*S7.4	0.107	0.120	0.43	0.395	0.93	1	1
MHC-I*S3.1	-0.104	0.095	0.57	0.537	0.986	1	1
MHC-I*S3.2	-0.079	0.110	0.311	0.436	0.929	1	1
MHC-I*S3.3	-0.047	0.110	0.14	0.442	0.947	1	1

Table S23: Power analysis results for specific MHC supertype associated with TT. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
DAB1*S1.1	0.004	0.089	0.061	0.576	0.989	1	1
DAB1*S1.2	-0.032	0.109	0.101	0.414	0.949	1	1
DAB1*S9.1	-0.019	0.081	0.072	0.662	0.999	1	1
DAB1*S9.2	0.095	0.106	0.393	0.43	0.952	1	1
DAB1*S13.1	-0.013	0.083	0.053	0.644	0.998	1	1
DAB1*S13.2	0.011	0.092	0.059	0.599	0.992	1	1
DAB2*S3.1	0.056	0.099	0.203	0.498	0.977	1	1
DAB2*S3.2	0.054	0.121	0.127	0.356	0.894	1	1
DAB2*S7.1	-0.030	0.081	0.111	0.672	0.998	1	1
DAB2*S7.2	0.089	0.099	0.405	0.488	0.976	1	1
DAB2*S10.1	-0.009	0.088	0.067	0.602	0.989	1	1
DAB2*S10.2	-0.102	0.088	0.606	0.583	0.992	1	1
DAB2*S6.1	-0.012	0.088	0.058	0.586	0.995	1	1
DAB2*S6.2	-0.047	0.088	0.18	0.624	0.999	1	1
MHC-I*S6.1	-0.030	0.102	0.091	0.477	0.958	1	1
MHC-I*S6.2	0.031	0.103	0.079	0.477	0.967	1	1
MHC-I*S6.3	-0.014	0.128	0.062	0.344	0.877	1	1
MHC-I*S7.1	-0.006	0.090	0.052	0.573	0.995	1	1
MHC-I*S7.2	-0.070	0.113	0.23	0.427	0.929	1	1
MHC-I*S7.3	-0.022	0.110	0.06	0.418	0.941	1	1
MHC-I*S7.4	0.022	0.116	0.067	0.423	0.936	1	1
MHC-I*S3.1	-0.097	0.086	0.572	0.603	0.994	1	1
MHC-I*S3.2	-0.043	0.106	0.132	0.462	0.942	1	1
MHC-I*S3.3	-0.044	0.103	0.137	0.488	0.965	1	1

Table S24: Power analysis results for specific MHC supertype associated with PHA. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
DAB1*S1.1	0.045	0.078	0.194	0.723	1	1	1
DAB1*S1.2	0.081	0.095	0.383	0.528	0.992	1	1
DAB1*S9	0.005	0.075	0.055	0.712	1	1	1
DAB1*S13.1	-0.061	0.077	0.32	0.682	0.999	1	1
DAB1*S13.2	-0.006	0.084	0.055	0.661	1	1	1
DAB2*S3.1	-0.075	0.086	0.399	0.611	0.996	1	1
DAB2*S3.2	-0.076	0.101	0.283	0.439	0.967	1	1
DAB2*S7	0.013	0.074	0.041	0.719	1	1	1
DAB2*S10.1	-0.001	0.073	0.049	0.761	1	1	1
DAB2*S10.2	0.080	0.080	0.466	0.654	0.999	1	1
DAB2*S6.1	-0.026	0.079	0.094	0.667	0.999	1	1
DAB2*S6.2	0.075	0.086	0.377	0.588	0.996	1	1
MHC-I*S6.1	-0.068	0.083	0.364	0.67	0.997	1	1
MHC-I*S6.2	-0.050	0.100	0.152	0.472	0.963	1	1
MHC-I*S6.3	-0.065	0.109	0.217	0.411	0.931	1	1
MHC-I*S7.1	-0.017	0.081	0.089	0.683	0.999	1	1
MHC-I*S7.2	-0.050	0.088	0.203	0.587	0.996	1	1
MHC-I*S7.3	-0.122	0.097	0.686	0.521	0.975	1	1
MHC-I*S3.1	0.004	0.080	0.051	0.692	0.999	1	1
MHC-I*S3.2	0.040	0.094	0.142	0.559	0.993	1	1
MHC-I*S3.3	0.015	0.099	0.069	0.503	0.983	1	1
MHC-I*S3.4	0.083	0.099	0.348	0.474	0.981	1	1

Table S25: Relationship between MHC supertype predictors and fecundity of the ectoparasite. The description of the header is found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Rank	939.844	0.000	0.009	
Rank+Date	940.143	0.300	0.007	1.162
Rank+BS	940.653	0.809	0.006	1.499
Date	940.913	1.069	0.005	1.707
Age	941.187	1.344	0.004	1.958
Date+Age	941.267	1.423	0.004	2.037
1	941.342	1.498	0.004	2.115
Age+BS	941.495	1.652	0.004	2.284
Rank+DAB1*S9	941.575	1.731	0.004	2.376
BS	941.590	1.747	0.004	2.395
Rank+Sex	941.683	1.840	0.003	2.509
Rank+MHC-I*S3	941.689	1.845	0.003	2.516
Rank+Date+DAB1*S9	941.714	1.870	0.003	2.548
Rank+DAB2*S7	941.751	1.908	0.003	2.596
Rank+DAB2*S6	941.913	2.070	0.003	2.814

Table S26: Power analysis results for specific MHC supertype associated with fecundity of the ectoparasite. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
DAB1*S1.1	-0.060	0.096	0.23	0	0.986	1	1
DAB1*S1.2	0.015	0.119	0.063	0.382	0.905	1	1
DAB1*S9.1	0.019	0.087	0.072	0.597	0.998	1	1
DAB1*S9.2	0.090	0.109	0.354	0.417	0.934	1	1
DAB1*S13.1	-0.002	0.087	0.05	0.611	0.997	1	1
DAB1*S13.2	-0.111	0.097	0.635	0.543	0.986	1	1
DAB2*S3.1	0.044	0.108	0.122	0.429	0.954	1	1
DAB2*S3.2	0.012	0.132	0.054	0.307	0.84	1	1
DAB2*S7.1	0.000	0.089	0.048	0.598	0.996	1	1
DAB2*S7.2	0.084	0.103	0.345	0.455	0.952	1	1
DAB2*S10.1	-0.036	0.093	0.125	0.557	0.993	1	1
DAB2*S10.2	-0.012	0.096	0.071	0.501	0.982	1	1
DAB2*S6	-0.007	0.095	0.057	0.533	0.981	1	1
MHC-I*S6.1	-0.057	0.113	0.169	0.413	0.929	1	1
MHC-I*S6.2	0.003	0.114	0.055	0.376	0.93	1	1
MHC-I*S6.3	0.006	0.140	0.056	0.295	0.809	1	1
MHC-I*S7.1	0.002	0.092	0.041	0.531	0.99	1	1
MHC-I*S7.2	0.101	0.115	0.407	0.399	0.935	1	1
MHC-I*S7.3	-0.016	0.115	0.056	0.418	0.927	1	1
MHC-I*S7.4	0.000	0.126	0.055	0.327	0.854	1	1
MHC-I*S3.1	0.059	0.095	0.245	0.537	0.986	1	1
MHC-I*S3.2	-0.072	0.114	0.235	0.395	0.937	1	1
MHC-I*S3.3	-0.057	0.107	0.201	0.475	0.963	1	1

Table S27: Relationship between MHC supertype predictors and haematocrit. The description of the header is found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Sex+Age	650.984	0.000	0.008	
Sex+Rank	651.197	0.213	0.008	1.112
Sex+DAB2*S6+Rank	651.262	0.277	0.007	1.149
Sex+Age+DAB2*S6	651.513	0.529	0.006	1.303
Sex+Age+DAB2*S6+DAB1*S9	651.828	0.844	0.006	1.525
Sex+Age+DAB1*S9	652.010	1.026	0.005	1.670
Sex+Age+Date	652.316	1.332	0.004	1.946
Sex+Age+Rank	652.324	1.340	0.004	1.954
Sex+DAB2*S6+DAB1*S9+Rank	652.347	1.362	0.004	1.976
Sex+Age+BS	652.471	1.487	0.004	2.103
Sex+Age+DAB2*S6+DAB1*S9+BS+Date	652.487	1.503	0.004	2.120
Sex+Age+DAB2*S6+Rank	652.545	1.561	0.004	2.182
Sex+Age+DAB2*S6+BS	652.613	1.628	0.004	2.257
Sex+Age+BS+Date	652.673	1.688	0.004	2.326
Sex+Age+DAB2*S6+DAB1*S9+BS	652.676	1.691	0.004	2.330
Age	652.740	1.755	0.004	2.405
Sex+Age+BC	652.827	1.842	0.003	2.512
Sex+Age+DAB2*S6+BS+Date	652.909	1.925	0.003	2.618
Sex+DAB1*S9+Rank	652.978	1.994	0.003	2.709
Sex+Age+DAB2*S6+DAB1*S9+DAB2*S7	653.007	2.023	0.003	2.750

Table S28: Relationship between MHC supertype predictors and amount of leucocytes. The description of the header is found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Date+Rank+Age+Sex	217.771	0.000	0.032	
Date+Rank+Age+Sex+DAB2*S6	217.816	0.044	0.032	1.022
Date+Rank+Age+Sex+DAB1*S1	219.064	1.293	0.017	1.909
Date+Rank+Age+Sex+DAB2*S6+DAB1*S1	219.186	1.415	0.016	2.029
Date+Rank+Age+Sex+BC	219.606	1.835	0.013	2.503
Date+Rank+Age+Sex+DAB2*S6+DAB1*S9	219.613	1.842	0.013	2.512
Date+Rank+Age+Sex+DAB2*S6+BS	219.665	1.894	0.013	2.578
Date+Rank+Age+Sex+BS	219.800	2.029	0.012	2.758

Table S29: Relationship between MHC supertype predictors and immunoglobulins proteins (Ig). The description of the header is found in the caption of the table S1.

Models	AICc	ΔAICc	wAICc	ER
Date+DAB1*S9+DAB2*S7	-736.374	0.000	0.017	
Date+DAB1*S9+DAB2*S7+DAB1*S13	-736.152	0.222	0.015	1.117
Date+DAB1*S9+DAB2*S7+Rank	-736.003	0.371	0.014	1.204
Date+DAB1*S9+DAB2*S7+DAB1*S13+Age	-735.744	0.630	0.012	1.370
Date+DAB1*S9+DAB2*S7+Age	-735.706	0.668	0.012	1.397
Date+DAB1*S9+DAB2*S7+DAB1*S13+Rank	-735.614	0.760	0.012	1.462
Date+DAB1*S9+DAB2*S7+BC	-735.419	0.955	0.011	1.612
Date+DAB1*S9+DAB2*S7+BC+Rank	-734.995	1.379	0.009	1.992
Date+DAB1*S9+DAB2*S7+DAB1*S1	-734.740	1.634	0.008	2.264
Date+DAB1*S9+DAB2*S7+BC+Age	-734.647	1.726	0.007	2.371
Date+DAB1*S9+DAB2*S7+DAB1*S13+BC	-734.593	1.781	0.007	2.436
Date+DAB1*S9+DAB2*S7+BS	-734.529	1.845	0.007	2.516
Date+DAB1*S9+DAB2*S7+Sex	-734.155	2.219	0.006	3.033

Table S30: Power analysis results for specific MHC supertype associated with haematocrit. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
DAB1*S1.1	0.013	0.081	0.076	0.687	0.998	1	1
DAB1*S1.2	0.035	0.100	0.126	0.512	0.982	1	1
DAB1*S9.1	0.060	0.069	0.437	0.81	1	1	1
DAB1*S9.2	-0.051	0.097	0.181	0.52	0.985	1	1
DAB1*S13.1	-0.035	0.072	0.161	0.812	1	1	1
DAB1*S13.2	-0.006	0.076	0.07	0.776	1	1	1
DAB2*S3.1	0.067	0.082	0.343	0.673	1	1	1
DAB2*S3.2	0.032	0.107	0.074	0.474	0.961	1	1
DAB2*S7.1	0.043	0.069	0.263	0.819	1	1	1
DAB2*S7.2	-0.050	0.092	0.189	0.558	0.991	1	1
DAB2*S10.1	0.018	0.077	0.08	0.732	1	1	1
DAB2*S10.2	-0.036	0.080	0.164	0.706	0.999	1	1
DAB2*S6	-0.048	0.083	0.222	0.661	0.997	1	1
MHC-I*S6.1	-0.064	0.091	0.318	0.597	0.995	1	1
MHC-I*S6.2	-0.040	0.093	0.161	0.563	0.991	1	1
MHC-I*S6.3	0.115	0.115	0.5	0.401	0.93	1	1
MHC-I*S7.1	0.030	0.080	0.11	0.699	0.999	1	1
MHC-I*S7.2	0.046	0.096	0.178	0.563	0.988	1	1
MHC-I*S7.3	0.055	0.103	0.202	0.495	0.963	1	1
MHC-I*S7.4	0.073	0.104	0.309	0.5	0.97	1	1
MHC-I*S3.1	0.016	0.083	0.064	0.637	0.996	1	1
MHC-I*S3.2	0.066	0.097	0.293	0.543	0.982	1	1
MHC-I*S3.3	-0.016	0.087	0.063	0.641	0.999	1	1

Table S31: Power analysis results for specific MHC supertype associated with amount of leucocytes. CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
DAB1*S1.1	-0.035	0.073	0.154	0.742	1	1	1
DAB1*S1.2	-0.081	0.090	0.415	0.573	0.981	1	1
DAB1*S9.1	0.028	0.067	0.144	0.845	1	1	1
DAB1*S9.2	-0.008	0.099	0.051	0.51	0.983	1	1
DAB1*S13.1	0.013	0.068	0.058	0.811	1	1	1
DAB1*S13.2	0.022	0.075	0.081	0.705	1	1	1
DAB2*S3.1	0.045	0.082	0.188	0.667	0.999	1	1
DAB2*S3.2	0.035	0.099	0.098	0.494	0.98	1	1
DAB2*S7.1	0.033	0.068	0.176	0.811	1	1	1
DAB2*S7.2	-0.008	0.094	0.05	0.561	0.989	1	1
DAB2*S10.1	-0.019	0.077	0.095	0.718	0.997	1	1
DAB2*S10.2	-0.002	0.075	0.049	0.741	1	1	1
DAB2*S6	-0.054	0.072	0.314	0.755	1	1	1
MHC-I*S6.1	0.027	0.086	0.104	0.615	0.99	1	1
MHC-I*S6.2	0.002	0.086	0.053	0.599	0.992	1	1
MHC-I*S6.3	0.016	0.109	0.057	0.431	0.947	1	1
MHC-I*S7.1	0.002	0.073	0.045	0.742	1	1	1
MHC-I*S7.2	-0.028	0.092	0.084	0.532	0.988	1	1
MHC-I*S7.3	-0.017	0.090	0.069	0.558	0.988	1	1
MHC-I*S7.4	-0.012	0.099	0.056	0.492	0.966	1	1
MHC-I*S3.1	-0.002	0.072	0.057	0.754	1	1	1
MHC-I*S3.2	-0.039	0.090	0.132	0.576	0.992	1	1
MHC-I*S3.3	-0.111	0.083	0.761	0.661	1	1	1

Table S32: Power analysis results for specific MHC supertype associated with immunoglobulins proteins (Ig). CI: confidence interval; CES: current effect size.

	Power for the observed effect			Power for effect size at			
	Current effect size	95%CI	Power CES	0.1	0.2	0.3	0.5
DAB1*S1.1	-0.005	0.078	0.06	0.685	0.999	1	1
DAB1*S1.2	-0.069	0.097	0.285	0.535	0.979	1	1
DAB1*S9	-0.023	0.076	0.079	0.688	1	1	1
DAB1*S13.1	0.064	0.076	0.362	0.696	0.998	1	1
DAB1*S13.2	0.048	0.085	0.184	0.618	0.998	1	1
DAB2*S3.1	0.007	0.087	0.053	0.625	0.991	1	1
DAB2*S3.2	0.009	0.103	0.064	0.444	0.966	1	1
DAB2*S7	0.048	0.075	0.254	0.741	1	1	1
DAB2*S10.1	-0.028	0.074	0.112	0.754	0.999	1	1
DAB2*S10.2	-0.030	0.079	0.109	0.721	0.999	1	1
DAB2*S6.1	-0.005	0.078	0.07	0.695	0.998	1	1
DAB2*S6.2	-0.031	0.084	0.11	0.645	0.994	1	1
MHC-I*S6.1	-0.025	0.082	0.075	0.646	0.998	1	1
MHC-I*S6.2	-0.028	0.097	0.068	0.502	0.982	1	1
MHC-I*S6.3	0.005	0.102	0.053	0.46	0.965	1	1
MHC-I*S7.1	-0.016	0.080	0.065	0.668	1	1	1
MHC-I*S7.2	0.095	0.087	0.553	0.596	0.993	1	1
MHC-I*S7.3	0.017	0.098	0.051	0.502	0.979	1	1
MHC-I*S3.1	-0.077	0.079	0.458	0.68	0.999	1	1
MHC-I*S3.2	-0.012	0.093	0.064	0.545	0.992	1	1
MHC-I*S3.3	-0.039	0.096	0.132	0.557	0.984	1	1
MHC-I*S3.4	-0.020	0.097	0.066	0.543	0.978	1	1

Chapter 4

Functional diversity of MHC genes in the face of the recent colonisation history of the European barn owl (*Tyto alba*)

Arnaud Gaigher, Sylvain Antoniazza, Felipe Siverio, Reto Burri*, Alexandre Roulin* & Luca Fumagalli*

Abstract

Understanding the interplay of the different evolutionary forces shaping adaptive genetic diversity at a spatial scale are of central interest in evolutionary and conservation biology. Insular populations and populations in post-glacially recolonized areas harbour usually a limited genetic diversity, which can in turn limit the adaptive potential of populations in the face of environmental changes. This can be particularly true at loci directly involved in individual fitness, such as the major histocompatibility complex (MHC) genes, which are essential in pathogen resistance. One possibility to maintain a high MHC diversity despite the effect of demographic events, is to maximise functional divergence over the remaining alleles in the population, a mechanism that currently is still poorly understood. Here, we sequenced MHC-I and MHC-IIB genes in a total of 16 populations across the European geographical range of the barn owl (*Tyto alba*) including several islands. In addition, all these populations were genotyped for 22 neutral microsatellite markers to contrast with MHC diversity and structure. We show that the recolonization history after the last glaciation from the Iberian Peninsula to Eastern Europe leave a strong signature on MHC genes, with a decline of adaptive diversity. Insular populations are also affected by demographic events, however a certain amount of diversity in terms of functional allelic divergence is preserved in the remaining genetic pool. Overall, the pattern of MHC diversity characterized in the barn owl is mainly driven by colonization history, although in some populations, certain aspects of the MHC diversity pattern suggest that selection may also play a role. Our study brings an important step forward in the understanding of evolutionary mechanisms driving the spatial pattern of MHC diversity, and shows that many possible scenarios can act simultaneously to shape this pattern, and consequently making their disclosure even more challenging.

Introduction

Investigating spatial variation in adaptive genetic diversity (i.e. genetic variation that promotes or may promote adaptation) can provide major insights into the evolutionary mechanisms influencing how organisms become adapted to their environment. This is so, because different regimes of selection are expected to produce distinct patterns of inter- and intra-population genetic diversity. Directional selection for adaptation to the surrounding environmental conditions (referred as local adaptation) will tend to decrease adaptive genetic variation at a local scale, although, at a global scale, genetic diversity across localities may increase if selective agents differ among such localities, leading to the fixation of alternative adaptive variants at each site. Contrarily, high adaptive genetic diversity within a locality may be indicative of mechanisms of selection for alternative adaptive variants (i.e. balancing selection), such as negative frequency-dependent or overdominant selection. If balancing selective agents are common to different populations, we can expect genetic diversity to be of similar magnitude across different populations. However, if selective pressures change in space but are still of balancing nature, overall adaptive genetic diversity will be maximized as it will be maintained within populations (individuals in a population carry different adaptive variants) and between populations (the adaptive pool differs from one locality to the next).

However, patterns of spatial variation in adaptive genetic diversity can also be modulated by other evolutionary forces such as gene flow and genetic drift, and in turn, by the complex interplay between these forces and natural selection. In addition, one particular challenge is to evaluate the relative contribution of each force, especially because natural selection can produce patterns of genetic variation that mimic the ones produced by demographic effects. For instance, local adaptation may be constrained by gene flow, which will slow down the fixation of adaptive variants and the degree of genetic differentiation between populations (Kawecki & Ebert 2004). Gene flow will have an enhancing effect on local diversity, which can be confounded with the effects of balancing selection. Similarly, genetic drift can constrain the balancing selection mechanisms owing to random loss of adaptive variants, and have a depleting effect on genetic diversity, which can mimic the loss of adaptive genetic diversity, owing to the action of selection. Despite its challenge, integrating the potential action of all different evolutionary forces is necessary to understand the mechanisms that actively shape adaptive genetic diversity, and to use this knowledge to better assess the capacity of species to adapt to environmental challenges.

Genes of the major histocompatibility complex (MHC) play a primordial role in immune recognition of pathogens and represent an ideal system to investigate the relative role of different evolutionary forces, especially because different types and intensities of selection as well as gene flow, and genetic drift are expected to act in concert to shape MHC diversity (Alcaide 2010; Spurgin & Richardson 2010). Classical MHC-I and -II genes encode cell surface proteins that, once bound to pathogen-derived antigens, trigger an immune response (Klein & Sato 2000; Jensen 2007). One of the most intriguing properties of MHC genes is their exceptional levels of functional genetic diversity. They are the most polymorphic genes in vertebrates (Gaudieri *et al.* 2000), with the majority of the diversity concentrated within the peptide-binding region (PBR), i.e. the amino acid residues directly involved in the recognition of pathogen-derived peptides. The maintenance of such uncommon diversity is generally assumed to be the result of selection imposed by pathogens (Apanius *et al.* 1997; Jeffreys *et al.* 2001; Spurgin & Richardson 2010). Pathogen-related selection on the MHC system is usually mediated by three non-exclusive means: (i) the heterozygote advantage, (ii) the rare-allele advantage, and (iii) the diversifying selection at a spatial or temporal scale (Sommer 2005; Piertney & Oliver 2006; Spurgin & Richardson 2010), which have received empirical support in previous studies(Froeschke & Sommer 2005; Meyer-Lucht & Sommer 2005; Bonneaud *et al.* 2006; Ekblom *et al.* 2007; Loiseau *et al.* 2009; Oliver *et al.* 2009; Kyle *et al.* 2014). On the other hand, theoretical and empirical evidence suggest that demographic events resulting in smaller population sizes (colonisation events or bottlenecks) can leave a strong signature at MHC genes in terms of loss of MHC diversity (Babik *et al.* 2009; Miller *et al.* 2010; Radwan *et al.* 2010; Ejsmond & Radwan 2011; Sutton *et al.* 2011; Strand *et al.* 2012; Sutton *et al.* 2015). Finally, the recent study in the greater flamingo (*Phoenicopterus roseus*) proposed that local adaptation at MHC genes may be constrained by high gene flow between populations (Gillingham *et al.* 2017).

Discerning between the role of selection, gene flow, and demography in shaping adaptive variation at the MHC system can be achieved by contrasting the spatial pattern of MHC diversity between and within populations with that of neutral markers, given that a discrepancy between them can reveal the signature of selection (Bernatchez & Landry 2003; Spurgin & Richardson 2010). Given the three different non-mutually exclusive scenarios that can explain the spatial pattern of MHC diversity: (i) demographic history, (ii) directional selection for local adaptation, and (iii) balancing selection, we can expect the following differences at the level of variation in neutral and MHC markers. First, if the MHC is mostly driven by stochastic or demographic factors, the level of population structuration between MHC and neutral markers should be similar. Second, if selection for local adaptation plays a major role, MHC genotypes can be expected to exhibit a more pronounced inter-population differentiation compared to neutral expectations. Third, if balancing selection is the main force acting on the MHC at a given geographic scale, a weaker genetic structure should be observed compared to the neutral population structure, if for instance prevailing pathogens are the same over the whole study area.

The European barn owl (*Tyto alba*) represents an ideal system to test the above-mentioned scenarios. Based on neutral makers and demographic simulations, the phylogeography and colonization history of this species in Europe are well understood (Antoniazza *et al.* 2010; Antoniazza *et al.* 2014; Burri *et al.* 2016). Previous studies found support for a ring-like population structure around the Mediterranean (Burri *et al.* 2016). Briefly, barn owls seem to have colonized Europe from the Middle East following two distinct routes. On the one hand, the barn owl populations spread from the Middle East through North Africa, entered Europe from the Iberian Peninsula, from where they spread all along Western and Central Europe and reached Eastern Europe. On the other hand, barn owls seem to have also colonized Europe from the Middle East, by spreading through Turkey and Greece, where they could be in secondary contact with the populations that colonized Europe through Northern Africa. Populations in Eastern Europe (i.e. at the longest tail of the ring-like colonization) and on islands, exhibit the lowest neutral genetic diversity (Burri *et al.* 2016). Studying populations of the European barn owl provides a unique framework to investigate the relative contribution of natural selection and historical demographic events on the evolution of adaptive genes, such as the MHC.

In the present study, using a total of 16 populations across the European geographical range of the barn owl, including several islands, we aimed to determine (i) how the MHC diversity is distributed among populations, (ii) what evolutionary forces could have shaped the diversity and structure of MHC genes, and (iii) whether selection could have counteracted the loss of diversity in populations with colonisation history. All these populations were sequenced for MHC-I and MHC-IIB genes with high-throughput sequencing technologies, and genotyped for 22 neutral microsatellite markers. The latter were used as a baseline to contrast with MHC diversity and genetic structure, in order to infer the role of selection in the evolution of MHC diversity.

Material and Methods

Sampling and MHC molecular methods

The present study includes a total of 431 barn owls from 16 localities including 10 different sites in mainland Europe, Middle East, three different islands in the Canary Islands, Balearic Islands and Crete (Fig. 1). Tissue samples (feathers, blood or muscle) were obtained mostly by collaborators working in survey or monitoring programmes, recovery centres and museums. These tissues were used to extract genomic DNA using a BioSprint 96 extraction robot with the BioSprint 96 DNA blood kit, or, in some cases, using the DNeasy blood and tissue kit following the manufacturer's instructions (Qiagen, Hilden, Germany).

The adaptive genetic variation was estimated from the exon 3 of MHC class Iα (MHC-I) genes and exon 2 of MHC class IIβ (MHC-IIB) genes. Both exons are known to encode for the polymorphic peptide-binding regions. The detailed protocols of primers development, PCR amplification and high-throughput sequencing are described in Gaigher *et al.* (2016), for MHC-I, and in Chapter 2 and Burri *et al.* (2008), for MHC-IIB. In brief, locus-specific primers were designed to amplify independently two MHC-IIB loci (DAB1 and DAB2 loci), whereas MHC-I primers covered the simultaneous co-amplification of two loci. In order to attribute the sequences to a given sample, MHC primers were modified by adding individual barcodes. In the following order, all samples were amplified, quantified, purified and finally pooled according to an equimolar concentration of purified PCR products. Libraries of MHC-I and MHC-IIB loci were sequenced with the Illumina MiSeq technology and the 454 Titanium pyrosequencing protocol respectively.

Raw data processing and MHC genotyping procedures were performed according to the current methods developed for high throughput sequencing (Galan *et al.* 2010; Oomen *et al.* 2013; Sommer *et al.* 2013; Lighten *et al.* 2014; Stutz & Bolnick 2014; Gaigher *et al.* 2016). Validated with replicated samples and family data, our methodologies have shown to provide a reliable MHC genotyping (all details in Gaigher *et al.* (2016) and in Chapter 2). In total for this study, 384, 364 and 380 samples were unambiguously genotyped for MHC-I, MHC-IIB DAB1 and MHC-IIB DAB2 respectively (Table 1).

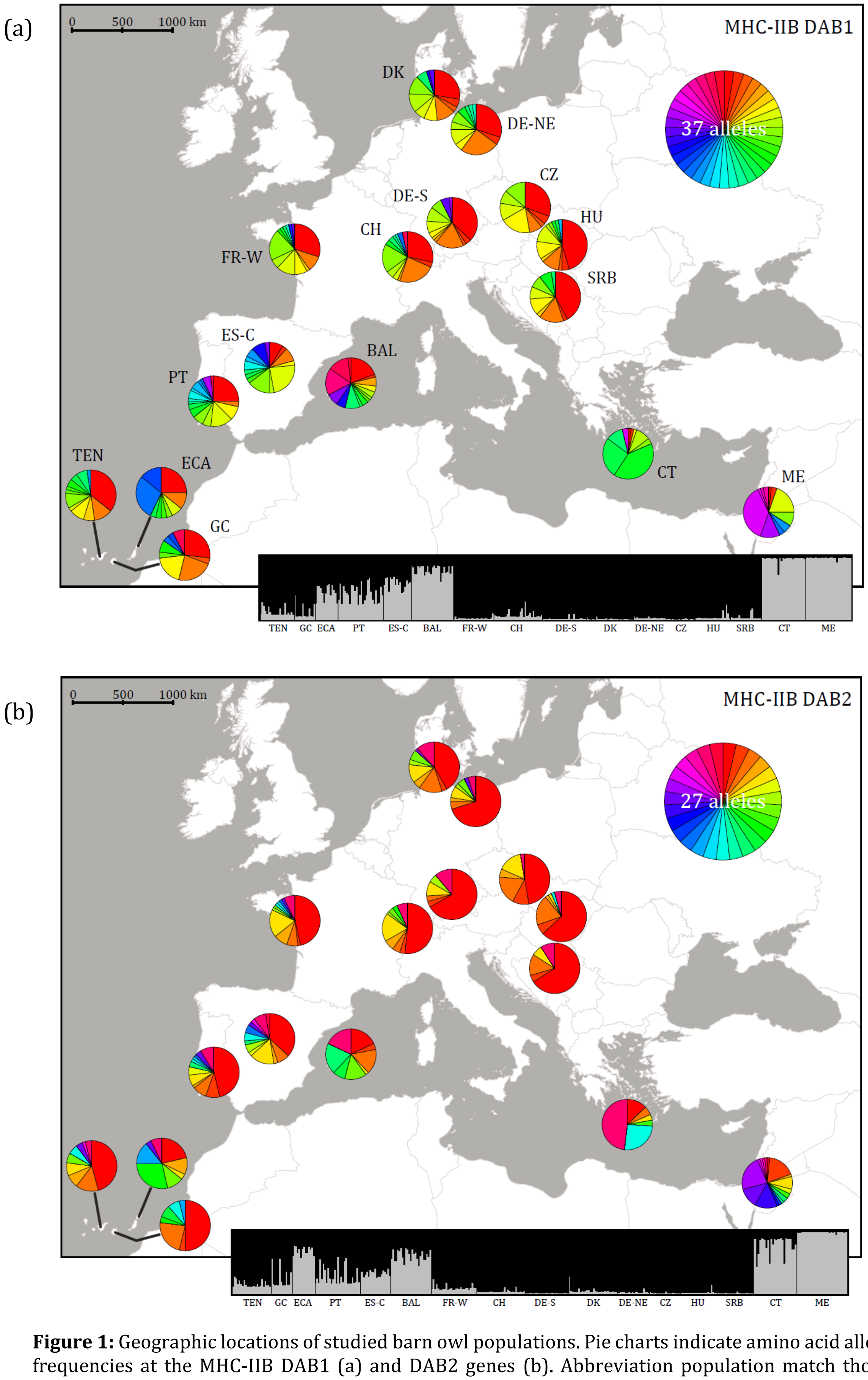


Figure 1: Geographic locations of studied barn owl populations. Pie charts indicate amino acid allele frequencies at the MHC-IIB DAB1 (a) and DAB2 genes (b). Abbreviation population match those given in Table 1. Barplots represent Bayesian analysis performed with STRUCTURE (each line represents one individual and each color an inferred cluster). Due to the high number of amino acid MHC-I alleles (123), the map is illegible, consequently not presented in the results.

MHC diversity analysis

Due to the specific amplification of MHC-IIB DAB1 and DAB2 genes, the zygosity at these genes can be known. Conversely, the co-amplification of the duplicated MHC-I genes made impossible to assigning alleles to the specific loci (i.e., one to four alleles were observed per individual at MHC-I exon 3). As a result, different population genetic approaches were needed for the different markers. In addition, if only the number of different alleles within individuals is taken into account, allele frequencies can be biased by underestimating the frequency of common alleles, and overestimating rare alleles. Consequently, for MHC-I, we determined how many copies of each allele an individual carries, by using the ratios of sequencing coverage observed for each allele. This method has already been shown to provide a robust genotyping in this species (Gaigher *et al.* 2016). For individuals with three different alleles, the allele presenting double the coverage of the other two alleles (2:1:1) was considered to be present in two copies. For 1:1:1 ratios, which can result from biased amplification or copy number variation (CNV) (Gaigher *et al.* 2016), no duplicated alleles were assigned. For individuals with two alleles exhibiting the same coverage ratio (2:2), both were considered to be duplicated. Finally, when only one allele was sequenced in individuals (only two cases), the allele was considered to be duplicated only once, i.e. we did not assume the allele was shared between loci.

Conformity of MHC genes to Hardy-Weinberg equilibrium was assessed with Fstat (Goudet 1995). MHC diversity was estimated with five different measurements: number of alleles (Na), gene diversity (Hs), allelic richness (Ar), nucleotide diversity (π) and functional diversity (Fd). Allelic richness was calculated using Hierfstat (Goudet 2005), with a standardization for 11 individuals (i.e. similar to neutral makers). The functional diversity was estimated by calculating the intra-allelic genetic distances within each individual. The functional distance was calculated by taking in account amino acid physicochemical properties as described in Agbali *et al.* (2010) and in Chapter 2, and was performed on three sequence partitions, namely the entire sequences, PBR codons, and codons inferred to be under positive selection (PSS). PBR codons were defined from Human HLA and Chicken BF for MHC-I (Bjorkman *et al.* 1987; Wallny *et al.* 2006) and from Human HLA for MHC-IIB (Brown *et al.* 1993). PSS were estimated with maximum likelihood site-models using CodeML implemented in PAML 4.7 (Yang 2007).

MHC population structure

MHC genetic differentiation between populations was estimated with pairwise F_{ST} in Arlequin 3.5.2 (Excoffier & Lischer 2010). Data were entered as nucleotide sequences with the number of individuals carrying that sequence within each population. For microsatellites, pairwise F_{ST} were calculated in Fstat (Goudet 1995).

The pattern of isolation by distance was inferred using Mantel and partial Mantel tests. First, we performed simple Mantel tests to test the correlation between geographical distance and population genetic differentiation (F_{ST}), at both MHC and neutral markers. Then, to account for the effect of gene flow and genetic drift, we used a partial Mantel test to assess the association between MHC and geographic distance while controlling for genetic differentiation at neutral makers. These tests were performed using the ring geographic distance, i.e. the shortest overland distance between populations within a clockwise ring starting in the Middle East locality. We used the ring distance because for neutral makers this distance was showed to fit better the colonization history of the European barn owl than the flight and shortest overland distances (Burri *et al.* 2016). Crete and the Middle East were not included in the analyses due to their geographical separation. All Mantel tests were performed with 10’000 permutations using the vegan package in R (Oksanen *et al.* 2017).

The number of genetic clusters was evaluated with the Bayesian clustering procedure implemented in STRUCTURE 2.3.4 (Pritchard *et al.* 2000). Number of assumed clusters (K) was ranged from 1 to 10, with 10 replicates for each K. Each simulation was based on 20^5 burn-in followed by 10^6 iterations using an admixture model with correlated allele frequencies and prior on sampling locations. The most likely number of clusters was inferred with the ΔK method (Evanno *et al.* 2005) implemented in Structure Harvester (Earl & Vonholdt 2012). Membership probability of individuals of the selected K was averaged over the 10 replicates in Clumpp 1.1.2 (Jakobsson & Rosenberg 2007) and clusters were visualised using Distruct 1.1 (Rosenberg 2004).

Neutral markers - molecular methods and analysis

As described above, the pattern of neutral diversity was previously investigated in Europe (Burri *et al.* 2016). However, the sample localities included in our study slightly differ from the previously published data. Consequently, we re-ran all the analyses at neutral markers on our set of individuals to directly contrast them with those conducted on the

MHC loci. The neutral variation was estimated with nuclear and mitochondrial markers. First, 426 individuals were genotyped with a panel of 22 microsatellite markers in GeneMapper 4.0 software (Applied Biosystems) with no departure from Hardy–Weinberg equilibrium and no null alleles. Detailed genotyping procedures are reported in Antoniazza *et al.* (2014) and Burri *et al.* (2016). Diversity at microsatellites was estimated by gene diversity (Hs) and allelic richness standardized for 11 individuals (Ar) using Hierfstat (Goudet 2005). Second, for 239 individuals a fragment of 411 bp of the mitochondrial ND6 gene was sequenced according to the protocol described in Burri *et al.* (2016). Diversity at the mitochondrial marker was obtained with the number of haplotypes and nucleotide diversity (π) using Arlequin 3.5.2 (Excoffier & Lischer 2010).

Results

Overall diversity at MHC markers

A total of 135 MHC-I alleles, 37 MHC-IIB DAB1 alleles and 28 MHC-IIB DAB2 alleles were identified from 16 localities (Table 1). None of these sequences showed evidence of non-functionality, owing to frameshift mutations or stop codons. Twelve out of the 135 MHC-I alleles were synonymous whereas none of the 37 MHC-IIB DAB1 alleles and only one of the 28 MHC-IIB DAB2 alleles were synonymous.

Spatial variation in MHC diversity: colonisation effects in mainland populations

MHC diversity and population structure followed a similar pattern as the one observed at neutral genetic markers. Considering only mainland localities, all indices of MHC diversity (i.e. allelic richness, gene diversity, nucleotide diversity and functional diversity) yielded similar results, with the highest diversity in the Iberian Peninsula (allelic richness for MHC-I, DAB1, and DAB2, respectively: Portugal, 27.25, 11.34, 8.28; Spain, 27.24, 11.83, 9.83) (Table 1, Fig. 2). MHC diversity decreased towards Central, North and Eastern Europe being the lowest in the Eastern localities (allelic richness in the Hungary and the Balkans are 21.29, 8.18, 5.03, and 19.29, 7.51, 4.58, respectively) (Table 1, Fig. 1). The Middle East locality, i.e. the supposed origin population of the ring colonization, showed a disparate pattern of diversity according to MHC loci. Whereas MHC-I and MHC-IIB DAB1 genes showed one of the lowest diversities, the MHC-IIB DAB2 harboured the highest diversity together with the Spanish locality (allelic richness in Middle East is 19.55, 7.77, 9.81) (Table 1, Fig. 1). In addition, regardless of the MHC loci, the Middle East locality

Table 1: Summary of population genetic diversity

	Microsatellites				ND6			MHC-I						MHC-IIB DAB1							MHC-IIB DAB2						
Populations (Abbrev.)	N	Pa	Hs	Ar	N	Na	π	N	Na	Pa	Ar	π	Fd	N	Na	Hs	Pa	Ar	π	Fd	N	Na	Hs	Pa	Ar	π	Fd
Tenerife (Ten)	26	0	0.66	4.8	14	3	0.004	24	27	3	18.9	0.03	0.396	21	12	0.84	0	9.35	0.06	0.573	24	9	0.76	1	7.4	0.04	0.367
Gran Canaria (GC)	16	3	0.7	5.1	15	5	0.006	16	24	1	19.2	0.03	0.410	13	9	0.85	0	8.35	0.06	0.642	13	7	0.71	0	6.5	0.04	0.267
Eastern Canaries (*T. a. gracilirostris*) (ECA)	17	2	0.66	5.1	14	4	0.006	15	20	0	17.5	0.03	0.396	14	9	0.84	0	8.1	0.06	0.590	14	8	0.85	0	7.5	0.04	0.434
Portugal (PT)	30	10	0.73	6.3	15	6	0.008	30	51	10	27.3	0.03	0.395	28	17	0.9	1	11.3	0.07	0.656	28	12	0.76	0	8.3	0.04	0.298
Spain Center (ES-C)	20	3	0.74	6.2	15	5	0.007	19	40	4	27.2	0.03	0.409	17	15	0.91	0	11.8	0.07	0.707	19	13	0.83	1	9.8	0.04	0.434
Baleares (BAL)	29	0	0.68	5.3	14	4	0.006	27	25	3	17.3	0.03	0.411	26	15	0.9	2	10.6	0.06	0.568	25	8	0.86	0	7	0.04	0.325
France West (FR-W)	28	1	0.7	5.5	15	5	0.007	27	45	6	26.5	0.03	0.411	25	13	0.85	1	8.75	0.05	0.531	28	12	0.75	0	7.4	0.04	0.285
Switzerland (CH)	30	0	0.7	5.4	16	5	0.006	30	36	5	23.1	0.03	0.379	30	13	0.84	0	8.47	0.06	0.508	30	9	0.7	0	6.6	0.03	0.263
Germany South (DE-S)	32	0	0.7	5.4	15	3	0.007	29	34	4	21.5	0.03	0.396	27	12	0.82	0	8.44	0.06	0.588	27	6	0.54	0	5	0.03	0.252
Denmark (DK)	33	0	0.69	5.3	15	7	0.007	30	31	2	21	0.03	0.382	29	11	0.87	1	8.94	0.06	0.546	30	9	0.78	0	7.1	0.04	0.336
Germany Northeast (DE-NE)	21	3	0.71	5.3	15	5	0.006	19	30	1	22.5	0.03	0.437	20	11	0.84	0	8.75	0.06	0.526	20	8	0.51	0	6	0.03	0.263
Czech Republic (CZ)	20	0	0.7	5.4	14	3	0.002	19	25	1	21.6	0.04	0.385	18	8	0.84	0	7.34	0.05	0.465	19	6	0.72	0	5.4	0.03	0.289
Hungary (HU)	30	3	0.69	5.1	15	4	0.007	20	27	1	21.3	0.03	0.355	22	12	0.77	0	8.18	0.05	0.399	23	7	0.57	0	5	0.03	0.218
Balkans (SRB)	28	2	0.69	5.1	15	3	0.007	22	27	3	19.3	0.03	0.392	19	9	0.79	0	7.51	0.05	0.450	22	5	0.54	0	4.6	0.03	0.238
Crete (CT)	34	2	0.7	5.3	17	2	0.001	27	19	6	14.9	0.03	0.407	27	9	0.76	0	6.43	0.07	0.658	27	6	0.69	0	5.1	0.03	0.253
Middle East (ME)	32	10	0.72	5.9	15	4	0.004	30	31	13	19.6	0.03	0.431	28	11	0.8	3	7.77	0.06	0.610	31	17	0.88	6	9.8	0.05	0.518
Total	426	-	-	-	239	12	-	384	135	-	-	-	-	364	37	-	-	-	-	-	380	28	-	-	-	-	-

N: number of samples, Na: number of alleles, Pa: private allele, Hs: Gene diversity, Ar: allelic richness based on 11 samples, π: nucleotide diversity, Fd: functional diversity (entire sequence).

exhibited a high functional diversity, i.e. the individuals carried highly divergent alleles (Fig. 2).

Regarding allelic composition, one allele of each MHC loci predominated in all mainland European localities. The alleles Tyal-UA*01, Tyal-DAB1*01, and Tyal-DAB2*01 occurred at more than 8%, 30% and 40% in all localities in recolonized areas (i.e. France to Balkans), as well as in the Iberian Peninsula but at lower frequencies (red alleles in Fig. 1). A strong difference regarding the allelic frequency and composition exists between mainland sites in Europe and the Middle East locality, where the largest number of private alleles was observed for each MHC marker (Table 1). Additionally, several of these privates were present at high frequencies in Middle East (Fig. 1) and some of them harbouring amino acid residues that are unique to the Middle East locality.

Although we found a relatively low neutral genetic structure in mainland Europe (pairwise F_{ST} ranged from 0 to 0.067), a significant isolation by distance was detected ($r = 0.64$, $p = 0.001$) (Fig. 3). Similarly, for all MHC loci, the genetic differentiation was overall low and non-significant among all mainland localities (Fig. 3; Fig. S1). Compared to microsatellites, a weaker isolation by distance was observed at MHC genes (Mantel tests for MHC-I, DAB1, and DAB2 respectively: $r = 0.25$, $p = 0.07$; $r = 0.36$, $p = 0.03$; $r = 0.32$, $p = 0.05$) (Fig. 3). These relationships were no longer significant after controlling for differentiation at neutral markers (partial Mantel tests for MHC-I, DAB1, and DAB2 respectively: $r = 0.27$, $p = 0.08$; $r = 0.08$, $p = 0.34$; $r = 0.18$, $p = 0.17$). Bayesian analyses of MHC population structure identified two main clusters (Fig. 1), in accordance with microsatellites (see Fig. 2 in Appendix). The first cluster included all localities from North and Central Europe (i.e., within the post-glacially recolonized area), with a high assignation probability to this cluster. The second cluster included Middle East locality also with a high assignation probability to this cluster, whereas the Iberian Peninsula localities showed an admixture pattern.

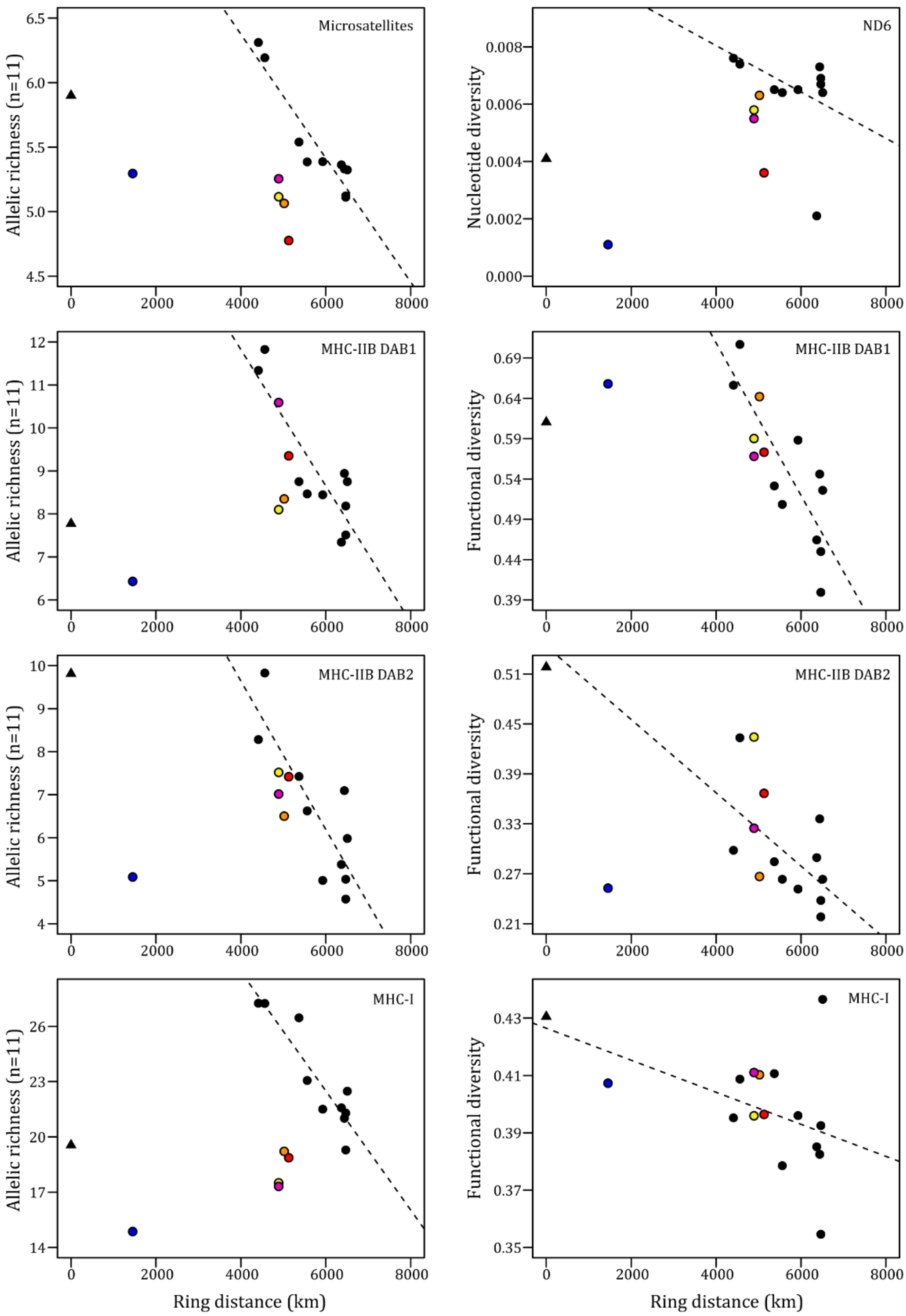


Figure 2: Geographic structure of neutral and adaptive diversity. The first line of graphs describes the neutral diversity in terms of allelic richness and nucleotide diversity for microsatellites (right) and mitochondrial DNA (left), respectively. The second, third and fourth lines of graphs illustrate the adaptive diversity for MHC-IIB DAB1, MHC-IIB DAB2 and MHC-I, respectively. The first and second columns show the allelic richness and the functional diversity on the entire sequence, respectively. Geographic distance is measured according the ring distance as described in the main text. Mainland populations are depicted by black circles, and the Middle East is represented by a triangle. Red: TEN, orange: GC, yellow: ECA, purple: BAL, and blue: CT. Regression lines are exclusively based on mainland populations.

Spatial variation in MHC diversity: island effects

A reduced genetic diversity is expected in islands, and in agreement with the neutral variation, all insular localities showed the lowest MHC-I diversity compared to continental localities (Fig. 2), although, a slightly different pattern was observed for MHC-IIB for some insular localities.

The Crete locality exhibited the smallest diversity for all MHC markers in term of allelic richness when compared to Middle East and all European localities (allelic richness in Crete for MHC-I, DAB1 and DAB2 is 14.87, 6.43 and 5.09, respectively) (Table 1, Fig. 2), in accordance with the neutral diversity (allelic richness for MS, and π for ND6 are respectively: 5.30, 0.001) (Table 1, Fig. S2). It also showed a divergent allelic composition compared to the rest of the localities (Fig. 1). The most common MHC alleles in Crete were found in other localities but at low frequency (less than 10%) and even absent in the closest localities (Balkans and Middle East) (Fig. 1). In this sense, Crete is the locality with the highest differentiation compared to the others, and although closer to each other, Crete and the Middle East showed the highest genetic differentiation at MHC genes (Fig. 3, Fig. S1). This corroborates with the STRUCTURE analysis, in which Middle East and Crete group together, although, by increasing the number of assumed clusters at K=3, Crete is split from the other localities (Fig. 1). Intriguing, whereas the Crete locality might have experienced strong demographic events resulting in low diversity at both neutral and MHC markers (Fig. 2), this locality possesses MHC alleles that combine into one of the highest functional divergence compared to other localities (Fig. 2). The two most common DAB1 alleles in Crete (DAB1*08 and *09), which together represent a third of all the genotypes found in Crete, combine a high genetic divergence (Fig. 2). This high functional diversity is overall characterized for all insular localities, and as illustrated in the Figure 2, these localities move in the graph from below, up to the regression line.

The Balearic locality showed a high level of MHC-IIB DAB1 diversity, similar to one of the closest mainland Iberian populations (allelic richness for DAB1 is 10.59) (Fig. 2), which contrasts with the neutral diversity. This locality also showed private alleles with one being present at high frequency and exhibiting a chimeric conformation (i.e. allele resulting from the mixing of two parental alleles; DAB1 allele: 17.3%). Concerning the Canary localities (including only Tenerife and Gran Canaria), they showed an intriguing allelic frequency distribution and composition, which was similar to the one observed for post-recolonized area localities in central Europe (Fig. 1). These results were supported

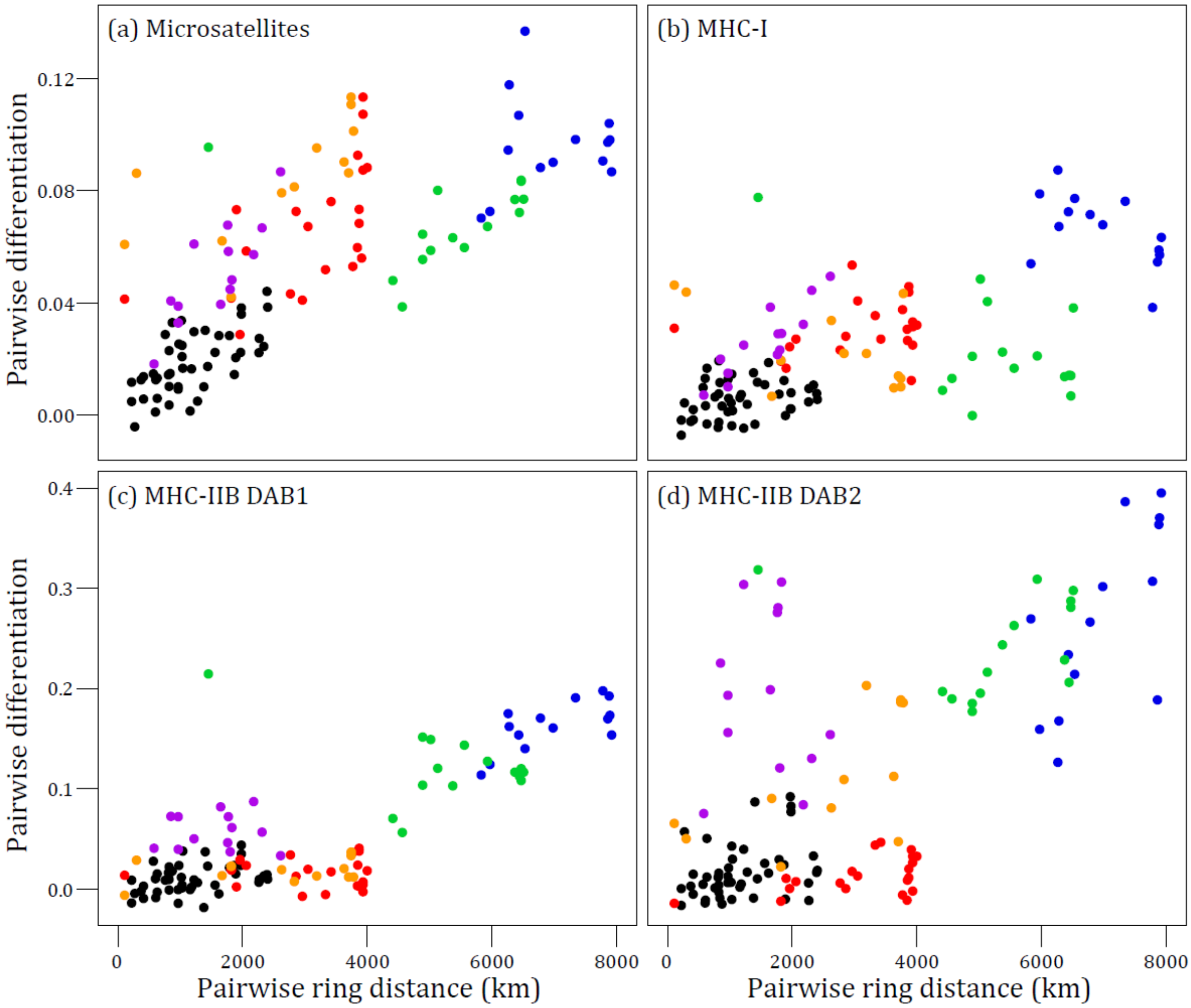


Figure 3: Neutral and adaptive differentiation between populations against geographic distance. Geographic distance is measured according the ring distance as described in the main text. Black circles: pairwise comparison between mainland populations; red circles: GC and TEN against mainland Europe; orange circles: ECA against mainland, GC and TEN populations; purple circles: BAL against mainland and canary populations; blue circles: CT against mainland and others island populations; green circles: ME against all others populations.

by the low pairwise genetic differentiation between the Canary localities and mainland Eastern localities (Fig. 3, Fig. S1). In line with this finding, whereas microsatellites showed strong isolation by distance when including the Canary localities ($r = 0.85$, $p < 0.001$), this relationship disappeared for MHC-IIB genes (DAB1 and DAB2 respectively: $r = 0.26$, $p = 0.05$; $r = 0.17$, $p = 0.12$) (Fig.3). Additionally, the Canary Islands are mostly assigned to the post-glacially recolonized area in the Bayesian structure analysis (Fig. 1), which contrasts with neutral markers.

Discussion

In the present study, we found, in parallel to neutral diversity, a cline of MHC-I and MHC-IIB diversity across Europe. The highest MHC diversity was found in South Western Europe (Iberian Peninsula), and decreased toward mainland Eastern Europe, suggesting that the recolonization history after the last glacial period left a strong signature on MHC diversity. Insular localities showed a low neutral diversity and in general low MHC diversity. Even if some differences exist between MHC loci and insular localities, our findings suggest that, in the same way than in the continent, insular MHC diversity is affected by demography. Intriguingly, the MHC functional diversity was not decreased in some insular populations, and despite the effect of colonisation, these localities have preserved MHC alleles combining high genetic divergence. We discuss the potential explanations involved in the spatial pattern of MHC variation in the barn owl across Europe.

European colonization history after the last glacial period has drastically shaped the distribution of intraspecific genetic diversity (Taberlet *et al.* 1998; Hewitt 2000), and it has been previously invoked to explain the reduction of diversity at MHC genes in post-recolonised areas (Babik *et al.* 2009). The Iberian Peninsula has been a refugium for several European species during the glacial periods and is the starting point in the expansion of these species across Europe (Taberlet *et al.* 1998). Previous studies supported that the barn owl followed this route after the last glaciation (Antoniazza *et al.* 2010; Antoniazza *et al.* 2014; Burri *et al.* 2016) and our results point out that the cline of MHC allelic diversity across mainland Europe is similar to the one observed at neutral markers. This suggests a major effect of the colonization event in shaping MHC diversity. This finding highlights the importance of taking into account the phylogeography and colonization history of a species, in order to understand the pattern of MHC diversity at a large spatial scale, given that differences in allele frequency distribution over the geographical range, may be perceived as the result of local adaptation, especially in the case of genes known to have a strong adaptive function, such as the MHC.

In insular systems, a reduction of genetic diversity is expected due to founder effects followed by genetic drift, which can in turn limit the capacity of populations to adapt to fast-evolving pathogens via MHC genes (Sommer 2005; Radwan *et al.* 2010). The recovery of MHC diversity of depauperate populations can arise by migration, which may be limited in insular populations. However, some studies proposed that balancing selection can counteract genetic drift and maintain high MHC diversity in bottlenecked populations (Aguilar *et al.* 2004; but see Hedrick 2004; van Oosterhout *et al.* 2006; Mona *et al.* 2008), or even by molecular mechanisms such as gene conversion (Spurgin *et al.* 2011). Alternatively, low MHC diversity remaining in bottlenecked populations may preserve alleles covering a high functional divergence and potentially recognize a broader range of pathogens (Radwan *et al.* 2010). This alternative scenario was observed for instance in the Arabian Oryx (*Oryx leucoryx*), which has only three alleles, but with a very large amino acid divergence (Hedrick *et al.* 2000). In the present study, we showed overall that founder effects and genetic drift might have negatively impacted the diversity at MHC genes in insular localities. This is supported by the constant low number of MHC-I alleles in all insular localities. However, we also observed that, despite this, a certain amount of diversity in terms of number of alleles or functional allelic divergence has been maintained in certain localities. For instance, in the Balearic Island, the high number of MHC-IIB DAB1 alleles, coupled with the fact that some private alleles exhibited a chimeric conformation (i.e. alleles with sequence portions originating from two parental alleles), may suggest that gene conversion has shaped MHC diversity toward high diversity, an observation in line with the study of Spurgin *et al.* (2011). Concordant with the scenario of Hedrick *et al.* (2000), and despite the lowest diversity observed in Crete for all types of markers, the most common alleles are the ones combining the highest functional diversity, a pattern particularly marked at the MHC-IIB DAB1 loci. In the margin of our sampling, the Middle East locality harboured a very specific adaptive genetic pool compared to the rest of Europe, even with its closest locality (i.e. the Crete). This locality has high functional diversity, but mainly located at MHC-IIB DAB2 loci.

All these intriguing patterns observed in insular localities or Middle East, with the maintenance of specific and high functional diversity resulting in a contrasted MHC diversity among localities, may have been promoted by local adaptation to specific pathogen communities, in agreement with the hypothesis of Hill (1991). In this sense, previous studies have invoked local adaptation to different ecological regions or directly to specific pathogens (Bonneaud *et al.* 2006; Ekblom *et al.* 2007; Loiseau *et al.* 2009; Eizaguirre *et al.* 2012). While the barn owl covers a large range distribution, little is known about its pathogen communities, which are often assumed to be the drivers of MHC diversity patterns (Sommer 2005; Radwan *et al.* 2010; Spurgin & Richardson 2010). However, pathogen communities are expected to vary across the barn owl distribution, given that this species is present in several climatic regions of Europe, including oceanic, humid continental, Mediterranean or semi-arid climates (Peel *et al.* 2007). Due to the different environmental conditions in which barn owl inhabits, this species may encounter different pathogens, which may be involved in the contrasting MHC diversity observed between localities. Interestingly, the diversity was not distributed equally among MHC loci within locality (in Crete, high functional diversity at DAB1 but not at DAB2) but also between localities (compared to Crete, in Middle East DAB2 maximise the functional diversity). We can speculate that due to the tight linkage between MHC-IIB loci (Chapter 2), if a certain allele at DAB1 loci or combination of alleles within DAB1 genotype confers an advantage against pathogens, the diversity at DAB2 may ultimately "just" follow the selective regime dominating at DAB1. This may explain the dissimilar MHC diversity between DAB1 and DAB2 in Crete.

However, all these patterns that we highlighted in the present study may also arise from neutral processes. For instance by chance, genetic drift in insular localities may have increased the frequency of alleles with high diversity. The high differentiation of the Middle East locality with other populations may be explained by the fact that this locality is probably isolated since a long time with limited gene flow, therefore a differentiation related to its ancient lineage (even considered a subspecies, *Tyto alba erlangeri*; Del Hoyo *et al.* 2014)). Consequently, a detailed characterization of pathogen communities is still necessary to properly understand the spatial variability at MHC genes. Indeed, although we revealed contrasting patterns between localities, comparing the pathogen prevalence and intensity between individuals within localities and between localities will help to disentangle the relative contribution of specific selective regimes and neutral processes (gene flow and genetic drift).

To conclude, the pattern of MHC diversity characterized in the barn owl is largely influenced by its colonization history, even if some intriguing aspects of MHC diversity in certain localities may allow us to think that selection may have also contributed to shape MHC diversity. Our study takes an important step forward in the understanding of the evolutionary mechanisms driving the spatial pattern of MHC diversity, and suggests that many possible scenarios can act simultaneously to shape this pattern, making their disclosure even more challenging. Coupling the sampling of many localities with pathogen information, should bring important insights in this context. Finally, our study has important implications in conservation management projects, given that the lack of homogeneity in the distribution of adaptive MHC diversity over the whole distribution in Europe may indicate that if re-introductions are conducted, they may break-up potential locally co-adapted genes.

References

Agbali M, Reichard M, Bryjová A, Bryja J, Smith C (2010) Mate choice for nonadditive genetic benefits correlate with MHC dissimilarity in the rose bitterling (*Rhodeus ocellatus*). *Evolution*, **64**, 1683-1696.

Aguilar A, Roemer G, Debenham S*, et al.* (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 3490-3494.

Alcaide M (2010) On the relative roles of selection and genetic drift in shaping MHC variation. *Molecular Ecology*, **19**, 3842-3844.

Antoniazza S, Burri R, Fumagalli L, Goudet J, Roulin A (2010) Local adaptation maintains clinal variation in melanin-based coloration of european barn owls (*Tyto alba*). *Evolution*, **64**, 1944-1954.

Antoniazza S, Kanitz R, Neuenschwander S*, et al.* (2014) Natural selection in a postglacial range expansion: the case of the colour cline in the European barn owl. *Molecular Ecology*, **23**, 5508-5523.

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, **17**, 179-224.

Babik W, Pabijan M, Arntzen JW*, et al.* (2009) Long-term survival of a urodele amphibian despite depleted major histocompatibility complex variation. *Molecular Ecology*, **18**, 769-781.

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

Bjorkman PJ, Saper MA, Samraoui B*, et al.* (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, **329**, 512-518.

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

Brown JH, Jardetzky TS, Gorga JC*, et al.* (1993) Three-dimensional structure of the human class-II histocompatibility antigen HLA-DR1. *Nature*, **364**, 33-39.

Burri R, Antoniazza S, Gaigher A*, et al.* (2016) The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean. *Evolution*, **70**, 140-153.

Burri R, Hirzel HN, Salamin N, Roulin A, Fumagalli L (2008) Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. *Molecular Biology and Evolution*, **25**, 1180-1191.

de Eyto E, McGinnity P, Consuegra S*, et al.* (2007) Natural selection acts on Atlantic salmon major histocompatibility (MH) variability in the wild. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 861-869.

Del Hoyo J, Collar NJ, Christie DA, Elliott A, Fishpool LDC (2014) *HBW and BirdLife International Illustrated Checklist of the Birds of the World* Lynx Edicions in association with BirdLife International, Barcelona, Spain and Cambridge, UK.

Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359-361.

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

Ejsmond MJ, Radwan J (2011) MHC diversity in bottlenecked populations: a simulation model. *Conservation Genetics*, **12**, 129-137.

Ekblom R, Saether SA, Jacobsson P*, et al.* (2007) Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology*, **16**, 1439-1451.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.

Froeschke Gt, Sommer S (2005) MHC class II DRB variability and parasite load in the striped mouse (*Rhabdomys pumilio*) in the southern Kalahari. *Molecular Biology and Evolution*, **22**, 1254-1259.

Gaigher A, Burri R, Gharib WH*, et al.* (2016) Family-assisted inference of the genetic architecture of major histocompatibility complex variation. *Molecular Ecology Resources*, **16**, 1353-1364.

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

Gaudieri S, Dawkins RL, Habara K, Kulski JK, Gojobori T (2000) SNP profile within the Human Major Histocompatibility Complex reveals an extreme and interrupted level of nucleotide diversity. *Genome Research*, **10**, 1579-1586.

Gillingham MAF, Béchet A, Courtiol A*, et al.* (2017) Very high MHC Class IIB diversity without spatial differentiation in the mediterranean population of greater Flamingos. *BMC Evolutionary Biology*, **17**, 56.

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.

Goudet J (2005) hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184-186.

Hedrick P (2004) Evolutionary genomics: Foxy MHC selection story. *Heredity*, **93**, 237-238.

Hedrick PW, Parker KM, Gutiérrez-Espeleta GA, Rattink A, Lievers K (2000) Major histocompatibility complex variation in the Arabian Oryx. *Evolution*, **54**, 2145-2151.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Hill AVS (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. 403-420. Springer Berlin Heidelberg, Berlin, Heidelberg.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801-1806.

Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics*, **29**, 217-222.

Jensen PE (2007) Recent advances in antigen processing and presentation. *Nature Immunology*, **8**, 1041-1048.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225-1241.

Klein J, Sato A (2000) The HLA System. *New England Journal of Medicine*, **343**, 702-709.

Kyle CJ, Rico Y, Castillo S*, et al.* (2014) Spatial patterns of neutral and functional genetic variations reveal patterns of local adaptation in raccoon (*Procyon lotor*) populations exposed to raccoon rabies. *Molecular Ecology*, **23**, 2287-2298.

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

Loiseau C, Richard M, Garnier S*, et al.* (2009) Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Molecular Ecology*, **18**, 1331-1340.

Meyer-Lucht Y, Sommer S (2005) MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). *Molecular Ecology*, **14**, 2233-2243.

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

Mona S, Crestanello B, Bankhead-Dronnet S*, et al.* (2008) Disentangling the effects of recombination, selection, and demography on the genetic variation at a major histocompatibility complex class II gene in the alpine chamois. *Molecular Ecology*, **17**, 4053-4067.

Oksanen J, Blanchet FG, Friendly M*, et al.* (2017) vegan: Community Ecology Package. R package version 2.4-3. https://CRAN.R-project.org/package=vegan.

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

Oomen RA, Gillett RM, Kyle CJ (2013) Comparison of 454 pyrosequencing methods for characterizing the major histocompatibility complex of nonmodel species and the advantages of ultra deep coverage. *Molecular Ecology Resources*, **13**, 103-116.

Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, **11**, 1633-1644.

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

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Radwan J, Biedrzycka A, Babik W (2010) Does reduced MHC diversity decrease viability of vertebrate populations? *Biological Conservation*, **143**, 537-544.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137-138.

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

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

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

Spurgin LG, van Oosterhout C, Illera JC*, et al.* (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, **20**, 5213-5225.

Strand TM, Segelbacher G, Quintela M*, et al.* (2012) Can balancing selection on MHC loci counteract genetic drift in small fragmented populations of black grouse? *Ecology and Evolution*, **2**, 341-353.

Stutz WE, Bolnick DI (2014) Stepwise threshold clustering: a new method for genotyping MHC loci using next-generation sequencing technology. *PLoS One*, **9**, e100587.

Sutton JT, Nakagawa S, Robertson BC, Jamieson IG (2011) Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. *Molecular Ecology*, **20**, 4408-4420.

Sutton JT, Robertson BC, Jamieson IG (2015) MHC variation reflects the bottleneck histories of New Zealand passerines. *Molecular Ecology*, **24**, 362-373.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453-464.

van Oosterhout C, Joyce DA, Cummings SM*, et al.* (2006) Blancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution*, **60**, 2562-2574.

Wallny H-J, Avila D, Hunt LG*, et al.* (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1434-1439.

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

Supplementary material

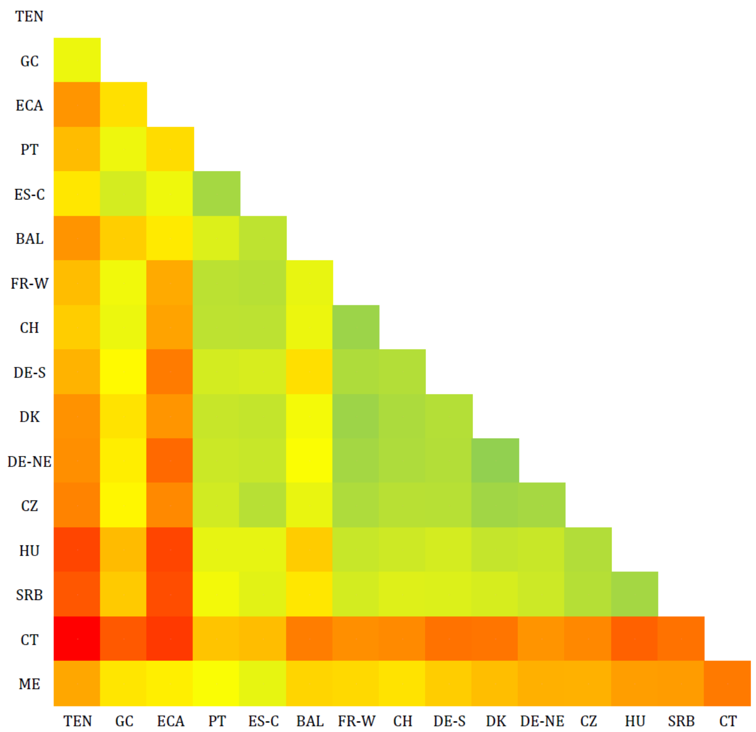
Supplementary materials include the following documents:

Figure S1: Heatmap of pairwise differentiation (F_{ST}) for microsatellites and MHC-IIB genes.

Figure S2: Spatial frequency distribution of mitochondrial ND6 haplotypes.

Figure S1: Heatmap of pairwise differentiation (F_{ST}) for microsatellites (a) and MHC-IIB genes (b). For MHC-IIB genes, below and above the diagonal correspond to pairwise differentiations for DAB1 and DAB2, respectively. Color range from green to red for low to high differentiation, respectively.

(a)



(b)

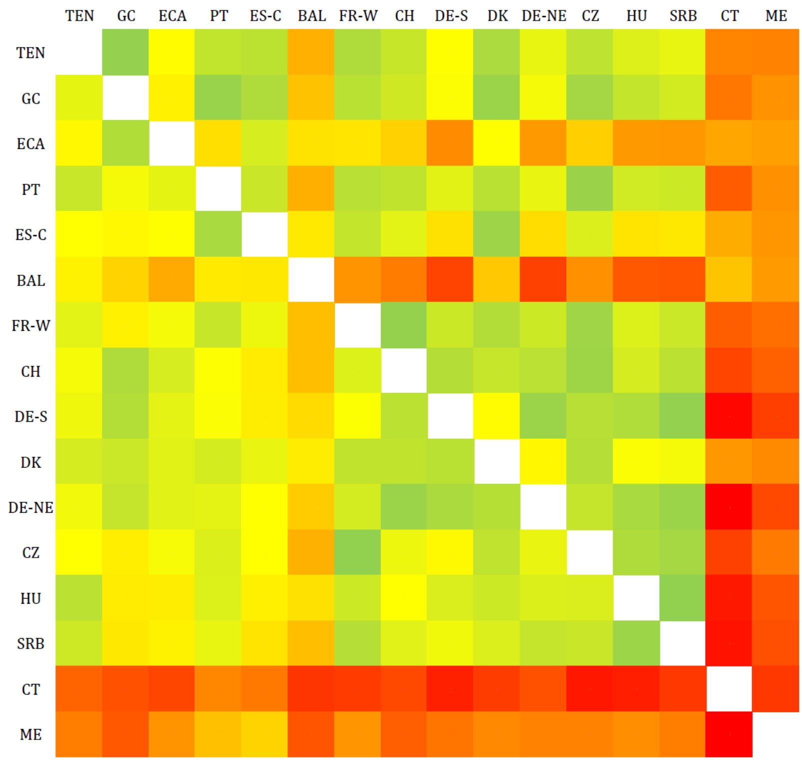
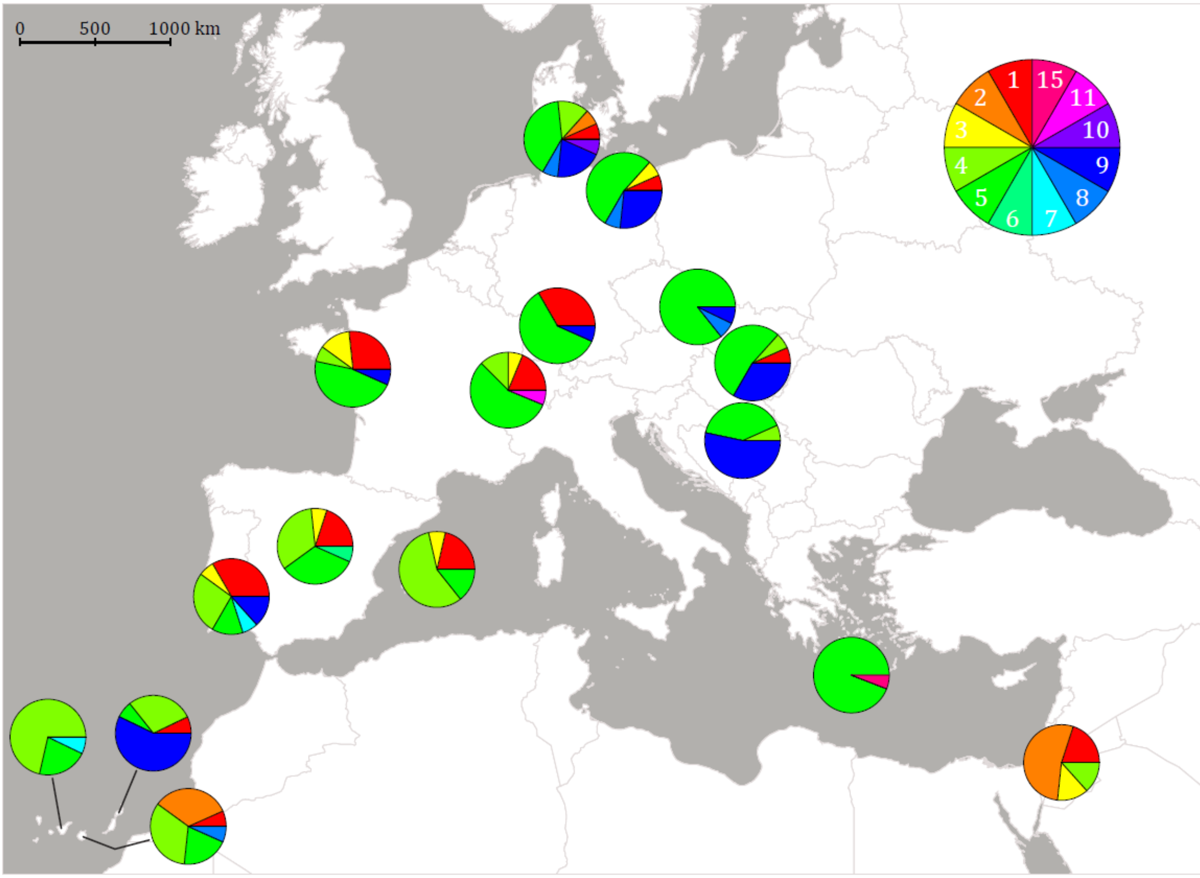


Figure S2: Spatial frequency distribution of mitochondrial ND6 haplotypes. Each colour represent a haplotype



GENERAL DISCUSSION

During my PhD, I investigated several key aspects related to the current research in the MHC field, including genotyping, molecular evolution, the link between MHC diversity and individual fitness-related traits such as immunocompetence, and finally the spatial distribution of MHC diversity. I will first summarize the main findings of each chapter, emphasizing the contribution of my work in the general research field of MHC and pointing out some aspects that I accentuated less in my chapters. In a second section, I will discuss those main aspects that will need to be explored in future studies to cover some of the remaining knowledge gaps in relation to both, the barn owl MHC and the MHC field in general.

Summary of main findings

During all my PhD, I have been particularly attracted to different methodological aspects associated to the MHC system, and a large investment was devoted to accurately genotype the MHC genes. As developed in the general introduction, due to the complex evolutionary dynamic of this system, the genotyping of MHC genes is particularly challenging, and it has been the center of attention since many years now. Indeed, providing an accurate genotyping is fundamental to avoid biases in the subsequent analyses. Including my own work, a large number of publications have been dedicated to improve the MHC genotyping (see general introduction). In chapters 1 and 2, I addressed this aspect by genotyping MHC-I and MHC-IIB genes with two different technologies (Illumina MiSeq vs. 454 pyrosequencing). Both yield considerable different coverage amplicon, with Illumina Miseq providing higher coverage than 454. Even if the 454 can perform as well as MiSeq sequencing when the number of duplicated genes is low, or when genes can be independently amplified, Illumina technology can outperform 454 protocol in individuals with higher diversity in terms of number of duplicated genes. Our data resulting from both technologies showed that family-based genotyping can be of great utility to confirm MHC genotyping, for simple or complex MHC genomic organization. I am aware that the accessibility to pedigree data in wild species can be challenging, however if only a subset of a larger dataset includes family information, this subset can be used to estimate the genotyping accuracy in the global dataset. Overall, sequencing technologies are in constant improvement, and sequencing companies frequently increase coverage and fragment lengths, while reducing sequencing errors. All these developments will inevitably continue to move the MHC field forwards, and make the MHC characterization in non-model species advance, particularly in those with several duplicated MHC genes.

Taking advantage of family data, I covered certain aspects of MHC diversity during this PhD that have been poorly demonstrated in wild species. In my first chapter, I showed that the reconstruction of haplotypes inferred from the pattern of allelic segregation within family data could provide essential information related to the architecture of MHC genes. In the barn owl, I revealed linkage disequilibrium between MHC genes, allele sharing among genes and variation in copy number (CNV). These results can in turn deeply contribute to improve research in the context of MHC evolutionary ecology. For instance, a study in Tasmanian devil populations (*Sarcophilus harrisii*) showed the contribution of CNV in MHC into disease susceptibility (Siddle *et al.* 2010). In the second chapter of this thesis, by using the previously acquired molecular knowledge from my first chapter, I investigated whether duplication may constitute the raw material for selection at an ecological level. Specifically, I tested whether the tight linkage of MHC duplicates enables selection for haplotypes with high divergence (under the assumption that higher MHC diversity grants a better immunity). I found low support for this hypothesis, perhaps because of the specific evolutionary dynamics of the two MHC classes in the barn owl, i.e. extreme homogenisation among MHC-I loci *versus* strong divergence among MHC-IIB loci. The granted inheritance of high diversity within MHC-IIB haplotypes may ensure the transmission of a large amount of individual MHC diversity to the offspring and maintain a large pool of MHC diversity in the population. Such evolutionary dynamic may have major implications in contexts such as for instance MHC-related mate choice, given that the inherently promoted high diversity within haplotypes may reduce the strength of sexual selection for complementary alleles, as recently postulated in the Leach's storm petrel (*Oceanodroma leucorhoa*) (Dearborn *et al.* 2016). My two first chapters highlight the necessity to properly characterize the magnitude of MHC molecular processes including duplication, recombination, gene conversion as well as the extent to which MHC loci are linked, in order to understand how these genomic changes may contribute to the adaptation for optimal MHC-mediated pathogen defence and conversely, how the selective pressure imposed by pathogens can affect the MHC genomic organization. An illustrative example is the co-evolution of TAPs with classical MHC-I genes in the chicken.

The minimalist MHC organization observed in chickens promoting limited recombination between TAPs and MHC-I genes ensures the maintenance of co-adapted alleles, which co-inheritance provides a better disease resistance (Kaufman *et al.* 1999; Wallny *et al.* 2006; Walker *et al.* 2011).

The resulting detailed MHC characterization from chapters 1 and 2 was used subsequently to investigate, in chapter 3, whether specific aspects of MHC diversity confer a higher immunocompetence. Overall, I found very little support in this sense, given that neither high diversity (in terms of number of alleles or the functional divergence between alleles) nor specific alleles predict the immune-related variables. Our findings, together with the current literature of the field, made wonder why there is a general little support for the link between MHC system and immunocompetence. Using a set of power analyses, I found in our study that if an effect exists, it would be expected to be of small-to-moderate effect size, which is in agreement with our estimations of the effect sizes reported in studies that found significant associations (see Bonneaud *et al.* 2005; Charbonnel *et al.* 2010). I hypothesized that low effect sizes are likely to exist in the context of MHC-immunocompetence associations, owing to the multiple genetic factors that influence the immune response. Consequently, coupling several candidate genes (for instance Toll-like receptors) in future studies of this field will probably contribute to better understand the variation in pathogen resistance in wild species. Most importantly, these findings have strong implications in future studies that want to investigate the link between MHC and immunocompetence. Indeed if small to moderate effect sizes are expected, researchers will need to increase their sampling efforts to achieve enough power to be able to detect small-effects associations. This view may ultimately be not restricted only to the link between MHC and immunocompetence, but extended to other traits related to individual fitness, such as survival, growth, reproductive success or the natural pathogens that the hosts are confronted with. I highlight the importance of conducting meta-analysis by compiling multiple studies over various taxa, which will permit to draw a more general conclusion about the relationship between MHC and immunocompetence, and ultimately, fitness. A recent study in this sense, but where the focus was given to MHC-dependent mating preferences, highlighted that in line with our expectations, a small effect size has been generally observed when investigating whether the mate choice is preferentially based on MHC diversity or dissimilar partners (Kamiya *et al.* 2014).

All these previous chapters were focused on a well-monitored Swiss population of barn owls. I took a more global view and investigated in a fourth chapter the pattern of MHC diversity over several localities in Europe. In general, I found that the pattern of MHC diversity has been mainly driven by the colonization history of the species in Europe, even if certain localities showed intriguing patterns that may allow thinking that selection has also contributed to shape MHC diversity. Our empirical study, highlights a major issue associated to studies investigating the spatial pattern of MHC variation, i.e. it is very challenging to distinguishing the relative contribution of evolutionary forces shaping the MHC diversity (see review in Spurgin & Richardson 2010). Indeed, the MHC may be influenced by different pathogen-mediated selective regimes, which may vary across the geographical range of the species but may also act simultaneously, as for instance local adaptation and balancing selection. In addition to that, non-adaptive processes such as genetic drift and gene flow may interact with selection, making the general picture more complex, especially as these processes can result in more similar patterns of diversity than those driven by selection. I expect that integrating many sampled localities over the distribution of the species, coupled with information on pathogen prevalence and intensity at each locality, will help to discern the evolutionary forces in action.

Finally, all together, these chapters revealed the difficulty of first capturing the MHC diversity at the molecular level, and of subsequently understanding the biological meaning of such diversity. In other words, when some pieces of the puzzle are missing, the final picture can be difficult to draw with certainty.

Missing pieces: pathogens, others genes, and other populations?

Since the MHC system is a key component of pathogen resistance, a fully comprehensive understanding of the selection behind MHC diversity, requires identifying the pathogens and inspecting their role as selective agents. In my present work, one important piece of missing information was related to the pathogen communities. I sequenced MHC genes from individuals for which immunocompetence challenges have been previously performed. Such tests provide overall information on the ability of individuals to mount an immune response against foreigner antigens, and on the general heath status of the individuals, and are of great utility because they can be easily applied in the field and in different species and contexts. However, they do not provide an insight into the natural association that can exist between the wild communities of pathogens that the host is facing, and MHC diversity. Surprisingly, in the Swiss population, and generally in the barn owl, little is known about parasites and diseases in terms of both qualitative and quantitative measurements. This species may be exposed to various common owl parasites including gastrointestinal and blood parasite protozoans, arthropods or helminths (Roulin 2002; Ferrer *et al.* 2004; Ishak *et al.* 2008; Santoro *et al.* 2013), and in Switzerland (at least) it is parasited by the common blood-sucking ectoparasite fly (*Carnus haemapterus*) (Roulin 1998). In wild bird populations, a gap still exists in associations between pathogens and MHC. Nevertheless, we can take advantage of the monitoring of the barn owl that has been performed in our group since many decades (Roulin group). For thousands of individuals, blood samples have been collected in our population and they can be used as a first step for the identification of different blood parasites (haematozoa). Using molecular methods (Knowles *et al.* 2011), the most represented haematozoa groups, i.e. *Leucocytozoon*, *Haemoproteus* and *Plasmodium spp* may be screened. The detection of the presence/absence of haematozoa parasites, coupled with posterior MHC sequencing of these individuals, will allow to test if the presence/absence of parasites or parasite diversity is related to MHC diversity in this species. Such relations have been previously shown in passeriforms (Westerdahl *et al.* 2012; Sepil *et al.* 2013) and are important to better understand the pathogen-mediated selection and, more generally, they are of great significance in both ecologically and epidemiologically contexts. Ultimately, these studies can reveal that the picture is even more complex, due to the probable co-infection by multiple pathogens (McClelland *et al.* 2003). From personal observation in Switzerland, the barn owl is infected by several arthropods (ectoparasite fly or ticks) or helminths (see also Roulin 2002 for detailed endo- and ectoparasites), which may also be important to take into account, in order to understand the dynamic of host-pathogen interactions during multiple co-infection.

In this PhD, I covered the most widely and well-known used immune system genes, i.e. the MHC genes. However, the immune system is amazingly complex and is compiled by hundreds of genes, all with specific roles in the immune pathway. Investigating only the MHC genes can explain why few relationships seem to exist between MHC diversity and immunocompetence in the barn owl (see Chapter 2). As developed in the introduction, the candidate gene approaches can be fruitful, but as the name indicates only targets specific genes. Investigating other immune genes can be relevant, and diversity at immune genes other than the MHC have been previously shown to be mediated by pathogen pressures (Alcaide & Edwards 2011; Bollmer *et al.* 2011; Quéméré *et al.* 2015; Bateson *et al.* 2016). However, most of these studies still used candidate gene approaches and capture only a subset of immune gene diversities. Rather, working at the genome scale for the detection of multiple functionally important immune genes may bring major insights in the relative contribution of different immune genes, including the MHC ones, in shaping pathogen resistance. New methodologies, including exon capture and transcriptomics may be interesting as they focus only on regions of the genome that may have adaptive diversity and or tissues specialized in mediating pathogen resistance. Some recent studies have shown, by using these methodologies, the power to detect many regulatory immune genes, which can be used to understand host-pathogen interactions (Lenz *et al.* 2013; Ekblom *et al.* 2014; Fijarczyk *et al.* 2016).

Finally, as developed in the introduction, the barn owl, *Tyto alba* (Tytonidae) is a remarkable species for its worldwide distribution, with many subspecies across the world (even if the actual taxonomic status of the different described subspecies is under review in our group). In addition, many other Tytonidae species closely related to the barn owl, seem to occur in both sympatry and allopatry (Del Hoyo *et al.* 2014). The large distribution of this species and the overlap with other *Tytonidae* species offers a unique system for research in MHC evolutionary biology. An intriguing observation in the MHC system, is the sharing of MHC alleles between species (i.e. often alleles within a species are more divergent than between species) (Klein *et al.* 1998). Such pattern can arise by the trans-species polymorphism (TSP), i.e. the result of long the persistence of allelic lineages beyond speciation events due to pathogen-mediated selection (Klein 1987). Alternatively, convergent evolution can yield a similar pattern, but for a different reason, i.e. independently, species can share a similar MHC polymorphism due to their adaptation to similar environments or exposure to similar pathogens (Yeager & Hughes 1999). Distinguishing these two scenarios is of central interest to understand long-term evolutionary relationships among MHC genes, and has been previously investigated, including in owls from the Strigidae family (Burri *et al.* 2008; Lenz *et al.* 2013; Gillingham *et al.* 2016). The Tytonidae family in this context can bring important insight due to their large distribution with apparently sympatric speciation. During this PhD, the MHC primers were tested and some individuals from the other *Tyto alba* subspecies over several countries outside Europe were sequenced. The MHC primers allowed sequencing these different subspecies, which makes it possible to use them to conduct studies comparing the MHC of different Tytonidae species and subspecies. In our group, a project concerning the world phylogeny of this species and closely related is being conducted in parallel, with a panel of seven neutral markers, and SNP data. I can expect that the confrontation of the MHC phylogeny against the neutral phylogeny can yield important insights into the processes of adaptation and speciation.

Final conclusion

The multilevel approach applied in this PhD, coupled with high throughput sequencing technologies, have permitted to bring new insight into the MHC evolutionary ecology, by using the barn owl as model species. In addition, our empirical study triggers new open issues in the MHC field, where hopefully future studies will take into consideration the several sampling-related factors I determined as important in this PhD: i) the family-based genotyping, ii) the caution in sample sizes, in order to achieve enough power to detect expected small-effect associations between MHC and fitness related traits, and iii) the sampling of pathogen communities in order to fully understand the outcomes of host-pathogen interactions.

References

Alcaide M, Edwards SV (2011) Molecular evolution of the Toll-like receptor multigene family in birds. *Molecular Biology and Evolution*, **28**, 1703-1715.

Bateson ZW, Hammerly SC, Johnson JA*, et al.* (2016) Specific alleles at immune genes, rather than genome-wide heterozygosity, are related to immunity and survival in the critically endangered Attwater's prairie-chicken. *Molecular Ecology*, **25**, 4730-4744.

Bollmer JL, Ruder EA, Johnson JA, Eimes JA, Dunn PO (2011) Drift and selection influence geographic variation at immune loci of prairie-chickens. *Molecular Ecology*, **20**, 4695-4706.

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

Burri R, Hirzel HN, Salamin N, Roulin A, Fumagalli L (2008) Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. *Molecular Biology and Evolution*, **25**, 1180-1191.

Charbonnel N, Bryja J, Galan M*, et al.* (2010) Negative relationships between cellular immune response, Mhc class II heterozygosity and secondary sexual trait in the montane water vole. *Evolutionary Applications*, **3**, 279-290.

Dearborn DC, Gager AB, McArthur AG*, et al.* (2016) Gene duplication and divergence produce divergent MHC genotypes without disassortative mating. *Molecular Ecology*, **25**, 4355-4367.

Del Hoyo J, Collar NJ, Christie DA, Elliott A, Fishpool LDC (2014) *HBW and BirdLife International Illustrated Checklist of the Birds of the World* Lynx Edicions in association with BirdLife International, Barcelona, Spain and Cambridge, UK.

Ekblom R, Wennekes P, Horsburgh GJ, Burke T (2014) Characterization of the house sparrow (Passer domesticus) transcriptome: a resource for molecular ecology and immunogenetics. *Molecular Ecology Resources*, **14**, 636-646.

Ferrer D, Molina R, Castellà J, Kinsella JM (2004) Parasitic helminths in the digestive tract of six species of owls (Strigiformes) in Spain. *The Veterinary Journal*, **167**, 181-185.

Fijarczyk A, Dudek K, Babik W (2016) Selective Landscapes in newt Immune Genes Inferred from Patterns of Nucleotide Variation. *Genome Biology and Evolution*, **8**, 3417-3432.

Gillingham MAF, Courtiol A, Teixeira M*, et al.* (2016) Evidence of gene orthology and trans-species polymorphism, but not of parallel evolution, despite high levels of concerted evolution in the major histocompatibility complex of flamingo species. *Journal of Evolutionary Biology*, **29**, 438-454.

Ishak HD, Dumbacher JP, Anderson NL*, et al.* (2008) Blood parasites in owls with conservation implications for the Spotted Owl (*Strix occidentalis*). *PLoS One*, **3**, e2304.

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

Kaufman J, Milne S, Gobel TWF*, et al.* (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, **401**, 923-925.

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

Klein J, Sato A, Nagl S, O'HUigin C (1998) Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics*, **29**, 1-21.

Knowles SCL, Wood MJ, Alves R*, et al.* (2011) Molecular epidemiology of malaria prevalence and parasitaemia in a wild bird population. *Molecular Ecology*, **20**, 1062-1076.

Lenz TL, Eizaguirre C, Kalbe M, Milinski M (2013) Evaluating patterns of convergent evolution and trans-species polymorphism at mhc immunogenes in two sympatric stickleback species. *Evolution*, **67**, 2400-2412.

McClelland EE, Penn DJ, Potts WK (2003) Major histocompatibility complex heterozygote superiority during coinfection. *Infection and Immunity*, **71**, 2079-2086.

Quéméré E, Galan M, Cosson J-F*, et al.* (2015) Immunogenetic heterogeneity in a widespread ungulate: the European roe deer (*Capreolus capreolus*). *Molecular Ecology*, **24**, 3873-3887.

Roulin A (1998) Cycle de reproduction et abondance du diptère parasite *Canus hemapterus* dans les nichées de chouettes effraies *Tyto alba*. *Alauda*, **66**, 265-272.

Roulin A (2002) The barn owl. In: *Update: Journal of the birds of Western Palearctic* pp. 115–138.

Santoro M, Mattiucci S, Nascetti G*, et al.* (2013) Helminth communities of owls (Strigiformes) indicate strong biological and ecological differences from birds of prey (Accipitriformes and Falconiformes) in southern Italy. *PLoS One*, **7**, e53375.

Sepil I, Lachish S, Hinks AE, Sheldon BC (2013) Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20130134.

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

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

Walker BA, Hunt LG, Sowa AK*, et al.* (2011) The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 8396-8401.

Wallny H-J, Avila D, Hunt LG, *et al.* (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1434-1439.

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

Yeager M, Hughes AL (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunological Reviews*, **167**, 45-58.

Appendix

The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean

R. Burri*, S. Antoniazza*, A. Gaigher, A.L. Ducrest, C. Simon, The European Barn Owl Network, L. Fumagalli, J. Goudet†, & A. Roulin†

* Joint first authors
† Joint senior authors

This Appendix is published in ***Evolution (2016) 70, 140-153.***

ORIGINAL ARTICLE

doi:10.1111/evo.12824

The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean

Reto Burri,[1,2,*] Sylvain Antoniazza,[3,4,*] Arnaud Gaigher,[5] Anne-Lyse Ducrest,[3] Céline Simon,[3] The European Barn Owl Network,[6] Luca Fumagalli,[5] Jérôme Goudet,[3,†] and Alexandre Roulin[3,†]

[1]*Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-75236 Uppsala, Sweden*

[2]*E-mail: burri@wildlight.ch*

[3]*Department of Ecology and Evolution, University of Lausanne, Biophore, CH-1015 Lausanne, Switzerland*

[4]*Swiss Ornithological Institute, Seerose 1, CH-6204 Sempach, Switzerland*

[5]*Laboratory for Conservation Biology, Department of Ecology and Evolution, University of Lausanne, Biophore, CH-1015 Lausanne, Switzerland*

[6]*Members and affiliations of this group author are listed in the Supplementary Information*

Received January 19, 2015
Accepted November 9, 2015

Uncovering the genetic basis of phenotypic variation and the population history under which it established is key to understand the trajectories along which local adaptation evolves. Here, we investigated the genetic basis and evolutionary history of a clinal plumage color polymorphism in European barn owls (*Tyto alba*). Our results suggest that barn owls colonized the Western Palearctic in a ring-like manner around the Mediterranean and meet in secondary contact in Greece. Rufous coloration appears to be linked to a recently evolved nonsynonymous-derived variant of the melanocortin 1 receptor (*MC1R*) gene, which according to quantitative genetic analyses evolved under local adaptation during or following the colonization of Central Europe. Admixture patterns and linkage disequilibrium between the neutral genetic background and color found exclusively within the secondary contact zone suggest limited introgression at secondary contact. These results from a system reminiscent of ring species provide a striking example of how local adaptation can evolve from derived genetic variation.

KEY WORDS: **Barn owl, clines, local adaptation, melanin-based coloration, ring species.**

Phenotypic variation is widespread in natural populations, and understanding how ecology-driven selection contributes to its evolution and maintenance is a central task in evolutionary biology. The demonstration of an involvement of selection in phenotypic evolution represents only a first step to this end, and leaves a wealth of questions concerning the origin, maintenance, and consequences of phenotypic variation. What are the agents of selection? When did alternative phenotypes evolve? May new phenotypes enable the colonization of new ranges? In such cases, can geographic isolation, ecology-driven ecological selection, or their interplay lead to ecological speciation (Nosil 2012)?

To gain a better understanding of these questions, insights into the phenotypes' ecology and the genetics underlying phenotypic variation are required. The ecological function of a phenotypic trait can point toward potential selective agents to be investigated, and may in some cases identify traits correlated to the focal phenotype that may be the targets of selection in its stead (e.g., Stearns 1992). Especially for complex phenotypes that may mirror multiple functions, such as melanin-based coloration (Ducrest et al. 2008), identifying the actual target of selection can pose challenges. Identifying the gene(s) underlying the observed

*These authors are joint first authors.
†These authors are joint senior authors.

phenotypic variation may provide important clues in this context, by providing information about the extent of alternative target traits linked via pleiotropy. Moreover, the population genetic study of these genes can help corroborate the evidence for the adaptive evolution of the phenotype of interest, and may provide insights into the relative ages of alternative phenotypes. The integrative study of species' colonization history, ecology, and quantitative trait locus (QTL) variation therefore offers opportunities to further our understanding of the role of novel genetic and phenotypic variation for the colonization of new ranges, and ultimately species formation.

Barn owls (*Tytonidae*) of the genus *Tyto* represent an ideal system to tackle these questions. These birds exhibit extraordinary examples of phenotypic variation, with pheomelanin-based plumage color clines across continents found in at least seven (sub)species (Roulin et al. 2009). The strongest cline is found in the European barn owl (*Tyto alba*), which varies from white in Iberia to dark rufous in Northeastern Europe (Roulin 2003; Antoniazza et al. 2010). The pronounced phenotypic population structure of this color cline contrasts with a significantly weaker pattern of isolation by distance observed at the neutral genetic level within Europe (Antoniazza et al. 2010). Spatially explicit modeling provides evidence that the species' European range was colonized in a single expansion out of Iberia following the last glaciation, and indicates that the color cline did not evolve as a neutral by-product of colonization (Antoniazza et al. 2014). These two lines of evidence strongly suggests that the barn owl color polymorphism is maintained by selection (Antoniazza et al. 2010, 2014), and correlations of coloration with other phenotypes may reinforce the pattern of color-related local adaptation.

Differences in the ecology correlated with color phenotypes indicate that rufous individuals are better adapted to the continental environment prevailing in Northeastern Europe (Antoniazza et al. 2010): Rufous owls more often exploit open habitats than white owls, and feed predominantly on voles rather than murids (Roulin 2004; Charter et al. 2012). Females choose specific breeding habitats that match their color phenotype, with rufous females producing more offspring in open habitats and white females more offspring in woodier habitats (Dreiss et al. 2012). Rufous males make higher reproductive investments and sire more fledglings in some years (Roulin et al. 2001). Rufous juveniles develop better with limited food conditions (Roulin et al. 2008) and show more altruistic behavior than nonrufous juveniles (Roulin et al. 2012). Moreover, several studies link rufous coloration to increased juvenile dispersal distance (Van den Brink et al. 2012; Roulin 2013) and phenotypic traits enhancing long-distance flight (longer wings, Roulin 2004; Charter et al. 2012) and intense flying (Charter et al. 2015), and may suggest that rufous coloration is connected to a disperser phenotype.

Because of the correlations of coloration with other phenotypes and the strong pleiotropy of the melanocortin system that underlies melanin-based coloration (Ducrest et al. 2008), conclusive evidence for the selective target(s) and agent(s) establishing the European barn owl color cline is lacking. Identifying the gene(s) underlying barn owl plumage coloration and their extent of pleiotropy may therefore provide clues to determine whether selection in the European barn owl color cline acts on coloration itself or instead may rather target one of the correlated life-history traits. Also, the so far restricted sampling of the species range—which does not include Southeastern Europe, the Middle East, and North Africa—precludes ultimate conclusions about the range-wide colonization history under/following which the cline evolved, because incomplete phylogeographic sampling may result in misleading conclusions about colonization histories (see, e.g., Yannic et al. 2012). Studying the evolution of color phenotypes and the underlying genetic variants in the context of the species' range-wide colonization history therefore promises to provide important insights into both the origin and evolutionary history of barn owl color phenotypes, and possible genome-wide consequences of ecology-driven selection.

Here, we investigated genetic data from 22 microsatellite markers and a mitochondrial marker (NADH dehydrogenase 6 gene, *ND6*) together with data on melanin-based coloration and from a candidate color gene (melanocortin 1 receptor, *MC1R*) in an extensive sampling of $>$700 European barn owls (*T. alba*) from 28 populations from Europe, North Africa, and the Middle East. We (i) reconstructed the circum-Mediterranean colonization history of barn owls, (ii) identified a major QTL underlying barn owl plumage coloration and studied the levels of variation linked to this QTL variant, and (iii) contrasted the population structure of coloration and of the color QTL with neutral genetic population structure to infer the role of selection in the evolution of the barn owl color cline.

Methods

SAMPLING AND MOLECULAR ANALYSES

The study includes 724 unrelated barn owls from 28 different localities around the Mediterranean and throughout Europe (Fig. 1, Table S1). Genomic DNA was extracted from the basal millimeter of breast feather quills, or from blood or muscle stored in 96% ethanol. DNA extractions were performed on a BioSprint 96 extraction robot using the BioSprint 96 DNA blood kit, or the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). Sex was determined using the molecular method by Py et al. (2006).

All 724 individuals were genotyped at 22 microsatellites with no evidence for null alleles and no constant deviation from Hardy–Weinberg equilibrium (Table S2) (Burri et al. 2008; Klein et al. 2009). The same loci were genotyped in outgroup populations from California, USA (*T. a. pratincola*, $N = 65$), Australia (*T. a. delicatula*, $N = 19$), and Singapore (*T. a. javanica*, $N = 11$). For 456 individuals plus 24 individuals from California ($N = 8$),

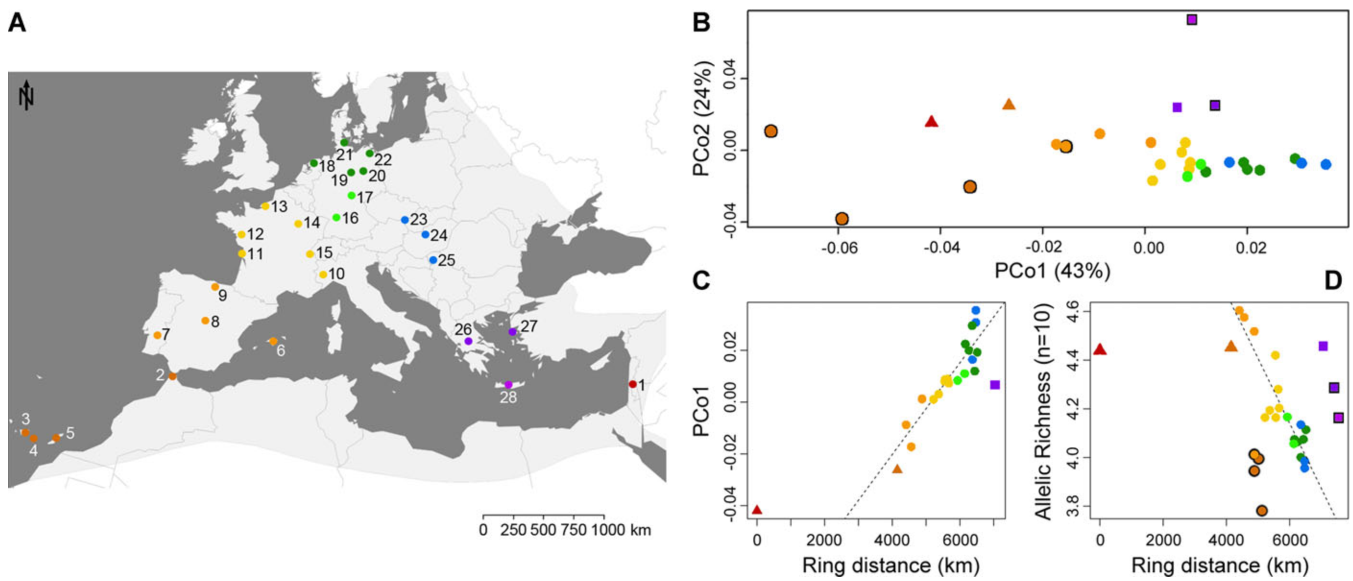


Figure 1. **Sampling and population structure. (A) Geographic distribution of sampling locations of barn owls. The distribution of barn owls in the study range is shaded in light gray. 1, Middle East; 2, North Africa; 3, Tenerife; 4, Gran Canaria; 5, Eastern Canaries (Lanzarote, Fuerteventura); 6, Baleares; 7, Portugal; 8, Central Spain; 9, Northern Spain; 10, Italy; 11, France La Rochelle; 12, France Nantes; 13, Northern France; 14, Eastern France; 15, Switzerland; 16, Southern Germany; 17, Germany Sachsen; 18, Netherlands; 19, Germany Niedersachsen; 20, Germany Brandenburg; 21, Denmark; 22, Northeastern Germany; 23, Czech Republic; 24, Hungary; 25, Balkans; 26, Greece; 27, Aegean; 28, Crete. (B) PCoA based on microsatellite markers. Coloration follows (A). (C) Correlation of the first PCo axis of mainland populations with ring distance from the Middle East. The regression line is based on all European populations but Greece. (D) Geographic structure of microsatellite allelic richness. The regression line is based on all European mainland populations. (C) and (D): Greece, the Aegean, and Crete are depicted with squares, Middle Eastern and North African populations with triangles, islands in (D) with a black border.**

Singapore ($N = 8$), and Australia ($N = 8$) a fragment of 411 bp of the mitochondrial *ND6* gene was sequenced (Table S2).

Extensive studies in mammals and birds (reviewed in Hoekstra 2006; Roulin and Ducrest 2013) place the *MC1R* gene as a prime candidate gene responsible for melanin-based color polymorphisms. To study whether *MC1R* variant explain color variation in barn owls, 543 bp ($N = 211$) or the whole 998 bp ($N = 417$) of the single exon were sequenced for 671 individuals from Europe (Table S1), and outgroup individuals from California ($N = 3$), Australia ($N = 1$), and Singapore ($N = 1$) (Supporting Information Methods). For museum samples and some low-quality DNA samples amplification of these fragments was impossible ($N = 43$), and we used an allelic discrimination assay for the Val126Ile mutation (Supporting Information Methods).

Microsatellites were genotyped in Gene Mapper version 4.0 (Applied Biosystems, Zug, Switzerland). *ND6* and *MC1R* sequences were assembled and edited in CodonCode Aligner version 3.7.1, and aligned using Clustal W (Thompson et al. 1994). *MC1R* polymorphisms were phased using PHASE (Stephens et al. 2001) in DnaSP 5.10.01 (Librado and Rozas 2009) using 1000 burn-in and 1000 post-burn-in iterations and a thinning interval of 10. A neighbor net haplotype network of *MC1R* alleles was constructed using SplitsTree 4.12.3 (Hudson and Bryant 2006).

NEUTRAL GENETIC POPULATION STRUCTURE AND ADMIXTURE ANALYSES

To infer neutral population structure and diversity, pairwise F_{ST} between populations and allelic richness (A_R) based on microsatellite and mitochondrial data were estimated using hierfstat version 4-10 (Goudet 2005). The same package was used to perform principle coordinate analyses (PCoA) based on pairwise F_{ST}. Patterns of isolation by distance inferred using Mantel tests were performed using ecodist 1.2.7 (Goslee and Urban 2007). A haplotype network for *ND6* was constructed using TCS 1.21 (Clement et al. 2000) to infer population structure in haplotype frequencies. Haplotypes from California, Singapore, and Australia were disconnected from the European haplotype network and thus not included in the results.

Linear models were used to test whether population genetic structure (as estimated by PCo axes coordinates) and genetic diversity (A_R) are related to geographic proximity measured as (i) flight and (ii) shortest overland distance, or to (iii) ring distance (see Fig. S1 for an example). Flight distance and shortest overland distance were used, because barn owls have good flying abilities and colonized many far off-shore islands, but may still avoid crossing large water bodies. Ring distance was modeled as the shortest overland distance between populations within a clockwise ring starting in the Middle East and ending in Crete. Start and end points of the ring were determined from the population structure depicted by PCoA (Figs. 1B, S2). Note that even with a colonization origin in Iberia, this distance depicts the ring scenario with a break in population structure situated between Greece and the Middle East. Distances were measured in Google Earth.

To estimate the number of genetic clusters and admixture proportions for each individual, clustering analyses implemented in STRUCTURE 2.3.4 (Pritchard et al. 2000) were performed using microsatellite data at K from $K = 1$ to $K = 10$. Analyses were run once for all populations, and once for mainland populations plus the Aegean exclusively. The latter was included with mainland populations because it was a focal population for admixture analyses and because of the Aegean islands' proximity to the mainland and their genetic diversity being very similar to the neighboring mainland population in Greece (all other island populations show much lower genetic diversity). Ten runs each with a burn-in of 10^5 and 10^6 post-burn-in generations were performed for each K using an admixture model with correlated allele frequencies. The number of clusters was inferred using the ΔK method (Evanno et al. 2005) implemented in STRUCTURE HARVESTER (Earl and vonHoldt 2012).

To estimate which populations/clusters exhibit allele frequencies closest to outgroup populations from California, Australia, and Singapore, we performed three analyses. First, we estimated a neighbor-joining (NJ) population tree using F_{ST} and Cavalli-Sforza and Edwards chord distance (Cavalli-Sforza and Edwards 1967). Second, we ran STRUCTURE (5×10^4 burn-in, and 5×10^5 post-burn-in generations) including outgroup populations, and estimated NJ trees based on net nucleotide distance among clusters to infer the relationships between the inferred clusters. We ran 10 iterations for each K, and increased K until no new cluster splitting off one or several populations was observed (at $K = 7$). Three microsatellites showed more than 10% missing data in the outgroup populations, and were excluded from both analyses (Tak.Oeo53, Ta.305, Ta.408). As island populations showed increased differentiation relative to the other populations, and to avoid problems connected with long-branch attraction of the NJ method, they were excluded from both analyses. Third, we reconstructed the phylogenetic relationships among European, American, and Australasian (Singapore and Australia) *ND6* haplotypes using MrBayes 3.2.5 (Ronquist et al. 2012). Haplotypes from *Phodilus badius* (Tytonidae; Genbank accession no. KF961183) and *Otus elegans* (Strigidae; Genbank accession no. EU123912) were included as outgroups. Two runs with four chains each were run for 2×10^6 generations, including 5×10^5 burn-in generations, using a GTR + G + I nucleotide substitution model. Convergence was evaluated based on the average standard deviation of split frequencies among runs (0.008).

To test whether the mean and variance in admixture proportions (as estimated using STRUCTURE with $Q = 2$) observed in the secondary contact zone in Greece and the Aegean (see Results) differed significantly from those in other populations with similar admixture proportions (Portugal, Central and Northern Spain), we performed nonparametric bootstrap tests over individuals as described below. As we observed no significant differentiation of the populations identified as belonging to the secondary contact zone (Greece, Aegean), these populations were pooled for the following analysis. Random samples of the same size as the combined Greek and Aegean populations ($N = 46$) were drawn from the admixture proportion (Q) distributions of the combined Iberian populations ($N = 61$, these populations show the most similar admixture patterns to Greece). The same procedure was performed using the hybrid index (HI) (Buerkle 2005) instead of Q. For each analysis 10^6 random samples were drawn with replacement to generate the expected distribution of the variance in admixture proportions. HI was estimated using the "introgress" package (Gompert and Buerkle 2010). To estimate HI, "southern" and "northeastern" parental populations were defined from STRUCTURE results, by using individuals with a respective $Q \geqq 0.97$ and without missing data from Northern and Eastern Europe (northeastern lineage), and from the Middle East, Crete, and the Canary Islands (southern lineage), respectively. The northeastern parental population thus consisted of 78 individuals from the Netherlands ($N = 10$), Niedersachsen ($N = 15$), Brandenburg ($N = 8$), Denmark ($N = 14$), Northeastern Germany ($N = 8$), Hungary ($N = 11$), Czech Republic ($N = 4$), and the Balkans ($N = 8$). The "southern" parental population consisted of 24 individuals from the Middle East ($N = 5$), Eastern Canaries ($N = 6$), and Crete ($N = 13$).

MEASUREMENT AND QUANTITATIVE GENETICS ANALYSES OF COLOR PHENOTYPES

Pheomelanin-based plumage color of 626 birds was measured from one to five breast feathers (mean: 4.15, SD: 1.08). Reflectance spectra from four points per feather were captured with a USB4000 spectrophotometer (Ocean Optics, Dunedin, FL) and a DH-2000-bal dual deuterium/halogen light source (Mikropackan, Ostfildern, Germany). For each reflectance spectrum, the brown chroma was calculated following Montgomerie (2006). For each individual, values were averaged per feather and then per individual. Color measurements were high repeatable (97.6% of among-individual variance), as shown by the repeated measurement in 14 individuals one year apart. Feathers are archived in the last author's collection.

To test whether color differentiation among populations evolved by genetic drift exclusively or whether local adaptation played a role in its evolution, we used two approaches. The first is derived from classic F_{ST}–Q_{ST} comparisons (Merilä and Crnokrak

2001; McKay and Latta 2002; Leinonen et al. 2008, 2013; Brommer 2011). We estimated F_{ST} as outlined above. P_{ST} (an analog of Q_{ST} based on phenotypic traits) for color differentiation was estimated using ANOVA with color as response variable and the populations of origin and sex as explanatory variables following Antoniazza et al. (2010). Variance components were then combined with the estimated heritability of coloration (Roulin and Dijkstra 2003). To estimate whether clinal variation in *MC1R* genotypes differs from clinal variation in other genes involved in the expression of coloration, we also estimated P_{ST} after taking into account *MC1R* genotype in a linear model. Following the general approach by Whitlock (2008), Whitlock and Guillaume (2009) proposed a parametric bootstrap approach to compare overall Q_{ST} and F_{ST}. This approach was not applicable to our system in which quantitative measurements are field based, because it requires a resampling of the variance components of Q_{ST} from a quantitative genetics breeding design. We thus used the general idea of Whitlock's (2008) proposition and directly compared empiric P_{ST} estimates to the theoretical χ^2 distribution of Lewontin and Krakauer (1973) to obtain a *P*-value for the F_{ST}–P_{ST} comparison. F_{ST} estimates for *ND6* (divided by four to account for the difference of effective population size of mitochondrial markers) and for *MC1R* were added for comparison. The second approach by Ovaskainen et al. (2011) compares Bayesian estimates of population differentiation at neutral genetic makers and quantitative traits. We applied this approach using the R packages RAFM (Karhunen and Ovaskainen 2012) and Driftsel (Karhunen et al. 2013).

RELATING COLORATION TO *MC1R* POLYMORPHISM, GEOGRAPHY, AND NEUTRAL GENETIC ANCESTRY

To investigate whether *MC1R* polymorphism is associated with individual variation in pheomelanin-based coloration, linear models were estimated with coloration (brown chroma) as the response variable, and *MC1R* genotype at amino acid position 126 and sex as factors (*MC1R* variants at position 126 related to the white and rufous phenotype are referred to as $MC1R_{RUFOUS}$ and $MC1R_{WHITE}$, respectively). Models were estimated for the entire dataset, taking into account population structure using each individual's position on the first principal component from an individual-based principal component analysis (PCA) on neutral genetic diversity as a covariate, and for each population separately. PCA was conducted using the adegenet 1.3-7 R package (Jombart et al. 2008). Population structure for the nonsynonymous polymorphism at amino acid position 126 was estimated by calculating F_{ST} using hierfstat. We investigated how population structure at *MC1R* correlated with color differentiation and spatial distances between populations using Mantel tests and partial Mantel tests as implemented in ecodist.

To test whether there were significantly more variable sites linked to $MC1R_{WHITE}$ (which was sampled more often) than to $MC1R_{RUFOUS}$, we applied a randomization test on all variation observed across the full sequence (10 polymorphic sites). Only sequences covering all variable sites were used for this analysis (618 $MC1R_{WHITE}$, 252 $MC1R_{RUFOUS}$). Then, we randomly sampled 252 of the 618 $MC1R_{WHITE}$ alleles for 10^6 times. Each time we estimated the mean number of pairwise differences (π) in the sample. The proportion of times the observed π is larger than those generated by randomization is an estimate of the sought *P*-value.

To test for linkage disequilibrium of coloration with the neutral genetic background in the secondary contact zone, linear models were used to examine the hypothesis that within the secondary contact zone rufous individuals have an elevated proportion of neutral genetic ancestry from Northeastern Europe. As the populations from Greece and the Aegean are not genetically differentiated, they were combined for this analysis. Models included coloration (brown chroma) as a response variable, and sex, *MC1R* genotype, and neutral genetic ancestry as explanatory variables. We used three approaches to estimate neutral genetic ancestry. The first models made use of the admixture proportions (*Q*) attributed to the "Northeastern" cluster by STRUCTURE for mainland populations (including the Aegean). Second, ancestry was approximated by individuals' first principal axis score of an individual-based PCA (PCA eigenvalue of first axis = 5.7). The PCA was conducted in R using the adegenet 1.3-7 package. Third, neutral genetic ancestry was expressed in terms of the HI estimated as outlined above.

Results

RING-LIKE POPULATION STRUCTURE AROUND THE MEDITERRANEAN

Genetic differentiation among barn owl populations follows a shallow pattern of isolation by distance (overall F_{ST}: microsatellites, 0.045; *ND6*, 0.134) (Table 1, Fig. S3). Strongest differentiation on the mainland was identified between the populations in Eastern Europe and the Middle East, discordant with these populations' spatial proximity (microsatellite F_{ST}: Hungary-Middle East, 0.085; Balkans-Middle East, 0.084; Fig. S4A). PCoA showed the same close genetic proximity of the Middle Eastern population with the geographically distant populations from the Canary Islands, North Africa, and Iberia, while placing it most distant from the geographically closer populations in Eastern Europe (Figs. 1B, S2), suggesting ring-like colonization around the Mediterranean. This hypothesis is consistent with the geographic distribution of admixture proportions estimated in Bayesian analyses of population structure (Pritchard et al. 2000) (Fig. 2A, B). These analyses found highest support for two

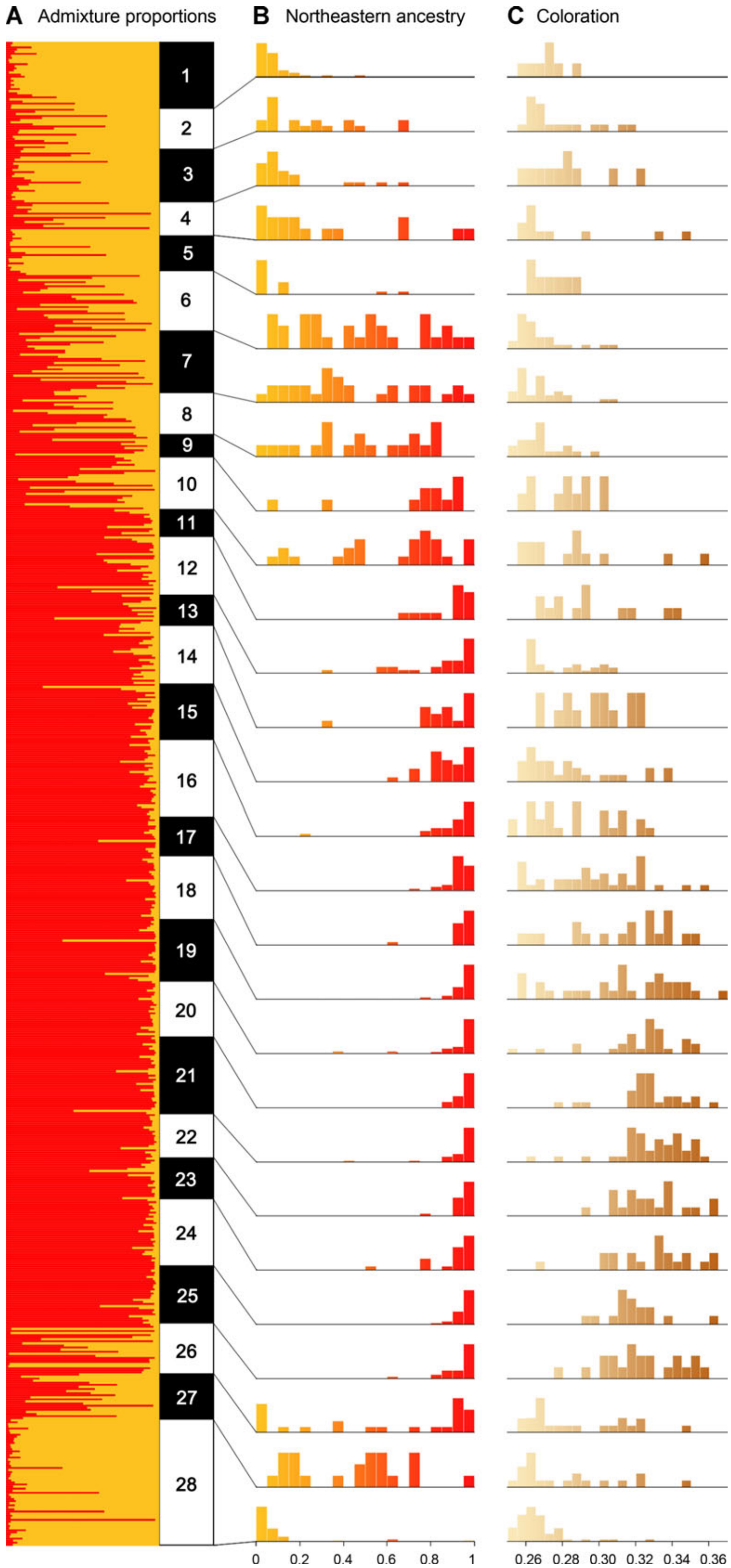


Figure 2. Admixture and color distribution per barn owl population. (A) STRUCTURE barplot. Each horizontal line represents one individual. Proportions to which each individual was assigned to the southern (orange) and northeastern (red) cluster are depicted. (B) Population-wise frequencies at which individuals were assigned to the southern (left, orange shading) and northeastern (right, red shading) clusters. (C) Population-wise color distribution (brown chroma). Legends to population numbers are provided in Figure 1 and Table S1.

Table 1. Mantel *r* for tests comparing classical isolation-by-distance and ring-colonization hypotheses.

	Mantel tests						Partial Mantel tests flight			
	Islands included			Islands excluded			Islands included		Islands excluded	
	Flight	Land	Ring	Flight	Land	Ring	Land	Ring	Land	Ring
Msat	0.616	0.616	0.674	0.751	0.777	0.845	0.079	0.381	0.326	0.59
	(−0.001)	(−0.001)	(−0.001)	(−0.001)	(−0.001)	(−0.001)	(−0.17)	(−0.002)	(−0.071)	(−0.001)
ND6	0.441	0.461	0.544	0.62	0.632	0.722	0.161	0.355	0.159	0.471
	(−0.001)	(−0.001)	(−0.001)	(−0.001)	(−0.001)	(−0.001)	(−0.059)	(−0.001)	(−0.12)	(−0.013)
MC1R	0.255	0.184	0.168	0.307	0.253	0.25	−0.349	−0.064	−0.324	−0.029
	(−0.004)	(−0.019)	(−0.061)	(−0.001)	(−0.003)	(−0.002)	(−1)	(−0.781)	(−1)	(−0.655)
Color	0.199	0.126	0.109	0.356	0.319	0.265	−0.361	−0.088	−0.34	−0.122
	(−0.01)	(−0.066)	(−0.119)	(−0.002)	(−0.001)	(−0.002)	(−1)	(−0.812)	(−1)	(0.902)

In Mantel tests and partial Mantel test taking into account flight distance insularity was taken into account when island populations were included. *P*-values are provided in parentheses. Flight distance is denoted "Flight," shortest overland distance "Land," and ring distance "Ring."

clusters among mainland populations ($\Delta K(2) = 251$, Fig. S5A–D), and for two and three clusters with islands included ($\Delta K(2) = 237$, $\Delta K(3) = 331$, Fig. S5E–H). Individuals from the Middle East, North Africa, the Canary Islands, and Crete were predominantly assigned to a "southern" lineage, whereas populations from Central, Northern, and Eastern Europe formed a "northeastern" lineage ($K = 2$; Fig. 2A, B). Populations from Iberia to Central Europe showed a gradual change from southern to northeastern ancestry (Fig. 2A, B). The island population from Crete was split off from the southern lineage into its own cluster at $K = 3$.

To corroborate the finding of ring colonization around the Mediterranean, we performed tests that contrast alternative colonization hypotheses. Classical hypotheses of isolation by distance were modeled using distance matrices representing (i) flight distance, and (ii) shortest overland distance; ring colonization with a secondary contact zone in Greece was modeled using (iii) ring distance, that is, distance within a clockwise ring starting in the Middle East, through North Africa, Iberia, and Europe, and ending in Crete (see Fig. S1), following PCoA results (Figs. 1B, S2). Two methods that compared these models unequivocally supported the ring-colonization scenario (Tables 1, 2). In linear models that related flight, overland, and ring distance from the Middle East to genetic structure measured by coordinates of the first PCo axis (PCo1), ring distance explained the highest amount of genetic structure ($R^2 = 0.74$, Table 2, see also Fig. 1C). Flight distance best explained the genetic structure along PCo2 ($R^2 = 0.40$, Table 2), indicating genetic exchange among spatially close populations following colonization. In Mantel tests, pairwise flight, overland, and ring distances all significantly correlated with genetic differentiation (Table 1, Fig. S3). However, ring distance explained most of the variance (Mantel $r = 0.674$, Table 1), also when accounting for dispersal (flight distance) among geographically close populations unrelated to colonization history (Mantel $r = 0.381$, Table 1). Results were concordant between microsatellites and mitochondrial data, and not sensitive to the inclusion or exclusion of island populations (Tables 1, 2).

GENETIC DIVERSITY AND ORIGIN OF THE RING-LIKE COLONIZATION

Colonization is expected to leave traces in the geographic distribution of neutral genetic diversity. Highest diversity was found in Southern Iberian populations (microsatellite allelic richness, A_R: Portugal, 4.60, Central Spain, 4.58) followed by Italy, Northern Spain, North Africa, and the Middle East (Fig. 1D). Lowest diversity on the mainland was found in Eastern Europe (microsatellite A_R: Hungary, 3.96; Balkans, 3.98; Brandenburg, 4.00) (Fig. 1D). Diversity is lower on islands (islands: median $A_R = 4.00$, range 3.78–4.28; mainland: median $A_R = 4.17$, range 3.96–4.60) and decreases eastwards in the north of the Mediterranean (latitude × longitude: $t = -3.84$, $P < 10^{-3}$; insularity: $t = 3.17$, $P = 0.004$; longitude: $t = 3.94$, $P < 10^{-3}$; latitude: -1.24, $P = 0.23$; $R^2 = 0.64$). Diversity also showed a strong tendency to decrease with increasing ring distance (islands included: insularity: $t = -2.13$, $P = 0.043$, $R^2 = 0.28$; ring distance $t = -1.80$, $P = 0.084$; mainland populations: $t = -2.97$, $P = 0.008$, $R^2 = 0.31$, Fig. 1D) but no relation to flight or shortest overland distance. Within the European continent, a strong significant decrease of diversity with distance from the Middle East was observed, with ring distance explaining most of the variance (ring distance: $t = -8.73$, $P < 10^{-6}$, $R^2 = 0.82$; flight: $t = 3.87$, $P = 0.001$, $R^2 = 0.47$; overland: $t = 5.00$, $P < 10^{-3}$, $R^2 = 0.60$) (Fig. 1D). The same relationship was found for mitochondrial diversity ($t = -2.24$, $P = 0.038$, $R^2 = 0.22$).

To infer the population with allele frequencies closest to the ancestral population, we introduced outgroup populations from

Table 2. Correlations of principle coordinate axes with geographic variables in barn owls (R2).

Marker	Distance	PCoA 1 Isl. included	Mainland	PCoA 2 Isl. included	Mainland
Msat	Flight	0.41 (0.034)	0.04 (0.351)	0.40 ($<10^{-3}$)	0.20 (0.040)
	Land	0.40 (0.043)	0.03 (0.476)	0.34 (0.003)	0.15 (0.075)
	Ring	0.74 ($<10^{-6}$)	0.76 ($<10^{-6}$)	0.05 (0.825)	0.01 (0.748)
ND6	Flight	0.23 (0.169)	0.06 (0.269)	0.01 (0.815)	0.04 (0.394)
	Land	0.24 (0.140)	0.04 (0.360)	0.02 (0.638)	0.04 (0.366)
	Ring	0.58 ($<10^{-4}$)	0.59 ($<10^{-4}$)	0.02 (0.561)	0.07 (0.218)

When islands were included, insularity was included in the model as a factor. P-values are provided in parentheses. Flight distance is denoted "Flight," shortest overland distance "Land," and ring distance "Ring."

California, Singapore, and Australia into analyses of population structure. Both a dendrogram depicting relationships among the clusters inferred by Bayesian analyses, and NJ trees based on pairwise F_{ST} (Fig. S6) or chord distance (not shown) between populations placed the populations from the Middle East at the root of the circum-Mediterranean colonization. Phylogenetic methods implementing nucleotide substitution models provide better estimations of long-term evolution than distance-based methods based on microsatellite data. However, Bayesian reconstruction of the phylogenetic relationships among mitochondrial *ND6* haplotypes provided no further insights into the origin of ring colonization, nodes among European haplotypes having no statistical support (Fig. S6F). According to the microsatellite-based population trees (Fig. 6A–E) barn owls would appear to have first colonized the region south of the Mediterranean out of the Middle East, and later spread through Europe in a more recent expansion from Iberia (Fig. 3, see also Antoniazza et al. 2014). However, this result is to some extent opposed by the observation of highest genetic diversity on the Iberian Peninsula (although higher habitat fragmentation in the Middle East/Northeastern Africa could explain the observed pattern), and not supported by phylogenetic analysis of mitochondrial sequence data, and has therefore to be treated with caution.

SECONDARY CONTACT ZONE IN SOUTHEASTERN EUROPE

The above results place the populations from Eastern Europe and the Middle East, respectively, at the ends of a genetic continuum (irrespective of the colonization origin being situated in the Middle East or Iberia). The geographic region in-between therefore likely holds a zone of secondary contact. In line with this hypothesis, populations from Greece, the Aegean, and Crete exhibit peculiar genetic compositions reminiscent of hybrid zones: (i) microsatellite data placed these three populations at intermediate positions in PCoA (Figs. 1B, S4A), rather than with the geographically closest populations from Eastern Europe (so did also mitochondrial data for Greece and the Aegean; Fig. S2). (ii) The Greek population does not fit the otherwise strong correlation of mainland population structure with ring distance (Fig. 1C); and (iii) for Crete, relationships with other populations were discordant between microsatellites (Figs. 1B, 2A) and mitochondrial *ND6* (Fig. S2).

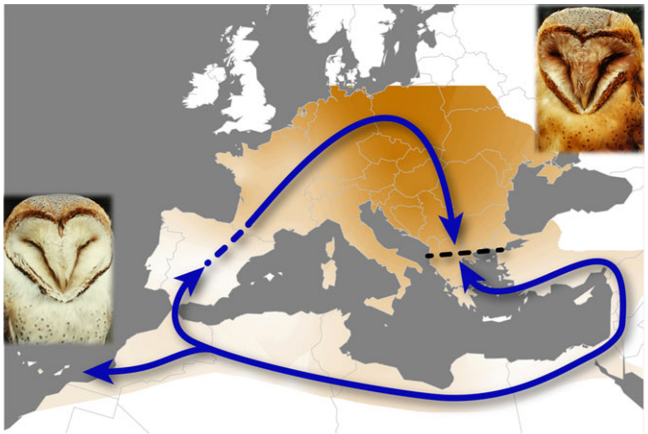


Figure 3. Circum-Mediterranean ring-like colonization of European barn owls. Arrows indicate the putative colonization routes, the black dashed line the region of secondary contact with limited introgression at microsatellites. The color gradient (interpolated using the Kriging algorithm) and typical white and rufous phenotypes are shown.

Results from the above Bayesian clustering analyses were used to obtain further insights into patterns of admixture in Greece, the Aegean, and Crete. All Cretan individuals but four of 61 showed southern ancestry (Fig. 2A, B). In contrast, the prevalence of predominantly Northeastern European haplotypes at mitochondria (Figs. S7, S8) placed Crete unambiguously with Eastern European populations, a pattern also found for one of 22 microsatellites (Ta-220). The presence of northeastern mitochondrial variation on a predominantly southern nuclear background provides evidence for a southern ancestry of Cretan barn owls, with mitochondrial introgression from the north and more limited introgression at the nuclear level.

The Greek and Aegean populations showed a fundamentally different genetic composition than the geographically close Cretan population. Different from all other populations, individual-level admixture appeared restricted: many individuals had either a predominantly northeastern or southern genotype (Fig. 2A, B). Although population-mean admixture was similar to Iberia (Fig. 2A, B), the population-level variance in admixture proportions in Greece and the Aegean were significantly higher than in Iberia (nonparametric bootstrap test, $P < 10^{-6}$). The same result was found using the HI (Buerkle 2005) instead of admixture proportions (nonparametric bootstrap test, $P < 10^{-6}$). This result indicates limited admixture between individuals from northeastern and southern origin within the secondary contact zone.

LOCALLY ADAPTED CLINAL COLOR VARIATION

Plumage coloration is sexually dimorphic with a bimodal distribution in both sexes (Fig. 4A), and showed a pronounced geographic structure. Barn owls in the south are white, whereas in Europe they get gradually darker rufous toward the east (latitude: $t = 5.44$, $P < 10^{-4}$; longitude: $t = -7.60$, $P < 10^{-7}$; interaction: $t = 8.02$, $P < 10^{-7}$; $R^2 = 0.89$) (Fig. 4B). Ring distance explained 22% of variation in population-mean coloration ($t = 2.72$, $P = 0.011$), and only the Middle East, Crete, Greece, and the Aegean did not follow this stark spatial pattern (Fig. S9). With these populations excluded, ring distance alone explained 85% of color variation ($t = 11.23$, $P < 10^{-9}$), which was more than explained by flight ($t = -3.76$, $P = 0.001$, $R^2 = 0.39$) or overland distance ($t = -5.28$, $P < 10^{-5}$, $R^2 = 0.56$).

Color differentiation among populations was marked, and followed an isolation-by-distance pattern best explained by spatial proximity (flight distance) of populations (Table 1). Although population-mean coloration reflects colonization history, color variance among populations appears to be strongly determined by the spatial proximity of populations that might reflect adaptation to a similar environment. In line with this interpretation, F_{ST}–P_{ST} comparisons between mainland populations showed that color differentiation (overall $P_{ST} = 0.40$) exceeds predictions from neutral genetic markers (overall microsatellite $F_{ST} = 0.045$) by an order of magnitude (Fig. 5). Concordantly, P_{ST} differed significantly from the expected distribution of neutral differentiation at microsatellites ($P < 10^{-15}$), demonstrating that genetic drift alone cannot explain color differentiation. Differentiation at mitochondrial *ND6* sequences corrected for mitochondrial effective population size was in the range of microsatellite markers, indicating that the shallow structure at microsatellites is not alone due to high diversity (Jakobsson et al. 2013). Bayesian quantitative genetic modeling (Ovaskainen et al. 2011) also rejected neutral evolution as the only driver of color differentiation ($S > 0.999$). These results provide evidence for diversifying selection on coloration at the scale of the Western Palearctic distribution of barn owls and confirm previous results found at a smaller scale (Antoniazza et al. 2010).

GENETIC BASIS AND ORIGIN OF COLORATION

Sequencing of the *MC1R* gene in 671 individuals revealed a frequent nonsynonymous Ile-Val polymorphism at amino acid position 126. Heterozygotes and Ile-homozygotes were significantly more rufous than Val-homozygotes (Fig. 6). *MC1R* genotype, sex, and population structure (coordinates of the first axes of an individual-based PCA) together explained 65% of color variance (sex: $t = -9.89$, $P < 10^{-15}$; individual PC1: $t = 6.65$, $P < 10^{-10}$; *MC1R*(Val-Ile): $t = -4.67$, $P < 10^{-5}$; *MC1R*(Val-Val): $t = -23.66$, $P < 10^{-15}$). Analyses by population after removing the sex effect confirmed this result (Fig. S10).

To further illustrate the strong link of *MC1R* with coloration, we analyzed the correlations of population structure at *MC1R* with coloration, neutral genetic and spatial population structure. *MC1R* allele frequencies showed a pronounced geographic structure (overall $F_{ST} = 0.383$, range 0–0.797, Figs. 5, 6B, S4B), and population structure (F_{ST}) at *MC1R* was strongly correlated with color differentiation (P_{ST}) (Mantel $R = 0.823$, $P = 0.001$), but much less with neutral genetic differentiation or spatial distances across the entire range (max 0.31, Table 1). Partial Mantel tests demonstrated that neither geography nor neutral genetic differentiation explain color differentiation nearly as well as *MC1R* (Table 1). Moreover, Mantel tests considering only the European continent, where basically all color variation is found (Fig. 4) showed that the color variance explained by geographic distance decreases by half after removing the effect of individual *MC1R* genotypes on coloration (Fig. 5A; without taking into account *MC1R*, Mantel $R = 0.755$, $P = 0.001$; after taking into account *MC1R*, Mantel $R = 0.360$, $P = 0.004$).

Together with the significantly stronger differentiation of *MC1R* than predicted from neutral genetic differentiation (Fig. 5B), this suggests that neutral evolution alone cannot explain population structure of coloration and of the underlying *MC1R* genotypes. The stronger correlation of geographic structure of coloration (and *MC1R*) with spatial proximity of populations (i.e.,

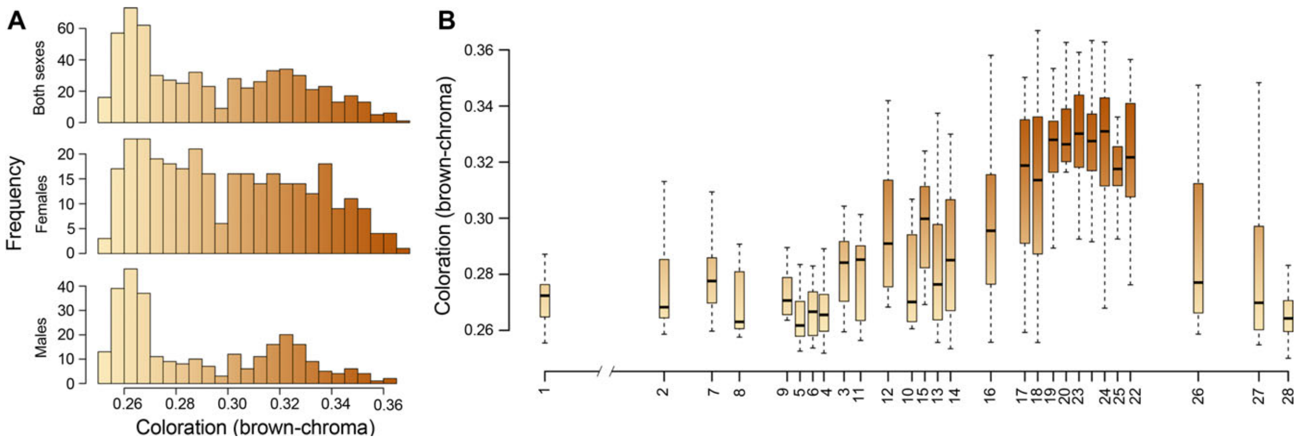


Figure 4. Color distributions in barn owls. (A) Color frequency distributions for both sexes separately and jointly. (B) Geographic distribution of coloration. Boxes are provided at approximate ring distance from the Middle East (exact distances overlap between some populations). Legends to population numbers are provided in Figure 1 and Table S1.

flight distance), rather than with ring distance (Table 1) suggests that selection mediated by locally prevailing environmental factors rather than colonization history determines the geographic structure observed for coloration and *MC1R*, and provides additional evidence for the local adaptation of color phenotypes.

To evaluate the relative ages of the nonsynonymous *MC1R* variants, we analyzed linked genetic variation. Ten low-frequency polymorphisms linked to $MC1R_{WHITE}$ (0.2–3.1% within $MC1R_{WHITE}$) were identified from full-length (998 bp) coding sequences ($N = 870$ alleles). Conversely, only two polymorphisms were found linked to $MC1R_{RUFOUS}$ ($N = 252$ alleles), both occurring at minimal frequency within $MC1R_{RUFOUS}$ (0.4% and 0.8%, respectively). These polymorphisms are a subset of the polymorphisms linked to $MC1R_{WHITE}$. Their higher frequencies on $MC1R_{WHITE}$ and their reticulate position in a haplotype network (Fig. S11) indicate that they originally evolved on $MC1R_{WHITE}$ and got linked to $MC1R_{RUFOUS}$ by recombination. Nonparametric bootstrap tests show that the over 10 times lower polymorphism linked to $MC1R_{RUFOUS}$ ($\pi = 2.16 \times 10^{-6}$) than to $MC1R_{WHITE}$ ($\pi = 2.66 \times 10^{-5}$) is not an artifact from unequal sample sizes ($P < 10^{-6}$). The significantly elevated variation linked to $MC1R_{WHITE}$ together with the match of this variant with the variant observed in outgroup individuals from California, Australia, and Singapore clearly establish $MC1R_{WHITE}$ as the ancestral variant.

Our results therefore strongly suggest that a derived nonsynonymous mutation at *MC1R* was involved in the evolution of a

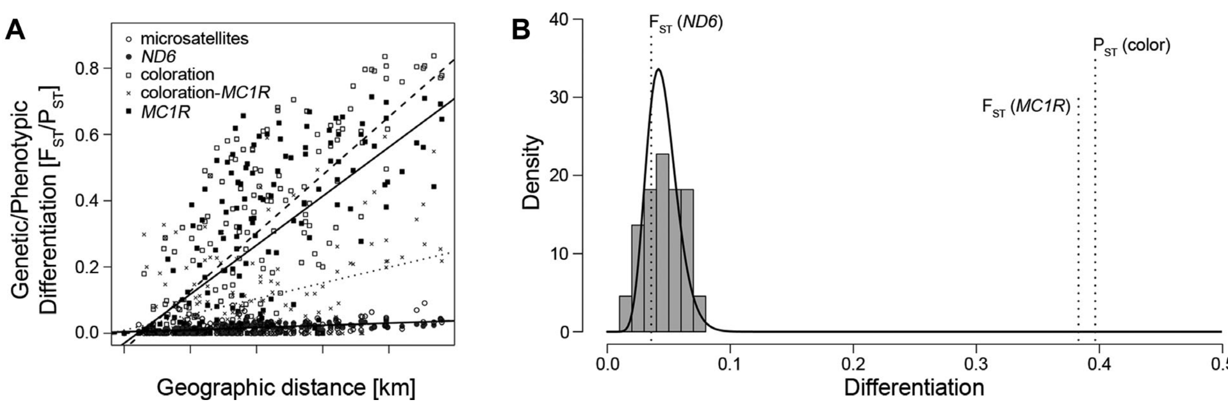


Figure 5. Comparisons of population structure at neutral genetic markers, *MC1R*, and coloration. (A) Pairwise genetic and phenotypic distances as a function of geographic distance within Europe. Regression lines are shown for illustrative purpose (lowest line: microsatellites and *ND6*, overlapping between these markers. Dashed upper line: coloration P_{ST}. Dotted line: coloration P_{ST} after taking into account *MC1R* genotype. Solid upper line: *MC1R*. (B) Overall differentiation in color (P_{ST}) and at *MC1R* are compared against neutral genetic differentiation at microsatellite markers (gray bars) and at the mitochondrial *ND6* locus. Gray bars: histogram of F_{ST} of the 22 microsatellite makers. Solid line: theoretical F_{ST} distribution (Lewontin and Krakauer 1973). Broken lines: overall differentiation at *ND6*, and differentiation in color and at *MC1R*. In both panels differentiation at *ND6* is corrected for differences in nuclear and mitochondrial effective population sizes.

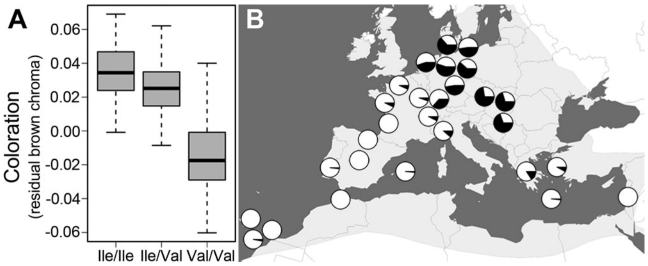


Figure 6. (A) Relationship of coloration with *MC1R* genotype in barn owls. Values provided for coloration are residuals after correcting brown chroma for population structure and sexual dimorphism. (B) Spatial frequency distribution of the $MC1R_{WHITE}$ (white) and $MC1R_{RUFOUS}$ variants (black).

novel rufous barn owl phenotype. The lack of private variation linked to the derived $MC1R_{RUFOUS}$ variant together with this variant's geographic distribution (Fig. 6) indicate that it evolved either shortly before or early during the colonization of Europe

LINKING COLORATION AND GENETIC ANCESTRY

If the rufous phenotype evolved early during the colonization of Europe and is locally adapted, selection may be expected to keep it restricted to a Northeastern European genetic background. Indeed, even after taking into account sexual dimorphism, *MC1R* genotype, and population structure, the genetic ancestry of individuals estimated in Bayesian analyses of population structure (Q) explained a significant amount of color variation (sex: $t = -10.08$, $P < 10^{-15}$; *MC1R*(Val-Ile): $t = -4.73$, $P < 10^{-5}$; *MC1R*(Val-Val): $t = -21.48$, $P < 10^{-15}$; F_{ST}: $t = 2.21$, $P = 0.027$; Q: $t = 7.65$, $P < 10^{-13}$; $R^2 = 0.67$). Still, this result may establish through the spatial correlation of coloration with genetic structure, and potentially insufficient correction for genetic structure in our model. However, especially in the secondary contact zone rufous and white phenotypes are expected to be associated with predominantly northern and southern genetic ancestry, respectively, if gene flow between the terminal populations within the ring is restricted. This prediction is supported by the observation that Greece is not only where individuals of northeastern and southern ancestry meet, but is also the region with highest variance in plumage coloration (Fig. 4B).

We therefore tested whether in the secondary contact zone in Greece and the Aegean rufous coloration was associated with northeastern ancestry by estimating genetic ancestry using (i) admixture proportions (Q), (ii) the first axis of an individual-based PCA, and (iii) the HI. All analyses confirmed that in the secondary contact zone darker rufous individuals have elevated northeastern ancestry (Fig. S12; Q: sex: $t = -2.95$, $P = 0.006$; *MC1R*(Val-Ile): $t = -0.60$, $P = 0.55$; *MC1R*(Val-Val): $t = -2.36$, $P = 0.025$; ancestry: $t = 2.53$, $P = 0.017$, $R^2 = 0.60$; PCA, sex: $t = -2.45$, $P = 0.021$, *MC1R*(Val-Ile): $t = -0.74$, $P = 0.464$; *MC1R*(Val-Val): $t = -2.49$, $P = 0.019$, ancestry: $t = 2.35$, $P = 0.026$, $R^2 = 0.59$; for HI, sex: $t = -2.53$, $P = 0.017$, *MC1R*(Val-Ile): $t = -0.86$, $P = 0.397$; *MC1R*(Val-Val): $t = -2.63$, $P = 0.013$, HI: $t = -3.15$, $P = 0.004$, $R^2 = 0.64$). No other population showed the same consistent correlation between color and genetic ancestry. The secondary contact zone in Greece and the Aegean is thus the only region with evidence for linkage disequilibrium of coloration with neutral genetic ancestry.

Discussion

We found that European barn owls likely colonized their Western Palearctic range in a ring around the Mediterranean (Fig. 3), and that a novel rufous phenotype evolved by a coding mutation in the *MC1R* gene likely around the onset of colonization of Europe. Analyses of the spatial distribution of coloration and the underlying *MC1R* variation provide evidence that the barn owl's color cline evolved by local adaptation during or following this colonization. Finally, admixture patterns and linkage disequilibrium of coloration with the neutral genetic background in Greece suggest that introgression at secondary contact is limited, uncovering a pattern reminiscent of ring species. We in turn discuss the evidence and implications of local adaptation of the color cline, and the hypothesis of circum-Mediterranean ring speciation in European barn owls.

LOCAL ADAPTION: TOWARDS IDENTIFYING THE TARGET AND AGENT OF SELECTION

Local adaptation of barn owl color phenotypes is highlighted by one of the most remarkable findings of the present study: Differentiation of color and of the underlying gene (*MC1R*) is strongest in the geographic region with the weakest—indeed almost absent—population structure (Central, Northern, and Eastern Europe). Moreover, the range-wide bimodal color distribution may suggest reduced fitness of intermediate phenotypes, or a lower frequency of the habitat occupied by this phenotype (or alternatively dominance of the rufous allele). Together with quantitative genetic analyses and previous results (Antoniazza et al. 2010, 2014) this indicates an important role for local adaptation in maintaining the geographic structure of coloration and/or linked phenotypic traits. Alongside these results the identification of *MC1R* as a major-effect color QTL and the demonstration of a steep geographic gradient in *MC1R* allele frequencies parallel to the color cline provide strong evidence for local adaptation, as environment cannot affect allele frequencies other than via natural selection.

Previous work on color-related adaptation identified *MC1R* as a major genetic determinant of color polymorphism in a wide range of vertebrates (Våge et al. 1997; Theron et al. 2001; Mundy et al. 2004; Hoekstra et al. 2006). The identification of this gene as major contributor to color variation in barn owls helps understanding the evolutionary history of color adaptation in this species. *MC1R* is thought to have minimal pleiotropic effects (Mundy 2005) as compared to other melanocortin receptors (Ducrest et al. 2008). Unless *MC1R* has additional, yet to be described phenotypic effects, it therefore appears likely that coloration itself is the target of selection. From this perspective and as indicated by color-related diet (Roulin 2004; Charter et al. 2012), predator–prey interactions seem a likely agent of selection. Adaptation to prey is likely to involve multiple traits related to hunting and to go along with additional habitat-related adaptations. Barn owl color phenotypes may therefore be part of alternative phenotypic syndromes connected to predation that encompass an entire suit of correlated traits (see Introduction). Unless the genes encoding these traits are tightly physically linked (e.g., as a supergene in an inversion, Joron et al. 2011; Thompson and Jiggins 2014), selection for local adaptation still must act on each gene separately. Based on the currently available data, we cannot exclude such a genetic architecture of the phenotypic syndrome correlated with coloration, although the abundant evidence for *MC1R*-related adaptation (e.g., Hubbard et al.; Nachman et al. 2003; Rosenblum et al. 2004; Hoekstra et al. 2006; Nadeau and Jiggins 2010) may suggest that direct selection on *MC1R* is a likely explanation for the strong population structure at this locus and the correlated color phenotypes. Future whole genome analyses to identify genetic variants that parallel clinal variation and explain individual variation in phenotypes (e.g., Poelstra et al. 2014) other than coloration will provide promising insights into the genetic architecture of the phenotypic syndromes.

ADAPTIVE EVOLUTION FROM A NOVEL GENETIC VARIANT

An important question in evolutionary biology addresses the origin of adaptive genetic variation: does local adaptation evolve from genetic variation segregating in populations (standing variation), or from novel, derived variants (Barrett and Schluter 2008)? Studies demonstrating local adaptation of color phenotypes and identifying the underlying genes have previously contributed remarkable insights in this respect (Hoekstra et al. 2006). The $MC1R_{RUFOUS}$ variant predominant in the most recently colonized northeast of Europe appears to be derived and have evolved more recently than the ancestral $MC1R_{WHITE}$ variant. Together with its almost complete absence in the south of the range, this suggests that the $MC1R_{RUFOUS}$ variant arose either shortly before or during the colonization of Central Europe. The supposition that rufous coloration is correlated to a disperser phenotype (Roulin 2004, 2013; Charter et al. 2012, 2015; Van den Brink et al. 2012) adds another interesting perspective in this context, as theoretical models suggest that increased dispersal propensity itself could have conferred rufous birds with an additional advantage during range expansion (Travis and Dytham 2002). This leads us to hypothesize that the rapid colonization of Northern and Eastern Europe by barn owls may have been triggered by the derived phenotype's ability to cope with continental environments, and accelerated by the phenotype's enhanced dispersal ability, or vice versa.

CIRCUM-MEDITERRANEAN RING SPECIATION IN BARN OWLS?

The ring-like isolation-by-distance pattern around the Mediterranean and limited introgression at secondary contact observed in European barn owls are reminiscent of ring species. Ideal ring species consist of populations with a ring-like distribution around a geographic barrier that are interconnected by gene flow throughout the ring, vary gradually in phenotypic traits, and display full reproductive isolation at secondary contact (Irwin et al. 2001a). As yet, none of the few proposed ring species withheld scrutinous examination (Irwin et al. 2001a,b, 2005; Joseph et al. 2008; Kuchta et al. 2009). Even in the prime example of greenish warblers (*Phylloscopus trochiloides*), recent genome-wide data demonstrated multiple breaks in gene flow in the history of the ring (Alcaide et al. 2014) . However, while for greenish warblers deep phylogeographic breaks in mitochondrial variation (Irwin et al. 2001b) gave hints into this direction, barn owls display only gradual shifts in mitochondrial haplotype frequencies. Current data thus suggest that barn owls may withstand the criterion of continuously interconnected populations. However evidence for reproductive isolation at secondary contact is limited and indirect. Future detailed investigations, including ecological and behavioral studies in the secondary contact zone and comprehensive coalescent modeling using genome-wide data will have to provide further insights into the extent and nature of reproductive isolation at secondary contact and refine the picture of population structure across the ring, before a revival of ring species.

ACKNOWLEDGMENTS

We thank Associate Editor R. Brumfield, A. Buerkle, T. Kawakami, P. Nosil, A. Shafer, and anonymous reviewers for comments on previous versions of the article; M. Stöck, A. Cibois, and R. Tomé for help with establishing contacts; and L. Clément and V. Uva for laboratory assistance. R. Alonso, M. Cozzo, the Natural History Museums in Frankfurt and Munich, A. Häller, B. Hartung, M. Hug, G. Klammer, G. Linde, J. Luge, G. Moyne, P. L. Pap, O. Schmidt, and F. Valera provided additional samples. We are indebted to the Wildlife Rehabilitation Centres La Tahonilla (Cabildo de Tenerife), Tafira (Cabildo de Gran Canaria), and the Biological Station of La Oliva (Cabildo de Fuerteventura), and the Natural History Museum of Tenerife and Gustavo Tejera (Lanzarote) for providing tissue samples from the Canary Islands. The authors thank ANIMA (Athens, Greece), the Natural History Museum Paris, and the Association CH-ENE for collaboration, and M. Karhunen for assistance with RAFM and Driftsel. Financial support was provided by the Swiss National Science Foundation, grants 31003A_120517 to AR, 31003A_138180 to JG, and PBLAP3-140171 and PBLAP3-134299 to RB, and the Fondation Agassiz to AR.

DATA ARCHIVING

Microsatellite, mitochondrial ND6, and MC1R genotypes, and coloration data available from the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.2d53k). Mitochondrial and MC1R haplotypes were deposed on GenBank (accession numbers KU143929-KU143967).

LITERATURE CITED

Alcaide, M., E. S. C. Scordato, T. D. Price, and D. E. Irwin. 2014. Genomic divergence in a ring species complex. Nature 511:83–85.

Antoniazza, S., R. Burri, L. Fumagalli, J. Goudet, and A. Roulin. 2010. Local adaptation maintains clinal variation in melanin-based coloration of European barn owls (*Tyto alba*). Evolution 64:1944–1954.

Antoniazza, S., R. Kanitz, S. Neuenschwander, R. Burri, A. Gaigher, A. Roulin, and J. Goudet. 2014. Natural selection in a postglacial range expansion: the case of the colour cline in the European barn owl. Mol. Ecol. 23:5508–5523.

Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. Trends Ecol. Evol. 23:38–44.

Brommer, J. E. 2011. Whither Pst? The approximation of Qst by Pst in evolutionary and conservation biology. J. Evol. Biol. 24:1160–1168.

Buerkle, C. A. 2005. Maximum-likelihood estimation of a hybrid index based on molecular markers. Mol. Ecol. Notes 5:684–687.

Burri, R., S. Antoniazza, F. Siverio, A. Klein, A. Roulin, and L. Fumagalli. 2008. Isolation and characterization of 21 microsatellite markers in the barn owl (*Tyto alba*). Mol. Ecol. Resour. 8:977–979.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Am. J. Hum. Genet. 19:233–257.

Charter, M., O. Peleg, Y. Leshem, and A. Roulin. 2012. Similar patterns of local barn owl adaptation in the Middle East and Europe with respect to melanic coloration. Biol. J. Linn. Soc. 106:447–454.

Charter, M., Y. Leshem, I. Izhaki, and A. Roulin. 2015. Pheomelanin-based colouration is correlated with indices of flying strategies in the barn owl. J. Ornithol. 156:309–312.

Clement, M., D. Posada, and K. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657–1660.

Dreiss, A. N., S. Antoniazza, R. Burri, L. Fumagalli, C. Sonnay, C. Frey, J. Goudet, and A. Roulin. 2012. Local adaptation and matching habitat choice in female barn owls with respect to melanic coloration. J. Evol. Biol. 25:103–114.

Ducrest, A.-L., L. Keller, and A. Roulin. 2008. Pleiotropy in the melanocortin system, coloration and behavioural syndromes. Trends Ecol. Evol. 23:502–510.

Earl, D. A., and B. M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620.

Gompert, Z., and A. C. Buerkle. 2010. introgress: a software package for mapping components of isolation in hybrids. Mol. Ecol. Resour. 10:378–384.

Goslee, S. C., and D. L. Urban. 2007. The ecodist package for dissimilarity-based analysis of ecological data. J. Stat. Softw. 22:1–19.

Goudet, J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes 5:184–186.

Hoekstra, H. E. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity 97:222–234.

Hoekstra, H. E., R. J. Hirshmann, R. A. Bundey, P. A. Insel, and J. P. Crossland. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313:101–104.

Hubbard, J. K., J. A. C. Uy, M. E. Hauber, H. E. Hoekstra, and R. J. Safran. 2010. Vertebrate pigmentation: from underlying genes to adaptive function. Trends Genet. 26:231–239.

Hudson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.

Irwin, D., J. Irwin, and T. Price. 2001a. Ring species as bridges between microevolution and speciation. Genetica 112–113:223–243.

Irwin, D. E., S. Bensch, J. H. Irwin, and T. D. Price. 2005. Speciation by distance in a ring species. Science 307:414–416.

Irwin, D. E., S. Bensch, and T. D. Price. 2001b. Speciation in a ring. Nature 409:333–337.

Jakobsson, M., M. D. Edge, and N. A. Rosenberg. 2013. The relationship between F_{ST} and the frequency of the most frequent allele. Genetics 193:515–528.

Jombart, T., S. Devillard, A.-B. Dufour, and D. Pontier. 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101:92–103.

Joron, M., L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477:203–206.

Joseph, L., G. Dolman, S. Donnellan, K. M. Saint, M. L. Berg, and A. T. D. Bennett. 2008. Where and when does a ring start and end? Testing the ring-species hypothesis in a species complex of Australian parrots. Proc. R. Soc. B: Biol. Sci. 275:2431–2440.

Karhunen, M., and O. Ovaskainen. 2012. Estimating population-level coancestry coefficients by an admixture F model. Genetics 192:609–617.

Karhunen, M., J. Merilä, T. Leinonen, J. M. Cano, and O. Ovaskainen. 2013. driftsel: an R package for detecting signals of natural selection in quantitative traits. Mol. Ecol. Resour. 13:746–754.

Klein, Á., G. J. Horsburgh, C. Küpper, Á. Major, P. L. M. Lee, G. Hoffmann, R. Mátics, and D. A. Dawson. 2009. Microsatellite markers characterized in the barn owl (*Tyto alba*) and of high utility in other owls (Strigiformes: AVES). Mol. Ecol. Resour. 9:1512–1519.

Kuchta, S. R., D. S. Parks, R. L. Mueller, and D. B. Wake. 2009. Closing the ring: historical biogeography of the salamander ring species *Ensatina eschscholtzii*. J. Biogeogr. 36:982–995.

Leinonen, T., R. B. O'Hara, J. M. Cano, and J. Merila. 2008. Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. J. Evol. Biol. 21:1–17.

Leinonen, T., R. J. S. McCairns, R. B. O'Hara, and J. Merila. 2013. QST-FST comparisons: evolutionary and ecological insights from genomic heterogeneity. Nat. Rev. Genet. 14:179–190.

Lewontin, R. C., and J. Krakauer. 1973. Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. Genetics 74:175–195.

Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

McKay, J. K., and R. G. Latta. 2002. Adaptive population divergence: markers, QTL and traits. Trends Ecol. Evol. 17:285–291.

Merilä, J., and P. Crnokrak. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. J. Evol. Biol. 14:892–903.

Montgomerie, R. 2006. Analysing colors. Pp. 90–147 *in* Hill G. E. and K. J. McGraw., eds. Bird coloration. Mechanisms and measurements. Harvard Univ. Press, London.

Mundy, N. I. 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. Proc. R. Soc. B: Biol. Sci. 272:1633–1640.

Mundy, N. I., N. S. Badcock, T. Hart, K. Scribner, K. Janssen, and N. J. Nadeau. 2004. Conserved genetic basis of a quantitative plumage trait involved in mate choice. Science 303:1870–1873.

Nachman, M. W., H. E. Hoekstra, and S. L. D'Agostino. 2003. The genetic basis of adaptive melanism in pocket mice. Proc. Natl. Acad. Sci. USA 100:5268–5273.

Nadeau, N. J., and C. D. Jiggins. 2010. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends Genet. 26:484–492.

Nosil, P. 2012. Ecological speciation. Oxford Univ. Press, Oxford, U.K.

Ovaskainen, O., M. Karhunen, C. Zheng, J. M. C. Arias, and J. Merilä. 2011. A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. Genetics 189:621–632.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–1414.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Py, I., A.-L. Ducrest, N. Duvoisin, L. Fumagalli, and A. Roulin. 2006. Ultraviolet reflectance in a melanin-based plumage trait is heritable. Evol. Ecol. Res. 8:483–491.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Rosenblum, E. B., H. E. Hoekstra, and M. W. Nachman. 2004. Adaptive color variation and the evolution of the MC1R gene. Evolution 58:1794–1808.

Roulin, A. 2003. Geographic variation in sexual dimorphism in the barn owl *Tyto alba*: a role for direct selection or genetic correlation? J. Avian Biol. 34:251–258.

———. 2004. Covariation between plumage color polymorphism and diet in the barn owl *Tyto alba*. Ibis 146:509–517.

———. 2013. Ring recoveries of dead birds confirm that darker pheomelanic barn owls disperse longer distances. J. Ornithol. 154:871–874.

Roulin, A., and C. Dijkstra. 2003. Genetic and environmental components of variation in eumelanin and phaeomelanin sex-traits in the barn owl. Heredity 90:359–364.

Roulin, A., and A.-L. Ducrest. 2013. Genetics of colouration in birds. Semin. Cell Dev. Biol. 24:594–608.

Roulin, A., C. Dijkstra, C. Riols, and A.-L. Ducrest. 2001. Female- and male-specific signals of quality in the barn owl. J. Evol. Biol. 14:255–266.

Roulin, A., J. Gasparini, P. Bize, M. Ritschard, and H. Richner. 2008. Melanin-based colorations signal strategies to cope with poor and rich environments. Behav. Ecol. Sociobiol. 62:507–519.

Roulin, A., M. Wink, and N. Salamin. 2009. Selection on a eumelanic ornament is stronger in the tropics than in temperate zones in the worldwide-distributed barn owl. J. Evol. Biol. 22:345–354.

Roulin, A., A. Da Silva, and C. A. Ruppli. 2012. Dominant nestlings displaying female-like melanin coloration behave altruistically in the barn owl. Anim. Behav. 84:1229–1236.

Stearns, S. C. 1992. The evolution of life histories. Oxford Univ. Press, Oxford, U.K.

Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

Theron, E., K. Hawkins, E. Bermingham, R. E. Ricklefs, and N. I. Mundy. 2001. The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. Curr. Biol. 11:550–557.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Thompson, M. J., and C. D. Jiggins. 2014. Supergenes and their role in evolution. Heredity 113:1–8.

Travis, J. M., and C. Dytham. 2002. Dispersal evolution during invasions. Evol. Ecol. Res. 4:1119–1129.

Våge, D. I., D. Lu, H. Klungland, S. Lien, S. Adalsteinsson, and R. D. Cone. 1997. A non-epistatic interaction of *agouti* and *extension* in the fox, *Vulpes vulpes*. Nat. Genet. 15:311–315.

Van den Brink, V., A. N. Dreiss, and A. Roulin. 2012. Melanin-based coloration predicts natal dispersal in the barn. Anim. Behav. 84:805–812.

Whitlock, M. C. 2008. Evolutionary inference from Q(ST). Mol. Ecol. 17:1885–1896.

Whitlock, M. C., and F. Guillaume. 2009. Testing for spatially divergent selection: comparing QST to FST. Genetics 183:1055–1063.

Yannic, G., R. Burri, V. G. Malikov, and P. Vogel. 2012. Systematics of snow voles (Chionomys, Arvicolinae) revisited. Mol. Phylogenet. Evol. 62:806–815.

Associate Editor: R. Brumfield
Handling Editor: R. Shaw

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. Population sample sizes, sample sizes for genetic markers and color phenotypes, and ring distance from the Middle East.
Table S2. Microsatellite multiplex sets.
Figure S1. Alternative distance models.
Figure S2. Principle coordinate analysis based on population differentiation (F_{ST}) at mitochondrial *ND6* in the barn owl.
Figure S3. Isolation-by-distance patterns for different distance models in the barn owl.
Figure S4. Neighbor-joining trees based on pairwise population differentiation (F_{ST}) estimated from (A) 22 nuclear microsatellite markers and (B) allele frequencies at *MC1R* amino acid position 126 in the barn owl.
Figure S5. STRUCTURE results for mainland populations only (A–D) and for all populations (E–H).
Figure S6. Population trees among European mainland populations rooted with outgroup populations from the United States and Oceania (Australia and Singapore).
Figure S7. Spatial frequency distribution of mitochondrial *ND6* haplotypes in the barn owl.
Figure S8. Mitochondrial *ND6* haplotype network in the barn owl.
Figure S9. Spatial distribution of populations' mean coloration along the circum-Mediterranean ring in the barn owl.
Figure S10. Population-wise relationships of coloration with *MC1R* genotype in the barn owl.
Figure S11. NeighborNet haplotype network of *MC1R* haplotypes.
Figure S12. Correlations of coloration with genetic ancestry in the secondary contact zone (Greece and Aegean).