School of Doctoral Studies in Biological Sciences

University of South Bohemia in České Budějovice

Faculty of Science

Conservation genetics of Galápagos mockingbirds: from immune genes to genomes

Ph.D. thesis

Mgr. Jakub Vlček

Supervisor: doc. RNDr. Jan Štefka, Ph.D.

Department of Parasitology, Faculty of Science, University of South Bohemia

Institute of Parasitology, Biology Centre of the Czech Academy of Sciences

Co-supervisor: doc. RNDr. Michal Vinkler, Ph.D.

Department of Zoology, Faculty of Science, Charles University

České Budějovice 2022

This thesis should be cited as:

Vlček, J., 2022: Conservation genetics of Galápagos mockingbirds: from immune genes to genomes. Ph.D. Thesis Series, No.17. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 244 pp

☆Annotation

In this thesis I have dealt with the population genetic processes of mockingbirds in the Galápagos Islands (*Mimus*) in relation to the limited island area, from the perspective of two types of immune genes and the genome-wide approach. The thesis starts with an introduction to population genetic concepts relevant to conservation genetics followed by description of immune genes: the major histocompatibility complex (MHC) and the Toll-like receptors (TLR). In the final part of the introduction, I address how genetic drift, selection and inbreeding shape genome-wide genetic patterns in small populations. The introduction is followed by four chapters, beginning with an examination of MHCIIβ polymorphism in populations of mockingbirds in the Galápagos Islands. The study of the relationship between MHCIIβ and the abundance of ectoparasites is the content of the second chapter. The third chapter shows how polymorphism of TLRs is shaped by interaction of genetic drift and natural selection. The last chapter deals with the effects of limited island size on inbreeding and genetic load, supported by inferences of the past demography of mockingbirds. The thesis concludes with a summary of the results and their potential implications in the field of conservation genetics.

☆Declaration

I hereby declare that I am the author of this dissertation and that I have used only those sources and literature detailed in the list of references

České Budějovice, 7.10. 2022

…….

Jakub Vlček

This thesis originated from a partnership of Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre of the ASCR, supporting doctoral studies in the Zoology study programme.

Jihočeská univerzita
v Českých Budějovicích
Přírodovědecká
fakulta

BIOLOGY
CENTRE
CAS

☆Financial support

The thesis was supported by following grant awards:

Czech Science Foundation (Project P506/12/P529)

Grant Agency of the University of South Bohemia (Project 048/2019/P)

Charles University (Grant PRIMUS/17/SCI/12)

Czech Science Foundation (Grant P502/19-20152Y)

Swiss friends of Galápagos

Cover image: Jakub Vlček

☆Acknowledgements

So many great beings have helped me along the way that they would not fit on this page. I thank them all!

Primarily, I thank my PhD supervisor Jan Štefka for his calm and steady support. He opened the door to so many great opportunities and introduced me to the world of science and population genetics.

I am grateful to Michal Vinkler as co-supervisor, expert in immune genes and my first host at Charles University. I am also grateful to his team who helped me a lot, namely Hana Velová and Martin Těšický.

I thank Filip Kolář for pushing my limits, hosting me at Charles University and practically being my third supervisor.

I thank Robert Wayne and Kirk Lohmueller at UCLA and their teams, namely Annabel Beichman and Daniel Chavez.

I thank my colleagues from České Budějovice and Prague, namely Marie Drábková, Michaela Fickerová, Alena Sucháčková, Tijana Cvetkovic, Massoud Nazarizadeh, Dragomir Damnjanovic, Milena Nováková, Magdalena Bohutínská and Nelida Padilla.

I thank the people behind the University of South Bohemia, Faculty of Science for the pleasant environment in which I was able to conduct my research and study for many great years. In the same way I'm grateful to the people at Biology Centre, Institute of Parasitology.

This work would not have been possible without the support of my parents, ancestors, friends and my wife and daughter. I dedicate the thesis to them.

Finally, I thank nature for allowing me to reflect on itself. ☆☆☆☆☆

☆List of publications and manuscripts and author's contribution

The thesis is based on the following publications and manuscripts (listed chronologically):

I. **Vlček J**, Hoeck P.E.A., Keller L.F., Wayhart J.P., Dolinová I., Štefka J. (2016) Balancing selection and genetic drift create unusual patterns of MHCIIβ variation in Galápagos mockingbirds. Molecular Ecology. 25, 4757–4772 (IF=6.6).
Vlček Jakub and ŠJ conceptualise ideas behind the study. VJ carried out library preparation and sequencing of MHC amplicons and all subsequent analyses. DI helped with MHC sequencing. HP, KL and WJ collected samples and genotyped microsatellite data. VJ wrote the manuscript with help of ŠJ and all the co-authors. VJ contribution 60%.

II. **Vlček J.**, Štefka J. (2020) Association between louse abundance and MHC II supertypes in Galápagos mockingbirds. Parasitology Research. 119, 1597–1605 (IF=2.3).
Vlček Jakub and ŠJ conceptualise ideas behind the study. VJ carried out all analyses and wrote the manuscript with help of ŠJ. VJ contribution 75%.

III. **Vlček J.**, Miláček M., Vinkler M., Štefka J. (2022) Effect of population size and selection on Toll-like receptor diversity in populations of Galápagos mockingbirds. Journal of Evolutionary Biology, In print (IF=2.5).
Vlček Jakub conceptualised ideas behind the study with help of VM and ŠJ. VJ carried out library preparation and sequencing of

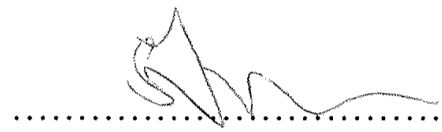
TLR amplicons and majority of subsequent analyses. MM carried out protein analysis. VJ wrote the manuscript with the help of VM and ŠJ. VJ contribution 70%.

IV. **Vlček J.**, Espinoza-Ulloa S., Cowles SA., Ortiz-Catedral L., Coutu C., Chaves J.A., Andrés J.A., Štefka J. Genomes of Galápagos mockingbirds reveal the impact of island size and past demography on inbreeding and genetic load in contemporary populations. (Manuscript to be submitted)

Vlček Jakub conceptualised ideas behind the study with help of ŠJ. VJ carried out bioinformatic analysis of resequencing data. ES sequenced, assembled and annotated the reference genome. VJ and ES carried out demographic inference and other analyses. VJ wrote the manuscript with ES helped by all other co-authors. VJ contribution 65%.

☆Co-author agreement

Jan Štefka, the supervisor of this thesis and co-author of all presented manuscripts, fully acknowledges the contribution of Jakub Vlček as the first author and his contributions as stated above.



doc. RNDr. Jan Štefka, Ph.D.

☆Contents

Introduction

Conservation genetics

The decline in abundance of many wildlife species is a common feature of current ecosystem change, termed 'Anthropocene defaunation' (Barnosky et al. 2011, Dirzo et al. 2014, Li et al. 2016, Ceballos et al. 2017, Sánchez-bayo & Wyckhuys 2019, Ceballos et al. 2020, Ceballos et al. 2017). Habitat loss and fragmentation associated with human expansion are responsible for the decline (Aguilar et al. 2006, Maxwell et al. 2016). Overexploitation, invasive species, as well as global climate change, are further triggers for population decline. (Maxwell et al. 2016). Importantly, fragmentation and reductions in population size entail second-order genetic and demographic processes that have the potential to further destabilise populations (Lande et al. 2003, Allendorf et al. 2013). Loss of genetic diversity and inbreeding depression are inevitable consequences of Anthropocene defaunation (Li et al. 2016, Schlaepfer et al. 2018, Leigh et al. 2019). There is strong evidence that these genetic factors influence the resilience of small populations (Spielman et al. 2004, Willi et al. 2006). A thorough understanding and appreciation of these factors can contribute to effective conservation management (Crandall et al. 2000, Frankham et al. 2014, Willi et al. 2022, Hohenlohe et al. 2020), with the ultimate aim to mitigate the impacts of human expansion on biodiversity. Understanding these genetic factors and applying genetic methods to conservation needs is the essence of the field of conservation genetics. Studying the size-limited populations of mockingbirds that have survived for thousands of years in the Galápagos Islands can provide a valuable perspective in this regard.

Reducing population size makes a population more vulnerable to extinction due to a number of factors. The decrease in population size is associated with an increase in demographic stochasticity, defined as the variability in population growth rate due to random sampling of births and deaths in a population of finite size (Lande et al. 2003). An increase

in demographic stochasticity leads to a decrease in the growth rate (e.g. due to an unequal sex ratio of offspring) and a further decrease in population size. Similarly, small populations are more vulnerable to environmental factors (environmental stochasticity) such as natural disasters, climate change or the introduction of invasive species, simply because they are smaller and therefore have a smaller range or less phenotypic variability to either run away or adapt. The extinction of several species of honeycreepers (Drepanidae) following the introduction of malaria to the Hawaiian Islands (Scott et al. 2001) or the extinction of the Lyall's Wren (*Traversia lyalli*) due to a lighthouse keeper's cats (Galbreath & Brown 2004) are illustrative examples of such vulnerability. The Allee effect is another factor that reduces the average fitness of an individual due to the reduction of population size (Stephens et al. 1999). There are several ecological mechanisms of the Allee effect, such as mate limitation or degradation of the cooperation in group animals (Kramer et al. 2009). Finally, there are genetic factors that make small populations more vulnerable (Allendorf et al. 2013). These factors are sometimes considered as an alternative mechanism of the Allee effect in a broader sense (Willi et al. 2005, Luque et al. 2016). Importantly, the synergistic interaction of these factors in a small population can reduce its viability, further reducing its size, which in turn amplifies the effects of these factors in a lethal feedback loop called the extinction vortex (Gilpin & Soule 1986).

Genetic factors were first considered in conservation biology in the 1970s. Understanding of these factors has advanced rapidly in the first two decades of the 21st century (Willi et al. 2022), along with the rapid development of modern evolutionary synthesis (Provine 2001) and deoxyribonucleic acid (DNA) sequencing (Padmanabhan et al. 1974). There are several aspects of how small population size affects microevolutionary forces, which in turn affect allele or genotype frequencies, which then condition phenotype and population viability. Genetic diversity as a summary of differences in DNA molecules

between individuals was the first important aspect to be considered as an evolutionary (and also industrial and agricultural) resource (Frankel 1974). Apart from this crude aspect of genetic diversity as a raw material for evolution, it is an overarching value that reflects the interplay of microevolutionary forces within populations. Therefore, genetic diversity is considered a valuable segment of biological diversity that requires the attention of conservation efforts (McNeely et al 1990, Hoban et al. 2022). Genetic diversity plays an important role in conservation and population genetics. It was the discovery of an unexpected level of genetic diversity (Lewontin & Hubby 1966, Harris 1966) that provided the impetus for the development of the nearly neutral theory of molecular evolution (Kimura 1983, Charlesworth et al. 2016).

Molecular genetic diversity is determined by the rates of DNA mutation, genetic drift, selection, and gene flow (Graur & Li 2000, Ellegren & Galtier 2010). Errors in the passing of DNA between generations (mutations) are the only definitive source of genetic diversity (Nei 2014). Several different types of mutations are recognised, usually categorised by their size. On one side, there are large-scale mutations that affect entire chromosomes, such as duplication, followed by shorter inversions, insertions, and deletions of DNA fragments (Graur & Li 2000). On the other side of this spectrum are the substitutions or deletions of single molecular units of DNA, the nucleotides. Nucleotide substitutions are one of the most commonly studied types of mutations (Graur & Li 2000, Nielsen et al. 2012). In this context, genetic diversity is increased by the emergence of mutations (new alleles) in a population. The resulting molecular differences are referred to as single nucleotide polymorphisms (SNPs) (Sherry et al. 1999). While new mutations increase genetic diversity, the fixation and loss of the alleles is a process that reduces genetic diversity.

Alleles at neutral sites, defined as sites with negligible phenotypic effects, are randomly lost in time mainly through genetic drift (Charlesworth

2009). This is an important concept, representing a process of fluctuation in allele frequencies due to random sampling in populations of limited size (Wright 1931, Charlesworth 2009, Masel 2011). In an idealised Wright-Fisher population where all individuals contribute equally to reproduction, the magnitude of genetic drift is inversely proportional to the population size (N) (Crow & Kimura 1970). Smaller populations therefore exhibit greater genetic drift, more rapid allelic fluctuations, higher fixation rates, and consequently lower diversity. Drift reduces heterozygosity by $\frac{1}{2N}$ per generation, which was already described by Mendel (1865), although the term drift was introduced only later by Wright (1931). If one considers only the mutation rate (μ) as the input and genetic drift as the sink, then the genetic diversity of nucleotides resulting from the balance between these two forces is mathematically described as $4N\mu$ (Ewens 1972), which is called population scaled mutation rate (θ). Simply put, if one imagines populations of different sizes whose size is constant, whose mutation rate is identical, and in which there is no gene flow, then the differences in neutral diversity between them are solely due to the differences in their population size.

However, population size is rarely stable in real populations, and the other Wright-Fisher assumptions are also rarely satisfied (Charlesworth 2009). For example, the population may experience a cyclical change in size, as is the case with small rodents (Andreassen et al. 2021), or it may experience a large reduction for a variety of demographic reasons, referred to as a bottleneck (Nei 2005). Another violation of Wright-Fisher assumptions is caused by unequal reproductive success of males and females (Janicke et al. 2016). All these factors amplify the effects of genetic drift by reducing the number of effectively reproducing individuals. For this reason, the concept of effective population size (N*e*) was developed early in the history of population genetics (Wright 1931). It is a quantity that reflects the extent of evolutionary change due to drift, rather than a temporary census size. For real populations, therefore, it is

more correct to say that their genetic diversity scales with N*e*, which is one way to account for the assumptions of the Wright-Fisher model. It is worth noting here that the genetic diversity of populations can be indicated by a dozen different metrics, such as heterozygosity as the average frequency of genotypes with two alleles in the case of diploid populations, allelic richness as the average number of alleles normalised by sample size (Foulley & Ollivier 2006), the number of polymorphic sites (K) or average number of pairwise nucleotide differences in a population (π). Each measure has a slightly different response to demography or N*e* in general. For example, a bottleneck will reduce allelic richness faster than heterozygosity (Allendorf 1986). Similar logic is behind the widely used demographic index Tajima's D, which compares K and π, as the two different perspectives on population scaled mutation rate (θ) (Tajima 1989).

While the effect of genetic drift shapes diversity at all genomic positions, the effects of natural selection are important for coding and regulatory genomic regions (Nielsen 2005, Craig et al. 2018). Coding regions (CDS) contain the heritable information stored in DNA that is actually translated into a phenotypic trait through transcription and translation (Crick 1958, Graur & Li 2000). Natural selection causes differential survival or reproductive success of individuals due to variation in these phenotypic traits (Darwin 1859). The variant of a phenotypic trait that increases survival and reproduction of an individual is also more likely to be passed on to the next generation compared to other variants. In other words, the frequency of the phenotype and the underlying genotype or alleles varies according to the fitness they confer. Natural selection is a key mechanism of evolution that drives the constant adaptation to the ever-changing conditions of life on Earth.

The effects of natural selection on genetic diversity depend on the type of natural selection. In most cases, selection reduces diversity by accelerating the fixation of advantageous and loss of disadvantageous

alleles, respectively (Crow & Kimura 1970). The effect of selection on the diversity of CDS is evident when comparing the genetic diversity of the sites where a nucleotide change does not alter an amino acid due to codon redundancy (Crick 1958) (four-fold degenerate sites) and the sites where each nucleotide change also alters an amino acid (zero-fold degenerate sites). Zero-fold diversity is generally lower than four-fold diversity (Chen et al. 2017). The reduction in diversity is mainly due to purifying selection that removes deleterious mutations, as this category of mutations is the most common (Eyre-Walker & Keightley 2007). In addition, a significant amount of diversity is also removed by the interaction between purifying selection and linkage disequilibrium, referred to as background selection (Charlesworth et al. 1993, Charlesworth & Charlesworth 2018). Importantly, there is also balancing selection that maintains polymorphism, either through heterozygote advantage or frequency-dependent selection (Gillespie 1991). Balancing selection is less common than other types of selection (Andres et al. 2009), although it was considered an important force for maintaining diversity in a population until the advent of the neutral theory of molecular evolution (Kimura 1983). Balancing selection is particularly important for immune genes (Andres et al. 2009), and thus I will discuss this topic in more detail in the following chapter of the thesis introduction.

Gene flow or migration is the last major microevolutionary force affecting genetic diversity in a population. This process is directly related to the subdivision of populations (Slatkin 1987). Natural populations are usually not panmictic, meaning that mate finding is not completely random with respect to geographic distance. More often, populations are subdivided into local, smaller panmictic units (demes). In extreme cases, demes are well-defined and almost completely isolated, for example, in species with low vagility on oceanic islands or in highly fragmented landscapes (Frankham 1997, Aguilar et al. 2006). Subdivision of populations leads to a lack of heterozygotes (i.e., lower diversity)

compared to Hardy-Weinberg equilibrium, which is known as the Wahlund effect (Allendorf et al. 2013). Subdivided demes show more inbreeding and genetic drift compared to a panmictic population containing the same total number of individuals (Charlesworth et al. 2003). In this way, genetic diversity is reduced and genetic differentiation between demes increases (Woolfit & Bromham 2005). In contrast, gene flow (i.e, migration and genetic mixing between differentiated demes) increases heterozygosity and dampens differentiation (Slatkin 1987).

In the previous sections, I introduced basic concepts of population genetics and showed that genetic diversity is largely shaped by four microevolutionary forces. Now, I will return to the problem of small populations and the genetic factors that influence their viability. Above I showed that genetic diversity is lower in smaller and more fragmented populations. In this respect, low diversity can serve as an indicator of endangerment (Frankham et al. 2010). But genetic diversity is not just a consequence of low population size; it might be a cause of loss of evolutionary potential as well. Populations with low genetic diversity show less adaptability to changing environments. This was exemplified in an experiment with fruit flies where larger populations were more likely to adapt to high salt concentration compared to smaller ones (Frankham et al. 1999). Since the fruit fly experiment, studies have been accumulating that demonstrate a significant relationship between genetic diversity and various components of fitness in wild populations (for a review, see DeWoody et al. 2021). Further theoretical and empirical support for this phenomenon comes from the field of quantitative genetics.

Quantitative genetic variation will erode in small populations as it depends on genetic diversity (Lande 1995). The amount of additive variance, together with heritability, determines the adaptive capacity of a population (Willi et al. 2006). Small populations are therefore less likely to adapt due to loss of diversity than larger populations. However, it remains questionable to what degree neutral genetic diversity as a proxy

of population genetic health reflects phenotypic variance. For this reason, a study of genes, where the diversity is adaptive, has been widely advocated in conservation genetics (Hedrick 2001, Höglund 2009).

Adaptive polymorphism in immune genes

Adaptive polymorphism within a population is a result of as well as a substrate for evolution (Cain & Sheppard 1954, Piertney & Oliver 2006). immune genes are of particular interest from the perspective of conservation genetics, as their adaptive polymorphism is associated with the defence against pathogens (Sommer 2005, Höglund 2009). The polymorphism of proteins that recognise pathogens, such as the major histocompatibility complex (MHC) or Toll-like receptors (TLR), influences the ability of an appropriate immune response and thus the evolutionary resilience of the population (Radwan et al. 2010, Morris et al. 2015).

Single alleles of immune genes recognise specific pathogens and thus provide some degree of resistance to them (Tschirren et al. 2013, Sepil et al. 2013). The polymorphism in these genes is to some extent generated and maintained by balancing selection (Richman 2000). Therefore, polymorphism in immune genes is a more appropriate indicator of the genetic health of populations than the non-functional polymorphism of neutral markers (Sommer 2005). In the following chapter, I will summarise what is known about MHC and TLR genes in general and with regard to the field of conservation genetics.

The major histocompatibility complex (MHC) is a diverse family of genes associated with immunological functions (Kasahara 2000). The classical MHC genes encode transmembrane glycoproteins, which are the essential components of the adaptive immune response (Murphy & Janeway 2008). The role of MHC is to present peptides derived from pathogens (antigens) to T cells. Presentation involves binding the peptide

to the peptide binding groove of the MHC molecule and presenting the antigens to the extracellular space where they stimulate the T cell via a T cell receptor. The T cell evaluates the signal and either destroys the signalling cell or further regulates the adaptive immune response to combat the pathogen (Murphy & Janeway 2008). Two MHC classes are distinguished based on their molecular structure, and the origin of the antigen presented and the type of T cells involved. MHC class I molecules present antigens derived from proteins in the cytosol (usually of viral or self origin) to CD8+ T cells (cytotoxic T cells). When a T cell recognises a particular antigen, it releases cytotoxins that eventually cause apoptosis of the signalling cell (Murphy & Janeway 2008). MHC class molecules II present peptides from endocytosed proteins of endoparasitic or bacterial origin to CD4+ T cells (helper T cells). When the antigen is recognised by the T cell, information about the pathogen is passed on and the adaptive immune response is regulated accordingly. Importantly, each protein variant of both MHC classes can only bind and present a limited number of antigens due to the specificity of the peptide binding groove (Falk et al. 1991). Therefore, individual MHC alleles can confer resistance to a limited number of pathogens (Martin & Carrington 2005, Milinski 2006), indicating the central role of adaptive polymorphism in the MHC.

MHC genes contain the highest polymorphism of all genes in jawed vertebrates. In some species, the total number of alleles can reach hundreds, as in the blue tit (*Cyanistes caeruleus*), the collared flycatcher (*Ficedula albicollis*) or humans (Zagalska-Neubauer et al. 2010, Sepil et al. 2012, de Bakker & Raychaudhuri 2012). Interestingly, the actual MHC polymorphism was the first thing scientists noticed when it was discovered in 1968, even before the function was known (Snell 1968). The high number of alleles in the population is also associated with a variable number of paralogues of a particular MHC gene. This feature is called copy number variation (CNV) (Schrider & Hahn 2010). Copy number of MHC genes is particularly high in passerine birds (O'Connor

et al. 2019). It is widely agreed that polymorphism in MHC genes is to some degree maintained by a form of pathogen-mediated balancing selection (PMBS) (Bernatchez & Landry 2003, Milinski 2006, Spurgin & Richardson 2010, Radwan et al. 2020). PMBS is thought not only to produce high levels of polymorphism within species, but also to maintain the same alleles (haplotypes) between species across wide evolutionary timescale. This phenomenon is called trans-species polymorphism (Klein et al. 2007).

Several mechanisms have been proposed for the maintenance of MHC diversity through PMBS: The first mechanism, termed heterozygote advantage or overdominance, (Doherty & Zinkernagel 1975) predicts that an individual heterozygous at MHC loci will have higher fitness than both homozygotes because it can recognise a wider range of pathogens. This form of balancing selection increases the average heterozygosity in the population. Divergent allele advantage, is a form of heterozygote advantage that presupposes that the expression of more divergent alleles within an individual provides selection advantage because such an individual can bind a broader range of pathogens (Wakeland et al. 1990). Second mechanism, the rare allele advantage (Slade & McCallum 1992), is based on a negative frequency dependence between the pathogens and the alleles of the MHC genes. In other words, pathogens will be highly selected to overcome the most common MHC allele, and subsequently the successful pathogen will dominate. But then a rare MHC allele that can recognise such a pathogen and provide good protection spreads throughout the population. The circle closes and fluctuations prevent rare alleles from disappearing from the population. Third, proponents of the so-called fluctuating (diversifying) selection argue that changes in the parasite regime in space and time lead to high polymorphism in MHC genes (Hill 1991).

These four possible mechanisms of pathogen-mediated balancing selection are not mutually exclusive and may interact, making it difficult

to disentangle their effects and roles (Spurgin & Richardson 2010). Studies linking MHC diversity to pathogen resistance provide quite robust evidence that frequency-dependent selection is the major type, as studies typically show that specific MHC alleles confer either resistance or susceptibility. The comprehensive study in great tits (*Parus major*) found that one allele was associated with resistance to *Plasmodium relictum* and another allele was associated with susceptibility to *Plasmodium circumflexum* (Sepil et al. 2013). However, instead of susceptibility, the second allele conferred a quantitative disease resistance whereas the first allele conferred a qualitative disease resistance, based on a theory and nomenclature introduced by Westerdahl et al. (2012). Several other studies have found alleles conferring either qualitative or quantitative disease resistance, e.g. in blue tit (Westerdahl et al. 2013), common yellowthroat (*Geothlypis trichas*) (Dunn et al. 2013), house sparrow (*Passer domesticus*) (Loiseau et al. 2008), european rabbit (*Oryctolagus cuniculus*) (Oppelt et al. 2010), red jungle-fowl (*Gallus gallus*) (Worley et al. 2010), or seychelles warbler (*Acrocephalus sechellensis*) (Brouwer et al. 2010). Interestingly, in each case only one or two alleles actually played a role in pathogen resistance. Perhaps these alleles were the earlier rare alleles that were victorious in the current strains of the pathogens. Finally, the fact that some of the studies found population-specific resistance alleles (Ekblom et al. 2007, Hawley & Fleischer 2012), together with a clear association between specific MHC alleles and survival rates in the Attwater's prairie chicken (*Tympanuchus cupido attwateri*) (Bateson et al. 2016) support the interplay of the rare allele advantage, fluctuating selection and perhaps also the divergent allele advantage (Agudo et al. 2011).

Apart from PMBS also other mechanisms can maintain MHC polymorphism. Assortative mate choice driven by MHC differences is a well known phenomenon (Jordan & Bruford 1998, Havlíček & Roberts 2009, Winternitz et al. 2016). This phenomenon alone may contribute to the maintenance of polymorphism (Eizaguirre et al. 2009). The

associative balancing complex is another mechanism that can potentially contribute to polymorphism in MHC (Oosterhout 2009). In this concept, recessive deleterious mutations associated with MHC alleles lead to the disadvantage of homozygous genotypes and consequently to the maintenance of polymorphism.

While PMBS or other mechanisms maintain polymorphism, the effect of genetic drift can erode it. This is of particular concern in small, fragmented populations where the effects of genetic drift are exacerbated. This raises a central question in MHC conservation genetics, coined by Miller & Lambert (2004): Whether genetic drift can outweigh the balancing selection. Since drift and balancing selection have opposing effects on MHC polymorphism, the question arises as to which force prevails in different demographic scenarios and under different selection constrains. This question has often been addressed by comparing neutral polymorphism, reflecting the effects of genetic drift, and MHC polymorphism, or reflecting the interaction of genetic drift and selection in populations of different sizes in which the strength of drift varied. Originally, it was assumed that balancing selection would preserve MHC polymorphism as opposed to neutral markers such as microsatellites (Hedrick 2001). However, there is an increasing empirical evidence that in small populations drift outweighs balancing selection and diversity in the MHC decreases at least proportionally to neutral diversity (Sommer & Tichy 1999, Miller & Lambert 2004, Radwan et al. 2007, Munguia-Vega et al. 2007, Siddle et al. 2007, Říčanová et al. 2011, Eimes et al. 2013, Pečnerová et al. 2016, Athrey et al. 2018, Arauco-Shapiro et al. 2020, Tang et al. 2022).

The role of genetic drift and bottleneck in shaping MHC polymorphism was summarised by Sutton et al. (2011). In a meta-analysis, they found that MHC polymorphism declined by 15% more than neutral polymorphism in populations following bottlenecks. This “post-bottleneck” pattern has also been found in several recent studies

(Alcaide 2010, Eimes et al. 2011, Arauco-Shapiro et al. 2020, Tang et al. 2022). Two mutually non-exclusive mechanisms have been proposed to explain this pattern. The hypothesis of drift across loci (Eimes et al. 2011) states that genetic drift can fix the same alleles in different paralogous loci and thus reduce the MHC variability twofold compared to unduplicated neutral loci. Alternatively, negative frequency-dependent selection diverts the frequency of rare alleles, which are then removed from small populations by genetic drift more likely than the alleles at neutral loci (Ejsmond & Radwan 2011). In the later article, in which the authors used simulations, it was found that when there was a pronounced bottleneck, variation in the MHC was lost to a greater extent than at neutral loci, but that about 40 generations after the bottleneck, MHC heterozygosity recovered.

Although it seems reasonable to say that genetic drift plays a greater role than balancing selection in small populations, there are particular cases where natural selection has maintained MHC polymorphism despite reduced neutral variability and the exacerbated effect of genetic drift. High MHC but low microsatellite polymorphism has been found in the San Nicolas Island fox (*Urocyon littoralis*) (Aguilar et al. 2004), the guppy (*Poecilia reticulata*) (Oosterhout et al. 2006) and the Berthelot's pipit (*Anthus berthelotii*) (Spurgin et al. 2011). In addition, a particularly straightforward study in water voles (*Arvicola amphibius*) has shown that diversity in the MHC was maintained by balancing selection despite a severe bottleneck (Oliver & Piertney 2012), and similar, albeit less clear-cut, evidence comes from a study in Hawaiian honeycreepers (Drepanidinae) (Jarvi et al. 2004). This suggests that in some cases genetic drift was indeed outweighed by balancing selection. Further clarification of the problem can be achieved by not only studying nucleotide MHC alleles, but by focusing on MHC protein variants clustered according to their effect on the phenotype (Doytchinova & Flower 2005, Lighten et al. 2017).

Toll-like receptors (TLR) have become the second most studied gene family (after MHC) when addressing questions of evolutionary immunogenetics in wild populations, as it also harbours adaptive polymorphism (Grueber et al. 2015, Minias & Vinkler 2022). The functional unit in TLR is formed by a transmembrane glycoprotein that recognises pathogens and mediates the innate immune response (Murphy & Janeway 2008). Upon binding of an antigen to an extracellular/ endosomal domain of TLR, a signalling cascade is initiated that leads to the expression of cytokines. This process subsequently triggers an inflammatory response and the first line of defence against pathogens (Murphy & Janeway 2008, Barreiro et al. 2009, Alcaide & Edwards 2011). This is one of the most important differences from MHC genes, which are associated with the adaptive or second line of defence. TLR recognises specific pathogen-associated molecular patterns, which are common pathogen structures such as lipoproteins or RNA. In birds, there are 10 known TLR receptors, each of which is specialised in recognising a slightly different set of molecular patterns (Alcaide & Edwards 2011). In general, TLR 1,2,4,5,15 recognise bacterial or fungal compounds such as lipoproteins, flagellin or peptidoglycans, while TLR 3,7,21 recognise nucleic acids, mostly of viral origin (Brownlie & Allan 2011).

Although TLR genes are not as polymorphic as the MHC, the evidence of pathogen-mediated selection has also been observed in this group of genes. Much of this evidence comes from the evolutionary analyses of synonymous to non-synonymous polymorphism or substitution rate (dN/dS) (Ferrer-Admetlla et al. 2008, Grueber et al. 2014, Králová et al. 2018). Such studies often showed an episodic occurrence of positive selection in most TLRs, which was interpreted as balancing selection by Minias & Vinkler (2022). The effect of selection is also supported by studies showing associations between single TLR alleles and susceptibility or resistance to some pathogens. In rodents, several TLR2 genotypes were associated with resistance to *Borrelia* infections

(Tschirren et al. 2013), and certain TLR4 genotypes provided resistance to several ectoparasites (Gavan et al. 2015). Associations between single alleles and different pathogens were also found in a bank vole (*Myodes glareolus*) (Kloch et al. 2018) and a roe deer (*Capreolus capreolus*) (Quéméré et al. 2021). In addition, two experiments have shown an association between the expression of TLR genes and infection by coccidian parasites in domestic chickens (*Gallus domesticus*) (Zhang et al. 2012) and wild populations of house sparrows (Martin et al. 2013). A great example is also the Attwater's prairie chicken, where an allele of TLR1B influenced the survival of reintroduced individuals (Bateson et al. 2016). These associations in individual alleles together with evolutionary analyses of molecular polymorphism suggest that TLR are under episodic positive selection. However, purifying selection that stabilises protein structure clearly dominates the evolution of TLR genes (Mukherjee et al. 2009, Darfour-Oduro et al. 2015, Vinkler et al. 2015, Wang et al. 2016, Raven et al. 2017, Nelson-Flower et al. 2018, Levy et al. 2020).

Few studies have examined the effects of genetic drift on TLR polymorphism. In particular, there were only three studies of TLR polymorphism in small endangered populations of wild birds when I began my PhD research. Grueber et al. (2013) showed that in a small population of Stewart's Island robins (*Petroica australis*), allele frequencies were dominated by genetic drift, although an allele of TLR4 was associated with better survival. Hartmann et al. (2014) showed that TLR diversity was extremely depleted in the Pale-headed brushfinch (*Atlapetes pallidiceps*), probably due to an extreme bottleneck in this population. Similarly, Gonzales-Quevedo et al. (2015) observed that genetic drift dominated the process of TLR genetic differentiation among insular populations of Berthelot's pipit (*Anthus berthelotii*). In addition, drift was observed to predominate in several warbler species (*Acrocephalus*) as TLR diversity correlated with population size (Gilroy et al. 2016). However, the dunnock (*Prunella modularis*) introduced to New Zealand has maintained similar levels of diversity to the original

European population in 8 TLR receptors, despite a relatively severe bottleneck 150 years ago (Lara et al. 2020).

Overall, there is compelling evidence that phenotypic and underlying genetic variation in both immune families is essential for successful defence against pathogens. Moreover, a functional and flexible immune response is a particularly important determinant of population resilience. In several cases, populations of endangered species went extinct due to disease outbreaks (Scott et al. 2001, Smith et al. 2006), which can generally be linked to low polymorphism in immune genes (for a review, see Smith et al. 2009). Therefore, it is of utmost importance to understand how this adaptive polymorphism is shaped by specific microevolutionary forces in small, vulnerable populations. Despite ample evidence, it is still unclear at what population size drift dominates in shaping immunogenetic diversity. Another question is: if drift affects the genetic polymorphism of TLR and MHC as shown above, how does it affect the resulting polymorphism of protein phenotypes? These questions motivated me to dedicate a large part of my PhD to the analysis of MHCIIβ and TLRs in different sized populations of Galápagos mockingbirds. I present the result of my efforts in the first three chapters of this PhD thesis.

Inbreeding depression and genetic load

Inbreeding depression is probably the most important genetic factor that can contribute to the deterioration of a population's viability (Keller & Waller 2002). The case of the decimated Florida panther (*Felis concolor coryi*), which exhibits various morphological defects (Roelke et al. 1993), is the most notable example of inbreeding depression in wild populations. Inbreeding depression is generally a term used to describe the reduction in fitness of offspring of closely related mates in contrast to the fitness of offspring of unrelated mates. This phenomenon was already observed by Darwin (1876) in his experiments with plants, where he noted a reduction in the vigour of seedlings produced by self-fertilisation in otherwise

outcrossing plants. There are two mechanisms by which inbreeding reduces the vigour or viability of a population, and both are related to the inflation of homozygosity. The first mechanism relates to genes under heterozygote advantage (e.g. MHC). If a heterozygote has a higher fitness than the two homozygotes, then the increase in homozygosity due to inbreeding inevitably reduces fitness (Charlesworth & Willis 2009). The heterozygote advantage is not limited to MHC and other immune receptors, but is also symptomatic of the self-incompatibility locus in plants (Charlesworth et al. 2005). Inbreeding depression may be mediated by low polymorphism of these genes in some specific cases (Willi et al. 2015) and to some extent (Charlesworth & Charlesworth 1999).

The second mechanism, which is much more likely to contribute to inbreeding depression, concerns recessive or partially recessive deleterious alleles, i.e. a class of alleles that affect the phenotype only in the homozygous state and whose effect is detrimental (Keller & Waller 2002, Hedrick & García-Dorado 2016). This class of alleles is widespread throughout the genome, as the majority of new mutations are deleterious (Eyre-Walker & Keightley 2007) and a significant proportion appear to be recessive (Crow 1993, Huber et al. 2018). Inbreeding exposes the deleterious phenotypic effects of these alleles as homozygosity increases throughout the genome, leading to debilitating phenotypic effects and higher mortality. The hidden fitness effect of these alleles in a population is called genetic load. The term genetic load was originally defined as the cost of selection in general (Muller 1950) and this particular aspect of recessive deleterious alleles has been aptly termed hidden (or masked) load (Bertorelle et al. 2022). From this perspective, the severity of inbreeding depression depends on the level of hidden load and the rate of inbreeding (Hedrick & García-Dorado 2016). The rate of inbreeding is determined by the extent of common ancestry, which depends on life history traits, ecology and also population size. The extent of inbreeding is a function of N*e* (Kimura 1970). In smaller populations, two mates are more likely to be related to some degree than in larger populations.

Importantly, inbreeding does not only concern consanguineous mating between very closely related individuals, but is a continuous effect that can be measured directly from genomic data based on the analysis of runs of homozygosity (Narasimhan et al. 2016, Ceballos et al. 2018).

Several authors have shown that the amount of masked load increases with N*e* (Hedrick & García-Dorado 2016, Kyriazis et al. 2019, van der Valk et al. 2019, Bertorelle et al. 2022). In contrast, smaller populations appear to carry less masked load (Bertorelle et al. 2022). Purging is the most likely explanation for this observation (García-Dorado 2012, Hedrick 1994, Crow 1970). Purging is a phenomenon of increased efficiency of purifying selection to remove recessive deleterious alleles thanks to inbreeding (Hedrick & García-Dorado 2016, García-Dorado 2012). Bertorelle et al. (2022) have shown that in populations with N*e* below 100, masked load is rapidly unmasked and contributes to the realised load, it becomes visible to purifying selection. Purging has the potential to reduce the accumulation of genetic load in some small populations of wildlife (Iberian lynx (*Lynx pardinus*) - Kleinman-Ruiz et al. 2022, vaquita porpoise (*Phocoena sinus*) - Robinson et al. 2022, domestic horses (*Equus caballus*) - Orlando & Librado 2019, Alpine ibex (Capra ibex)(Grossen et al. 2020), island foxes (*Urocyon littoralis*) - Robinson et al. 2018, mountain gorilla (*Gorilla beringei beringei*) - Xue et al. 2015). The effect of purging was also demonstrated in an experiment with fruit flies (*Drosophila melanogaster*) (Pekkala et al. 2012). However, this effect is not universal, as suggested by a review of experimental work (Crnokrak & Barret 2002) and the absence of purging in a tiny population of Chatham Island robins (*Petroica traversi*) (Kennedy et al. 2014). Since genetic load determines the severity of inbreeding depression (Van Oosterhout 2020), understanding its dynamics in relation to N*e* and purging is of urgent importance for current conservation genetics.

The fact that both microevolutionary forces, genetic drift and natural selection, modulate the allele frequencies of populations implies particular consequences of the accumulation of genetic load in small populations. The rate of genetic drift, reflected in N*e*, increases with decreasing population size, while the effect of selection depends only on the fitness difference between genotypes (selection coefficient - s). The fixation probability (PF) of a new mutation with initial frequency 1/2N in a diploid population of size N, assuming that the effect on phenotype is additive, is given by this formula: $PF(1/2N) = \frac{1-e^{-s}}{1-e^{-2Ns}}$ where e is the base of the natural logarithm (Kimura 1962). It states that the probability of fixation is determined by the ratio of the selection coefficient and population size. If we consider deleterious mutations with |s| << 1/N they may even reach fixation due to drift, although their effect is detrimental. The smaller the population, the stronger the effect of mutations that behave neutrally. This phenomenon of less efficient selection in a small population theoretically leads to an accumulation of genetic load (Lynch et al 1995) reffered to as drift load (Hedrick & García-Dorado 2016), and recently several studies have empirically confirmed this phenomenon in wild populations (Robinson et al 2016, Kutschera et al 2020, Leroy et al 2021, Mathur et al 2021). Ultimately, drift load can reduce population viability, which in turn reduces N*e* and further increases genetic load accumulation in a process known as mutational meltdown (Lynch et al. 1995). Although small populations are able to purge masked genetic load as shown above, the question is whether this can outweigh the accumulation of genetic load in the long run. This question, together with the demographic inference of my focal species, motivated the final chapter of this thesis.

Mockingbirds of the Galápagos Islands

Peculiar birds (Aves) of the Americas, possessing a curved bill, the size of the European black bird (*Turdus merula*) and observed both on the mainland and on various islands, triggered Darwin's doubts about the immutability of species (Darwin 1839, Steinheimer 2004). The birds that sparked Darwin's evolutionary thinking are called mockingbirds (*Mimus*) because they tend to imitate the voices of other species. Mockingbirds belong to the superfamily Muscicapoidea and their closest European relatives are starlings (*Sturnus*) (Jetz et al. 2012, Selvatti et al. 2015). In total, there are 14 species of mockingbirds, four of which are endemic to the Galápagos Islands and were previously placed in a separate genus *Nesomimus* (Swarth 1931). The first molecular phylogenetic studies revealed that mockingbirds in Galápagos are monophyletic (Arbogast et al. 2006) and that they are in fact closely related to other mockingbirds, namely the Bahama mockingbird (*Mimus gundlachii*) and other continental species (e.g. Northern mockingbird, *M. polyglottos*). For this reason, despite their uniqueness within this genus, their scientific name was changed back to *Mimus* (Lovette et al. 2012).

Almost all Galápagos Islands harbour a population of these birds (Figure 1).The Hood or Española mockingbird (*M. macdonaldi*, Ridgway 1890) inhabits the southernmost island of Española that has an area of 6048.0 ha, and hosts approximately 600 - 1700 mature individuals. The species is classified as vulnerable (BirdLife International 2020a). The San Cristóbal mockingbird (*M. melanotis*, Gould 1837) is endemic to the island of the same name, located about 50 km northeast of Española. It is a much larger island, covering an area of 55808.6 ha, with 20 - 38 thousand mature individuals. The species is listed as near threatened (BirdLife International 2020b). The Floreana mockingbird (*M. trifasciatus*, Gould 1837) currently occurs only on two small islets near Floreana island 70 km west of Española. The species became extinct from Floreana about

150 years ago by the introduction of cats and rats (Curry 1986). On Isla Gardner by Floreana, with its 81.2 ha, an average of 131∓46 individuals lived between 2003 and 2008, and on Champion, a small eroded tuft cone, with its 9.5 ha, lived only 38∓12 individuals (Jiménez-Uzcáteguy et al. 2011). The divergence between these two populations is only shallow, as they were connected in the past via Floreana (Hoeck et al. 2010a), but the fixation index Fst between the populations based on microsatellites is high due to genetic drift (Hoeck et al. 2010b). This species is considered endangered (BirdLife International 2018a). The Galápagos mockingbird (*M. parvulus*), the last and most widespread species, inhabits the rest of the archipelago and is the least concerned from a conservation perspective (BirdLife International 2018b).

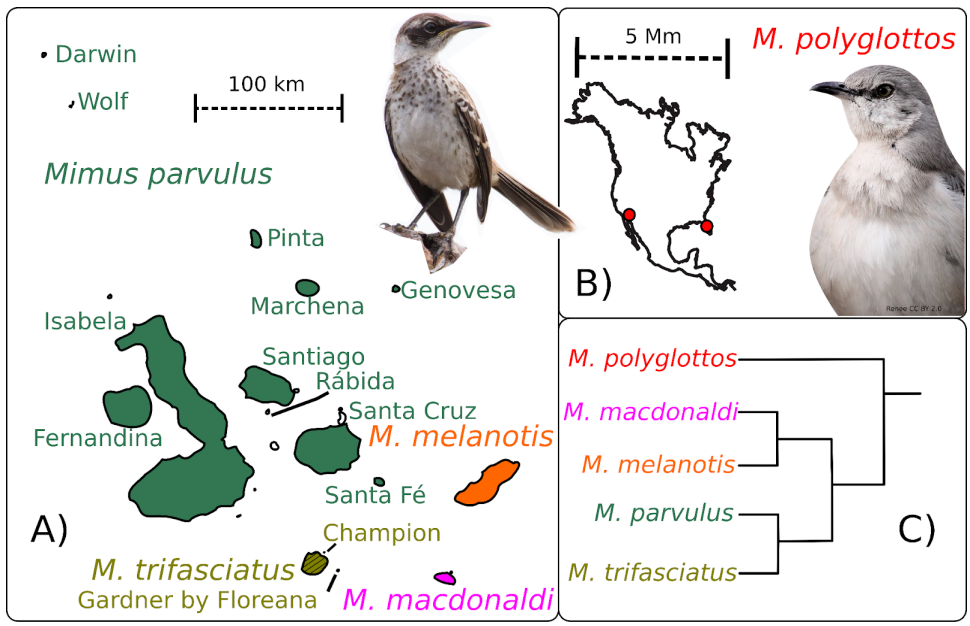


Figure 1 Distribution of species and populations of mockingbirds in the Galápagos Islands (A) with their continental relative (B) as outgroup in an illustrative species phylogenetic tree based on Lovette et al. (2012) (C). Photo of Galápagos mockingbird (*M. parvulus*) by Jan Štefka and

northern mockingbird (*M. polyglottos*) by Renee CC BY 2.0, Flickr. The abbreviation Mm stands for millions of metres.

Galápagos mockingbird is the only polytypic mockingbird in Galápagos with 6 subspecies occurring in the central, western and northern islands. *M. p. hulli* (Rothschild 1898) inhabits the northernmost small island of Darwin (110 ha). *M. p. wenmani* (Swarth 1931) occurs on Wolf (130 ha), which is about 38 km south-east of Darwin. *M. p. personatus* (Ridgway 1890) is endemic to Pinta (5940 ha), Marchena (12996 ha), Santiago (58465 ha) and Rábida (499), islands north of Santa Cruz (Figure 1). *M. p. bauri* (Ridgway 1894) is endemic to Genovesa (1411 ha) and is probably a hybrid between *M. parvulus* and *M. melanotis* (Nietlisbach et al. 2013). *M. p. parvulus* (Gould 1837) inhabits the central and westernmost islands of Santa Cruz (98555 ha), Isabela (458812 ha), Fernandina (64248 ha) and nearby islets. Finally, *M. p. barringtoni* is found on Santa Fé (2413 ha), which lies between Santa Cruz and San Cristóbal. With this detailed description, I wanted to illustrate that each mockingbird population inhabits an island of different size, which is an essential feature with regard to this thesis.

Long-term studies of population ecology, mainly on Genovesa but also on other islands, provided detailed information on the factors affecting the demography of mockingbirds in Galápagos. Curry & Grant (1989) followed the population on Genovesa for more than 11 years and were able to show that rainfall has a decisive influence on breeding phenology and reproductive success. In dry years, the birds did not breed at all, while in wet years there was an upsurge of breeding as the food supply increased. Population density was consequently influenced by the continuing climatic oscillation, driven by the El Niño-Southern Oscillation (ENSO) phenomenon (Curry & Grant 1989). The average generation time has been estimated at 4.5 years, at least for the Floreana mockingbird (Grant et al. 2000).

Mockingbirds are food generalists.

They forage from centipedes
(*Scolopendra galapagoensis*)
to tissue surrounding opuntia seeds
(*Opuntia megasperma*) (Grant & Grant 1979)
and they can also drink vital fluids
from sea lions' wounds
(*Zalophus californianus* - Curry & Anderson 1987).

Mockingbirds are also known for their cooperative breeding strategy (Kinnaird & Grant 1982), which is probably due to the limited territories on a range-restricted island (Curry 1989). The extent of helping behaviour also depends on ENSO, with helpers usually being males and their contribution increasing reproductive success (Curry & Grant 1989). Cooperative breeding has been observed in all species except the San Cristóbal mockingbird (Curry 1989). On the tiny Champion Island, which can typically accommodate only 11 territories, cooperative breeding and brooders' tolerance of non-breeders allow the presence of twice as many sexually mature individuals as would otherwise fit in the limited number of territories (Curry 1989, Grant et al. 2000). Thus, apart from island size, these demographic fluctuations must also influence the rate of genetic stochasticity (N*e*) experienced by mockingbird populations.

Mortality is another demographic phenomenon that determines population density and N*e*. The source of mortality is largely unknown in mockingbirds, but the pox virus or other viruses introduced to Galápagos with poultry (Deem et al. 2012) were most likely responsible for an epizootic outbreak in 1982-83 on Genovesa (Curry & Grant 1989). In contrast, fledgling success in mockingbirds does not appear to be affected by an invasive fly parasite, *Philornis downsi*, which severely impacts

Darwin's finch survival (Knutie et al. 2016). The remnant populations of Floreana mockingbirds are under special protection (Charles Darwin Foundation 2008) and therefore complex surveys for parasites and haematology were conducted to assess potential threats (Deem et al. 2011). No serious viral or bacterial pathogens such as mycoplasma, chlamydophila or paramyxovirus were found in the populations and only coccidian parasites were present in the Gardner by Floreana. Although the Champion population appears to be less healthy compared to Gardner by Floreana based on haematology (Deem et al. 2011), Champion, being an extremely small population, has no abnormalities that would indicate inbreeding depression (Grant et al. 2000). However, further risk assessment through immunogenetic diversity analysis has been called for (Hoeck & Keller 2012).

Population genetic studies of mockingbirds in the Galápagos Islands revealed that each island forms an allopatric deme. Hoeck et al. (2010b) laid the foundation for this thesis by showing that most populations are well separated and that island size correlates with N*e* based on 17 microsatellite loci. They also showed that populations are in drift-mutation equilibrium based on historical samples from older expeditions, and that diversity decreased only in the smallest population of Champion. In a follow-up study, Hoeck & Keller (2012) investigated whether more inbred populations also had higher susceptibility to pathogens or lower immunocompetence to test the hypothesis that inbreeding and low genetic diversity in small populations reduce their viability (McCallum & Dobson 1995). Several indices of immune health based on white blood cell count, haemolysis-haemagglutination test and parasite count, showed no correlation with inbreeding coefficients. Hoeck & Keller (2012) speculated on several explanations as to why the birds on the smaller islands are not less healthy than their conspecifics on the larger islands. One of the hypotheses was that populations that are small over a long period of time may have already purged the genetic load or adapted their immune systems. These were the triggering ideas for my

research. At the same time, my PhD supervisor was involved in research on mockingbirds, which showed that two ectoparasite species share the same phylogeographic patterns as their hosts (Štefka et al. 2011). This confirmed the findings of Hoeck et al. (2010b) that inter-island migration is very limited and populations are well-defined demes where island size determines the extent of genetic drift.

Aims of the thesis

My main aim was to investigate how specific demographic conditions in mockingbirds in the Galápagos Islands (i.e. limited area and isolation) affect the evolution of immune genes, the degree of inbreeding and genetic load.

As mentioned earlier, small population size introduces specific second-order genetic, ecological and demographic factors that have the potential to destabilise the population by further reducing its viability (Lande et al. 2003, Willi et al. 2006, Spielman et al. 2004, Allendorf et al. 2013). I focused on genetic factors and particularly immunogenetic factors because at the beginning of my study, theory prevailed over empirical evidence in this research area. Furthermore, I focused on genetic factors because newly developed methods for high-throughput DNA sequencing became available (Ekblom & Galindo 2011). It must be noted here that genetic factors are unlikely to deliver the coup de grâce, yet they certainly contribute to the potential extinction vortex of a population (Spielman et al. 2004). Loss of genetic diversity is the first factor, which is considerable, especially at loci where polymorphism is essential for pathogen recognition (Sommer 2005). The second genetic factor is the introduction of deleterious mutations that impair the function of a particular protein. The third factor is an increase in homozygosity, which induces the expression of recessive deleterious mutations and potentially leads to purging (Crow 1970, Hedrick 1994). Theoretically, the degree of influence of these factors is directly related to population

size. My general objective was to test empirically whether the molecular indices reflecting these three factors covary with population size in the case of mockingbirds in Galápagos. In other words, how is the interaction of genetic drift, inbreeding and natural selection modulated by population size.

*The first specific objective addressed in **Chapter I** was to show how the polymorphism of the beta subunit of the major histocompatibility complex class II (MHCIIβ) evolved in populations of mockingbirds of different sizes in Galápagos*. The MHCIIβ subunit, and in particular the region we sequenced, is in direct contact with pathogenic particles (Murphy & Janeway 2008). These particles need to be recognised and bound to MHCIIβ, which triggers an adaptive immune response (Murphy & Janeway 2008). For this reason, MHCIIβ is under strong pathogen-mediated balancing selection (PMBS), which maintains high polymorphism (Spurgin & Richardson 2010). We compared Galápagos populations with a mainland mockingbird population of the closely related Northern mockingbird (*M. polyglottos*). We hypothesised that various indices of polymorphism would be smaller in the island populations than in the larger mainland population, and that within the archipelago these indices would correlate with island area and neutral microsatellite diversity when genetic drift predominates over balancing selection. We used indices reflecting both genetic polymorphism and protein physicochemical traits assessed by grouping of functionally similar protein variants into supertypes (Doytchinova & Flower 2005). We also examined molecular polymorphism for evidence of balancing selection, indicated either by dN/dS ratios or trans-species polymorphism. *In addition, we examined the relationship between louse abundance and MHCIIβ supertypes to show the functional impact on ectoparasite load, which is covered in **Chapter II***. In our natural setting, with populations of different sizes, we further tested specific hypotheses on MHC evolution. We tested for the presence of a divergent allele advantage (Wakeland et al. 1990), i.e. simply put, whether in smaller populations more divergent

alleles occur, to offset the effect of lower allele count. We have tried to find evidence for the drift-across-loci hypothesis, which states that due to the presence of multiple paralogues, diversity in the MHC is lost more rapidly in smaller and bottlenecked populations (Eimes et al. 2011).

*The second specific target covered in **Chapter III** is similar to the previous one, except that we have focused on various Toll-like receptors (TLR), which also carry adaptive polymorphism relevant in innate immune response.* We calculated the diversity of synonymous sites (i.e. without selection, neutral) and non-synonymous sites (i.e. under selection, non-neutral) and hypothesised that neutral diversity is shaped by genetic drift, while the linear relationship between island area and non-neutral diversity is biassed due to the effect of selection (or the interaction of both microevolutionary forces). In addition, we investigated the phenotypic consequences of non-synonymous mutations by analysing the distribution of protein variants across the archipelago and their physicochemical properties such as tertiary structure and surface charge, which was inspired by the work of Těšický et al. (2020). In addition, we also included the Northern mockingbird, a relative on the mainland, to compare the extent of polymorphism and to test whether selection constrains have changed between the mainland and the Galápagos Islands.

*The third specific objective, covered in **Chapter IV**, was to clarify the past demography and colonisation history of Galápagos mockingbirds and to show how demography, in combination with limited island area, influenced inbreeding, genomic diversity and the extent of genetic load.* Specifically, we resequenced whole genomes of three individuals from each of the eight populations of Floreana and Galápagos mockingbirds representing small (Champion, Gardner by Floreana, Darwin, Wolf), medium (Pinta and Marchena) and large population size categories (Santa Cruz and Isabela). We mapped the resequencing data to an original genome assembly and determined genotypes for each genomic position.

Using these data, we inferred historical effective population size and hypothesised that there was a single colonisation bottleneck, followed by a divergence of populations and a decline in their N*e*. We then calculated nucleotide diversity, N*e*, heterozygosity, the levels of inbreeding and two indices representing masked and realised load, and we tested the effects of island size on these indices. Based on population genetic theory (Kimura 1983), we assumed that the indices of diversity and N*e* depend predominantly on island area due to genetic drift, and that inbreeding is an inverse function of island area due to the limited number of unrelated mates. Finally, we hypothesised to see lower masked (additive) genetic load and higher realised (recessive) load in smaller populations compared to larger and medium populations if purging is able to prevent the accumulation of deleterious alleles, as the theory suggests (Crow 1970, Hedrick 1994).

References

Agudo R, Alcaide M, Rico C et al. (2011) Major histocompatibility complex variation in insular populations of the Egyptian vulture: inferences about the roles of genetic drift and selection. Molecular Ecology, 20, 2329–40.

Aguilar R, Ashworth L, Galetto L, Aizen MA (2006) Plant reproductive susceptibility to habitat fragmentation: Review and synthesis through a meta-analysis. Ecology Letters, 9, 968–980.

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, 3490–4.

Alcaide M (2010) On the relative roles of selection and genetic drift in shaping MHC variation. Molecular ecology, 19, 3842–4.

Alcaide M, Edwards S V. (2011) Molecular evolution of the toll-like receptor multigene family in birds. Molecular biology and evolution, 28, 1703–15.

Alcaide M, Edwards S V., Negro JJ, Serrano D, Tella JL (2008) Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (*Falco naumanni*). Molecular ecology, 17, 2652–65.

Allendorf FW (1986) Genetic drift and the loss of alleles versus heterozygosity. Zoo Biology, 5, 181–190.

Allendorf FW, Luikart G, Aitken SN (2013) Conservation and the genetics of populations. Wiley-Blackwell, Chichester.

Andreassen HP, Sundell J, Ecke F et al. (2021) Population cycles and outbreaks of small rodents: ten essential questions we still need to solve. Oecologia, 195, 601–622.

Andres AM, Hubisz MJ, Indap A et al. (2009) Targets of Balancing Selection in the Human Genome. Molecular Biology and Evolution, 26, 2755–2764.

Arauco-Shapiro G, Schumacher KI, Boersma D, Bouzat JL (2020) The role of demographic history and selection in shaping genetic diversity of the Galápagos penguin (*Spheniscus mendiculus*). PLoS ONE, 15, 1–20.

Arbogast BS, Drovetski S V, Curry RL (2006) The origin and diversification of Galápagos mockingbirds. Evolution, 60, 370–382.

Athrey G, Faust N, Charlotte Hieke AS, Brisbin IL (2018) Effective population sizes and adaptive genetic variation in a captive bird population. PeerJ, 2018, 1–21.

de Bakker PIW, Raychaudhuri S (2012) Interrogating the major histocompatibility complex with high-throughput genomics. Human molecular genetics, 21, R29-36.

Barnosky AD, Matzke N, Tomiya S et al. (2011) Has the Earth's sixth mass extinction already arrived? Nature, 471, 51–57.

Barreiro LB, Ben-Ali M, Quach H et al. (2009) Evolutionary Dynamics of Human Toll-Like Receptors and Their Different Contributions to Host Defense. PLoS Genetics, 5, e1000562.

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.

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, 363–77.

Bertorelle G, Raffini F, Bosse M et al. (2022) Genetic load: genomic estimates and applications in non-model animals. Nature Reviews Genetics, 23, 492–503.

BirdLife International. 2020a. *Mimus macdonaldi*. The IUCN Red List of Threatened Species 2020: e.T22711070A182738354. https://dx.doi.org/10.2305/IUCN.UK.2020-3.RLTS.T22711070A182738354.en. Accessed on 03 October 2022.

BirdLife International. 2020b. *Mimus melanotis*. The IUCN Red List of Threatened Species 2020: e.T22711078A180990415. https://dx.doi.org/10.2305/IUCN.UK.2020-3.RLTS.T22711078A180990415.en. Accessed on 03 October 2022.

BirdLife International. 2018a. *Mimus trifasciatus*. The IUCN Red List of Threatened Species 2018: e.T22711063A132093428. https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22711063A132093428.en. Accessed on 03 October 2022.

BirdLife International. 2018b. *Mimus parvulus*. The IUCN Red List of Threatened Species 2018: e.T22711053A132093260. https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22711053A132093260.en. Accessed on 03 October 2022.

Brouwer L, Barr I, van de Pol M et al. (2010) MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. Molecular Ecology, 19, 3444–55.

Brownlie R, Allan B (2011) Avian toll-like receptors. Cell and Tissue Research, 343, 121–130.

Cain A, Sheppard PM (1954) The Theory of Adaptive Polymorphism. The American Naturalist, 88, 321–326.

Ceballos G, Ehrlich PR, Dirzo R (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proceedings of the National Academy of Sciences, 114, 201704949.

Ceballos G, Ehrlich PR, Raven PH (2020) Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. Proceedings of the National Academy of Sciences, 117, 13596–13602.

Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF (2018) Runs of homozygosity: windows into population history and trait architecture. Nature Reviews Genetics, 19, 220–234.

Charles Darwin Foundation (2008) The reintroduction of the Floreana Mockingbird to its island of origin. Charles Darwin Foundation and Galápagos National Park, Puerto Ayora.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics, 10, 195–205.

Charlesworth B, Charlesworth D (1999) The genetic basis of inbreeding depression. Genetical Research, 74, 329–340.

Charlesworth B, Charlesworth D (2018) Neutral Variation in the Context of Selection. Molecular Biology and Evolution, 35, 1359–1361.

Charlesworth B, Charlesworth D, Barton NH (2003) The effects of genetic and geographic structure on neutral variation. Annual Review of Ecology Evolution and Systematics, 34, 99–125.

Charlesworth B, Charlesworth D, Coyne JA, Langley CH (2016) Hubby and Lewontin on protein variation in natural populations: When molecular genetics came to the rescue of population genetics. Genetics, 203, 1497–1503.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. Genetics, 134, 1289–1303.

Charlesworth D, Vekemans X, Castric V, Glémin S (2005) Plant self-incompatibility systems: A molecular evolutionary perspective. New Phytologist, 168, 61–69.

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nature Reviews Genetics, 10, 783–796.

Chen J, Glémin S, Lascoux M (2017) Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. Molecular Biology and Evolution, 34, 1417–1428.

Craig RJ, Suh A, Wang M, Ellegren H (2018) Natural selection beyond genes: Identification and analyses of evolutionarily conserved elements in the genome of the collared flycatcher (*Ficedula albicollis*). Molecular Ecology, 27, 476–492.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. Trends in Ecology & Evolution, 15, 290–295.

Crick FH (1958) On Protein Synthesis. Symposia of the Society for Experimental Biology, 12, 138–163.

Crnokrak P, Barret C (2002) Perspective: purging the genetic load: a review of the experimental evidence. Evolution, 56, 2347–2358.

Crow JF (1970) Genetic loads and the cost of natural selection. In: Mathematical topics in population genetics (ed Levin SA), pp. 128–177. Springer, Berlin.

Crow JF (1993) Mutation, mean fitness, and genetic load. Oxford Surv. Evol. Biol, 9, 3–42.

Crow JF, Kimura M (1970) An Introduction to Population Genetics Theory. Harper and Row Publishers, New York.

Curry R (1986) Whatever happened to the Floreana mockingbird. Noticias de Galápagos, 43, 13–15.

Curry RL (1989) Geographic variation in social organization of Galápagos mockingbirds: ecological correlates of group territoriality and cooperative breeding. Behaviroal Ecology and Sciobiology, 24, 147–160.

Curry RL, Anderson DJ (1987) Interisland Variation in Blood Drinking by Galápagos Mockingbirds. Auk, 104, 517–521.

Curry R, Grant P (1989) Demography of the cooperatively breeding Galápagos mockingbird, *Nesomimus parvulus*, in a climatically variable environment. The Journal of Animal Ecology, 58, 441–463.

Darfour-Oduro K, Megens H-J, Roca AL, Groenen M, Schook LB (2015) Adaptive Evolution of Toll-Like Receptors (TLRs) in the Family Suidae. PLOS ONE, 10, e0124069.

Darwin CR (1839) Narrative of the surveying voyages of His Majesty's Ships Adventure and Beagle between the years 1826 and 1836, describing their examination of the southern shores of South America, and the Beagle's circumnavigation of the globe. Journal and remarks 1832–1836. Henry Colburn, London.

Darwin CR (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray.

Darwin CR (1876) The Effects of Cross and Self Fertilization in the Vegetable Kingdom. John Murray, London.

Deem SL, Cruz MB, Higashiguchi JM, Parker PG (2012) Diseases of poultry and endemic birds in Galápagos: Implications for the reintroduction of native species. Animal Conservation, 15, 73–82.

Deem SL, Parker PG, Cruz MB, Merkel J, Hoeck PEA (2011) Comparison of blood values and health status of Floreana Mockingbirds (*Mimus trifasciatus*) on the islands of Champion and Gardner-by-Floreana, Galápagos Islands. Journal of wildlife diseases, 47, 94–106.

DeWoody JA, Harder AM, Mathur S, Willoughby JR (2021) The long-standing significance of genetic diversity in conservation. Molecular Ecology, 30, 4147–4154.

Dirzo R, Young HS, Galetti M et al. (2014) Defaunation in the Anthropocene. Science, 345, 401–406.

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

Doytchinova I, Flower DR (2005) In silico identification of supertypes for class II MHCs. Journal of immunology, 174, 7085–7095.

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

Eimes J a, Reed KM, Mendoza KM et al. (2013) Greater prairie chickens have a compact MHC-B with a single class IA locus. Immunogenetics, 65, 133–44.

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

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity, 107, 1–15.

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, 1439–51.

Ellegren H, Galtier N (2016) Determinants of genetic diversity. Nature Reviews Genetics, 17, 422–433.

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theoretical Population Biology, 3, 87–112.

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. Nature reviews. Genetics, 8, 610–8.

Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee HG (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature, 351, 290–6.

Ferrer-Admetlla A, Bosch E, Sikora M et al. (2008) Balancing Selection Is the Main Force Shaping the Evolution of Innate immune genes. The Journal of Immunology, 181, 1315–1322.

Foulley JL, Ollivier L (2006) Estimating allelic richness and its diversity. Livestock Science, 101, 150–158.

Frankel OH (1974) Genetic conservation: our evolutionary responsibility. Genetics, 78, 53–65.

Frankham R (1997) Do island populations have less genetic variation than mainland populations? Heredity, 78, 311–327.

Frankham R, Ballou J, Briscoe DA (2010) Introduction to Conservation Genetics, 2nd Edition. Cambridge University Press, Cambridge.

Frankham R, Bradshaw CJA, Brook BW (2014) Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation, 170, 56–63.

Frankham R, Lees K, Montgomery ME et al. (1999) Do population size bottlenecks reduce evolutionary potential? Animal Conservation, 2, 255–260.

Galbreath R, Brown D (2004) The tale of the lighthouse-keeper's cat: Discovery and extinction of the Stephens Island wren (*Traversia Iyali*). Notornis, 51, 193–200.

García-Dorado A (2012) Understanding and predicting the fitness decline of shrunk populations: Inbreeding, purging, mutation, and standard selection. Genetics, 190, 1461–1476.

Gavan MK, Oliver MK, Douglas A, Piertney SB (2015) Gene dynamics of toll-like receptor 4 through a population bottleneck in an insular population of water voles (*Arvicola amphibius*). Conservation Genetics, 16, 1181–1193.

Gillespie JH (1991) The causes of molecular evolution. Oxford University Press, New York.

Gilpin ME, Soule ME (1989) Minimum Viable Populations: Processes of Species Extinction (ME Soule, Ed,). Sinauer, Sunderland.

Gilroy DL, Van Oosterhout C, Komdeur J, Richardson DS (2017) Toll-like receptor variation in the bottlenecked population of the endangered Seychelles warbler. Animal Conservation, 20, 235–250.

Gonzalez-Quevedo C, Spurgin LG, Illera JC, Richardson DS (2015) Drift, not selection, shapes toll-like receptor variation among oceanic island populations. Molecular Ecology, 24, 5852–5863.

Grant P, Curry R, Grant B (2000) A remnant population of the Floreana mockingbird on Champion island, Galápagos. Biological Conservation, 92, 285–290.

Grant PR, Grant N (1979) Breeding and feeding of Galápagos Mockingbirds, *Nesomimus parvulus*. The Auk, 96, 723–736.

Graur D, Li WH (2000) Fundamentals of molecular evolution 2nd edition. Sinauer Associates, Sunderland, MA.

Grueber CE, Knafler GJ, King TM et al. (2015) Toll-like receptor diversity in 10 threatened bird species: relationship with microsatellite heterozygosity. Conservation Genetics, 16, 595–611.

Grueber CE, Wallis GP, Jamieson IG (2013) Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. Molecular ecology, 22, 4470–82.

Grueber CE, Wallis GP, Jamieson IG (2014) Episodic positive selection in the evolution of avian toll-like receptor innate immune genes. PloS one, 9, e89632.

Harris H (1966) Enzyme Polymorphisms in Man. Proceedings of the Royal Society of London. Series B, Biological Sciences, 164, 298–310.

Hartmann S a, Schaefer HM, Segelbacher G (2014) Genetic depletion at adaptive but not neutral loci in an endangered bird species. Molecular Ecology, 23, 5712–5725.

Havlíček J, Roberts SC (2009) MHC-correlated mate choice in humans: A review. Psychoneuroendocrinology, 34, 497–512.

Hawley DM, Fleischer RC (2012) Contrasting epidemic histories reveal pathogen-mediated balancing selection on class II MHC diversity in a wild songbird. PloS one, 7, e30222.

Hedrick PW (1994) Purging inbreeding depression and the probability of extinction: Full-sib mating. Heredity, 73, 363–372.

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

Hedrick PW, Garcia-Dorado A (2016) Understanding Inbreeding Depression, Purging, and Genetic Rescue. Trends in Ecology and Evolution, 31, 940–952.

Hoban S, Archer FI, Bertola LD et al. (2022) Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables

(EBVs) for genetic composition. Biological Reviews, 97, 1511–1538.

Hoeck PEA, Beaumont MA, James KE et al. (2010a) Saving Darwin's muse: evolutionary genetics for the recovery of the Floreana mockingbird. Biology letters, 6, 212–215.

Hoeck PEA, Bollmer JL, Parker PG, Keller LF (2010b) Differentiation with drift: a spatio-temporal genetic analysis of Galápagos mockingbird populations (*Mimus* spp.). Philosophical Transactions of the Royal Society B: Biological Sciences, 365, 1127–1138.

Hoeck PEA, Keller LF (2012) Inbreeding, immune defence and ectoparasite load in different mockingbird populations and species in the Galápagos Islands. Journal of Avian Biology, 43, 423–434.

Höglund J (2009) Evolutionary conservation genetics. Oxford University Press, Oxford.

Hohenlohe PA, Funk WC, Rajora OP (2021) Population genomics for wildlife conservation and management. Molecular Ecology, 30, 62–82.

Huber CD, Durvasula A, Hancock AM, Lohmueller KE (2018) Gene expression drives the evolution of dominance. Nature Communications, 9, 2750.

Janicke T, Häderer IK, Lajeunesse MJ, Anthes N (2016) Evolutionary Biology: Darwinian sex roles confirmed across the animal kingdom. Science Advances, 2, 1–10.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers a O (2012) The global diversity of birds in space and time. Nature, 491, 444–8.

Jiménez-Uzcátegui G, Llerena W, Bryan Milstead W, Lomas EE, Wiedenfield DA (2011) Is the population of Floreana Mockingbird *Mimus trifasciatus* declining? Cotinga, 33, 34–40.

Jordan WC, Bruford MW (1998) New perspectives on mate choice and the MHC. Heredity, 81, 239–245.

Kasahara M (2000) Major Histocompatibility Complex (M Kasahara, Ed,). Springer Japan, Tokyo.

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends in Ecology and Evolution, 17, 230–241.

Kennedy ES, Grueber CE, Duncan RP, Jamieson IG (2014) Severe inbreeding depression and no evidence of purging in an extremely inbred wild species-the chatham island black robin. Evolution, 68, 987–995.

Kimura M (1962) On the probability of fixation of mutant genes in a population. Genetics, 47, 713–719.

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Kinnaird MF, Grant PR (1982) Cooperative breeding by the Galápagos mockingbird, *Nesomimus parvulus*. Behavioral Ecology and Sociobiology, 10, 65–73.

Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. Annual review of genetics, 41, 281–304.

Kleinman-Ruiz D, Lucena-Perez M, Villanueva B et al. (2022) Purging of deleterious burden in the endangered Iberian lynx. Proceedings of the National Academy of Sciences of the United States of America, 119, e2110614119.

Knutie SA, Owen JP, Mcnew SM et al. (2016) Galápagos mockingbirds tolerate introduced parasites that affect Darwin's finches. Ecology, 97, 940–950.

Králová T, Albrecht T, Bryja J et al. (2018) Signatures of diversifying selection and convergence acting on passerine Toll-like receptor 4 in an evolutionary context. Molecular Ecology, 27, 2871–2883.

Kyriazis CC, Wayne RK, Lohmueller KE (2019) High genetic diversity can contribute to extinction in small populations. bioRxiv, 678524.

Lande R (1995) Mutation and Conservation. Conservation Biology, 9, 782–791.

Lande R, Engen S, SÆther B-E (2003) Demographic and environmental stochasticity. In: Stochastic Population Dynamics in Ecology and Conservation (eds Lande R, Engen S, Saether BE), pp. 1–24. Oxford University Press, Oxford.

Lara CE, Grueber CE, Holtmann B et al. (2020) Assessment of the dunnocks' introduction to New Zealand using innate immune-gene diversity. Evolutionary Ecology, 34, 803–820.

Leigh DM, Hendry AP, Vázquez-Domínguez E, Friesen VL (2019) Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. Evolutionary Applications, 12, 1505–1512.

Levy H, Fiddaman SR, Vianna JA et al. (2020) Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird. Molecular biology and evolution, 37, 1708–1726.

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.

Li H, Xiang-Yu J, Dai G et al. (2016) Large numbers of vertebrates began rapid population decline in the late 19th century. Proceedings of the National Academy of Sciences of the United States of America, 113, 14079–14084.

Lighten J, Papadopulos AST, Mohammed RS et al. (2017) Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. Nature Communications, 8, 1294.

Loiseau C, Zoorob R, Garnier S et al. (2008) Antagonistic effects of a Mhc class I allele on malaria-infected house sparrows. Ecology letters, 11, 258–65.

Lovette IJ, Arbogast BS, Curry RL et al. (2012) Phylogenetic relationships of the mockingbirds and thrashers (Aves: Mimidae). Molecular phylogenetics and evolution, 63, 219–29.

Luque GM, Vayssade C, Facon B et al. (2016) The genetic Allee effect: a unified framework for the genetics and demography of small populations. Ecosphere, 7.

Lynch M, Conery J, Burger R (1995) Mutational meltdowns in sexual populations. Evolution, 49, 1067–1080.

Lynch M, Gabriel W (1990) Mutation Load and the Survival of Small Populations. Evolution, 44, 1725.

Martin MP, Carrington M (2005) Immunogenetics of viral infections. Current opinion in immunology, 17, 510–6.

Martin LB, Coon C a C, Liebl AL, Schrey AW (2013) Surveillance for microbes and range expansion in house sparrows. Proceedings of the Royal Society B: Biological Sciences, 281, 20132690–20132690.

Masel J (2011) Genetic drift. Current Biology, 21, R837–R838.

Maxwell SL, Fuller RA, Brooks TM, Watson JEM (2016) Biodiversity: The ravages of guns, nets and bulldozers. Nature, 536, 143–145.

McCallum H, Dobson A (1995) Detecting disease and parasite threats to endangered species and ecosystems. Trends in Ecology & Evolution, 10, 190–194.

McNeely J, Miller KR, Reid WV, Mittermeier RA, Werner TB (1990) Conserving the world’s biological diversity. IUCN, Washington, DC.

Mendel G (1865) Versuche über Pflanzen-Hybriden. Verhandlungen des naturforschenden Vereines in Brünn, 4, 3–57.

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

Miller HC, Lambert DM (2004) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). Molecular ecology, 13, 3709–21.

Minias P, Vinkler M (2022) Selection Balancing at Innate Immune Genes: Adaptive Polymorphism Maintenance in Toll-Like Receptors. Molecular Biology and Evolution, 39, 1–9.

Morris KM, Wright B, Grueber CE, Hogg C, Belov K (2015) Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil (*Sarcophilus harrisii*). Molecular Ecology, 24, 3860–3872.

Mukherjee S, Sarkar-Roy N, Wagener DK, Majumder PP (2009) Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. Proceedings of the National Academy of Sciences of the United States of America, 106, 7073–7078.

Muller H (1950) Our load of mutations. The American Journal of Human Genetics, 2, 111–176.

Munguia-Vega A, Esquer-Garrigos Y, Rojas-Bracho L et al. (2007) Genetic drift vs. natural selection in a long-term small isolated population: major histocompatibility complex class II variation in the Gulf of California endemic porpoise (*Phocoena sinus*). Molecular ecology, 16, 4051–65.

Murphy KP, Janeway C (2008) Janeway's Immunobiology. Garland Science, London.

Narasimhan V, Danecek P, Scally A et al. (2016) BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics, 32, 1749–1751.

Nei M (2005) Bottlenecks, genetic polymorphism and speciation. Genetics, 170, 1–4.

Nei M (2014) Mutation-Driven Evolution. Oxford University Press, Oxford.

Nelson-Flower MJ, Germain RR, MacDougall-Shackleton EA, Taylor SS, Arcese P (2018) Purifying Selection in the Toll-Like Receptors of Song Sparrows *Melospiza melodia*. Journal of Heredity, 109, 501–509.

Nielsen R (2005) Molecular signatures of natural selection. Annual Review of Genetics, 39, 197–218.

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. PLoS ONE, 7, e37558.

Nietlisbach P, Wandeler P, Parker PG et al. (2013) Hybrid ancestry of an island subspecies of Galápagos mockingbird explains discordant gene trees. Molecular Phylogenetics and Evolution, 69, 581–592.

O'Connor EA, Westerdahl H, Burri R, Edwards S V (2019) Avian MHC Evolution in the Era of Genomics: Phase 1.0. Cells, 8, 1152.

Ohta T (1992) The Nearly Neutral Theory Of Molecular Evolution. Annual Review of Ecology and Systematics, 23, 263–286.

Oliver MK, Piertney SB (2012) Selection maintains MHC diversity through a natural population bottleneck. Molecular biology and evolution, 29, 1713–20.

Van Oosterhout C, Joyce D, Cummings S (2006) Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). Evolution, 60, 2562–2574.

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

Van Oosterhout C (2020) Mutation load is the spectre of species conservation. Nature Ecology & Evolution, 4, 1004–1006.

Oppelt C, Starkloff A, Rausch P, Von Holst D, Rödel HG (2010) Major histocompatibility complex variation and age-specific endoparasite load in subadult European rabbits. Molecular ecology, 19, 4155–67.

Orlando L, Librado P (2019) Origin and Evolution of Deleterious Mutations in Horses. Genes, 10, 649.

Padmanabhan R, Jay E, Wu R (1974) Chemical synthesis of a primer and its use in the sequence analysis of the lysozyme gene of bacteriophage T4. Proceedings of the National Academy of Sciences of the United States of America, 71, 2510–2514.

Pečnerová P, Díez-del-Molino D, Vartanyan S, Dalén L (2016) Changes in variation at the MHC class II DQA locus during the final demise of the woolly mammoth. Scientific Reports, 6, 25274.

Pekkala N, Knott KE, Kotiaho JS, Puurtinen M (2012) Inbreeding rate modifies the dynamics of genetic load in small populations. Ecology and Evolution, 2, 1791–1804.

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

Provine WB (2001) The origins of theoretical population genetics: with a new afterword. University of Chicago Press, Chicago.

Radwan J, Babik W, Kaufman J, Lenz TL, Winternitz J (2020) Advances in the Evolutionary Understanding of MHC Polymorphism. Trends in Genetics, 36, 298–311.

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

Radwan J, Kawałko a, Wójcik JM, Babik W (2007) MHC-DRB3 variation in a free-living population of the European bison, *Bison bonasus*. Molecular ecology, 16, 531–40.

Raven N, Lisovski S, Klaassen M et al. (2017) Purifying selection and concerted evolution of RNA-sensing toll-like receptors in migratory waders. Infection, Genetics and Evolution, 53, 135–145.

Říčanová Š, Bryja J, Cosson JF et al. (2011) Depleted genetic variation of the European ground squirrel in Central Europe in both microsatellites and the major histocompatibility complex gene: Implications for conservation. Conservation Genetics, 12, 1115–1129.

Richman A (2000) Evolution of balanced genetic polymorphism. Molecular Ecology, 9, 1953–1963.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK (2018) Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. Current Biology, 28, 3487-3494.e4.

Robinson JA, Kyriazis CC, Nigenda-Morales SF et al. (2022) The critically endangered vaquita is not doomed to extinction by inbreeding depression. Science, 376, 635–639.

Roelke ME, Martenson JS, O'Brien SJ (1993) The consequences of demographic reduction and genetic depletion in the endangered Florida panther. Current Biology, 3, 340–350.

Sánchez-Bayo F, Wyckhuys KAG (2019) Worldwide decline of the entomofauna: A review of its drivers. Biological Conservation, 232, 8–27.

Schlaepfer DR, Braschler B, Rusterholz H-P, Baur B (2018) Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere, 9, e02488.

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

Scott JM, Conant S, van Riper III C (2001) Evolution, ecology, conservation and management of Hawaiian birds: a vanishing avifauna. Allen Press, Inc, Lawrence.

Selvatti AP, Gonzaga LP, Russo CA de M (2015) A Paleogene origin for crown passerines and the diversification of the Oscines in the New World. Molecular Phylogenetics and Evolution, 88, 1–15.

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.

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

Sherry ST, Ward M, Sirotkin K (1999) dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome research, 9, 677–679.

Siddle H V, Kreiss A, Eldridge MDB et al. (2007) Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. Proceedings of the National Academy of Sciences of the United States of America, 104, 16221–6.

Slade R, McCallum H (1992) Overdominant vs. frequency-dependent selection at MHC loci. Genetics, 132, 861–862.

Slatkin M (1987) Gene Flow and the Geographic Structure of Natural Populations. Science, 236, 787–792.

Smith KF, Acevedo-Whitehouse K, Pedersen a. B (2009) The role of infectious diseases in biological conservation. Animal Conservation, 12, 1–12.

Smith KF, Sax DF, Lafferty KD (2006) Evidence for the role of infectious disease in species extinction and endangerment. Conservation biology : the journal of the Society for Conservation Biology, 20, 1349–57.

Snell GD (1968) The H-2 locus of the mouse: observations and speculations concerning its comparative genetics and its polymorphism. Folia biologica, 14, 335–58.

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

Sommer S, Tichy H (1999) Major histocompatibility complex (MHC) class II polymorphism and paternity in the monogamous *Hypogeomys antimena*, the endangered, largest endemic Malagasy rodent. Molecular ecology, 8, 1259–72.

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. Proceedings of the National Academy of Sciences of the United States of America, 101, 15261–15264.

Spurgin LG, Oosterhout C Van, Illera JC et al. (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. Molecular Ecology, 20, 5213–25.

Štefka J, Hoeck PE, Keller LF, Smith VS (2011) A hitchhikers guide to the Galápagos: Co-phylogeography of Galápagos mockingbirds and their parasites. BMC Evolutionary Biology, 11, 284.

Steinheimer FD (2004) Charles Darwin’s bird collection and ornithological knowledge during the voyage of H.M.S. “Beagle”, 1831-1836. Journal of Ornithology, 145, 300–320.

Stephens PA, Sutherland WJ, Freckleton RP (1999) What Is the Allee Effect? Oikos, 87, 185.

Swarth HS (1931) The avifauna of the Galápagos Islands, by Harry S. Swarth. California Academy of Sciences San Francisco.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123, 585–595.

Tang Y, Liu G, Zhao S et al. (2022) Major Histocompatibility Complex (MHC) Diversity of the Reintroduction Populations of Endangered Przewalski's Horse. Genes, 13.

Teixeira JC, Huber CD (2021) The inflated significance of neutral genetic diversity in conservation genetics. Proceedings of the National Academy of Sciences of the United States of America, 118, 1–10.

Těšický M, Velová H, Novotný M et al. (2020) Positive selection and convergent evolution shape molecular phenotypic traits of innate immunity receptors in tits (Paridae). Molecular Ecology, 29, 3056–3070.

Tschirren B, Andersson M, Scherman K et al. (2013) Polymorphisms at the innate immune receptor TLR2 are associated with *Borrelia* infection in a wild rodent population. Proc Biol Sci, 280, 20130364.

van der Valk T, de Manuel M, Marquez-bonet T, Guschanski K (2019) Estimates of genetic load suggest extensive genetic purging in mammalian populations. bioRxiv, 696831.

Vinkler M, Bainová H, Bryjová A et al. (2015) Characterisation of Toll-like receptors 4, 5 and 7 and their genetic variation in the grey partridge. Genetica, 143, 101–112.

Wakeland EK, Boehme S, She JX et al. (1990) Ancestral polymorphisms of MHC class II genes: Divergent allele advantage. Immunologic Research, 9, 115–122.

Wang J, Zhang Z, Liu J, Zhao J, Yin D (2016) Ectodomain Architecture Affects Sequence and Functional Evolution of Vertebrate Toll-like Receptors. Scientific Reports, 6, 26705.

Westbury M V., Hartmann S, Barlow A et al. (2018) Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the brown Hyena. Molecular Biology and Evolution, 35, 1225–1237.

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.

Westerdahl H, Stjernman M, Råberg L, Lannefors M, Nilsson J-Å (2013) MHC-I Affects Infection Intensity but Not Infection Status with a Frequent Avian Malaria Parasite in Blue Tits. PloS one, 8, e72647.

Willi Y, Van Buskirk J, Fischer M (2005) A threefold genetic allee effect: Population size affects cross-compatibility, inbreeding depression and drift load in the self-incompatible *Ranunculus reptans*. Genetics, 169, 2255–2265.

Willi Y, Van Buskirk J, Hoffmann AA (2006) Limits to the Adaptive Potential of Small Populations. Annual Review of Ecology, Evolution, and Systematics, 37, 433–458.

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

Woolfit M, Bromham L (2005) Population size and molecular evolution on islands. Proceedings of the Royal Society B: Biological Sciences, 272, 2277–2282.

Worley K, Collet J, Spurgin LG et al. (2010) MHC heterozygosity and survival in red junglefowl. Molecular Ecology, 19, 3064–75.

Wright S (1931) Evolution in mendelian populations. Genetics, 16, 97–159.

Xue Y, Prado-Martinez J, Sudmant PH et al. (2015) Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science, 348, 242–245.

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.

Zhang P, Huang K, Zhang B et al. (2018) High polymorphism in MHC-DRB genes in golden snub-nosed monkeys reveals balancing selection in small, isolated populations. BMC Evolutionary Biology, 18, 1–13.

Zhang L, Liu R, Ma L et al. (2012) *Eimeria tenella*: Expression profiling of toll-like receptors and associated cytokines in the cecum of infected day-old and three-week old SPF chickens. Experimental Parasitology, 130, 442–448.

Chapter I

Balancing selection and genetic drift create unusual patterns of MHCIIβ variation in Galápagos mockingbirds

Molecular Ecology (2016) 25, 4757–4772

DOI 10.1111/mec.13807

Chapter I

Balancing selection and genetic drift create unusual patterns of MHCIIβ variation in Galápagos mockingbirds

Vlček Jakub[1,2], Hoeck Paquita E A[3], Keller Lukas F[4], Wayhart Jessica[3], Dolinová Iva[5] & Štefka Jan[1,2,*]

1 Faculty of Science, University of South Bohemia, Branišovská 1670, 37005, České Budějovice, Czech Republic.

2 Biology Centre CAS, Branišovská 31, 37005, České Budějovice, Czech Republic

3 Institute for Conservation Research, San Diego Zoo Global, 15600 San Pasqual Valley Road, Escondido, CA 92027, USA

4 Department of Evolutionary Biology and Environmental Studies,University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

5 Institute for Nanomaterials, Advanced Technologies and Innovation, Technical University of Liberec, Bendlova 7, 46001, Liberec, Czech Republic

* corresponding author

Abstract

The extracellular subunit of the major histocompatibility complex MHCIIβ plays an important role in the recognition of pathogens and the initiation of the adaptive immune response of vertebrates. It is widely accepted that pathogen-mediated selection in combination with neutral micro-evolutionary forces (e.g. genetic drift) shape the diversity of MHCIIβ, but it has proved difficult to determine the relative effects of these forces. We evaluated the effect of genetic drift and balancing selection on MHCIIβ diversity in 12 small populations of Galápagos mockingbirds belonging to four different species, and one larger population of the Northern mockingbird from the continental USA. After genotyping MHCIIβ loci by high-throughput sequencing,we applied a correlational approach to explore the relationships between MHCIIβ diversity and population size by proxy of island size. As expected when drift predominates, we found a positive effect of population size on the number of MHCIIβ alleles present in a population. However, the number of MHCIIβ alleles per individual and number of supertypes were not correlated with population size. This discrepancy points to an interesting feature of MHCIIβ diversity dynamics: some levels of diversity might be shaped by genetic drift while others are independent and possibly maintained by balancing selection.

Keywords

genetic diversity, major histocompatibility complex, *Mimus*, population size, trans-species polymorphism

Introduction

The major histocompatibility complex comprises a family of genes that play a crucial role in the adaptive immune system. MHC class II encodes glycoproteins that bind and present antigens to the helper T cells that eventually trigger an adaptive immune response (Neefjes et al. 2011). The MHCIIβ subunit is part of the trans-membrane MHC II complex that is involved particularly in binding and presenting exogenous antigens. The number of MHCIIβ variants coding for different glycoproteins in an individual defines the range of antigens (pathogens) that the immune system can recognize and fight off (Spurgin & Richardson 2010). Consequently, the MHCIIβ is one of the most variable loci found in jawed vertebrates (e.g. 1829 alleles in the human MHC II DRB gene; de Bakker & Raychaudhuri 2012), and its polymorphism is considered an adaptive trait maintained by balancing selection (Oliver & Piertney 2012; Tobler et al. 2014).

Pathogen-mediated balancing selection (PMBS) is a main force shaping the MHC polymorphism as a consequence of an arms race between pathogens and the host's immune system (Spurgin & Richardson 2010). Several types of PMBS that act upon MHC have been recognized: heterozygote advantage (Doherty & Zinkernagel 1975; Hedrick 2012), rare allele advantage (Slade & McCallum 1992; Sepil et al. 2013), divergent allele advantage (Wakeland et al. 1990; Lenz et al. 2013) and fluctuating selection (Hill 1991; Ekblom et al. 2007). PMBS can also lead to trans-species polymorphism when the same advantageous MHC alleles are conserved across distinct evolutionary units in spite of differentiating evolutionary processes (Klein et al. 2007). Furthermore, a growing body of evidence shows that PMBS can be better detected in diversity of MHC supertypes, clusters of MHC variants grouped by functional properties, rather than in diversity of individual alleles (Trachtenberg et al. 2003; Sepil et al. 2013). However, the extent of PMBS' ability to maintain

functional variation in wild populations is not yet clearly understood, as MHC variation can be simultaneously affected by various evolutionary forces that conceal the direct effect of PMBS (sexual selection: Bonneaud et al. 2006; non-pathogen-mediated selection: Van Oosterhout 2009; random genetic drift: Sutton et al. 2011; gene conversion: Spurgin et al. 2011).

Particularly, with decreasing population size genetic drift becomes a very important force. In populations with small effective population size (N*e*), genetic drift has been shown to reduce MHC allelic diversity (Sutton et al. 2011, 2015), thus counteracting the effect of PMBS (Robertson 1962; Kimura 1983; Miller & Lambert 2004a). In several very small populations, MHC diversity was found to be even more reduced than neutral diversity (Sutton et al. 2011; Strand et al. 2012). This pattern has been explained by the fixation of the same alleles across multiple MHC loci (drift across loci; Eimes et al. 2011), or by the effects of PMBS itself when PMBS distorts allele frequency and the alleles are fixed/lost with higher frequency than alleles in neutral loci (Ejsmond & Radwan 2011). Drift was shown to be a predominant force in several studies that investigated the dynamics between genetic drift and PMBS in populations that experienced a recent bottleneck event (Miller & Lambert 2004b; Eimes et al. 2011; Strand et al. 2012; Grueber et al. 2013, 2015; Sutton et al. 2015; Gonzalez-Quevedo et al. 2015). On the other hand, several studies have shown that PMBS may counteract the effect of random genetic drift and maintain MHC diversity in spite of small population size (Aguilar et al. 2004; Van Oosterhout et al. 2006) or a recent bottleneck (Oliver & Piertney 2012). This discrepancy in the findings of these studies might be caused by various idiosyncratic factors (e.g. different diversity measures used or different demographic status of the populations).

Our aim was to analyse the effect of PMBS on various MHC diversity indices while controlling for the demographic status of the populations in

our study system. This setting can provide generally interpretable results in the discussion about the strength of PMBS vs. genetic drift in MHC diversity in natural populations.

Galápagos Mockingbirds (hereafter referred to as GM) comprise four species of the genus *Mimus* showing an allopatric distribution in the Galápagos archipelago (Figure 1). GM form a monophyletic lineage within mockingbirds and thrashers (Lovette et al. 2012), which implies a single colonization event of the archipelago. Based on the mutation rate of mtDNA, it was estimated that the colonization took place between 1.6 and 5.5 Ma (Arbogast et al. 2006). Following colonization, GM evolved into four species: the San Cristóbal mockingbird (*Mimus melanotis*), the Española mockingbird (*Mimus macdonaldi*), the Floreana mockingbird (*Mimus trifasciatus*) and the Galápagos mockingbird (*Mimus parvulus*; Nietlisbach et al. 2013) separated into 17 populations. Because of low dispersal, the majority of the populations is effectively isolated and gene flow between the islands is restricted (Hoeck et al. 2010b; Stefka et al. 2011) with the exception of extremely proximate islands (Gardner by Española and Española, and to a lesser extent also Rabida and Santiago, and Isabela and Fernandina). Microsatellite data showed that estimates of N*e* vary greatly between different populations (from 43 on Champion to 1590 on Isabela, Table 1) and that N*e* is strongly positively correlated with island size (Hoeck et al. 2010b). Therefore, island size is a suitable proxy for population size. Moreover, based on an analysis of neutral genetic diversity in historical and contemporary samples, Hoeck et al. (2010b) concluded that populations have been in mutation–drift equilibrium for at least 100 years. This finding does not directly imply that MHC is in mutation–drift equilibrium as well, due to the different type of mutational processes involved. However, the information obtained from different neutral diversity measures suggests that the effective population size in the majority of GM populations has been stable for at least 100 years, which is an important factor in the dynamics of the drift vs. PMBS effect.

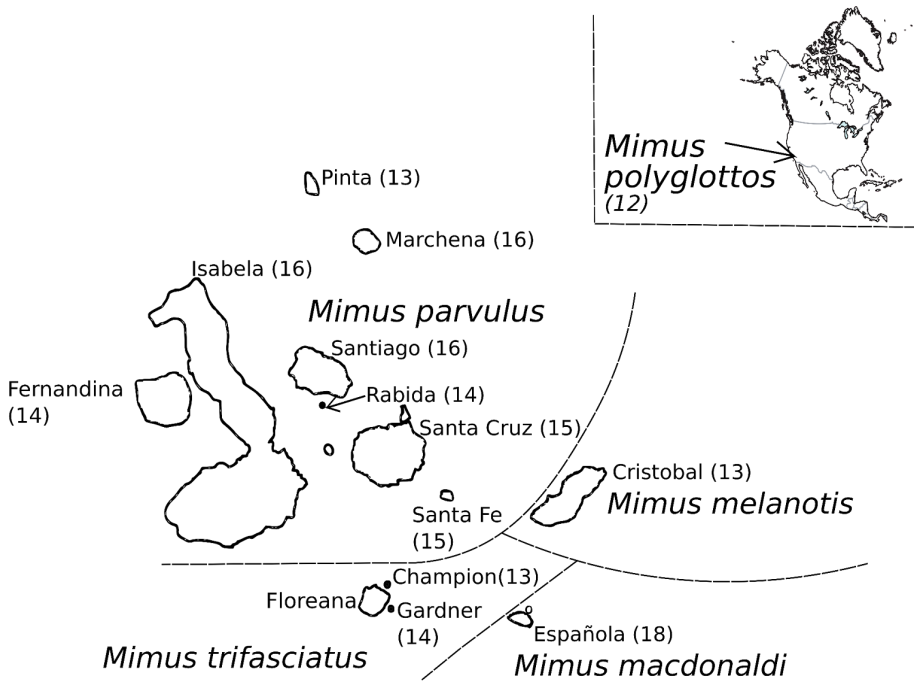


Figure 1 Distribution of populations and species of Galápagos Mockingbirds with numbers of sampled individuals in parentheses. The number of sampled individuals of *Mimus polyglottos* with the approximate collection site is shown in the inserted box.

In this study, we aimed to test the effect of drift and balancing selection on MHCIIβ diversity in GM. We investigated MHCIIβ diversity in 12 distinct populations of GM, covering the four different GM species, and compared observed patterns with one population of their closely related

continental species, the Northern Mockingbird (*Mimus polyglottos*) [time of divergence: 5.7 Ma (Jetz et al. 2012)]. We focused our analyses on the population rather than the species level, because populations represent natural evolutionary units that reflect differences in the distribution of genetic diversity. We predicted: (i) higher MHCIIβ diversity in *M. polyglottos* compared to GM due to weaker genetic drift and stronger selection pressure on the continental population, where the population size is larger and pathogen diversity is higher (Wikelski et al. 2004; Hellgren et al. 2011), (ii) a positive relationship between MHCIIβ diversity and population size, (iii) a greater decrease in individual MHCIIβ diversity than in microsatellite diversity with decreasing population size, because of the fixation of the same MHCIIβ allele across multiple loci (drift across loci hypothesis; Eimes et al. 2011). Furthermore, we inferred the effect and nature of PMBS based on an analysis of several molecular patterns: (i) the rate of accumulation of nonsynonymous and synonymous mutations (dN/dS) in MHCIIβ locus was used to evaluate whether historical balancing selection was present in the system (Garrigan & Hedrick 2003), (ii) sharing of MHC alleles between populations was analysed to detect the presence of trans-species and transpopulation evolution (Klein et al. 2007), and (iii) the relationship between the genetic distance and population size was used to test divergent allele advantage hypothesis – the case when small populations have more divergent MHC alleles than expected due to PMBS (Wakeland et al. 1990; Agudo et al. 2011).

Table 1 Population genetic diversity and other population parameters

Population (sp.)	Isl.Size (Ha)	Ne	№ind.	Ai	Ai_sd	Ap	Priv._Ap	Di	Di_sd	AR	HS	HS_sd	super_No
Champion (M.tri)	9.50	43.39	13	4.23	1.36	13	0	0.154	0.029	1.161	0.065	0.029	5
Gardner-by-Floreana (M.tri)	81.00	132.74	14	7.36	1.15	17	5	0.194	0.017	1.989	0.288	0.081	6
Rabida (M.par)	499.30	238.76	14	6.00	2.88	29	3	0.185	0.017	2.966	0.541	0.057	6
StFe (M.par)	2413.00	513.13	15	3.20	2.01	14	1	0.261	0.052	2.021	0.282	0.054	5
Pinta (M.par)	5940.00	NA	13	5.77	2.09	22	2	0.167	0.049	2.425	0.364	0.109	5
Española (M.mac)	6048.00	424.69	18	6.33	2.06	24	4	0.135	0.030	1.901	0.254	0.072	5
Marchena (M.par)	12996.00	808.69	16	5.06	1.34	17	2	0.166	0.047	2.808	0.458	0.092	4
SanCristobal (M.mel)	55808.60	1206.46	13	2.92	1.32	14	3	0.125	0.041	2.685	0.378	0.078	5
Santiago (M.par)	58465.00	1363.40	16	6.69	1.99	40	10	0.186	0.029	3.923	0.553	0.059	6
Fernandina (M.par)	64248.00	NA	14	7.71	1.86	28	5	0.189	0.025	3.394	0.445	0.094	6
StCruz (M.par)	98555.00	1116.88	15	5.80	1.32	35	7	0.154	0.029	3.829	0.605	0.119	6
Isabela (M.par)	458812.00	1590.86	16	6.81	2.51	48	10	0.177	0.037	3.430	0.517	0.094	6
Continent (M.pol)	NA	NA	12	6.33	2.71	32	30	0.273	0.017	5.827	0.651	0.092	6

Isl. Size (Ha), size of island in hectares; Ne, effective population size estimate adopted from Hoeck et al. (2010b); No ind., number of sampled individuals; Ai, number of MHCIIß alleles per individual averaged per population; Ai_sd, standard deviation of Ai; Ap, total number of true MHCIIß alleles per population; Priv._Ap, number of private MHCIIß alleles; Di, estimate of distance of alleles within an individual averaged per population; Di_sd, standard deviation of Di; AR, microsatellite allelic richness; HS, observed microsatellite heterozygosity averaged per population; HS_sd, standard deviation of HS; super_No, number of supertypes per population.

Materials and methods

Samples and microsatellite data

Samples used in this study (189 individuals) comprise a subset of samples (177) from GM used in Hoeck et al. (2010b) and 12 individuals from *Mimus polyglottos*. DNA was obtained from blood samples collected on 12 Galápagos Islands (Figure 1) between 2003 and 2008 covering the range of the four species (Hoeck et al. 2009, 2010b). In each population, the samples were collected within a maximum time span of 2 years, with the exception of Isabela where five samples were collected in 2003 and 11 in 2008. All samples of *M. polyglottos* were collected in 2013 from a population near San Diego, California. DNA extraction was performed as described previously (Hoeck et al. 2009). All individuals from Galápagos and the mainland were genotyped at 26 microsatellite loci following previously established protocols (Hoeck & Keller 2012), using primers that were designed based on four individuals from three GM populations (Hoeck et al. 2009). For 12 of the islands, the microsatellite data were adopted from Hoeck & Keller (2012). New microsatellite genotypes for this study were generated for the population of *M. parvulus* from Pinta and for *M. polyglottos* using the same primers as in Hoeck & Keller (2012). Microsatellite genotypes are available in Table S1 (Supporting information).

MHC primer design and testing

To design reliable primers for MHCII exon 2β (MHCIIβ) in GM, we first amplified the exon and surrounding introns with primers 1a and 2a designed for Little Greenbul (*Andropadus virens*; Aguilar & Edwards 2006). The resulting products from 12 different populations of GM (1 individual per population) were cloned into *Escherichia coli* plasmids using PGEM-T Easy Vector Systems (Promega). A reliable sequence was

obtained only for *M. melanotis*. This sequence was aligned with homologous songbird MHCIIβ fragments (*Ficedula albicollis*, *Erithacus rubecula*, *Turdus iliacus* and *Luscinia svecica*) in GENEIOUS 7.0.6 (Biomatters Ltd.). Based on this alignment, specific primers (Moc325F/R) were designed at conserved positions, corresponding to previously known primers 325/326 (Edwards et al. 1995). The forward primer site lies +48 bp from the 5' end, and the reverse primer site 57 bp from the 3' end of the MHCIIβ exon 2. In the final step of the primer design, we verified that our primers Moc325F/R amplified products in the desired range (between 163 and 164 bp) consistently across all species and populations of GM and *M. polyglottos*.

High-throughput sequencing of MHCIIβ

Sequencing of the MHC fragment was carried out in two runs on the Ion Torrent sequencing platform (Rothberg et al. 2011). The Ion Torrent specific primers Moc325*ion were composed of a platform-specific adaptor sequence, individual tags (only in forward primers) and MHC-specific primers Moc325F/R (Document S1, Supporting information).

We selected 12–18 samples from each population, resulting in 189 unique samples in total (Table S2, Supporting information). A total of 21 samples were replicated in an independent PCR under the same conditions, except that different individual tags were used. These replicated samples were used to estimate the accuracy of the genotyping. The sample set was divided into two batches for sequencing in the two Ion Torrent runs.

In the initial PCR amplification of the amplicon library (Document S1, Supporting information), we followed the guidelines of Lenz & Becker (2008) to minimize the occurrence of artificial chimeras (e.g. usage of nonproofreading enzyme, reduced number of PCR cycles and increased elongation time). Concentration of the products was measured using

QUBIT 2.0 Fluorometer High Sensitivity ds DNA kit (Life Technologies, Grand Island, NY, USA), and amplicons were pooled equimolarly. The resulting library was purified using the AGENCOURT AMPURE system (Agencourt Bioscience Corporation, Beverly, MA, USA). Precise measurement of library size and concentration was performed on Bio analyser instrument (Agilent Technologies), and the final libraries were diluted to a required concentration of 26 pM of DNA.

The Ion Torrent sequencing was performed at the Institute for Nanomaterials at the Technical University of Liberec, Czech Republic. The emulsion PCR was prepared using the ION ONETOUCHTM 2 System (Life Technologies) with a 400-bp chemistry kit according to the manufacturer's manual. The enriched Ion Spheres were loaded onto two 318TM Ion Torrent Chips, and one directional sequencing was conducted on the Ion Torrent Personal Genome Machine (Life Technologies).

MHCIIβ allele and supertype identification

Demultiplexing of the sequence reads was performed in the ION REPORTER TM software (4.0, Life Technologies). Low-quality reads (Document S1, Supporting information) were discarded. Identical sequence reads within the amplicon (variants) were collapsed while their frequency was recorded by fastx_collapser script from the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/ index.html). Singletons were discarded as possible PCR artefacts. The investigation of read length distributions and amino acid sequences showed that reads longer or shorter than the expected 164 bp were either homopolymer errors or putative pseudogenes with frameshift mutations. Based on this finding, only the reads of 164 bp passed to the next step because we aimed to capture only functional variation.

True alleles were initially identified based on the assumption that their sequencing depth (frequency) is considerably higher than the sequencing

depth of artefacts. The threshold depth discriminating true alleles from artefacts was assessed individually for each amplicon using the degree of change (DOC) method (Lighten et al. 2014b). This approach was complemented by the chimera detection and MHC allele filtering pipeline developed by Sommer et al. (2013). All details about allele filtering are accessible in Document S1 (Supporting information). Also, we made a BLAST search using our true MHCIIβ variants against sequences in GenBank (http://www.ncbi.nlm.nih.gov/genbank/) in order to detect possible cross-contamination and to reveal the closest available species containing homologous MHCIIβ sequences.

Accuracy of the genotyping was analysed based on 20 replicated samples (one replicate was removed due to poor amplification). For each replicated sample, accuracy was calculated as the ratio between the number of alleles that showed up in both runs and the total number of alleles. The measure was averaged across all replicates.

Recognition of different antigens by MHC is a trait affected by balancing selection. Groups of different MHC alleles can recognize similar antigen motifs, and such groups are called supertypes. Therefore, balancing selection may affect functional variation via the number of supertypes rather than the number of individual alleles (Trachtenberg et al. 2003; Huchard et al. 2013; Sepil et al. 2013). To test whether selection maintained similar supertype diversity across the studied populations, we determined the supertypes according to Doytchinova & Flower (2005). Initially, we identified amino acid sites under positive selection by SLAC test (Kosakovsky Pond & Frost 2005) on the Datamonkey webserver (Delport et al. 2010). The test was conducted on alignment of all MHCIIβ variants for the whole 164-bp MHC fragment. Subsequently, the supertypes were determined based on physiochemical properties of the nine amino acid sites that were found under positive selection by SLAC. Clustering of alleles was performed by discriminant analysis of principle components of the amino acid physiochemical properties

(hydrophobicity, steric bulk, polarity and electronic effect) adopted from Sandberg et al. (1998). The analysis was performed using package adegenet 2.0.1 in R (Jombart 2008; R Core Team 2014). K-means clustering algorithm was used to identify the optimal number of clusters with ΔBIC ≤2 rule as in Lillie et al. (2015).

MHCIIβ diversity in GM and *M. polyglottos*

To ascertain the differences in MHCIIβ diversity between insular and continental populations (prediction i), we used five measures of MHCIIβ diversity: (i) the number of unique alleles per population (population allelic diversity, Ap); (ii) the number of alleles per individual (Ai); (iii) mean evolutionary distance between amino acid variants in an individual representing functional diversity of the MHCIIβ protein (Di). Di was estimated in MEGA6 (Tamura et al. 2013) as the number of amino acid differences over all sequence pairs within each individual using a p-distance model; (iv) the number of MHCIIβ supertypes per population; and (v) the number of supertypes per individual. Ai and Ap were generated by SQL query of the data set (Table S3, Supporting information). Differences in functional MHCIIβ diversity patterns were contrasted with diversity patterns of neutral microsatellites – individual heterozygosity (HS) and allelic richness on the population level (AR). HS was calculated in the program GENHET as the ratio of the number of heterozygous loci divided by the number of genotyped loci (Coulon 2010). AR was calculated using package hierfstat (Goudet 2005) in R. One way ANOVA was used to test for differences in the MHCIIβ diversity indices Ai and Di and microsatellite HS among populations. Multiple comparisons of differences between populations were conducted by Tukey's HSD test in R (R Core Team 2014).

Effects of genetic drift on MHCIIβ diversity

To test whether individual MHC diversity was affected by population size (prediction ii), we constructed a regression between the number of MHCIIβ alleles per individual (Ai) and island size (natural log-transformed, as a proxy of population size). To control for study design, we used population identity as a random factor. We performed a generalized linear model with mixed effects (GLMM) in package LME4 (Bates et al. 2015). Because Ai is a count represented by natural numbers, we used Poisson error distribution in the model. Overdispersion in this model was negligible (Document S1, Supporting information). To assess how neutral genetic diversity was affected by population size, we analysed regression between HS and population size using a linear model with population identity as a random factor. HS showed a normal distribution; therefore, a linear model with a Gaussian distribution was utilized. The slopes of the regressions of MHC Ai and microsatellite HS were compared to assess the relative role of genetic drift on functional and neutral genetic diversities. If drift across loci in MHCIIβ is present, we should observe a steeper positive slope in the regression of Ai on population size than of HS on population size (prediction iii). The regressions between population size and genetic diversities were only performed for GM populations (subset of Table S3, Supporting information), because an estimate of population size was lacking for the *M. polyglottos* population. Regression between the number of MHC supertypes and population size was analysed in the same way.

To test whether the population-level MHC diversity (Ap) was affected by population size, we performed a regression between the number of alleles found in the population (Ap) and the natural logarithm of island size in a linear model with Gaussian distribution. Individual populations were taken as data points, and therefore, the sample size was only 12. Thus, we examined this relationship bearing in mind that an absence of a correlation might be caused by a lack of statistical power. A similar linear model was constructed to test for the effect of island size on

microsatellite AR at the population level. To compare the effect of island size on AR and Ap, we constructed linear regression models between the proportional reduction in each genetic diversity measure and island size. To quantify the proportional reduction in genetic diversity for each population, we calculated a ratio between the Ap or AR value for a given population and the value found on the largest island, Isabela (Ap_{pop}/Ap_{Isa}; AR_{pop}/AR_{Isa}). Transformation of Ap and AR into the proportional reduction allowed us to directly compare regression slopes between the two linear models. All the models were constructed in R version 3.2.4 (R Core Team 2014).

Effects of balancing selection on MHCIIβ diversity

We used dN/dS test to assess whether historical balancing selection shaped MHCIIβ diversity (Garrigan & Hedrick 2003). The test was performed in the program MEGA6 (Tamura et al. 2013) by a Z-test for positive selection (H0:dN–dS ≤ 1) using the Nei–Gojobori method (Nei & Gojobori 1986) and Jukes–Cantor correction. SLAC was used as an alternative, maximum-likelihood method for detection of historical positive selection (Kosakovsky Pond & Frost 2005). Both tests were conducted on alignment of all MHCIIβ variants for the whole 164-bp MHC fragment. A codon-based approach was utilized to identify individual codons under selection. We investigated whether the rates of nonsynonymous and synonymous substitutions (dN–dS) differed between codons that are involved in binding/presenting pathogen (antigen binding sites, ABS) and codons that are not involved in the presentation (non-ABS). Positioning of the ABS in mockingbird sequences was inferred from a human MHC II molecular model (Brown et al. 1993). Also, we assessed whether the position of codons under selection in our system matched patterns found in other songbirds (Document S1, Supporting information).

Phylogenetic relationships of MHCIIβ alleles were constructed to investigate the pattern of allele clustering that would imply the presence of trans-population and trans-species evolution (Klein et al. 2007). We constructed a haplotype network of all 150 MHCIIβ alleles in the program SPLITSTREE4 (Huson & Bryant 2006) using the neighbour-Net method (Bryant & Moulton 2004) to show possible reticular relationships originating from gene duplication and recombination in MHCIIβ. Sharing of alleles found in GM populations was visualized by an haplotype network of MHCIIβ alleles constructed by TCS (Clement et al. 2000) in POPART (http://popart.otago.ac.nz).

To test whether the sharing of alleles between populations was affected by balancing selection or whether the observed pattern could be explained by the retention of ancestral polymorphism, we compared the number of shared MHC alleles with the divergence in neutral markers. If the sharing of MHC alleles were due to retention of ancestral polymorphism, then, according to the neutral theory (Kimura 1983), we would observe a decreasing number of shared MHCIIβ alleles with an increase in population divergence. To test this empirically, we analysed the correlation between the numbers of MHC alleles shared by each pair of populations and their pairwise Fst values based on microsatellite data (Hoeck et al. 2010b). We analysed the correlation of the two matrices by Mantel test using the VEGAN package in R (Oksanen et al. 2015). The number of permutations was set to 9999. We used the inverse of the MHC sharing matrix to test for a positive correlation with Fst. A matrix of pairwise comparisons of common alleles between all populations of GM was generated in R. Because one population pair shared no alleles, the matrix was incremented by one before the inversion.

In small populations, PMBS sometimes preserves more divergent MHC alleles than expected (Agudo et al. 2011). To test for this scenario, we constructed linear regression between amino acid distance (Di) and the

natural logarithm of island size. To control for dependency on intrapopulation individual variation, we used population identity as a random factor. Because Di showed a normal distribution, we constructed a linear model with mixed effects in the R package LME4.

Results

Sequencing and genotyping

Two runs of the Ion Torrent PGM sequencing produced 8.2M sequence reads from which 1.84M reads passed the first set of filtering steps (base-call quality, length, singletons). One amplicon retained no reads after these steps, and it was discarded from further analysis. The final amplicon coverage ranged from 35 to 76 910 reads, with an average of 7775 reads.

A total of 98% of amplicons showed a prominent inflexion point in the cumulative plots of sequencing depths. Therefore, we considered the DOC method suitable for selecting the threshold value between artificial and true alleles. Five amplicons were discarded either because of a low number of reads, or a smooth cumulative plot of sequencing depths. After correction for PCR bias and identification of chimeras, we obtained 150 true MHC alleles (GenBank Accession nos: KT123899– KT124105; Table S4, Supporting information) corresponding to 149 peptide variants. The MHC genotypes were obtained for 189 individuals and 20 replicates. On average, the accuracy between the replicated samples was 88.4%. The most similar MHCIIβ alleles found by blast search belonged to the Thrushes family (Turdidae; phylogenetically the closest relatives of GM with MHCIIβ sequenced, Table S5, Supporting information).

MHCIIβ diversity in GM and *Mimus polyglottos*

The number of MHCIIβ alleles per individual (Ai) ranged from one to 12 with a mean of 5.67 (median 6). Ai varied significantly among

populations (Figure 2a; ANOVA, d.f. = 12/176, P < 0.0001, F = 8.03). However, statistically significant differences were only found between three populations with highly reduced Ai: St Fe (3.2), San Cristóbal (2.92), Champion (4.23) and almost all other populations (Table S6, Supporting information). The continental population (*M. polyglottos*) with Ai = 6.3 was not statistically different from the GM populations except for St Fe and San Cristóbal.

The total number of MHC alleles (Ap) varied substantially across populations from 13 on Champion to 48 on Isabela (Figure 2b, Table 1), and it was not correlated with the number of individuals sampled per population (F = 1.065, d.f. = 11, P = 0.324, r = 0.09). *M. polyglottos* showed seven alleles more (Ap = 32) than the mean number of Ap in GM (25.08). However, three populations of GM showed higher Ap (Isabela, Santiago and St Cruz) than *M. polyglottos* (Table 1). The amino acid distance of MHCIIβ alleles per individual (Di) ranged from 0.019 to 0.315 with a mean of 0.181. Averaged Di varied significantly among populations (Figure 3; ANOVA, d.f. = 12/174, P < 0.0001, F = 21.29). Interestingly, two populations, St Fe (0.260) and *M. polyglottos* (0.273), had average Di values almost two times higher than all other populations (Table S7, Supporting information).

Clustering of 149 MHC peptide variants showed the presence of eight supertypes, with each supertype grouping five to 47 original alleles (Figure S1, Supporting information). The number of supertypes per population was within a narrow range from five to six with the exception of Marchena where only four supertypes were found (Table 1). The population of *M. polyglottos* contained six supertypes in total (with one private supertype 6) while supertype 2 was present only in Galápagos. The average number of supertypes per individual ranged considerably from 1.3 on San Cristóbal to 5.36 on Gardner by Floreana (Figure S1, Supporting information). *M. polyglottos* with 3.75 was very close to the

average number of supertypes per individual in GM (3.2) and showed no significant difference from it.

Microsatellite individual heterozygosity (HS) showed much higher variation among populations than Ai (Figure 2c; ANOVA, d.f. = 12/176, $P < 0.0001$, $F = 55.52$). *M. polyglottos* showed a higher average heterozygosity (0.651) than the average heterozygosity across GM populations (0.396), but the St Cruz population (HS = 0.605) showed no significant difference from the continental population (Table S8, Supporting information). On the other hand, microsatellite allelic richness (AR) was distinctly higher in the continental population (5.827) both compared to the average AR detected in GM populations (2.711) and to the GM population with the highest AR (3.923, Santiago; Figure 2d).

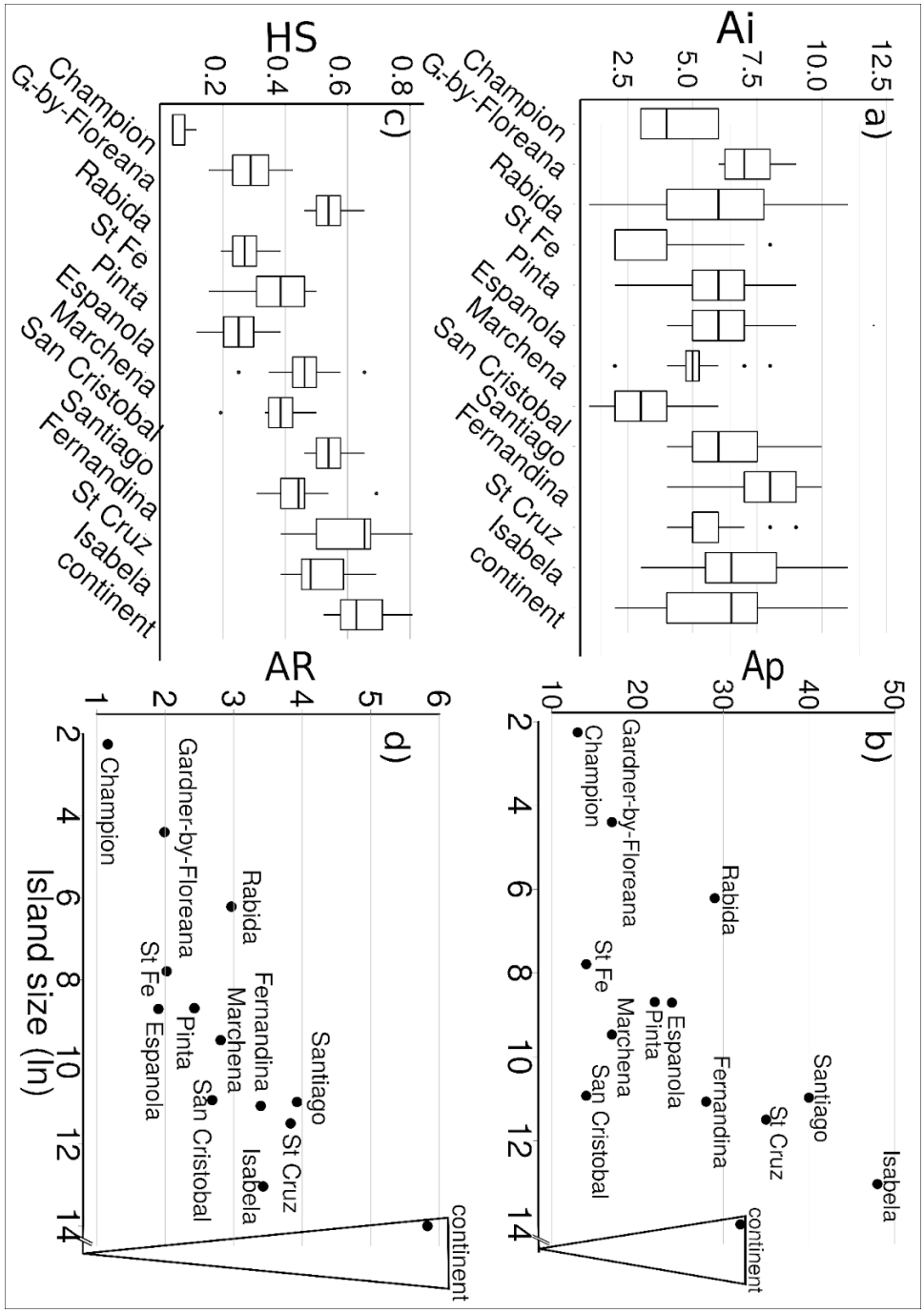


Figure 2 Genetic diversity in the populations of Galápagos mockingbirds and *Mimus polyglottos* (continent). (a) Number of MHCIIβ alleles per

individual (Ai) per population, populations are ordered from smallest to largest; (b) number of MHCIIβ alleles present in a population (Ap) with respect to natural logarithm of island size in hectares (continent not in scale); (c) observed microsatellite heterozygosity (HS) within individuals per population; (d) microsatellite allelic richness (AR) with respect to natural logarithm of island size in hectares. Box-and-whisker plots show the median (horizontal line), upper and lower quartiles (box) and maximum and minimum values (excluding outliers represented by a dot).

Population size and MHCIIβ diversity in GM

At the individual level, we found no significant relationship between the number of alleles per individual (Ai) and island size (slope = 0.015, 95% confidence interval [CI] = 0.041 to 0.071, LRT = 0.315, P = 0.575), whereas the regression of microsatellite heterozygosity (HS) on island size was significant (slope = 0.035, CI = 0.016–0.055, LRT = 9.505, P = 0.002). As the regression slope for Ai was not larger than that for HS, there was no evidence of drift across loci.

At the population level, we found a significant relationship between the number of alleles per population (Ap) and island size (slope = 2.272, CI = 0.332–4.212, R2-adj = 0.346, P = 0.026). The correlation between microsatellite allelic richness and population size was also significant (slope = 0.211, CI = 0.092–0.329, R2-adj = 0.572, P = 0.003). Interestingly, the effect of island size was not different between microsatellites and MHC (Figure 5). The slope between the proportional reduction in microsatellite AR and island size was 0.064 (CI = 0.028–0.100, P = 0.003, R2-adj = 0.572), while the slope between the proportional reduction in Ap and island size was 0.047 (CI = 0.007–0.088, P = 0.026, R2- adj = 0.346). A Z-test showed no significant difference between the slopes (Z-test = 0.686, P = 0.248). The numbers of supertypes per population or individual were not correlated with population size (P = 0.597 and P = 0.289, respectively).

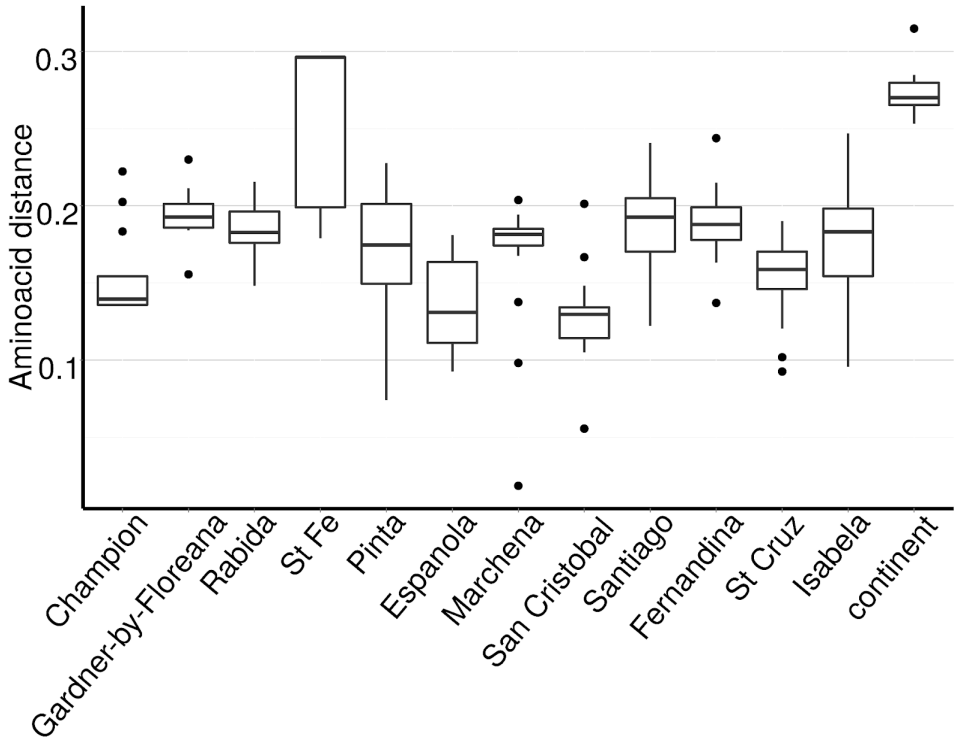


Figure 3 Intra-individual amino-acid distance of MHCIIβ alleles per population (Di). Box and whisker plots show the median (horizontal line), upper and lower quartiles (box) and maximum and minimum values (excluding outliers represented by a dot). Populations are ordered from smallest to largest.

Effects of balancing selection on MHCIIβ diversity

The comparison of synonymous to nonsynonymous changes showed that the whole MHCIIβ locus was under balancing selection (mean dN–dS = 1.74, P = 0.042). Moreover, the codons involved in antigen binding showed a significantly higher number of non synonymous changes (mean dN–dS = 1.56) than all other codons (mean dN–dS = 0.21). The codons with elevated numbers of nonsynonymous changes (nine of them also identified by SLAC to be under BS) corresponded to the sites that were

found under balancing selection in other songbird species rather than to the ABS identified in the human model (Figure 4). In total, we found 18 codons with dN–dS > 1, from which 13 sites matched with homologous codons under positive selection found in other species of songbirds and only nine corresponded to human antigen binding sites.

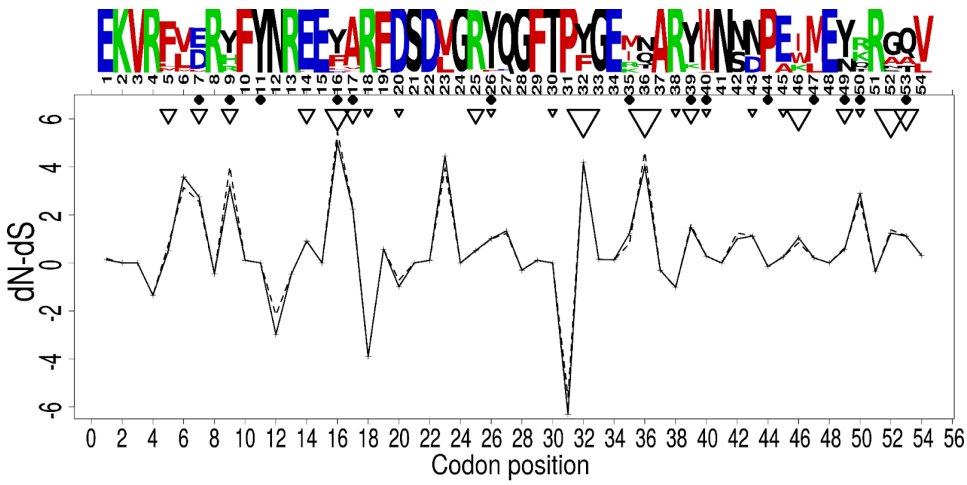


Figure 4 Variation of dN-dS in MHCIIβ alleles in the populations of GM and *Mimus polyglottos*. The solid line shows normalized dN-dS values based on MEGA, the dashed line is the same variable analysed by SLAC. The sequence logo above the top X axis represents the frequencies of each amino acid in the MHCIIβ alleles. Triangles denote codons under positive selection that were found in previous studies of MHCIIβ in five other songbird species: *Philesturnus carunculatus* and *Philesturnus rufusater* (Sutton et al. 2013); *Luscinia svecica* (Anmarkrud et al. 2010); *Ficedula albicollis* (Zagalska-Neubauer et al. 2010); and *Passer domesticus* (Borg et al. 2011). The size of the triangles reflects the number of matches across studies. Squares denote antigen binding sites identified in the human MHC molecule (Brown et al. 1993).

The phylogenetic network of MHCIIβ alleles showed five distinctive clusters. But the clustering did not follow the neutral marker-based

phylogeny of GM and *M. polyglottos*. MHC alleles from all populations were scattered among all clusters (Figures S2-S4, Supporting information). One cluster contained only alleles specific for *M. polyglottos* population, but several other alleles from *M. polyglottos* population were clustered together with alleles from GM population suggesting trans-species evolution of MHCIIβ.

Sharing of alleles between species and populations was high within GM (Figures S2-S3, Supporting information). We found 4.3 private alleles and 17.8 shared alleles on average per Galápagos population. Sharing between GM was not limited to species boundaries, each species shared at least several alleles with the other species (Table S9, Supporting information, e.g. 11 shared alleles between *M. macdonaldi* (Española) and *M. parvulus* (Isabela). However, we observed a general pattern that more diverged populations shared less MHC alleles, because the inverse number of shared MHC alleles was positively correlated with microsatellite Fst (Pearson's product–moment correlation = 0.323, P = 0.0072). The continental population of *M. polyglottos* showed a very high number of private alleles (30 from a total of 32 alleles) compared to GM.

Populations of GM, except for St Fe, showed no among-population differences in average amino acid distance (Di) as a proxy of functional divergence between alleles. Our data did not support general presence of divergent allele advantage in small populations as we found no effect of island size on Di (slope = 0.081, CI = 0.021 to 0.371, LRT = 0.344, P = 0.558). St Fe with its high Di value was the only outlier in this analysis.

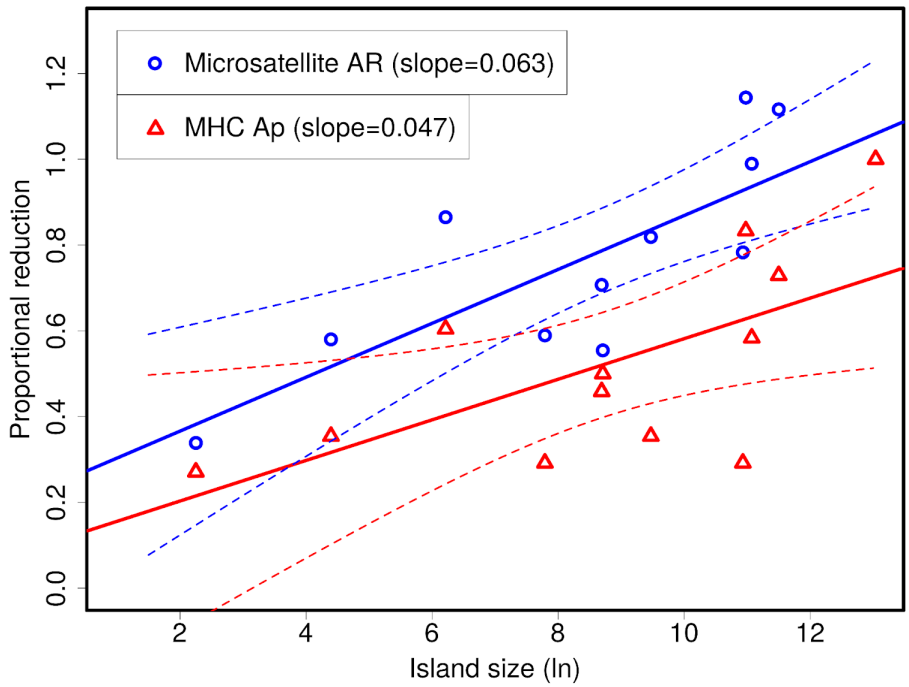


Figure 5 Regression of the proportional reduction in two genetic diversity indices (microsatellite allelic richness (in blue), the number of MHCIIβ alleles (in red) and natural logarithm of island size in hectares). The proportional reduction in both indices was calculated as the proportion of the individual index in relation to the largest island Isabela. Regression slope of microsatellite AR is slightly steeper compared to MHC Ap, but the difference was not statistically significant. Regression function AR: y = 0.240 + 0.063x; Ap: y = 0.108 + 0.047x.

Discussion

We genotyped MHCIIβ in four species of mockingbirds in Galápagos and one of their close continental relatives, *Mimus polyglottos*. We found substantial diversity in MHCIIβ: 150 unique alleles translated to 149 unique peptides, 1–12 alleles per individual (Ai) suggesting at least six

MHCIIβ loci and 13–48 alleles per population (Ap). These indices of MHCIIβ diversity are within the range of diversities found in other songbirds by high throughput sequencing (Zagalska-Neubauer et al. 2010; Sutton et al. 2013). Also, as expected, the closest homologous alleles were found in the phylogenetically closest species (Table S3, Supporting information). The distribution of MHC diversity across populations of GM and *M. polyglottos* suggests that the two evolutionary forces, genetic drift and PMBS, have dissimilar effects on the different levels of MHC diversity.

MHCIIβ diversity in GM and *M. polyglottos*

Populations of GM and continental *M. polyglottos* differed in their MHCIIβ supertype and allelic composition (1 unique supertype and 30 unique alleles in *M. polyglottos*) and in the evolutionary distance within individuals (Di; on average two times higher in *M. polyglottos*). Surprisingly, the two lineages did not differ systematically in the genetic diversity represented by either the number of MHCIIβ alleles/supertypes in an individual (Ai) or in a population (Ap). This finding contrasts with our first prediction (i) that MHC diversity in the continental population should be higher because of milder effects of genetic drift and higher pathogen diversity, similarly to results from the majority of studies that compared insular and continental populations (Miller & Lambert 2004b; Bollmer et al. 2011). However, there were a few exceptions to the general pattern of similar numbers of MHCIIβ alleles between GM and *M. polyglottos*. Two small populations, Champion (N*e* = 43), and St Fe (N*e* = 513), and one larger, San Cristóbal (N*e* = 1206; N*e* values from Hoeck et al. 2010b), showed much lower Ai values (Figure 2), while the biggest island Isabela (N*e* = 1591) showed a higher Ap compared to *M. polyglottos*. In the smaller populations, genetic drift might have caused the decrease in Ai. However, this pattern was not consistent for all GM populations, and it is possible that these three populations (Champion, St Fe and San Cristóbal) went through a bottleneck event that caused the

drop in Ai. This scenario is plausible at least in the case of the tiny Champion population, which diverged from the former and now extinct population on Floreana around 270 years ago and has retained low N*e* ever since (Hoeck et al. 2010a).

In contrast to Ai and Ap, the microsatellite diversity indices HS and AR were considerably higher in the continental compared to the insular GM populations (Table 1). The higher levels of neutral diversity in the continental population support our premise of larger N*e* and weaker effects of genetic drift in that population. The discrepancy between neutral and MHC diversity suggests that some additional mechanism maintained similar numbers of MHC alleles in large- to mid-sized insular populations as in the continental population notwithstanding the population size differences.

Despite the similar number of MHCIIβ alleles and supertypes, insular and continental populations showed remarkable differences in the intra-individual amino acid distance (Di). Divergent allele advantage hypothesis proposes that individuals with higher Di can recognize wider range of pathogens and thus have a higher fitness (Wakeland et al. 1990). This hypothesis was empirically supported several times (Richardson & Westerdahl 2003; Lenz et al. 2013). However, in our case we can only speculate whether the effect of divergent allele advantage in the continental population was stronger due to PMBS or other evolutionary mechanism. One argument for this assertion is that insular population usually harbours only impoverished pathogen communities (Hochberg & Møller 2001; Blumstein & Daniel 2005), which would cause lower selection pressure for Di in GM. Currently, we cannot support or refute this argument due to lack of empirical data on pathogen communities in our system. However, the hypothesis of lower pathogen pressure in GM becomes controversial when we consider supertype diversity that does not show strong contrast between the continent and GM. Therefore, we hypothesize that the difference in Di is a result of differences in

demographic histories between the GM and continental populations rather than difference in the PMBS pressure. Di diversity is accumulated during longer evolutionary timescale and can be relatively easily reduced by bottleneck event (Ejsmond & Radwan 2011) compared to supertype diversity that is less affected by random loss of alleles due to allele redundancy in individual supertypes. Thus, the discrepancy between Di and supertype diversity illuminates the fact that the two measures reflect different features of the MHCIIβ molecular diversity and can be affected by evolutionary forces in a different way.

Only one insular population, St Fe, showed exceptionally high Di (0.260), which reached almost the same level as in the continental population (0.273). This pattern suggests that a divergent allele advantage selection (Wakeland et al. 1990) might have shaped genetic diversity in this small population substantially. However, Di is averaged over the number of alleles per individual (Ai) and St Fe showed very low Ai (median Ai = 2). Thus, the observed pattern may merely represent a statistical bias. Nevertheless, no similar pattern was seen, for example on San Cristóbal where Ai was also very low (median Ai = 3) but Di was low as well (0.125). Therefore, it is likely that on St Fe a high effect of drift caused loss of copy number variation, but a few highly divergent alleles were retained as they possessed an evolutionary advantage to the population. The St Fe pattern seemingly fits the prediction of small Ai but higher allelic divergence in smaller populations, a pattern seen, for example in the Egyptian Vulture (Agudo et al. 2011). However, we found this pattern only in the St Fe population, whereas the MHCIIβ allelic distance (Di) across all GM populations showed no consistent correlation with population size.

MHCIIβ studies of insular populations of birds usually revealed reduced numbers of alleles compared to their continental counterparts (Miller & Lambert 2004b; Alcaide et al. 2010; Bollmer et al. 2011; Agudo et al. 2011). Only few empirical examples exist where the same numbers of

alleles were maintained in small populations, possibly because of PMBS in MHCIIβ (Borg et al. 2011) and MHCI (Oliver & Piertney 2012; New house & Balakrishnan 2015). The level of intra-individual allelic distance across several insular and continental populations of birds has been assessed in the Egyptian vulture (Agudo et al. 2011), where insular populations had lower numbers of MHCIIβ alleles but higher Di. Similar patterns of higher allelic divergence in small populations compared to larger populations as an adaptation to pathogens are well known from mammals (Hedrick et al. 2002; Sommer 2005), but in birds the evidence is scarce. Our study points to an opposite pattern, that is a similar number of alleles and supertypes in insular and continental populations (Ai, Ap with some exceptions) and a generally higher divergence (Di) in the continental population, which is a unique finding among avian MHCIIβ studies so far (Miller & Lambert 2004b; Alcaide et al. 2010; van Rensburg et al. 2012; Whittaker et al. 2012).

Population size and MHCIIβ diversity in GM

Variation of size of GM populations allowed us to test for the effects of genetic drift on MHCIIβ diversity across multiple populations. While we found no significant relationship between the numbers of MHCIIβ alleles present in an individual (Ai) and population size, there was a significant positive correlation between population size and number of alleles at the population level (Ap). This discrepancy (Figure 2) in the effect of population size between individual and population diversity of MHCIIβ is another unconventional finding of our study. To our knowledge, the two levels of MHC genetic diversity (represented by the same indices) were compared only in the studies of a Greater prairie chicken (*Tympanuchus cupido*; Eimes et al. 2011) and Black grouse (Strand et al. 2012). Both indices were negatively affected by genetic drift in these studies, while our study shows that only Ap decreased significantly with decreasing population size (Figure 2). Ai decreased markedly only in the populations where Ap dropped below 15 (Champion, St Fe and San Cristóbal, see

Figure 2). A plausible explanation for this pattern might be that the loss of an allele from a population (decrease in Ap) will not affect Ai proportionally, because of an uneven allele frequency distribution. The loss of a single rare allele from a population will affect Ap directly. However, it will not necessarily cause a similar reduction in Ai (which takes a maximum value of 12), because it is the common alleles that contribute most to Ai. This pattern could also be connected to bottleneck effects, which distort allele frequency distribution, and it would be in line with the idea developed above that the three populations with low Ai experienced some bottleneck event.

An alternative explanation involves balancing selection mitigating the effect of genetic drift on Ai. For example, nonrandom, preferential mating of individuals that are capable of recognizing MHC allelic diversity in their counterparts as observed in house sparrows (Griggio et al. 2011) will result in the maintenance of optimal Ai in a population as expected by theory (Woelfing et al. 2009). The effect of balancing selection is also supported by the observation of no significant relationship between supertype diversity measures and population size. Although we cannot determine which explanation is more plausible here, these patterns demonstrate some of the unique behaviour of MHCIIβ loci in natural populations, which has implications in both theoretical research and conservation management.

The observation of a positive relationship between Ap and population size (Figure 2b) is consistent with our prediction (ii) and it contributes to the growing body of empirical evidence that genetic drift negatively affects MHC genetic diversity proportionally to population size (Radwan et al. 2010; Sutton et al. 2011). Moreover, the effect of genetic drift on the reduction in microsatellite allelic richness and Ap was not significantly different (Figure 5). The observation of microsatellite diversity being lost to a similar degree as MHC diversity is in contrast to studies based on analysis of genetic diversities after strong bottleneck events, where MHC

diversity decreased more substantially than microsatellite diversity (Eimes et al. 2011; Sutton et al. 2011, 2015). This difference might be caused by the fact that GM represent different-sized populations in a mutation–drift equilibrium (for neutral loci), rather than heavily bottlenecked populations where the loss of MHC allelic diversity might be reinforced by the interaction between drift and balancing selection (Ejsmond & Radwan 2011).

Surprisingly, high genetic diversity, seemingly defying the effect of genetic drift, was observed in the small population of Rabida, where indices of neutral and MHC diversity (Ap, AR and HS) were inflated compared to the size of the island. Rabida lies very close (shore to shore distance = 4.4 km) to the much larger island of Santiago, which possesses one of the highest indices of genetic diversity in GM. Despite significant genetic differentiation between the two populations (Fst = 0.131, Hoeck et al. 2010a), occasional gene flow between the two islands is likely and provides a possible explanation for the pattern observed on Rabida.

The hypothesis of drift across loci (prediction iii) was not supported by our data set. We found no significant relationship between Ai and population size, while both our and an earlier study (using less loci, Hoeck et al. 2010) found a strong positive correlation between microsatellite diversity and population size. On the contrary, Eimes et al. (2011) found that Ai of MHCIIβ in Greater prairie chicken (*Tympanuchus cupido*) dropped by 44% compared to only an 8% decrease in microsatellite heterozygosity after a severe bottleneck. Eimes et al. (2011) therefore concluded that drift across loci (i.e. fixation of the same allele across loci) caused a greater decrease in MHCIIβ variation compared to neutral variation. It is questionable whether we can directly compare the results of our population study to a study that compared a pre- and postbottleneck population. Nevertheless, we conclude that in our study, which investigated small but not drastically bottlenecked populations, Ai was in general not significantly affected by genetic drift. Two

populations, Champion and St Fe, did not follow this pattern and we suggest that stronger reduction in Ai compared to other GM populations might be caused by a combination of low effective population size and balancing selection (Ejsmond & Radwan 2011).

Effects of balancing selection on MHCIIβ diversity

Several lines of evidence suggest that natural selection has been acting on MHCIIβ variation in GM and *M. polyglottos*, supporting the view that balancing selection reduced the effects of genetic drift. First, we observed a significant excess of nonsynonymous changes in antigen binding codons (Figure 4), which suggests that historically balancing selection shaped MHC variation (Garrigan & Hedrick 2003; Spurgin & Richardson 2010). Interestingly, positively selected codons in GM corresponded to codons that were also found to be under balancing selection in other songbirds (Zagalska-Neubauer et al. 2010; Anmarkrud et al. 2010; Borg et al. 2011; Sutton et al. 2013) rather than to codons known to be ABS in humans (Brown et al. 1993). This observation suggests that the MHCIIβ molecule in songbirds may have a different conformation than what has been observed in humans and that different codons can assume the crucial role of binding and presenting antigens.

A second indication of balancing selection (Garrigan & Hedrick 2003) shaping MHCIIβ in the recent past is the high volume of shared alleles found between GM populations. Almost three-fourth of the MHCIIβ alleles were shared among populations and species with alleles from all populations distributed across all of the five MHC lineages (Figures S1, S2, Supporting information). This is in stark contrast to the pattern at neutral markers where strong genetic divergence between GM populations was evident (Stefka et al. 2011). These patterns suggest that some type of balancing selection counter acted the neutral evolutionary processes (lineage sorting) in MHCIIβ and maintained the trans-species and trans-population polymorphisms after the colonization of the

archipelago. However, the sharing of MHC alleles was negatively correlated with divergence in microsatellites, that is in general more divergent populations shared a lower number of alleles. Hence, the picture that emerges is that of partial retention of ancestral MHCIIβ polymorphisms through balancing selection and increasing divergence through genetic drift and mutations. Thus, both neutral and adaptive evolutionary processes shaped current MHCIIβ diversity.

High levels of MHC allele sharing were observed also in a group of closely related finch species (*Nesospiza*, *Rowettia*), which colonized the Tristan da Cunha and Gough islands (van Rensburg et al. 2012). No evidence for pathogen-mediated selection was found in these populations, but the level of allele sharing between two divergent species was comparable to what we found in the GM.

Potential caveats

Using high-throughput sequencing techniques to genotype MHC loci still represents a relatively complex approach (Lighten et al. 2014a). Hence, we discuss potential problems in our data set that could bias the interpretation of the results. First, the accuracy between replicated amplicons (88.4%) was somewhat lower than in other MHCIIβ studies (98%: Sepil et al. 2012; 94%: Sutton et al. 2013). This suggests that we may have underestimated the number of alleles per individual. While this artefact may bias Ai, the population-level MHC allelic richness, Ap (as a sum of unique alleles over all individuals within a population), should be more robust against this problem because even if an allele is not detected in one individual, there is a high probability that it is detected in another individual from the same population. Second, we did not analyse whether the amplified MHCIIβ alleles are transcribed or not and we inferred that if an allele does not contain a stop codon, it is putatively functional. However, even alleles without stop codons may not be transcribed (Borg et al. 2011). On the other hand, the fact that many of the amplified alleles

showed evidence of balancing selection suggests that amplified alleles were functional at least for some time in the past. To overcome the problems with functional alleles, we have carried out supertype analysis, where putatively functional MHC units were characterized and the diversity indices based on the supertypes were used to corroborate the results based on allelic diversities.

Conclusions

Our study brings several novel findings to the long standing debate about the effects of genetic drift and balancing selection on MHCII variation. In the 12 different-sized populations of GM studied here, the effect of drift was detected only on the level of population allelic diversity (Ap), whereas all other diversity indices (Ai, supertypes and Di) were independent of population size and possibly maintained by balancing selection. These results point to an interesting and overlooked issue concerning drift/PMBS studies: different levels of genetic diversity might actually be affected differently. More such studies are needed to assess the Ai/Ap inconsistency and stochastic models should be constructed to test the mechanisms involved.

The comparison between GM and the continental population of *Mimus polyglottos* brought another interesting finding. Continental population showed no difference in MHC diversity represented either by Ai, Ap or supertype indices from GM. However, amino acid distances in MHCIIβ were higher in *M. polyglottos*. This provides support for selection of optimal number of MHC alleles across all populations, while the high amino acid distance in *M. polyglottos* is associated rather with stable population size than with stronger divergent allele advantage in the continental population.

Acknowledgement

We thank the Galápagos National Park Service and the Charles Darwin Foundation for providing research permits and logistic support in the field (permit no. PC-48-10) and our laboratory assistant Glauco Camenisch. We thank Marie Jeschek, Camila Mazzoni and Felix Heeger from the Berlin Center for Genomics in Biodiversity, for help with bioinformatics. Luboš Piálek, Jiří Košnar and Simone Sommer helped with the design of MHC amplicon sequencing. Petr Šmilauer advised on statistical data processing. We thank four anonymous reviewers for useful comments and suggestions. This study was funded by the Czech Science Foundation (project no. P506/12/P529). Ion Torrent sequencing was supported by the project CZ.1.05/2.1.00/ 01.0005 of the OPR&DI Project Centre for Nanomaterials, Advanced Technologies and Innovation.

Data accessibility and supplementary material

Sequences of MHCII genes were submitted to GenBank under Accession nos: KT123899–KT124105. Complete data set is published in Supporting Information files.

Additional supporting information may be found in the online version of this article (doi: 10.1111/mec.13807) which is available to authorised users.

References

Agudo R, Alcaide M, Rico C et al. (2011) Major histocompatibility complex variation in insular populations of the Egyptian vulture: inferences about the roles of genetic drift and selection. Molecular Ecology, 20, 2329–2340.

Aguilar A, Edwards S (2006) Patterns of variation in MHC class II β loci of the little greenbul (*Andropadus virens*) with comments on MHC evolution in birds. Journal of Heredity, 97, 133–142.

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, Lemus J, Blanco G et al. (2010) MHC diversity and differential exposure to pathogens in kestrels (Aves: Falconidae). Molecular Ecology, 19, 691–705.

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

Arbogast BS, Drovetski SV, Curry RL (2006) The origin and diversification of Galápagos mockingbirds. Evolution, 60, 370–382.

de Bakker PIW, Raychaudhuri S (2012) Interrogating the major histocompatibility complex with high-throughput genomics. Human Molecular Genetics, 21, R29–R36.

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

Blumstein DT, Daniel JC (2005) The loss of anti-predator behaviour following isolation on islands. Proceedings of the Royal Society B: Biological Sciences, 272, 1663–1668.

Bollmer JL, Hull JM, Ernest HB, Sarasola JH, Parker PG (2011) Reduced MHC and neutral variation in the Galápagos hawk, an island endemic. BMC Evolutionary Biology, 11, 143.

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

Borg AA, Pedersen SA, Jensen H, Westerdahl H (2011) Varia- tion in MHC genotypes in two populations of house sparrow (*Passer domesticus*) with different population histories. Ecology and Evolution, 1, 145–159.

Brown J, Jardetzky T, Gorga J (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA DR1. Nature, 364, 33–39.

Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Molecular Biology and Evolution, 21, 255–265.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Molecular Ecology, 9, 1657–1659.

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

Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics (Oxford, England), 26, 2455–2457.

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

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

Edwards SV, Grahn M, Potts WK (1995) Dynamics of Mhc evolution in birds and crocodilians: amplification of class II genes with degenerate primers. Molecular Ecology, 4, 719–729.

Eimes JA, Bollmer JL, Whittingham LA et al. (2011) Rapid loss of MHC class II variation in a bottlenecked population is explained by drift and loss of copy number variation. Journal of Evolutionary Biology, 24, 1847–1856.

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.

Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. Evolution, 57, 1707–1722.

Gonzalez-Quevedo C, Spurgin LG, Illera JC, Richardson DS (2015) Drift, not selection, shapes toll-like receptor variation among oceanic island populations. Molecular Ecology, 24, 5852–5863.

Goudet J (2015) HIERFSTAT, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Resources, 5, 184–186.

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

Grueber CE, Wallis GP, Jamieson IG (2013) Genetic drift out weighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. Molecular Ecology, 22, 4470–4482.

Grueber CE, Peel E, Gooley R, Belov K (2015) Genomic insights into a contagious cancer in Tasmanian devils. Trends in Genetics, 31, 528–535.

Hedrick PW (2012) What is the evidence for heterozygote advantage selection? Trends in Ecology & Evolution, 27, 698– 704.

Hedrick PW, Lee RN, Garrigan D (2002) Major histocompatibility complex variation in red wolves: evidence for common ancestry with coyotes and balancing selection. Molecular Ecology, 11, 1905–1913.

Hellgren O, Križanauskiené A, Hasselquist D, Bensch S (2011) Low haemosporidian diversity and one key-host species in a bird malaria community on a mid-Atlantic island (São Miguel, Azores). Journal of Wildlife Diseases, 47, 849–859.

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. 403–419. Springer, Berlin.

Hochberg ME, Møller AP (2001) Insularity and adaptation in coupled victim-enemy associations. Journal of Evolutionary Biology, 14, 539–551.

Hoeck PEA, Keller LF (2012) Inbreeding, immune defence and ectoparasite load in different mockingbird populations and species in the Galápagos Islands. Journal of Avian Biology, 43, 423–434.

Hoeck PEA, Bucher TB, Wandeler P, Keller LF (2009) Microsatellite primers for the four Galápagos mockingbird species (*Mimus parvulus*, *Mimus macdonaldi*, *Mimus melanotis* and *Mimus trifasciatus*). Molecular Ecology Resources, 9, 1538– 1541.

Hoeck PEA, Beaumont MA, James KE et al. (2010a) Saving Darwin's muse: evolutionary genetics for the recovery of the Floreana mockingbird. Biology Letters, 6, 212–215.

Hoeck PEA, Bollmer JL, Parker PG, Keller LF (2010b) Differentiation with drift: a spatio-temporal genetic analysis of Galápagos mockingbird populations (*Mimus* spp.). Philosophical Transactions of the Royal Society B: Biological Sciences, 365, 1127–1138.

Huchard E, Baniel A, Schliehe-Diecks S, Kappeler PM (2013) MHC-disassortative mate choice and inbreeding avoidance in a solitary primate. Molecular Ecology, 22, 4071–4086.

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

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. Nature, 491, 444–448.

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24, 1403–1405.

Kimura M (1983) The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. Annual Review of Genetics, 41, 281–304.

Kosakovsky Pond SL, Frost SDW (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular Biology and Evolution, 22, 1208– 1222.

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.

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, 20130714.

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, Bent zen P (2014b) Ultra-deep Illumina sequencing accurately identifies MHC class IIβ 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 Hochstet ter's frog. BMC Evolutionary Biology, 15, 63.

Lovette IJ, Arbogast BS, Curry RL et al. (2012) Phylogenetic relationships of the mockingbirds and thrashers (Aves: Mimidae). Molecular Phylogenetics and Evolution, 63, 219– 229.

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

Miller HC, Lambert DM (2004b) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). Molecular Ecology, 13, 3709–3721.

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, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution, 3, 418–426.

Newhouse DJ, Balakrishnan CN (2015) High major histocompatibility complex class I polymorphism despite bottlenecks in wild and

domesticated populations of the zebra finch (*Taeniopygia guttata*). BMC Evolutionary Biology, 15, 265.

Nietlisbach P, Wandeler P, Parker PG et al. (2013) Hybrid ancestry of an island subspecies of Galápagos mockingbird explains discordant gene trees. Molecular Phylogenetics and Evolution, 69, 581–592.

Oksanen J, Blanchet FG, Kindt R et al. (2015) vegan: Community Ecology Package. http://cran.r-project.org/package=vegan. Accessed 18 March 2016.

Oliver MK, Piertney SB (2012) Selection maintains MHC diversity through a natural population bottleneck. Molecular Biology and Evolution, 29, 1713–1720.

R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

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

van Rensburg AJ, Bloomer P, Ryan PG, Hansson B (2012) Ancestral polymorphism at the major histocompatibility complex (MHCIIβ) in the Nesospiza bunting species complex and its sister species (*Rowettia goughensis*). BMC Evolutionary Biology, 12, 143.

Richardson DS, Westerdahl H (2003) MHC diversity in two Acrocephalus species: the outbred Great reed warbler and the inbred Seychelles warbler. Molecular Ecology, 12, 3523– 3529.

Robertson A (1962) Selection for heterozygotes in small populations. Genetics, 47, 1291–1300.

Rothberg JM, Hinz W, Rearick TM et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature, 475, 348–352.

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.

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

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.

Slade R, McCallum H (1992) Overdominant vs. frequency dependent selection at MHC loci. Genetics, 132, 861–862. 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 CJ (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. BMC Genomics, 14, 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 OC, Illera JC et al. (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. Molecular Ecology, 20, 5213–5225.

Štefka J, Hoeck PEA, Keller LF, Smith VS (2011) A hitchhikers guide to the Galápagos: cophylogeography of Galápagos mockingbirds and their parasites. BMC Evolutionary Biology, 11, 284.

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.

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

Sutton JT, Robertson BC, Grueber CE, Stanton J-AL, Jamieson IG (2013) Characterization of MHC class II B polymorphism in bottlenecked New Zealand saddlebacks reveals low levels of genetic diversity. Immunogenetics, 65, 619–633.

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

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution, 30, 2725–2729.

Tobler M, Plath M, Riesch R et al. (2014) Selection from para sites favours immunogenetic diversity but not divergence among locally adapted host populations. Journal of Evolutionary Biology, 27, 960–974.

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

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

Van Oosterhout C, Joyce D, Cummings S (2006) Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). Evolution, 60, 2562–2574.

Wakeland EK, Boehme S, She JX et al. (1990) Ancestral polymorphisms of MHC class II genes: divergent allele advantage. Immunologic Research, 9, 115–122.

Whittaker DJ, Dapper AL, Peterson MP, Atwell JW, Ketterson ED (2012) Maintenance of MHC Class IIB diversity in a recently established songbird population. Journal of Avian Biology, 43, 109–118.

Wikelski M, Foufopoulos J, Vargas H, Snell H (2004) Galápagos birds and diseases: invasive pathogens as threats for island species. Ecology and Society, 9, 5.

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

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.

Chapter I

Chapter II

Association between louse abundance and MHC II supertypes in Galápagos mockingbirds

Parasitology Research (2020) 119, 1597-1605

DOI 0.1007/s00436-020-06617-3

Chapter II

Association between louse abundance and MHC II supertypes in Galápagos mockingbirds

Vlček Jakub[1,2,*], Štefka Jan[1,2]

1 Faculty of Science, University of South Bohemia, Branišovská 1670, 37005, České Budějovice, Czech Republic.

2 Biology Centre CAS, Branišovská 31, 37005, České Budějovice, Czech Republic

* corresponding author

Abstract

Major histocompatibility complex class II (MHC II) is an essential molecule triggering the adaptive immune response by the presentation of pathogens to helper T cells. The association between individual MHC II variants and various parasites has become a frequent finding in studies of vertebrate populations. However, although bird ectoparasites have a significant effect on their host's fitness, and the host's immune system can regulate ectoparasitic infections, no study has yet investigated the association between MHC II polymorphism and ectoparasite infection in the populations of free-living birds. Here we test whether an association exists between the abundance of a chewing louse (*Myrsidea nesomimi*) and MHC II polymorphism of its hosts, the Galápagos mockingbirds (*Mimus*). We have found that the presence of two MHC II supertypes (functionally differentiated clusters) was significantly associated with louse abundance. This pattern supports the theory that a co-evolutionary interaction stands behind a maintenance of MHC polymorphism. Moreover, we have found a positive correlation between louse abundance and heterophil/lymphocyte ratio (an indicator of immunological stress), that serves as an additional piece of evidence that ectoparasite burden is affected by immunological state in Galápagos mockingbirds.

Keywords

arms-race, co-evolution, immunity, ectoparasite, supertype

Introduction

The immune system is a multilevel defence scheme composed of innate and adaptive immune pathways that interact towards the common goal of controlling pathogens (Medzhitov 2007; Murphy and Janeway 2008). Reciprocal interactions between pathogens attempting to maximize their fitness at the expense of hosts, and hosts controlling pathogens by their immune system, is a key force driving the evolution of the immune system (Schneider and Ayres 2008). All immune pathways are initiated by a pathogen recognition that is provided by various kinds of host receptors. Among the most important recognition receptors are the molecules of the Major Histocompatibility Complex (MHC), which represent an essential part of the adaptive immune system (Rock et al. 2016). More specifically, MHC class II is a complex of proteins that serves as the detector of extracellular pathogens. Several MHC II proteins form a trans-membrane molecule that binds an antigen in a peptide-binding groove and presents it to helper T cells (CD4+). Helper T cells subsequently initiate a pathway of adaptive immune responses against the presented antigen with the ultimate goal of clearing the extracellular pathogen (Murphy and Janeway 2008; Rock et al. 2016). The binding of the antigen in the MHC peptide-binding groove (PBG) is the crucial moment for the subsequent immune response.

The biochemical properties of the PBG, primarily defined by the aminoacid sequence, determine the capacity to bind a specific antigen. Some MHC variants may overlap in their binding capacity due to their aminoacid chains possessing similar biochemical properties (Sette and Sidney 1998). For this reason MHC supertypes reflecting PBG binding properties have been used instead of individual variants in the majority of association studies (Trachtenberg et al. 2003; Schwensow et al. 2007; Sepil et al. 2013; Pilosof et al. 2014; Buczek et al. 2016). Supertypes are defined by clustering of the MHC variants with similar binding properties

based on the physicochemical properties of individual aminoacid sites that are in contact with antigens (Doytchinova and Flower 2005). Supertypes thus represent a cluster of variants that are very likely to recognize a similar set of antigens and allow a substantial reduction of the number of predictors in association models in situations where there are too many alleles, as in the case of songbirds (Bollmer et al. 2010).

The high levels of intra-specific and trans-species polymorphism of MHC indicate adaptive importance of the locus (Klein et al. 2007). More specifically, the polymorphism is maintained by balancing selection propelled either by sexual or natural selection, or most likely by a combination of both (Apanius et al. 1997). Considerable attention has been focused on the effects of natural selection mediated by pathogens. Several specific hypotheses have been developed to understand the exact evolutionary mechanisms that maintain such polymorphism (Spurgin and Richardson 2010). The heterozygote advantage (overdominance) hypothesis predicts that individuals heterozygous at the MHC will have higher fitness and therefore more variants will persist in a population and their frequencies will not vary considerably in time (Doherty and Zinkernagel 1975). On the other hand, the negative frequency-dependent selection (rare allele advantage) hypothesis depends on the co-evolutionary arms race between hosts and pathogens. It assumes that one MHC variant increases in frequency due to its ability to recognize a particular antigen until the antigen mutates and subsequently a different MHC variant comes under positive selection due to its ability to recognize the new antigen. These fluctuations continue ad infinitum and a loss of MHC alleles is precluded by the episodic directional selection drives of the rare alleles (Slade & McCallum 1992). Both hypotheses have found support in a wide range of reports. Evidence for overdominance comes from studies where heterozygous individuals showed higher pathogen resistance compared to homozygous individuals (Oliver et al. 2009; Worley et al. 2010; Savage and Zamudio 2011; Bolnick et al. 2014). Whereas associations between specific MHC

variants and pathogen resistance or fitness in birds (Westerdahl et al. 2013; Dunn et al. 2013; Sepil et al. 2013; Bateson et al. 2016) and mammals (Oppelt et al. 2010; Pilosof et al. 2014; Buczek et al. 2016), together with a simulation study (Ejsmond & Radwan 2015) support rare allele advantage as the prevalent mode of co-evolution between MHC and individual pathogens. However, it must be noted that the two modes of balancing selection are non-exclusive with the polymorphism of MHC being a result of both forces, especially when multiple pathogens are taken into account (Apanius et al. 1997; Oliver et al. 2009).

Compared to the examples linking PBG polymorphism with severe pathogens or endoparasites, we have focused on the association between PBG polymorphism and bird louse infections. Although ectoparasites live on a host's body surface, their effect on the host's fitness is considerable (Lehmann 1993; Richner et al. 1993; Clayton et al. 2015). A large body of evidence exists on the interaction between ectoparasites and immune system in birds and other vertebrates (Wikel 1982; James 1999; Evidence in birds reviewed in: Owen et al. 2010). Ectoparasites can trigger an immune response (e.g. Northern Fowl Mite (*Ornithonyssus sylviarum* Canestrini and Fanzago, 1877), King et al. 2011) and a bird's immune response can affect the survival of ectoparasites (e.g. Great Tit (*Parus major* Linnaeus, 1758) vs. Hen Flea (*Ceratophyllus gallinae* Schrank, 1803), Walker et al. 2003). Not only the strictly blood sucking ectoparasites (e.g. ticks or anopluran lice), but also chewing lice from the order Amblycera can interact with the host's immune system as they feed on the skin and come into direct contact with the blood by chewing growing pin feathers (Marshall 1981). Thus, chewing lice were found to be associated with high a production of eosinophiles in the Ring-billed Gull (*Larus delawarensis* Ord, 1815) (Fairn et al. 2012) and the biodiversity of ambyceran lice was associated with the intensity of immune response (Møller and Rózsa 2005).

Despite a large body of evidence for the interaction between ectoparasites and immune response (see Clayton et al. 2015 chap. 3 for a review), only one study has analysed the effect of MHC polymorphism on the resistance to ectoparasites in birds. Owen et al. (2008) found that domestic chicken (*Gallus gallus* Linnaeus, 1758) carrying a specific MHC variant showed a reduced abundance of Northern Fowl Mite due to their stronger inflammation of the skin compared to hens carrying other MHC variant (Owen et al. 2009). In mammals, the evidence is more abundant from both domestic and free-living species (Untalan et al. 2007; Oliver et al. 2009; Schad et al. 2012). Finally, an association between an ectoparasite and MHC variant has been found also in poikilotherms, e.g. in Sand Lizards (*Lacerta agilis* Linnaeus, 1758) (Olsson et al. 2005). But, to our knowledge, no study has yet focused on the effect of MHC polymorphism on ectoparasites in free-ranging populations of birds.

Here, we aim to assess whether MHC II polymorphism affects the abundance of louse *Myrsidea nesomimi* Palma and Price, 2010 in the populations of Galápagos mockingbirds (*Mimus*). We also include heterophil/lymphocyte ratio (H:L), a measure of stress and inflammation (Davis et al. 2008) to control for bird stress condition. More specifically we aim to test the following questions:

1. Does louse abundance correlate with the stress index represented by H:L ratio and (or) other host-determined environmental factors (sex, and weight)?

2. Does louse abundance correlate with the number of MHC II alleles, the number of supertypes in an individual or microsatellite heterozygosity?

3. Does louse abundance correlate with the presence or absence of individual MHC II supertypes?

4. Does H:L correlate with the presence/absence of a supertype?

In order to answer the questions, we have focused on the populations of four species of Galápagos mockingbirds (GM): *Mimus macdonaldi* Ridgway, 1890, *Mimus melanotis, Mimus parvuls* & *Mimus trifasciatus* Gould, 1837. Allopatric populations occupy islands of different size and there is a limited gene flow between them (Hoeck et al. 2010). GM were initially used to test how restricted population size affects genetic diversity and health status (Hoeck et al. 2010; Hoeck and Keller 2012). Although neutral genetic diversity was largely shaped by population size, the health status represented by H:L ratio, lysis, agglutination and number of ectoparasites was neither correlated with population size nor with the level of inbreeding. It has been argued that the lack of correlation was caused by the fact that the neutral genetic variation does not reflect variation in the genes that code for immune traits. In a follow up study (Vlček et al. 2016) we have described the diversity in MHCIIβ subunit and showed that MHC diversity, represented by the number of alleles in an individual, and by the number of supertypes, was not affected by population size to the same degree as neutral microsatellite diversity. We have also found supporting evidence for balancing selection in the excess of non-synonymous mutations in antigen binding sites and in extensive trans-species polymorphism. Observed patterns implied that genetic drift was partially outweigh by balancing selection, however the effect of MHC on fitness or parasite load was not analysed. Here we fuse previously published data on MHC with data on ectoparasite abundance to understand the effect of MHC polymorphism on ectoparasite load.

Material and methods

In order to analyse the effect of MHC polymorphism we tested for statistical associations between ectoparasite load, MHCIIβ polymorphism, neutral genetic diversity, and H:L ratio in 121 individuals of GM. The samples from 10 different populations and 4 species of GM were collected between 2006 and 2008. Microsatellite data (Hoeck et al. 2009; Hoeck and Keller 2012; Vlček et al. 2016) were used to calculate mean observed heterozygosity for each individual as an index of neutral genetic diversity. Ectoparasite load was assessed by dust-ruffling, a method where bird feathers are treated with an insecticide and ectoparasites that fall out within a predefined treatment duration (5 minutes) are collected and counted (Walther and Clayton 1997; Hoeck and Keller 2012). This method has proved generally reliable in estimating chewing lice numbers (Koop & Clayton 2013). In GM we found three species of ectoparasites, *Analgid* mites Linnaeus, 1758, *Brueelia galapagensis* Kellog and Kuawana, 1902 and *Myrsidea nesomimi* lice (Štefka et al. 2011), but for the purpose of our study we used only *Myrsidea* louse due to its widespread occurrence and high variance in abundance. Moreover this species displays the closest contact with the bird's skin and due to its blood feeding habit it is more likely to interact with the bird's immune system than *Brueelia* louse or *Analgid* mite (Møller and Rózsa 2005; Møller et al. 2010). *Myrsidea* abundance (sum of the number of nymphs and adults) was assessed for 121 samples from our dataset with only 6 birds being completely louse free.

Numbers of MHCIIβ alleles per individual were used as an index of MHC genetic diversity. MHCIIβ genotypes per individual were adopted from (Vlček et al. 2016). The dataset contained 120 MHCIIβ alleles in total with 1 - 12 alleles per individual. The reason behind such a wide variation in the number of alleles per individual is the co-occurrence of at least six MHCIIβ paralogs and the substantial copy number variation that

is a common feature of MHCII in song birds (Bollmer et al. 2010). Therefore, to express MHC genetic diversity in our models we used the total number of alleles per individual. This index combines both aspects of MHC variation – heterozygosity of individual loci and their total number.

Supertypes, functionally distinct MHCIIβ clusters, were used to represent individual functional variants. The supertypes (adopted from Vlček et al. 2016) were identified by clustering all MHCIIβ peptide variants based on physicochemical properties of nine amino acid sites in PBG according to the approach developed by Doytchinova and Flower (2005). We used only nine amino acid sites that were found under balancing selection, thus indicating importance of these sites in direct interaction with antigens (details in Vlček et al. 2016). For the purpose of this study, we excluded one of the eight originally found supertypes because of its extremely low frequency.

Heterophil/lymphocyte ratio (H:L) was used as an index of a health status of individual birds. An increased number of heterophils relative to lymphocytes indicates stress and inflammation in birds (Tompkins et al. 2006; Davis et al. 2008). We assumed that H:L could be linked with louse abundance (e.g. more stressed birds being more louse susceptible), but also we considered a situation in which H:L was affected by MHCIIβ supertypes. Therefore we included H:L in the whole model with ectoparasite load as a response (question 1), but we also tested whether H:L can be explained by MHCIIβ supertype presence in a separate linear model (question 4). H:L values, based on white blood cell counts, were adopted from a previous study (Hoeck and Keller 2012). One extreme outlier of H:L was removed from the dataset and the H:L ratio was normalized by natural logarithm.

To test if louse abundance is affected by H:L, or host-determined environmental factors (question 1), or overall MHC diversity (number of

MHC supertypes and alleles) (question 2) we constructed a single generalized linear mixed model with package lme4 (Bates et al. 2015). We used louse abundance on individual birds as the response variable explained by sex of the host, its body weight (log scaled), and scaled body weight (ratio of body weight to tarsus length reflecting condition), H:L ratio, number of MHC alleles per individual (Ai), number of MHC supertypes, and microsatellite heterozygosity. We used negative binomial error distribution in the model because the response variable showed a non-random aggregation of high values in some individuals causing a high level of over-dispersion, which is typical for parasite abundance data (Clayton et al. 2015, chap. 2). To account for differences caused by the sampling from 10 different populations, we supplied population identity as a factor with random effect. Species identity was not used because it largely overlapped with the population identity, moreover it did not explain any extra variation and it did not change general outcome when added as another random factor. Furthermore, a previous study (Štefka et al. 2011) showed that populations of both the birds and their parasites from individual islands function as highly related but genetically separated units, whether at inter- or intra-specific level. In order to determine relevant predictors, we used a backward stepwise model selection, in which predictors were eliminated based on the changes in Akaike's information criterion (ΔAIC). If the elimination of a predictor caused an increase of AIC by more than 2 the predictor was retained in the model, otherwise the predictor was removed. Two pairs of the predictors showed an elevated level of cross-correlation (number of MHC alleles and number of supertypes, and weight and scaled weight variance inflation factor (VIF) = 1.65, 9.5 respectively), but the number of alleles and the scaled weight were eliminated early in the process of model selection so the cross-correlation of explanatory variables did not bias the final model. Predictors that showed statistical significance in this exploratory analysis were further used in the statistical model for question 3.

To test question 3, we constructed a generalized linear model with louse abundance as a response term and the presence or absence of each individual supertype as a predictor. Moreover, we added H:L as an explanatory variable, because it was the only significant predictor from the previous model. As above, the model was constructed with a negative binomial error distribution and population identity as an explanatory factor with random effect. The best predictors were selected by backward selection using ΔAIC (see above). Because of the need for higher stringency in gene disease association studies (Manly 2005), several follow-up tests were run on the final model predictor selection. First, we evaluated association as robust if 95% confidence intervals of coefficient estimates did not approach zero. Second, we estimated effect sizes of the final predictors by Cohen's D method (Cohen 1988). Finally, we also tested a scenario in which supertypes confer a dissimilar effect on louse abundance in individual populations. Such a scenario would be expected under negative-frequency dependent selection, where independent associations are assumed to evolve. This hypothesis was tested by a comparison of the AIC of two models. In the first model, louse abundance was the response variable and the presence of individual supertype was the explanatory variable with fixed effect. Population identity was supplied as a random effect of the louse abundance. In the second model we added the random effect of the interaction between supertype and population. If the second model explained more variance it would have suggested that the supertype confers dissimilar effect in different populations.

The fourth question dealing with the association of MHC supertype and H:L ratio was tested in a separate linear mixed model. In this model H:L was explained by the presence or absence of the supertypes, with population identity used as a factor with a random effect.

All statistical analyses were performed in R 3.2.3 (R Core Team 2015).

Results

The first model, testing associations between louse abundance, H:L, host-determined environmental factors (question 1), and MHCIIβ and microsatellite diversity (question 2), revealed that H:L is the only statistically significant predictor of louse abundance (Table 1). The relationship was positive indicating that birds with higher H:L also showed higher louse abundance. No significant relationship was evident between louse abundance, and the number of MHC alleles, number of supertypes, and microsatellite heterozygosity.

Table 1 Results of the first statistical model testing associations between louse abundance, MHC diversity (№ MHC alleles, № MHC supertypes), microsatellite diversity (Microsat. Ho), heterophil/lymphocyte ratio (H:L) and other factors determining louse niche parameters (scaled weight, weight and host sex). Predictors are ordered according their elimination from the generalized linear model. Coefficient estimate of each predictor is given together with standard error (SE) of the estimate, Z test statistics, probability of Z test being significant (P-value) and the difference in Akaike's information criterion between the model without and with the predictor (ΔAIC). Statistically significant predictors are in italic.

Predictor	Coefficient	SE	Z-value	P-value	ΔAIC
№ MHC alleles	0.001	0.039	0.021	0.983	-2.000
Scaled weight	1.087	1.436	0.757	0.449	-1.428
Sex	-0.229	0.175	-1.308	0.191	-0.286
№ MHC supertypes	-0.026	0.066	-0.399	0.690	-1.839
Microsat. Ho	-1.083	0.484	-2.237	0.025	1.783
Weight	-0.693	0.572	-1.213	0.225	-0.586
H:L	*0.209*	*0.088*	*2.357*	*0.018*	*3.378*

The second model, testing association between louse abundance and individual MHC supertypes (question 3), revealed that the incidence of supertype 2, supertype 4, and H:L were statistically significant predictors of louse abundance (Table 2). Moreover, each association of the supertypes showed a significant effect size (Cohen's D_2 = 0.415, D_4 = 0.381) and the confidence intervals of the estimates did not reach zero (Table 2). Individuals with supertype 2 had significantly lower number of

lice (5.58 lice on average) compared to individuals without the supertype (8.08 lice on average). On the other hand, individuals with supertype 4 had significantly higher number of lice (8.93 on average) compared to individuals without the supertype (6.61 on average) (Figure 1). Nevertheless, we have to consider that supertype 4 was present only in 15 individuals out of 121, and thus one can consider this result less reliable due to unequal group size compared to supertype 2, which was present in 57 individuals. Furthermore, we found no dissimilar effect of the two supertypes among GM populations as the models with random effect of supertype and population interaction explained no more variance than the models without this random effect ($\Delta AIC < 2$). H:L ratio as the index of stress remained the only significant predictor of louse abundance apart from the supertypes. Finally, the third model (question 4) showed that the presence or absence of any supertype did not significantly affect the H:L ratio (removal of any of the predictors did not show $\Delta AIC > 2$).

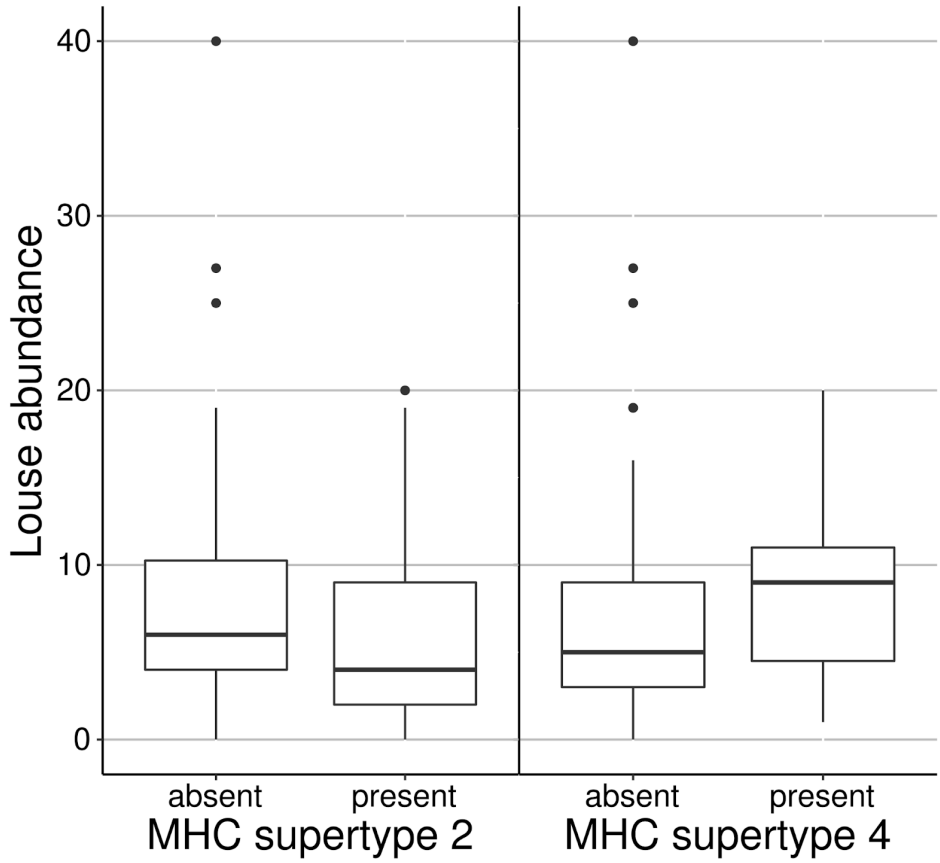


Figure 1 Abundance of *Myrsidea* louse on Galápagos mockingbirds is lower in individuals carrying MHCIIβ supertype 2 and higher in individuals carrying supertype 4. Distribution of the abundance is expressed by box-and-whiskers plots that show the median (horizontal line), upper and lower quartiles (box) and maximum and minimum values (excluding outliers represented by a dot).

Table 2 Results of the second statistical model testing association between louse abundance, individual MHC supertypes and heterophil/lymphocyte ratio (H:L). Predictors are ordered according their elimination from the generalized linear model model. Coefficient estimate of each predictor is given together with standard error (SE) of the estimate, 95% confidence interval (Confint), Z test statistics, probability of Z test being significant (P-value) and the difference in Akaike's information criterion between the model without and with the predictor (ΔAIC). Statistically significant predictors are in italic.

Predictor	**Coefficient**	**SE**	**Confint**	**Z-value**	**P-value**	**ΔAIC**
Supertype 3	-0.009	0.212	–	-0.043	0.966	-1.998
Supertype 8	0.080	0.216	–	0.371	0.711	-1.860
Supertype 1	-0.134	0.245	–	-0.546	0.585	-1.698
Supertype 5	-0.123	0.189	–	-0.654	0.513	-1.563
H:L	*0.214*	*0.084*	*0.050 – 0.384*	*2.551*	*0.011*	*4.529*
Supertype 4	*0.636*	*0.247*	*0.166 – 1.159*	*2.579*	*0.010*	*5.140*
Supertype 2	*-0.441*	*0.168*	*-0.771 – -0.096*	*-2.626*	*0.009*	*4.142*

Discussion

Replicated populations of the Galápagos mockingbird study system allowed us to test the effects of various genetic indices on the abundance of a widespread ectoparasite species. We observed no effect of neutral microsatellite heterozygosity or number of MHC alleles and supertypes (Table 1), which suggests that heterozygote advantage does not play any important role in the co-evolution between *Myrsidea nesomimi* and MHCIIβ in GM. However, we found a statistically significant association between the incidence of two MHCIIβ supertypes and the abundance of *Myrsidea* louse (Table 2). The presence of supertype 2 was associated with a reduced louse abundance, while the presence of supertype 4 was associated with a higher louse abundance. Associations between individual MHC variants, either positive or negative, and parasites have recently become a common finding (e.g. Sepil et al. 2013; Pilosof et al. 2014), however, this study is the first to document an MHC-ectoparasite association in free-living populations of birds.

The observed associations are based on a correlation, thus it should be noted that correlation does not necessarily mean causality. Other factors (such as other pathogens or environmental conditions) could drive similar patterns of louse abundance, with MHC not being directly involved. On the other hand, we controlled for many major factors that can determine louse abundance like host body weight, sex and population identity. Moreover, correlations with incidence of supertypes showed significant effect size and confidence intervals that did not reach zero. This provides our results relevant credibility of correlative associations that are worth discussing. Nevertheless, only a manipulative experiment would provide the power to unambiguously disentangle a causal relationship.

It might seem unexpected for a variation in a signalling molecule of the adaptive immune system to be linked with the abundance of an

ectoparasite, yet there are several plausible pathways that can explain our result. There is some evidence that MHC haplotypes affect ectoparasite community. Majority of the evidence comes from the studies of mammals and their blood sucking ectoparasites (Untalan et al. 2007; Schad et al. 2012), or fish and their gill ectoparasites (Seifertová et al. 2016). To our knowledge, only one very compelling study exists in birds. Owen et al. (2008) showed that domestic chickens carrying a specific MHC haplotype harbour less Northern Fowl Mites. Apart from the scarce MHC association studies, evidence of a general interaction between a host immune system and blood sucking ectoparasites is well documented in birds (Owen et al. 2010). The louse species in our study belongs to the ambliceran taxon of lice that usually live near their host's skin and feed on blood only occasionally (Marshall 1981). Therefore, the contact with the host will definitely be less intimate than between domestic chicken and Northern Fowl Mites, mainly due to the lack of a piercing mouth apparatus. Nevertheless, it has been observed that even non-bloodsucking lice can stimulate an immune response, although this result comes from sheep and their lice (James 1999). Furthermore, ambliceran lice were found to be capable of transmitting microfilarial worms via blood (Cohen et al. 1991). This evidence would allow us to explain the effect of the MHC supertype by a direct interaction between the louse and the immune system. In such case, individuals with supertype 2 would be able to recognize louse antigens penetrating via bites, initiating a proper immune response that would hinder louse fitness by skin inflammation. On the other hand, individuals with supertype 4 would be more susceptible to louse infection because of the supertype inability to recognize louse antigens and initiate an appropriate reaction.

Indirect effect of MHCIIβ on louse abundance is another plausible explanation of the observed pattern. MHCIIβ can affect composition of uropygial gland wax or skin and feather microbial communities that can subsequently affect louse abundance. Uropygial gland produces various waxes and other chemical compounds that keep a bird's plumage in good

condition (Jacob and Ziswiler 1982; Haribal et al. 2005). It affects microbial communities (Soler et al. 2012; Jacob et al. 2014) and to some degree also ectoparasites (Moreno-Rueda 2010). MHC variants were found to alter preen wax composition and related odours in seabirds (Leclaire et al. 2015) and songbirds (Slade et al. 2016). Leclaire et al. (2019) also found that the distribution of MHCIIβ functional diversity affected feather microbial community in Blue Petrel (*Halobaena caerulea* Gmelin, 1789). Furthermore, Leclaire et al. showed strong association with a bacterium from the genus *Arsenophonus* that is known as a symbiont of avian ectoparasites (Nováková et al. 2009). Such indirect evidence allows us to argue that supertypes 2 and 4 can affect either the composition of preen glands waxes and/or bacterial communities, which subsequently affects *Myrsidea* abundance.

Theoretically, indirect effect of MHCIIβ on louse abundance through other pathogens is also possible. If a supertype conferred resistance to a virulent pathogen, resistant individuals would have more energy to combat ectoparasites than susceptible individuals. But this situation is unlikely in GM where screening of a wide range of infectious disease agents in one of the GM species by Deem at al. (2011) showed absence of virulent endoparasites (e.g. haemoparasites or infectious bacteria) that could impair bird's health.

On a more general level, we have tried to investigate the link between MHCIIβ, louse abundance and heterophil/lymphocyte ratio (H:L), a measure indicating infection and stress in birds (Davis et al. 2004, 2008). We have found a significant positive correlation between H:L and *Myrsidea* abundance, but no link between MHCIIβ supertypes and H:L. Two alternative hypothesis can explain the observed pattern. Either a presence of higher number of lice causes more stress and an increase in H:L, or the birds are stressed by something else (e.g. other pathogen, ecological factors) allowing the lice to proliferate on such impaired individuals. We cannot discern between these two options based on our

correlative approach. Several studies showed a link between white blood cell derived indices and ectoparasite abundance. Fairn et al. (2012) found that louse abundance in Ring-Billed Gulls does not correlate with H:L ratio but only with the proportion of eosinophils in the blood. In the study of Eurasian Bee-eaters (*Merops apiaster* Linnaeus, 1758), manipulation with louse abundance affected bird weight as well as blood sedimentation and haematocrit (Hoi et al. 2012). Notwithstanding the causality, both studies, together with our results, indicate the existence of a link between immunity and louse abundance in wild populations of birds. However, the exact pathways for this phenomenon remain unresolved. An experimental approach considering all possible factors, from immunogenetic diversity or host morphology and exhaustive information on major pathogens to environmental factors should be used to evaluate this interaction.

In Galápagos hawks (*Buteo galapagoensis* Gould, 1837), a negative correlation was found between the level of genetic diversity, louse abundance, and the levels of natural antibodies, showing that a reduction in population size negatively affected immunocompetence (Whiteman et al. 2006). Such patterns were not observed in GM, where indices of immune response and ectoparasite abundance were independent of neutral genetic diversity and population size (Hoeck and Keller 2012). Current study has a potential to explain the lack of the population-size dependent pattern in Galápagos Mockingbirds as we found that the louse abundance is affected by incidence of individual MHCIIβ variants rather than by the total amount of neutral or MHCIIβ genetic diversity.

In the previous study (Vlček et al. 2016) we have also found that MHCIIβ supertype diversity is partially resistant to genetic drift, possibly due to the effect of balancing selection. One of the patterns typically generated by balancing selection is trans-species polymorphism (Spurgin and Richardson 2010; Lighten et al. 2017). Trans-species polymorphism is maintained if a common pathogen is present in different species and shared MHC alleles confer resistance to that common pathogen. In the

previous study (Vlček et al. 2016) we have observed a lot of shared polymorphism between different species and populations, but we did not know whether common supertypes confer some effect. Current observation of individual supertypes associated with *Myrsidea* abundance notwithstanding population identity serve as an evidence that observed trans-species polymorphism is functional, and thus balancing selection can be considered as an evolutionary force maintaining diversity of MHCIIβ supertypes.

Considering the type of balancing selection, our association results correspond to frequency-dependent selection rather than heterozygote advantage. In the case of heterozygote advantage, we should observe negative relationship between the number of MHC alleles and abundance of a parasite (Doherty and Zinkernagel 1975), whereas frequency-dependent selection usually results in either positive or negative associations with individual variants (Westerdahl et al. 2012). Frequency-dependent selection is also more prevalent when considering the arms-race between a single pathogen and its host, whereas MHC diversity becomes proportionally more important when multiple pathogens are considered (Apanius et al. 1997; Oliver et al. 2009). Therefore, because we have not sampled all possible pathogens, we cannot tell what is the exact type of balancing selection that maintains MHCIIβ diversity in Galápagos Mockingbirds.

In conclusion, our study provides the first evidence for an association between MHC variants and ectoparasite abundance in the populations of free-living birds. On top of the louse abundance being associated with variation in the signalling molecule of the adaptive immune system, we have found a positive correlation with the ratio of heterophils/lymphocytes. These observations indicate that louse infection can be linked with the state of the host's immune system. Additionally, the GM study system comprises populations highly restricted in size, where genetic drift may have the potential to counterbalance the forces of

selection. However, in agreement with several other studies (Aguilar et al. 2004; Vlček et al. 2016; Marmesat et al. 2017) the results presented here provide another piece of evidence for the capability of balancing selection to maintain functional diversity in MHC despite reduced population size.

Acknowledgements

We thank Prof. Lukas Keller and anonymous reviewers for valuable comments on the manuscript. We also thank the Galápagos National Park Service for permission to conduct this research (Permits No. PC-48-10 and PC-08-14) and the Charles Darwin Foundation for assistance in the field. We thank Czech Science Foundation and Grant Agency of the University of South Bohemia for funding.

Data accessibility

The dataset generated during the current study is available in the Mendeley Data repository, https://data.mendeley.com/datasets/75rn24pbvz/1

References

Aguilar A, Roemer G, Debenham S, et al (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proc Natl Acad Sci U S A 101:3490–4. doi: 10.1073/pnas.0306582101

Apanius V, Penn D, Slev PR, et al (1997) The nature of selection on the major histocompatibility complex. Crit Rev Immunol 17:179–224. doi: 10.1615/CritRevImmunol.v17.i2.40

Bates D, Maechler M, Bolker B, et al (2015) Fitting Linear Mixed-Effects Models Using {lme4}. Journal of Statistical Software 67:1–48. doi: 10.18637/jss.v067.i01

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. Mol Ecol 25:4730–4744. doi: 10.1111/mec.13793

Bollmer JL, Dunn PO, Whittingham LA, Wimpee C (2010) Extensive MHC class II B gene duplication in a passerine, the common Yellowthroat (*Geothlypis trichas*). J Hered 101:448–60. doi: 10.1093/jhered/esq018

Bolnick DI, Snowberg LK, Caporaso JG, et al (2014) Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. Mol Ecol 23:4831–4845. doi: 10.1111/mec.12846

Buczek M, Okarma H, Demiaszkiewicz AW, Radwan J (2016) MHC, parasites and antler development in red deer: No support for the Hamilton & Zuk hypothesis. J Evol Biol 29:617–632. doi: 10.1111/jeb.12811

Clayton DH, Bush SE, Johnson KP (2015) Coevolution of Life on Hosts. University of Chicago Press, Chicago

Cohen J (1988) Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, New York

Cohen S, Greenwood MT, Fowler JA (1991) The louse *Trinoton anserinum* (Amblycera: Phthiraptera), an intermediate host of Sarconema eurycerca (Filarioidea: Nematoda), a heartworm of swans. Med Vet Entomol 5:101–110. doi: 10.1111/j.1365-2915.1991.tb00527.x

Davis AK, Cook KC, Altizer S (2004) Leukocyte Profiles in Wild House Finches with and without Mycoplasmal Conjunctivitis, a Recently Emerged Bacterial Disease. EcoHealth 1:362–373. doi: 10.1007/s10393-004-0134-2

Davis AK, Maney DL, Maerz JC (2008) The use of leukocyte profiles to measure stress in vertebrates: A review for ecologists. Functional Ecology 22:760–772. doi: 10.1111/j.1365-2435.2008.01467.x

Deem SL, Parker PG, Cruz MB, et al (2011) Comparison of blood values and health status of Floreana Mockingbirds (*Mimus trifasciatus*) on the islands of Champion and Gardner-by-Floreana, Galápagos Islands. Journal of wildlife diseases 47:94–106. doi: 10.7589/0090-3558-47.1.94

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

Doytchinova I a, Flower DR (2005) In silico identification of supertypes for class II MHCs. J Immunol 174:7085–7095. doi: 10.4049/jimmunol.174.11.7085

Dunn PO, Bollmer JL, Freeman-Gallant CR, Whittingham L a (2013) MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats. Evolution 67:679–87. doi: 10.1111/j.1558-5646.2012.01799.x

Ejsmond MJ, Radwan J (2015) Red Queen Processes Drive Positive Selection on Major Histocompatibility Complex (MHC) Genes.

PLOS Computational Biology 11: e1004627. doi: 10.1371/journal.pcbi.1004627

Fairn ER, McLellan NR, Shutler D (2012) Are Lice Associated with Ring-Billed Gull Chick Immune Responses? Waterbirds 35:164–169. doi: 10.1675/063.035.0118

Haribal M, Dhondt AA, Rosane D, Rodriguez E (2005) Chemistry of preen gland secretions of passerines: Different pathways to same goal? why? Chemoecology 15:251–260. doi: 10.1007/s00049-005-0318-4

Hoeck PEA, Bollmer JL, Parker PG, Keller LF (2010) Differentiation with drift: a spatio-temporal genetic analysis of Galápagos mockingbird populations (*Mimus* spp.). Philosophical transactions of the Royal Society of London. Series B, Biological sciences 365:1127–1138. doi: 10.1098/rstb.2009.0311

Hoeck PEA, Bucher TB, Wandeler P, Keller LF (2009) Microsatellite primers for the four Galápagos mockingbird species (*Mimus parvulus*, *Mimus macdonaldi*, *Mimus melanotis* and *Mimus trifasciatus*). Mol Ecol Resour 9:1538–41. doi: 10.1111/j.1755-0998.2009.02704.x

Hoeck PEA, Keller LF (2012) Inbreeding, immune defence and ectoparasite load in different mockingbird populations and species in the Galápagos Islands. Journal of Avian Biology 43:423–434. doi: 10.1111/j.1600-048X.2012.05725.x

Hoi H, Krištofík J, Darolová A, Hoi C (2012) Experimental evidence for costs due to chewing lice in the European bee-eater (*Merops apiaster*). Parasitology 139:53–59. doi: 10.1017/S0031182011001727

Jacob J, Ziswiler W (1982) Avian Biology: The Uropygial Gland. Academic Press, New York

Jacob S, Immer A, Leclaire S, et al (2014) Uropygial gland size and composition varies according to experimentally modified

microbiome in Great tits. BMC Evol Biol 14:134. doi: 10.1186/1471-2148-14-134

James PJ (1999) Do sheep regulate the size of their mallophagan louse populations? Int J Parasitol 29:869–875. doi: 10.1016/S0020-7519(99)00055-7

Kellogg VL, Kuawana SI (1902) Papers from the Hopkins Stanford Galápagos Expedition, 1898–1899. X. Entomological results (8). Mallophaga from birds. Proceedings of the Washington Academy of Sciences 4:457–499

King MO, Owen JP, Schwabl H (2011) Injecting the mite into ecological immunology: measuring the antibody response of house sparrows (*Passer domesticus*) challenged with hematophagous mites. Auk 128:340–345. doi: 10.1525/auk.2011.10253

Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. Annu Rev Genet 41:281–304. doi: 10.1146/annurev.genet.41.110306.130137

Koop JAH, Clayton DH (2013) Evaluation of two methods for quantifying passeriform lice. Journal of field ornithology 84:210–215. doi: 10.1111/jofo.12020

Leclaire S, Strandh M, Dell'Ariccia G, et al (2019) Plumage microbiota covaries with the major histocompatibility complex in blue petrels. Mol Ecol 28:833–846. doi: 10.1111/mec.14993

Leclaire S, van Dongen WFD, Voccia S, et al (2015) Preen secretions encode information on MHC similarity in certain sex-dyads in a monogamous seabird. Scientific Reports 4:6920. doi: 10.1038/srep06920

Lehmann T (1993) Ectoparasites: Direct impact on host fitness. Parasitology today 9:13–17. doi: 10.1016/0169-4758(93)90154-8

Lighten J, Papadopulos AST, Mohammed RS, et al (2017) Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. Nat Commun 8:1294. doi: 10.1038/s41467-017-01183-2

Manly KF (2005) Reliability of statistical associations between genes and disease. Immunogenetics 57:549–558. doi: 10.1007/s00251-005-0025-x

Marmesat E, Schmidt K, Saveljev AP, et al (2017) Retention of functional variation despite extreme genomic erosion: MHC allelic repertoires in the *Lynx* genus. BMC Evol Biol 17:158. doi: 10.1186/s12862-017-1006-z

Marshall AG (1981) The Ecology of Ectoparasitic Insects. Academic Press, London

Medzhitov R (2007) Recognition of microorganisms and activation of the immune response. Nature 449:819–26. doi: 10.1038/nature06246

Møller AP, Erritzøe J, Rózsa L (2010) Ectoparasites, uropygial glands and hatching success in birds. Oecologia 163:303–311. doi: 10.1007/s00442-009-1548-x

Møller AP, Rózsa L (2005) Parasite biodiversity and host defenses: chewing lice and immune response of their avian hosts. Oecologia 142:169–76. doi: 10.1007/s00442-004-1735-8

Moreno-Rueda G (2010) Uropygial gland size correlates with feather holes, body condition and wingbar size in the house sparrow *Passer domesticus*. Journal of Avian Biology 41:229–236. doi: 10.1111/j.1600-048X.2009.04859.x

Murphy KP, Janeway C (2008) Janeway's Immunobiology. Garland Science, London

Nováková E, Hypša V, Moran NA (2009) Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiol 9:1–14. doi: 10.1186/1471-2180-9-143

Oliver MK, Telfer S, Piertney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). Proc Biol Sci 276:1119–28. doi: 10.1098/rspb.2008.1525

Olsson M, Wapstra E, Madsen T, et al (2005) Costly parasite resistance: a genotype-dependent handicap in sand lizards? Biol Lett 1:375–377. doi: 10.1098/rsbl.2005.0339

Oppelt C, Starkloff A, Rausch P, et al (2010) Major histocompatibility complex variation and age-specific endoparasite load in subadult European rabbits. Mol Ecol 19:4155–67. doi: 10.1111/j.1365-294X.2010.04766.x

Owen JP, Delany ME, Cardona CJ, et al (2009) Host inflammatory response governs fitness in an avian ectoparasite, the northern fowl mite (*Ornithonyssus sylviarum*). Int J Parasitol 39:789–799. doi: 10.1016/j.ijpara.2008.12.008

Owen JP, Delany ME, Mullens B a. (2008) MHC haplotype involvement in avian resistance to an ectoparasite. Immunogenetics 60:621–31

Owen JP, Nelson AC, Clayton DH (2010) Ecological immunology of bird-ectoparasite systems. Trends Parasitol 26:530–539. doi: 10.1016/j.pt.2010.06.005

Palma R, Price R (2010) The species of Myrsidea Waterston (Insecta: Phthiraptera: Menoponidae) from the Galápagos Islands, with descriptions of new taxa. Tuhinga 21:135–146

Pilosof S, Fortuna M a, Cosson J-F, et al (2014) Host-parasite network structure is associated with community-level immunogenetic diversity. Nat Commun 5:5172. doi: 10.1038/ncomms6172

R Core Team (2015) R: A Language and Environment for Statistical Computing

Richner H, Oppliger A, Christe P (1993) Effect of an Ectoparasite on Reproduction in Great Tits. Journal of animal ecology 62:703. doi: 10.2307/5390

Rock KL, Reits E, Neefjes J (2016) Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol 37:724–737. doi: 10.1016/j.it.2016.08.010

Savage AE, Zamudio KR (2011) MHC genotypes associate with resistance to a frog-killing fungus. Proc Natl Acad Sci U S A 108:16705–10. doi: 10.1073/pnas.1106893108

Schad J, Dechmann DKN, Voigt CC, Sommer S (2012) Evidence for the "good genes" model: Association of MHC Class II DRB alleles with ectoparasitism and reproductive state in the neotropical lesser bulldog bat, Noctilio albiventris. PLoS One 7: e37101 doi: 10.1371/journal.pone.0037101

Schneider DS, Ayres JS (2008) Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases. Nat Rev Immunol 8:889–895. doi: 10.1038/nri2432

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 (Edinb) 99:265–77. doi: 10.1038/sj.hdy.6800993

Seifertová M, Jarkovský J, Šimková A (2016) Does the parasite-mediated selection drive the MHC class IIB diversity in wild populations of European chub (*Squalius cephalus*)? Parasitol Res 115:1401–1415. doi: 10.1007/s00436-015-4874-4

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. Proc R Soc B Biol Sci 280:20130134. doi: 10.1098/rspb.2013.0134

Sette A, Sidney J (1998) HLA supertypes and supermotifs: A functional perspective on HLA polymorphism. Curr Opin Immunol 10:478–482. doi: 10.1016/S0952-7915(98)80124-6

Slade JWG, Watson MJ, Kelly TR, et al (2016) Chemical composition of preen wax reflects major histocompatibility complex similarity in songbirds. Proc R Soc B Biol Sci 283: 20161966. doi: 10.1098/rspb.2016.1966

Slade R, McCallum H (1992) Overdominant vs. frequency-dependent selection at MHC loci. Genetics 132:861–862

Soler JJ, Peralta-Sánchez JM, Martín-Platero AM, et al (2012) The evolution of size of the uropygial gland: Mutualistic feather mites and uropygial secretion reduce bacterial loads of eggshells and hatching failures of European birds. J Evol Biol 25:1779–1791. doi: 10.1111/j.1420-9101.2012.02561.x

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc Biol Sci 277:979–88. doi: 10.1098/rspb.2009.2084

Štefka J, Hoeck PEA, Keller LF, Smith VS (2011) A hitchhikers guide to the Galápagos: cophylogeography of Galápagos mockingbirds and their parasites. BMC Evol Biol 11:284. doi: 10.1186/1471-2148-11-284

Tompkins DM, Mitchell R a, Bryant DM (2006) Hybridization increases measures of innate and cell-mediated immunity in an endangered bird species. J Anim Ecol 75:559–64. doi: 10.1111/j.1365-2656.2006.01076.x

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

Untalan PM, Pruett JH, Steelman CD (2007) Association of the bovine leukocyte antigen major histocompatibility complex class II DRB3*4401 allele with host resistance to the Lone Star tick,

Amblyomma americanum. Vet Parasitol 145:190–195. doi: 10.1016/j.vetpar.2006.12.003

Vlček J, Hoeck PEA, Keller LF, et al (2016) Balancing selection and genetic drift create unusual patterns of MHCIIβ variation in Galápagos mockingbirds. Mol Ecol 25:4757–4772. doi: 10.1111/mec.13807

Walker M, Steiner S, Brinkhof MWG, Richner H (2003) Induced responses of nestling great tits reduce hen flea reproduction. Oikos 102:67–74. doi: 10.1034/j.1600-0706.2003.12208.x

Walther B a., Clayton DH (1997) Dust-ruffling: a simple method for quantifying. Journal of field ornithology 68:509–518. doi: 10.2307/4514260

Westerdahl H, Asghar M, Hasselquist D, Bensch S (2012) Quantitative disease resistance: to better understand parasite-mediated selection on major histocompatibility complex. Proc Biol Sci 279:577–84. doi: 10.1098/rspb.2011.0917

Westerdahl H, Stjernman M, Råberg L, et al (2013) MHC-I Affects Infection Intensity but Not Infection Status with a Frequent Avian Malaria Parasite in Blue Tits. PLoS One 8:e72647. doi: 10.1371/journal.pone.0072647

Whiteman NK, Matson KD, Bollmer JL, Parker PG (2006) Disease ecology in the Galápagos Hawk (*Buteo galapagoensis*): host genetic diversity, parasite load and natural antibodies. Proc Biol Sci 273:797–804. doi: 10.1098/rspb.2005.3396

Wikel SK (1982) Immune responses to arthropods and their products. Annu Rev Entomol 27:21–48. doi: 10.1146/annurev.en.27.010182.000321

Worley K, Collet J, Spurgin LG, et al (2010) MHC heterozygosity and survival in red junglefowl. Mol Ecol 19:3064–75. doi: 10.1111/j.1365-294X.2010.04724.x

Chapter II

Chapter III

Effect of population size and selection on Toll-like receptor diversity in populations of Galápagos mockingbirds

Journal of Evolutionary Biology (2022), in print

DOI 10.1111/jeb.14121

Chapter III

Effect of population size and selection on Toll-like receptor diversity in populations of Galápagos mockingbirds

Vlček Jakub[1,2,3,*], Miláček Matěj[1,2], Vinkler Michal[4], Štefka Jan[1,2]

1 Faculty of Science, University of South Bohemia, Branišovská 1670, 37005, České Budějovice, Czech Republic.

2 Biology Centre CAS, Branišovská 31, 37005, České Budějovice, Czech Republic

3 Charles University Faculty of Science, Department of Botany, Benátská 2, 128 01 Prague, Czech Republic

4 Charles University Faculty of Science, Department of Zoology, Viničná 7, 128 43 Prague, Czech Republic

* corresponding author

Abstract

The interactions of evolutionary forces are difficult to analyse in free-living populations. However, when properly understood, they provide valuable insights for evolutionary biology and conservation genetics. This is particularly important for the interplay of genetic drift and natural selection in immune genes that confer resistance to disease. The Galápagos Islands are inhabited by four closely related species of mockingbirds (*Mimus* spp.). We used 12 different-sized populations of Galápagos mockingbirds and one population of their continental relative northern mockingbird (*Mimus polyglottos*) to study the effects of genetic drift on the molecular evolution of immune genes, the Toll-like receptors (TLRs: TLR1B, TLR4, TLR15). We found that neutral genetic diversity was positively correlated with island size, indicating an important effect of genetic drift. However, for TLR1B and TLR4, there was little correlation between functional (e.g. protein) diversity and island size, and protein structural properties were largely conserved, indicating only a limited effect of genetic drift on molecular phenotype. In contrast, TLR15 was less conserved and even its putative functional polymorphism correlated with island size. The patterns observed for the three genes suggest that genetic drift does not necessarily dominate selection even in relatively small populations, but that the final outcome depends on the degree of selection constraint that is specific for each TLR locus.

Keywords

Conservation genetics, Purifying selection, Island birds, Innate immunity, Molecular phenotype.

Introduction

Genetic diversity is a crucial factor for the survival and adaptation of natural populations to a changing environment. Both theoretical and empirical evidence indicates that a reduction in population size increases the rate of genetic drift, which in turn leads to a loss of genetic diversity (Kimura 1983; Charlesworth 2009). Small population size also impairs the efficiency of selection and leads to a higher rate of fixation of deleterious alleles (Ohta 1992; Galtier 2016). Small populations are also prone to inbreeding depression (Frankham 2005). Overall, a reduction in population size can reduce the fitness of a population and worsen its chances of survival (Spielman et al. 2004; Willi et al. 2006). The effect of population size is particularly important for genes involved in pathogen recognition and determining the host's ability to respond appropriately to infection (DeCandia et al. 2018). Several authors have reported that these loci are shaped in small populations mainly by genetic drift, regardless of their functional importance (Miller & Lambert 2004; Grueber et al. 2013).

Our study focuses on single locus genes encoding Toll-like receptors (TLRs), molecules responsible for direct recognition of pathogen ligands. These first line of defence molecules trigger both innate and adaptive immune mechanisms (Iwasaki & Medzhitov 2010), and their variation is associated with variable resistance to various infectious diseases in humans (Ferwerda et al. 2007; Azad et al. 2012; Li et al. 2016) and birds (Leveque et al. 2003). TLRs are an evolutionarily conserved family of transmembrane glycoproteins characterised by their horseshoe-shaped exodomain structures (Wang et al. 2016). Each member of the TLR family recognises a specific group of ligands (Kumar et al. 2011). For example, TLR4 recognises lipopolysaccharide (LPS) (Nagai et al. 2002) and TLR1 recognises lipopeptides (Jin et al. 2007), while TLR15 is thought to be activated by bacterial and fungal proteases (de Zoete et al.

2011). Non-covalent binding of the ligands in the ligand binding region (LBR) leads to dimerisation of the receptor, which activates downstream signalling pathways (Kumar et al. 2011).

The predominant type of natural selection governing the overall molecular evolution of TLRs is purifying (i.e. stabilising) selection. This process leaves a distinct pattern in the non-synonymous to synonymous substitutions (dN/dS) that is lower than one across the vertebrate phylogeny (e.g. TLR4 ~ 0.425; TLR1B ~ 0.420; TLR15 ~ 0.325) (Wang et al. 2016). The prevalence of purifying selection in TLRs is supported by several studies focusing on the analysis of TLR polymorphism at the population level or in closely related taxa (Mukherjee et al. 2009; Darfour-Oduro et al. 2015; Vinkler et al. 2015; Raven et al. 2017; Nelson-Flower et al. 2018; Levy et al. 2020). Evidence of locus-specific positive (i.e. directional) selection has been found in several taxa (e.g. Alcaide & Edwards 2011; Tschirren et al. 2012; Quéméré et al. 2015; Králová et al. 2018), and the effect of balancing selection has also been demonstrated (Kloch et al. 2018; Quéméré et al. 2021; Minias & Vinkler 2022). The association between single TLR alleles and resistance to pathogens in natural populations is further evidence of the adaptive importance of TLR variation (Gavan et al. 2015; Tschirren 2015; Bateson et al. 2016).

Several authors have shown that a reduction in population size affects various aspects of TLR molecular evolution through increased stochasticity in sampling variance. For example, Grueber et al. (2013) reported that in a small population of Stewart Island robins (*Petroica australis rakiura*, Fleming), drift outweighed positive selection in TLR4 because an advantageous allele did not increase in frequency despite its positive effect on survival. Furthermore, Gonzales-Quevedo et al. (2015) observed that drift dominated the process of TLR genetic differentiation between island populations of Berthelot's pipit (*Anthus berthelotii*, Bolle). Finally, Hartmann et al. (2014) showed that the interaction of drift and

purifying selection depleted TLR diversity in a population of Pale-headed Brush-Finch (*Atlapetes pallidiceps*, Sharpe) severely affected by bottleneck. Interestingly, Brush-Finch's TLR diversity was negatively correlated with survival, suggesting that some deleterious alleles were segregating in the population. Although these studies indicate that the interaction of drift and selection is important and that drift is often prevalent, it remains unclear how population size affects the maintenance of adaptive variability over longer evolutionary timescales. Here we attempt to fill this gap by analysing TLR LBR diversity in a range of populations of different sizes.

Our study system consisted of 12 mockingbird populations from the Galápagos Islands (hereafter all mockingbird species on Galápagos are collectively abbreviated as GM). GM descended from a single colonisation event that occurred between 1.6 and 5.5 million years ago, based on mitochondrial DNA (Arbogast et al. 2006). Currently, four allopatric species have been described: the Floreana mockingbird *(Mimus trifasciatus*, Gould), the Hood mockingbird (*M. macdonaldi*, Ridgway), the San Cristóbal mockingbird (*M. melanotis*, Gould) and the Galápagos mockingbird (*M. parvulus*, Gould) (Figure 1). The Galápagos mockingbird occurs on several islands in the central and northwestern part of the archipelago, while the other species inhabit only one island or a few smaller islets. Importantly, there is only one allopatric population on each island. Regardless of their taxonomic species status, microsatellite data indicate that each population forms a distinct evolutionary unit within the GM monophyletic lineage (Hoeck et al. 2010). Migration between major populations is limited and effective population size is generally determined by island size (Hoeck et al. 2010). The Floreana mockingbird is an endangered species that survives in two very small populations, with population sizes on the small islet of Champion (0.095 km2) ranging from 20 to 50 individuals and those on Gardner-by-Floreana Island (0.81 km2) ranging from 60 to 180 individuals (Jiménez-Uzcátegui et al. 2011). The Galápagos mockingbird

population on the 4588.12 km2 island of Isabela is on the other side of the size spectrum, with an estimated effective population size (Ne) of several thousand (Hoeck et al. 2010). The other 9 populations bridge these extremes, forming a gradual Ne sequence ranging from a few tens to a few thousands (Hoeck et al. 2010) (S1, Table S12). The GM populations were supplemented by a population of Northern mockingbird (*Mimus polyglottos*, Linnaeus; hereafter abbreviated as NM), a widespread mainland relative of GM (Lovette et al. 2012) that diverged from GM ancestors about 5.7 million years ago. The Californian NM population selected for our research is the largest population in the study system (Vlček et al. 2016).

Based on this system, we test the extent to which population size influences molecular evolution in TLRs by comparing the relationships between island size and different indices of genetic diversity. We hypothesise that "neutral" (synonymous) TLR diversity variance can be attributed to population size and largely explained by the effect of drift, similar to microsatellite diversity (Hoeck et al. 2010). On the other hand, the correlation between "non-neutral" (e.g. non-synonymous, protein) diversity and population size should be weaker when natural selection dominates. To better understand the effect of island size on phenotypic diversity, we examine the distribution of protein variants and their physicochemical properties (tertiary structure, surface charge) throughout the study system. We also compare GM with a larger continental population of NM and test whether selection constraints have changed in these groups.

Methods

The DNA samples used in this study were adopted from Hoeck et al. (2010) and Vlček et al. (2016). For details on sampling and DNA extraction, see Appendix S1, SN1.1. To establish a baseline for our study, we first examined TLR1A, TLR1B, TLR2A, TLR2B, TLR3, TLR4, TLR5, TLR7, TLR15 and TLR21 diversity in all five species of our study system. This was done using Sanger amplicon sequencing, which allowed us to compare patterns of nucleotide polymorphism distribution between loci (details in Appendix S1, Chapter SN1.2). We selected TLR1B, TLR4 and TLR15 for population-wide analysis using high-throughput sequencing after an initial Sanger-based screen demonstrated reliability of amplification at these loci in all populations and detectable levels of polymorphism that allowed comparisons between populations (details in S1, Table S3).

Amplicon sequencing was used to sequence the TLR1B, TLR4 and TLR15 loci at 235 mockingbirds using Illumina MiSeq in pair-end mode. We used a dual-barcoded two-step PCR to prepare the amplicon library. In the first PCR, we amplified TLRs with specific primers designed based on preliminary sequencing data for the LBRs in TLR1B (Jin et al. 2007) and TLR4 (Park et al. 2009) and the vicinity of the putative cleavage site in TLR15 (de Zoete et al. 2011; Wang et al. 2016) in an analogous approach to Králová et al. (2018). Since the length limitation of MiSeq sequencing did not allow us to sequence the entire target region of TLR4 and TLR15 in one amplicon, we designed three primer pairs that produced shorter overlapping amplicons (S1, Table S1-S2). Since the LBR in TLR1B spans only a short region, we used only one amplicon. In the second PCR, we used primers that complemented the amplicon with sample-specific barcodes. Sequencing was performed at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany.

Detailed information on library preparation and sequencing can be found at S1, chapter SN1.3.

To resolve the TLR LBR haplotypes and analyse their genetic and phenotypic diversity, we processed the sequence data using the following procedures. First, we removed all low-quality sequences with Trimmomatic (Bolger et al. 2014) and amplicons with coverage less than six. The read pairs were then assembled using PEAR (http://www.exelixis-lab.org/pear). We then used the frequencies of unique nucleotide sequences (variants) per amplicon to distinguish between sequencing errors and true alleles and genotypes. If the ratio of the second most frequent to the first most frequent variant was greater than ⅓, we considered the amplicon to be heterozygous and included both variants. If, on the other hand, the ratio was smaller than ⅓, we considered the amplicon to be homozygous and used only the most common variant in the next steps (see details and rationale in S1, chapter SN1.3, Figure S1 and Table S4). Homozygous genotypes were automatically assembled using an in-house Python script, while heterozygous genotypes were manually assembled. Unresolved haplotypes resulting from amplification of the three partially overlapping amplicons in TLR4 and TLR15 were inferred by the programme PHASE using 1000 iterations, a thinning interval of 10 and a burn-in of 1000 (Stephens et al. 2001). In the case of TLR1B LBR, it was not necessary to phase haplotypes. Analysis of sequencing and phasing accuracy was performed by comparing 20 replicates included in the amplicon sequencing. The error rate of 0.056 confirms the reliability of the approach. The orthology of the sequenced loci was confirmed by BLAST. Basic diversity indices (nucleotide, haplotype, synonymous and non-synonymous diversity) and descriptive statistics (Tajima's D & Hardy-Weinberg equilibrium tests) were calculated based on the final dataset per population, per GM and over the entire sample set (including NM) using Dendropy (Sukumaran & Holder 2010), the SNAP programme (https://www.hiv.lanl.gov/content/sequence/ SNAP) and

Genepop (http://genepop.curtin.edu.au/, Rousset 1995). The full bioinformatics pipeline is available on Github (Vlček 2022) (for more details see S1, chapter SN1.4 and supporting data in Vlček et al. (2020).

We used linear regression models that fitted population diversity indices to island size to assess whether population size affects the genetic and phenotypic diversity of TLR LBR in GM (Snell et al. 1996). Species identity was not considered in the regression analysis for several reasons. Regardless of the taxonomic identity of the species, which is based on an arbitrary classification of divergence and morphology, all insular populations in fact represent genuine evolutionary units that descend from a recent common ancestor and therefore share most of their life-history traits. Previous research (Hoeck et al. 2010) has shown that GM on individual islands form genetically distinct evolutionary units whose N*e*s are determined by island size, independent of species identity. To exclude the potential effect of species identity, we repeated the analysis with a subset belonging to the single most widespread species Galápagos mockingbird (*M. parvulus*).

As an explanatory variable, island size was normalised with a natural logarithm. Nucleotide diversity, haplotype diversity, nucleotide diversity based on synonymous and non-synonymous sites and number of nucleotide haplotypes and protein variants were used as response variables in the subsequent independent models. Because the sample size per population was unbalanced, we also calculated linear models with diversity values derived from a sample set in which the number of individuals per population was randomly downsampled to match the population with the fewest individuals. Linear models were built in R v.3.5.2 (R Core Team 2016) using the lme4 package (Bates et al. 2015) and results plotted using the ggplot package (Wickham 2009).

The distribution of nucleotide and protein variants was analysed across populations. Comparison of the haplotype networks between the TLR loci

allowed us to understand the relative roles of different evolutionary constraints on each TLR locus. We created a haplotype network of all nucleotide haplotypes as a minimum spanning network in PopArt (Leigh & Bryant 2015) and then highlighted the most common protein variants. To investigate how divergence in TLR was affected by evolutionary constraints compared to neutral loci, we tested the extent to which pairwise genetic Jost's D distances between populations derived from TLR LBR correlated with distances derived from 26 microsatellites adopted from Vlček et al. (2016) (details in S1, chapter SN1.5.).

To understand whether the observed non-synonymous changes could significantly alter the phenotype of the receptors, we analysed several physicochemical features of the most common protein variants. First, we extracted segregating residues and categorised their conservatism based on a Grantham scale (Grantham 1974; Rudd et al. 2005). Second, we characterised protein variants based on the physicochemical properties of their segregating residues and used principal component analysis (PCA) to examine the relative differences between proteins in the composite physicochemical space (details in S1 Chapter SN1.6). Third, we investigated the effect of residue substitution on the functionality of TLR by analysing the tertiary protein structures. We modelled the three-dimensional structure of TLR LBRs using I- TASSER software (Yang & Zhang 2015). We then superimposed the structural models and calculated the root-mean-square deviations of atomic positions (RMSD) as a measure of structural difference. We also clustered the protein variants according to their surface charge, an important protein feature that affects the binding of pathogen-derived ligands (Walsh et al. 2008, Vinkler et al. 2014). The surface charge (electrostatic potential) of TLR LBR haplotypes was analysed using the programme PIPSA (Richter et al. 2008). We then used the R package pvclust (Suzuki & Shimodaira 2006) to perform hierarchical clustering of the proteins based on their surface charge distance and evaluated the uncertainty of the clustering by bootstrap resampling. All statistical analyses were performed in R version

3.5.2 (R Core Team 2016). For more details, see Appendix S1 chapter SN1.6.

Finally, we used several approaches to evaluate the role of natural selection in the molecular evolution of TLRs in GM compared to NM. First, we tested a scenario in which colonisation of Galápagos led to a significant accumulation of adaptive substitutions in TLR LBRs. We used the webpage platform MKT, to perform a McDonald-Kreitman test (Egea et al. 2008) to analyse whether there was an excess of non-synonymous substitutions between GM and NM compared to polymorphism within GM. For this test, we used the common starling (Sturnus vulgaris, Linnaeus) as a more divergent outgroup. Secondly, we tested an alternative scenario in which the selection pressure was relaxed after colonisation. For this, we used the programme RELAX (Wertheim et al. 2015), which compares evolutionary rates in a gene between test and reference lineages (alleles) by testing whether selection constraints are strengthened or relaxed in the test lineages compared to the reference branches. We assigned branches leading exclusively to GM alleles as test lineages and branches leading to NM as reference lineages, testing the change in selection constraints in GM relative to NM. Third, we used the codon-based methods FUBAR (Murrel et al. 2013) and FEL (Kosakovsky Pond & Frost 2005) to assess whether particular sites were significantly affected by a selection constraint. A site was only considered under selection if it was detected by both methods. For more details on the selection tests, see Appendix S1, Chapter SN1.7.

Results

We genotyped the LBRs of three TLRs in 229 individuals from GM and their continental relative NM to reveal patterns of functional immunogenetic diversity in small insular populations. The number of individuals sequenced per population varied between 10 and 25 (S1, Table S5). The individual TLR LBR loci differed considerably in their diversity levels, with TLR1B having the lowest values, TLR4 being in the middle range and TLR15 having the highest values across all estimated parameters (Table 1). All loci had an overall negative Tajima's D and an excess of synonymous over non-synonymous diversity. All populations and loci were in Hardy-Weinberg equilibrium, with the exception of TLR15 in Fernandina (S1, Table S5).

Table 1 Quantitative summary of the individual TLR LBR loci in all populations.

Locus	Len. (bp)	№ inds.	№ nuc. alleles	№ prot. var.	π	Hd	πN/πS	TajD
TLR1B	472	212	8	6	0.00096	0.163	0.138	-1.265
TLR4	889	228	34	18	0.00218	0.768	0.200	-1.655
TLR15	970	214	39	22	0.00274	0.832	0.225	-1.495

NOTE. — Len. (bp) = length in nucleotide base pairs; inds. = individuals; nuc. = nucleotide; prot. var. = protein variant; π = nucleotide diversity; Hd = haplotype diversity; pN/pS = ratio of non-synonymous to synonymous nucleotide diversity; TajD = Tajima’s D.

Island size was an important predictor of several diversity indices in the regression tests, but there were significant differences between loci (Figure 2; S1, Figure S2). In the TLR15 LBR, all diversity indices showed a statistically significant positive correlation with island size, while the correlation for all diversity-related traits was significant for TLR4 except non-synonymous diversity and number of protein variants.

For TLR1B, none of the diversity indices showed a significant relationship due to the low diversity at this locus. The regression results remained qualitatively unchanged when based on a downsampled dataset, with the exception of TLR4 nucleotide diversity, where the relationship was no longer statistically significant (the results of each statistical model are available in S1, Table S6). The results also remained similar in the analysis restricted to the widespread *M. parvulus* species only. The significance levels only changed for some of the tests for TLR4 (S1, Table S6).

The pattern of distribution of nucleotide and protein variants across populations also suggests that TLR4 and TLR1B in GM are substantially shaped by purifying selection. Although we observed 14 nucleotide haplotypes in GM TLR4, 96% of individuals carried only one of the two most common protein variants (P10, P4; Figure 1). TLR1B was almost completely monomorphic in GM populations, with protein variant P1 carried by 99% of individuals. TLR15, on the other hand, was less conservative with 20 nucleotide haplotypes and four major protein variants carried by 90% of individuals (Figure 1). We also observed a significant positive correlation between the pairwise genetic distance of TLRs and microsatellites, although this correlation was weaker for TLR4 than for TLR15 (correlation coefficients 0.573 and 0.797, respectively, Mantel $p < 0.0001$ for both; details in S1, Chapter SN1.4, Figure S4 - S5). In addition, we found that the synonymous diversity of the population was significantly higher than the non-synonymous diversity in TLR4 ($T11.4 = -4.16$, $P = 0.0015$; SI Figure S3), but not in TLR15 and TLR1B, supporting the existence of differences in selection constraints between loci.

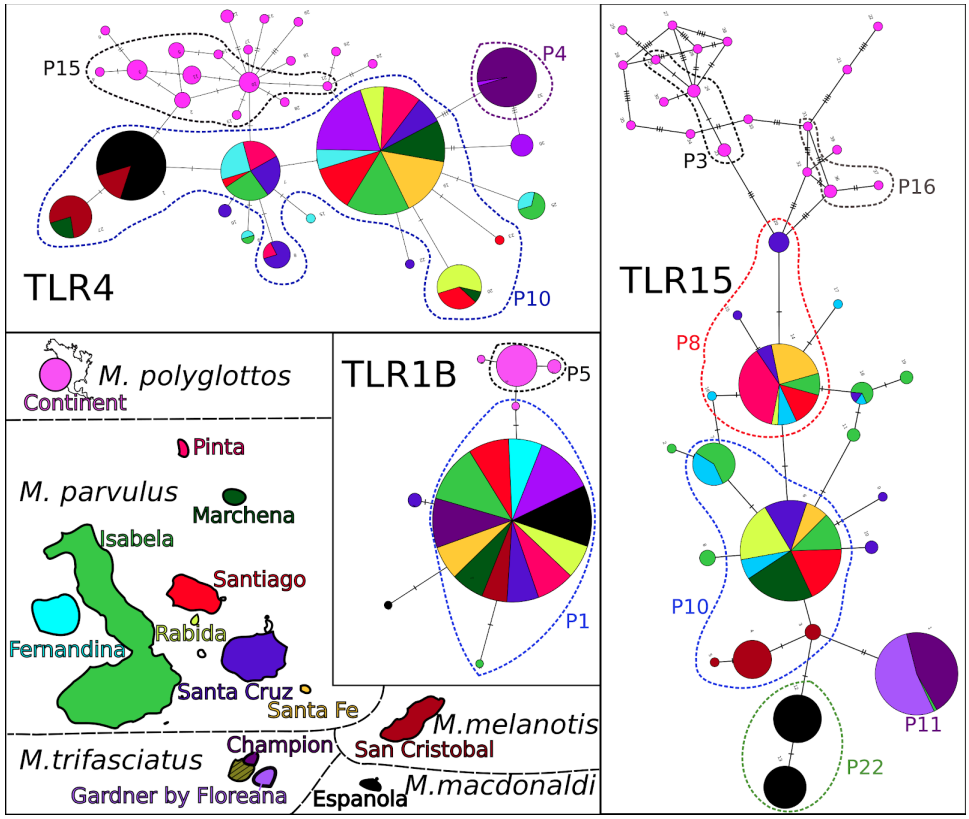


Figure 1 Map of the Galápagos and continental populations with haplotype networks of all TLR nucleotide haplotypes. Each pie chart indicates a unique haplotype, with colours indicating population identity and size indicating haplotype frequency. The most common protein variants are marked by dashed lines and their identifiers (e.g. P1). The size of the tiny Champion and Gardner by Floreana on the map is not scaled for clarity.

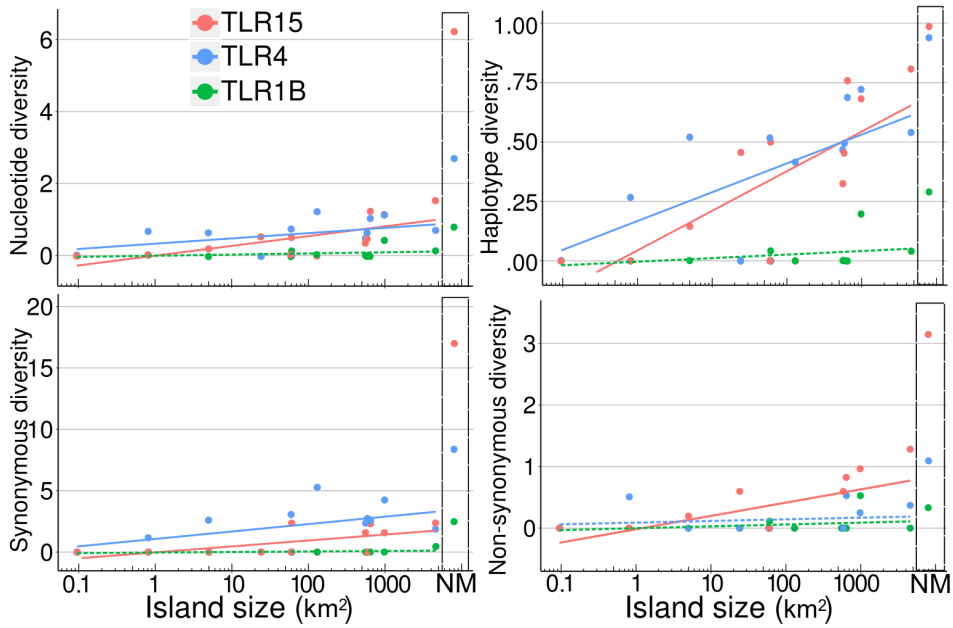


Figure 2 Relationship between population molecular diversity at three TLR loci and island size on a logarithmic scale. Most relationships are positive and statistically significant (indicated by a solid line). Statistically non-significant relationships are indicated by dashed lines. The lines represent predictions based on a linear regression model calculated for the Galápagos populations only. The values of molecular diversity for northern mockingbirds (NM) are shown on the right side of the diagrams in the dashed rectangle. Nucleotide, synonymous and non-synonymous diversity were multiplied by 1000. A figure showing the number of haplotypes can be found in S1, figure S2, the test values in table S6 and the size of the individual populations in table S12.

The physicochemical and structural differences between the most abundant protein variants in GM were insignificant for both TLR4 and TLR15 LBR. In TLR4, the two most abundant haplotypes differed in three residues with conservative (Leu > Val; Thr > Ser) or moderately non-conservative substitutions that differed in charge (Lys > Glu). For TLR15, the three most common haplotypes had only a single pairwise

substitution in residues (Arg > His; Leu > Val; Leu > Phe), which were all conservative (SI Table S7). Differences in protein structure, as measured by root mean square distance (RMSD), were negligible between all protein variants in TLR4 and TLR15 (SI: Table S11); clustering based on surface charge also showed no significant differences (Figure 3; details in S1 SN1.5, Figure S7-S11).

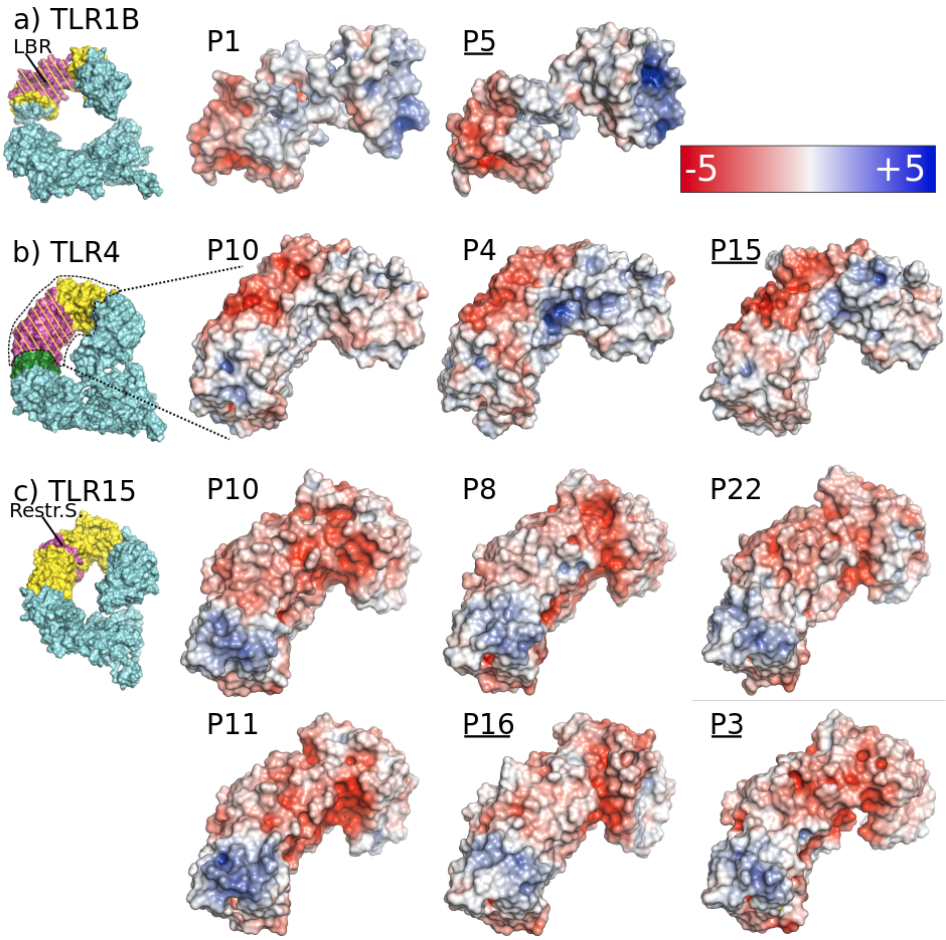


Figure 3 Visualisation of surface charge for sequenced segments of the most common protein variants. The colour gradient indicates the surface charge (red = negative, blue = positive). Haplotypes found in northern mockingbirds are indicated by the underlined text. The leftmost column shows a complete TLR molecule with the ligand binding region (LBR;

Jin et al. 2007, Park et al. 2009) or putative cleavage site (Restr. S.; Wang et al. 2016, de Zoete et al. 2011) highlighted in purple and the sequenced regions highlighted in yellow (hatching indicates overlap between the LBR and sequenced region). The green ribbon in TLR4 indicates a part of the LBR that was not sequenced.

A comparison of GM with NM showed that insular populations had lower diversity, but also that the selection constraints in NM matched the patterns observed in GM. Nucleotide and protein diversity was always higher in NM, whether compared to individual populations of GM or to diversity calculated across all GM, neglecting population structure (Table 2, Figure 2). Conversely, the ratio of non-synonymous to synonymous diversity was higher in GM. Interestingly, TLR4 was functionally conserved, even in NM, where the majority of DNA haplotypes encoded one protein variant (P15) and other protein variants were rare (Figure 1). A similar situation was observed in TLR1B, where one major and two rare protein variants occurred. A different pattern was seen at TLR15, where multiple protein variants segregated in intermediate frequencies. Consistent with the network of protein variants (Figure 1), we also observed that TLR4 haplotypes formed two tight clusters in physicochemical space specific to GM and NM (S1, Figure S6). Conversely, TLR15 did not form such tight clusters, but was clearly scattered in NM.

Table 2 Comparison of diversity indices for mockingbirds inhabiting the Galápagos Islands (GM) and the northern mockingbird (NM).

Locus	Strata	№ inds.	№ nuc. alleles	№ prot. var.	π	Hd	πN/πS	Charge	RMSD
TLR15	GM tot.	203	20	10	0.0022	0.815	0.25	0.567	2.1
TLR15	GM avg.	16.9	3.1	2.3	0.0005	0.344	0.71†	NA	NA
TLR15	NM	11	19	11	0.0062	0.985	0.19	0.597	2.1
TLR4	GM tot.	211	14	8	0.0019	0.730	0.18	0.633	1.5
TLR4	GM avg.	17.6	2.8	1.8	0.0006	0.386	0.33†	NA	NA
TLR4	NM	17	18	10	0.0027	0.938	0.13	0.615	1.64
TLR1B	GM tot.	196	4	3	0.0001	0.025	1	0.615	2.87
TLR1B	GM avg.	16.3	1.3	1.2	0.0001	0.023	0.92†	NA	NA
TLR1B	NM	16	4	3	0.0008	0.289	0.13	0.7	4.31

NOTE. — GM tot. = diversity indices calculated based on a pool of all GM individuals; GM avg. = diversity indices calculated based on individual GM populations and then averaged; inds. = individuals; nuc. = nucleotide; prot. var. = protein variants; π = Nucleotide diversity; Hd = haplotype diversity; pN/pS = ratio of non-synonymous to synonymous nucleotide diversity; Charge = average surface charge distance; RMSD = root mean square distance between protein models. † = within population pS was often 0, meaning that pN/pS was undefined. To address this, we arbitrarily assigned population pN/pS as value 1, neglecting any potential excess of non-synonymous mutations. Hence, the GM average is a conservative estimate as the real value is likely to be rather higher.

The McDonald-Kreitman test comparing polymorphism in GM with outgroups, was not significant at any locus or level of divergence (GM - NM and GM - common starling; Table 3). This low rate of non-synonymous change between NM and GM indicates an overall conservativeness of TLR loci in mockingbird evolution. Analysis of reciprocal evolutionary rate by RELAX showed that selection was neither relaxed nor strengthened at GM compared to NM (marginally non-significant at TLR15, where the values of the selection parameter k were less than one, indicating relaxed selection (k=0.23, P=0.077)).

Finally, codon-based tests indicated that several sites in TLR1B, TLR4 and TLR15 were subject to purifying selection, with one site in TLR15 likely subject to positive selection (details in Appendix S1 Chapter SN.1.6, Tables S8-S10).

Table 3. Results of McDonald-Kreitman tests for two different divergence levels and each TLR locus.

Locus	Outgroup	PN	PS	DN	DS	MKT Chisq	p-value
TLR15	NM	9	8	1	2.01	0.398	0.527
TLR15	Starling	9	8	21.4	13.54	0.324	0.568
TLR4	NM	8	7	1	1	0.008	0.928
TLR4	Starling	9	6	25.74	24.21	0.333	0.563
TLR1B	NM	2	1	1	2	0.68	0.409
TLR1B	Starling	2	1	13.31	15.5	0.456	0.499

NOTE. — PN = Number of non-synonymous polymorphisms in Galapágos mockingbirds (GM); PS = number of synonymous polymorphisms in GM; DN = number of non-synonymous fixed differences between GM and the outgroup; DS = number of synonymous fixed differences between GM and the outgroup; NM = Northern Mockingbird, MKT Chisq = McDonald-Kreitman Chi Square statistic.

Discussion

Our study provides a new perspective on the evolutionary processes behind the diversity of pattern recognition receptors in relation to population size. We found that neutral genetic diversity in TLR LBRs is affected by population size, but the effect of population size on functional diversity is attenuated by selection constraints. Importantly, selection constraints vary between TLR loci, affecting the extent to which functional diversity is shaped by genetic drift. Our results suggest that TLR15 is less constrained by selection, and is therefore more strongly influenced by genetic drift. On the other hand, strong purifying selection has shaped the phenotypic diversity of TLR1B and TLR4, despite the variable degree of stochasticity in the different-sized GM populations. Below, we discuss our results in light of current views on the molecular evolution of pathogen recognition receptors.

Purifying selection modulates TLR diversity in GM and NM

Natural selection is one of the four evolutionary forces that determine the distribution of genetic and phenotypic diversity (Lynch 2007). Since the distribution of TLR diversity has been reliably determined across several lineages, we investigated the extent to which it was shaped by selection and which type of selection was predominant. Several lines of evidence suggest that purifying selection was dominant in the evolutionary history of all three TLRs studied (TLR15, TLR4, TLR1B) in GM and NM. Tajima's D indicates an excess of rare variants, which can be interpreted as an effect of purifying selection, as we can rule out an alternative scenario of GM population growth based on the available microsatellite data (Hoeck et al. 2010). Furthermore, we found that the majority of non-synonymous mutations were purged, as we observed that the pN/pS ratio was less than one. Codon-based models showed that multiple positions were affected by purifying selection, with only one site in

TLR15 subject to positive selection. McDonald-Kreitman tests did not support our hypothesis that adaptive molecular changes occurred as a result of Galápagos colonisation, and the low rate of non-synonymous changes between NM and GM confirmed the predominant effect of purifying selection. Despite potential biases inherent in any of the dN/dS methods (Kryazhimskiy & Plotkin 2008; Mugal et al. 2014), we conclude that the LBRs of the three TLRs studied were substantially shaped by purifying selection. This result is consistent with other recent reports indicating a dominant role of purifying selection in TLRs at the intraspecific level or between closely related taxa (Mukherjee et al. 2009; Raven et al. 2017; Nelson-Flower et al. 2018). This conservatism could be due to shared pathogen communities in closely related taxa (Eichler 1948; Clark & Clegg 2017), an explanation that is even more plausible on islands with depauperate pathogen diversity (Wikelski et al. 2004).

Alleles escaping selection in the smallest GM population

Although purifying selection was predominant in all TLRs, some private or nearly private variants with non-synonymous substitutions reached relatively high frequencies (TLR4-P4, TLR15-P11, P22). By analysing their physicochemical properties, we tried to assess whether these variants have a significant impact on the phenotypic function of the receptor. We hypothesised that some deleterious variants may occur in extremely small populations due to less efficient selection, which is outweighed by genetic drift (Ohta 1992). In general, the residue changes were rather conservative in all TLRs. However, we observed one protein variant of TLR4-P4 that has become fixed in the most endangered and smallest population on Champion island. This variant shows several differences in surface charge pattern compared to variant P10 that is shared by all other GM populations (Figure 3). Assuming that the populations share the same pathogens, as suggested by Deem et al. (2011), the P4 haplotype on Champion island can be considered a deleterious mutation that got fixed and shifted the receptor structure away

from its optimum. Furthermore, Champion Island had significantly lower MHCIIβ diversity (Vlček et al. 2016) compared to the other GM populations, and poorer health than a larger population of the same species on Gardner-by-Floreana island (Deem et al. 2011). These observations suggest that in extremely small populations with a census size of less than 50, drift may outweigh selection constraints, even at loci involved in pathogen recognition, supporting the conclusion of Grueber et al. (2013).

Effect of population size differs between TLR loci

Based on the assumption that genetic drift plays an important role in GM populations of different sizes (Hoeck et al. 2010), we hypothesised that island size would be a significant predictor of both neutral and non-neutral diversity (estimated as synonymous and non-synonymous diversity, respectively). While our regression models supported the population size-related effects of genetic drift on TLR diversity, each TLR locus differed in the degree of constraint, limiting the effects of drift on non-synonymous diversity. At TLR4, island size was a statistically significant predictor of synonymous but not non-synonymous diversity (Figure 2), while at TLR15 non-synonymous diversity was also positively correlated with island size. Furthermore, only at TLR4 did we observe significantly increased synonymous diversity compared to non-synonymous diversity (Figure S3). TLR1B had a low amount of polymorphism and here island size was not a valid predictor for any of the diversity indices. The scenario of different selection constraints in the different TLRs was also supported by the haplotype network analysis (Figure 1), which showed only one major protein variant (P10) in TLR4 that was shared by almost all GM populations, while in TLR15 there were four major protein variants, two of which were species-specific (P22 and P11). In addition, genetic distances between populations in TLR15 were more consistent with distances based on neutral microsatellites than in TLR4. These patterns suggest that genetic variation in TLR15 is more

strongly shaped by genetic drift than in TLR4 and TLR1B, as these are more constrained by selection. Selection did not restrict the accumulation of non-synonymous mutations in TLR15, whereas the same class of mutations was purged from TLR4. TLR1B was purged to an even greater extent and is virtually free of non-synonymous polymorphisms. Alternatively, the lack of diversity in TLR1B could be due in part to the shorter length of the sequenced fragment (Table 1) combined with the effect of linkage. However, the overall effect of fragment size should be negligible in our statistical analyses due to the use of per-site diversity measurements.

The differences in genetic diversity between NM and GM provide further evidence of differential selection constraints between TLR loci, with a common pattern of reduced TLR genetic diversity in GM compared to NM. This was most likely caused by a colonisation bottleneck and subsequent evolution in small populations (founder event), with random drift affecting TLR diversity more intensely in GM compared to NM. This scenario is consistent with previously reported differences in microsatellite diversity between NM and GM (Vlček et al. 2016). Here we argue that some patterns of TLR diversity loss suggest a joint effect of natural selection and genetic drift. If no selection was involved in reducing genetic diversity, the mutation rate at each TLR locus was the same, and populations were in drift-mutation equilibrium, as indicated by Hoeck et al. (2010), then we would expect roughly identical levels of diversity at each TLR locus. In contrast to this prediction, we observed higher nucleotide and haplotype diversity at TLR15 compared to the other loci in both GM and NM. In NM, the nucleotide diversity of TLR15 was ~2 and ~7 times higher than TLR4 and TLR1B, respectively (Table 3). The difference between GM and NM was also much greater for TLR15 (0.004) than for TLR4 (0.0008) and TLR1B (0.0007), suggesting that TLR15 evolves under less stringent selection constraints. Similar patterns showing more dramatic changes in TLR15 diversity compared to other loci were also observed in a study of mainland and island

populations of the song sparrow (*Melospiza melodia*) (Nelson-Flower et al. 2018).

The differences in selection constraints observed between the TLR loci can be explained by functional differences in TLRs (Iwasaki & Medzhitov 2010). TLR15 is apparently adapted to a different mechanism of pathogen recognition than TLR4 and TLR1B. For TLR15, TLR4 and TLR1B, the sequenced regions contained 10%, 58% and 53% of sites known to be involved in ligand binding or putative cleavage activation, respectively (Jin et al. 2007; Park et al. 2009; de Zoete et al. 2011; Wang et al. 2016). Although the exact mechanism leading to TLR15 activation is not well known, it appears that its constrained region is either smaller compared to TLR4 and TLR1B or that selection to maintain its function is weaker. For TLRs, several studies have demonstrated the importance of positive selection for interspecific variation (Alcaide & Edwards 2011; Areal et al. 2011; Greuber et al 2014; Velová et al. 2018; Khan et al. 2019) and positive or balancing selection for intraspecific polymorphism (Tschirren et al. 2012; Davies et al. 2021). However, at GM we do not assume that the non-synonymous variation in TLR15 is due to some form of stronger diversifying or balancing selection. This is because our analysis revealed little evidence of functional differentiation or positive selection in the TLR15 alleles, while also indicating a marginally non-significant relaxation of selection acting in TLR15. The selection patterns acting in TLRs are lineage specific (Wlasiuk et al. 2009). Similar to our results, TLR15 has often been shown to have increased non-synonymous population polymorphism (e.g. in the domestic chicken, Ruan et al. 2012, in the Seychelles warbler, *Acrocephalus sechellensis*, Gilroy et al. 2017). Recent finding of cryptic functional pseudogenisation of TLR15 in *Eudyptes* penguins (Fiddaman et al. 2022), also suggests relaxed purifying selection in this gene. And relaxed selection in TLR15 is also consistent with the observation of fewer sites under positive selection in this gene on a broader phylogenetic scale (Velová et al. 2018). These findings illustrate that each TLR locus is under different

selection constraint and that increased polymorphism does not necessarily indicate balancing selection.

Conclusions

Several studies based on free-living island populations have claimed that TLRs in small populations are predominantly shaped by random genetic drift. Gonzales-Quevedo et al. (2015) found that the divergence of TLRs between populations was caused by demographic history rather than selection, and Grueber et al. (2013) observed that the frequency of an allele, even if it showed an advantage, was determined by genetic drift. While both studies concluded that drift as an evolutionary process outweighed natural selection on islands, our results are at odds with these findings. Using a study system in which differential island size serves as a proxy for the intensity of genetic drift, we show that while genetic drift affects neutral polymorphism, variability of molecular phenotype is constrained by purifying selection, the strength of which varies between TLR loci.

Our results also provide conservation genetic insight into the interplay of evolutionary processes that shape the diversity of important receptors for pathogen recognition. While the stochasticity of allele sampling generally increases with decreasing population size, our results suggest that the detrimental effect of fixing suboptimal (or poorly functioning) alleles in small populations can be mitigated by purifying selection, but not when the population falls to an extremely small size of a few dozen individuals. The efficiency of purifying selection might be actually enhanced in small populations by the effect of purging, as shown by Robinson et al. (2018). On the other hand, it should be noted that both increased intensity of purification and drift lead to a reduction in diversity (Charlesworth 2009). Thus, even if deleterious alleles are purged even in small populations, the future of these small populations is still precarious due to the resulting reduction in genetic diversity (Spielman et al. 2004).

Data Accessibility

The DNA haplotypes of TLR1B, TLR4 and TLR15 obtained by Illumina sequencing and subsequent bioinformatic analysis are available in Genbank under accession numbers MT260894 - MT260972.

Batch of supporting data have been published in a documented Figshare archive https://doi.org/10.6084/m9.figshare.12180027.v4 (Vlček et al. 2020). It contains DNA haplotypes obtained by original screening by Sanger sequencing, final genotype data for TLR1B, TLR4 and TLR15 linked with the DNA accession numbers, amplicon filtering statistics and amplicon frequencies, physicochemical properties of protein variants and their structural models and other properties.

Additional notes, figures and tables are accessible in the Appendix S1.

Scripts to generate genotypes from sequencing data and scripts to analyse the results are available at https://github.com/vlkofly/TLR-amplicon archived at Zenodo: https://doi.org/10.5281/zenodo.7248527 (Vlček 2022)

References

Alcaide, M., and S. V. Edwards. 2011. Molecular evolution of the toll-like receptor multigene family in birds. Molecular biology and evolution 28:1703–15.

Arbogast, B. S., S. V Drovetski, and R. L. Curry. 2006. The origin and diversification of Galápagos mockingbirds. Evolution 60:370–382.

Areal, H., Abrantes, J. and Esteves, P.J. 2011. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. BMC Evol Biol 11:368.

Azad, A. K., W. Sadee, and L. S. Schlesinger. 2012. Innate immune gene polymorphisms in tuberculosis. Infection and Immunity 80:3343–3359.

Bates, D., M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, et al. 2015. Fitting Linear Mixed-Effects Models Using {lme4}. Journal of Statistical Software 67:1–48.

Bateson, Z. W., S. C. Hammerly, J. A. Johnson, M. E. Morrow, L. A. Whittingham, and P. O. Dunn. 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.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics 10:195–205.

Clark, N. J., & Clegg, S. M. (2017). Integrating phylogenetic and ecological distances reveals new insights into parasite host specificity. Molecular Ecology, 26(11), 3074–3086. Portico.

Darfour-Oduro, K. a., H.-J. Megens, A. L. Roca, M. a. M. Groenen, and L. B. Schook. 2015. Adaptive Evolution of Toll-Like Receptors (TLRs) in the Family Suidae. PLOS ONE 10:e0124069.

Davies, C. S., Taylor, M. I., Hammers, M., Burke, T., Komdeur, J., Dugdale, H. L., and Richardson, D. S. 2021. Contemporary evolution of the innate immune receptor gene TLR3 in an isolated vertebrate population. Molecular ecology 30:2528–2542.

de Zoete, M. R., L. I. Bouwman, A. M. Keestra, and J. P. M. van Putten. 2011. Cleavage and activation of a Toll-like receptor by microbial proteases. Proceedings of the National Academy of Sciences 108:4968–4973.

DeCandia, A. L., A. P. Dobson, and B. M. VonHoldt. 2018. Toward an integrative molecular approach to wildlife disease. Conservation Biology 32:798–807.

Deem, S. L., P. G. Parker, M. B. Cruz, J. Merkel, and P. E. A. Hoeck. 2011. Comparison of blood values and health status of Floreana Mockingbirds (*Mimus trifasciatus*) on the islands of Champion and Gardner-by-Floreana, Galápagos Islands. Journal of wildlife diseases 47:94–106.

Egea, R., S. Casillas, and A. Barbadilla. 2008. Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic acids research 36:157–162.

Eichler W (1948) Annals and Magazine of Natural History. Annals and Magazine of Natural History, 1, 588–598.

Ferwerda, B., McCall, M. B. B., Alonso, S., Giamarellos-Bourboulis, E. J., Mouktaroudi, M., Izagirre, N., Syafruddin, D., Kibiki, G., Cristea, T., Hijmans, A., Hamann, L., Israel, S., ElGhazali, G., Troye-Blomberg, M., Kumpf, O., Maiga, B., Dolo, A., Doumbo, O., Hermsen, C. C., … Netea, M. G. (2007). TLR4 polymorphisms, infectious diseases, and evolutionary pressure during migration of

modern humans. Proceedings of the National Academy of Sciences, 104(42), 16645–16650.

Fiddaman SR, Vinkler M, Spiro SG, Levy H, Emerling CA, Boyd AC, Dimopoulos EA, Vianna JA, Cole TL, Pan H, Fang M, Zhang G, Hart T, Frantz LAF, Smith AL. 2022. Adaptation and Cryptic Pseudogenization in Penguin Toll-Like Receptors. Mol Biol Evol. 39:msab354.

Frankham, R. 2005. Genetics and extinction. Biological Conservation 126:131–140.

Galtier, N. 2016. Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis. PLoS Genetics 12:1–23.

Gavan, M. K., M. K. Oliver, A. Douglas, and S. B. Piertney. 2015. Gene dynamics of toll-like receptor 4 through a population bottleneck in an insular population of water voles (*Arvicola amphibius*). Conservation Genetics 16:1181–1193.

Gilroy, D. L., C. Van Oosterhout, J. Komdeur, and D. S. Richardson. 2017. Toll-like receptor variation in the bottlenecked population of the endangered Seychelles warbler. Animal Conservation 20:235–250.

Gonzalez-Quevedo, C., L. G. Spurgin, J. C. Illera, and D. S. Richardson. 2015. Drift, not selection, shapes toll-like receptor variation among oceanic island populations. Molecular Ecology 24:5852–5863.

Grantham, R. 1974. Amino Acid Difference Formula to Help Explain Protein Evolution. Science 185:862–864.

Grueber, C. E., G. P. Wallis, and I. G. Jamieson. 2013. Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. Molecular ecology 22:4470–82.

Grueber, C. E., G. P. Wallis, and I. G. Jamieson. 2014. Episodic positive selection in the evolution of avian toll-like receptor innate immune genes. PloS one 9:e89632.

Hartmann, S. a, H. M. Schaefer, and G. Segelbacher. 2014. Genetic depletion at adaptive but not neutral loci in an endangered bird species. Molecular Ecology 23:5712–5725.

Hoeck, P. E. A., J. L. Bollmer, P. G. Parker, and L. F. Keller. 2010. Differentiation with drift: a spatio-temporal genetic analysis of Galápagos mockingbird populations (*Mimus* spp.). Philosophical Transactions of the Royal Society B: Biological Sciences 365:1127–1138.

Iwasaki, A., and R. Medzhitov. 2010. Regulation of Adaptive Immunity by the Innate Immune System. Science 327:291–295.

Jiménez-Uzcátegui, G., W. Llerena, W. Bryan Milstead, E. E. Lomas, and D. A. Wiedenfield. 2011. Is the population of Floreana Mockingbird Mimus trifasciatus declining? Cotinga 33:34–40.

Jin, M. S., S. E. Kim, J. Y. Heo, M. E. Lee, H. M. Kim, S. G. Paik, H. Lee, et al. 2007. Crystal Structure of the TLR1-TLR2 Heterodimer Induced by Binding of a Tri-Acylated Lipopeptide. Cell 130:1071–1082.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge.

Khan I, Maldonado E, Silva L, Almeida D, Johnson WE, O'Brien SJ, Zhang G, Jarvis ED, Gilbert MTP, Antunes A. 2019. The Vertebrate TLR Supergene Family Evolved Dynamically by Gene Gain/Loss and Positive Selection Revealing a Host–Pathogen Arms Race in Birds. Diversity 11:131

Kloch, A., Wenzel, M. A., Laetsch, D. R., Michalski, O., Bajer, A., Behnke, J. M., & Piertney, S. B. 2018. Signatures of balancing selection in toll-like receptor (TLRs) genes–novel insights from a free-living rodent. Scientific Reports, 8:1-10.

Kosakovsky Pond, S. L., and S. D. W. Frost. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular biology and evolution 22:1208–22.

Králová, T., T. Albrecht, J. Bryja, D. Hořák, A. Johnsen, J. T. Lifjeld, M. Novotný, et al. 2018. Signatures of diversifying selection and convergence acting on passerine Toll-like receptor 4 in an evolutionary context. Molecular Ecology 27:2871–2883.

Kryazhimskiy, S., and J. B. Plotkin. 2008. The Population Genetics of dN/dS. PLoS Genetics 4:e1000304.

Kumar, H., T. Kawai, and S. Akira. 2011. Pathogen recognition by the innate immune system. International Reviews of Immunology 30:16–34.

Leveque, G., V. Forgetta, S. Morroll, A. L. Smith, N. Bumstead, P. Barrow, J. C. Loredo-Osti, et al. 2003. Allelic variation in TLR4 is linked to susceptibility to *Salmonella enterica* serovar typhimurium infection in chickens. Infection and Immunity 71:1116–1124.

Levy, H., S. R. Fiddaman, J. A. Vianna, D. Noll, G. V. Clucas, J. K. H. Sidhu, M. J. Polito, et al. 2020. Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird. Molecular biology and evolution 37:1708–1726.

Li, Y., M. Oosting, S. P. Smeekens, M. Jaeger, R. Aguirre-Gamboa, K. T. T. Le, P. Deelen, et al. 2016. A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans. Cell 167:1099-1110.e14.

Lovette, I. J., B. S. Arbogast, R. L. Curry, R. M. Zink, C. A. Botero, J. P. Sullivan, A. L. Talaba, et al. 2012. Phylogenetic relationships of the mockingbirds and thrashers (Aves: Mimidae). Molecular phylogenetics and evolution 63:219–29.

Lynch, M. 2007. The origins of genome architecture. Sinauer Associates, Sunderland.

Miller, H. C., and D. M. Lambert. 2004. Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). Molecular ecology 13:3709–21.

Minias P., Vinkler M. 2022. Selection balancing at innate immune genes: adaptive polymorphism maintenance in Toll-like receptors. Mol Biol Evol. msac102.

Mugal, C. F., J. B. W. Wolf, and I. Kaj. 2014. Why time matters: Codon evolution and the temporal dynamics of dN/dS. Molecular Biology and Evolution 31:212–231.

Mukherjee, S., N. Sarkar-Roy, D. K. Wagener, and P. P. Majumder. 2009. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. Proceedings of the National Academy of Sciences of the United States of America 106:7073–7078.

Murrell, B., S. Moola, A. Mabona, T. Weighill, D. Sheward, S. L. Kosakovsky Pond, and K. Scheffler. 2013. FUBAR: A fast, unconstrained bayesian AppRoximation for inferring selection. Molecular Biology and Evolution 30:1196–1205.

Nagai, Y., S. Akashi, M. Nagafuku, M. Ogata, Y. Iwakura, S. Akira, T. Kitamura, et al. 2002. Essential role of MD-2 in LPS responsiveness and TLR4 distribution. Nature Immunology 3:667–672.

Nelson-Flower, M. J., R. R. Germain, E. A. MacDougall-Shackleton, S. S. Taylor, and P. Arcese. 2018. Purifying Selection in the Toll-Like Receptors of Song Sparrows Melospiza melodia. Journal of Heredity 109:501–509.

Ohta, T. 1992. The Nearly Neutral Theory Of Molecular Evolution. Annual Review of Ecology and Systematics 23:263–286.

Park, B. S., D. H. Song, H. M. Kim, B.-S. Choi, H. Lee, and J.-O. Lee. 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature 458:1191–1195.

Quéméré, E., Hessenauer, P., Galan, M., Fernandez, M., Merlet, J., Chaval, Y., ... & Charbonnel, N. 2021. Pathogen-mediated selection favours the maintenance of innate immunity gene polymorphism in a widespread wild ungulate. Journal of Evolutionary Biology 34:1156-1166.

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

R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria.

Raven, N., S. Lisovski, M. Klaassen, N. Lo, T. Madsen, S. Y. W. Ho, and B. Ujvari. 2017. Purifying selection and concerted evolution of RNA-sensing toll-like receptors in migratory waders. Infection, Genetics and Evolution 53:135–145.

Richter, S., A. Wenzel, M. Stein, R. R. Gabdoulline, and R. C. Wade. 2008. webPIPSA: a web server for the comparison of protein interaction properties. Nucleic acids research 36:276–280.

Robinson, J. A., C. Brown, B. Y. Kim, K. E. Lohmueller, and R. K. Wayne. 2018. Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. Current Biology 28:3487-3494.e4.

Rousett, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8:103–106.

Ruan, W. K., Wu, Y. H., An, J., & Zheng, S. J. 2012. Polymorphisms of chicken Toll-like receptors 4, 15, and 21 in different breeds. Poultry science, 91:2512–2516.

Rudd, M. F., R. D. Williams, E. L. Webb, S. Schmidt, G. S. Sellick, and R. S. Houlston. 2005. The predicted impact of coding single nucleotide polymorphisms database. Cancer Epidemiology Biomarkers and Prevention 14:2598–2604.

Snell, H. M., P. A. Stone, and H. L. Snell. 1996. A summary of geographical characteristics of the Galápagos Islands. Journal of Biogeography 23:619–624.

Spielman, D., B. W. Brook, and R. Frankham. 2004. Most species are not driven to extinction before genetic factors impact them. Proceedings of the National Academy of Sciences of the United States of America 101:15261–15264.

Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. American journal of human genetics 68:978–989.

Sukumaran, J., and M. T. Holder. 2010. DendroPy: A Python library for phylogenetic computing. Bioinformatics 26:1569–1571.

Suzuki, R., and H. Shimodaira. 2006. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22:1540–1542.

Tschirren, B. 2015. Borrelia burgdorferi sensu lato infection pressure shapes innate immune gene evolution in natural rodent populations across Europe. Biology Letters 11:20150263.

Tschirren, B., M. Andersson, K. Scherman, H. Westerdahl, and L. Råberg. 2012. Contrasting patterns of diversity and population differentiation at the innate immunity gene toll-like receptor 2 (tlr2) in two sympatric rodent species. Evolution 66:720–731.

Velová, H., M. W. Gutowska-Ding, D. W. Burt, and M. Vinkler. 2018. Toll-Like Receptor Evolution in Birds: Gene Duplication,

Pseudogenization, and Diversifying Selection. (M. Yeager, ed.)Molecular Biology and Evolution 35:2170–2184.

Vinkler, M., H. Bainová, and J. Bryja. 2014. Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. Genetics Selection Evolution 46:72.

Vinkler, M., H. Bainová, A. Bryjová, O. Tomášek, T. Albrecht, and J. Bryja. 2015. Characterisation of Toll-like receptors 4, 5 and 7 and their genetic variation in the grey partridge. Genetica 143:101–112.

Vlček, J. 2022. vlkofly/TLR-amplicon: Scripts for a publication on Toll-like receptors in Galápagos mockingbirds. Zenodo. https://doi.org/10.5281/zenodo.7248527

Vlček, J., M. Miláček, M. Vinkler, and J. Štefka. 2020. Supplementary tables for: Effect of population size and selection on Toll-like receptor diversity in populations of Galápagos mockingbirds. figshare. https://doi.org/10.6084/m9.figshare.12180027.v4

Vlček, J., P. E. A. Hoeck, L. F. Keller, J. Wayhart, I. Dolinová, and J. Štefka. 2016. Balancing selection and genetic drift create unusual patterns of MHCIIβ variation in Galápagos mockingbirds. Molecular Ecology 25:4757–4772.

Walsh, C., M. Gangloff, T. Monie, T. Smyth, B. Wei, T. J. McKinley, D. Maskell, et al. 2008. Elucidation of the MD-2/TLR4 Interface Required for Signaling by Lipid IVa. The Journal of Immunology 181:1245–1254.

Wang, J., Z. Zhang, J. Liu, J. Zhao, and D. Yin. 2016. Ectodomain Architecture Affects Sequence and Functional Evolution of Vertebrate Toll-like Receptors. Scientific Reports 6:1–10.

Wertheim, J. O., B. Murrell, M. D. Smith, S. L. K. Pond, and K. Scheffler. 2015. RELAX: Detecting relaxed selection in a phylogenetic framework. Molecular Biology and Evolution 32:820–832.

Wickham, H. 2009. ggplot2: elegant graphics for data analysis. Springer New York.

Wikelski, M., J. Foufopoulos, H. Vargas, and H. Snell. 2004. Galápagos birds and diseases: Invasive pathogens as threats for island species. Ecology and Society 9:5.

Willi, Y., J. Van Buskirk, and A. A. Hoffmann. 2006. Limits to the Adaptive Potential of Small Populations. Annual Review of Ecology, Evolution, and Systematics 37:433–458.

Wlasiuk, G., Khan, S., Switzer, W. M., & Nachman, M. W. 2009. A history of recurrent positive selection at the toll-like receptor 5 in primates. Molecular biology and evolution 26:937–949.

Yang, J., and Y. Zhang. 2015. I-TASSER server: New development for protein structure and function predictions. Nucleic Acids Research 43:W174–W181.

Chapter IV

Genomes of Galápagos mockingbirds reveal the impact of island size and past demography on inbreeding and genetic load in contemporary populations

Manuscript to be submitted

Genomes of Galápagos mockingbirds reveal the impact of island size and past demography on inbreeding and genetic load in contemporary populations

Vlček Jakub[1,2,3,†], Espinoza-Ulloa Sebastian[4,5,†], Cowles Sarah[6], Ortiz-Catedral Luis[7], Coutu Cathy[8], Cháves Jaime[9], Jose Andrés[4], Štefka Jan[1,2,*]

1 Faculty of Science, University of South Bohemia, Branišovská 1670, 37005, České Budějovice, Czech Republic.

2 Biology Centre CAS, Branišovská 31, 37005, České Budějovice, Czech Republic

3 Charles University Faculty of Science, Department of Botany, Benátská 2, 128 01 Prague, Czech Republic

4 Department of Biology, University of Saskatchewan, Saskatoon, Canada

5 Pontificia Universidad Católica del Ecuador, Quito, Ecuador

6 University Miami, Department of Biology, Coral Gables, USA

7 Massey University, School of Natural Sciences, Ecology & Conservation Lab, Auckland, New Zealand

8 Agriculture & Agri-Food Canada, Saskatoon, Canada

9 Universidad san Francisco de Quito, Quito, Ecuador

* corresponding author; † shared first author position

Abstract

Restricted range size brings about noteworthy genetic consequences that may affect the viability of a population and eventually its extinction. Varying ranges and population sizes, however, interact with population genetic indices such as genetic diversity, inbreeding and genetic load in a non-trivial way due to stochasticity and different demographic history. Particularly, the question if an increase in inbreeding can avert the accumulation of genetic load via purging is hotly debated in the conservation genetic field. Insular populations with limited range sizes represent an ideal setup for an investigation relating range size to these genetic factors. Here we used a set of populations of Galápagos mockingbirds (*Mimus*) that occupy islands of different size ranging from 9.5 to thousands of hectares to investigate how island size shaped effective population size (N*e*), inbreeding, and genetic load. We assembled a genome of *M. melanotis* and genotyped 3 individuals per 8 populations by whole genome resequencing. We inferred population demography by several methods and show that the ancestral population colonised the archipelago 1 - 2 Mya, but surprisingly the N*e* of most populations was high during their insular history. The final decline in *M. parvulus* happened only 10 - 20 Kya while *M. trifasciatus* showed a longer history of N*e* below 10,000. Despite these historical fluctuations, current island size determines N*e* in a linear fashion. In contrast, significant inbreeding coefficients, derived from runs of homozygosity, were identified only in the 4 smallest populations. Additive genetic load, quantified as the number of non-synonymous derived alleles, indicated that purging might play a role in the case of *M. parvulus* as the smallest populations showed the lowest load. By contrast, *M. trifasciatus* carried both the highest additive and recessive genetic loads from all studied populations. This indicated less efficient purging despite higher autozygosity in this species, which might be linked with a recent rapid

bottleneck. Overall, our study demonstrates a complex interplay of area and population size variation on population genomic proxies of genetic drift and inbreeding that has implications in conservation genetics in general and in an ongoing conservation project of endangered *M. trifasciatus* in particular.

Keywords

Genetic diversity, Conservation genetics, Demographic inference, Genetic load

Introduction

Species inhabiting oceanic islands have played an important role in evolutionary biology with their distinctiveness inspiring the first evolutionary thoughts (Darwin 1859). They are particularly interesting from a conservation perspective due to their high speciation (Johnson & Stattersfiled 1990) and extinction rates (Ricketts et al. 2005). Compared to mainland populations, insular populations are usually smaller and thus more vulnerable to environmental and genetic stochasticity (Frankham 1997, Lande 1993). However, as populations of continental species are becoming fragmented and declining (Ceballos et al. 2017), the genetic consequences related to small population size can influence their survival as well. Understanding the genetic consequences of isolation and small range size in islands can thus provide us valuable insight into conservation genetics problems that may be relevant in the near future across the globe.

Insular populations are mechanistically limited by the area of the island they inhabit, and such constraint on population size has genetic consequences. The rate of change caused by genetic drift is inversely proportional to population size, so an insular populations' genepool will be affected by genetic drift to a higher degree compared to a continental

population (Frankham 1997). The first negative consequence of genetic drift in the insular population is a substantial loss of genetic diversity (Allendorf et al. 2013) which may reduce fitness and affect future adaptability (DeWoody et al. 2021, Spielman et al. 2004). The second negative consequence of genetic drift assumes the nearly neutral theory of molecular evolution (Ohta 1992), in which more slightly deleterious alleles accumulate in insular populations and this effect is proportional to population size due to a lower efficacy of purifying selection (Lanfear at al. 2014). This view is supported by the observation that insular populations have a higher non-synonymous to synonymous substitution rate compared to mainland populations (Woolfit & Bromham 2005, Leroy et al. 2021). In the long run, the accumulation of deleterious mutations can lead to mutational meltdown and extinction of the population (Lynch et al. 1995). A third detrimental genetic consequence is increased inbreeding in small populations due to the lack of unrelated mates (Woodworth et al. 2002). Inbreeding leads to the expression of recessive deleterious alleles, which increases mortality in the population (Charlesworth & Willis 2009). These three main issues may compromise insular populations from a genetic point of view, however their importance for the survival of a species and how they interact are a matter of heated debate (Willi et al. 2022, Teixeira & Huber 2021, Kardos et al. 2021, DeWoody et al. 2021, García-Dorado & Caballero 2021).

The main dispute is whether low neutral genetic diversity also indicates lowered fitness due to the accumulation of deleterious mutations and inbreeding depression (Teixeira & Huber 2021). The positive link between genetic diversity and fitness has been empirically observed in dozens of cases (DeWoody et al. 2021), however, several genomic studies showed that small populations with low genetic diversity were able to overcome the detrimental genetic effects related to their size (Robinson et al. 2022, Westbury et al. 2018, Robinson et al. 2016, Xue et al. 2015). Purging (Hedrick 1994, Crow 1970) was proposed as the mechanism to explain these observations, as increased inbreeding in small populations

exposes recessive deleterious alleles to purifying selection, which in turn becomes more efficient (Charlesworth 2018). Simulations (Kyriazis et al. 2019) as well as studies across several taxa (van der Valk et al. 2019) support the importance of purging in small populations and advocate indices of genetic load (e.i., net fitness cost of alleles under selection) as a more informative index of genetic health in contrast to neutral genetic diversity (Teixeira & Huber 2021).

In the context of this dispute, successful colonisation and long term survival of insular populations depends on the balance between accumulation and purging of deleterious alleles (Hedrick & Dorado 2016, James et al. 2016), if not driven to extinction due to demographic or environmental causes (Lande 1993). This balance is affected by multiple factors, of which effective population size (N*e*), reflecting the amount of genetic drift and its changes in the past, is one of the most important (Van Oosterhout 2020, Charlesworth 2009). Other factors like mutation rate, distribution of fitness effects of mutations, and each mutation's dominance also play a role in establishing the balance (Kyriazis et al. 2020, Mathur & DeWoody 2021, Robinson et al. 2016, van der Valk et al. 2019), however these factors are rather invariable within a species. Migration is also important as it can affect the balance either by inflating N*e* (Welles & Dlugosch 2018) or by the introduction of highly deleterious recessive alleles, which can cause extinction as in the case of the Royal Island wolves (Mathur & DeWoody 2021, van der Valk et al. 2019). In order to assess the prospects of insular or otherwise size restricted species, it is of uttermost importance to understand how genetic load is shaped by these aforementioned factors. It is not clear if purging is efficient enough to avert substantial accumulation of genetic load in small populations (Teixeira & Huber 2021): while some empirical evidence supports the significance of purging (Robinson et al. 2022, Westbury et al. 2018, Robinson *et al.* 2016), other studies do not (Kutschera et al. 2020, Kennedy et al. 2014, Ballou 1997).

Galápagos mockingbirds (*Mimus* spp.), in which populations occur on many differently sized islands of the Galápagos archipelago, offer a great opportunity to investigate the effect of dynamic demography and restricted population size on inbreeding and genetic load accumulation. The Galápagos mockingbirds are a monophyletic group that consist of four species, of which three are restricted to a single island and surrounding islets (Española – *M. macdonaldi*, San Cristóbal – *M. melanotis* , two islets around Floreana – *M. trifasciatus*) and the fourth Galápagos mockingbird (*M. parvulus*) is distributed in almost all the other islands of the archipelago (Arbogast et al. 2006, Lovette et al. 2012). Mitochondrial DNA analyses suggest the one ancestral mockingbird population colonized the archipelago and began its diversification between 1.6 – 5.5 Mya, thus being of a similar age as the formation of the oldest islands of the current archipelago configuration (Arbogast et al. 2006, Geist et al. 2014). The colonisation is deemed to follow the progression rule (Shaw & Gillespie 2016), putatively starting in the southeast, where phylogenetically basal *M. macdonaldi* and *M. melanotis* occur, with colonisation reaching northwest islands later with *M. trifasciatus* and *M. parvulus* as terminal sister clades (Nietlisbach et al. 2013, Štefka et al. 2011, Arbogast et al. 2006).

Despite living on islands 1,000 km away from the shore of Ecuador, mockingbirds are characterised as terrestrial species with poor long-distance flight ability. Hoeck et al. (2010b) showed that migration events between islands are rare, based on microsatellite markers. In this way, each population, even within *M. parvulus* and *M. trifasciatus*, are diverging mainly due to the effects of genetic drift (Hoeck et al. 2010b). Each population is therefore bound to a particular island, whose geographic size determines their carrying capacity and therefore their N*e* (Hoeck et al. 2010b). Importantly, each island has a unique geological history and have undergone historical changes in its shape and size caused mainly by volcanic activity, erosion and eustasy (sea level

changes) (Geist et al. 2014, Ali & Aitchison 2014). Moreover, during the Pleistocene glaciation periods some islands were interconnected (Ali & Aitchison 2014). We hypothesise that such changes had a significant impact on historical range size, population structure, and gene flow in Galápagos mockingbirds, which in turn has shaped N*e*, inbreeding, and genetic load in each population.

Here, we examined the effects of dynamic demography and restricted population size on inbreeding and genetic load accumulation in mockingbirds from the Galápagos. We first inferred past demography of two populations of the critically endangered *M. trifasciatus* with a current census population size below 200 individuals (Jiménez-Uzcátegui et al. 2010) and six populations of *M. parvulus*, including the small and distinct northernmost populations of Darwin and Wolf that were not included in a previous microsatellite study (Hoeck et al. 2010b). Second, we generated an original reference genome and, together with the representatives of the four small populations, we resequenced genomes of individuals from intermediate and large *M. parvulus* populations to test the effect of current island size on N*e*, inbreeding, and genetic load. Third, we linked the demography of selected populations with geological history to bring novel insights about the colonisation history of *M. parvulus* and *M. trifasciatus*. Finally, we interpreted the amount of genetic load and inbreeding with respect to the past demography, in order to understand how the indices were shaped by dynamic micro-evolutionary forces in extremely small insular populations.

Methods

De-novo genome assembly and annotation.

We assembled a reference genome using 10x Genomics Chromium technology. Firstly, we prepared a DNA library according to the manufacturer's protocol from a blood sample of a single individual of a San Cristóbal mockingbird (*Mimus melanotis*) (see Genome supplementary notes). High molecular weight DNA was extracted using MagAttract HMW DNA Kit (Qiagen). 10x Genomics library construction was set to have an effective coverage of 56X based on paired-end (PE) short-reads of 150 bp. The library was sequenced on a single Illumina-Nextseq500 run at the Cornell Bio-Resource Center (BRC) in paired-end mode. With the resulting four sequenced libraries, we performed a quality control in FastQC for each PE library. We looked for any signal of adapters initially with FastQC (ver. 0.11.3; Andrews, 2010) and later we ran Trimmomatic (ver. 0.39; Bolger et al., 2014) to remove them using default settings. Then, we used all four PE libraries to assemble the genome using the 10x Supernova assembler (Weisenfeld et al. 2017) and later applied Satsuma (ver. 3.1.0; Grabherr et al., 2010) to assemble Supernova scaffolds based on a synteny with the collared flycatcher’s (*Ficedula albicollis*) genome (FicAlb1.5) to get pseudo-chromosome level genome assembly. Using this final assembly, we proceeded to perform the gene annotation with Maker (Cantarel et al. 2008), implementing both *ab initio* prediction and homology-based gene annotation. The completeness of the assembled genome was assessed by the presence of nearly universal lineage-specific single-copy orthologous genes using BUSCO (ver. 3.1.0; Simão et al., 2015) and the Aves obd9 gene set from the Metazoa database (OrthoDB v10; Kriventseva et al., 2019). All genome methodology is described in detail in Genome supplementary notes.

Population sampling and variant calling.

In order to understand the effect of island size on inbreeding and genetic load, we selected eight allopatric populations inhabiting differently sized islands (Figure 1, Table S9). For the widespread *M. parvulus* we selected Isabela (**I**) and Santa Cruz (**S**) populations with nominal subspecies as the largest geographic size category (tens of thousand hectares). Populations of *M. p. personatus* on Pinta (**P**) and Marchena (**M**) were selected as an intermediate island size category (thousand hectares). *M. p. hulli* and *M. p. wenmani* occupy the northernmost Galápagos Islands Darwin (**D**) and Wolf (**W**) respectively and they fall in a small island size category (lower than 130 ha) together with two remnant populations of *M. trifasciatus*. *M. trifasciatus* is considered a sister species of *M. parvulus* based on majority of the studies (Nietlisbach et al. 2013; Štefka et al. 2011) and occurs on south-west islets Champion (**C**) and Gardner by Floreana (**F**). Three unrelated individuals per population were selected based on microsatellite data (Vlček et al. 2016). Most of the populations are rather panmictic based on microsatellite data, except for the largest islands (Hoeck et al. 2010b). Therefore, for Isabela we selected individuals from Alcedo deme and also for Santa Cruz we selected individuals from the southern deme around Puerto Ayora to overcome potential Wahlund effect. We also selected three samples of *M. polyglottos* from the Miami (USA) area as an outgroup. For sampling details see Supplementary notes.

We extracted genomic DNA using MasterPure DNA Purification Kit (Lucigen). PCR-free sequencing libraries were prepared by the Norwegian Sequencing Centre and sequenced on six lanes of Illumina HiSeqX10 in a pair-end mode. We used Trimmomatic (v. 0.36, Bolger et al. 2014) to remove adapters and trim low quality base pairs from raw

reads. Quality filtered reads of at least 120bp (minimum base quality Phred score of 15) were mapped onto our *M. melanotis* reference genome by BWA – mem (v. 0.7.15) (Li & Durbin 2009). We removed duplicated reads by Picard (v. 2.8.1) (Broad Institute 2018). Variant calling was done by GATK (v. 3.7) (McKenna et al. 2010) with a series of stringent filtering steps to minimise the presence of false genotypes. Reliable biallelic SNPs were obtained following the GATK best practices for hard filtering as described in DePristo et al. (2011). Variants failing the following filters were excluded (QD < 8 || FS > 15 || MQ < 50 || MQRankSum < −5 || RPRSlow < −8.0 || RPRShigh > 3 || SOR > 3.0). We also excluded sites that either exceeded average coverage depth per individual by more than two standard deviations, showed excess heterozygosity ($p < 0.05$ as determined by VCFTools (v. 0.1.16; Danecek et al. 2011), or had more than 25% missing data. We also excluded sites found within repetitive regions and, finally, invariant sites with low quality (QUAL < 15). Such filtering resulted in a vcf file containing ~8 M bi-allelic SNPs and ~879 M invariant sites. Additional filters were applied to retain only reliable sites. We removed genotypes with depth of coverage lower than 7 and for most of the analyses we used only sites with no missingness. We used only sites on autosomes, i.e. removing sites that mapped on the Z chromosome. The neutralome region was defined as a union of non-coding regions (10kbp offset from a coding region) and non-coding conserved sites (100 Kbp offset) that were lifted over from *F. albicollis* (Craig et al. 2018).

Phylogenetic relationships and population structure

We inferred the phylogenetic relationships between populations using several methods. First, we constructed a maximum likelihood tree based on 7245 neutral SNPs using SNPhylo (Lee et al. 2014). The neutralome SNPs were pruned for linkage disequilibrium (ld threshold 0.2) and minor allele frequency (MAF) 0.1. The dataset included three *M. polyglottos* individuals as an outgroup. Confidence of the nodes was calculated based

on 1,000 bootstrap iterations. Second, because reticular interconnections are often common between connected and/or recently split populations we also constructed a split network using Nei's genetic distances, as implemented in StAMPP (Pembleton *et al.* 2013), and the Neighbour-Net algorithm of Splitstree (Huson & Bryant 2006). Third, we estimated the relative contributions of all possible topologies among islands based on nuclear genomes using TWISST (Martin and Belleghem 2017). After phasing all the autosomal SNPs (details in Supplementary notes) we computed ML trees in 100 SNPs windows using PhyML. Finally, we used topology weighting to compute the proportion of subtrees matching each possible topology. TWISST analyses were implemented according to the author's guidelines in a custom script. We inferred population structure based on 16,632 neutral LD-pruned SNPs with MAF 0.05 using a principal component analysis (PCA) in the program Snprelate (Zheng et al. 2012) and using clustering analysis in the program ADMIXTURE V. 1.3.0 (Alexander & Novembre 2009) assuming 1 to 9 clusters (K). Cross-validation error rate was used to identify the number of clusters that best matched genetic structure in our data.

Demographic inference

We inferred the demographic history of selected populations using four different approaches. First, we used Multiple Sequentially Markovian Coalescent (MSMC2) to calculate the coalescent rate and thus effective population size (N*e*) in the past (Schiffels & Durbin 2014). To minimise the potential bias caused by non-neutral molecular processes, we used only the neutralome, as defined above. We ran MSMC2 with time segments 1*2+15*1+1*2 and the confidence interval of time estimates was calculated based on 20 bootstrap iterations. Second, similarly we used another Markovian coalescent method SMC++ (v1.15.2; Terhorst et al. 2017) with individuals pooled per population. Third, as an alternative, to estimate both N*e* and the coalescence rates between populations (that reflect past connectivity of the populations), we used the program Relate

(Speidel et al. 2019). We ran Relate (--mode All) to estimate genome-wide genealogies with population size -N 1,000 and mutation rate -m $4.6x10^{-9}$ (direct estimate from relatively closely related *F. albicollis* (Smeds et al. 2016) and using a polarised genome (details in Supplementary notes). Coalescence rates between and within populations were estimated using 9 iterations and generation time was set to 4.5 based on the empirical estimate by Grant et al. (2000). Initially, we ran Relate for each chromosome separately, and subsequently, we obtained joint estimates for all autosomal chromosomes excluding Chromosome 4a, Z and shorter scaffolds. Relative pairwise coalescence values were calculated as the proportion of the cross-coalescence rate between populations and the intra-coalescence rate for the larger population from each pair.

Fourth, complementary demographic analyses were carried out using G-PhoCS (Generalized Phylogenetic Coalescent Sampler; Gronau et al. 2011) to determine ancestral population sizes, divergence times, and post-divergence migration between populations. To generate G-PhoCS input, genomic regions with the least constraint from non-neutral genomic processes (selection, linked selection and GC-biased gene conversion) were selected to avoid inference bias (Gronau et al. 2011). First of all, we used neutralome as defined above. In order to avoid the effect of within locus recombination, we filtered these loci by size with lengths between 1,300 and 1,500 bp (based on LD analysis) resulting in 8,429 neutral loci. This final input was used in G-PhoCS in combination with the most supported phylogenetic topology obtained from our previous phylogenetic analysis (Figure 1A) as a backbone. To explore post-divergence migration events between populations, the migration bands were tested bidirectionally between all *M. parvulus* populations (Figure SN2). Since G-PhoCS lacks statistical tests for output parameters, nine independent runs were performed based on nine subsets of migrations models (Figure SN2). Default parameters described by Gronau et al. (2011) were used to run the analysis (α= 1.0 and a β = 10,000 for

mutation-scaled population sizes and divergence times, and α = 0.002 and β = 0.00001 for mutation-scaled migration rates). The finetunes argument was set on TRUE in the control file, and the MCMC was set to run 200,000 iterations. The first 100,000 were burn-in iterations, leaving the last 100,000 iterations for analysis; the analysis samples were taken each ten iterations. Additionally, *tau-theta-print*=10,000 was used to scale up the raw estimates because the smallest estimates were below the resolution threshold. Later, to calibrate G-PhoCS raw estimates, we adopted the approach by Freedman *et al*. (2014), and we used the mutation rate of related species *F. albicollis* (μ=4.6E-9; Smeds et al. 2016) and the estimated Galápagos mockingbirds generation time of 4.5 years (Grant et al. 2000).

Effect of island size on genetic diversity, inbreeding and genetic load

To assess the effect of island size on genetic diversity, inbreeding, and genetic load, we used regression models fitting various population genetic indices by island size. We calculated the number of heterozygous positions per callable site within the neutralome as a measure of neutral genetic diversity per individual. We calculated population nucleotide diversity in a window based analysis using a 10kb window size in VCFTools (see details in Supplementary notes). Individual level of inbreeding was assessed by analysis runs of homozygosity (ROH) identified using the BCFtools/ROH method in bcftools v. 1.9 (Narasimhan et al. 2016). ROHs with a quality lower than 30 were removed. The inbreeding index (F_{ROH}) was calculated based on the BCFtools result as a fraction of the genome covered by long ROHs over 1 million base pairs (Mbp) (McQuillan et al. 2008). As an alternative approach to estimate the effect of inbreeding, we calculated linkage disequilibrium for each population using PopLDdecay v3.41 (Zhang et al. 2019) with default parameters. We also analysed the effect of island size on these population genetic indices in the same way as above.

There are several indices of genetic load that capture the potential costs of selection (Bertorelle et al. 2022). We selected two relatively straightforward indices that reflect two modes of dominance and assume that the majority of non-synonymous mutations are partially deleterious (Eyre-Walker & Keightley 2007). First, we used the number of derived non-synonymous alleles per individual, assuming an additive effect of deleterious alleles (homozygous genotypes are counted twice). We are aware that the majority of deleterious alleles could be recessive (Crow 1993, Huber et al. 2018) and therefore this 'additive load' reflects both the masked potential of deleterious alleles as well as their realised effect (i.e. total load in the sense of Bertorelle et al. 2022). Second, we used the total number of homozygous genotypes with derived non-synonymous allele per individual (counting only homozygous genotypes). If we assume that deleterious mutations are recessive, this index reflects realized or recessive load (Simons & Sella 2016, Bertorelle *et al.* 2022). As a distribution of dominance is a rather unknown factor in genetic load estimation, these two indices reflect two edge cases. For the load calculation we inferred allelic state (derived or ancestral) by Est-sfs v. 2.03 (Keightley & Jackson, 2018) using information from common starling (*Sturnus vulgaris*) and northern mockingbird (*M. polyglottos*) as outgroups (See details of polarisation in Supplementary notes). The effect of derived alleles was analysed by Snpeff 4.3 (Cingolani et al. 2012) and the counts of derived alleles were normalised by the number of callable sites to account for the variance in sequencing coverage.

Finally, we carried out a regression between the island size and the aforementioned indices in R ver. 3.5 (R core team 2018). A linear regression model was used for all indices as they all showed relatively normal distribution. In the case of inbreeding coefficient, we found an exponential relationship with island size and therefore we used a nonlinear regression model (Nonlinear Least Squares model).

Results

Genome assembly, annotation and re-sequencing

We assembled the genome of *M. melanotis* and subsequently called genotypes of selected 27 individuals of *M. parvulus*, *M. trifasciatus* and *M. polyglottos*. For the genome assembly, we generated ~830 million paired-end (PE) reads with a mean length of 139.5 bp resulting in a genome effective coverage of 42.4X. We generated initial genome assembly consisting of 830 long scaffolds (>10kb), and an N50 of contig and scaffold sizes of 74.5 Kb, and 7.55 Mb respectively. The total length of our assembled genome was ~980 Mb. Low fragmentation of the draft allowed us to assemble the genome to a chromosomal level by whole-genome synteny alignment using the genome of *F. albicollis* (FicAlb1.5) as a reference. Chromosomal assembly resulted in 29 autosomal pseudo-chromosomes, a pseudo-chromosome Z, a single linkage group (LGE22), and four super-scaffolds. Our annotation approach established the homologous location of 15,750 genes along the chromosomes with a gene length mean of ~ 11.5 Kb. Almost all these annotations (~99%) had Annotation Edit Distance (AED) scores of ≤ 0.5, revealing a high accuracy in the prediction of gene models exclusively based on homology. After a transcriptome-based gene model adjustment and prediction of unidentified genes, the final assembly resulted in 18,669 genes with a mean gene length of 11.7 Kb. From 18,669 gene models, 95,4% were under the 0.5 AED score. BUSCO assessment showed 89.21% gene annotation completeness, the rest of the gene annotations were either fragmented (4.4%) or missing (6.44%). The amount of repetitive elements was 6.59%. All assembly and annotation results are described in detail in Genome supplementary notes.

High-throughput sequencing and variant calling pipeline resulted in a reliable genotype dataset for all 24 individuals of *M. parvulus* and *M.*

trifasciatus and three individuals of *M. polyglottos*. Our final dataset consisted of 8.6M biallelic SNPs with an average coverage of 24.8X, and the sequencing data covered 85.9% of all sites in our reference genome (details of resequencing and variant calling in Supplementary Table S1 and S2).

Phylogenetic relationships and population structure.

The maximum likelihood tree showed *M. trifasciatus* and *M. parvulus* to form reciprocally monophyletic groups, with two highly (BS>99) supported clades within the latter species (Figure 1C). The first clade within *M. parvulus* includes individuals from Pinta, Marchena and Santa Cruz islands (PMS clade), while the second clade groups Isabela samples with the northern islands of Darwin and Wolf (DWI clade). The relationships among islands in the first clade ((M, S) P) are highly supported (BS ≥ 97), with the topology of the second clade not well resolved (BS < 80). The network analysis (Figure 1B) revealed small basal reticulations restricted to the DWI clade. Twisst analysis showed that phylogenetic relationships within PMS and DWI clades are highly unresolved with almost equal proportions of genomic windows supporting all possible topologies within both clades (Figure S1, S2). In spite of the incomplete lineage sorting, ADMIXTURE analysis supported our initial hypotheses that each population forms a distinct cluster, suggesting the populations represent a set of discrete units characterised by different frequencies of segregating alleles (Figure 1A). The model with the lowest cross validation error (0.65), K=7, shows each population as a group, with the exception of Santa Cruz and Marchena which were merged. However, the model in which each of the islands represents a unique genetic cluster (K=8) exhibited the second lowest error rate (0.70), markedly lower than any other K values (>0.75). The fact that each island forms a distinct population was also supported by clustering samples from each island into discrete clusters in a principal component analysis (Figure S4).

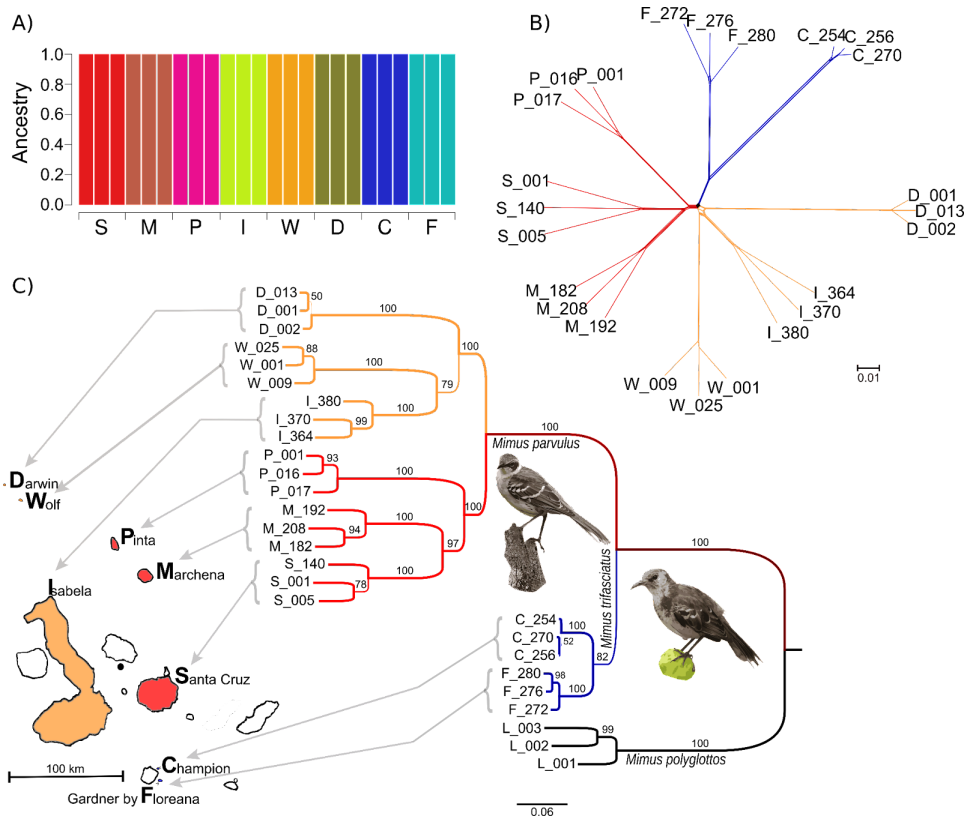


Figure 1 A) Population structure bar plot at K=8 generated by ADMIXTURE. Individual bars represent ancestry proportions. In our case populations form completely non-admixed demes B) Split network of mockingbird individuals based on Nei's genetic distances built by the Neighbour-Net algorithm. C) Maximum likelihood tree for *Mimus* mockingbirds based on 7,245 unlinked putatively neutral SNPs. Bootstrap values above each branch are based on 1,000 iterations. Three major clades are recognized: the *M. parvulus* north-west populations clade (orange - DWI), the *M. parvulus* east populations clade (red - PMS), and the *M. trifasciatus* clade (blue).

Demography

All three methods we used to estimate past effective population size (N*e*) (i.e., Relate, MSMC2 and SMC++) pointed to a relatively similar result (Figure 2, Figure S5, Figure S6), concerning colonisation bottleneck. Between 10 – 1 Mya, all populations followed an almost identical N*e* trajectory with a potential colonisation bottleneck between 1–2 million years ago (Mya) (Figure 2A). In that period, N*e* fell from around 100,000 to 50,000. Shortly after, populations started to diverge in their N*e* trajectories. In all *M. parvulus* populations, there is an increase in size following the bottleneck (in some cases back to 100,000), but later around 100 thousand years ago (Kya) all populations except Santa Cruz and Isabela start to decline. In *M. trifasciatus* there is evident decline right after the colonisation bottleneck and further decline follows in more recent time (50 Kya or 10 Kya depending on the method). The relative cross-coalescence rate between populations showed one panmictic population before 1 Mya where the cross-coalescent rate equals the population coalescent rate. But after 1 Mya the cross-coalescent rate rapidly decreases indicating decreasing connectivity between islands, with the latest split occurring between the two populations of the *M. trifasciatus* around 100 Kya (Figure 2).

Recent N*e* estimates were slightly different across all methods used, however the general pattern of reduction of size in the last 100 Kya was recovered by each of the approaches. While Relate showed a gradual decrease in population size between 50 to 3 Kya followed by a stabilisation towards the present, SMC++ suggested a more rapid decrease especially in Isabela and Santa Cruz starting in the last 10 Kya. Pinta, Marchena, Wolf and Darwin populations showed an increase in population sizes between 300 to 10 Kya based on SMC++ (Figure S6). Estimates based on MSMC are also more erratic towards present. N*e* estimates in small populations could not be estimated beyond 200 Kya,

most likely due to the limited amount of genealogical information and the methodological limitations. Importantly, the harmonic mean of N*e* estimated using *Relate* showed a positive correlation with island size (Pearson's rho 0.66). In addition, the harmonic means estimates of N*e* were also congruent with G-PhoCS (Table S9).

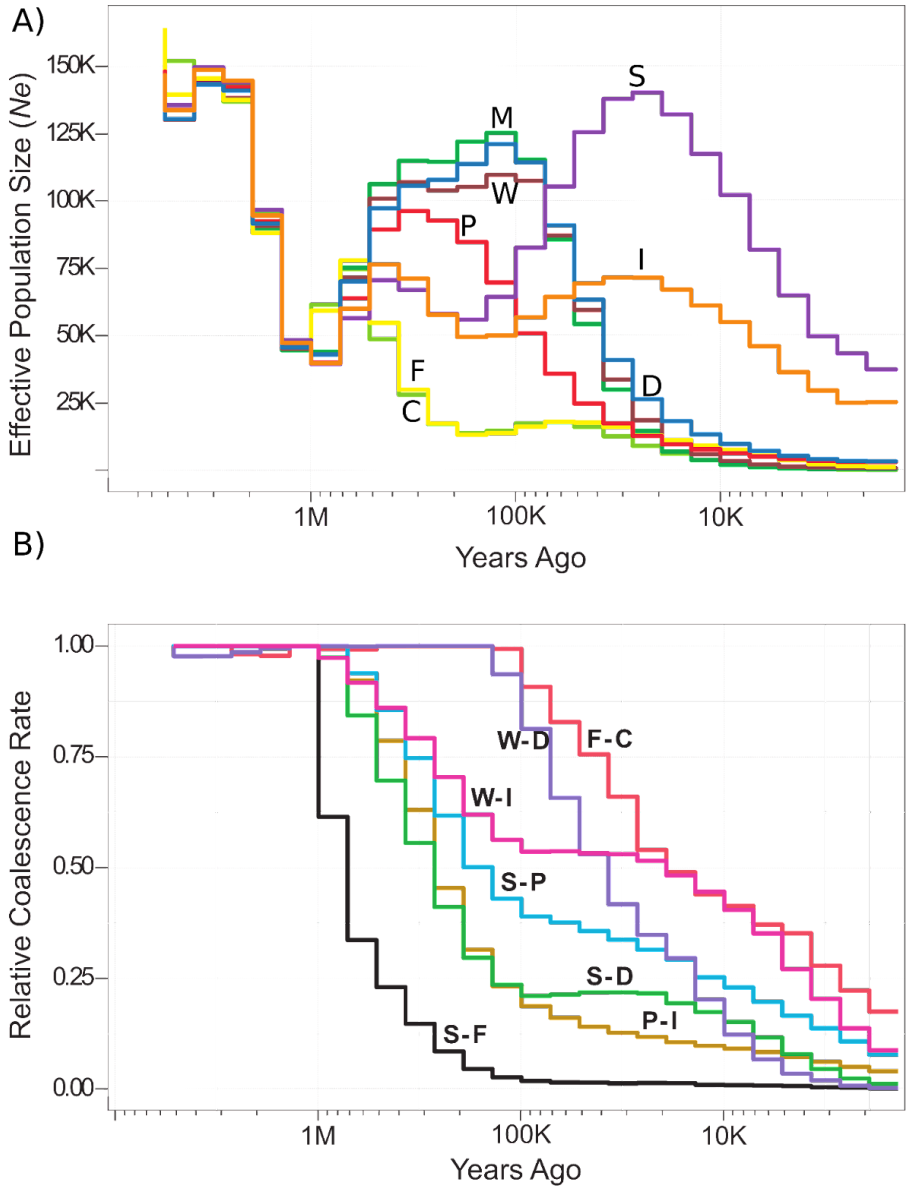


Figure 2 Demographic history inferred by the program Relate for each mockingbird population and their. A) Estimate of N*e* based on the joint information from all autosomal chromosomes (results from individual

chromosomes in Figure S7). Each line in the graph represents a different population as indicated in the legend. B) Past levels of interconnection between a subset of population pairs. The interconnection is represented by the ratio of within-population coalescent rate over between-population coalescent rate. Equity between these rates before 1 Mya indicates that the pair formed a panmictic population while the decrease towards the present indicates increasing separation. Relative coalescent rates between all the pairs are available in supporting information (Figure S8). Population abbreviations are as in Figure 1.

While Relate showed a continual decrease of relative cross-coalescent rates (connectedness) of populations (Figure 2B), with G-PhoCS we were able to estimate the split times more precisely and also analyse subsequent migration rates. Importantly, G-PhoCS estimates for theta (θ) and tau (τ) were consistent for all migration models tested, indicating good reliability of the obtained results. Effective population size of the past ancestral populations estimated by G-PhoCS roughly followed the trajectories from Relate and SMC++ (Figure 3). Once large ancestral population gave rise to smaller populations with N*e* tightly linked with island size (Figure 3b). Time splits of the populations are also consistent with the sequence of splits indicated by Relate. The split between *M. parvulus* and *M. trifasciatus* occurred 428 (±10) Kya, with the first divergence within the *M. parvulus* clade estimated at 64 (±3) Kya, splitting the central populations (PMS clade) and the western populations (DWI clade). Immediately after, 62 (±3) Kya, the PMS clade diverged resulting in the current population of Santa Cruz and the ancestral population for Pinta and Marchena. At 43 (±4) Kya, the DWI clade diverged giving rise to the current populations on Isabela and the ancestral populations on Darwin and Wolf. The splits between Pinta and Marchena and Darwin and Wolf occurred 15 (±2) and 5 (±1) Kya, respectively (Figure 3c). Finally, the most recent split occurred between the remnant populations of *M. trifasciatus* Champion and Gardner by Floreana 247 (±129) years ago. Interestingly, post divergence migration

rates were extremely low (Table S8) with the total migration rate for all migrations averaging lower than 10^{-3} migrants per total duration of a migration band.

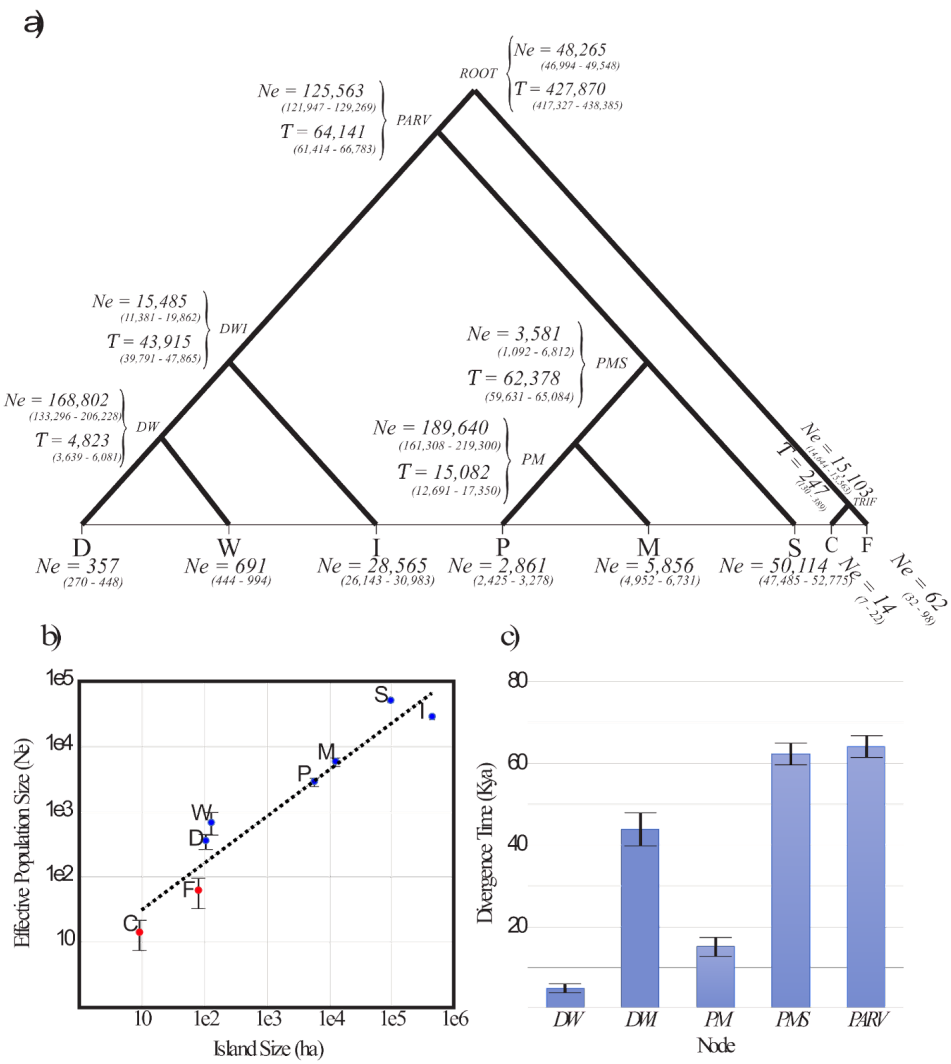


Figure 3 Overall estimates from all G-PhoCS tested models for population size and divergence. A) Calibrated estimates for population size and divergence times within the tested phylogenetic topology. B)

Correlation between the effective population size of the current populations and island size ($p<0.000$, $R=0.953$). C) Divergence times within the *M. parvulus* clade. All values are shown in Tables S3–S8.

Island size determines, genetic diversity, inbreeding and genetic load

We found that island size was a significant predictor of all indices of genetic diversity, inbreeding, and genetic load. Neutralome heterozygosity spanned two orders of magnitude from 4.3e-5 in Champion to 1.3e-3 in Santa Cruz and positively correlated with island size ($p<0.000$, $R=0.89$, Figure 5). This result is mirrored by genome-wide nucleotide diversity ($p<0.000$, $R = 0.93$, Figure S9). On the other hand, inbreeding inferred by runs of homozygosity showed a negative correlation with island size and the relationship was exponential, with small island, Champion characterised by extreme levels of inbreeding ($F_{ROH} = 0.62$). Other small populations (F, D, W) also showed high levels of inbreeding ($F_{ROH} = 0.22, 0.23, 0.12$, respectively) with intermediate and large populations showing almost zero inbreeding ($F_{ROH} < 0.009$). The only birds with a ROH over 3 Mbp are those of the small islands (C, F, D, and W). On Champion, individuals contained between 75–100 long ROH (>2 Mbp) on average, whereas birds from Darwin and Wolf contained 26 ROH on average per individual. These later populations, on the other hand, were characterised with a high frequency of shorter ROHs (0.1–0.5 Kbp) on average 1224 per individual (Figure 4). Analysis of linkage disequilibrium showed that the amount of linkage is correlated with island size ($R= -0.905$, $p < 0.01$, Figure S10).

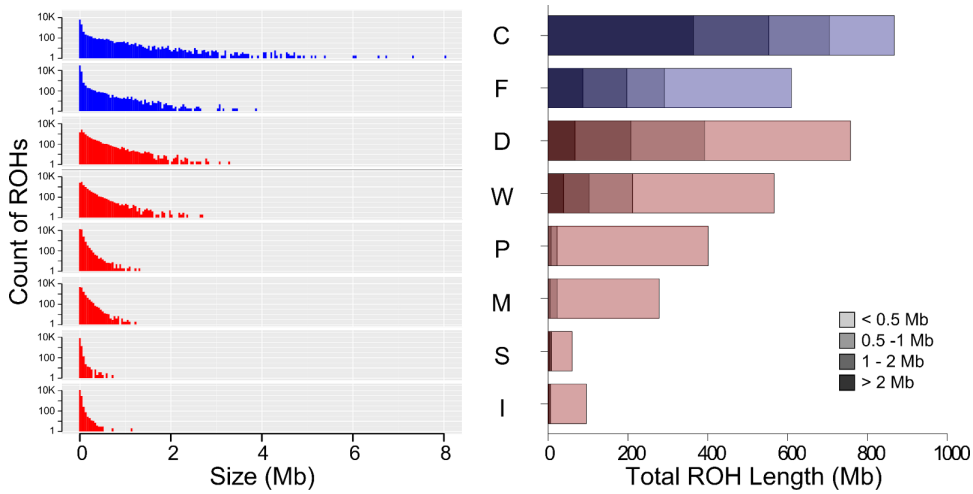


Figure 4 Runs of Homozygosity (ROH) distribution by island (red: *M. trifasciatus* populations, blue: *M. parvulus* populations). A) Distributions are arranged by island size with largest islands on the top. Island name as follows: C: Champion, F: Gardner by Floreana, D: Darwin, W: Wolf, P: Pinta, M: Marchena, S: Santa Cruz, I: Isabela) B) Total ROHs stacked bars for each island divided in four length categories each (light to dark: <0.5, 0.5–1, 1–2, >2 Mbp) longest ROHs category correspond to >2 Mbp).

We observed a negative correlation between the number of derived homozygous non-synonymous genotypes and island size; in other words recessive genetic load decreases with island size (Figure 5, $p<0.000$, R-adj=0.89). Conversely, the total number of non-synonymous derived alleles showed a positive correlation with island size when *M. trifasciatus* was excluded ($p<0.000$, R-adj=0.62), indicating that additive genetic load increases with population size. *M. trifascitus* showed similar high values, which on average overcomes the highest value estimated for *M. parvulus* (Figure 5D).

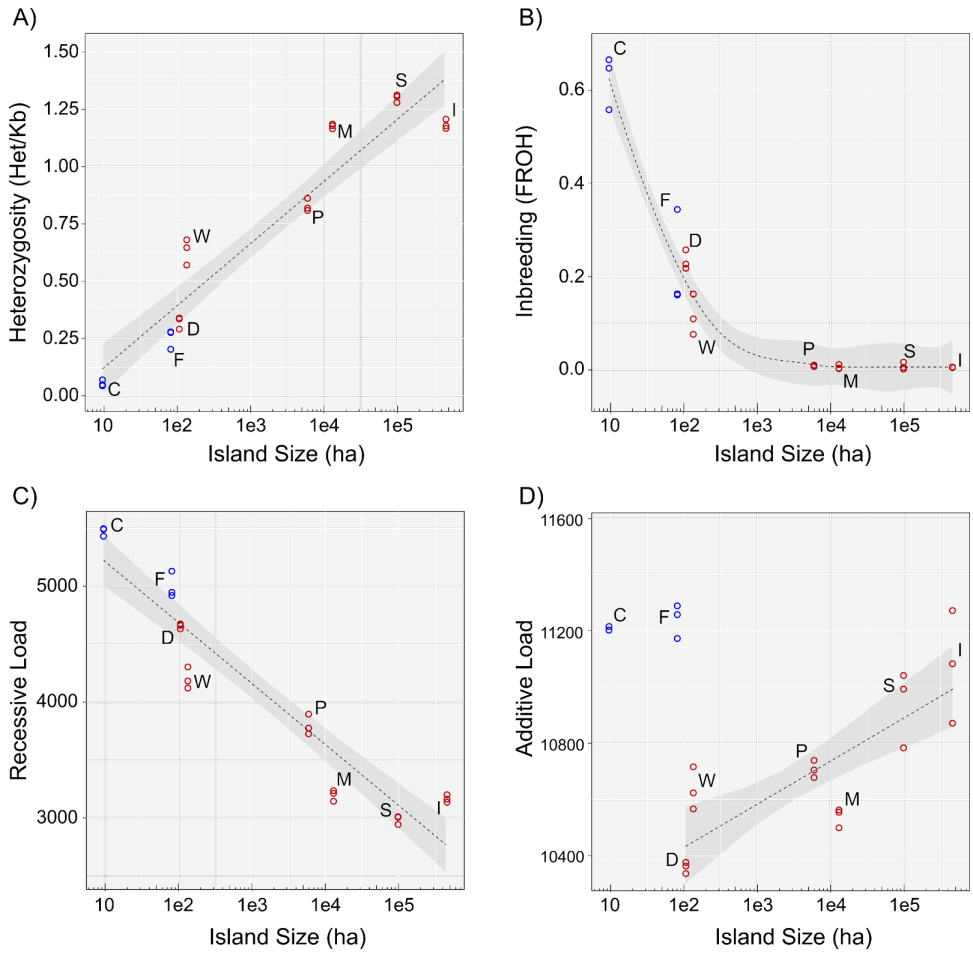


Figure 5 Regression models between individual genetic indices and island size. A) Mean number of heterozygous SNPs per kb). B) Inbreeding coefficient based on the frequency of long runs of homozygosity (> 1Mbp) in the genome; C) Recessive load: the total number of homozygous genotypes with a derived non-synonymous allele; D) Additive load: the total number of derived non-synonymous alleles (*M. trifasciatus* populations were excluded from additive load regression

model). For reference, each species was colour coded, red for *M. trifasciatus* and black for *M. parvulus*.

Discussion

We generated one high quality reference genome and 27 resequenced genomes to investigate genetic and demographic indices across the populations of mockingbirds from the Galápagos inhabiting island of different sizes. Analyses of past effective population size (N*e*) revealed a bottleneck that can be putatively linked to colonisation of Galápagos. Subsequently, populations diverged from each other, and N*e* decreased to the current state, interestingly, not directly following the colonisation event but later, within the last 100–20 Kya. We found that nucleotide diversity and current N*e* is correlated with island size; however, traces of substantial inbreeding in the genome are noticeable only in the smallest populations of *M. parvulus* (D, W) and *M. trifasciatus* (C, F). Interestingly, additive load is lowest in the smallest *M. parvulus* populations suggesting an effect of purging in these populations. Overall, our study brings a novel perspective to the colonisation history of the mockingbirds of the Galápagos and sheds light on the genomic consequences of evolution on isolated oceanic islands.

Demography and colonisation history

Colonisation of an oceanic island is a crucial period in the evolutionary history of an insular population. This event is characterised by a severe population bottleneck resulting from the initial founder event, that usually leaves a noticeable genomic footprint. Several genetic studies have investigated the colonisation history of the Galápagos mockingbirds. Arbogast et al. (2006) based on mtDNA data, showed that the origin of all four species is the result of a single colonisation of the archipelago by one ancestral mockingbird population 1.6–5.5 Mya,. Additionally, Nietlisbach et al. (2013) reported the first split between *M. melanotis–M. macdonaldi*

and *M. parvulus–M. trifasciatus* to have occurred between 146 Kya–1.4 Mya. Our genomic data corroborate the hypothesis of this single origin. Although we included only two out of the four species genomic data from just a few individuals can reveal past demography, because they contain not only the traces of DNA mutation, but also recombination events that allow robust inference of the past coalescent rate (inverse of effective population size N*e*) by the reconstruction of an ancestral recombination graph (Schiffels & Durbin 2014; Terhorst et al. 2017; Speidel et al. 2019). Furthermore, we found that between 10–1 Mya, the ancestral population of *M. parvulus* and *M. trifasciatus* showed the same coalescent rates, thus forming a single panmictic population (Figure 2A) followed by a drop in N*e* (halving in size) between 2–1 Mya, after which each population started to follow different trajectories, corroborating the timing of the single colonisation event reported by Arbogast et al. (2006).

Our genome-wide phylogenetic reconstruction recovered two monophyletic clades for *M. parvulus* and *M. trifasciatus* (Figure 1), supporting previous multilocus-based topologies recovered by Arbogast et al. (2006) and Nietlisbach et al. 2013 over that of Štefka et al. (2011), with *M. trifasciatus* nested within *M. parvulus.* Although a phylogenetic tree based on microsatellite distance data placed *M. trifasciatus* sister to all Galápagos mockingbirds (Hoeck et al. 2010) we adopted the assumption that *M. trifasciatus* and *M. parvulus* are sister species (Arbogast et al. 2006, Nietlisbach et al. 2013, but also Štefka et al. 2011 by mtDNA) in order to further interpret their demographic past in comparative manner. The N*e* of the ancestral population of these two species roughly follows the same trajectory between 5–1 Mya based on Relate. The other two methods, MSMC and SMC++, either do not show the coalescent rate that distant in the past, or show a suspicious growth signal in the Champion population (Figure S5, S6), probably due to the fact that the recent extreme bottleneck distorts the signal of past bottlenecks, which is a known problem of these methods (Terhorst & Song 2015). We therefore preferably interpret the results from Relate,

which is more robust to this issue (Speidel et al. 2019). Moreover, it allows not only a reliable inference of a within population coalescent rate, but also inference of the coalescent rate between populations (cross-coalescent rate), which is able to indicate past connectivity. The equity of these two coalescent rates between all the population pairs from 5–1 Mya indicate a panmictic population (Figure 2B). After 1 Mya the cross-coalescent rate between *M. trifasciatus* and *M. parvulus* (Figure 2B) drops rapidly. The relative coalescent rate (cross/intra) decreases from 1 to 0.25 in between 1–0.5 Mya and reaches almost zero (full divergence) at 100 Kya. GPhoCS inference indicated a split time between the species of 428 Kya, which is congruent with the cross-coalescent rate, that captures the graduality of the divergence process. N*e* of both populations of *M. trifasciatus* grew for a shorter period of time after the bottleneck between 1–0.5 Mya and then decreased rapidly. The split between *M. trifasciatus* and *M. parvulus* can be linked to the paleogeography of the islands that was reconstructed based on lava ageing and interpolation from tectonic movement. Geist et al. (2014) suggesting a continuous westernmost landmass in the Galápagos which subsequently separated into several islands. We suggest that the ancestral population of *M. parvulus* and *M. trifasciatus* occupied this proto-island, resulting in population growth in both species. Later at 500 Kya, speciation of *M. trifasciatus* was mediated by vicariance when Floreana island became isolated and smaller resulting in the decrease of N*e* in *M. trifasciatus*.

Population growth after an initial bottleneck was observed in each of the *M. parvulus* populations (Figure 2A, Figure S5, S6). Time span of the growth was specific to each population with an exception of northern populations (W, D, M). The trajectory of their N*e* was noticeably similar (Figure 2A) with an N*e* peak of 120,000 at 100 Kya followed by a rapid decline. Isabela and Santa Cruz, the two currently largest populations, showed double peaks of their N*e*. The first peak was similar in the two populations. The second growth started 100 Kya and N*e* in Santa Cruz reached almost 150,000, which is similar to the period before the

colonisation of the Galápagos. Timing of the growth in Isabela is the same, but the peak reached only 70,000. Both populations entered their final bottleneck at 20 Kya. This final constriction, which occurred only 10 Kya, was when the populations of *M. parvulus* reached their current N*e* likely constrained by island size .

The most striking pattern reported here is that populations were not separated on small islands since the initial colonisation of the archipelago, but instead, they formed larger populations with N*e* reaching above 100,000 in the period between 1 Mya to at least 40 Kya. The young divergence of each isolated population also explains the high amount of incomplete lineage sorting seen in the TWISST analysis (Figure S2). This observation becomes less striking when the demography is linked with the paleography of the islands. Geist et al. (2014) reported that ~1 Mya the core Galápagos Islands (around Santa Cruz) formed one larger landmass, with repeated inter-island connections due to eutasy (sea level fluctuations) between 1 Mya–20 Kya (Ali & Aitchison 2014). This large landmass probably hosted a larger ancestral population (what is currently *M. parvulus*) that colonised the northern islands of Pinta, Marchena, Darwin and Wolf, which were never connected with the core island (Geist et al. 2014). Interestingly, these populations started to diverge between 700–500 Kya (cross-coalescence) except for Darwin and Wolf, which started diverging 100 Kya (Figure S8). This initial phase of divergence of the northern populations is reflected in their population growth estimates, which indicates that rather than a colonisation of the northern islands, which were already present at that time (Geist et al. 2014), these populations were already diverging on this core landmass, which was intermittently fragmented. Northern islands were only colonised later as suggested by the timing of the split obtained between PMS and DWI clades, which happened only ~64 Kya (GPhoCS analysis). An alternative explanation would posit that the northern islands were colonised early and constant gene flow from the core inflated their past population size substantially. Given that the northern islands, especially Darwin and

Wolf, are quite far (~150 km from current northern edge of Isabela) from the core to sustain considerable gene flow, we support the former scenario that for most of the Pleistocene *M. parvulus* populations were large and to some degree interconnected. Our proposed mode and tempo of divergence for mockingbirds also matches Racer snakes (Ali & Aitchison 2014) highlighting the roles of allopatric divergence and hybridization events in shaping their evolutionary history. In mockingbirds, the connectivity could have possibly been augmented by their dispersal ability, although it seems to be limited (Hoeck et al. 2010b, Curry & Grant 1990, Grant et al. 2000, but see Ortiz-Catedral et al. 2021). In sum, the general demographic scenario supported by our data is in contrast with our hypothesis that mockingbird populations on each island have been isolated for a long time characterised by low population sizes since the initial founder colonisation, which is partially the case only for *M. trifasciatus*.

Noticeable effect of the last glacial maximum (LGM) is another interesting pattern derived from our demographic inference. It is best seen in the populations currently occupying the largest islands Isabela and Santa Cruz. For a substantial amount of time (~600 Ky) their N*e* was smaller than that of the populations currently present in the northern islands (D, W, M). N*e* of Isabela and Santa Cruz peaks above 75,000 only between 30–10 Kya matching the LGM and subsequent deglaciation (Clark et al. 2009). This results can be attributed to the suggested increase of land surface of the islands and interconnections of central islands with Isabela due to a substantial recess of the water level (Ali & Aitchison 2014). There is also a noticeable peak in the Northern populations (D, W, M, P), revealed by SMC++, which might also indicate that LGM influenced the carrying capacity of the islands. These interpretations are only valid if we assume mockingbirds to be widely distributed throughout the majority of the land they occupy on each island. Given that these species are recognized as generalist species with enough adaptive plasticity to exploit a large number of niches present throughout the

island (Curry & Grant 1990, Grant et al. 2000, Jiménez-Uzcátegui 2011), we suggest a plausible link between island size and N*e* of each population.

Genomic diversity, inbreeding and genetic load

Despite the turbulent changes of N*e* in the past, current populations of Galápagos mockingbirds showed a clear link between genomic diversity, expressed as heterozygosity, and island size. The population of *M. trifasciatus* on Champion islet has an average heterozygosity of 0.000043, which is two orders of magnitude lower than the population on Santa Cruz. Genome-wide heterozygosity in *M. trifasciatus* is to our knowledge the lowest value observed among all vertebrates with sequenced genomes (e.g. Robinson et al. 2022, Brüniche-Olsen et al. 2022, Leroy et al. 2021, Bruniche-Olsen et al. 2019, Abascal et. al 2016, Zhou et al. 2013). The diversity on Champion is followed by Gardner by Floreana, Darwin, Wolf, Pinta, Marchena, Isabela and Santa Cruz (Figure 5A), which is the same order if ordered by island size, with Isabela as a single exception. The relationship between heterozygosity and island size on a logarithmic scale showed a statistically significant linear trend and it was also mirrored by nucleotide diversity (Figure S9) and current N*e* (Figure 3B). This result is expected only if the populations on each island were panmictic and the size of an island determined the carrying capacity (suitable habitat), while gene flow between populations contributed negligibly and past demography was stable (for at least 4N*e* generations). The first assumption of no structure and habitat preferences would be met in generalist mockingbirds that occupy the whole area of islands (Curry & Grant, 1998), as Curry (1989) eloquently described: “... the only unoccupied habitat available to dispersing mockingbirds on Genovesa, Champion, and Espaniola is the Pacific Ocean.” On larger islands with more vegetation zones, like Santa Cruz or Isabela, mockingbirds may prefer the arid and transition zones, but it seems that the effect of habitat preference is only minor as it does not break the overall island-diversity

correlation. Gene flow also contributed negligibly since the divergence of current populations based on our demographic analysis (Table S7, S8). However, the divergence times between populations are younger than we expected, ranging from 64–4 Kya. Also, the N*e* of each population gained its respective value ranked by island size only in the last 50–60 Kya. Our result raises the question of whether island size could modulate genetic diversity in such a short time, whereas before that time period N*e* fluctuated rather independently of the current island sizes (Figure 2A), with their demography modulated by eustasy (Ali & Aitchison 2014) and under different island geography settings (Geist et al. 2014).

The diversity observed now should reflect processes in the population that trace back up to the most recent common ancestor, approximately 4N*e* generations ago (Wakeley 2009). In vertebrates, the time to the most recent common ancestor (MRCA) was estimated to span thousands to millions of years (Vijay *et al.* 2017). In our case, MRCA spans from 306 years ago in Champion to 902,052 years ago in Santa Cruz, based on the N*e* estimate from GPhoCS using generation time of 4.6 (Grant *et al.* 2000). MRCA is no older than 51,516 in the small and intermediate populations (F, D, W, P), which means their nucleotide diversity captures only the post divergence demography on individual islands. On the contrary, genetic diversity in Santa Cruz and Isabela reflects demographic events that happened already 902 and 514 Kya, respectively. Whilst higher than expected diversity in Santa Cruz (given the island size trend) can be explained by the large ancestral population, the lower than expected heterozygosity in Isabela can be related to a weak but non-negligible population structure detected there by Hoeck *et al.* (2010b).

Our results support the view brought in by Frankham et al. (1997) that area is a crucial determinant of genetic diversity and genetic drift. This view is an unsettled issue with several empirical studies supporting its validity (Exposito-Alonso et al. 2022, Brüniche-Olsen et al. 2021,

Bruniche-Olsen et al. 2019, Lino et al. 2019, Hunt et al. 2022) while others refuting it (Mackintosh et al. 2019, James et al. 2016, Romiguier et al. 2014). The evidence for the area effect was also found in Darwin's Finches and allies based on whole genome sequencing (Bruniche-Olsen et al. 2019) and a previous microsatellite study (Petren et al. 2005). Our result also resonates with the more general observation that genomic diversity follows a power law with geographic area (Exposito-Alonso et al. 2022). The studies that did not find a relationship between range area and genetic diversity rather focused on a between-species comparison, where other factors like propagule size, or other life history traits are more important determinants of genetic diversity than their range (Mackintosh et al. 2019, Chen et al. 2017, Romiguier et al. 2014). Hence, our study contributes to the mounting evidence that geographic area determines the amount of genetic diversity and genetic drift in populations within one species (or closely related species), where life history traits differences are negligible.

In this sense we corroborate the results of Hoeck et al. (2010) that showed island size as the determinant of genetic diversity in microsatellites. However, using whole genome data allowed us to relate the genetic load and island size. Our hypothesis based on theory (Hedrick 1994, Crow 1970) and former empirical studies (Kyriazis et al. 2019, van der Valk et al. 2019) was to observe a positive correlation between genetic load and island size if purging was more efficient in the smaller populations. We used the number of derived, non-synonymous homozygous genotypes (recessive load) and the total number of derived non-synonymous alleles per population (additive load) as indices of genetic load, corresponding to the methodology of Simons & Sella (2016). These indices reflect two different modes of dominance which is an important, but rather elusive factor modulating estimates of genetic load (Bertorelle *et al.* 2022, Kyrizais *et al.* 2022). We found a negative correlation between the recessive load and island size, indicating higher recessive load or realised load (in the terms of Bertorelle *et al.* 2022) in smaller populations (Figure

5C). This is driven by the fact that homozygosity is determined by island size and therefore more non-synonymous derived alleles appear in homozygous state in smaller islands. Simons & Sella (2016) already showed that recessive load is affected by recent bottlenecks, which would also be the case in our species. They advocated the use of additive load as a more robust index for long-term selection efficiency (purging). In mockingbirds, additive load actually shows a positive correlation with island size, but only for *M. parvulus* populations where the smallest populations have the lowest additive load, which suggests an effect of purging in this species.

By contrast, the two populations of *M. trifasciatus* on Gardner by Floreana and Champion do not fit the relationship as they show the highest additive load from all the populations (Figure 5D). The result can be explained by considering the past demography of mockingbirds in the light of recent findings in the population genomics field. Additive load indicates masked load in the terms of Bertorelle et al. (2022) if we assume that deleterious alleles are recessive (Huber et al. 2018). In our case it suggests that purging is more efficient in the smaller populations of *M. parvulus* compared to larger populations of the same species, but also compared to the smaller populations of *M. trifasciatus*. Although the size of *M. trifasciatus* was small for a longer period (Figure 2A) than any *M. parvulus* population, it was heavily affected by the extinction of *M. trifasciatus* on Floreana island approximately 140 years ago (Curry 1986). Both current *M. trifasciatus* populations were to some degree interconnected with the larger Floreana population, especially Champion, and the extinction resulted in a strong bottleneck (Hoeck et al. 2010a). Higher masked load in *M. trifasciatus* can be therefore attributed to the fact that its isolation and extremely small population size is only recent (~140 years) and there was not enough time for purging to remove the genetic load shared with the extinct parental population (García-Dorado 2012). We are aware that our indices of genetic load are relatively crude, depending on several assumptions about dominance and lack of positive

selection and we call for further examination of genetic load using simulation as suggested by Kyrizais et al. (2022).

Analysis of runs of homozygosity (ROH), long stretches of DNA identical by descent, showed that the levels of inbreeding are high only in the smallest populations of both species (D, W, C, F). It has been shown that the proportion of long ROHs in the genome gives very precise estimates of inbreeding coefficients (Ceballos et al. 2018, McQuillan et al. 2008). The total length of ROH in Champion was unprecedented as the ROHs over 1Mbp spanned around 62 % of the genome (F_{ROH} = 0.62). It is even higher than what has been observed in the Isle Royale wolf population during its final decline before genetic rescue (Hedrick et al. 2019), or in the heavily bottlenecked kakapo (Foster et al. 2021). Such a level of inbreeding is typical for partially selfing plants (Dole & Ritland 1993) rather than for outcrossing animals. The distribution of lengths of ROHs in Champion indicate a long and ongoing history of consanguineous mating, as long ROHs over 2Mbp are common (Figure 4) and even dozens of ROHs over 6 Mbp were found. This pattern is related to the extremely small population size, possibly combined with the recent disconnection of the Champion population from the extinct population on Floreana island, which induced a strong bottleneck. Gardner by Floreana shows a much lower inbreeding coefficient on average (F_{ROH} = 0.22), but there is also higher variation, with two individuals showing F_{ROH} 0.16 and one individual showing F_{ROH} 0.34. The inbreeding coefficient on Darwin is similar to Gardner by Floreana (F_{ROH} = 0.23), but the number of long ROHs over 2 Mbp is substantially smaller (Figure 4), indicating that inbreeding is less common, or older than in *M. trifasciatus*. This result can indicate some past period of intensive consanguineous mating as we see a high frequency of ROHs in the 0.1–0.5 Mbp category. The Inbreeding coefficient in Wolf is smaller than in Darwin, but the distribution in ROH is similar, indicating a common recent demography. The other intermediate or large populations have inbreeding coefficients close to 0. Therefore, the relationship between

inbreeding coefficient and island size is rather exponential (Figure 5B) and the inbreeding is considerable only if population size, reflected by island size is really small now (Woodworth et al. 2002) as in *M. trifasciatus* (Figure 3), or if it shrank in the recent past (Ceballos et al. 2018).

Although both the masked and realised load is high in *M. trifasciatus* and the levels of inbreeding are unprecedented, especially in Champion, we do not see any acute inbreeding depression in the populations of this critically endangered species (Grant *et al.* 2000). Immune defence indices are not lower than in other populations of Galápagos mockingbirds (Hoeck & Keller 2012), although Deem et al. (2011) showed that the Champion population is less clinically healthy compared to Gardner-by-Floreana. Importantly, the census population size of both populations did not show any decreasing trend (Jiménez-Uzcátegui et al. 2011). Apart from our finding of the highest masked and realised load across the genome, a previous study of immune genes showed the *M. trifasciatus* populations to have the lowest diversity in the major histocompatibility complex II (Vlček et al. 2016), and the Champion population to carry a potentially deleterious allele of the TLR 4 receptor (Vlček et al. 2022). Although these genetic factors are unfavourable in the tiny populations, the populations seem not to be affected substantially by inbreeding depression. Nonetheless, the species could benefit from the planned reintroduction project that aims to repopulate Floreana island from Champion and Gardner by Floreana. Our genomic investigation revealed only shallow divergence between the populations, similar to Hoeck et al. (2010a). Moreover, similar levels of genetic load in the populations imply that crossing between them should not cause a problem from the conservation genetic point of view.

Acknowledgement

The authors are grateful to several colleagues for their help in fieldwork and for sharing their samples, namely Lukas Keller, Paquita Hoeck, Kenneth Petren, Glauco Camenish, Francisco Rubianes, Enzo Rodriguez-Reyes, Sabina Ascencio Ramírez, Danny Rueda, and Christian Sevilla. We are grateful to Annabel Beichmann, Kirk Lohmueller, Robert Wayne and Daniel Chavez for help with data analysis. Further we thank Filip Kolář for support and comments on earlier versions of the manuscript. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic. Additional as well as the Swiss friends of Galápagos for financial support, and Charles Darwin Foundation for assistance during fieldwork. We also thank the Ecuadorian Ministry of the Environment and Directorate of the Galápagos National Park for assistance with logistics, and permits to collect the relevant samples. All samples were collected following approved protocols by the Directorate of the Galápagos National Park (permits PC 74-17 and PC-27-19). All material for the reference genome was accessed under MAE-DNB-CM-2016-0041 to JAC .

Data accessibility and supplementary material

Document with supplementary notes, tables and figures is accessible online upon request to the author. https://docs.google.com/document/d/1V8BRRt1BVjfg3WQ_5pidMvtOys4LYXnQWG9xyLrt4OE/edit?usp=sharing

Genome supplementary notes document is accessible online upon request to the author. https://docs.google.com/document/d/1FpfwkjAzqqmFDnJQiUlFEz95Tkw9bV0hzxjdrCd7JPw/edit?usp=sharing

Scripts related to the work will be published in Github.

Resequencing data are available on Genbank under accession numbers: SRPXXXX: SRRXXXX-SRRXXXXX.

References

Abascal F, Corvelo A, Cruz F et al. (2016) Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biology, 17, 251.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19, 1655–1664.

Ali JR, Aitchison JC (2014) Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos (JM Fernández-Palacios, Ed,). Journal of Biogeography, 41, 1227–1241.

Allendorf FW, Luikart G, Aitken SN (2013) Conservation and the genetics of populations, 2nd edition. Wiley-Blackwell.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arbogast BS, Drovetski S V, Curry RL (2006) The origin and diversification of Galápagos mockingbirds. Evolution, 60, 370–382.

Ballou JD (1997) Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. Journal of Heredity, 88, 169–178.

Bertorelle G, Raffini F, Bosse M et al. (2022) Genetic load: genomic estimates and applications in non-model animals. Nature Reviews Genetics, 23, 492–503.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170

Brüniche-Olsen A, Kellner KF, Belant JL, Dewoody JA (2021) Life-history traits and habitat availability shape genomic diversity in

birds: Implications for conservation. Proceedings of the Royal Society B: Biological Sciences, 288.

Brüniche-Olsen A, Kellner KF, DeWoody JA (2019) Island area, body size and demographic history shape genomic diversity in Darwin's finches and related tanagers. Molecular Ecology, 28, 4914–4925.

Cantarel BL, Korf I, Robb SMC et al. (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research, 18, 188–196.

Ceballos G, Ehrlich PR, Dirzo R (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proceedings of the National Academy of Sciences, 201704949.

Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF (2018) Runs of homozygosity: windows into population history and trait architecture. Nature Reviews Genetics, 19, 220–234.

Charlesworth B (2018) Mutational load, inbreeding depression and heterosis in subdivided populations. Molecular Ecology, 27, 4991–5003.

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nature Reviews Genetics, 10, 783–796.

Chen J, Glémin S, Lascoux M (2017) Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. Molecular Biology and Evolution, 34, 1417–1428.

Cingolani P, Platts A, Wang LL et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly, 6, 80–92.

Clark PU, Dyke AS, Shakun JD et al. (2009) The Last Glacial Maximum. Science, 325, 710–714.

Craig RJ, Suh A, Wang M, Ellegren H (2018) Natural selection beyond genes: Identification and analyses of evolutionarily conserved

elements in the genome of the collared flycatcher (*Ficedula albicollis*). Molecular Ecology, 27, 476–492.

Crow JF (1970) Genetic loads and the cost of natural selection. In: Mathematical topics in population genetics , pp. 128–177. Springer.

Crow JF (1993) Mutation, mean fitness, and genetic load. Oxford Surv. Evol. Biol, 9, 3–42.

Curry R (1986) Whatever happened to the Floreana mockingbird. Noticias de Galápagos, 43, 13–15.

Curry RL (1989) Geographic variation in social organization of Galápagos mockingbirds : ecological correlates of group territoriality and cooperative breeding. Behaviroal Ecology and Sciobiology, 24, 147–160.

Curry R, Grant P (1989) Demography of the cooperatively breeding Galápagos mockingbird, Nesomimus parvulus, in a climatically variable environment. The Journal of Animal Ecology, 58, 441–463.

Danecek P, Auton A, Abecasis G et al. (2011) The variant call format and VCFtools. Bioinformatics, 27, 2156–2158.

Darwin C (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray.

Deem SL, Parker PG, Cruz MB, Merkel J, Hoeck PEA (2011) Comparison of blood values and health status of Floreana Mockingbirds (*Mimus trifasciatus*) on the islands of Champion and Gardner-by-Floreana, Galápagos Islands. Journal of wildlife diseases, 47, 94–106.

Depristo MA, Banks E, Poplin R et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics, 43, 491–501.

DeWoody JA, Harder AM, Mathur S, Willoughby JR (2021) The long-standing significance of genetic diversity in conservation. Molecular Ecology, 30, 4147–4154.

Dole J, Ritland K (1993) Inbreeding depression in two *Mimulus* taxa measured by multigenerational changes in the inbreeding coefficient. Evolution, 47, 361–373.

Exposito-Alonso M, Booker TR, Czech L et al. (2022) Genetic diversity loss in the Anthropocene. bioRxiv, 464000.

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. Nature reviews. Genetics, 8, 610–8.

Foster Y, Dutoit L, Grosser S et al. (2021) Genomic signatures of inbreeding in a critically endangered parrot, the kākāpō (P Morrell, Ed,). G3 Genes|Genomes|Genetics, 11.

Frankham R (1997) Do island populations have less genetic variation than mainland populations? Heredity, 78, 311–327.

Freedman AH, Gronau I, Schweizer RM et al. (2014) Genome Sequencing Highlights the Dynamic Early History of Dogs (L Andersson, Ed,). PLoS Genetics, 10, e1004016.

García-Dorado A (2012) Understanding and predicting the fitness decline of shrunk populations: Inbreeding, purging, mutation, and standard selection. Genetics, 190, 1461–1476.

García-Dorado A, Caballero A (2021) Neutral genetic diversity as a useful tool for conservation biology. Conservation Genetics, 22, 541–545.

Geist DJ, Snell H, Snell H, Goddard C, Kurz MD (2014) A Paleogeographic Model of the Galápagos Islands and Biogeographical and Evolutionary Implications. In: The Galápagos: A Natural Laboratory for the Earth Sciences (eds Harpp K, Mittelstaedt E, D'Ozouville N, Graham DW), pp. 145–166. John Wiley & Sons.

Grant P, Curry R, Grant B (2000) A remnant population of the Floreana mockingbird on Champion island, Galápagos. Biological Conservation, 92, 285–290.

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. Nature Genetics, 43, 1031–1035.

Hedrick PW (1994) Purging inbreeding depression and the probability of extinction: Full-sib mating. Heredity, 73, 363–372.

Hedrick PW, Garcia-Dorado A (2016) Understanding Inbreeding Depression, Purging, and Genetic Rescue. Trends in Ecology and Evolution, 31, 940–952.

Hedrick PW, Robinson JA, Peterson RO, Vucetich JA (2019) Genetics and extinction and the example of Isle Royale wolves. Animal Conservation, 22, 302–309.

Hoeck PEA, Beaumont MA, James KE et al. (2010a) Saving Darwin's muse: evolutionary genetics for the recovery of the Floreana mockingbird. Biology letters, 6, 212–215.

Hoeck PEA, Bollmer JL, Parker PG, Keller LF (2010b) Differentiation with drift: a spatio-temporal genetic analysis of Galápagos mockingbird populations (*Mimus* spp.). Philosophical Transactions of the Royal Society B: Biological Sciences, 365, 1127–1138.

Hoeck PEA, Keller LF (2012) Inbreeding, immune defence and ectoparasite load in different mockingbird populations and species in the Galápagos Islands. Journal of Avian Biology, 43, 423–434.

Huber CD, Durvasula A, Hancock AM, Lohmueller KE (2018) Gene expression drives the evolution of dominance. Nature Communications, 9, 2750.

Hunt DAGA, DiBattista JD, Hendry AP (2022) Effects of insularity on genetic diversity within and among natural populations. Ecology and Evolution, 12, 1–12.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution, 23, 254–67.

James JE, Lanfear R, Eyre-Walker A (2016) Molecular Evolutionary Consequences of Island Colonization. Genome Biology and Evolution, 8, 1876–1888.

Jiménez-Uzcátegui G, Llerena W, Bryan Milstead W, Lomas EE, Wiedenfield DA (2011) Is the population of Floreana Mockingbird Mimus trifasciatus declining? Cotinga, 33, 34–40.

Johnson TH, Stattersfield AJ (1990) A global review of island endemic birds. Ibis, 132, 167–180.

Kardos M, Armstrong EE, Fitzpatrick SW et al. (2021) The crucial role of genome-wide genetic variation in conservation. Proceedings of the National Academy of Sciences, 118.

Keightley PD, Jackson BC (2018) Inferring the Probability of the Derived vs. the Ancestral Allelic State at a Polymorphic Site. Genetics, 209, 897–906.

Kennedy ES, Grueber CE, Duncan RP, Jamieson IG (2014) Severe inbreeding depression and no evidence of purging in an extremely inbred wild species-the chatham island black robin. Evolution, 68, 987–995.

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic acids research, 47(D1), D807-D811.

Kutschera VE, Poelstra JW, Botero-Castro F et al. (2020) Purifying Selection in Corvids Is Less Efficient on Islands. Molecular Biology and Evolution, 37, 469–474.

Kyriazis CC, Robinson JA, Lohmueller KE (2022) Using computational simulations to quantify genetic load and predict extinction risk. bioRxiv, 2022.08.12.503792.

Kyriazis CC, Wayne RK, Lohmueller KE (2019) High genetic diversity can contribute to extinction in small populations. bioRxiv, 678524.

Lande R (1993) Risks of Population Extinction from Demographic and Environmental Stochasticity and Random Catastrophes. The American Naturalist, 142, 911–927.

Lanfear R, Kokko H, Eyre-Walker A (2014) Population size and the rate of evolution. Trends in Ecology and Evolution, 29, 33–41.

Lee TH, Guo H, Wang X, Kim C, Paterson AH (2014) SNPhylo: A pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics, 15, 1–6.

Leroy T, Rousselle M, Tilak M et al. (2020) Endemic island songbirds as windows into evolution in small effective population sizes. bioRxiv, 2020.04.07.030155.

Leroy T, Rousselle M, Tilak M-K et al. (2021) Island songbirds as windows into evolution in small populations. Current Biology, 31, 1303-1310.e4.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754–1760.

Lino A, Fonseca C, Rojas D, Fischer E, Ramos Pereira MJ (2019) A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. Mammalian Biology, 94, 69–76.

Lynch M, Conery J, Burger R (1995) Mutational meltdowns in sexual populations. Evolution, 49, 1067–1080.

Mackintosh A, Laetsch DR, Hayward A et al. (2019) The determinants of genetic diversity in butterflies. Nature Communications, 10, 3466.

Martin SH, Van Belleghem SM (2017) Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. Genetics, 206, 429–438.

McKenna A, Hanna M, Banks E et al. (2010) The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20, 1297–1303.

McQuillan R, Leutenegger AL, Abdel-Rahman R et al. (2008) Runs of Homozygosity in European Populations. American Journal of Human Genetics, 83, 359–372.

Narasimhan V, Danecek P, Scally A et al. (2016) BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics, 32, 1749–1751.

Nietlisbach P, Wandeler P, Parker PG et al. (2013) Hybrid ancestry of an island subspecies of Galápagos mockingbird explains discordant gene trees. Molecular Phylogenetics and Evolution, 69, 581–592.

Ohta T (1992) The Nearly Neutral Theory Of Molecular Evolution. Annual Review of Ecology and Systematics, 23, 263–286.

Van Oosterhout C (2020) Mutation load is the spectre of species conservation. Nature Ecology and Evolution, 16–18.

Ortiz-Catedral L, Lichtblau A, Anderson MG, Sevilla C, Rueda D (2021) First report of co-occurrence of two species of mockingbird in the Galápagos Islands: a San Cristóbal Mockingbird Mimus melanotis in a population of Floreana Mockingbird M. trifasciatus. Galápagos Research, 70, 41–44.

Pembleton LW, Cogan NOI, Forster JW (2013) StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources, 13, 946–952.

Petren K, Grant PR, Grant BR, Keller LF (2005) Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. Molecular ecology, 14, 2943–57.

R Core Team (2018) R: A Language and Environment for Statistical Computing. Vienna, Austria.

Ricketts TH, Dinerstein E, Boucher T et al. (2005) Pinpointing and preventing imminent extinctions. Proceedings of the National Academy of Sciences of the United States of America, 102, 18497–18501.

Robinson JA, Kyriazis CC, Nigenda-Morales SF et al. (2022) The critically endangered vaquita is not doomed to extinction by inbreeding depression. Science, 376, 635–639.

Robinson JA, Ortega-Del Vecchyo D, Fan Z et al. (2016) Genomic Flatlining in the Endangered Island Fox. Current Biology, 26, 1183–1189.

Romiguier J, Gayral P, Ballenghien M et al. (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature, 515, 261–263.

Schiffels S, Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. Nature Genetics, 46, 919–925.

Shaw KL, Gillespie RG (2016) Comparative phylogeography of oceanic archipelagos: Hotspots for inferences of evolutionary process. Proceedings of the National Academy of Sciences of the United States of America, 113, 7986–7993.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31(19), 3210-3212.

Simons YB, Sella G (2016) The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. Current Opinion in Genetics & Development, 41, 150–158.

Smeds L, Qvarnström A, Ellegren H (2016) Direct estimate of the rate of germline mutation in a bird. Genome Research, 26, 1211–1218.

Speidel L, Forest M, Shi S, Myers SR (2019) A method for genome-wide genealogy estimation for thousands of samples. Nature Genetics, 51, 1321–1329.

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. Proceedings of the National Academy of Sciences of the United States of America, 101, 15261–15264.

Štefka J, Hoeck PE, Keller LF, Smith VS (2011) A hitchhikers guide to the Galápagos: Co-phylogeography of Galápagos mockingbirds and their parasites. BMC Evolutionary Biology, 11, 284.

Teixeira JC, Huber CD (2021) The inflated significance of neutral genetic diversity in conservation genetics. Proceedings of the National Academy of Sciences of the United States of America, 118, 1–10.

Terhorst J, Kamm JA, Song YS (2017) Robust and scalable inference of population history from hundreds of unphased whole genomes. Nature Genetics, 49, 303–309.

Terhorst J, Song YS (2015) Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. Proceedings of the National Academy of Sciences of the United States of America, 112, 7677–7682.

van der Valk T, de Manuel M, Marquez-bonet T, Guschanski K (2019) Estimates of genetic load suggest extensive genetic purging in mammalian populations. bioRxiv, 696831.

Vijay N, Weissensteiner M, Burri R et al. (2017) Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. Molecular Ecology, 26, 4284–4295.

Vlček J, Hoeck PEA, Keller LF et al. (2016) Balancing selection and genetic drift create unusual patterns of MHCIIβ variation in Galápagos mockingbirds. Molecular Ecology, 25, 4757–4772.

Vlček J, Miláček M, Vinkler M, Štefka J (2022) Effect of population size and selection on Toll-like receptor diversity in populations of Galápagos mockingbirds. Journal of evolutionary biology, In print.

Wakeley J (2009) Coalescent Theory: An Introduction. Macmillan Learning.

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB (2017) Direct determination of diploid genome sequences. Genome Research, 27, 757–767.

Welles SR, Dlugosch KM (2019) Population Genomics of Colonization and Invasion. In: Population Genomics: Concepts, Approaches and Applications (ed Rajora OP), pp. 655–683. Springer International Publishing, Cham.

Westbury M V., Hartmann S, Barlow A et al. (2018) Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the brown Hyena. Molecular Biology and Evolution, 35, 1225–1237.

Willi Y, Van Buskirk J, Hoffmann AA (2006) Limits to the Adaptive Potential of Small Populations. Annual Review of Ecology, Evolution, and Systematics, 37, 433–458.

Willi Y, Kristensen TN, Sgro CM et al. (2022) Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. Proceedings of the National Academy of Sciences of the United States of America, 119, 1–10.

Woodworth LM, Montgomery ME, Briscoe DA, Frankham R (2002) Rapid genetic deterioration in captive populations: Causes and conservation implications. Conservation Genetics, 3, 277–288.

Woolfit M, Bromham L (2005) Population size and molecular evolution on islands. Proceedings of the Royal Society B: Biological Sciences, 272, 2277–2282.

Xue Y, Prado-Martinez J, Sudmant PH et al. (2015) Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science, 348, 242–245.

Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L (2019) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files (R Schwartz, Ed,). Bioinformatics, 35, 1786–1788.

Zheng X, Levine D, Shen J et al. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28, 3326–3328.

Zhou X, Sun F, Xu S et al. (2013) Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. Nature Communications, 4, 1–6.

Summary

Summary of results

Population declines and fragmentations as a result of human activities are leading to a precarious future for many taxa (Li et al. 2016, Schlaepfer et al. 2018, Leigh et al. 2019). Dwindling populations isolated from each other by agricultural landscapes face a variety of threats ranging from air pollution to overexploitation in the first place (Maxwell et al. 2016) and from demographic stochasticity to ecological and genetic Allee effects in the second place (Willi et al. 2005, Allendorf et al. 2013). The challenges of maintaining such populations in the long term are reminiscent of those faced by small populations isolated on oceanic islands. In this respect, my work provides a valuable case study of the genetic challenges faced by naturally fragmented and size-limited populations of mockingbirds that have survived on the Galápagos Islands for thousands of years. It offers a rather optimistic perspective, as it seems that natural processes offer some kind of relief from the detrimental genetic challenges.

Below I show how MHCIIβ, TLR:1B,4,15, and genome-wide patterns of genetic and phenotypic polymorphism scale with population size, which is determined by the size of the islands on which my focal populations of mockingbirds live.

In **Chapter I**, my colleagues and I showed that while the total number of MHCIIβ alleles per population is affected by island size, the number of functionally distinct peptide supertypes and the number of alleles per individual are maintained. Moreover, the number of supertypes per population and the number of alleles per individual in Galápagos mockingbirds are not lower than in the larger mainland mockingbird population. Trans-species polymorphism, a high rate of non-synonymous to synonymous polymorphisms at binding sites and a statistically significant, albeit weak, association of the MHCIIβ supertypes with ectoparasite load (**chapter II**) suggest an active role of

pathogen-mediated balancing selection in MHCIIβ. Thus, it appears that although individual alleles are lost, functional diversity is maintained by balancing selection despite the differential extent of N*e*, which is an advocated solution to the ongoing dilemma of whether drift can outweigh balancing selection (Lighten et al. 2017). I realise that the absence of a correlation between island area and the number of alleles or supertypes does not prove the absence of causality. I also recognise that both selection and genetic drift affect allele frequencies and diversity, and I do not argue to the contrary. I merely conclude that the theoretical range of pathogens detected by MHCIIβ in mockingbird populations is not seriously restricted by variation in N*e*, which is contrary to the general view in the field that the health of small populations degenerates due to lower immunogenetic diversity (Hedrick 2001, Höglund 2009).

In **chapter III** on the diversity of three TLR genes, my colleagues and I showed that neutral synonymous diversity clearly correlates with island area, but again, functional non-synonymous diversity, at least in the two genes TLR1B and TLR4, shows no correlation. Moreover, there is only one major protein variant in these genes, which is shared by almost all populations. Considering that the sampling is almost identical to the previous chapter, the difference between the MHCIIβ with 149 peptide variants and TLR1B and TLR4 with only one or two major protein variants is striking, even despite the fact that at least six different MHCIIβ loci were co-amplified. This contrast suggests that the TLR loci are subjected to a strong stabilising selection, in contrast to MHCIIβ, which is subjected to balancing selection. Despite the strong stabilising selection on TLR4, we found a protein variant P4 in the smallest population of Champion that exhibits remarkable physicochemical differences compared to the predominant widespread variant P10. It is therefore questionable whether this is a fixed deleterious variant or an adaptation. I tend to interpret it as a deviation from the optimal protein conformation, which is probably a case of negative genetic effect related to small population size, yet experimental validation would be needed to clarify

this. TLR15 appears to be under less pressure from stabilising selection, as both neutral and adaptive diversity correlate with island size and there were four major protein variants. This is actually consistent with a recent finding that TLR15 has been cryptically pseudogenised and therefore lacks functional immunogenetic polymorphism in birds (Fiddaman et al. 2022). Previous studies (Grueber et al. 2013, Hartmann et al. 2014) claimed that TLRs are predominantly shaped by drift in small populations, but my results show a different perspective. Although the neutral diversity of TLRs is shaped by genetic drift, the functional diversity of TLR4 and TLR1B is strongly constrained by purifying selection. This finding also suggests some resilience to deleterious genetic factors in mockingbird populations in Galápagos, while Champion remains a possible exception to this general finding.

In **chapter IV** my colleagues and I present new details on the population demography and fluctuation of N*e* in a subset of eight populations of Galápagos and Floreana mockingbirds. My colleagues and I show that the bottleneck in Galápagos colonisation occurred between 1-2 Mya, constraining previous estimates based on mitochondrial DNA (Arbogast et al. 2006). In contrast to our earlier view of populations being small and separated over long periods of time, the subdivision of Galápagos mockingbird populations was relatively recent and the decline of their N*e* occurred within the last 100,000 years. The connection and separation of islands due to eustasis and erosion (Ali & Aitchison 2014, Geist et al. 2014) fits well the inferred demographic history. Despite the fluctuation of N*e* in the past, we found that both current N*e* and nucleotide diversity correlate linearly with island area. This confirms the findings from the previous chapters and from Hoeck et al. (2010b) that genetic drift is highly determined by current island area. In contrast, the degree of inbreeding detected by the analysis of runs of homozygosity was considerable only in the smallest populations (Champion, Darwin, Wolf and Gardner by Floreana). Champion's population in particular showed an exceptionally high degree of inbreeding, which is more typical of

partially selfing plants than of outcrossing vertebrates (Dole & Ritland 1993).

The analysis of genetic load as part of chapter IV suggested a possible role of purging (Crow 1970, Hedrick 1994). While realised (recessive) genetic load was an inverse function of island size, additive (total) load was lowest in the smallest population of Galápagos mockingbird, Darwin. Furthermore, a positive correlation was found between additive load and island area in the Galápagos mockingbird. This is a possible indication of more efficient selection against deleterious mutations in smaller populations compared to larger populations, a phenomenon known as purging. Our study therefore complements a growing body of evidence for the importance of purging (e.g. Robinson et al. 2018). Nevertheless, additive load was highest in the two extant populations of *M. trifasciatus*. This pattern may be related to the fact that the populations experienced a rapid bottleneck 150 years ago when the species went extinct on Floreana and there was not enough time to remove the burden (Hoeck et al. 2010a). Overall, the results of the last chapter suggest that although the effect of genetic drift increases as the island area is reduced, it does not lead to accumulation of a higher volume of deleterious mutations in small populations of the Galápagos mockingbird, whereas there is evidence of an accumulation of genetic load in the case of the Floreana mockingbird, possibly due to its particular demography.

The work offers a new perspective on long-standing questions in the field of conservation genetics, but also provides sources of information relevant to the actual conservation of mockingbirds in Galápagos. Using genomic data, my colleagues and I have made an estimate of the N*e* of the two remaining populations of Floreana mockingbird. Both populations have N*e* below 100, which is considered too low to avoid inbreeding depression and far too low to maintain evolutionary potential for fitness in the long term (Frankham et al. 2014). Even if we consider a disapproval of a "100/1000 Ne" rule by García-Dorado (2015), who

argues that the recommended levels can be lowered thanks to purging, the population size is critically low, especially in Champion with the N*e* of 14∓7. Both populations of this species have high levels of genetic load, which, together with a putative deleterious variant of TLR4 on Champion, indicates inefficiency of purifying selection. Together with the fact that divergence between populations is low, this provides sufficient justification for the planned genetic rescue of this species (Charles Darwin Foundation 2008). Genomic data such as the assembled and annotated genome of the San Cristóbal mockingbird and 24 resequenced genomes will serve as a resource for further conservation efforts in Galápagos mockingbirds.

Future perspectives

Herewith I would like to point out some future perspectives in the field of conservation genetics with regard to my thesis. By focusing only on correlative approaches between the determined genetic indices and island area, I realised that there is a desperate lack of experimental validation in this work, but also in the field in general. For example, in a major review on the use of population genomics for wildlife conservation and management, there is no mention of experimental validation (Hohenlohe et al. 2021). The effect of individual alleles could nowadays be tested directly with transgenic methods (Barrangou and Doudna 2016), and I would be tempted to estimate the fitness effects of some of the immune variants under different pathogenic conditions. But in the case of Galápagos mockingbirds, experiments are rather unattainable and we have to rely on natural experiments or work with *Drosophila melanogaster* (Pekkala et al. 2012) or plants where experimental manipulations are more common in conservation genetics (Willi et al. 2005, Allendorf et al. 2013). More reachable advance would be a validation of the expression of specific MHCIIβ alleles (Babik 2010), which would greatly enhance the credibility of my findings. Indeed, we spent some time with my undergraduate student on RNA extraction from

RNA-later fixed samples collected by our team, but unfortunately we failed due to RNA degradation.

Indices of genetic load that have been validated in model organisms (Simons & Sella 2016) should also be revalidated in a focal species group, as the distribution of fitness effects and the distribution of dominance are variable across species (Eyre-Walker & Keightley 2007, Huber et al. 2018, Bertorelle et al. 2022). In addition, other genetic load indices such as the Genomic Evolutionary Rate Profiling (GERP) (Cooper et al. 2005) or sorting tolerant from intolerant mutations (SIFT) (Ng & Henikoff 2003) could be used to validate our results (Bertorelle et al. 2022).

Simulation approaches will be another fruitful avenue in the field of conservation genetics. This approach of in silico evolution, where all parameters can be controlled (Haller & Messer 2017, Kyriazis et al. 2022), has great potential to validate results that reflect evolutionary changes over thousands of generations. Simulation would mimic evolution under specific demographic scenarios derived from genomic data. In this way, it is possible to test which selection and which demographic constraints reproduce the same patterns of empirically observed genetic load (Kyriazis et al. 2022). Furthermore, once validated, such simulations have the potential to predict future population persistence under hypothetical scenarios, which is a powerful tool for decision-making in conservation biology (Robinson et al. 2022).

In summary, I echo the recent arguments of Teixeira & Huber (2021a), who suggest that the importance of neutral diversity as an index of genetic health is inflated and that more attention should be paid to functional genetic diversity instead. By this statement, I do not mean to say that neutral markers are useless in conservation genetics, which is a common misunderstanding of Teixeira & Huber's arguments (DeWoody et al. 2021, García-Dorado & Caballero 2021, Teixeira and Huber 2021b).

Finally, I have been concerned about a hypothetical question that is looming above the field of conservation genetics: whether to fund a genomic study or rather buy a piece of land for the purpose of nature protection. With current empirical knowledge, it is clear that larger populations have better prospects (Allendorf et al. 2022), we do not need more population genetics to prove that. Perhaps it is still worth understanding how genetic rescue works (Hedrick & García-Dorado 2016) if we are to use it as a conservation measure. However, another question is whether genetic rescue is a sustainable approach (Hedrick et al. 2019), and I tend not to comment on the sustainability of de-extinction in this regard (Richmond et al. 2016). After spending a few years researching the genetic consequences of small populations, I realised that it is curiosity that motivates me to dive deep into the genomes of endangered species and into the landscapes of past changes, rather than the conservation aspect. I have uncovered interesting consequences of population genetics in populations of mockingbirds in the Galápagos Islands, thereby hopefully expanding human knowledge. However, I also realised that increasing the space for natural processes would bring more benefits to endangered wildlife than mere accumulation of subjective human knowledge about genetic processes in these small populations.

References

Ali JR, Aitchison JC (2014) Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos (JM Fernández-Palacios, Ed,). Journal of Biogeography, 41, 1227–1241.

Allendorf FW, Funk WC, Aitken SN, Byrne M, Luikart G (2022) Conservation and the Genomics of Population. John Wiley and Sons, New York.

Allendorf FW, Luikart G, Aitken SN (2013) Conservation and the genetics of populations. Wiley-Blackwell, Chichester.

Babik W (2010) Methods for MHC genotyping in non-model vertebrates. Molecular Ecology Resources, 10, 237–51.

Barrangou R, Doudna JA (2016) Applications of CRISPR technologies in research and beyond. Nature Biotechnology, 34, 933–941.

Bertorelle G, Raffini F, Bosse M et al. (2022) Genetic load: genomic estimates and applications in non-model animals. Nature Reviews Genetics, 23, 492–503.

Charles Darwin Foundation (2008) The reintroduction of the Floreana Mockingbird to its island of origin. Charles Darwin Foundation and Galápagos National Park, Puerto Ayora.

Cooper GM, Stone EA, Asimenos G et al. (2005) Distribution and intensity of constraint in mammalian genomic sequence. Genome Research, 15, 901–913.

Crow JF (1970) Genetic loads and the cost of natural selection. In: Mathematical topics in population genetics (ed Levin SA), pp. 128–177. Springer, Berlin.

DeWoody JA, Harder AM, Mathur S, Willoughby JR (2021) The long-standing significance of genetic diversity in conservation. Molecular Ecology, 30, 4147–4154.

Dole J, Ritland K (1993) Inbreeding depression in two Mimulus taxa measured by multigenerational changes in the inbreeding coefficient. Evolution, 47, 361–373.

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. Nature reviews. Genetics, 8, 610–8.

Fiddaman SR, Vinkler M, Spiro SG et al. (2022) Adaptation and Cryptic Pseudogenization in Penguin Toll-Like Receptors. Molecular Biology and Evolution, 39, 1–17.

Frankham R, Bradshaw CJA, Brook BW (2014) Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation, 170, 56–63.

García-Dorado A (2015) On the consequences of ignoring purging on genetic recommendations for minimum viable population rules. Heredity, 115, 185–187.

García-Dorado A, Caballero A (2021) Neutral genetic diversity as a useful tool for conservation biology. Conservation Genetics, 22, 541–545.

Geist DJ, Snell H, Snell H, Goddard C, Kurz MD (2014) A Paleogeographic Model of the Galápagos Islands and Biogeographical and Evolutionary Implications. In: The Galápagos: A Natural Laboratory for the Earth Sciences (eds Harpp K, Mittelstaedt E, D'Ozouville N, Graham DW), pp. 145–166. John Wiley & Sons.

Grueber CE, Wallis GP, Jamieson IG (2013) Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. Molecular ecology, 22, 4470–82.

Haller BC, Messer PW (2017) SLiM 2: Flexible, Interactive Forward Genetic Simulations. Molecular Biology and Evolution, 34, 230–240.

Hartmann SA, Schaefer HM, Segelbacher G (2014) Genetic depletion at adaptive but not neutral loci in an endangered bird species. Molecular Ecology, 23, 5712–5725.

Hedrick PW (1994) Purging inbreeding depression and the probability of extinction: Full-sib mating. Heredity, 73, 363–372.

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

Hedrick PW, Robinson JA, Peterson RO, Vucetich JA (2019) Genetics and extinction and the example of Isle Royale wolves. Animal Conservation, 22, 302–309.

Hoeck PEA, Beaumont MA, James KE et al. (2010a) Saving Darwin's muse: evolutionary genetics for the recovery of the Floreana mockingbird. Biology letters, 6, 212–215.

Hoeck PEA, Bollmer JL, Parker PG, Keller LF (2010b) Differentiation with drift: a spatio-temporal genetic analysis of Galápagos mockingbird populations (Mimus spp.). Philosophical Transactions of the Royal Society B: Biological Sciences, 365, 1127–1138.

Höglund J (2009) Evolutionary conservation genetics. Oxford University Press, Oxford.

Hohenlohe PA, Funk WC, Rajora OP (2021) Population genomics for wildlife conservation and management. Molecular Ecology, 30, 62–82.

Huber CD, Durvasula A, Hancock AM, Lohmueller KE (2018) Gene expression drives the evolution of dominance. Nature Communications, 9, 2750.

Kyriazis CC, Robinson JA, Lohmueller KE (2022) Using computational simulations to quantify genetic load and predict extinction risk. bioRxiv, 2022.08.12.503792.

Leigh DM, Hendry AP, Vázquez-Domínguez E, Friesen VL (2019) Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. Evolutionary Applications, 12, 1505–1512.

Li H, Xiang-Yu J, Dai G et al. (2016) Large numbers of vertebrates began rapid population decline in the late 19th century. Proceedings of the National Academy of Sciences of the United States of America, 113, 14079–14084.

Lighten J, Papadopulos AST, Mohammed RS et al. (2017) Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. Nature Communications, 8, 1294.

Maxwell SL, Fuller RA, Brooks TM, Watson JEM (2016) Biodiversity: The ravages of guns, nets and bulldozers. Nature, 536, 143–145.

Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Research, 31, 3812–3814.

Pekkala N, Knott KE, Kotiaho JS, Puurtinen M (2012) Inbreeding rate modifies the dynamics of genetic load in small populations. Ecology and Evolution, 2, 1791–1804.

Richmond DJ, Sinding MHS, Gilbert MTP (2016) The potential and pitfalls of de-extinction. Zoologica Scripta, 45, 22–36.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK (2018) Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. Current Biology, 28, 3487-3494.e4.

Robinson JA, Kyriazis CC, Nigenda-Morales SF et al. (2022) The critically endangered vaquita is not doomed to extinction by inbreeding depression. Science, 376, 635–639.

Schlaepfer DR, Braschler B, Rusterholz H-P, Baur B (2018) Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere, 9, e02488.

Simons YB, Sella G (2016) The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. Current Opinion in Genetics & Development, 41, 150–158.

Teixeira JC, Huber CD (2021a) The inflated significance of neutral genetic diversity in conservation genetics. Proceedings of the National Academy of Sciences of the United States of America, 118, 1–10.

Teixeira JC, Huber CD (2021b) Authors' Reply to Letter to the Editor: Neutral genetic diversity as a useful tool for conservation biology. Conservation Genetics, 22, 547–549.

Willi Y, Van Buskirk J, Fischer M (2005) A threefold genetic allee effect: Population size affects cross-compatibility, inbreeding depression and drift load in the self-incompatible *Ranunculus reptans*. Genetics, 169, 2255–2265.

Author’s Curriculum vitae

Author’s Curriculum vitae

Name: Mgr. Jakub Vlček

Date of Birth: 15/05/1990

Nationality: Czech

Address: Třebín 01, Dubné, 37384, Czech Republic

Phone: +420 773 124 333

Email: k.vlcak@gmail.com

ORCID: 0000-0002-2174-9374

Education

2015 – now PhD candidate in Zoology, Faculty of Science, University of South Bohemia.

2012 – 2015 Master degree in Zoology, Faculty of Science, University of South Bohemia.

2009 – 2012 Bachelor degree in Biology, Faculty of Science, University of South Bohemia.

2001 – 2009 Gymnázium Jiřího Wolkera Prostějov– biology and chemistry.

Employment History

2017 – now Technician and research assistant, Unit of Evolution and Diversity of Vascular Plants, Department of botany, Faculty of Science, Charles University.

2012 – now Technician and research assistant, Laboratory of Molecular Ecology, Institute of Parasitology, Biology Centre CAS.

2015 – now PhD student, Institute of Parasitology, Biology Centre CAS.

2010 – now Field ornithologist: Bird monitoring in Important Bird Areas (IBAs) and NATURA 2000, Czech Society for Ornithology.

2015 – 2016 Field ornithologist: Bird breeding atlas census in the Czech Republic, Czech University of Life Sciences Prague.

Internship

May – September 2018 Kirk Lohmueller and Robert Wayne labs, Department of Ecology and Evolutionary Biology at the University of California, Los Angeles, USA.

June 2017 Vladimír Beneš, EMBL, Genomics Core Facility, Heidelberg, Germany.

September 2014 Camila Mazzoni, BeGenDiv, Berlin.

Teaching

2016 – 2018 teaching assistant, Molecular ecology, Faculty of Science, University of South Bohemia.

2016 – 2018 teaching assistant, Conservation genetics, Faculty of Science, University of South Bohemia.

2017 – 2018 teaching assistant, Population genetics, Faculty of Science, University of South Bohemia.

2015 – 2020 teaching assistant, Introductory bachelor field course, Faculty of Science, University of South Bohemia.

Supervision of theses:

2019 Bc. Matěj Miláček

Variability of molecular diversity and protein structure

of Toll-like receptors in Galápagos mockingbirds. Bc. Thesis, in Czech, Faculty of Science, University of South Bohemia.

Publications

Bohutínská M, **Vlček J**, Monnahan P, Kolář F (2023) Population genomic analysis of diploid - autopolyploid species. Polyploidy, Methods and Protocols, edited by Yves Van de Peer, Humana New York, to be published in 2023, 515.

Vlček J, Miláček M, Vinkler M, Štefka J (2022) Effect of population size and selection on Toll-like receptor diversity in populations of Galápagos mockingbirds. Journal of evolutionary biology, In print.

Riegert J, Chmel K, **Vlček J** et al. (2021) Alarming declines in bird abundance in an Afromontane global biodiversity hotspot. Biodiversity and Conservation, 30, 3385–3408.

Bohutínská M, **Vlček J**, Yair S, Laenen B, Konečná V, Fracassetti M, Slotte T, et al. (2021) Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. Proceedings of the National Academy of Sciences 118, e2022713118.

Konečná V, Bray S, **Vlček J**, Bohutínská M, Požárová D, Choudhury RR, Bollmann-Giolai A, et al. (2021) Parallel adaptation in autopolyploid *Arabidopsis arenosa* is dominated by repeated recruitment of shared alleles. Nature Communications 12.

Kubelka V, Havlíček J, **Vlček J** (2021) Zajímavá a vzácná ornitologická pozorování v jižních Čechách VIII Interesting and rare ornithological records in South Bohemia VIII. Sborník Jihočeského muzea v Českých Budějovicích, Přírodní vědy, 61, 46–68.

Vlček J and Štefka J. (2020). Association between louse abundance and MHC II supertypes in Galápagos mockingbirds. Parasitology Research, 119, 1597–1605.

Holicová T, Sedláček F, Mácová A, **Vlček J**, Robovský J. (2018) New record of *Microtus mystacinus* in eastern Kazakhstan: phylogeographical considerations. ZooKeys, 781, 67–80.

Vlček J , Hoeck PEA, Keller LF , Wayhart J, Dolinová I, Štefka J. (2016) Balancing selection and genetic drift create unusual patterns of MHCIIβ variation in Galápagos mockingbirds. Molecular Ecology 25, 4757–4772.

Workshops and Conferences

2019 Congress of The European Society for Evolutionary Biology (Talk: Genetic load in populations of Galápagos Mockingbirds)

2016 European Meeting of PhD Students in Evolutionary Biology (Talk: Immuno-genetic variation, pathogen mediated selection and genetic drift in populations of Galápagos mockingbirds.)

2016 Workshop on Population and Speciation Genomics (Český Krumlov, Czechia)

2013 Workshop on Molecular Evolution, Europe (Český Krumlov, Czechia)

2013 Congress of The European Society for Evolutionary Biology (Poster: Variation in MHC II genes in the context of Galápagos Mockingbirds study system).

Other

Languages: Czech (native), English (fluent), Russian (fluent), Spanish (basic).

Programming languages: R, bash, Python

Github: https://github.com/vlkofly

E-mail: k.vlcak@gmail.com

Conservation genetics of Galápagos mockingbirds: from immune genes to genomes.
Ph.D. Thesis Series, 2022, No. 17

Printed in the Czech Republic by Typodesign
Edition of 15 copies

University of South Bohemia in České Budějovice
Faculty of Science
Branišovská 1760
CZ-37005 České Budějovice, Czech Republic

Phone: +420 387 776 201
www.prf.jcu.cz, e-mail: sekret-fpr@prf.jcu.cz