University of South Bohemia in České Budějovice

Faculty of Science

Variability of MHC class II β gene in Galápagos mockingbirds

Master thesis

Bc. Jakub Vlček

Supervisor: Jan Štefka, PhD

České Budějovice 2014

Vlček J., 2014: Variability of MHC class II β gene in Galápagos mockingbirds. Mgr. Thesis, in English. – 61 p., Faculty of Science University of South Bohemia in České Budějovice.

Anotace:

Understanding the dynamics of functional genetic variability in small populations can have important implications in their conservation. I screened the variation of MHC II β gene in Galapagos mockingbirds to evaluate the evolutionary forces that shaped the genetic variation. I found out that genetic drift affected the MHC variation together with a specific form of natural selection. Although the MHC is supposed to be under a pathogen-mediated selection I found no evidence for this theory in the mockingbird study system.

Prohlašuji, že svoji diplomovou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury. Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, 10.12.2014

Acknowledgements:

I thank everyone who helped me to generate this thesis, special thanks goes to Jirka Kosnar and Iva Dolinova for introduction to the world of second-generation sequencing; Camila Mazzoni, Marie Jeschek and Simone Sommer for introduction to the world of bioinformatics; Paquita Hoeck, Lukas Keller, and Glauco Camenisch for developing and sharing the study system; and to the study objects for sharing their DNA.

I thank my supervisor Honza Štefka who has the lion's share in the formation of this thesis.

Acknowledgement also goes to the Student Grant Agency, the Faculty of Scince, Czech Science Foundation and Václav Hypša for finantial support of the research.

Contents

1.Introduction and aims....1
2.Literature review....3
2.1.Major histocompatibility complex....3
2.2.Study system....12
2.3.MHC genotyping method overview....17
3.Methods....18
3.1.Samples and DNA....18
3.2.Primer design....19
3.3.Amplicon preparation and sequencing....21
3.4.Allele identification....23
3.5.Statistical analysis of MHC alleles....24
3.6.Molecular variation....25
3.7.Selection....26
4.Results....26
4.1.MHC sequencing and allele filtering....26
4.2.MHC diversity....28
4.4.MHC variation and environmental factors in the GM....31
5.Discussion....34
5.1.Variation and genetic drift....35
5.2.Allelic divergence....37
5.3.Ancestral polymorphism....38
5.4.MHC and immunity....40
5.5.MHC and data quality....40
6.Conclusions....43
7.References....44
8.Appendix....59

1. Introduction and aims

Endangered species surviving in small populations often experience a loss of genetic variation because of inbreeding and random genetic drift (Höglund 2009). The loss of genetic variation can lead to severe reduction of their viability. In the long run it can cause a loss of evolutionary potential, i.e. the ability of populations to adapt to environmental changes. Thus, understanding the causes and consequences of the loss of genetic variation in endangered populations is an essential issue in conservation biology (Höglund 2009).

Substantial attention in conservation genetics was devoted to the studies of *neutral genetic variation* (eg. microsatellites) in endangered populations (Ouborg *et al.* 2010). However, the current issues concerning the loss of adaptive potential and impairment of population viability are much more related to the functional (adaptive) genetic variation. Therefore, studying the *adaptive genetic variation* can provide another level of knowledge useful in effective conservation of endangered animals (Sommer 2005; Ouborg *et al.* 2010).

Coding genes of the Major Histocompatibility Complex (MHC) play a crucial role in the adaptive immune system through MHC encoded glycoproteins that present antigens to the T-cells, which eventually trigger an immune response. The number of MHC variants in an individual defines the range of pathogens which the immune system can recognize and fight off (Murphy & Janeway 2008). Consequently, the MHC contains the most polymorphic genes in jawed vertebrates (de Bakker & Raychaudhuri 2012) and the polymorphism is considered adaptive. Natural selection is an important force shaping the MHC polymorphism. This is the main difference between the neutral variation, where only the stochastic micro-evolutionary forces (eg. genetic drift) act, and the adaptive MHC variation, where the natural selection and the stochastic forces interplay. A relative role of the selection and the random genetic drift on the MHC variation was studied either using individual based model simulations or by comparisons of the MHC and the neutral variation in natural populations (Sutton *et al.* 2011). A clear consensus on this issue has not been reached yet, although it seems that genetic drift in a small population will reduce the variation even in the loci under natural selection. Sufficient evidence is lacking to support or refute general validity of many claims. For example, it is not clear weather lower MHC variation is directly linked with lowered viability of the populations (Radwan *et al.* 2010) and higher pathogen prevalences. It rather seems that particular MHC alleles confer some form of resistance, but the entire variation is not as important for the immunocompetence (Westerdahl *et al.* 2013; Sepil *et al.* 2013).

The Galapagos mockingbirds (GM) comprise 4 endemic species distributed across the Galapagos archipelago. The populations vary in effective population size, which is linked with the size of the island. Consequently, there is proportional effect of the genetic drift and inbreeding in each population, shaping the variation in neutral loci. Because of low dispersal the populations are effectively isolated and gene flow between the islands is restricted. The effect of the genetic drift on neutral loci was assessed by microsatellites (Hoeck *et al.* 2010) and the intensity of ectoparasite infestation together with immunological indices were analysed (Hoeck & Keller 2012). Thus the set of the insular GM populations offers an ideal model for the investigation of the effect of selection on MHC genes. Because the gene flow is absent and the influence of genetic drift on neutral loci is known, I can analyse the patterns of variation caused exclusively by the natural selection. Also, individual islands differ in their parasite regime which can drive the natural selection in MHC. Finally, there are depauperate populations of the Floreana mockingbird (*Mimus trifasciatus*) with only less than 200 individuals surviving. This species is critically endangered and an assessment of the variation in its MHC genes can complement the information on health of its populations, which is available from direct observations and studies of neutral loci (Hoeck & Beaumont 2010; Deem *et al.* 2011).

The aims of this thesis are following:

- Review current knowledge of relevant aspects of the structure, function and evolution of the MHC.
- Characterise variation in the functional MHC genes in the GM and in a population of a related species (Northern mockingbird) using second-generation sequencing.
- Describe patterns of MHC variation across species, populations and individuals.
- Analyse the effect of random genetic drift and natural selection on the MHC variation.
- Explore possible correlations between immunological/ecological factors and the MHC variation.

2. Literature review

2.1. Major histocompatibility complex

The Major histocompatibility complex (MHC) is a diverse multi-gene family found in all jawed vertebrates(Kasahara 2000). The MHC is best known for its function in the immune system, although not all genes are strictly linked with immunological function (Kasahara 2000). Nevertheless the most studied MHC genes (so called "classical") code for glycoproteins which are the cardinal component of the innate immune response (Murphy & Janeway 2008). Through this text only these classical MHC genes are referred to using the abbreviation "MHC". The physiological function of MHC lies in presentation of peptides derived from proteins to T-cells. The presentation comprises binding the peptide onto peptide binding groove of the MHC molecule and exposing this part into extracellular space where it stimulates the T-cell via a T-cell receptor. The T-cell evaluates the signal and either exterminates the signalling cell or further regulates adaptive immune response (Murphy & Janeway 2008).

Two MHC classes are distinguished according to molecular structure, origin of the presented peptide and type of the involved T-cell. The MHC class I molecules present peptides (antigens) derived from proteins found in the cytosol (mostly viral origin) to CD8+ T-cells (cytotoxic T-cells). If the particular T-cell recognizes the particular antigen it releases cytotoxins which eventually cause an apoptosis of the signalling cell (Murphy & Janeway 2008). The MHC class II molecules present peptides from endocytosed proteins of endoparasitic or bacterial origin to CD4+ T-lymphocytes (helper T-cells). When the antigen is recognized by the T-cell the information about the pathogen is passed to a B-cell and the adaptive immune response is regulated accordingly. Importantly, each variant (allele) of classical MHC gene can bind and present only a limited range of pathogenic proteins (pathogens) based on the peptide binding groove specificity (Falk *et al.* 1991). Thus different variants of MHC genes can confer resistance to a restricted range of pathogens (Martin & Carrington 2005; Milinski 2006) -which implies the central role of polymorphism in the function of MHC genes.

The MHC genes show the highest polymorphism from all genes in jawed vertebrates. In some species the total number of alleles can reach hundreds, like in Blue tit (*Cyanistes caeruleus*), Collared flycatcher (*Ficedula albicollis*), or human (Zagalska-Neubauer *et al.* 2010; Sepil *et al.* 2012; de Bakker & Raychaudhuri 2012). Interestingly, the actual MHC

polymorphism was the first thing which struck the scientists in 1968 even before they knew its function (Snell 1968). The number of alleles in the population is probably linked to a variable number of copies of a given MHC gene within individual genome. This feature is known as copy number variation (CNV) (Schrider & Hahn 2010).

Because of the high polymorphism, importance in immune pathways and putative effect on various traits in a species' life history, the MHC became widely studied model of adaptive molecular evolution (Piertney & Oliver 2006; Spurgin & Richardson 2010). The main goal of adaptive evolutionary studies and, particularly the MHC studies, is to understand the effect of selection on molecular variation and the correlated effect of molecular variation on individual fitness. Specific MHC molecules involved in the binding of pathogenic peptides, like the molecule coded by exon 2 of subunit β of MHC II (hereafter MHCII β2), are often used in such studies, as this variability is considered adaptive.

It is widely agreed that polymorphism in MHC genes is maintained by some forms of balancing selection which is a result of arms race between hosts and parasites (Bernatchez & Landry 2003; Milinski 2006; Spurgin & Richardson 2010). Currently, there are at least three plausible hypotheses specifying the form of pathogen mediated balancing selection: First, Heterozygote advantage (Doherty & Zinkernagel 1975) proposes that when an individual is heterozygous at MHC loci it has higher fitness than both homozygotes (overdominance) because it can recognize a wider range of pathogens. Under this form of balancing selection there will be higher overall heterozygosity in the population leading to higher number of MHC alleles. Second, rare allele advantage hypothesis (Slade & McCallum 1992) is based on a negative frequency dependence between the pathogens and the alleles of the MHC genes. Pathogens are strongly selected to overcome the most frequent allele, and subsequently the successful pathogen will dominate. But then a rare allele, which can recognize such pathogen, provides good protection and high fitness and therefore it spreads and eventually dominates in the population. The circle is closed and fluctuations continue to maintain all rare alleles on the brink of loss from the population. Third, fluctuating (diversifying) selection hypothesis argues that changes of parasite regime in space and time will result in the high polymorphism at the MHC genes (Hill 1991). These three possible mechanisms of pathogen mediated balancing selection are not mutually exclusive and they can interact, thus it is hard to disentangle their effects and roles (Spurgin & Richardson 2010).

Moreover a selection against deleterious mutations (Oosterhout 2009b) developed under scenario of Associative Balancing Complex (ABC) can cause high MHC polymorphism even the without effect of pathogens. The recessive deleterious mutations are being cumulated around the functional MHC genes and the high polymorphism shelters these mutations, because the more MHC alleles the lower the probability that the deleterious mutations will happen to be in homozygous state and possibly lethal. Particularly the ABC hypothesis highlights the fact that previously mentioned forms of selection based on single gene model are not sufficient to explain the whole complexity of selection in the multigene family (Oosterhout 2009a).

MHC polymorphism was also proven to be affected by sexual selection (Penn & Potts 1999). Mate choice is often MHC-dependent in a way that females choose males with compatible MHC genotypes to produce a viable progeny (Milinski 2006). MHC dependent mate choice is often reported in mammals and fish, but the similar patterns were found also in birds where the issue was until recently controversial (Richardson *et al.* 2005; Bonneaud *et al.* 2006; Griggio *et al.* 2011; Baratti *et al.* 2012; Alcaide *et al.* 2012; Strand *et al.* 2012; Dunn *et al.* 2013).

MHC is also involved in maternal-foetal interactions (Ober *et al.* 1998) although this mediator of selection is relevant only in mammals.

Upper limit of MHC variability within an individual is determined by a negative T-cell selection, and this selective force can counteract the polymorphism generating forces described above. The more MHC alleles an individual bears the more T-cell variants will be removed from its repertoire because of auto-reaction of the T-cells and the MHC alleles during T-cell maturation (Nowak 1992; Woelfing *et al.* 2009). A hypothetical individual possessing excessively high polymorphism in the MHC would suffer a shortage of T-cells and impaired function of the immune system, therefore rather individuals with intermediate (optimal) levels of MHC variation are thought to be the fittest (Forsberg *et al.* 2007; Kalbe *et al.* 2009). This is a possible explanation for the observations of females choosing optimal rather than maximal levels of the MHC polymorphism (Woelfing *et al.* 2009).

Apart from the variety of natural selection forms, the MHC polymorphism is also shaped by stochastic molecular processes (recombination and mutation) and demography-dependent processes like genetic drift and gene flow. 1.) Recombination is known to affect polymorphism at MHC on various levels. Classical meiotic recombination is typical for this

multi-gene family, taking place in recombination hotspots located in intronic parts and interspersed with stretches of high linkage disequilibria (Jeffreys & May 2004). Polymorphism on this level is observed as differences in hotspot positions across human populations (Lam *et al.* 2013). Non-allelic homologous recombination (NHEJ) (Hurles & Lupski 2006) causes MHC multiplications and it is one of the mechanism generating the CNV in the MHC (Schrider & Hahn 2010). No less important is gene conversion which is defined as a recombinatorial transfer of DNA between alleles or loci without crossover (Jeffreys & May 2004). It causes a reshuffling of the present MHC variation and formation of new alleles (Ohta 1991). Although it is still surrounded by controversy, whether the recombination and gene conversion produce more polymorphism than point mutations (Nei & Rooney, 2006; Jan Klein, Sato, & Nikolaidis, 2007), a growing number of studies are supporting a substantial effect of recombination on MHC polymorphism in birds (Hosomichi *et al.* 2008; Chaves *et al.* 2010; Burri *et al.* 2010; Spurgin *et al.* 2011; Promerová *et al.* 2013; Strand *et al.* 2013). Recombination also plays a major role in so called birth-and-dead model (Nei *et al.* 1997) which is often applied to explain evolution of MHC genes. The model assumes that new genes emerge by duplication (specific form of recombination) and while some duplicates are maintained for long periods of time (even across speciations) others are either deleted or become nonfunctional. It is an opposite hypothesis to the concerted evolution model (Nei & Rooney 2005) where the duplicated genes are unified by the recombination and in one population evolve in concert. Also this concerted model is sometimes linked with evolution of the MHC (Burri *et al.* 2008).

2.) Mutations represent another main source of variation in MHC (Hosomichi *et al.* 2008). In human it was revealed that mutation rate at MHC genes is higher than the mean genomic mutation rate (Gaudieri *et al.* 2000). This feature of MHC was observed also in birds (Wild turkey) (Chaves *et al.* 2010). Point mutations seem to be predominant because of the high frequency of point non-synonymous mutations observed in the peptide binding region (Wang *et al.* 2013; Alcaide *et al.* 2014). Although in the chicken (Hosomichi *et al.* 2008) it was found that indels were pervasive.

3.) Genetic drift is of special concern to wildlife immunogenetics. In a genetically depleted (and endangered) population genetic drift can lead to further reduction of vital variability in the MHC genes (Alcaide 2010). In general, genetic drift theory posits that allelic frequencies are affected by random sampling during reproduction. And with the decreasing (effective) population size the strength of the drift increases, ultimately causing

an allele loss (Kimura 1983). In the studies of the MHC in non-model and wild species a central question often appears: Whether the drift can outweigh balancing selection (Miller & Lambert 2004b). As the drift and the balancing selection have contradicting effect on MHC polymorphism it is interesting to ask which force will dominate under different scenarios. The question was often tested by comparison of the neutral variability (effect of drift) and the MHC variability (interaction of drift and balancing selection) in populations of different sizes where the strength of the drift varies. Initially it was suggested that the balancing selection will maintain MHC polymorphism in contrast to the neutral markers like microsatellites (Hedrick 2001). This hypothesis recently turned controversial as there is growing empirical evidence that in small populations the drift will outweigh the balancing selection and the variability in the MHC will decrease proportionally with neutral variability (Sommer & Tichy 1999; Miller & Lambert 2004b; Radwan *et al.* 2007; Munguia-Vega *et al.* 2007; Siddle *et al.* 2010; Eimes *et al.* 2013). The issue was also reviewed several times. In the meta-analysis of post-bottlenecked populations Sutton et al. (2011) found that MHC variability drops by ~15% more than the neutral variability. This pattern was also found in several recent studies (Alcaide *et al.* 2010; Eimes *et al.* 2011). Two mutually non-exclusive explanations of this pattern exist. Drift across loci hypothesis (Eimes *et al.* 2011) proposes that drift can fix the same duplicated alleles across loci and thus reduce MHC variability two-fold compared to the non-duplicated neutral loci. Alternatively, the negative frequency dependent selection deflects frequencies of rare alleles which are then removed from small populations by the drift more frequently than alleles at neutral loci (Ejsmond & Radwan 2011). In the later article where authors used simulation approach, it was found that under a pronounced bottleneck the variation in MHC was lost to a greater degree than at neutral loci, but some 40 generations after the bottleneck the MHC diversity again recovered with the increasing number of individuals. Although the situation seems to be close to a consensus that the genetic drift in small populations has a major role over the balancing selection, there are particular cases in which putative natural selection maintained the MHC variability despite the depleted neutral variability. The high MHC but low microsatellite polymorphism was found in San Nicolas Island fox (*Urocyon littoralis*) (Aguilar *et al.* 2004), guppy (*Poecilia reticulata*) (Oosterhout *et al.* 2006) and Berthelot's pipit (*Anthus berthelotii*) (Spurgin *et al.* 2011) .

The last neutral process which affects polymorphism of the MHC is gene flow. The gene flow, defined as a transfer of alleles from one population to another, has some

peculiarities when MHC with its multiplicated loci is concerned. An incoming individual can introduce twice as many alleles as is the number of its MHC loci. It causes much more pronounced effect of gene flow compared to the introgression of alleles at non-multiplicated loci. Based on this theory McMullan & van Oosterhout (2012) argued that rapid MHC allelic turnover in guppies is not necessarily caused by balancing selection, but it can be explained solely by the demographic changes without the effect of pathogens.

All of the processes, from the various forms of selection to the gene flow, can interplay and shape the intricate patterns of MHC diversity in wild populations. Thus, it is often tricky or impossible to disentangle the roles of the separate forces which are dependent on various factors in the population. Moreover, the function of the MHC can be substituted by another branch of the immune system (Acevedo-Whitehouse & Cunningham 2006). Therefore, it is not surprising that empirical studies of the effect of MHC diversity on viability of populations or resistance to pathogens often produce ambiguous results and even more equivocal explanations.

In general not many studies supported the importance of MHC diversity on long term viability or survival of populations (reviewed in Radwan *et al.* 2010). The examples of viable populations with very low MHC diversity can be found in Eurasian beaver (*Castor fiber*) (Babik *et al.* 2005), Mountain goat (*Oreamnos americanus*) (Mainguy *et al.* 2006) or Black robin (*Petroica traversi*) (Miller & Lambert 2004b). In the last two examples the intensity of infection by parasites was not affected by MHC polymorphism. This pattern is not so rare and it implies that the diverse immune system can possibly substitute the function of the MHC by its other component (Acevedo-Whitehouse & Cunningham 2006). On the other side stands the Tasmanian devil (*Sarcophilus harrisii*) whose low MHC diversity was shown to be connected to the spread of contagious face cancer (Siddle *et al.* 2007). Similar pattern was also found in the Collared flycatcher where the prevalence of malaria decreased with increasing functional diversity of the MHC (Radwan *et al.* 2012). And the last compelling example of positive dependence between MHC variability and pathogen resistance was found in House finch (*Carpodacus mexicanus*), where the bacterial infections (*Mycoplasma gallisepticum*) were less severe in individuals with intermediate to high MHC diversity (Hawley & Fleischer 2012). The last example together with results from similar studies on House sparrows (Bonneaud *et al.* 2006) and Bank voles (Kloch *et al.* 2010) corroborate the optimal variability hypothesis, which proposes that higher fitness should be observed in individuals with intermediate MHC diversity.

Apparently, the most frequently observed pattern when investigating links between the MHC diversity and pathogens is that particular MHC alleles confer either a resistance or a susceptibility. The comprehensive study in Great tits (*Parus major*) found one allele linked with resistance to *Plasmodium relictum* and another allele with susceptibility to *Plasmodium circumflexum* (Sepil *et al.* 2013). But rather than the susceptibility, the second allele conferred quantitative disease resistance and the first allele conferred qualitative disease resistance based on a theory introduced by Westerdahl et al.(2012). Various other studies found alleles conferring either the qualitative or the quantitative disease resistance eg.: in Blue tit (Westerdahl *et al.* 2013), Common yellowthroat (*Geothlyps trichas*) (Dunn *et al.* 2013), House sparrow (*Passer domesticus*) (Loiseau *et al.* 2008), European rabbit (*Oryctolagus cuniculus*) (Oppelt *et al.* 2010) and 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 the pathogen resistance. Maybe these alleles were the former rare alleles which were winning the current strains of pathogens? Also, the fact that some of the studies found population specific resistance alleles(Ekblom *et al.* 2007; Hawley & Fleischer 2012) supports the interplay of the rare allele advantage and the fluctuating selection. However, so far any consensus is within sight.

Concerning links between the MHC diversity and the distribution of ecto-parasites in birds, not much work has been done so far. Only in chicken the specific MHC alleles confered either susceptibility or resistance to bloodsucking mites (Owen *et al.* 2008, 2009). Owen et al. also suggested the physiological basis of the resistance through MHC dependent inflammatory response, which causes thicker epidermis hindering bloodsucking by the mites.

General function of the MHC genes is identical across vertebrate classes, although some differences exist in function and especially in genomic organization (Kelley *et al.* 2005). First analysis of the bird's MHC genes suggested pronounced difference to the mammal's MHC as the chicken genes were extremely simple (Kaufman *et al.* 1995). But later was found that the genetic structure of MHC in birds varies from the so called minimal essential MHC in *Galliformes* (Kaufman *et al.* 1995) to highly variable in songbirds (*Passeriformes*) (Westerdahl *et al.* 2000; Balakrishnan *et al.* 2010; Ekblom *et al.* 2011). The minimal essential stands for extremely small, non-multiplicated and linked MHC genes with no pseudogens which were found apart from the *Galliforms* (Shiina *et al.* 2004; Chaves *et al.* 2007) also in kestrels (Alcaide *et al.* 2007), some parrots (Hughes *et al.* 2008) and

penguins (Bollmer *et al.* 2007). On the other hand, songbird MHC structure resembles more the organization observed in mammals, where longer introns and more multiplications are present together with higher recombination rate and many pseudogens (Westerdahl 2007; Balakrishnan *et al.* 2010; Ekblom *et al.* 2011). At least 3 and up to 20 loci were found in the MHCII β2 of songbirds (Tab. 1). Originaly it seemed like the more basal lineages of birds had the minimal essential MHC as the *Galliforms*, and throughout bird evolution the number of MHC genes increased, but recent studies showed that the number and structure of MHC genes is simply lineage variable. In the Little spotted Kiwi (*Apteryx owenii*) at least 5 MHCII β2 alleles were observed (Miller *et al.* 2011), in Eurasian coot (*Fulica atra*) at least 3 (Alcaide *et al.* 2014), 4 loci were found in Chinese egret (*Egretta eulophotes*) (Wang *et al.* 2013) and 2 putative loci were found found in Kakapo (*Strigops habroptilus*) (Knafler *et al.* 2014). Although it should be noted that the number of loci can be biased by the use of different primers across the studies.

Understanding of the evolutionary history of individual MHCII β2 lineages in birds is hindered by a high degree of concerted evolution which unifies the lineages within species and conceals orthologous relationships (Hess & Edwards 2002). However, ancient gene duplication, which preceded major avian radiation, was revealed by Burri et al. (Burri *et al.* 2010)From the two lineages (DAB1, DAB2), which are jointly present in owls, only one (DAB1) was traced in the songbirds. Therefore all allelic lineages of the MHCII β2 in songbirds are considered monophyletic. Clustering of the MHC alleles within songbirds often does not reflect the species tree, and many alleles are shared across genera showing some level of trans species polymorphism (Sato *et al.* 2011).

The last typical feature of MHCII β2 in songbirds is occurrence of the same alleles across loci caused probably by gene conversion (Westerdahl 2007). Therefore it is impossible to differentiate between loci according to allele similarity and thus multi-locus amplification of MHCII β2 is unavoidable approach for MHC genotyping in songbirds.

Species (family)	Method	№ of loci (alleles/ind.)	№ of alleles	Citation
Collared flycatcher (Muscicapidae)	Roche 454 seq.; 237 ind.	9 (NA)	194	(Zagalska-Neubauer *et al.* 2010)
Bluethroat (Turdidae)	Sanger seq. and clon.; only transcribed; 20 ind.	11 (3-20)	61	(Anmarkrud *et al.* 2010)
Nesospiza buntings in Tris. Da Cunha (Emberizidae)	Sanger seq. and clon.;16 ind.	4 (3-7)	23	(van Rensburg *et al.* 2012)
Common yellowthroat (Parulidae)	RFLP of 4 ind.	8 transcribed; 20 non-transcribed (NA)	NA	(Bollmer *et al.* 2010)
Zebra finch (Estrildidae)	Seq. of genome and BAC	5 (NA)	NA	(Balakrishnan *et al.* 2010)
New Zealand Robin (Petroicidae)	RT-PCR and RFLP; only transcribed; 2 ind	4 (NA)	NA	(Miller & Lambert 2004a)
Dark eyed junco (Emberizidae)	Sanger seq. and clon.; 100 ind.	4 (1-7)	73	(Whittaker *et al.* 2012)
South Island saddleback (Callaeidae)	Ion Torrent seq.; 20 ind.	2 (1-4)	4	(Sutton *et al.* 2013)
North Island saddleback (Callaeidae)	Ion Torrent seq.; 19 ind.	6 (3-12)	16	(Sutton *et al.* 2013)
Darwin Finches + others (39 spp Emberizidae, Parulidae, Icteridae, Cardinalidae, Thraupidae, Fringilidae)	Sanger seq. and clon. + southernblot; 117 ind.	7 (1-14)	321	(Sato *et al.* 2011)
House finch (Fringilidae)	Sanger seq. and SSCP	5 (1-10)	NA	(Hawley & Fleischer 2012)
Blue tit (Paridae)	SSCP and Southernblot; only transcribed; 22 ind.	4 (4-7)	13	(Aguilar *et al.* 2013)
House sparrow (Passeridae)	DGGE; 4 ind.	3 (1-5)	13	(Bonneaud *et al.* 2004)

Table 1: Number of loci and alleles found in studies of MHCII β2 in songbirds. Study specific methods are briefly described (seq.=sequencig, clon.=clonning, ind.=number of individuals genotyped).

2.2. Study system

Galapagos islands are inhabited by 4 species of mockingbirds (*Mimus*) with allopatric distribution (Fig.1). Galapagos Mockingbirds (GM) comprise a monophyletic lineage within mockingbirds and trashers (Lovette *et al.* 2012), which implies a singe colonization event of the archipelago. Based on the mutation rate of mtDNA it was estimated that the colonization took place between 1.6 – 5.5 million years ago (Arbogast *et al.* 2006). Since that time the GM formed 4 traditionally described species scattered over the archipelago. San Cristobal Mockingbird (*Mimus melanotis*, Gould 1837) occurs on a single, easterly placed island of identical name. The species is considered as endangered because of decreasing population trend caused by habitat degradation, introduced diseases and predators like cats and black rats (BirdLife International 2013). The number of mature individuals is estimated to c. 5300. Espanola Mockingbird (*Mimus macdonaldi*, Ridgway 1890) occurs, as it name suggests, on the Espanola island and additionally on a small islet Gardner by Espanola. This species is vulnerable with c. 600-1700 mature individuals (BirdLife International 2012). Fortunately, Espanola is free of introduced parasites and predators. The third species, Floreana mockingbird (*Mimus trifasciatus*, Gould 1837), was extirpated from the Floreana island around 1880 (Curry 1986) and now only two minute populations survive on adjacent islets Gardner and Champion by Floreana. The conservation status of this species is critically endangered with population of c. 50 mature individuals subjected to extreme climate-induced fluctuations (BirdLife International 2013). The last species, Galapagos mockingbird (*Mimus parvulus*, Gould 1837), inhabits all other major Galapagos islands and comprises 6 traditionally described subspecies (Tab. 2). Its conservation status is qualified as the least concern, because the populations are stable and the species is considered as common, although the global population size has not been quantified (BirdLife International 2012).

Subspecies name	Description	Distribution
M. p. hulli	Rothschild 1898	Darwin
M. p. wenmani	Swarth 1931	Wolf
M. p. personatus	Ridgway 1890	Pinta, Marchena, Santiago, Rabida
M. p. bauri	Ridgway 1894	Genovesa
M. p. parvulus	Gould 1837	Isabela, Fernandina, Daphne, Santa Cruz
M. p. barringtoni	Rothschild 1898	Santa Fe

Table 2: Subspecies of Galapagos mockingbird (*Mimus parvulus*) with author of description and distribution.

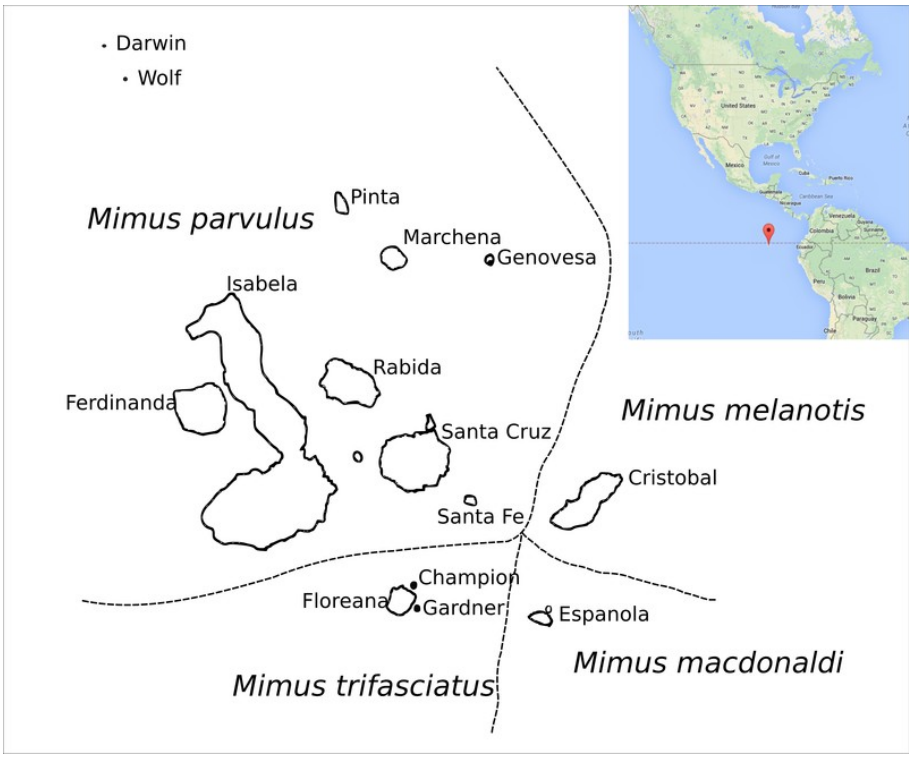


Figure 1: Illustrative map of Galapagos archipelago showing distribution of Galapagos mockingbirds.

Ecology of the GM is interesting from several aspects. First, the breeding success is tightly dependant on weather conditions as the mockingbirds breed only during rainy season (Grant & Grant 1979). Specially the El Nino Southern Oscillation can strongly affect breeding success in the whole archipelago. During dry La Nina periods there are almost no new fledglings, on the other hand the wet and warm El Nino periods cause high breeding success and better survival (Curry & Grant 1989). These weather fluctuations can have serious detrimental effect especially on small endangered populations causing high fluctuations in population size (Grant *et al.* 2000). Second interesting aspect of the GM is, that all species show some degree of cooperative breeding (Curry & Grant 1989). The helpers are often recruited from the offspring of the main breeding pair and these "families" can comprise as many as 24 individuals. The cooperative breeding in the GM is explained by a limited availability of appropriate breeding territories for younger birds. From the view of population genetics all these peculiarities of mockingbird's ecology will rather negatively affect the effective population size (Ne) and consequently the genetic variation in the populations.

The evolutionary history of the GM was uncovered thanks to many in-depth studies. The phylogenetic relationship was first reconstructed on the basis of mitochondrial DNA

variation (Arbogast *et al.* 2006). Apart from the first support of a single colonization event Arbogast et al. found a basal separation of two phylogenetic clusters of the GM (Fig. 2). The first cluster included two species from the eastern part of the archipelago (Espanola and San Cristobal mockingbirds) and a Genovesa population of the Galapagos mockingbird (*M. parvulus bauri*). The second cluster held all other populations with basal split between Floreana mockingbird and all other populations of the Galapagos mockingbird. The paraphyly among populations of the Galapagos mockingbird (Genovesa population clustered with eastern species) was not found in a phylogenetic analysis based on nuclear DNA variation of microsatellites (Hoeck *et al.* 2010). This study supported clustering according to the species affiliation, and stressed out the difference in the topologies based on nuclear and mitochondrial DNA. Recently this discordance between the gene trees was explained by a hybrid origin of the Genovesa population formed by cross-breeding of San Cristobal and Galapagos mockingbirds (Nietlisbach *et al.* 2013). Another independent view on the evolution of the GM brought the study of Stefka at al. (2011) where the phylogenies of the GM and their ecto-parasites were compared. The study showed that 2 ectoparasites (*Analgid* mite and *Amblyceran* louse) closely followed the evolution of the mockingbirds. The resulting common phylogenetic tree supported the split between the older eastern species (San Cristobal and Espanola mockingbird) and the younger western species (Galapagos mockingbird and Floreana mockingbird, which clustered inside the Galapagos mockingbird). The general phylogenetic consensus of the GM thus implies the gradual colonization of the archipelago from the eastern, older, islands (>2,5My) through the central islands (1-2.5My) toward the western and the youngest islands (<0.5My).

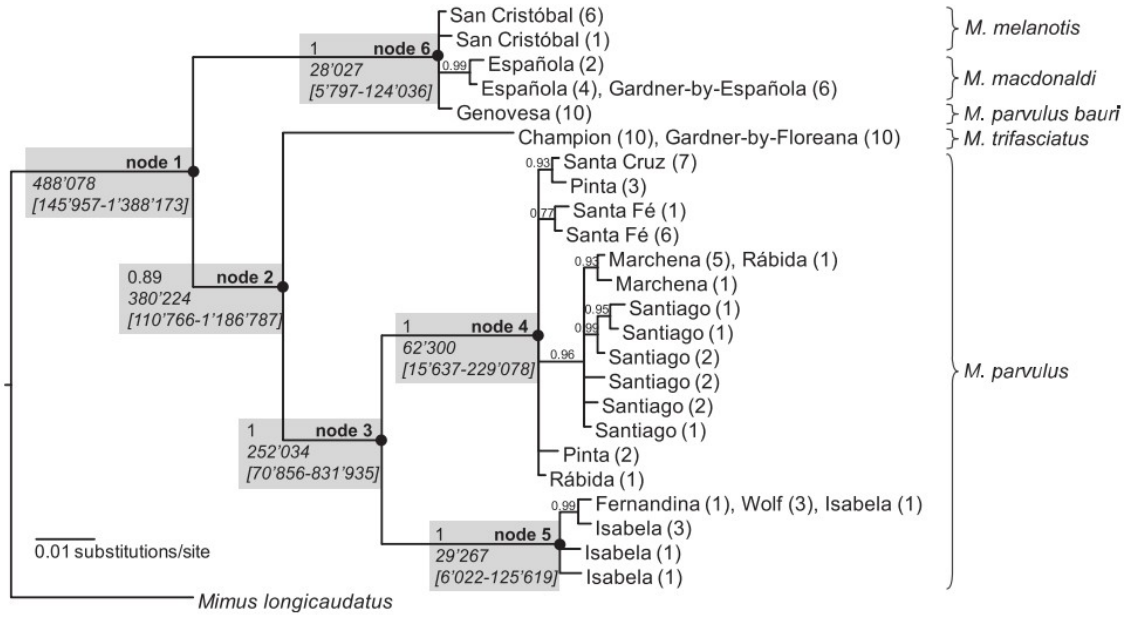


Figure 2: Mitochondrial phylogeny of the GM. Phylogenetic tree derived from mitochondrial ND2 sequences, above branches are Bayesian posterior probabilities and below branches in italics the mode and 95% highest posterior density intervals of split date estimates as derived from a relaxed clock model. Adopted from (Nietlisbach *et al.* 2013)

Apart from the evolutionary history the previous studies (Hoeck *et al.* 2010; Štefka *et al.* 2011) also revealed that there is almost no recent gene flow between the populations of the GM defined by islands. This observation corroborates the sedentary nature of mockingbirds. Because of absent gene flow and varying population sizes Hoeck et. al. (2010) also observed pronounced differences in neutral genetic polymorphism between the populations of the mockingbirds. Size of the island was the main factor determining effective population size and consequently the effect of random genetic drift and inbreeding in the population. Therefore the genetic diversity was significantly lower in smaller islands than in larger islands. For example, microsatellite allelic richness was twice smaller in the restricted population of the Champion islet compared to the population on the biggest island of Isabela (Fig 3.). These clearly isolated populations with remarkable variation in the strength of the genetic drift and inbreeding represent a unique model system perfectly suitable for investigation of the effect of neutral forces on the adaptive variation in the MHC genes.

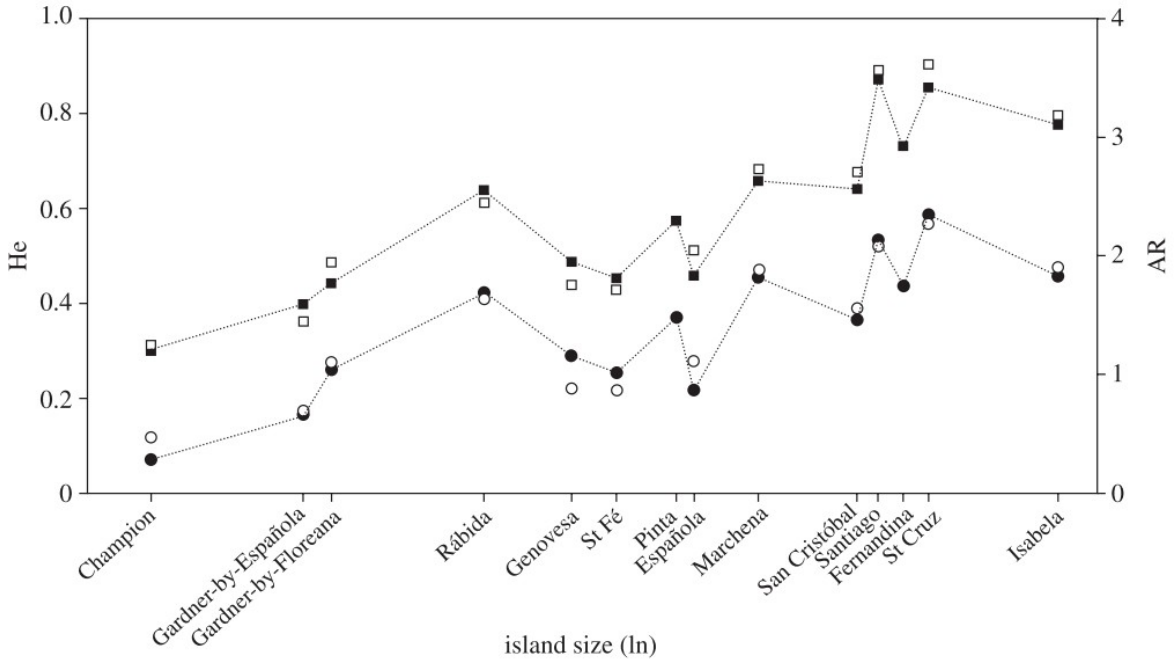


Figure 3: Expected heterozygosity (He) and allelic richness (AR) of contemporary and historic populations as a function of the natural logarithm of island size (in ha). To improve visual representation, points that overlapped were slightly moved (filled circle, contemporary He; open circle, historic He; filled square, contemporary AR; open square, historic AR. Adopted from (Hoeck *et al.* 2010)

Hoeck & Keller (2012) also investigated the effect of the microsatellite genetic diversity on immunocompetence and ecto-parasite load. The link between low genetic diversity and inferior immune defence was often reported from several species (Pearman & Garner 2005; Reid *et al.* 2007). Nevertheless none such correlations between any indices of immunocompetence or parasite prevalence and microsatellite diversity were found in the mockingbirds. Several indices of immune health were considered: ratio of heterocytes to lymphocytes, which is known to reflect infection status and physiological stress in birds (Maxwell 1993; Tompkins *et al.* 2006), then cell lysis and mean agglutination (measures directly linked with number of natural antibodies), which reflected state of innate immune system (Matson *et al.* 2005). The negative results of Hoeck & Keller (2012) invoked the idea of investigation of functional genetic diversity, that could be maintained in, otherwise, genetically depleted populations.

Special concern was paid to the immunocompetence and the genetic diversity of the critically endangered populations of Floreana mockingbird. The strongest effect of inbreeding and genetic drift were found in a tiny population on the Champion islet, where only between 20-53 individuals survive (Grant *et al.* 2000). The second population on the

Gardner by Floreana is slightly bigger with 200-500 birds and it showed also higher genetic diversity (Hoeck & Beaumont 2010). Coalescent model, based on microsatellite variation, showed that divergence between these two populations is not older than 800 years, and Hoeck et al. (2010) suggested that mixing of these two populations could lead to heterosis and increase of genetic variation. On the other hand, Deem et al. (2011) found that the populations are under different parasite regimes. For example the feather lice *Bruelia galapagoensis* was found only in the mockingbirds from the Gardener by Floreana and also overall prevalence and intensity of endoparasites was higher in the Gardner by Floreana than in the Champion. Moreover, the Champion mockingbird population was less clinically healthy based on immunological measures (Deem *et al.* 2011). Therefore, interesting view on immunological health and differences between these populations can be gained by screening the adaptive variation in MHC genes.

2.3. MHC genotyping method overview

Because primers for genotyping of MHCII β2 usually bind to more paralogous loci, individual PCRs result in a mixture of fragments comprising many alleles. Phasing of the product of one PCR (hereafter amplicon) can be accomplished either by laborious bacterial cloning methods or by second generation sequencing (SGS) (Babik *et al.* 2009). The advantage of the SGS lies in efficient separation of each fragment within amplicon by clonal amplification and high throughput sequencing. On the contrary to bacterial cloning, the SGS can produce many more separated sequences (reads) for better price (Babik *et al.* 2009). But the drawback of the SGS methods lies in generating erroneous sequences which can be difficult to separate from real MHC alleles (Lighten *et al.* 2014a). Unsuccessful separation of the artefacts from real alleles can lead to under- or over-estimation of the MHC polymorphism.

The errors can originate in two levels of the experiment. 1) Errors generated during PCR as a result of heteroduplex formation or polymerase slippage and incorporation of incorrect bases. If these errors appear in the first cycles of a PCR their frequency within amplicon can be even higher than of some rare MHC alleles (Sommer *et al.* 2013). 2) Errors specific to the SGS platform. Most of these errors (e.g. homopolymer misincorporations in 454 and Ion Torrent sequencing) can be filtered out by manufacturers' pipelines and thresholds set for high quality (Lighten *et al.* 2014a). Despite all such measures the MHC sequencing by SGC is an uneasy process due to its high GC content and many homopolymers.

Several methods were designed to solve this problem, although none of the methods was generally approved and the situation is currently unsettled (Lighten *et al.* 2014a). Most of the methods are based on slightly different assumptions about the frequency of errors and the number of reads necessary for a reliable genotyping. The proportions of errors are naturally different for each experiment because of the differences in SGS platforms and PCR conditions. The first set of methods estimates a threshold value of the number of reads per variant according to the putative number of the MHC loci. These methods are based on an assumption that all erroneous variants have lower number of reads than real variants. The variants above the threshold are considered as putative alleles and the others as artefacts (Galan *et al.* 2010). These methods were improved by adding a factor of similarity between the variants, based on an assumption that less frequent variant, which is similar to the more frequent, is a possible erroneous derivation. In this way the rare (below the original threshold) but divergent alleles could be revealed (Sommer *et al.* 2013). The second method utilises the degree of change of a read depth per variant to distinguish errors from putative alleles (Lighten *et al.* 2014b). This method is possibly the most general and applicable to the widest spectrum of sequencing platforms and MHC systems. However, until now it worked well only for its developers and further improvement and testing is required (Marie Jeschek, in litt, 2014).

This short methods overview only outlined the basic methodological problems connected with MHC genotyping for a reader unfamiliar with the topic, for an in-depth current review see Lighten 2014 (Lighten *et al.* 2014a).

3. Methods

3.1. Samples and DNA

DNA samples of all studied populations of the GM were obtained from our Swiss colleagues (Dr. Hoeck and Prof. Keller, University of Zurich), who collected blood samples from 2003 to 2008 (Hoeck *et al.* 2010) and also performed the DNA extractions (Hoeck *et al.* 2009). Sample collection and DNA extraction of the continental Northern mockingbird (*Mimus polyglottos*) was performed by Paquita Hoeck in 2013. The collected Northern mockingbirds originated from one population occurring near San Diego (US, California).

All DNA samples were stored in 96-well PCR plates covered by Eppendorf adhesive storage foil in a -20°C freezer. But, after some time, the DNA solution dried out in several wells. In these cases, 10 µl of Milli-Q water was added, and the concentration was

subsequently measured (details below). If very low DNA concentrations were found then I extracted DNA of the particular samples from the original blood stained filter papers. The extraction was done using DNeasy® Blood & Tissue Kit (Qiagen) according to manufacturer's instructions.

3.2. Primer design

Substantial effort in primer design is crucial for obtaining a representative sample of MHC variants (Babik 2010). Because of the rapid evolutionary changes in MHC, the primer site can be polymorphic not only among populations of one species, but also between loci within one individual. For consistent MHC genotyping, there is a need for primers which can amplify the most variants of selected fragment possibly at all loci and equally across individuals, populations and species. The steps below were done with respect to these problems, following current literature.

First I tested primers 1a and 2a (Aguilar & Edwards 2006) originally designed for the Little greenbul (*Andropadus virens*) amplifying except the MHCII β exon 2 also surrounding introns. I used these primers in PCR amplification across all species and major populations of the GM. The resulting products were cloned into *E. coli* plasmids using pGEM®-T Easy Vector Systems (Promega), according to manufacturer's instructions. At least three colonies per sample were collected and sequenced by Sanger method in a commercial laboratory (Macrogen Europe). However, the results were unsatisfactory with only one sequence from San Cristobal mockingbird being obtained. This sequence was aligned with corresponding passerine MHCII β2 fragments [from Common yellowthroat (Geothlypis trichas), Collared flycatcher (Ficedula alba), Common bluethroat (Luscinia svecica), New Zealand robin (Petroica australis) and Woodpecker finch (Cactospiza pallida)] in Geneious 7.0.6 (Biomatters Ltd.). Based on this alignment, specific primers Moc325F/R (Tab. 3) were designed at monomorphic positions, identical to previously known primers 325/326 (Edwards *et al.* 1995). The primer sites lie between the polymorphic subunits of MHCII β2, and specific primers used for songbirds of *Muscicapidae* family were often designed at this position (Zagalska-Neubauer *et al.* 2010; Anmarkrud *et al.* 2010). Later testing showed that these primers were amplifying the same length cca 164bp products consistently across all mockingbird individuals.

I used Roche GS Junior 454 sequencing platform for a preliminary analysis of MHCII β2 variation. The design of the platform specific, fusion primers was based on the

manufacturer's manual. Multiplex sequencing was conducted using identification of individuals by tag combination on both sides of the DNA fragment (Galan *et al.* 2010). Fusion primers Moc325*454 were composed of adaptor A, multiplex identifier (tag) and Moc325F/R primers (Tab. 3). For the preliminary analysis I used only 7 forward and 7 reverse primers with different tags. By combining them I could multiplex 49 samples into one run.

The final analysis of MHC variation was done on an Ion Torrent sequencing platform where different fusion primers were used. The main difference from the 454 platform was that the tags could be present only in the forward primer, which meant that the number of multiplexed samples in one run had to be equal to the number of forward primers with unique tags. The Ion Torrent specific primers Moc325*ion (Tab. 3) were composed of platform specific adaptor sequence, tags (only in forward primers) and primers Moc325F/R.

For the Ion torrent analysis I used one reverse and 115 forward primers with unique tags provided by the manufacturer (Ion Xpress barcode) to multiplex 115 samples per one run. The barcode adapter (GAT) was omitted, but care was taken to avoid homopolymers following the barcode. The primers were HPLC purified after synthesis (Sigma Aldrich) to obtain 100% base-pair accuracy in the tags.

Primer name	Sequence 5'->3'
Moc325F	GAGTGTCACTTCATTAACGGCAC
Moc325R	GTAGTTGTGCCGGCAGTAATTGT
Moc325F454	CGTATCGCCTCCCTCGCGCCATCAGNNNNNNNNNNGAGTGTCACTTCATTAACGGCAC
Moc325R454	CTATGCGCCTTGCCAGCCCGCTCAGNNNNNNNNNGTAGTTGTGCCGGCAGTAATTGT
Moc325Fion	CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNNNNNNNGAGTGTCACTTCATTAACGGCAC
Moc325Rion	CCTCTCTATGGGCAGTCGGTGATGTAGTTGTGCCGGCAGTAATTGT

Table 3: Primers used in amplification of MHCII β2 in the Galapagos mockingbirds. Blue letters denote MHC specific primers, red colour denotes platform specific primers and black Ns denote variable tag sequence.

3.3. Amplicon preparation and sequencing

3.3.1. GS Junior 454 preliminary analysis

MHCII β2 exon was amplified in 33 individuals with unique combinations of the fusion primers designed for the GS Junior 454 platform. The PCR was carried out in a 50 µl volume which contained 4 mM MgCl2 (Fermentas), 0.2 mM each dNTPs (Fermentas), 2 µM of each primer, 1X Fermentas PCR buffer, 2 U Fermentas Taq DNA polymerase and approx. 20 ng of DNA. A proofreading enzyme was not used deliberately, because it can cause formation of artificial chimeras (Zylstra & Rothenfluh, 1998). Reaction conditions on the Eppendorf Mastercycler Pro machine were: 2 min of initial denaturation at 94°C then 35 cycles of: 30 sec denaturation at 94 °C, 30 sec annealing at 60 °C, and 30 sec elongation at 72 °C. Final elongation step was run for 9 min. at 72 °C. The PCR products were visualised by electrophoresis on a 1 % agarose gel with DNA stained by GelRed (Biotium). Because there were two bands of different sizes from which the shorter one was possibly a primer heterodimer, I had to extract the fragment of correct size using the JET quick gel extraction spin kit / 250 (Genomed). Subsequently, concentrations of products were measured on Bioanalyzer (Agilent Technologies) and samples were equimolarly pooled. Then the concentration of the final library was re-measured on the Bioanalyzer, and diluted to the 1x 107 molecules/µl required for emulsion PCR of the GS Junior Titanium platform.

All 454 sequencing procedures were done in the frame of a university course, Practice of 454 sequencing, at the Faculty of Science of University of South Bohemia. The emulsion PCR was carried out according to manufacturer's protocol (emPCR Amplification Method Manual – Lib-A). The emulsion PCR is an essential step for the separation of individual DNA molecules and their clonal amplification across primer-coated beads. The beads were later purified, loaded onto one GS Junior Titanium PicoTiterPlate (Roche), and all preparatory sequencing steps followed the protocol (Sequencing Method Manual GS Junior Titanium Series March 2012). The sequencing results were cursorily checked in GS Browser and subsequently outputted to an .SFF file.

3.3.2. Ion Torrent final analysis

Prior to the preparation of amplicons for Ion Torrent sequencing I performed measuring of concentration of DNA in the mockingbird samples using the Qubit 2.0

Florometer Broad Range ds DNA kit (Life Technologies). Based on the concentrations and the associated data from previous studies, I selected 5-15 samples for each of the 15 mockingbird's populations resulting in 204 unique samples in total. Additionally 21 samples were amplified twice with a unique primer combination as independent replicates. The sample set was divided in two batches, because each amplicon had to have its unique forward primer (only 115 primers) for pooling to one Ion torrent run. And then two rounds of PCR amplification and two Ion Torrent runs were conducted.

The PCR was prepared on two 96-well PCR plates in a total volume of 25 µl containing: 2 mM MgCl2 (AmpliTaq Gold), 0.2 mM each dNTPs (Fermentas), 2 µM of each primer, 1X AmpliTaq Gold 360 Buffer, 1.5 U AmpliTaq Gold 360 DNA polymerase and approx. 16 ng of DNA. The PCR reaction was performed on the same machine as the 454 PCR, albeit with different thermal conditions. Initialization of HotStart polymerase together with the initial denaturation was run for 8 min. at 95°C and followed by 27 cycles of: 30 sec at 95 °C, 30 sec at 56.8 °C, 30 sec at 72 °C, and final elongation 7 min. at 72 °C.

Guidelines of Lenz & Becker (2008) were followed to minimize the occurrence of PCR artefacts, like artificial chimeras and basepair mismatches, when amplifying highly polymorphic loci. The measures included the usage of non proofreading enzyme, low concentration of primers, reduced number of PCR cycles and increased elongation time.

Excision from the gel was omitted because the modified PCR conditions did not result in any longer products but the targeted fragment. Then the concentration of each amplicon (product of one PCR marked with a unique tag) was measured by the Qubit 2.0 Florometer High Sensitivity ds DNA kit (Life Technologies, Grand Island, NY, USA) and the amplicons were pooled equimolarly. The resulting library was purified by the Agencourt AMPure system (Agencourt Bioscience Corporation, Beverly, MA), analysed on the Bioanalyzer (Agilent Technologies), and diluted to the required concentration of 26 pM.

The Ion Torrent sequencing was performed at The Institute for Nanomaterials, Advanced Technology and Innovation at the Technical University of Liberec. The emPCR was done using the Ion One TouchTM 2 System (Life Technologies) with a 400 bp chemistry kit according to the manufacturers manual. The enriched Ion Spheres were loaded onto two 318TM Ion Torrent Chips, and one-directional sequencing was done on the Ion Torrent Personal Genome MachineTM System (Life Technologies). The demultiplexing of sequences based on barcode was done in Ion Reporter TM software (4.2). And sequence data

were subsequently exported to “fastq” format.

3.4. Allele identification

Allele identification in the 454 dataset was based on a threshold method similar to (Galan *et al.* 2010). First, the low quality reads were discarded (lower 20 PHRED) and the sequence reads were subsequently uploaded to the online sorting program SESAME (Meglécz *et al.* 2011). This software allowed demultiplexing of the sequence reads according to the tags and trimming the specific primers. Only the reads with length of 163-165 bp passed to the next step where they were aligned and ordered according to their frequency. Singletons (reads encountered only once in the amplicon) were discarded as potential artefacts. The first threshold of minimal number of reads per amplicon (50) was passed by all amplicons. The second threshold dealt with the frequency of reads within an amplicon. The reads which occurred in an amplicon at least 3 times were classified as putative alleles. But the variants comprising only three reads and present in only one amplicon were discarded. Low frequency variants which were obvious recombinants of two more frequent variants were classified as artificial chimeras. This step was based on the methodology developed by Sommer et al. (2013). The alignments were also checked for obvious homo-polymer errors. The chimera and homo-polymer checking was done manually in Geneious 7.0.6 (Biomatters Ltd.). After translation to the protein I discovered that all reads shorter than 164 bp actually contained a stop codon so these reads were separated as putative pseudogenes.

For allele identification in the Ion torrent dataset, I used the backbone of a clustering method developed by Sommer et al. (2013), which was modified for my specific dataset. Demultiplexing of the sequence reads based on tags was done in Ion Reporter TM software (4.0, Life Technologies). Data were subsequently exported to the fastq format. Low quality reads (lower than 20 (PHRED) in more than 2% of the basepairs) were discarded from both Ion Torrent runs in order to avoid sequencing errors. Only the reads containing the exact MHC specific primer sequence were retained and primer site was excised by a python script. Identical sequence reads within the amplicon (=clusters) were collapsed while their frequency was recorded into their name by fastx_collpaser script from FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Investigation of read length distributions showed that classes higher and lower than the expected size of 164 bp were either homo-polymer errors or putative pseudogenes. Based on this finding, only the reads of 164 bp passed to the next step. Singletons as possible artefacts were again removed.

After this preprocessing stage I used scripts for putative allele identification adjusted from Sommer et al. (2013), this bioinformatic part was done during a research stay at the BeGenDiv Centre in Berlin, where the original scripts were written. In the first step, the clusters were ordered according to their frequency within an amplicon and categorised to 3 groups. Clusters which could be explained as a chimera of two more frequent clusters within the amplicon were classified as putative chimeras. The second group, called "1-2bp diff", were clusters which differed only by 1 or 2 bp from higher frequency clusters, and the third group, ">2bp diff", were clusters with more than 2 differences from the higher frequency clusters. All these steps were based on the assumption that artefacts (either chimeras or mismatches) had lower frequencies than the true alleles from which they possibly originated.

In the second step, the clusters were compared between the amplicons. The clusters labelled "chimera" or "1-2bp diff" were classified as artefacts when they were absent in their replicate. The ">2bp diff" group was classified as an artefact only when it was not found in the other amplicons. Because this pipeline is very strict for "1-2bp diff" which were absent in the second replicate I had to adjust the script for my dataset since there were only 20 individuals duplicated (difference from original methods). So the clusters "1-2bp diff" were classified back as non-artefact if they had a coverage of at least 3 reads and occurred at least in 12 other replicates. Also, chimeras were treated less strictly, because many of the chimeric variants occurred in other amplicons as high frequency clusters (cluster1 or cluster2). These variants also passed to the last step of putative allele identification. In the third step, all non-artefact "1-2bp diff" and ">2bp diff" were classified as unclassified variants if they appeared in some amplicon with a lower frequency than the artefacts. The other variants from these two groups were classified as putative alleles. All "clusters one" were also classified as putative alleles. Chimeras were checked manually, and were classified as putative alleles when 1) they appeared as high frequency cluster, or 2) had a high coverage (more than 10) and were found in other amplicons, or 3) no obvious chimeric origin was found after a visual check with possible parent clusters.

3.5. Statistical analysis of MHC alleles

The aggregated numbers of putative alleles per individual, total number of alleles per population and corresponding classification data (species, population, number of ectoparasites, the immunological indices per individual and total coverage per amplicon) were exported from data tables using a SQL query. First, I tested correlation between the number of alleles discovered in each amplicon and the total coverage of the amplicon using

linear regression model in the program R (R Core Team 2014). In this test I wanted to assess the reliability of sequencing (i.e. sufficiency of the amplicon coverage). Also the basic plots of number of alleles per population were produced in R with use of the package ggplot2 (Wickham 2009). Sharing of alleles was analysed by dedicated R script that generated matrix of pairwise comparisons of common alleles between populations and species. This matrix was subsequently visualised by a table-viewer tool in the program circos (Krzywinski *et al.* 2009).

To test the effect of demographic evolutionary processes on MHCII β2 at the population level I used generalized linear model with the total number of MHC alleles per population as a "response" and island size, island age and microsatellite allelic richness as "explanatory" variables. The microsatellite allelic richness, island size and age data were adopted from a previous literature on mockers (Hoeck *et al.* 2010). The Northern mockingbird population was removed from these analyses because of the lack of corresponding microstellite data. In the GLM model I used a quasipoisson distribution.

Correlation of the MHC diversity (represented by the responses above), with ectoparasite and imunological indices were tested in a similar manner as above, but 3 more populations were excluded, because of a lack of the imunological and parasitological data (Marchena, Pinta, Genovesa). The ectoparasite intensity was represented by either total mean number of all ectoparasites (Myrsidea, Bruelia and mites), or the mean total number of Myrsidea parasites per population. Immunological indices (agglutination, lysis and number of lymphocytes) were adopted from the immunological study on the GM (Hoeck & Keller 2012).

At the level of individuals I tested whether the number of MHC alleles present in an individual is correlated to immunological indices and parasite intensity. In the model I implemented island (affiliation to population) as covariate with random effect in order to filter out the between island differences. These models were tested also in program R (R Core Team 2014) however in this case I used simple linear models and function "aov".

3.6. Molecular variation

Mean evolutionary divergence of the MHCII β2 alleles within populations was estimated in MEGA6 (Tamura *et al.* 2013) using Jukes-Cantor distance model. To visualize the relationships between mockingbird MHC alleles and alleles from other species I constructed a haplotype network in the program SplitsTree4 (Huson & Bryant 2006). I used

Neighbour-Net method (Bryant & Moulton 2004) to show possible reticular relationships originating from gene duplication and recombination.

3.7. Selection

Selection in MHC alleles was first tested in the program MEGA6 (Tamura *et al.* 2013) by a Z test for positive selection (H0:dN/dS<=0). This method compares the number of synonymous and nonsynonymous substitutions using Nei Gjobori method (Nei & Gojobori 1986) and Jukes-Cantor correction. The test was conducted for whole sequences and also for antigen binding sites (ABS) and non-ABS sites separately. The positions of these sites, important for recognition of pathogenic peptides, were taken from a human MHC molecule model (Brown *et al.* 1993).

As an alternative test I also used the SLAC method (Kosakovsky Pond & Frost 2005) implemented in the HyPhy package (Pond *et al.* 2005) and accessed on-line through a datamonkey web page (Delport *et al.* 2010). The method uses maximum likelihood to find appropriate model and estimate substitution rates for each branch separately, subsequently it estimates the dN/ dS ratio for each codon based on the substitution rates.

4. Results

4.1. MHC sequencing and allele filtering

454 sequencing of 33 tagged amplicons (corresponding to 33 individuals) resulted in 55717 sequence reads. Only 24877 reads were retained after quality filtering, demultiplexing and length selection. Collapsing of these reads in SESAME revealed 9652 variants from which more than 80% (8062) were discarded as singletons. From the rest I classified only 145 variants as putative alleles with the length 164 bp. A large portion of the variants (566) was 163 bp long and because of the stop codon occurrence, I classified these variants as pseudo-genes. Also, artificial chimeras were present, but their number was not substantial (30). The mean amplicon coverage was 731.7 (+-296,9) and ranged from 312 to 1450 reads. In this preliminary genotyping I found that amplicon coverage was not sufficient for revealing the complete MHCII β2 allelic diversity, because the number of alleles per amplicon was strongly correlated to the coverage (R^2=0.550, p=1.326e-05, Appendix 3).

Two runs of the Ion Torrent PGM sequencing of 225 uniquely tagged amplicons produced 8.2 M sequence reads. 1882391 reads were extracted after discarding low quality reads, reads with incomplete primers or tags and reads with a length different from 164

(Fig. 4). One amplicon retained no reads after these steps, so it was discarded from further analysis. The amplicon coverage ranged from 35 to 76910 reads with an average of 7775 reads.

I obtained 13076 clusters after collapsing the reads and singletons removal. The intra amplicon comparisons classified 11091 clusters as "1-2bp diff", 308 clusters as chimeras, 224 clusters as "cluster 1" and 1453 clusters as ">2bp diff". In the second step of classification, between amplicons, 1040 clusters were labelled as putative alleles, 1378 as unclassified variants and 10658 as artefacts (10307 as 1-2bp artefacts, 308 as chimeras, and 43 others).

The final classification re-assessed the state of some artefactual clusters, so at the end I obtained 2106 clusters classified as putative alleles and 10970 unclassified or artefactual clusters. Based on this allele classification, I revealed MHCII β2 genotypes for 203 individuals and 21 replicates.

However, the credibility of the genotyping is open to question. First, the mean repeatability for the Ion torrent replicates (percentage of common alleles in both replicates from all alleles) was 38.2 %. On average 12.8 alleles were unique in both replicates. Also there were inconsistencies between individual genotypes obtained from the 454 and Ion torrent sequencing (N=11 replications). On average, 52.3% of Ion torrent alleles were found also in the 454 dataset. The repeatability (same measure as above) between the 454 and Ion datasets was very low at 18.3%, but the two datasets are not directly comparable, because of the different methods of allele classification used. Second, a statistically significant positive correlation was found between the number of putative alleles and amplicon coverage (R^2=0.1471, p=9.73e-09, Appendix 3). Lastly, the coverage of particular alleles (amplification efficiency) varied substantially among individuals. Range of coverage for all alleles was 46.7 – 3105.3 with standard deviation (995.4) almost ¾ times higher than mean allele coverage (646.8).

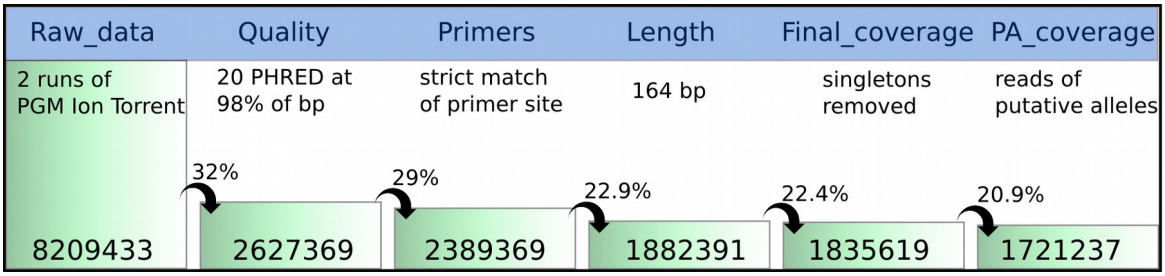


Figure 4: Diagram showing steps of filtering of the reads. Green columns are proportional to the number of reads retained after each filtering step noted in the headline.

4.2. MHC diversity

Based on Ion torrent data I discovered 161 unique alleles from the 5 species of mockingbirds. The mean number of alleles per individual (9.62) varied among species (ANOVA, Df=4, F=2.568, p= 0.039) (Tab. 4) from 5.92 in the San Cristobal mockingbird to 11.08 in the Espanola mockingbird. Number of alleles per individual ranged from 2 to 28 (Appendix 2) meaning that my primers amplified as many as 14 loci. On the other hand the 28 alleles were observed only in one outlying individual, so it is possible that the number of loci is slightly lower. The mean number of alleles per individual varied substantially also among populations (ANOVA, Df=14, F=10.32, p= < 2e-16). The lowest number of alleles per individual were found in Santa Fe (4.0) and the highest in Fernandina (17,2) (Fig. 5). The total number of alleles per population also showed substantial variation. On average 37.5 distinct alleles were present in each population with the lowest number in the Santa Fe (19 alleles) and the highest in Isabela and Santa Cruz (Fig. 5). The population of Northern mockingbird surprisingly showed slightly subnormal number of distinct alleles (35).

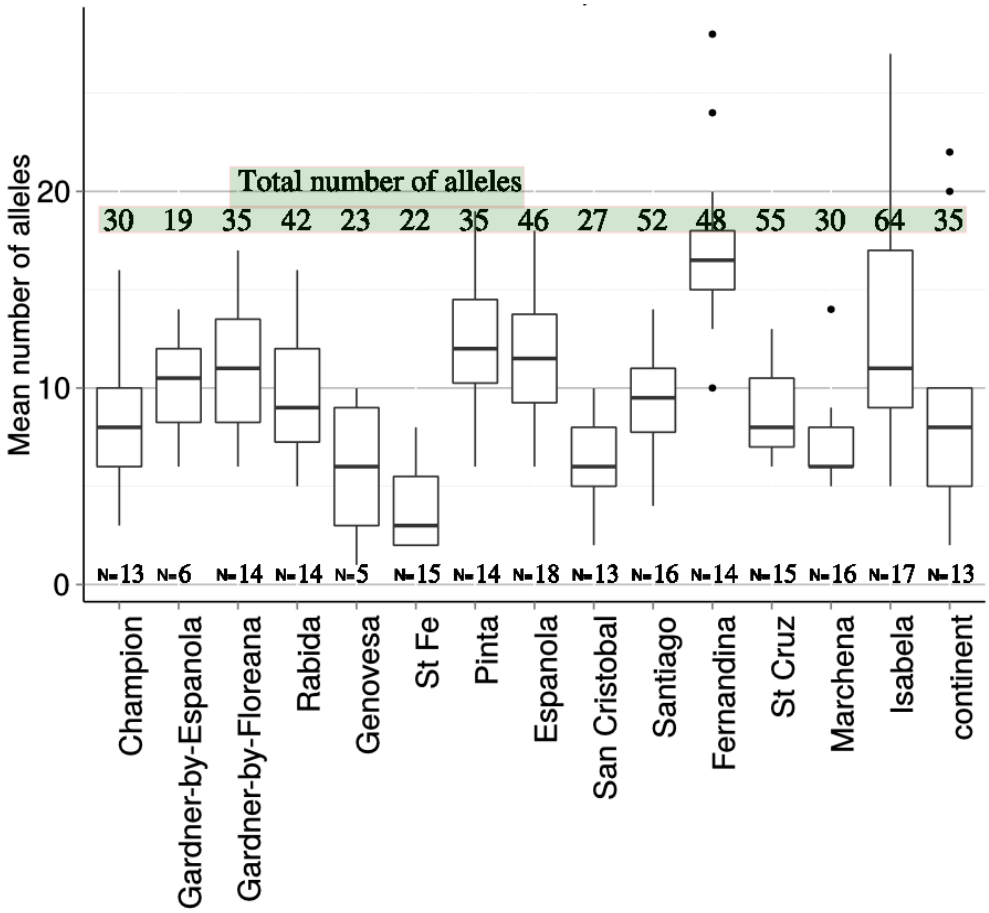


Figure 5: Boxplots visualising the mean number of MHCII β2 alleles per mockingbird population. Numbers below the boxplots denote numbers of genotyped individuals and row above the boxplots shows total number of distinct alleles per population. The population of Northern mockingbird is called “continent”.

Sharing of the same alleles across populations and species was very high. 71,4 % of all the MHCII β2 alleles were shared between population and only 46 alleles were private (Appendix 1). Expectedly the biggest portion of private alleles was found in the Northern mockingbird (25). Within Galapagos populations the portion of private alleles was much lower with the maximum number of private alleles found in the San Cristobal (5). The MHCII β2 alleles were equally shared among the populations, moreover, the species boundaries were not reflected by the sharing at all (Appendix 1, 4).

One of the aim of this thesis was to test the adaptive divergence between the two populations of the critically endangered Floreana mockingbird. But even despite the different parasite regimes the two populations shared 18 MHCII β2 alleles. Moreover no private alleles were found on Champion and only one private allele was on Gardner by Floreana (Appendix 1,5). The total number of alleles can slightly reflect the population differences, where 30 alleles were present in smaller Champion populations and 35 in Gardner by Floreana. But the numbers of shared alleles suggest that there is no adaptive divergence in MHCII β2 neither in Floreana mockingbird nor in any other population of the GM.

Mean overall divergence of all MHCII β2 alleles was 0.123. This index showed interesting pattern on a population level, where the allelic divergence was almost two times higher in the Northern mockingbird (0.173) compared to an average GM population (0.093) (Fig.6).

In phylogenetic network mockingbirds' MHCII β2 alleles fromed a few apparent clusters interspersed with more divergent alleles (Fig. 10). Interestingly the more divergent alleles clustered together with other MHCII β2 alleles from other passerines. This pattern suggest some degree of MHC transspecies polymorphism in songbirds.

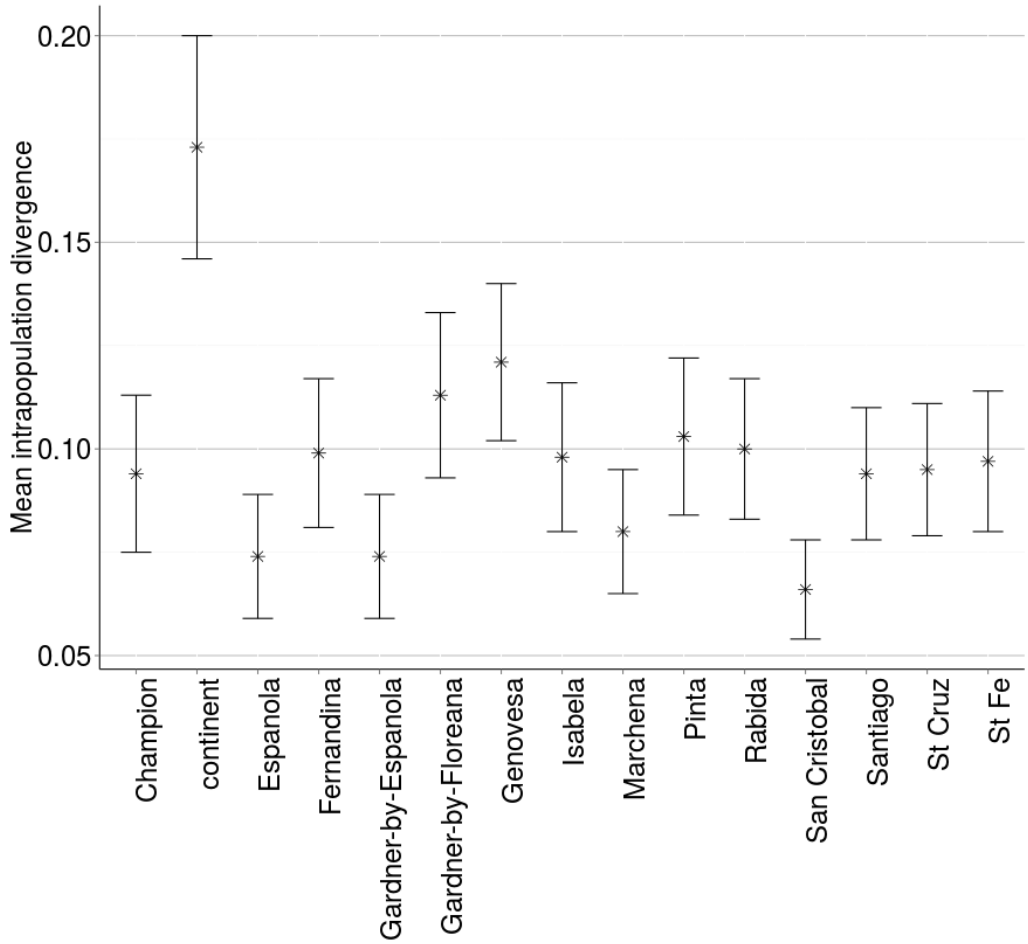


Figure 6: Estimates of average evolutionary divergence over sequence pairs of MHCII β2 alleles within populations of the GM and Northern mockingbird (continent).

4.3.Natural selection at MHC

The overall proportion of nonsynonymous (0.126) substitutions in MHCII β2 exceeded that of the synonymous substitutions (0.074). This pattern suggests historical effect of positive selection acting at the MHCII β exon 2 (Ztest=2.023,p=0.023). But the selection acted rather site specifically, because even stronger signal of positive selection was observed in the antigen binding sites (ABS) (Ztest=2.488,p=0.007). The same hypothesis of positive selection was not statistically significant in non-ABS (Ztest=0.654,p=0.257). Codon based approach by the SLAC method showed similar results, where 8 codons were found under positive selection, but only 4 of them overlapped with the ABS (Fig. 7). Also the number of nonsynonymous changes in the ABS was significantly higher than in the non-ABS. The positive selection observed at the MHCII β2 means that selection played an important role in

shaping the variability in this exon during the evolutionary history of mockingbirds. By the same (SLAC) method I also discovered 2 sites under purifying selection (where synonymous changes had higher frequency than under neutrality). These sites are possibly important for some invariable bounding within the MHCII molecule and thus its amino-acid state is conserved by the selection.

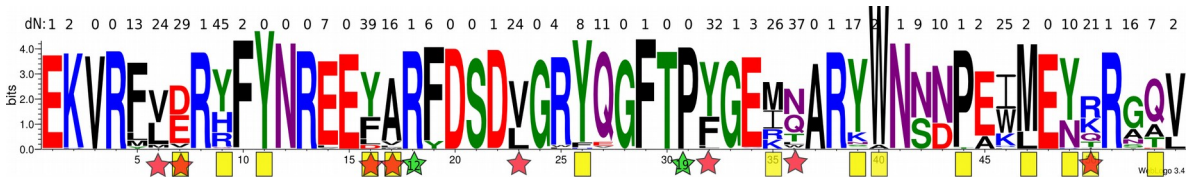


Figure 7: Amino-acid consensus sequence of MHCII β2 from the Galapagos Mockingbirds. The size of amino-acid letter is in proportion with its frequency. Yellow boxes below the sequence denote antigen binding sites characterized by Brown et. al. (1993). Red stars denote sites under positive selection and green stars denote purifying selection both identified by SLAC. The row above the sequence shows number of non-synonymous changes for each codon and numbers in green stars denote synonymous changes per codon.

4.4. MHC variation and environmental factors in the GM

microsatellite allelic richness per population showed close positive correlation with the number of MHC alleles per population (=MHC allelic diversity) (GLM,F=11.53,Df=12,p=0.00532,R^2=0.493, Fig. 9). Island size and the MHC allelic diversity showed also positive correlation, although the effect of the island size was slightly smaller (GLM,F=7.015,Df=12, p=0.0212,R^2=0.356, Fig. 10). The correlation between island age and MHC allelic diversity was not statistically significant (GLM,F=2.85,Df=12,p=0.117,R^2=0.193). These findings suggest that neutral forces have a substantial effect on MHC variation. At the population level I have not found any statistically significant correlation between the allelic diversity and immunological or parasitological indices.

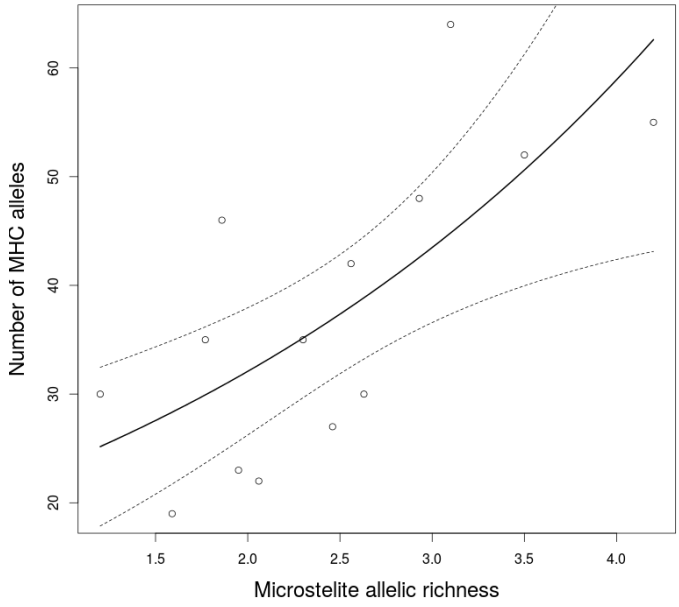


Figure 8: Correlation of MHC and neutral variability at population level in the Galapagos mockingbirds. Number of distinct MHC alleles is used as index of MHC variability and mean microsatellite allelic richness as index of neutral variability. GLM,F=11.53,df=12,p=0.00532, R^2=0.493.

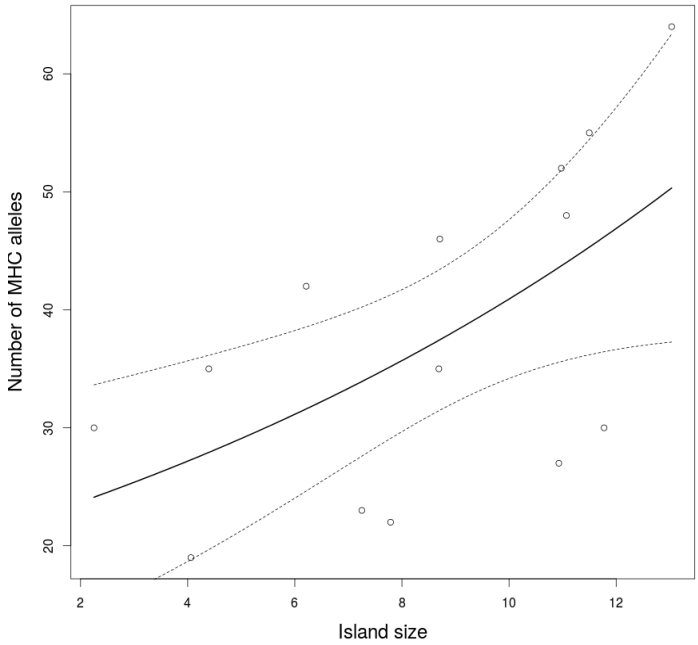


Figure 9: Correlation of MHC variability and island size in populations of the Galapagos mockingbirds. GLM,F=7.015, Df=12, p=0.0212, R^2=0.356.

At the individual level I tested for possible correlations between the number of alleles per individual as a response and several ecological indices as explanatory variables (Tab. 5). In general, I found no robust correlation, but a few ecological factors seemed to be in some weak relation to the number of MHC alleles. For example, I found negative correlation between the number of alleles and number of lymphocytes. However, the correlation became statistically non-significant when I included population (island) as a covariate. Positive correlation was also found between the number of MHC alleles and the weight of the birds, but similarly the model was not statistically significant after inclusion of island as a covariate. The only statistically significant correlation in the model with island as a covariate was found between the index of agglutination and number of alleles. However, this positive correlation was very weak (R^2=0.054) which leaves the effect of the number of MHC alleles on this immunological factor questionable.

Explanatory	Error term	F value	Residual DF	p
Lymphocytes	Island	0.000	132	0.993
Lymphocytes	NO	5.080	145	0.006
All parasites	Island	0.115	120	0.735
Weigth	NO	7.797	145	0.006
Weigth	Island	0.000	135	0.984
Heterophils	Island	0.199	132	0.656
Heterophils	NO	8.493	142	0.004
Lysis	Island	0.083	72	0.774
Agglutination	NO	5.745	82	0.019
Agglutination	Island	5.978	72	0.017
Myrsidea	Island	0.323	121	0.571
Bruelia	Island	1.998	121	0.160
Mites	Island	0.108	121	0.743
Heterozygosity	Island	0.315	175	0.575

Table 5: Statistical results of linear models where I used the number of MHC alleles per individual as a response and several immunological, parasitological and others indices as explanatory variables. Island was included as a random error factor.

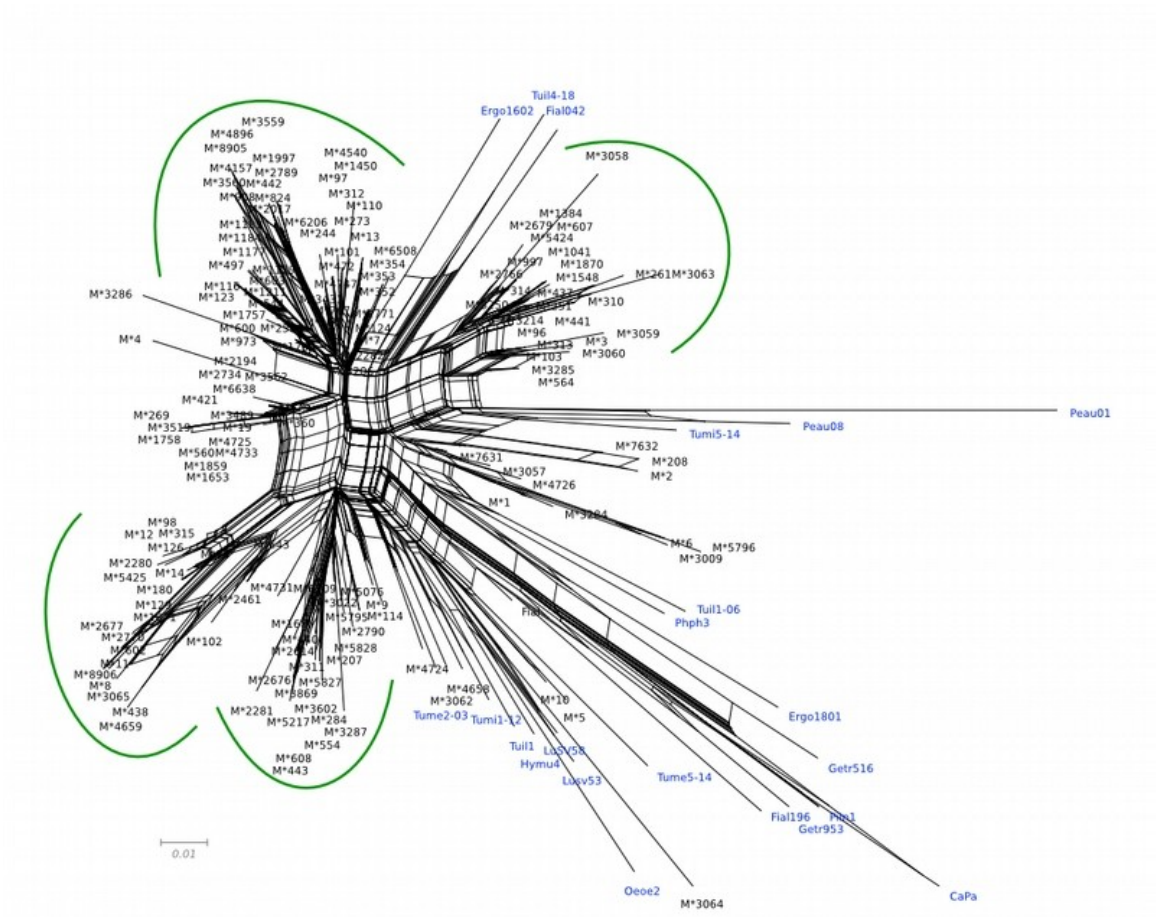


Figure. 10: Phylogenetic network of MHCII β2 alleles of Galapagos mockingbirds (M* number) and MHCII β2 alleles from other songbirds (in blue). Closely related clusters of MHCII β2 alleles are highlighted by green brackets.

5. Discussion

Genotyping of the MHCII β2 in 4 species of GM and their relative, Northern mockingbird, revealed high level of polymorphism comparable to that found in other songbirds (Tab. 1). In total I revealed 161 putative alleles, with 9.56 alleles found on average per individual. Extreme variation was observed in the number of alleles per individual ranging from 2 to 28. This finding signifies that in mockingbirds can be up to 14 MHCII β2 loci per individual when assuming that the individual is heterozygous at all loci. Such number of loci in the MHCII β2 is among the highest observed in songbirds. 11 loci were detected in Bluethroat, phylogenetically the closest genotyped species, and 20 loci, the highest number from all songbirds, was revealed in a Common yellowthroat (Tab. 1). On the other side, our genotyping showed several inconsistencies (discussed later), so I interpret the

estimate of the number of loci and all other results conservatively bearing in mind potential pitfalls.

5.1. Variation and genetic drift

One of the most compelling results of my study was the observation of a very strong correlation between the island size, the microsatellite allelic richness and the total number of MHC alleles per population. In smaller populations I observed lower total number of MHC alleles, which implies that the MHC variability was negatively affected by genetic drift or inbreeding in the same manner as the neutral variability. This pattern aligns with the growing evidence that the MHC behave as evolutionary neutral loci when the population size is small (Radwan *et al.* 2010; Sutton *et al.* 2011). The chance that this strong correlation in my data is caused by some genotyping errors is small. Therefore this observation support the reliability of MHC genotyping in my study system and also the view that the force of balancing selection can not outweigh genetic drift in such small populations.

A similar study investigated the effect of drift on MHCII β2 in populations of Black and South island robins (*Petroicidae*) in New Zealand(Miller & Lambert 2004b). The populations differed by size (and historical bottlenecks) and the study showed that total numbers of the MHCII β2 alleles were highly correlated to the size of the populations and minisatelite variation. Comparison of microsatellite and MHCII β2 in continental and insular populations of the Egyptian Vulture showed the same correlation between the neutral and adaptive variability. However, in this case it was speculated that divergent alleles were maintained in the insular populations by natural selection(Agudo *et al.* 2011). Another congruent result comes from a Galapagos hawk (*Buteo galapagoensis*) study, where both MHCII β2 and microsatellite diversity were much lower than in a mainland relative, the Swainson's hawk (*Buteo swainsoni*). On contrary, in Dark-eyed juncos (DEJ) a small recently established population had the same diversity and numbers of MHCII β2 alleles as the large ancestral population, despite it showed lower microsatellite variability (Whittaker *et al.* 2012). Conversly to my study system, the founder event in DEJ happened only cca 30 years ago and the isolation is probably not as complete as in the GM. For these reasons the effect of genetic drift in DEJ can be still marginal.

In songbirds there are almost no studies supporting the hypothesis that the effect of selection can maintain variability in small populations. This pattern is often explained by high effect of genetic drift in small populations reinforced by drift across loci (Eimes *et al.*

2011), a process specific to multiplicated gene families. Moreover in my case the impoverished insular parasite and pathogen fauna imposes lower selection pressure on the hosts (Hochberg & Møller 2001; Blumstein & Daniel 2005). The neutral theory posits that a gene is effectively neutral when selection coefficient is lower than 1/2Ne (effective population size) (Kimura 1983). Consequently, it is not surprising that the variation in MHCII β2 in the GM was shaped rather by stochastic neutral processes than selection. Also the observation of selection based on dN/dS ratio is not contradictory to the low effect of current selection as the dN/dS measure rather reflects selection events from past evolutionary history of the MHCII β2 (Hedrick 2012).

A shift in defence mechanisms from adaptive immune system (and the MHC) to the innate immune system (Matson 2006)was observed in studies of insular populations (Matson 2006) . The numbers of natural antibodies as a measure of immunocompetence based on innate immune response were higher in insular populations according Matson's study, where also the GM were treated. The high numbers of natural antibodies and high complement enzyme titres were found consistently across the populations of the GM (Deem *et al.* 2011; Hoeck & Keller 2012). Consequently it is possible that MHC genes are not so important for a short term survival as their function can be replaced by another part of the diverse immune system. But the adaptive variability of MHC can become essential when a highly virulent pathogen attacks the population (Hawley & Fleischer 2012).

We also found that ectoparasite load was not affected by the number of MHCII β2 alleles per individual, and also that the mean ectoparasite load per population was not affected by the number of MHC alleles per population. This is probably because of very weak selective pressure from the ectoparasites and the weak effect of selection pressures in insular populations overall. Despite it was found that *Myrsidea* lice can interact with the immune system due to their blood feeding habits (Møller & Rózsa 2005), their presence does not seem to affect, or be in relation with, the genetic composition of MHCII β2 according to the results observed in our study. The endoparasites can have much higher effect on immune response. To find out more about their role in the GM system, I plan to screen the endoparasites for possible correlation with the MHC in our future work.

Preliminary data on endoparasite prevalences are available from the two populations of Floreana mockingbird (Deem *et al.* 2011). The two populations provide a proxy of a natural experiment testing the effect of two different parasite regimes on the variability of MHC. Higher prevalence of endoparasites like coccidians was seen in the Gardner by Floreana

population than in the other population on Champion. Also the total number of MHCII β2 alleles and the number of MHCII β2 alleles per individual was higher in the more parasitised but immunologically healthier population (Deem *et al.* 2011). As in some cases above, the results require careful interpretation, because the differences between the two islands were relatively subtle (5 alleles) and thus potentially affected by the quality of genotyping. Moreover, individual alleles were highly shared between the two islands, although not to a greater degree than between some populations of the Galapagos mockingbird. The conclusion is that there is no adaptive divergence between these two populations. But, to fully assess the effect of different parasite regimes on differentiation in MHC, better quality of data will be needed, ideally combined with more in depth study of other ecological factors like breeding success.

5.2. Allelic divergence

Notable difference was observed in the levels of intra-population allelic divergence between the Northern and the Galapagos mockingbirds. The Northern mockingbird population showed almost twice higher allelic divergence than an average Galapagos population. The divergence in the MHCII β2 can actually represent an important adaptive advantage because when transformed to proteins these more diverged alleles can confer resistance to a wider range of pathogens (Spurgin & Richardson 2010). Based on this assumption I can say that natural selection at the MHCII β2 shaped the divergence in the mainland population. I presume that similar divergence was present in the mainland ancestor of the GM and that this divergence was lost due to a combination of founder effect, weakened selection and genetic drift. Interestingly, the numbers of alleles in the insular populations were either the same or even higher than in the mainland. This pattern implies that substantial MHCII β2 variation can be generated even without obvious pathogen-mediated selection and maybe the ABC hypothesis and the selection against deleterious mutations plays a role in shaping the adaptive variability in GM (Oosterhout 2009b). To validate this hypothesis flanking regions of MHC would have to be screened for linkage disequilibrium and deleterious mutations.

In several similar MHCII β2 studies the insular populations of birds always showed reduced number of alleles (Miller & Lambert 2004b; Alcaide *et al.* 2010; Bollmer *et al.* 2011; Agudo *et al.* 2011) . Unfortunately the allelic divergence was assessed only in the study of Egyptian vulture (Agudo *et al.* 2011), where the insular population had lower numbers of MHCII β2 alleles but more pronounced divergence compared to mainland

populations. The authors argued that the divergence is maintained by pathogen-mediated selection. Similar pattern of higher divergence in small population as an adaptation to pathogens is well known from mammals (Sommer 2005), but in birds it is rather scarce or overlooked. Eventually I have not found any pattern of divergence in MHCII β2 studies in birds (Miller & Lambert 2004b; Alcaide *et al.* 2010; van Rensburg *et al.* 2012; Whittaker *et al.* 2012) that would be comparable to what was found in my study system. Therefore I should be cautious in the interpretation of the data, specially because of inconsistencies of my genotyping which are discussed below. However, although biased genotyping could affect the number of alleles, especially across species, the divergence of alleles should be more robust in that respect.

5.3. Ancestral polymorphism

Unprecedented volume of allele sharing was detected among the populations of all GM. Almost ¾ of the MHCII β2 alleles were shared among the populations and species. Only the Northern mockingbird was divergent enough, possessing 25 private alleles. Neutral markers (microsatellites and mtDNA) showed complete lineage sorting across the four GM species and high divergence between the populations within species (Štefka *et al.* 2011). Hence, some force, which maintain the trans-species and trans-population polymorphism, has to be present in the system to counteract the neutral evolutionary processes in MHCII β2. Above, I developed the idea that pathogen mediated selection was weak in the Galapagos archipelago. Thus, there has to be another form of selection which conserved the MHC alleles across the species and populations. Here, again the ABC hypothesis provides explanation for the observed pattern (Oosterhout 2009b). I can assume that a new allele (originated by mutation as a neutral stochastic event) shares the same recessive mutation in the flanking region of the MHC exon with its ancestral allele. Therefore, when both alleles meet in one genome, the lethal effect can cause inability of such individual to pass its genes on. Then the new allele, which is present at a very low frequency, is purged and lost from population. In this way the ancestral alleles are conserved across speciation or population separation events even without a pathogen mediated selection(Oosterhout 2009b). Despite such scenario involves relatively complex assumptions, the feasibility of the ABC mechanism was validated in a simulation experiment(Oosterhout 2009b).

High levels of allele sharing were observed also in a closely related group of finch species (Nesospiza), 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 sharing of alleles between two divergent species was comparable to the situation in the GM. Authors explained this pattern by a retention of ancestral polymorphism, but unfortunately neutral markers were not screened to check whether the ancestral polymorphism was pronounced also there. It can be speculated that these two species are probably more divergent in microsatellites than in MHC as the separation is more than 3 Mya old (Jetz *et al.* 2012). Hence, the trans-species polymorphism could be maintained by the ABC also in this case, although more evidence is needed to confirm such hypothesis.

Sharing of MHCII β2 alleles among different species was also found in Darwin finches (Vincek *et al.* 1997). But many more studies found trans-species sharing (TSS) of not identical but very similar and probably orthologous alleles (Jarvi *et al.* 2004; Alcaide *et al.* 2007; Kikkawa *et al.* 2009; Anmarkrud *et al.* 2010). On the other hand, in some sister species the sharing of alleles was not observed at all, like in the New Zealand robins or Saddlebacks (Miller & Lambert 2004b)(Sutton *et al.* 2013), where the absence of TSS was explained by a severe bottleneck and loss of the shared variation. Some cases even showed that selection can produce local adaptations in MHCII β2 to different environments and consequently a higher differentiation of MHC alleles than microsatellites between two populations (Ekblom *et al.* 2007). Proper comparison of different systems to reveal the cause of the TSS is very tricky, because in some cases there is higher effect of drift, whereas in other various forms of selection can play the main role. In my study system the effect of pathogen mediated selection was weak, yet the TSS polymorphism was pronounced. While I revealed that the polymorphism can be maintained despite the lack of an important evolutionary force (pathogen mediated selection), the question "how it is maintained" remains opened for future testing.

The sharing of alleles also showed that populations of the GM lack any strong local adaptations differing between the islands. Such observation has important implications for the endangered Floreana mockingbird. Despite the two populations are evolving separately for more than 800 Ky and under slightly different conditions (Hoeck & Beaumont 2010; Deem *et al.* 2011), there are no pronounced differences in their MHCII β2 variation. Therefore the implementation of the reintroduction to Floreana as a measure to prevent this species from extinction can be performed with flocks of individuals from both remnant populations. No problems should occur as a result of a pronounced divergence neither in neutral loci (Hoeck & Beaumont 2010) nor in the MHC. The two endangered populations showed the same level of MHC polymorphism as other populations of the GM that live on

human inhabited islands bearing higher pathogen diversity from poultry (Deem *et al.* 2012). Consequently, from the point of the MHC variation as a measure of immunocompetence, the Floreana mockingbirds seem sufficiently fit to survive the interaction with parasite fauna on Floreana island. However, as it was also revealed in this study, the MHC variation in the mockingbirds does not provide a direct measure of their immunocompetence and screening of other branches of the immune system could provide a different view.

5.4. MHC and immunity

The correlations between MHC diversity and 3 immunological indices produced one statistically significant result. The agglutination, as a measure reflecting the number of natural antibodies, was positively correlated with the total number of alleles per individual. Interestingly, the level of natural antibodies was only recently found to be affected by the composition of MHC II DR gene in human (Pozsonyi *et al.* 2009). Despite the article is the only empirical evidence of link between the MHC II polymorphism and the natural antibodies, it was showed that people with some specific MHC II alleles possessed also significantly elevated levels of the natural antibodies. This genetic determination can work similarly also in birds. Although an in-depth immunological study would be needed to prove this. In my case I can only speculate that with rising number of alleles also the amount of natural antibodies rose. But specially these correlations can be affected by genotyping error as they are weak and slight changes in number of alleles would change the significance.

5.5. MHC and data quality

The genotyping of MHCII β2 showed several indices pointing to a relatively low reliability of the dataset. In the preliminary analysis (454 sequencing) I found that the number of detected alleles per amplicon was correlated with the total number of reads obtained per that amplicon. This suggested that some portion of alleles remained undetected when the mean amplicon coverage was 731.7. Somehow surprisingly, the same problem appeared also in the final analysis (Ion torrent sequencing), where the mean amplicon coverage was elevated to 7775 sequences. Despite the effect of amplicon coverage on the number of detected alleles was lower in the final analysis than in the 454 dataset, it was still pronounced (R^2=0.1471, p= 9.73e-09) compared to other studies (R=2 x10 ^-6, p=0.98: Zagalska-Neubauer *et al.* 2010; R=0.0038, p=0.07: Sepil *et al.* 2012). This problem could stem from the level of coverage insufficient for the detection of all alleles in the MHCII β2. Alternatively, it could be explained by a generation of undetected errors proportionally to the

number of reads.

Another weakness of my final genotyping was revealed by low repeatability between the replicates (38%). Again, in other studies this measure was much higher (98%: Sepil *et al.* 2012; 94%: Sutton *et al.* 2013). Last measure pointing to less reliable genotyping of MHC in the mockingbirds was the variability in amplification efficiency of alleles across individuals. Differences in this measure could originate either due to different number of copies of a given allele across individuals, when an allele is highly amplified in one sample and much less in another. Alternatively it could be explained by a PCR amplification bias (Lighten *et al.* 2014a) related to primer binding under specific conditions of the PCR. Primers can bind with different specificity to different variants across individuals. Given the relatively high coverage of Iontorrent sequencing, which should uncover even low frequency alleles, the PCR bias provides more probable explanation of the observed inconsistencies between the replicates and of the high variability of amplification efficiencies.

The PCR bias could also be linked to potential primer binding differences between the individuals of 5 different species and 16 populations. The primers Moc325F/R could amplify a different set of alleles and loci across species and populations, because the primer binding site was analysed only in one individual of San Cristobal mockingbird prior to the high throughput sequencing. As a result the numbers of recorded MHCII β2 alleles across species and populations could be slightly biased.

On the other hand, several facts indicate that the bias is probably only moderate. The amplified alleles contained no stop codons and a sign of positive historical selection was found in the codons known to bind the antigens (ABS). The pattern of selection was very similar to that found in other studies of songbird MHC (Zagalska-Neubauer *et al.* 2010; Sutton *et al.* 2013). Also the phylogenetic analysis (Fig. 10) revealed clustering of Mockingbird's alleles with other functional alleles from songbirds. Moreover, most of the alleles were shared among the Galapagos populations, which suggests that the amplification was not affected by the population differences in the primer binding site to a dramatic degree. It also implies that if some portion of variability remained undetected, this portion was equal among populations. Under these assumptions I see the dataset as a representative snapshot of the adaptive variability in the GM.

Nevertheless, the problems mentioned above point to the fact, that my genotyping approach was not flawless. Probably I should have implemented more thoroughly the testing

of primers and PCR conditions to avoid the PCR bias (Burri *et al.* 2014). For example, Canal et al (Canal *et al.* 2010) designed primers which amplify only the non-pseudogeneic MHCII β2 in the *Muscicapidae* family, which is relatively closely related to *Mimide*. Utilisation of these primers could also make the amplification of functional MHCII β2 alleles more efficient as the pseudo-genes consumed at least 1/10 of the amplicon coverage.

The clustering method used here for artefact detection may not be the best choice for my dataset and alternative adjustments of the clustering or some different approach could provide more accurate results. For example, a notable difference was observed in the number of clusters classified as "1-2 bp diff " between my study and the study where the clustering method was originally developed (Sommer *et al.* 2013). In my dataset I received almost 84.8% of clusters classified as "1-2 bp diff" and only 2.4% of chimeras, whereas in the dataset of Sommer at al. the percentage was 17.3% and 66% respectively (Tab. 6). Higher frequency of the small bp differences could be natural in the mockingbirds' MHCII β2 because of a higher complexity, but the method was originally designed for less complex mammal MHC, where the 1-2bp difference was considered as artefact. Therefore, I possibly underestimated the number of MHCII β2 alleles. However, as in the case of the PCR bias, if the underestimation of allele number is present, it is probably spread evenly across all samples. In a future study I should definitely try to re-analyse the data by other methods, which were recently developed and are based on more general assumptions (Stutz & Bolnick 2014; Lighten *et al.* 2014b).

Cluster class	My_total	Sommer_total	My_proportion	Sommer_proportion
1-2 bp diff	11091	713	0.848	0.173
>2 bp diff	1453	615	0.111	0.150
chimera	308	2712	0.024	0.660
total	13076	4112	1.000	1.000

Table 6: Comparison of numbers and proportions of clusters after first level of classification between my dataset and dataset from Sommer *et al.* (2013)

6. Conclusions

- Variation in functional alleles of the MHCII β2 was revealed for 4 species of the GM and Northern mockingbird.
- The variation found across all populations was comparable to other studies, mockingbirds possess up to 14 loci of the MHCII β2.
- Based on the distribution of MHCII β2 variation I came to a conclusion that its evolution is not shaped by pathogen mediated selection. The MHCII β2 variation could be rather shaped by stochastic evolutionary processes like drift, or another form of selection e.g. selection against deleterious mutations (ABC).
- In the Northern mockingbird the numbers of alleles were similar to an average Galapagos population, but the intra-population allelic divergence was much higher. This corroborates the previous point that pathogen mediated selection is weak in the GM and that other form of selection is shaping the MHC variation.
- I have found no significant correlations between the MHCII β2 variation and ectoparasites. Also the correlations between the imunological measures and MHCII β2 variation were not strong, only the agglutination representing natural antibodies was slightly affected by the number of MHC alleles per individual.
- The endangered Floreana mockingbirds shared most of the MHC alleles among their two populations, suggesting that there is no significant genetic divergence which would cause problems in the planned reintroduction project.
- Finally, I discussed potential impact of the sequencing artefacts on the results of my study.

7. References

Acevedo-Whitehouse K, Cunningham A (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in ecology & evolution*, **21**, 433–8.

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 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**, 133–42.

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.

Aguilar JR, Schut E, Merino S *et al.* (2013) MHC class II B diversity in blue tits: a preliminary study. *Ecology and evolution*, **3**, 1878–89.

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, Negro JJ (2007) Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. *Journal of molecular evolution*, **65**, 541–54.

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

Alcaide M, Muñoz J, Martínez-de la Puente J, Soriguer R, Figuerola J (2014) Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot Fulica atra (Aves: Rallidae). *Ecology and evolution*, **4**, 688–98.

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

Anmarkrud J a, 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**, 1206–17.

Arbogast BS, Drovetski S V, Curry RL (2006) The origin and diversification of Galapagos

mockingbirds. *Evolution*, **60**, 370–382.

Babik W (2010) Methods for MHC genotyping in non-model vertebrates. *Molecular ecology resources*, **10**, 237–51.

Babik W, Durka W, Radwan J (2005) Sequence diversity of the MHC DRB gene in the Eurasian beaver (Castor fiber). *Molecular ecology*, **14**, 4249–57.

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

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

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

Baratti M, Dessì-Fulgheri F, Ambrosini R *et al.* (2012) MHC genotype predicts mate choice in the ring-necked pheasant Phasianus colchicus. *Journal of evolutionary biology*, **25**, 1531–42.

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.

BirdLife International (2012) The IUCN Red List of Threatened Species. Version 2014.3. www.iucnredlist.org. Downloaded on 25 November 2014.

BirdLife International (2013) The IUCN Red List of Threatened Species. Version 2014.3. www.iucnredlist.org. Downloaded on 25 November 2014.

Blumstein DT, Daniel JC (2005) The loss of anti-predator behaviour following isolation on islands. *Proceedings of the Royal Society B*, **272**, 1663–8.

Bollmer JL, Dunn PO, Whittingham LA, Wimpee C (2010) Extensive MHC class II B gene duplication in a passerine, the common Yellowthroat (Geothlypis trichas). *The Journal of heredity*, **101**, 448–60.

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.

Bollmer JL, Vargas FH, Parker PG (2007) Low MHC variation in the endangered Galápagos penguin (Spheniscus mendiculus). *Immunogenetics*, **59**, 593–602.

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*, **273**, 1111–6.

Bonneaud C, Sorci G, Morin V *et al.* (2004) Diversity of Mhc class I and IIB genes in house sparrows (Passer domesticus). *Immunogenetics*, **55**, 855–65.

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.

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

Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution*, **21**, 255–65.

Burri R, Hirzel HN, Salamin N, Roulin A, Fumagalli L (2008) Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. *Molecular biology and evolution*, **25**, 1180–91.

Burri R, Promerová M, Goebel J, Fumagalli L (2014) PCR-based isolation of multigene families: lessons from the avian MHC class IIB. *Molecular ecology resources*, **14**, 778–88.

Burri R, Salamin N, Studer RA, Roulin A, Fumagalli L (2010) Adaptive divergence of ancient gene duplicates in the avian MHC class II beta. *Molecular biology and evolution*, **27**, 2360–74.

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

Chaves LD, Faile GM, Krueth SB, Hendrickson J a, Reed KM (2010) Haplotype variation, recombination, and gene conversion within the turkey MHC-B locus. *Immunogenetics*, **62**, 465–77.

Chaves LD, Krueth SB, Reed KM (2007) Characterization of the turkey MHC chromosome through genetic and physical mapping. *Cytogenetic and genome research*, **117**, 213–20.

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

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.

Deem SL, Cruz MB, Higashiguchi JM, Parker PG (2012) Diseases of poultry and endemic birds in Galapagos: implications for the reintroduction of native species (T Katzner, K Acevedo-Whitehouse, Eds,). *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.

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**, 2455–7.

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

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; international journal of organic evolution*, **67**, 679–87.

Edwards S V., Grahn M, Potts WK (1995) Dynamics of Mhc evolution in birds and crocodilians: amplification of class II genes with degenerate primers. *Molecular ecology*, **4**, 719–29.

Eimes J a, Bollmer JL, Whittingham L a *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**, 1847–56.

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, 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.

Ekblom R, Stapley J, Ball A, Birkhead T (2011) Genetic mapping of the major histocompatibility complex in the zebra finch (Taeniopygia guttata). *Immunogenetics*, **63**, 523–30.

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.

Forsberg L a, Dannewitz J, Petersson E, Grahn M (2007) Influence of genetic dissimilarity in the reproductive success and mate choice of brown trout - females fishing for optimal MHC dissimilarity. *Journal of evolutionary biology*, **20**, 1859–69.

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

Gaudieri S, Dawkins R, Habara K (2000) SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. *Genome research*, **10**, 1579–1586.

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

Grant P, Grant N (1979) Breeding and feeding of Galapagos mockingbirds, Nesomimus parvulus. *Auk*, **96**, 723–736.

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.

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

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

Hess CM, Edwards S V. (2002) The Evolution of the Major Histocompatibility Complex in Birds. *BioScience*, **52**, 423.

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. Berlin.

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

Hoeck PEA, Beaumont M (2010) Saving Darwin's muse: evolutionary genetics for the recovery of the Floreana mockingbird. *Biology Letters*, **6**, 212–5.

Hoeck PEA, Bollmer JL, Parker PG, Keller LF (2010) Differentiation with drift: a spatio-temporal genetic analysis of Galapagos mockingbird populations (Mimus spp.). *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 1127–38.

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**, 1538–41.

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.

Hosomichi K, Miller MM, Goto RM *et al.* (2008) Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. *Journal of immunology*, **181**, 3393–9.

Hughes CR, Miles S, Walbroehl JM (2008) Support for the minimal essential MHC hypothesis: a parrot with a single, highly polymorphic MHC class II B gene. *Immunogenetics*, **60**, 219–31.

Hurles ME, Lupski JR (2006) Recombination Hotspots in Nonallelic Homologous Recombination. In: *Genomic Disorders* (eds Lupski JR, Stankiewicz P), pp. 341–355. Humana Press, Totowa, NJ.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies.

Molecular biology and evolution, **23**, 254–67.

Jarvi SI, Tarr CL, McIntosh CE, Atkinson CT, Fleischer RC (2004) Natural selection of the major histocompatibility complex (Mhc) in Hawaiian honeycreepers (Drepanidinae). *Molecular ecology*, **13**, 2157–68.

Jeffreys AJ, May CA (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nature genetics*, **36**, 151–6.

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.

Kalbe M, Eizaguirre C, Dankert I *et al.* (2009) Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings of the Royal Society B*, **276**, 925–34.

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

Kaufman J, Völk H, Wallny HJ (1995) A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunological reviews*, **143**, 63–88.

Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics*, **56**, 683–95.

Kikkawa EF, Tsuda TT, Sumiyama D *et al.* (2009) Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus Spheniscus). *Immunogenetics*, **61**, 341–52.

Kimura M (1983) *The neutral theory of molecular evolution.* Cambridge.

Kloch A, Babik W, Bajer A, Siński E, Radwan J (2010) Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole Myodes glareolus. *Molecular ecology*, **19 Suppl 1**, 255–65.

Knafler GJ, Fidler A, Jamieson IG, Robertson BC (2014) Evidence for multiple MHC class II β loci in New Zealand's critically endangered kakapo, Strigops habroptilus. *Immunogenetics*, **66**, 115–21.

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**,

1208–22.

Krzywinski M, Schein J, Birol İ (2009) Circos: an information aesthetic for comparative genomics. *Genome research*, 1639–1645.

Lam TH, Shen M, Chia J-M, Chan SH, Ren EC (2013) Population-specific recombination sites within the human MHC region. *Heredity*, **111**, 131–8.

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.

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

Lighten J, Oosterhout C Van, Paterson IG, McMullan M, Bentzen P (2014b) Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (Poecilia reticulata). *Molecular ecology resources*, **14**, 753–67.

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.

Mainguy J, Worley K, Côté SD, Coltman DW (2006) Low MHC DRB class II diversity in the mountain goat: past bottlenecks and possible role of pathogens and parasites. *Conservation Genetics*, **8**, 885–891.

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

Matson KD (2006) Are there differences in immune function between continental and insular birds? *Proceedings of the Royal Society B*, **273**, 2267–74.

Matson KD, Ricklefs RE, Klasing KC (2005) A hemolysis-hemagglutination assay for characterizing constitutive innate humoral immunity in wild and domestic birds. *Developmental and comparative immunology*, **29**, 275–86.

Maxwell MH (1993) Avian blood leucocyte responses to stress. *World's Poultry Science*

Journal, **49**, 34–43.

McMullan M, Oosterhout C Van (2012) Inference of selection based on temporal genetic differentiation in the study of highly polymorphic multigene families. *PloS one*, **7**, e42119.

Meglécz E, Piry S, Desmarais E *et al.* (2011) SESAME (SEquence Sorter & AMplicon Explorer): genotyping based on high-throughput multiplex amplicon sequencing. *Bioinformatics (Oxford, England)*, **27**, 277–8.

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

Miller HC, Bowker-Wright G, Kharkrang M, Ramstad K (2011) Characterisation of class II B MHC genes from a ratite bird, the little spotted kiwi (Apteryx owenii). *Immunogenetics*, **63**, 223–33.

Miller HC, Lambert DM (2004a) Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics*, **56**, 178–91.

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**, 3709–21.

Møller AP, Rózsa L (2005) Parasite biodiversity and host defenses: chewing lice and immune response of their avian hosts. *Oecologia*, **142**, 169–76.

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.

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular biology and evolution*, **3**, 418–426.

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 7799–806.

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families.

Annual review of genetics, **39**, 121–52.

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.

Nowak M (1992) The optimal number of major histocompatibility complex molecules in an individual. *Proc Natl Acad Sci*, **89**, 10896–10899.

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

Ohta T (1991) Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. , **88**, 6716–6720.

Oosterhout C Van (2009a) Trans-species polymorphism, HLA-disease associations and the evolution of the MHC. *Communicative & integrative biology*, **2**, 408–410.

Oosterhout C Van (2009b) A new theory of MHC evolution: beyond selection on the immune genes. *Proceedings of the Royal Society B*, **276**, 657–65.

Oosterhout C Van, 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.

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.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. *Trends in genetics*, **26**, 177–87.

Owen JP, Delany ME, Cardona CJ, Bickford A a., Mullens B a. (2009) Host inflammatory response governs fitness in an avian ectoparasite, the northern fowl mite (Ornithonyssus sylviarum). *International Journal for Parasitology*, **39**, 789–799.

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

Pearman PB, Garner TWJ (2005) Susceptibility of Italian agile frog populations to an emerging strain of Ranavirus parallels population genetic diversity. *Ecology Letters*, **8**,

401–408.

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

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

Pond SLK, Frost SDW, Muse S V (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics (Oxford, England)*, **21**, 676–9.

Pozsonyi E, György B, Berki T *et al.* (2009) HLA-association of serum levels of natural antibodies. *Molecular immunology*, **46**, 1416–23.

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

R Core Team (2014) R: A Language and Environment for Statistical Computing.

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.

Radwan J, Zagalska-Neubauer M, Cichoń M *et al.* (2012) MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Molecular ecology*, **21**, 2469–79.

Reid JM, Arcese P, Keller LF *et al.* (2007) Inbreeding effects on immune response in free-living song sparrows (Melospiza melodia). *Proceedings of the Royal Society B*, **274**, 697–706.

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

Richardson DS, Komdeur J, Burke T, von Schantz T (2005) MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. *Proceedings of the Royal Society B*, **272**, 759–67.

Sato A, Tichy H, Grant PR *et al.* (2011) Spectrum of MHC class II variability in Darwin's

finches and their close relatives. *Molecular biology and evolution*, **28**, 1943–56.

Schrider DR, Hahn MW (2010) Gene copy-number polymorphism in nature. *Proceedings of the Royal Society B*, **277**, 3213–21.

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*, **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.

Shiina T, Shimizu S, Hosomichi K *et al.* (2004) Comparative Genomic Analysis of Two Avian (Quail and Chicken) MHC Regions. *The Journal of Immunology*, **172**, 6751–6763.

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.

Siddle H V, Marzec J, Cheng Y, Jones M, Belov K (2010) MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. *Proceedings of the Royal Society B*, **277**, 2001–6.

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

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, 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.

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.

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.

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B*, **277**, 979–88.

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, Wang B, Meyer-Lucht Y, Höglund J (2013) Evolutionary history of black grouse major histocompatibility complex class IIB genes revealed through single locus sequence-based genotyping. *BMC genetics*, **14**, 29.

Strand TM, Westerdahl H, Pontarp M *et al.* (2012) Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. *Proceedings of the Royal Society B*, **279**, 4457–63.

Stutz WE, Bolnick DI (2014) Stepwise Threshold Clustering: A New Method for Genotyping MHC Loci Using Next-Generation Sequencing Technology. , **9**, 25–27.

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

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**, 619–33.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*, **30**, 2725–9.

Tompkins DM, Mitchell R a, Bryant DM (2006) Hybridization increases measures of innate and cell-mediated immunity in an endangered bird species. *The Journal of animal ecology*, **75**, 559–64.

Vincek V, O'Huigin C, Satta Y *et al.* (1997) How large was the founding population of

Darwin's finches? *Proceedings of the Royal Society B*, **264**, 111–118.

Wang Z, Zhou X, Lin Q, Fang W, Chen X (2013) Characterization, polymorphism and selection of major histocompatibility complex (MHC) DAB genes in vulnerable Chinese egret (Egretta eulophotes). *PloS one*, **8**, e74185.

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

Westerdahl H, Asghar M, Hasselquist D, Bensch S (2012) Quantitative disease resistance: to better understand parasite-mediated selection on major histocompatibility complex. *Proceedings of the Royal Society B*, **279**, 577–84.

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.

Westerdahl H, Wittzell H, von Schantz T (2000) Mhc diversity in two passerine birds: no evidence for a minimal essential Mhc. *Immunogenetics*, **52**, 92–100.

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.

Wickham H (2009) *ggplot2: elegant graphics for data analysis*. Springer New York.

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**, 117–28.

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

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.

Zylstra P, Rothenfluh HS, Weiller GF, Blanden R V, Steele EJ (1998) PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. *Immunology and cell biology*, **76**, 395–405.

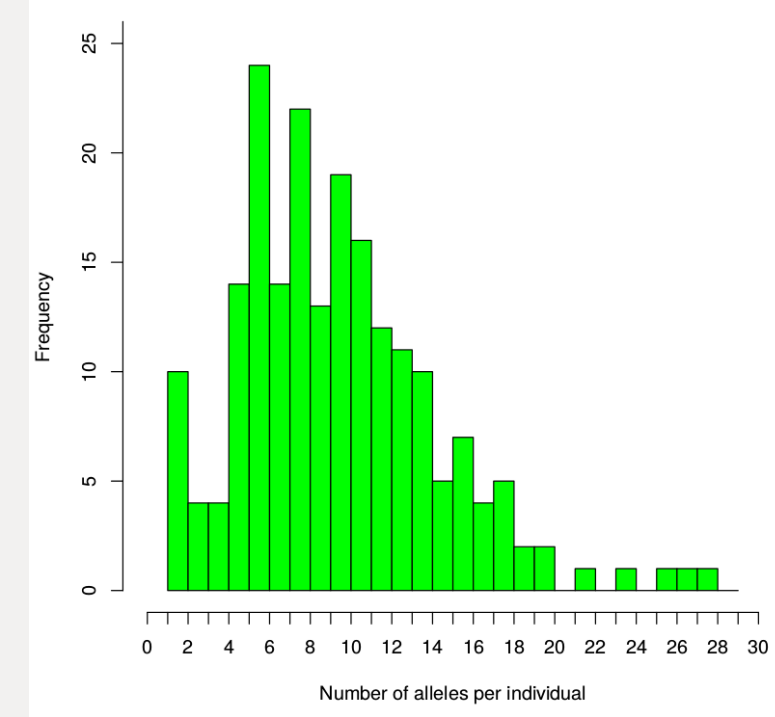
8. Appendix

Appendix 1: Variables describing MHC variation and ecological factors for studied populations of mockingbirds.

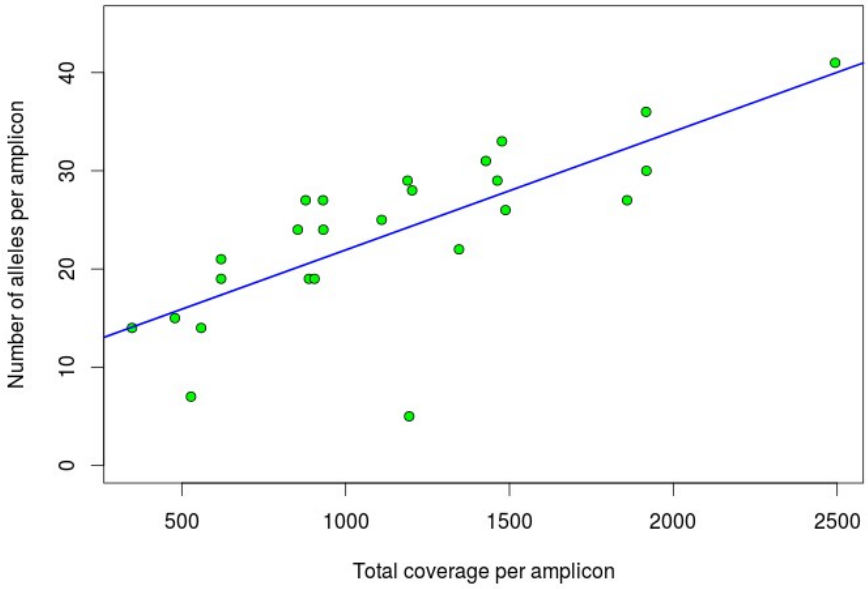
		Mean characteristics of MHCII β2 alleles per population						Ecological factors					
Population (Spp)	№ Ind.	№ A. per Ind.	s.d.	Tot. № A.	№ Priv. A.	Dist. of A.	Dist. s.d.	Isl. Age (My)	Isl. size (Ha)	SSR. A. rich.	Lysis	Agglut.	Tot. par.
Champion (M.tri)	13	8.54	3.57	30	0	0.094	0.019	1.50	9.50	1.20	3.27	9.33	11.27
Continent (M.pol)	13	9.08	5.95	35	25	0.173	0.027	NA	NA	NA	NA	NA	NA
Espanola (M.mac)	18	11.72	3.43	46	0	0.074	0.015	3.00	6048.00	1.86	3.58	8.48	11.31
Fernandina (M.par)	14	17.21	4.51	48	3	0.099	0.018	0.04	64248.00	2.93	NA	NA	NA
G.b.E. (M.mac)	6	10.17	2.99	19	0	0.074	0.015	3.00	58.00	1.59	4.33	9.33	22.40
G.b.F. (M.tri)	14	10.93	3.52	35	1	0.113	0.020	1.50	81.00	1.77	3.68	9.18	9.83
Genovesa (M.par)	5	5.80	3.83	23	1	0.121	0.019	0.30	1410.80	1.95	NA	NA	NA
Isabela (M.par)	17	13.29	6.55	64	4	0.098	0.018	0.50	458812.00	3.10	2.95	9.19	9.91
Marchena (M.par)	16	7.06	2.26	30	1	0.080	0.015	0.60	129966.00	2.63	3.55	8.45	10.88
Pinta (M.par)	14	12.29	3.27	35	0	0.103	0.019	0.70	5940.00	2.30	NA	NA	NA
Rabida (M.par)	14	9.64	3.27	42	1	0.100	0.017	1.30	499.30	2.56	NA	9.13	13.56
San Crist. (M.mel)	13	6.23	2.31	27	1	0.066	0.012	2.40	55808.60	2.46	3.10	9.35	5.55
Santiago (M.par)	16	9.19	2.83	52	3	0.094	0.016	0.80	58465.00	3.50	3.70	7.37	2.50
St Cruz (M.par)	15	9.13	2.33	55	5	0.095	0.016	1.10	98555.00	4.20	3.50	8.50	9.27
St Fe (M.par)	15	4.00	2.27	22	1	0.097	0.017	2.90	2413.00	2.06	3.71	9.71	10.67

Names of populations abbreviated (G.b.E=Gardner by Espanola, G.b.F=Gardner by Floreana, San Cris.=San Cristobal). Other abbreviations: Ind.=Individual; A.=allele; Tot=total; Priv.=private; Dist.=genetic distance; Isl.=island; SSR=microsatellites; rich.=richness; Agglut.=agglutination; Tot. par.= number of all ecto-parasites per individual.

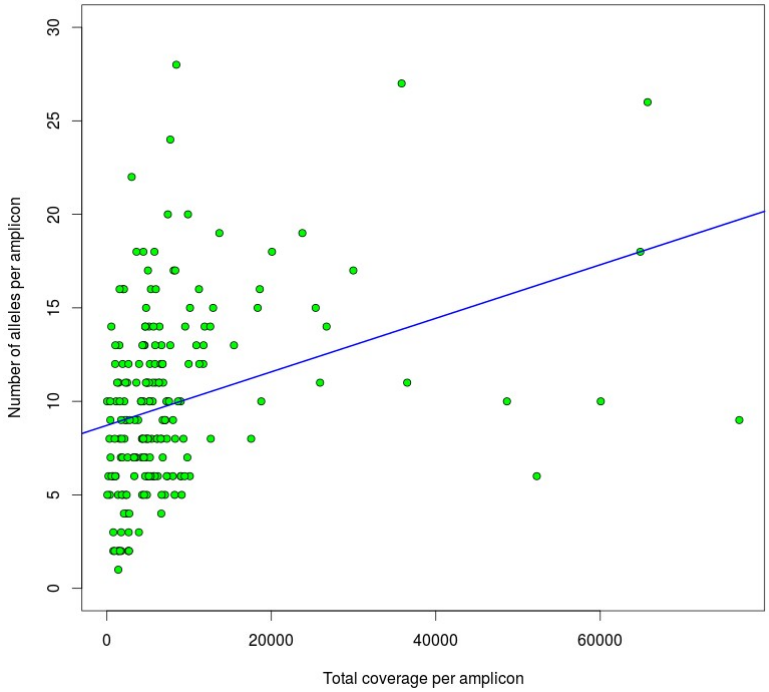
Appendix 2: Frequency of the number of MHCII β2 alleles found across individuals.



Appendix 3: Correlation between the number of MHCII β2 alleles and amplicon coverage in the preliminary 454 sequencing and the final Ion torrent sequencing.

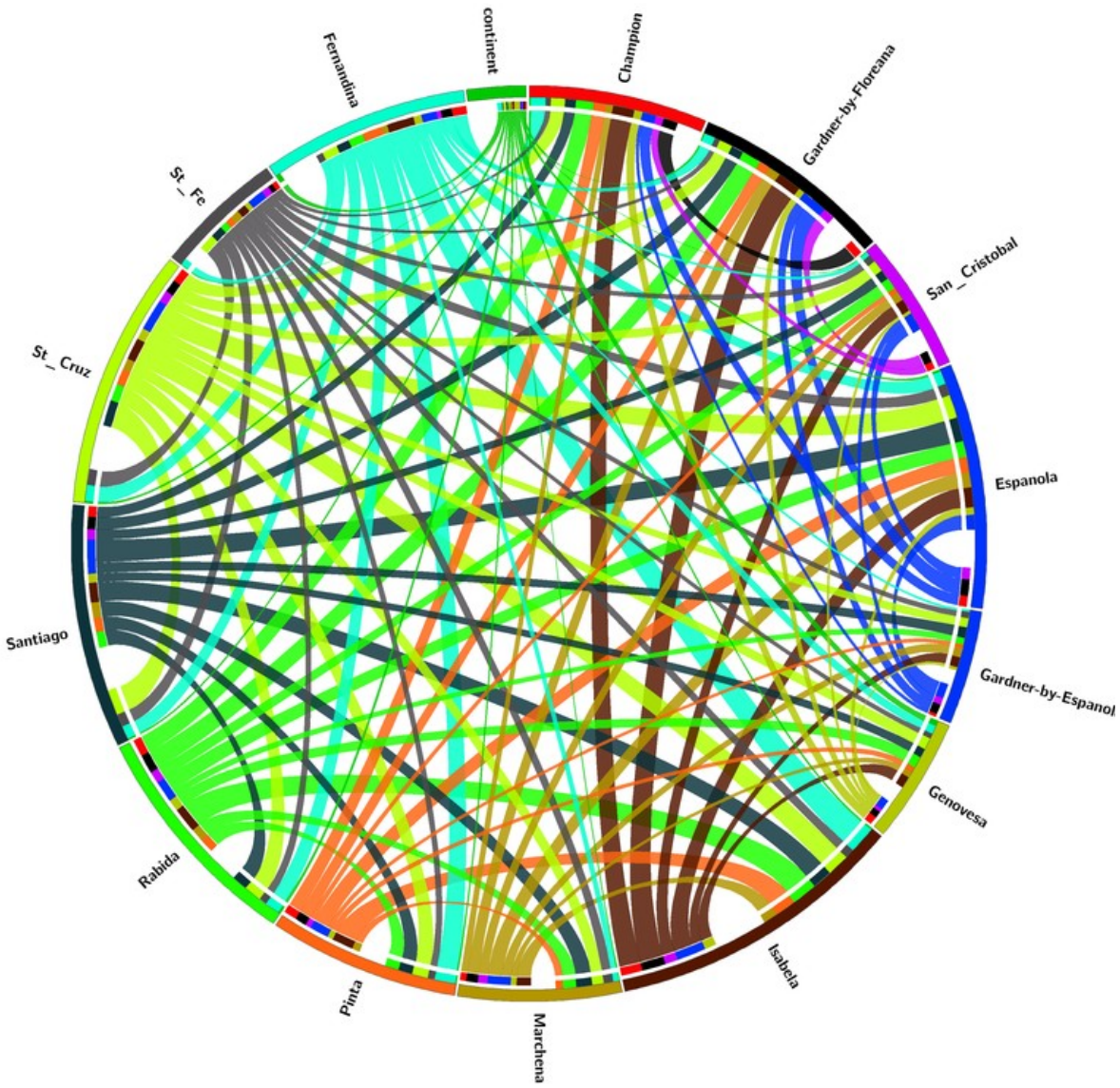


454 sequencing R^2=0.550, p=1.326e-05



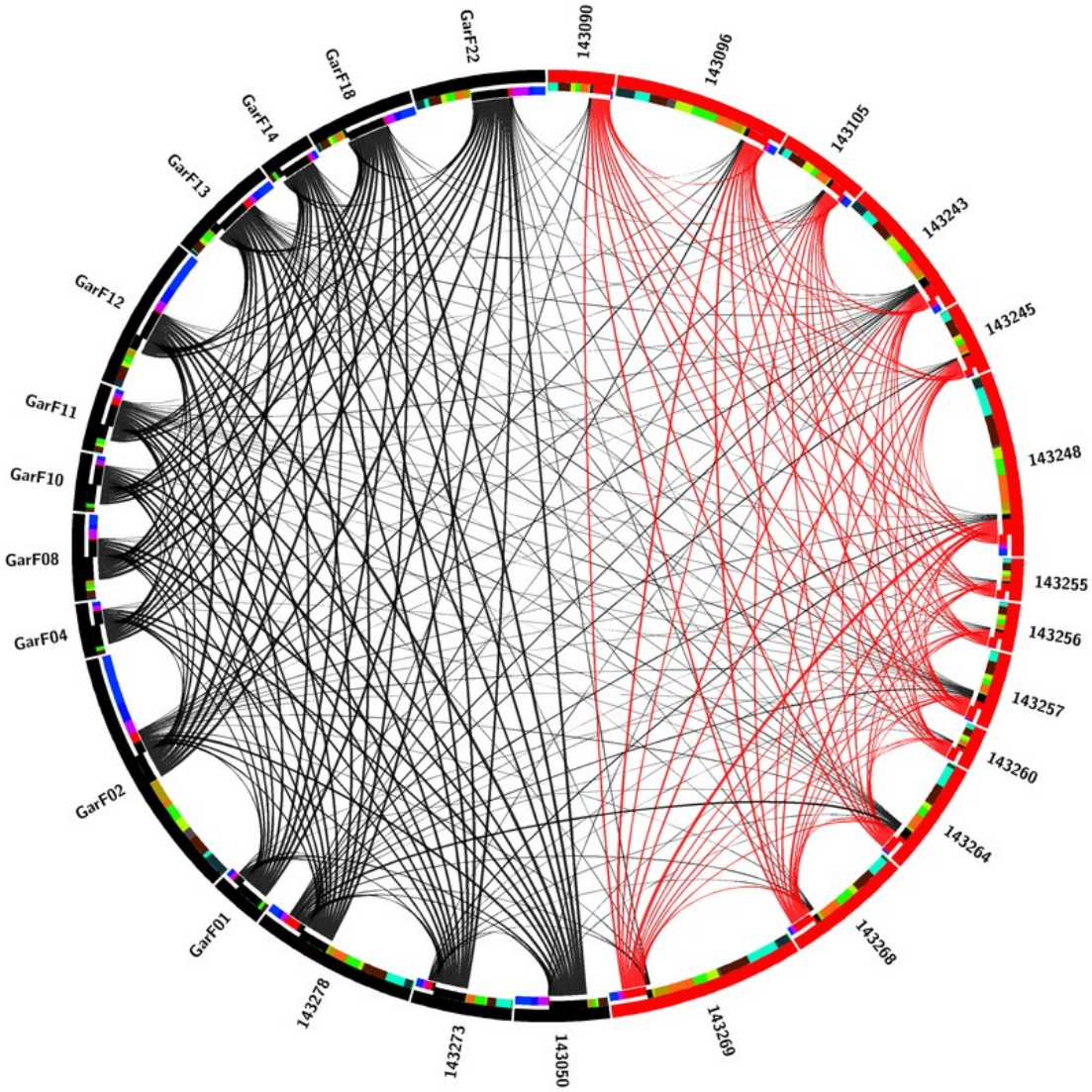
Ion Torrent sequencing R^2=0.1471, p= 9.73e-09

Appendix 4: Sharing of MHCII β2 alleles between populations.



Sharing of alleles between populations is depicted by ribbons connecting the populations. Width of the ribbon is proportional to the number of shared alleles between the populations. Outer bars around the circular graph denote the populations and the inner bars show proportions of alleles shared between the populations. The visualization was made by the program Circos.

Appendix 5: Sharing of alleles between individuals from two populations of the endangered Floreana mockingbird.



Similar representation of the sharing of alleles as in the previous picture. In this case the outer bars represent individuals. Populations are defined by colour. Red colour denotes individuals from Champion and black from Gardner by Floreana. The proportion of alleles shared with other populations is also visible in the inner bar.