The spatial scale of immune gene variation within and among bottlenecked populations

Photo by Karl P. Phillips

Catalina González-Quevedo

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

School of Biological Sciences

September 2014

Abstract

The general aim of this thesis is to explore different spatial scales at which pathogen-mediated selection drives the evolution of immune genes across and within populations of Berthelot's pipit (*Anthus berthelotii*), a historically bottlenecked passerine endemic to the oceanic islands of the Canary, Selvagens and Madeira archipelagos. I first investigated the evolution of key innate immune genes among the populations that the pipits inhabit. I found that while demographic history has played the major role in shaping patterns of among population variation at toll-like receptor loci, balancing selection (possibly pathogen-mediated) appears to have helped maintain functional variation at some specific loci. Second, I assessed the contribution of environmental factors to pathogen distribution and their subsequent effects on the major histocompatibility complex (MHC) class I genes of the acquired immune system within the population on Tenerife. I found a high prevalence of malaria in this population, the presence of which was correlated with climatic and anthropogenic variables: temperature, distance to poultry farms and distance to artificial water sources. Within the MHC I found evidence of trans-species polymorphism and gene conversion, and signatures of positive selection. Using landscape genetic analysis methods I found no evidence for overall within population patterns of structure at either neutral markers or the MHC. However, one MHC allele was associated to malaria infection risk and its distribution was (more strongly) associated with distance to poultry farms. These results suggest that demographic processes are the most important evolutionary force shaping variation at functional loci in isolated, bottlenecked populations. Nevertheless, selection can also shape patterns of variation at immunity loci, both at the coarser and the finer landscape scale, apparently in response to pathogens. This study therefore highlights the importance of considering different spatial scales when studying the evolutionary processes that shape functional genetic variation within populations.

Acknowledgements

I start by thanking my supervisory team. I would like to thank David Richardson for being a great friend, mentor, for being patient and for always finding ways to make me laugh. Thank you for making me feel at home. Richard Davies, thank you for teaching me so many things that I never knew I could understand and for being so patient. Lewis Spurgin, your never-ending input was invaluable for this PhD, thanks for always having (or making) time to explain stuff and for helping me though the difficult times. I would also like to thank David Wright and Karl Phillips, you were not part of my supervisory team, but I learned so much from you. Thanks for always putting up with my questions and troubles, and for always finding the time to help me go through stuff. I would also like to thank COLCIENCIAS for funding my PhD.

Being away from home for four years is not easy, you miss family, friends, food, weather… (the list goes on), but it's all easier when you have wonderful people around you. UEA friends, you were an incredibly lovable lot. I apologise if I was not very sociable, but believe me, I wouldn't have made it through this PhD if I hadn't had you around to make me laugh, bring me tea, coffee, cake (yum!) and biscuits. Special thanks to Alicia, Aly, Amanda, Amy, Becca, Becky, Cat, Chris, Clare, Claudia, Conrad, Damian, Danni G, Danni P, Dave C, Dave W, Elli, Emma, Freydís, Helena, Jacob, James, Jo, José, Karl, Kat, Kelly, Lewis, Liam, Liz, Lucy, Mike, Phil, Sarah, Sian, Tim, Tom, Vero, Wayne, Will. Special thanks go to my fellow Cabbages, that office was the place I spent most of my time during the past four years, and you never let me get bored. I will miss you all very much. I would also like to thank the faculty at UEA for all the helpful discussions. Special thanks to Jenny, Matt, Tracey, Iain, Martin, Cock, Brent and Tove.

I would like to thank everyone that made fieldwork in Tenerife possible and extremely enjoyable. Special thanks to Hen Bellman for putting up with me for three whole months! To David Padilla and Juan Carlos Illera, thank you for teaching me so much and for being such good friends even after fieldwork was done.

I wouldn't be anywhere near where I am today if it wasn't for the support of my family. Even when I was far away I always felt your warmth and support. Thank you Mom, Adri and little princess Emma for supporting me through these four years; thank you for always being there for me, comforting me when times were difficult and making it possible for me to achieve this. I missed you so much! I hope I've made you proud. I also thank my dear brother Andrés and my Dad, I miss you but it makes me happy to think that you would be proud of me.

Samuel, I couldn’t have gone through this journey without you. Thank you for always being there for me even during the difficult times when we were far apart, and for putting up with the PhD-related bad mood. Without your love and support this wouldn’t have been possible. Thanks for always encouraging me in times of doubt (PhD wise, off course!) and for being a wonderful partner.

Contents

Author contributions

One of the chapters of this thesis has been published, and three are manuscripts currently in preparation or in review. All chapters have involved collaboration, I am the lead author on all manuscripts, and have made the largest contribution to all. Below I include the full citation for each chapter, and highlight my contribution.

Chapter 2: González-Quevedo C, Spurgin LG, Illera JC & Richardson DS (In prep.)

CGQ did all the lab work, analysed the data and drafted the manuscript (80%).

Chapter 3: González-Quevedo C, Davies RG & Richardson DS (2014). *Journal of Animal Ecology* 83, 1091-1102.

CGQ collected the samples, did all the lab work, analysed the data and drafted the manuscript (80%).

Chapter 4: González-Quevedo C, Spurgin LG, Phillips KP & Richardson DS (In prep.)

CGQ collected the samples, did most of the lab work, analysed most of the data and drafted the manuscript (60%).

Chapter 5: González-Quevedo C, Davies RG, Spurgin LG, Phillips KP & Richardson DS (In prep.)

CGQ collected the samples, did most of the lab work, analysed the data and drafted the manuscript (80%).

Chapter 1

Introduction

Western slopes of Tenerife (Canary Islands), overlooking La Gomera

1.1 Evolutionary ecology

Evolutionary ecology deals with the study of how present-day ecological processes are both causes and consequences of evolutionary change (Losos 1994; Pianka 2011). Understanding the mechanisms that promote evolutionary change in wild populations - one of the central issues of evolutionary ecology - helps explain why variation exists in populations and what the consequences of such variation are (Schmid-Hempel 2003). Natural selection is one of the drivers of evolutionary change (Darwin 1859); it acts on functional genetic variation and leads to the adaptation of organisms to their environment (Fisher 1930). However, random fixation of selectively neutral variants in some instances is more important in shaping variation than selection (Kimura 1968). Evolutionary and ecological processes occur at different spatial scales, and so considering the different scales at which such processes vary is important for understanding how populations evolve in a spatially explicit manner (Levin 1992). A particular focus of evolutionary ecology is therefore to understand natural selection and the adaptive evolution of populations at different spatial scales (Chave 2013; Richardson *et al.* 2014).

One area of research in which evolutionary ecology approaches are particularly important is disease ecology. Disease causing agents, or pathogens, have very close ecological relationships with their hosts. They have negative impacts on survival and on reproductive success of their hosts and are thus strong selective agents (Haldane 1949). Evolution of both hosts and pathogens is dependent on one another, and the outcomes of such relationship depend on ecological (i.e. climatic factors, species interactions) and evolutionary processes (i.e. genetic variation at genes involved in the host-pathogen interaction, selection, gene flow and stochastic processes) (Altizer *et al.* 2003; Hawley & Altizer 2011). Host-pathogen interactions in the landscape are of particular interest since variable features of habitat and environment associated with different spatial scales affect patterns of pathogen transmission (Archie *et al.* 2009). Studying such interactions at these different spatial scales is crucial in order to recognize disease threats to wild populations, and to provide information relevant for species conservation (Hawley & Altizer 2011). Although many studies have focused on understanding spatial variation of host-pathogen interactions in the landscape, these have mainly been done at large spatial scales, across discrete populations (reviewed in Biek & Real 2010). There is a clear need for studies that assess how pathogen and host interactions vary at small spatial scales.

1.2 Genetic diversity

Genetic diversity provides the raw material for evolution. Genetic differences among individuals, populations and species begin with changes in the genetic material of individuals, generated by several non-mutually exclusive mechanisms, including mutation, recombination, gene duplication and gene conversion. If such changes occur in germ line cells, the genetic variant generated can be inherited and passed on to the next generation. Mutations principally arise naturally and randomly when errors occur in the process of DNA replication (Loeb *et al.* 1974), but also by insertion of mobile genetic elements (Georgiev 1984; Boehne *et al.* 2008) and chromosomal breaks (Malkova & Haber 2012). Recombination involves the exchange of genetic material between two homologous sequences, and occurs during meiosis in gamete development, when homologous chromosomes are joined (Whitehouse 1970). Gene conversion is a form of recombination which involves unidirectional transfer of genetic material from a 'donor' sequence to a highly homologous sequence (Slightom *et al.* 1980; Chen *et al.* 2007). Gene conversion is initiated by double strand chromosomal breaks and it occurs often between duplicated loci (Chen *et al.* 2007). Gene duplication (Bridges 1936) can result by unequal crossing over, retroposition or chromosomal (or genome) duplication (Zhang 2003).

The highest levels of genetic diversity are generally found in DNA sequences with little or no functional significance; such regions do not contain coding DNA or changes in them do not result in functional change of the molecule coded. In genes or coding DNA many of such nucleotide changes are found in third codon positions (Kimura 1977). Due to the redundancy of the genetic code, roughly 2/3 of random nucleotide substitutions at the third position of the codon are synonymous. These changes are thus invisible to selection (Kimura 1968). Such genetic diversity is referred to as 'neutral genetic variation'. Conversely, genetic diversity generated in DNA regions that code for functional molecules might result in functional changes of the molecule, i.e. a change in the phenotype; thus, the levels of functional genetic diversity are lower than those of neutral genetic diversity (Kimura 1991), although there are some exceptions as I shall cover below. Studying genetic diversity in wild populations is important because it provides information on their ability to evolve and adapt to changing environmental conditions. It is therefore imperative to understand how different evolutionary forces act upon genetic diversity.

Once genetic diversity is generated, three evolutionary forces act to promote evolutionary change in a population from one generation to another (Wright 1931). First, random events unrelated to the characteristics of the genetic material may result in reduction in genetic variation because by chance not all alleles will be passed on to the next generation. These random events include accidental death of individuals and random variation in reproductive success (not dependent on the genotype/phenotype). This process, known as 'genetic drift', leads to either loss or fixation of alleles resulting in a reduction of genetic diversity in a population (Wright 1948). Second, the movement of individuals between populations with different genetic backgrounds, and their subsequent reproduction within those populations, results in 'gene flow'. This process results in an increase of genetic variation in the 'receptor' population and reduces differences among populations (Slatkin 1987). Finally, if different genetic variants encode phenotypic differences which result in the differential reproductive success of those individuals that carry them, then the more successful variants will increase in frequency in the following generation, while the less successful ones will decrease (Fisher 1930). This process is known as 'natural selection' and results in adaptation of populations to their environment (Darwin 1859). When the environmental conditions change, the population adapts to the new conditions contingent on the presence of the appropriate genetic variants in the population (Fisher 1930). Selection can reduce or increase genetic diversity, depending on the mechanism of selection in operation. For example, purifying selection reduces genetic diversity by removing genetic variants that are disadvantageous (Loewe 2008). Balancing selection, however, may result in the increase of genetic diversity because multiple genetic variants are advantageous, therefore they are all favoured by selection and kept in the gene pool of the population (Hedrick & Thomson 1983; Hedrick 1998; See section 1.4).

The environmental conditions to which individuals are exposed determine the mechanisms through which selection operates in nature, and constitute the selective pressures that act upon genetic variation. These selective pressures might be abiotic or biotic. Abiotic selective pressures include climatic conditions (e.g. temperature, rainfall, altitude). Biotic selective pressures include intra-specific interactions (i.e. competition for mates and resources); and inter-specific interactions (i.e. competition, predation and parasitism). Among the inter-specific interactions pathogens are potent agents of natural selection (Haldane 1949) that are highly relevant to wild populations.

1.3 Pathogens as selective agents

Pathogens use other organisms for their growth and survival and have detrimental effects on the intrinsic growth rates of their host at the individual and the population level (Anderson & May 1978, 1979). They are a very diverse group of organisms including viruses, bacteria, protozoa, fungi, flatworms, nematodes and arthropods that parasitize many species of plants and animals (reviewed in Price 1980; Noble *et al.* 1989). The effects that pathogens have can vary depending on the characteristics of the host and the pathogen, but from a spatial perspective, pathogens have been shown to influence every aspect of the population biology of the host (Tompkins *et al.* 2002; Lion *et al.* 2006). The costs of pathogenic infection may be a consequence of the effects that pathogens have on the metabolic efficiency of the host, or on the allocation of host resources to dealing with the pathogen at a cost to reproduction (Møller 1997). Pathogens are known to affect various physiological processes such as nutrient absorption, efficiency of the circulatory system and locomotion of their hosts (Noble *et al.* 1989). Pathogens have also been shown to have considerable impacts on the population structure of wild populations and to have a pivotal role in ecosystem functioning (Altizer *et al.* 2003; Duncan & Little 2007; Riordan *et al.* 2007). A wide range of pathogens have been detected in wild animals and they have been linked to recent declines of wildlife populations (Smith *et al.* 1998; Riordan *et al.* 2007; Blaustein *et al.* 2012). Thus, infectious diseases are likely to play a role in future species endangerment (Smith *et al.* 2006; Smith *et al.* 2009).

For all the reasons discussed above, pathogens are rightly regarded as detrimental, but what is not always acknowledged is that they are probably one of the strongest drivers of evolution (Haldane 1949; Antonovics 1993; Fumagalli *et al.* 2011). The ecological relationship between a pathogen and its host is very close and this is reflected in the evolutionary processes involving both host and pathogen. Hosts evolve mechanisms to defend themselves from pathogens, while pathogens, in turn, evolve ways to overcome host defences, resulting in a co-evolutionary arms race (Slade & McCallum 1992; Danilova 2006). This process means that changes in gene frequencies as a result of selection in the host population induce changes in gene frequencies in the pathogen population (Lively & Dybdahl 2000; Woolhouse *et al.* 2002; Ebert 2008).

1.4 The immune response and evolution of host immune genes

The impacts that pathogens have on the evolution of hosts can be evidenced in the defensive strategies that hosts have evolved. Hosts have evolved different behaviours to decrease the probability of infection (Minchella 1985; Loehle 1995; Behringer *et al.* 2006; Parker *et al.* 2010). However, infection is not always avoided, and once infection occurs, other mechanisms come into play. The metabolic, physiological and molecular mechanisms that the host displays in reaction to pathogenic infection are collectively known as the immune response. In vertebrates there are two types of immune response: the innate and the acquired immune responses (Wakelin 1984; Roitt *et al.* 2001). The innate immune system is an ancient defence mechanism that uses germline encoded receptors for the recognition of pathogens in a non-specific manner (Wakelin 1984; Medzhitov & Janeway 1997; Wakelin & Apanius 1997). The receptors of the innate immune system are called pattern recognition receptors (PRR). Because of the non-specific nature of their action, the PRRs recognise conserved microbial molecular signatures known as pathogen associated molecular patterns (PAMPs). One type of PRR was first described in *Drosophila*: The protein Toll was found to be associated with resistance to infections with fungi and Gram-positive bacteria in this genus, suggesting it had a role in immune defence (Ip & Levine 1994; Lemaitre *et al.* 1996). Since then, molecules with very similar characteristics and structure have been identified in vertebrates and are known as Toll-like receptors (TLRs) (Medzhitov *et al.* 1997; Roach *et al.* 2005).

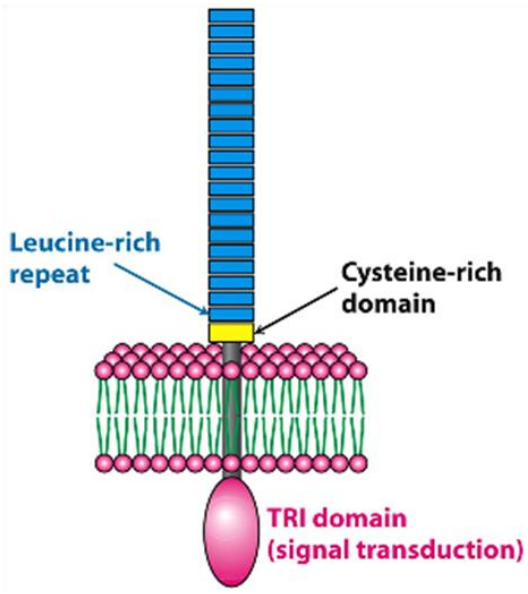


Figure 1.1 Schematic representation of the Structure of Toll-like receptors. *Unmodified figure reproduced from Berg et al. (2012)*

TLRs are type I transmembrane proteins with an extracellular domain constituted by a horseshoe-like structure consisting of Leucine-rich repeats (LRR) that mediate direct protein-protein interactions (Buchanan & Gay 1996). The cytoplasmic domain of TLRs is homologous to that of the interleukin I (IL-I) receptor and is known as the toll/IL-I receptor (TIR; Fig. 1.1; Gay & Keith 1991). When the extracellular domain of the TLR molecule binds a PAMP, an intracellular signalling pathway is initiated via the TIR that activates transcription factors that induce expression of cytokines and activation of macrophages (Belvin & Anderson 1996; Akira 2003). PAMP binding by TLRs also stimulate phagocytic cells and the subsequent presentation of pathogen peptides to T lymphocytes via the major histocompatibility complex (MHC) (Schnare *et al.* 2001). There are six major TLR families that vary considerably in the length of the LRR domain, and in the general class of PAMP they recognize (Roach *et al.* 2005). For example, TLR1LA, TLR1LB and TLR2 bind bacterial lipoproteins (Lien *et al.* 1999; Takeuchi *et al.* 2002), TLR3 binds viral RNA (Yoneyama & Fujita 2010), TLR21 binds bacterial DNA motifs (Keestra *et al.* 2010) and TLR4 binds bacterial lipopolysacharide (Poltorak *et al.* 1998). TLRs originated in the eumetazoan ancestor more than 600 million years ago (Leulier & Lemaitre 2008) and evolved independently by gene duplication before the evolutionary divergence of protostomes and deuterostomes (Roach *et al.* 2005). Although vertebrate TLRs have been regarded as an example of evolutionary conservation (Roach *et al.* 2005), recent studies have shown evidence of positive selection in TLRs in a range of organisms including birds (Downing *et al.* 2010; Alcaide & Edwards 2011; Grueber *et al.* 2014), fish (Chen *et al.* 2008), and mammals (Nakajima *et al.* 2008; Areal *et al.* 2011; Tschirren *et al.* 2011). Most of the sites identified as being under positive selection in these studies are located in the TLR extracellular domain which directly binds the PAMPS (Mikami *et al.* 2012). Specific polymorphisms within TLRs have been associated with differential pathogen resistance (Schröder & Schumann 2005; Misch & Hawn 2008; Basu *et al.* 2010), and TLR genes have been associated with pathogenic diseases through assays of differential expression (Farnell *et al.* 2003; Iqbal *et al.* 2005; Higuchi *et al.* 2008; Karpala *et al.* 2008). These data thus support the idea that pathogens are selective agents that play a role in determining patterns of variation at TLR loci. Evolutionary changes found in TLR genes may reflect changes in pathogen genes that code for PAMPs and the extant variation found at TLRs might be the consequence of the need to adapt quickly in response to evolving pathogens, or novel ones encountered in new environments (Downing *et al.* 2010). Nevertheless, direct evidence of pathogen-mediated selection (PMS) on TLRs in wild systems is

yet to be explored, thus, not much is known about the evolutionary mechanisms determining variation at these loci. Studies that assess TLR variation across populations with differing demographic histories and pathogenic pressures, and that correlate individual TLR variation with pathogenic disease are still needed to determine the relative effects of the different evolutionary forces on these genes.

Unlike the innate immune response, the acquired immune system can generate highly specific and repeatable responses to particular pathogens. The responses are initiated through the binding of pathogen derived peptides (antigens) by a specialized group of host cell molecules, encoded by a set of genes known as the major histocompatibility complex (MHC) (Bach 1976; York & Rock 1996; Roitt *et al.* 2001). There are two major classes of MHC genes, class I code for molecules that detect intracellular pathogens and present them on the cell surface of all cells in the organism. MHC class II genes code for molecules that can only be found in certain types of cells and recognize extracellular pathogens (Fig. 1.2; Frank 2002). After an MHC molecule binds an antigen and presents it at the cell surface, the T cells can recognize and bind the MHC-antigen complexes. If the T cells recognize these complexes as non self, changes in the T cells are induced that lead to the production of clones with the same T cell receptor that recognizes that particular MHC-antigen complex. The T cells are then "activated" and produce chemical substances that result in the destruction of cells that are expressing the specific MHC-antigen complex. Some of these clones become memory cells, which allow an enhanced (acquired) response if re-exposure to the antigen occurs (Wakelin 1984; Wakelin & Apanius 1997; Roitt *et al.* 2001).

The MHC is thought to have evolved less than 500 million years ago and has only been found in gnathostomes (jawed vertebrates) (Klein & Sato 1998). There is considerable variation in the organization and size of the MHC among vertebrates (Höglund 2009). Among birds, the chicken (*Gallus gallus*) MHC has been particularly extensively studied. In this species, and other closely related galliforms (Shiina *et al.* 2004; Chaves *et al.* 2009) the MHC genes are positioned very close together and form what has been called a minimal essential MHC (Kaufman *et al.* 1999). The MHC of passerines has only been well characterized in one species, the zebra finch, where it appears to differ markedly from that of fowl (Balakrishnan *et al.* 2010). For example, there is evidence of gene duplication and conversion (Miller & Lambert 2004), pseudogene formation and distribution of genes across multiple chromosomes, unlike observed in the

compact MHC of fowl (Kaufman *et al.* 1999). In general, studies across passerines show that they do not have a minimal essential MHC as seen in galliformes (Westerdahl *et al.* 1999, 2000; Balakrishnan *et al.* 2010).

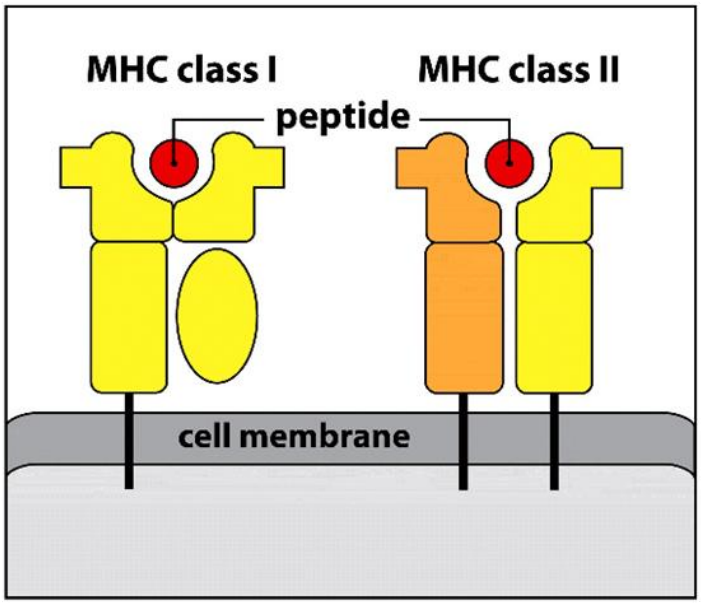


Figure 1.2 Schematic representation of the Structure of MHC class I and class II molecules. *Unmodified figure reproduced from Parham (2009).*

MHC genes have been put forward as the best candidates for the study of adaptive genetic diversity because they are extraordinarily variable (Bodmer *et al.* 1997), and are of obvious ecological relevance due to their role in detecting pathogens (Hill 1991; Hedrick 1994; Prugnolle *et al.* 2005; Höglund 2009). Evidence from studies on MHC genes in outbred wild populations supports the idea that these genes are highly variable (Mona *et al.* 2008; Lenz *et al.* 2009a), and that MHC variation in wild populations is normally greater than variation at neutral loci (Wittzell *et al.* 1998; Drake *et al.* 1999; Westerdahl *et al.* 1999; Freeman-Gallant *et al.* 2002; Richardson & Westerdahl 2003; Bonneaud *et al.* 2004). High levels of MHC variation have been found even in species shown to have low overall genetic diversity due to bottlenecks or founder effects, possibly because selection has maintained high polymorphism (Richardson & Westerdahl 2003; Aguilar *et al.* 2004; Jarvi *et al.* 2004; Wynne *et al.* 2007). However, if bottlenecks are strong enough, balancing selection is outweighed by other evolutionary forces like genetic drift (Bollmer *et al.* 2007; Babik *et al.* 2009).

The extraordinarily high variation normally observed at MHC genes suggests there are mechanisms that favour genetic diversity at these loci. Various mutational mechanisms may occur to give rise to new alleles in the MHC, from point mutations to gene conversion (Ohta 1995; Edwards & Hedrick 1998; Hedrick 1998). Once generated, other mechanisms act to select for high variation. Behavioural mechanisms that contribute towards the maintenance of MHC diversity have evolved to maintain high variation. For example, mate choice has been shown to favour higher MHC diversity; e.g. choosy females select for specific male haplotypes that generate heterozygosity in the progeny (Wedekind 1994; Brown 1998; Reusch *et al.* 2001; Penn 2002; Brouwer *et al.* 2010). The MHC has also been linked to kin recognition, allowing conspecifics to recognise (and avoid mating with) kin, through cues that reveal their MHC haplotypes, therefore favouring higher MHC diversity in the progeny (Potts *et al.* 1994; Radwan *et al.* 2008). These mechanisms contribute to shaping variation at the MHC (Hedrick 1998) but alone they cannot drive the extraordinary levels of variation seen at the MHC.

1.5 Pathogen-mediated selection on genes of the MHC

Pathogen mediated selection (PMS) has been suggested to be the main evolutionary force that drives high diversity at MHC loci (Doherty & Zinkernagel 1975). PMS can cause balancing selection (Hedrick 1998; Spurgin & Richardson 2010) whereby no specific allele is always the fittest over time, thus resulting in a more even frequency distribution of alleles in the population than that expected under the neutral model (Hedrick & Thomson 1983). Three main, non-mutually exclusive, mechanisms of PMS have been suggested: (i) heterozygote advantage (ii) negative frequency dependent selection (or rare allele advantage), and (iii) spatially heterogeneous selection (or fluctuating selection) (Potts & Wakeland 1990; Spurgin & Richardson 2010). The heterozygote advantage model (Doherty & Zinkernagel 1975) proposes that individuals heterozygous at MHC genes will be fitter than homozygotes because they will be able to recognize a wider range of pathogens, or better able to recognize a single pathogen. Various studies have shown evidence of heterozygote advantage (Hughes & Nei 1989; Penn *et al.* 2002; Evans & Neff 2009; Worley *et al.* 2010). Negative frequency dependent selection (Takahata & Nei 1990) occurs when a new allele that arises by mutation (or an old allele that has reduced its frequency in a population due to an original high susceptibility to infection has a selective advantage because the prevalent pathogens have not evolved the ability to infect host cells with this allele. Finally, the fluctuating selection hypothesis proposes that spatial and

temporal variation in the presence and/or abundance of pathogens due to environmental variation may play an important role in maintaining high variation at the MHC genes (Hill 1991; Hedrick 2002). Fluctuating selection leads to spatio-temporal differences in the intensity of selection at MHC genes, e.g. an individual carrying an MHC allele (or alleles) could be fitter in a certain point in time and space, but less fit in another point in time and space. This results in a high variation at MHC alleles across subpopulations. A theoretical study has shown that temporal variation in pathogen resistance can by itself be a strong mechanism for the maintenance of MHC polymorphism (Hedrick 2002). It is difficult to disentangle the effects of the three mechanisms of PMS in nature. For example, rare alleles would be overrepresented in heterozygous individuals because of their low frequency, and it would be difficult to assess whether the individual is resistant for carrying a rare allele or because it is heterozygous. Moreover, if not all host populations are sampled across spatial and temporal scales, it is difficult to find empirical evidence for fluctuating selection. Detailed molecular analyses to test for correlation between pathogens and heterozygosity or specific alleles are required to determine which mechanism or combination of mechanisms is operating (Schad *et al.* 2004; Meyer-Lucht & Sommer 2005; Lenz *et al.* 2009b). In order to explore the mechanisms of PMS several studies have investigated the population genetic structure on MHC genes (Landry & Bernatchez 2001; Aguilar & Garza 2006; Alcaide *et al.* 2008; Loiseau *et al.* 2009; Evans *et al.* 2010; Cammen *et al.* 2011). However these studies have been done at coarse spatial scales across very discrete populations. To date, we are lacking studies that explicitly assess the population genetic structure on the MHC variation at fine spatial scales, specifically at the individual host level. Such studies would give insight into the nature of the selection operating at the MHC and the role that environmental heterogeneity can have in this process.

1.6 The role of the environment in pathogen-host interactions and PMS

The way that mechanisms of PMS generate balancing selection is highly dependent on the specific interactions between hosts and pathogen in the context of the environment (Steinhaus 1960). The spatio-temporal distribution of species is influenced by biotic and abiotic factors, and this has important evolutionary implications, since evolutionary forces act differently in spatially structured populations and over different time frames (Whitlock 2004; Freedman *et al.* 2010). Thus, taking into account the quantitative and qualitative characteristics of a landscape in population genetics is important (Holderegger *et al.* 2006),

especially for understanding how selection operates in a spatially explicit manner. The way in which the characteristics of the environment play a role in the interaction between hosts and pathogens is complex and can occur at different levels. For example, one level at which these characteristics have an effect is on the distribution of the species. Only when hosts and pathogens coexist spatially and temporally can the interaction occur.

Once pathogen and host come into contact, the environment has further effects. One host-pathogen interaction can have different outcomes in different habitat types or different environments (Biek & Real 2010), and infectious disease patterns can be affected by heterogeneous features of habitat and environment associated with different spatial scales (Archie *et al.* 2009; States *et al.* 2009). For example, if a habitat is poor in resources for the host, then infection could be more likely, perhaps because less resources are allocated by the host to immune defences (Seppälä & Jokela 2010), or because primary defences might fail. In line with this, strong differentiation in host resistance that correlates with different habitat types has been reported, even when both host and pathogens coexist in these habitats (Laine *et al.* 2011). Crucial to infection and also dependent on the environment characteristics are the way pathogens are transmitted from one host to another, and the way hosts disperse in and between populations (Biek & Real 2010). Spatial patterns of these processes result in heterogeneous disease distribution across different scales (Kleindorfer & Dudaniec 2006; Wood *et al.* 2007).

Temporal variation is also important in shaping host-pathogen interactions, thus, the effects of selection can vary temporally. Key ecological factors influencing the population dynamics of hosts and parasites can vary in ecological time scales, or even in smaller time scales (from one year to another, or seasonally), generating variation in both host immunocompetence and abundance, and pathogen distribution and virulence (Roulin *et al.* 2007). Seasonal variation in pathogen load can be due to variation in the host immune function and host exposure to infective stages of the pathogens; this occurs because the pathogen can be sensitive to seasonal variation in a specific climatic variable, for example, humidity or temperature (Wilson *et al.* 2003). Temporal variation in pathogen presence in a host population has been reported in several studies (Montgomery & Montgomery 1989; Oesterholt *et al.* 2006; Xiao *et al.* 2010). In spite of the large amount of studies that show a relationship between environmental

variation and disease distribution, to my knowledge no study to date has linked this to immune gene variation within a single host population.

The discipline of landscape genetics, which aims to link landscape ecology with population genetics, has been very useful in understanding the interactions that occur between hosts and pathogens (Manel *et al.* 2003; Holderegger *et al.* 2006; Biek & Real 2010). Landscape genetic data collected at fine spatial scales contributes to our understanding of the microevolutionary processes that generate local adaptation and thus genetic structure within populations (Manel *et al.* 2003; Richardson *et al.* 2014). Although some studies have used this approach to link genetic data with landscape and environmental variables (reviewed in Manel *et al.* 2003; Sork & Waits 2010), the great majority of these have used neutral genetic markers or, alternatively, genome-wide sequence data to identify regions of the genome as candidates of adaptive evolution. Correlations between genetic variation at known adaptive loci and fine-scale environmental characteristics have been less well studied (Freedman *et al.* 2010; Sork & Waits 2010), even though the importance of these types of studies has been highlighted (Holderegger *et al.* 2006; Holderegger & Wagner 2008). The interaction between landscapes and microevolutionary processes has recently been included in analyses of host-pathogen interactions. In a recent review Biek & Real (2010) listed a number of studies that used the landscape genetics approach to analyse interactions between pathogen genetic structure, host genetic structure and environmental variables. The authors point out that disease emergence and spread can be predicted with host (neutral) genetic structure data, but they overlook the fact that structure at adaptive loci (specifically immune gene variation) could be more informative in this regard. Variation in host immune genes can influence the way a pathogen is transmitted within and among populations, because not all individuals are equally susceptible to infection by the pathogen (Wilson *et al.* 2003). More studies are needed in this area if we are to understand how environmental variables and geographic barriers are related to both the distribution of pathogens and of immune gene variants in a population of hosts.

The key to identifying environmental variables that drive the distribution of pathogens, and potentially host immune gene variation, lies in studying the pathogens across different spatial and temporal scales. Advances in Geographic Information Systems (GIS) have provided a powerful toolkit to use in landscape genetics (Kozak *et al.* 2008). GIS technology has been useful to map the distribution of diseases worldwide, mostly in relation to human diseases

(Baptista-Rosas *et al.* 2007; Aikembayev *et al.* 2010; Xiao *et al.* 2010). It has also been useful to detect disease foci or areas that are suitable for pathogens (Cortinas *et al.* 2002; Aikembayev *et al.* 2010; Xiao *et al.* 2010). However, although the means to assess how the environment can affect the way PMS acts on immune genes are widely available, no studies have done this at the individual host level. Assessment of immune gene variation at the individual host level in relation to the pathogens to which they are exposed, and the environmental variables that determine those host-pathogen interactions, is still needed. Studies of this nature promise to provide important insight into the way PMS acts on immune genes.

1.7 Berthelot's pipit populations as a study system

Oceanic islands provide excellent systems for testing evolutionary biology hypotheses, because of a number of key features (Emerson 2002): First, they represent discrete geographical entities with known boundaries. Second, species inhabiting islands are generally isolated with reduced gene flow among neighbouring islands. Third, their generally small size and lower biodiversity makes it easier to account for the species inhabiting it (including pathogens), also the ecological relationships among the species are less complex than in continental systems, all of which makes analyses more tractable. Fourth, despite their small size, islands can still contain a diversity of habitats (especially mountainous ones), making it possible to test effects of environmental variation on evolutionary processes. Consequently, populations of species that inhabit islands are often the best systems we have in which to explore complex evolutionary questions within a natural context (Emerson 2002).

Berthelot's pipit (*Anthus berthelotii*; Fig. 1.3) is a small, insectivorous passerine endemic to the north Atlantic archipelagos of Madeira, Selvagens and the Canary Islands where it occurs in relatively isolated populations of varying size (Cramp & Perrins 1977; Fig. 1.4). The species is split into two subspecies: *A. berthelotii berthelotii* on Selvagens and the Canary Islands, and *A. berthelotii madeiriensis* on Madeira (Clarke 2006; but see Illera *et al.* 2007). The pipit is one of the most abundant passerines throughout its range; it occurs in all habitats except for thick woodland and humid areas, from sea level to 2500 metres above sea level (Illera 2007). The breeding population of Berthelot's pipit is estimated to range between 20,000 to 100,000 breeding pairs across its range (Birdlife International 2004). Berthelot's pipits have a generation time of approximately three years; their breeding season spans from February to

August, they build the nest on the ground and clutch size ranges from two to four eggs (Garcia-Del-Rey & Cresswell 2007).

Figure 1.3 A Berthelot's pipit (*Anthus berthelotii*) on top of a clap net baited with a *Tenebrio molitor* larva. Photo by Karl P. Phillips.

The tawny pipit (*Anthus campestris*), the sister species of the Berthelot's pipit, is widespread across Europe, Asia and Northwest Africa. The two species diverged approximately 2.5 million years ago (Arctander *et al.* 1996; Voelker 1999). Berthelot's pipit is thought to have colonized the Macaronesian archipelagos from Africa during the Pleistocene thus causing the evolutionary split from the tawny pipit. Previous work on Berthelot's pipit found that this species has very low levels of genetic diversity at mitochondrial DNA, with only one and four haplotypes in the Control region and cytochrome b gene, respectively, across all populations (Illera *et al.* 2007). These data suggest a small founding population of pipits in these islands. There is also evidence that the Madeira and Selvagens archipelagos were colonised independently from the Canary Islands between 1,000 to 26,000 years ago (Spurgin *et al.* 2014). Furthermore, population bottlenecks of differing severity, which occurred during the colonization of each archipelago, have been dominant in shaping neutral genetic variation across the populations (Illera *et al.* 2007; Spurgin *et al.* 2014). There is absence of gene flow

among archipelagos of Berthelot's pipits, and the populations form three genetic clusters which correspond to the three archipelagos (Spurgin *et al.* 2014).

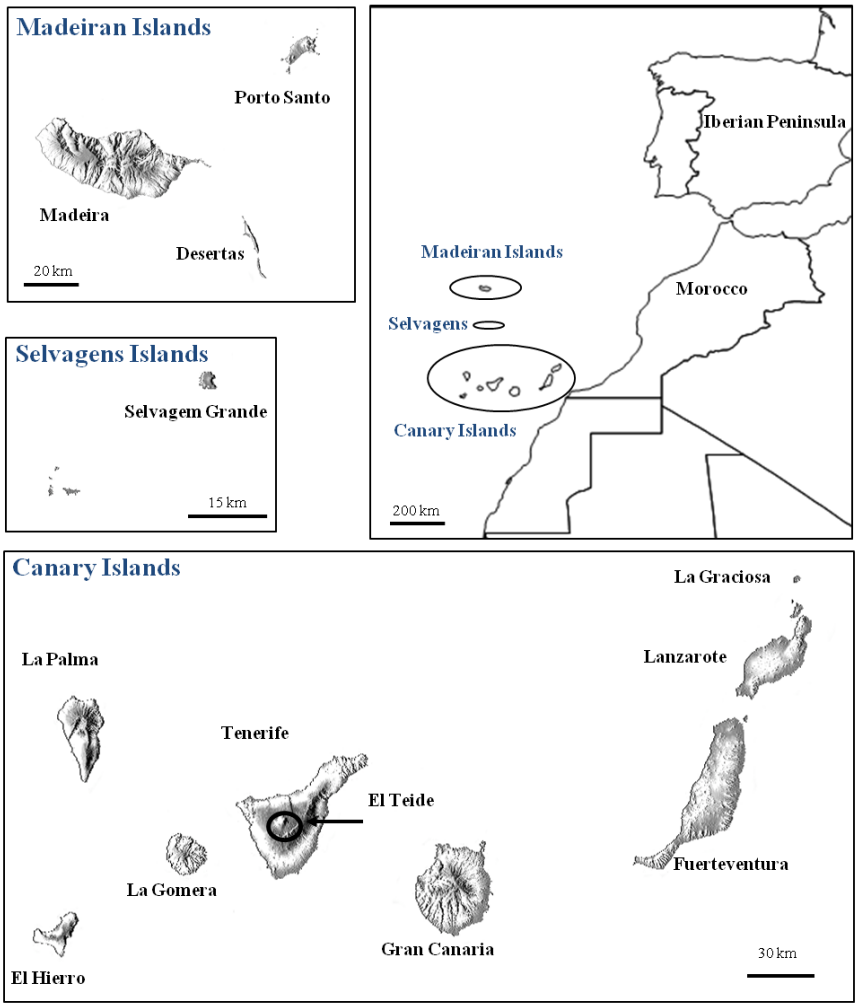


Figure 1.4 Distribution of Berthelot's pipit populations in the Macaronesian archipelagos in the North Atlantic. For this study thirteen populations from twelve islands were sampled. In Tenerife two populations were sampled (the coastal population and the volcano El Teide).

The pathogens that infect Berthelot's pipits have been studied previously. Illera *et al.* (2008) examined the prevalence of avian pox (*Poxvirus avium*) and avian malaria (*Haemoproteus* and *Plasmodium* spp) across the 12 main Berthelot's pipit populations. Overall, 8% of individuals showed evidence of pox lesions and 16% were infected with avian malaria. Marked differences in the prevalence of parasites among islands both within and between archipelagos were observed. The diversity of pathogens detected was low: only two genetic lineages of avian malaria and one lineage of avian pox were found to infect the pipit throughout its range. Interestingly, both avian malaria parasites found were *Plasmodium* lineages that had not been previously reported in the Macaronesian avifauna (but that had been observed in the lesser kestrel, *Falco naumannii*). The avian pox lineage found is a host-specific lineage that had previously been reported in two of the Canary Islands (Illera *et al.* 2008). Furthermore, another study showed that the spatially varying pathogen pressures that the populations of Berthelot's pipits are subject to are consistent over time and that considerable spatial variation in the distribution of avian malaria and avian pox exists within one of the populations, Tenerife (Spurgin *et al.* 2012).

Population-level variation at the MHC has also been characterised in Berthelot's pipits by genotyping the MHC class I in all the populations (Spurgin *et al.* 2011). This study reported a total of 41 MHC class I exon 3 alleles across the populations, with 9 – 14 alleles per population. It inferred that only a limited number of MHC class I alleles persisted when the species dispersed across the archipelagos, but that functional MHC variation was quickly restored by gene conversion. These results suggest that MHC genes are rapidly evolving across the populations of Berthelot's pipit as a result of gene conversion and positive selection (Spurgin *et al.* 2011).

The islands Berthelot's pipits inhabit have differing environmental characteristics and some populations, such as that of Tenerife, exist within very environmentally heterogeneous territories. Because of their recent colonization, isolation and consequent genetic differentiation, these populations provide an ideal study system to test the relative role of selection verses other evolutionary forces in shaping variation at functional loci at different spatial scales. Furthermore, the existence of a limited diversity of pathogens within and among the populations makes the assessment of pathogen-mediated selection on immune genes more tractable. The presence of multiple discrete populations enables assessment of patterns

of immune gene variation at large scales (across populations) while the large, environmentally heterogenous population on Tenerife make this population a suitable study system in which to assess patterns of selection at immune genes at fine spatial scales.

1.8 Aims of the thesis

In this thesis, I investigate selection at two sets of immune loci at different spatial scales. In chapter 2, I assess genetic variation at a set of innate immunity loci, the toll-like receptors (TLR) across the populations of Berthelot's pipit. I also compare the levels of diversity at TLRs in Berthelot's pipit with those of its sister species the tawny pipit, in order to investigate how these loci have evolved after the colonization of the archipelagos. In chapter 3 I investigate the distribution of avian malaria within the population of Tenerife, assess malaria risk at small spatial scales and determine the role of key anthropogenic and natural factors in predicting that risk. In chapter 4, I use 454 next generation sequencing to efficiently genotype major histocompatibility complex (MHC) class I variation in a large cohort of individuals from the Tenerife population, and use the data to assess signatures of selection at these loci. Finally, in chapter 5 I explore whether structure exist at neutral and MHC class I loci in the Tenerife population of pipits. I then assess the potential role of environmental factors, including malaria, in shaping the distribution of MHC alleles at fine scales within this population. Finally in chapter 6 I bring together the findings described in the previous chapters, discuss their overall significance and propose some ideas for future research.

1.9 References

Aguilar A, Garza JC (2006) A comparison of variability and population structure for major histocompatibility complex and microsatellite loci in California coastal steelhead (*Oncorhynchus mykiss walbaum*). *Molecular Ecology* **15**, 923-937.

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

Aikembayev AM, Lukhnova L, Temiraliyeva G*, et al.* (2010) Historical distribution and molecular diversity of *Bacillus anthracis*, Kazakhstan. *Emerging Infectious Diseases* **16**, 789-796.

Akira S (2003) Toll-like receptor signaling. *Journal of Biological Chemistry* **278**, 38105-38108.

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

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

Altizer S, Harvell D, Friedle E (2003) Rapid evolutionary dynamics and disease threats to biodiversity. *Trends in Ecology & Evolution* **18**, 589-596.

Anderson RM, May RM (1978) Regulation and stability of host-parasite population interactions 1. Regulatory processes. *Journal of Animal Ecology* **47**, 219-247.

Anderson RM, May RM (1979) Population biology of infectious-diseases. *Nature* **280**, 361-367.

Antonovics J (1993) The interplay of numerical and gene-frequency dynamics in host-pathogen systems. In: *Ecological Genetics* (ed. Leslie AR). Princeton University Press, Princeton.

Archie EA, Luikart G, Ezenwa VO (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends in Ecology & Evolution* **24**, 21-30.

Arctander P, Folmer O, Fjeldsa J (1996) The phylogenetic relationships of Berthelot's pipit *Anthus berthelotii* illustrated by DNA sequence data, with remarks on the genetic distance between rock and water pipits *Anthus spinoletta*. *Ibis* **138**, 263-272.

Areal H, Abrantes J, Esteves PJ (2011) Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *Bmc Evolutionary Biology* **11**, 368.

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

Bach FH (1976) Genetics of transplantation - Major Histocompatibility Complex. *Annual Review of Genetics* **10**, 319-339.

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

Baptista-Rosas RC, Hinojosa A, Riquelme M (2007) Ecological niche modeling of *Coccidioides* spp. in Western North American deserts. In: *Coccidioidomycosis: Sixth International Symposium* (eds. Clemons KV, LaniadoLaborin R, Stevens DA), pp. 35-46.

Basu M, Maji AK, Chakraborty A*, et al.* (2010) Genetic association of Toll-like-receptor 4 and tumor necrosis factor-alpha polymorphisms with *Plasmodium falciparum* blood infection levels. *Infection Genetics and Evolution* **10**, 686-696.

Behringer DC, Butler MJ, Shields JD (2006) Avoidance of disease by social lobsters. *Nature* **441**, 421-421.

Belvin MP, Anderson KV (1996) A conserved signaling pathway: The *Drosophila* Toll-Dorsal pathway. *Annual Review of Cell and Developmental Biology* **12**, 393-416.

Berg J, Tymoczko J, Stryer L, Gatto G (2012) *Biochemistry*, Seventh edn. WH Freeman and Company, New York.

Biek R, Real LA (2010) The landscape genetics of infectious disease emergence and spread. *Molecular Ecology* **19**, 3515-3531.

Birdlife International (2004) *Birds in Europe: population estimates, trends and conservation status* Birdlife International, Cambridge, U.K.

Blaustein AR, Gervasi SS, Johnson PTJ*, et al.* (2012) Ecophysiology meets conservation: understanding the role of disease in amphibian population declines. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**, 1688-1707.

Bodmer JG, Marsh SGE, Albert ED*, et al.* (1997) Nomenclature for factors of the HLA system, 1996. *Tissue Antigens* **49**, 297-321.

Boehne A, Brunet F, Galiana-Arnoux D, Schultheis C, Volff J-N (2008) Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Research* **16**, 203-215.

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

Bonneaud C, Sorci G, Morin V*, et al.* (2004) Diversity of MHC class I and IIB genes in house sparrows (*Passer domesticus*). *Immunogenetics* **55**, 855-865.

Bridges CB (1936) The Bar "gene" a duplication. *Science* **83**, 210-211.

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**, 3444-3455.

Brown JL (1998) The new heterozygosity theory of mate choice and the MHC. *Genetica* **104**, 215-221.

Buchanan SGS, Gay NJ (1996) Structural and functional diversity in the leucine rich repeat family of proteins. *Progress in Biophysics & Molecular Biology* **65**, 1-44.

Cammen K, Hoffman JI, Knapp LA, Harwood J, Amos W (2011) Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (*Halichoerus grypus*). *Molecular Ecology* **20**, 740-752.

Chave J (2013) The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecology Letters* **16**, 4-16.

Chaves LD, Krueth SB, Reed KM (2009) Defining the turkey MHC: Sequence and genes of the B locus. *Journal of Immunology* **183**, 6530-6537.

Chen J-M, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP (2007) Gene conversion: mechanisms, evolution and human disease. *Nature Reviews Genetics* **8**, 762-775.

Chen JSC, Wang TY, Tzeng TD, Wang CY, Wang D (2008) Evidence for positive selection in the TLR9 gene of teleosts. *Fish & Shellfish Immunology* **24**, 234-242.

Clarke T (2006) *Birds of the Atlantic islands* Helm, London.

Cortinas MR, Guerra MA, Jones CJ, Kitron U (2002) Detection, characterization, and prediction of tick-borne disease foci. *International Journal of Medical Microbiology* **291**, 11-20.

Cramp S, Perrins CM (1977-1994) *Handbook of the birds of Europe, the Middle East and Africa. The birds of the western Palearctic* Oxford University Press, Oxford.

Danilova N (2006) The evolution of immune mechanisms. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* **306B**, 496-520.

Darwin C (1859) *On The Origin of Species by Means of Natural Selection* John Murray, London.

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

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

Drake BM, Goto RM, Miller MM, Gee GF, Briles WE (1999) Molecular and immunogenetic analysis of major histocompatibility haplotypes in northern bobwhite enable direct identification of corresponding haplotypes in an endangered subspecies, the masked bobwhite. *Zoo Biology* **18**, 279-294.

Duncan AB, Little TJ (2007) Parasite-driven genetic change in a natural population of *Daphnia*. *Evolution* **61**, 796-803.

Ebert D (2008) Host-parasite coevolution: Insights from the *Daphnia*-parasite model system. *Current Opinion in Microbiology* **11**, 290-301.

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

Emerson BC (2002) Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology* **11**, 951-966.

Evans ML, Neff BD (2009) Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology* **18**, 4716-4729.

Evans ML, Neff BD, Heath DD (2010) MHC genetic structure and divergence across populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Heredity* **104**, 449-459.

Farnell MB, Crippen TL, He HQ, Swaggerty CL, Kogut MH (2003) Oxidative burst mediated by toll like receptors (TLR) and CD14 on avian heterophils stimulated with bacterial toll agonists. *Developmental and Comparative Immunology* **27**, 423-429.

Fisher RA (1930) *The genetical theory of natural selection* Clarendon Press, Oxford.

Frank SA (2002) *Immunology and evolution of infectious disease* Princeton University Press, Princeton.

Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology* **19**, 3773-3788.

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

Fumagalli M, Sironi M, Pozzoli U*, et al.* (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *Plos Genetics* **7**, e1002355.

Garcia-Del-Rey E, Cresswell W (2007) The breeding biology of the endemic Berthelot's Pipit, *Anthus berthelotii*, in a harsh oceanic island environment (Tenerife, Canary Islands). *Ostrich* **78**, 583-589.

Gay NJ, Keith FJ (1991) *Drosophila* Toll and IL-1 receptor. *Nature* **351**, 355-356.

Georgiev GP (1984) Mobile genetic elements in animal cells and their biological significance. *European Journal of Biochemistry* **145**, 203-220.

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

Haldane JBS (1949) Disease and evolution *La Ricerca Scientifica Suplemento* **19**, 1-11.

Hawley DM, Altizer SM (2011) Disease ecology meets ecological immunology: understanding the links between organismal immunity and infection dynamics in natural populations. *Functional Ecology* **25**, 48-60.

Hedrick PW (1994) Evolutionary genetics of the major histocompatibility complex. *American Naturalist* **143**, 945-964.

Hedrick PW (1998) Balancing selection and MHC. *Genetica* **104**, 207-214.

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

Hedrick PW, Thomson G (1983) Evidence for balancing selection at HLA. *Genetics* **104**, 449-456.

Higuchi M, Matsuo A, Shingai M*, et al.* (2008) Combinational recognition of bacterial lipoproteins and peptidoglycan by chicken Toll-like receptor 2 subfamily. *Developmental and Comparative Immunology* **32**, 147-155.

Hill AS (1991) HLA associations with malaria in Africa: some implications for MHC evolution. In: *Molecular Evolution of the Major Histocompatibility Complex* (eds. Klein J, Klein D), pp. 403-420. Springer Berlin Heidelberg.

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

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology* **21**, 797-807.

Holderegger R, Wagner HH (2008) Landscape genetics. *Bioscience* **58**, 199-207.

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci - evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 958-962.

Illera JC (2007) Bisbita Caminero *Anthus berthelotii*. In: *Atlas de las aves nidificantes en el archipiélago Canario (1997-2003)* (ed. JA L), pp. 344-347. General de Conservación de la Naturaleza-Sociedad Española de Ornitología, Madrid.

Illera JC, Emerson BC, Richardson DS (2007) Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology* **16**, 4599-4612.

Illera JC, Emerson BC, Richardson DS (2008) Genetic characterization, distribution and prevalence of avian pox and avian malaria in the Berthelot's pipit (*Anthus berthelotii*) in Macaronesia. *Parasitology Research* **103**, 1435-1443.

Ip YT, Levine M (1994) Molecular genetics of *Drosophila* immunity. *Current Opinion in Genetics and Development* **4**, 672-677.

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**, 2157-2168.

Karpala AJ, Lowenthal JW, Bean AG (2008) Activation of the TLR3 pathway regulates IFN beta production in chickens. *Developmental and Comparative Immunology* **32**, 435-444.

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

Keestra AM, de Zoete MR, Bouwman LI, van Putten JPM (2010) Chicken TLR21 is an innate CpG DNA receptor distinct from mammalian TLR9. *The Journal of Immunology* **185**, 460-467.

Kimura M (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetical Research* **11**, 247-269.

Kimura M (1977) Preponderance of synonymous changes as evidence for neutral theory of molecular evolution. *Nature* **267**, 275-276.

Kimura M (1991) The neutral theory of molecular evolution - a review of recent evidence. *Japanese Journal of Genetics* **66**, 367-386.

Klein J, Sato A (1998) Birth of the major histocompatibility complex. *Scandinavian Journal of Immunology* **47**, 199-209.

Kleindorfer S, Dudaniec RY (2006) Increasing prevalence of avian poxvirus in Darwin's finches and its effect on male pairing success. *Journal of Avian Biology* **37**, 69-76.

Kozak KH, Graham CH, Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution* **23**, 141-148.

Laine A-L, Burdon JJ, Dodds PN, Thrall PH (2011) Spatial variation in disease resistance: from molecules to metapopulations. *Journal of Ecology* **99**, 96-112.

Landry C, Bernatchez L (2001) Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). *Molecular Ecology* **10**, 2525-2539.

Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA (1996) The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973-983.

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

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

Leulier F, Lemaitre B (2008) Toll-like receptors - taking an evolutionary approach. *Nature Reviews Genetics* **9**, 165-178.

Levin SA (1992) The problem of pattern and scale in ecology. *Ecology* **73**, 1943-1967.

Lien E, Sellati TJ, Yoshimura A*, et al.* (1999) Toll-like receptor 2 functions as a pattern recognition receptor for diverse bacterial products. *Journal of Biological Chemistry* **274**, 33419-33425.

Lion S, van Baalen M, Wilson WG (2006) The evolution of parasite manipulation of host dispersal. *Proceedings of the Royal Society B-Biological Sciences* **273**, 1063-1071.

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

Loeb LA, Springga.Cf, Battula N (1974) Errors in DNA-replication as a basis of malignant changes. *Cancer Research* **34**, 2311-2321.

Loehle C (1995) Social barriers to pathogen transmission in wild animal populations. *Ecology* **76**, 326-335.

Loewe L (2008) Negative selection. *Nature Education* **1**, 59.

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

Losos JB (1994) Integrative approaches to evolutionary ecology - Anolis lizards as model systems. *Annual Review of Ecology and Systematics* **25**, 467-493.

lqbal M, Philbin VJ, Withanage GSK*, et al.* (2005) Identification and functional characterization of chicken Toll-like receptor 5 reveals a fundamental role in the biology of infection with *Salmonella enterica serovar typhimurium*. *Infection and Immunity* **73**, 2344-2350.

Malkova A, Haber JE (2012) Mutations arising during repair of chromosome breaks. *Annual Review of Genetics* **46**, 455-473.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

Medzhitov R, Janeway CA (1997) Innate immunity: The virtues of a nonclonal system of recognition. *Cell* **91**, 295-298.

Medzhitov R, Preston-Hurlburt P, Janeway CA (1997) A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394-397.

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

Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N (2012) Molecular evolution of vertebrate Toll-like receptors: Evolutionary rate difference between their leucine-rich repeats and their TIR domains. *Gene* **503**, 235-243.

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

Minchella DJ (1985) Host life-history variation in response to parasitism. *Parasitology* **90**, 205-216.

Misch EA, Hawn TR (2008) Toll-like receptor polymorphisms and susceptibility to human disease. *Clinical Science* **114**, 347-360.

Møller AP (1997) Parasitism and the evolution of host life history. In: *Host-parasite evolution, general principles and avian models* (eds. Clayton DH, Moore J). Oxford University Press, Oxford.

Mona S, Crestanello B, Bankhead-Dronnet S*, et al.* (2008) Disentangling the effects of recombination, selection, and demography on the genetic variation at a major

histocompatibility complex class II gene in the alpine chamois. *Molecular Ecology* **17**, 4053-4067.

Montgomery SSJ, Montgomery WI (1989) Spatial and temporal variation in the infracommunity structure of helminths of *Apodemus sylvaticus* (rodentia, muridae). *Parasitology* **98**, 145-150.

Nakajima T, Ohtani H, Satta Y*, et al.* (2008) Natural selection in the TLR-related genes in the course of primate evolution. *Immunogenetics* **60**, 727-735.

Noble ER, Noble GA, Schad GA, McInnes AJ (1989) *Parasitology: The biology of animal parasites*, Sixth edn. Lea and Febiger, London.

Oesterholt MJAM, Bousema JT, Mwerinde OK*, et al.* (2006) Spatial and temporal variation in malaria transmission in a low endemicity area in northern Tanzania. *Malaria Journal* **5**, 98.

Ohta T (1995) Gene conversion vs. point mutation in generating variability at the antigen recognition site of major histocompatibility complex loci. *Journal of Molecular Evolution* **41**, 115-119.

Parham P (2009) *The immune system*, Third edn. Garland Science, New York.

Parker BJ, Elderd BD, Dwyer G (2010) Host behaviour and exposure risk in an insect-pathogen interaction. *Journal of Animal Ecology* **79**, 863-870.

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

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

Pianka ER (2011) *Evolutionary ecology*, Seventh edn. Eric R. Pianka.

Poltorak A, Smirnova I, He XL*, et al.* (1998) Genetic and physical mapping of the LPS locus: Identification of the Toll-4 receptor as a candidate gene in the critical region. *Blood Cells Molecules and Diseases* **24**, 340-355.

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

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

Price PW (1980) Evolutionary biology of parasites. *Monographs in population biology* **15**, 1-237.

Prugnolle F, Manica A, Charpentier M*, et al.* (2005) Pathogen-driven selection and worldwide HLA class I diversity. *Current Biology* **15**, 1022-1027.

Radwan J, Tkacz A, Kloch A (2008) MHC and preferences for male odour in the bank vole. *Ethology* **114**, 827-833.

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

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

Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution* **29**, 165-176.

Riordan P, Hudson P, Albon S (2007) Do parasites matter? Infectious diseases and the conservation of host populations. In: *Key topics in conservation biology* (eds. Macdonald DW, Service K). Blackwell, Oxford.

Roach JC, Glusman G, Rowen L*, et al.* (2005) The evolution of vertebrate Toll-like receptors. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9577-9582.

Roitt IM, Brostoff J, Male DK (2001) *Immunology* Mosby, Philadelphia.

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

Schad J, Sommer S, Ganzhorn JU (2004) MHC variability of a small lemur in the littoral forest fragments of southeastern Madagascar. *Conservation Genetics* **5**, 299-309.

Schmid-Hempel P (2003) Variation in immune defence as a question of evolutionary ecology. *Proceedings of the Royal Society B-Biological Sciences* **270**, 357-366.

Schnare M, Barton GM, Holt AC*, et al.* (2001) Toll-like receptors control activation of adaptive immune responses. *Nature Immunology* **2**, 947-950.

Schröder NW, Schumann RR (2005) Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *Lancet Infectious Diseases* **5**, 156-164.

Seppälä O, Jokela J (2010) Maintenance of genetic variation in immune defence of a freshwater snail: role of environmental heterogeneity. *Evolution* **64**, 2397-2407.

Shiina T, Shimizu S, Hosomichi K*, et al.* (2004) Comparative genomic analysis of two avian (quail and chicken) MHC regions. *Journal of Immunology* **172**, 6751-6763.

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

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.

Slightom JL, Blechl AE, Smithies O (1980) Human fetal Gγ-globin and Aγ-globin genes - complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes *Cell* **21**, 627-638.

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

Smith KF, Sax DF, Lafferty KD (2006) Evidence for the role of infectious disease in species extinction and endangerment. *Conservation Biology* **20**, 1349-1357.

Smith RN, Cain SL, Anderson SH, Dunk JR, Williams ES (1998) Blackfly-induced mortality of nestling red-tailed hawks. *The Auk* **115**, 368-375.

Sork VL, Waits L (2010) Contributions of landscape genetics - approaches, insights, and future potential. *Molecular Ecology* **19**, 3489-3495.

Spurgin LG, Illera JC, Jorgensen TH, Dawson DA, Richardson DS (2014) Genetic and phenotypic divergence in an island bird: isolation by distance, by colonisation or by adaptation? *Molecular Ecology* **23**, 1028-1039.

Spurgin LG, Illera JC, Padilla DP, Richardson DS (2012) Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia* **168**, 691-701.

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

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

States SL, Hochachka WM, Dhondt AA (2009) Spatial variation in an avian host community: Implications for disease dynamics. *Ecohealth* **6**, 540-545.

Steinhaus EA (1960) Symposium - Selected topics in microbial ecology 2. Importance of environmental factors in the insect-microbe ecosystem. *Bacteriological Reviews* **24**, 365-373.

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

Takeuchi O, Sato S, Horiuchi T*, et al.* (2002) Cutting edge: Role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *Journal of Immunology* **169**, 10-14.

Tompkins DM, Arneberg P, Begon ME*, et al.* (2002) Parasites and host population dynamics In: *The ecology of wildlife diseases* (ed. Hudson PJ), pp. 45-62. Oxford University Press, Oxford.

Tschirren B, Raberg L, Westerdahl H (2011) Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents. *Journal of Evolutionary Biology* **24**, 1232-1240.

Voelker G (1999) Dispersal, vicariance, and clocks: Historical biogeography and speciation in a cosmopolitan passerine genus (*Anthus*: motacillidae). *Evolution* **53**, 1536-1552.

Wakelin D (1984) *Immunity to Parasites: How animals control parasitic infections* Edward Arnold Publishers, London.

Wakelin D, Apanius V (1997) *Immune defence: genetic control* Oxford University Press, Oxford.

Wedekind C (1994) Mate choice and maternal selection for specific parasite resistances before, during and after fertilization. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **346**, 303-311.

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

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

Whitehouse HL (1970) Mechanism of genetic recombination. *Biological Reviews of the Cambridge Philosophical Society* **45**, 265-313.

Whitlock MC (2004) Selection and drift in metapopulations In: *Ecology, genetics and evolution of metapopulation* (eds. Hanski I, Gaggiotti OE). Elsevier Academic Press, London.

Wilson K, Bjørnstad ON, Dobson AP (2003) Heterogeneities in macroparasite infections: patterns and processes. In: *The ecology of wildlife diseases* (ed. Hudson P). Oxford University Press, Oxford.

Wittzell H, Madsen T, Westerdahl H, Shine R, von Schantz T (1998) MHC variation in birds and reptiles. *Genetica* **104**, 301-309.

Wood MJ, Cosgrove CL, Wilkin TA*, et al.* (2007) Within-population variation in prevalence and lineage distribution of avian malaria in blue tits, *Cyanistes caeruleus*. *Molecular Ecology* **16**, 3263-3273.

Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* **32**, 569-577.

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

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

Wright S (1948) On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution* **2**, 279-294.

Wynne JW, Cook MT, Holmes BH, Elliott NG (2007) Allelic and haplotypic diversity at the major histocompatibility class II within domesticated Australian Atlantic salmon (*Salmo salar L.*). *Journal of Fish Biology* **70**, 45-59.

Xiao D, Long Y, Wang S*, et al.* (2010) Spatiotemporal distribution of malaria and the association between its epidemic and climate factors in Hainan, China. *Malaria Journal* **9**, 185.

Yoneyama M, Fujita T (2010) Recognition of viral nucleic acids in innate immunity. *Reviews in Medical Virology* **20**, 4-22.

York IA, Rock KL (1996) Antigen processing and presentation by the class I major histocompatibility complex. *Annual Review of Immunology* **14**, 369-396.

Zhang JZ (2003) Evolution by gene duplication: an update. *Trends in Ecology & Evolution* **18**, 292-298.

Chapter 2

The role of drift and selection in shaping variation at innate immune genes in oceanic island populations

The tawny pipit, *Anthus campestris*, Berthelot's pipit sister species. Photo by Karl Phillips

2.1 Abstract

Understanding the relative role of different evolutionary forces in shaping the level and distribution of functional genetic diversity among natural populations is a key issue in evolutionary and conservation biology. To do so accurately, genetic data must be analyzed in conjunction with an unambiguous understanding of the historical processes that have acted upon the populations. Here we assessed variation in the innate immunity toll-like receptor (TLR) loci within and among 13 island populations (grouped into three archipelagos) of Berthelot's pipit, *Anthus berthelotii*, for which population history has previously been ascertained. We also compared the variation observed with that found in its widespread sister species, the tawny pipit, *Anthus campestris.* Our results suggest that demographic history has played the major role in shaping patterns of TLR variation in Berthelot's pipit. Within each archipelago, diversity at each TLR locus corresponded closely with the severity of the bottleneck that occurred during colonization. Despite this, signatures of selection were found; at TLR4 one site showed evidence of positive selection and in one population - that with the highest levels of pathogen infection - two out of the only four haplotypes identified are new functional variants that appear to have arisen *in situ*. In TLR3 one codon also showed evidence of positive selection. Our study indicates that while founder effects have greatly reduced TLR variation in populations of Berthelot's pipit, balancing selection may have helped to maintain functional variation at some TLR loci. This study therefore suggests that TLR variation may be important within genetically depauperate, bottlenecked populations.

2.2 Introduction

Genetic variation provides the fundamental building blocks for evolution. Consequently, understanding the levels and distribution of functional genetic diversity among individuals and populations, and what forces drive these patterns, is a central component of evolutionary biology. Given that genetic variation is critical to the adaptive potential of populations and species, this understanding also has important implications for conservation (Frankham *et al.* 1999).

Population bottlenecks result in losses of functional genetic diversity (Cabe 1998; Gautschi *et al.* 2002; Sutton *et al.* 2013), and are therefore a key force in shaping the future evolution and persistence of populations (Frankham 1996; England *et al.* 2003). When populations undergo bottlenecks genetic drift is usually the dominant force, reducing genetic diversity within populations and driving differentiation across them (Hartl & Clark 2007). However, if strong natural selection acts on a specific region of the genome it can either counteract, or alternatively, reinforce the effects of drift (Aguilar *et al.* 2004; Miller & Lambert 2004). If balancing selection operates, genetic diversity will be maintained within the bottlenecked populations, reducing the amount of differentiation one might expect under drift alone (Hedrick & Thomson 1983; Hughes & Nei 1988) at least at the specific loci involved. Alternatively, under purifying or constant directional selection, genetic diversity will be reduced and the effects of drift and the resulting population differentiation will be reinforced (Jiggins & Hurst 2003; Winternitz & Wares 2013).

When investigating genetic variation, loci involved in the immune system are of particular interest, not least because of their obvious importance for individual and population survival (reviewed in Sommer 2005; Acevedo-Whitehouse & Cunningham 2006), but also because they are expected to be under strong and direct selection from pathogens (Trowsdale & Parham 2004; Ekblom *et al.* 2010). Over the last few decades, genes of the major histocompatibility complex (MHC), which code for receptors central to the acquired immune system, have been the focus of studies into functional genetic diversity and pathogen-mediated selection among wild populations (reviewed in Piertney & Oliver 2006; Spurgin & Richardson 2010). Only recently has attention spread to investigating other immune genes (Acevedo-Whitehouse & Cunningham 2006; Jensen *et al.* 2008; Grueber *et al.* 2012; Turner *et al.* 2012), and, as yet, the roles of selection and drift in shaping among-population variation at these genes remains poorly understood (but see Bollmer *et al.* 2011; Tschirren *et al.* 2012; Grueber *et al.* 2013).

Toll-like receptors (TLRs) are essential components of the innate immune response in all vertebrates (Roach *et al.* 2005). They bind pathogen associated molecular patterns (PAMPs), thus triggering an intracellular signaling cascade that results in an inflammatory response and activation of macrophages, which attack the infection (Belvin & Anderson 1996; Akira 2003). Vertebrate TLRs are divided into six families that vary in the type of PAMPs they recognize (Roach *et al.* 2005). For example, TLR1LA and TLR1LB bind bacterial lipoproteins (Takeuchi *et al.* 2002), TLR3 binds viral RNA (Yoneyama & Fujita 2010), TLR21 binds bacterial DNA motifs (Keestra *et al.* 2010) and TLR4 binds bacterial lipopolysacharide (Poltorak *et al.* 1998). Recent studies have shown evidence of positive selection in TLRs in a range of organisms including birds (Downing *et al.* 2010; Alcaide & Edwards 2011; Grueber *et al.* 2014), fish (Chen *et al.* 2008), and mammals (Nakajima *et al.* 2008; Areal *et al.* 2011; Tschirren *et al.* 2011). Most of the sites identified as being under positive selection in these studies are located in the TLR extracellular domain which directly binds the PAMPS (reviewed in Mikami *et al.* 2012), and specific polymorphisms within TLRs have been associated with differential disease resistance (Schröder & Schumann 2005; Misch & Hawn 2008; Basu *et al.* 2010). These data thus support the idea that pathogen-mediated selection plays a role in determining patterns of variation at TLR loci.

While the relationship between individual-level TLR variation and an organism's ability to resist infection is becoming clearer, the forces that drive TLR variation at the level of populations and species are relatively poorly understood. Studying population-level variation at these critical genes, especially in small bottlenecked populations and/or endangered species, is important from both an evolutionary and conservation perspective (Grueber *et al.* 2012). Although there have been a few studies of TLR variation across species (Nakajima *et al.* 2008; Wlasiuk & Nachman 2010; Areal *et al.* 2011; Mikami *et al.* 2012) to our knowledge, only one study has investigated TLR variation in a bottlenecked wild population (Grueber *et al.* 2012), and only one study has looked at how TLR variation is distributed across multiple natural populations of a single species (Tschirren *et al.* 2012).

Berthelot's pipit, *Anthus berthelotii*, is a small, insectivorous passerine endemic to the North Atlantic archipelagos of Madeira, Selvagens and the Canary Islands where it occurs on relatively isolated island populations of varying size (Cramp & Perrins 1977, Fig. 2.1). The species is thought to have colonized the Macaronesian archipelagos from Africa during the Pleistocene before the evolutionary split from its sister species, the tawny pipit, *Anthus*

campestris (Voelker 1999). Previous work on Berthelot's pipit has provided evidence of how population bottlenecks of differing severity, which occurred during the colonization of each archipelago, have been dominant in shaping neutral variation across the populations (Illera *et al.* 2007; Spurgin *et al.* 2014). These studies have shown that the populations of Berthelot's pipit display a high level of genetic structure across, but not within, archipelagos, and that genetic variation at microsatellites is highest in the Canary Islands and lowest in Selvagens, with intermediate levels of genetic diversity found in Madeira. Another study has shown that these populations are exposed to spatially varying pathogen pressures that are consistent over time (Spurgin *et al.* 2012), and that MHC genes are rapidly evolving across the populations as a result of gene conversion and positive selection (Spurgin *et al.* 2011). Thus, this species provides an ideal study system in which to investigate the roles of selection versus drift in determining patterns of functional genetic diversity in fragmented bottlenecked populations. With that as our overall aim, here we characterized variation at five TLR loci in Berthelot's pipit. Specifically we: (i) determined genetic variation at TLR loci in Berthelot's pipit and compared it to that found in other bird species, including its sister species, the tawny pipit, (ii) investigated how TLR genetic diversity is distributed within and among the 13 Berthelot's pipit populations and, (iii) assessed the relative roles of selection and drift in shaping the patterns of variation we observed in these important immune loci.

2.3 Methods

2.3.1 Study populations and sampling

We screened a minimum of five individuals from all islands sampled as part of earlier studies (Illera *et al.* 2007; Spurgin *et al.* 2012; Fig. 2.1). A 13th population was sampled from El Teide, a volcano situated at the center of Tenerife rising 3,700 m above sea level. The El Teide population exists at >2000 m above sea level (a.s.l.) across an alpine plateau (*ca.* 190 km^2), separated from the rest of the island population by its altitude and a ring of pine forest that extends from approximately 1600-2000 m.a.s.l., which is not suitable habitat for pipits (Illera 2007). Birds were captured using clap nets baited with *Tenebrio molitor* larvae. Blood samples were collected by brachial venipuncture, diluted in 800 µl absolute ethanol in screw-cap microcentrifuge tubes and stored at room temperature. We also obtained blood samples from 12 tawny pipits (two from Morocco, four from Mauritania and six from the Iberian Peninsula), captured using the same methods as for Berthelot's pipits.

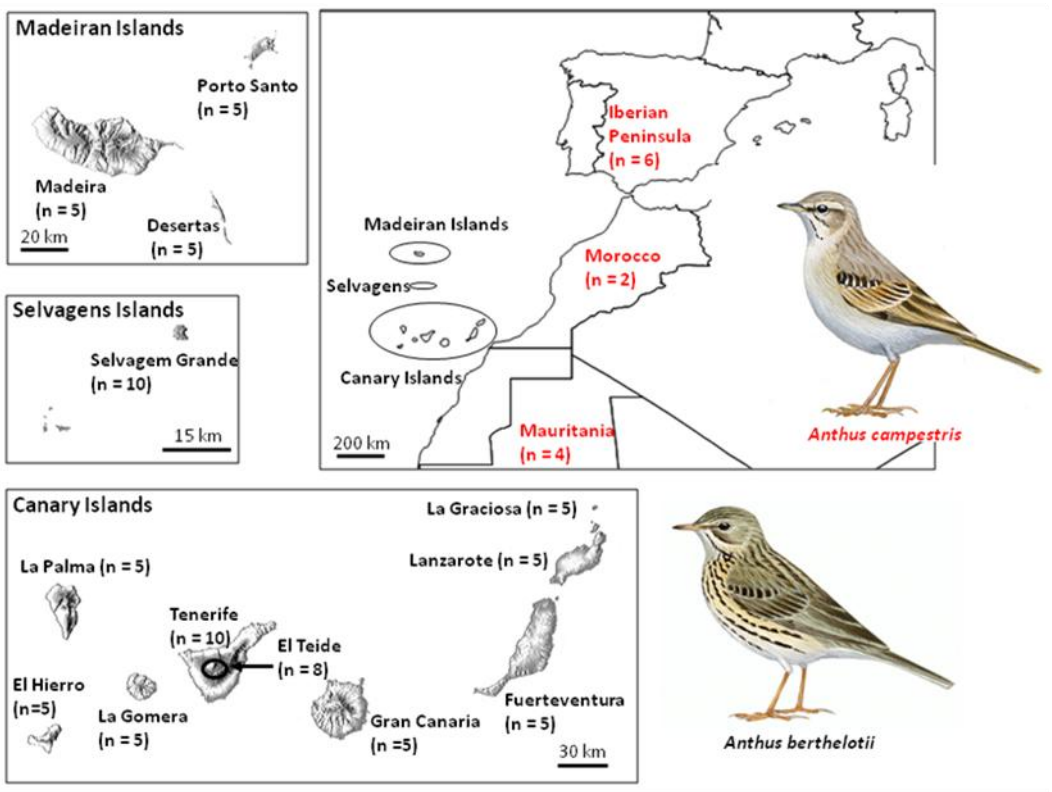


Figure 2.1 Distribution of Berthelot's pipit, *Anthus berthelotii*, across the Islands of the Macaronesian archipelagos, and populations of tawny pipits, *Anthus campestris*, used in this study. The identity of populations and the number of individuals sampled (in brackets) per population are shown.

2.3.2 Estimation of divergence time between Berthelot's and tawny pipits

Previously, the timing of divergence between Berthelot's and tawny pipits had been estimated using DNA from a tawny pipit from Denmark (Voelker 1999). However, a phylogenetic analysis suggests that the migratory behaviour of tawny pipits evolved after the split between Berthelot's and tawny pipits (Outlaw & Voelker 2006). Thus Berthelot's pipits likely evolved from an ancestral sedentary population of tawny pipits that existed near to the Canary Islands and, therefore, the divergence time might have been overestimated. We therefore re-estimated divergence time based on mitochondrial gene cytochrome oxidase subunit I (COI) sequence data. A 655 bp fragment of the COI gene was amplified in nine tawny pipits (four from north Africa, including two confirmed resident birds from Morocco, and five from the

Iberian Peninsula) and sixteen Berthelot's pipits (ten from the Canary Islands, three from Selvagens and three from Madeira) using primers PasserF1 and PasserR1 (Lohman *et al.* 2009), following thermal conditions described in Lohman *et al.* (2009). Amplification was carried out in 10 µl reactions using a Tetrad thermocycler (MJ Research), including TopTaq polymerase master mix (Qiagen), 0.5 µM of each primer and *ca.* 30 ng of DNA. Amplified fragments were visualized in 2% agarose gels stained with ethidium bromide and purified using a mixture of recombinant alkaline phosphatase and exonuclease I, incubating at 37°C for 30 mins, followed by enzyme inactivation at 95°C for 5 mins. Amplified fragments were sequenced using the BigDye terminator kit (Applied Biosystems) using the following thermal profile: 96°C for 2 mins, followed by 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 mins. Products were visualized in an ABI genetic analyzer (Applied Biosystems). Sequences were aligned and edited in BioEdit 7.0.9.0 (Hall 1999). For divergence time estimates we included a further five sequences of tawny pipit COI published in Genebank (Accession numbers GQ481330-34). The software BEAST v 1.8.0 (Drummond & Rambaut 2007) was used to estimate divergence times between COI sequences of Berthelot's and tawny pipits using constant size population priors, the Hasegawa-Kishino-Yano (HKY) nucleotide substitution model and a strict clock with a divergence rate of 2.1% per million years (Weir & Schluter 2008). The software was run for 10 million generations with a burn in of 1 million generations. The software Tracer v 1.5 (Rambaut & Drummond 2009) was then used to assess convergence of the chains and to obtain the mean and 95% intervals of highest posterior density (HPD) estimates of divergence time.

2.3.3 TLR genotyping

DNA was extracted using a salt extraction method following Richardson *et al.* (2001). In both Berthelot's and tawny pipits we amplified TLR1LA, TLR1LB and TLR3 using the primers published by Alcaide & Edwards (2011), and TLR4 and TLR21 with the primers published by Grueber *et al.* (2012). These primers amplify large fragments (mean = 862, ranging from 622 to 1041 bp) of the TLRs extracellular domains, the regions directly involved in pathogen recognition. We were unable to successfully amplify a further four TLRs (TLR2A, TLR2B, TLR5 and TLR15) using either the primers available in Alcaide & Edwards (2011) or Grueber *et al.* (2012). Polymerase chain reactions (PCR) were performed in a Tetrad thermocycler (MJ Research) using the following profile: initial denaturing at 94°C for 4 mins, followed by 35 cycles of denaturing at 94°C for 40 s, annealing at 60°C (TLR1LA, TLR3, TLR4 and TLR21) or 64°C (TLR1LB) for 40 s and extension at 72°C for 80 s; a final extension step was performed at 72°C

for 10 mins. Reactions were conducted in 10 µl volumes using the TopTaq polymerase master mix kit (Qiagen), 0.5 µM of each primer and *ca*. 30 ng of DNA. Amplified fragments were visualized on 2% agarose gels stained with ethidium bromide and purified using a mixture of recombinant alkaline phosphatase and exonuclease I, incubating at 37°C for 30 mins, followed by enzyme inactivation at 95°C for 5 mins. Amplified fragments were sequenced using the BigDye terminator kit (Applied Biosystems) following the same procedure as for COI (section 2.3.2). Single nucleotide polymorphisms (SNPs) were detected by visually examining chromatograms. The International Union of Pure and Applied Chemistry code for degenerate nucleotides was used for heterozygous positions. Individual haplotypes were resolved using the PHASE algorithm (Stephens *et al.* 2001) implemented in DnaSP 5.10.01 (Librado & Rozas 2009). All haplotypes were confirmed by repeat PCR and sequencing from at least two samples.

2.3.4 Analyses

All analyses were carried out in R version 3.0.2 (R Development Core Team 2011), unless stated otherwise. To determine the relationship between alleles from each TLR across different bird species, we constructed maximum likelihood trees for each locus using 1,000 bootstrap replications and the general time reversible substitution model using MEGA 6 (Tamura *et al.* 2013). These trees included sequences from the two pipit species obtained in this study and three species obtained from GenBank: the house finch, *Carpodacus mexicanus,* lesser kestrel, *Falco naumanni* (Alcaide & Edwards 2011), and the New Zealand robin, *Petroica australis rakiura* (Grueber *et al.* 2012). To further visualize relationships between TLR genes in Berthelot's and tawny pipits, we built haplotype networks using HapStar 0.5 (Teacher & Griffiths 2011).

The ratio (ω) of nonsynonymous to synonymous substitutions per site (d_N/d_S) was calculated in MEGA 6 (Tamura *et al.* 2013) for each locus using the haplotype sequences identified in Berthelot's and tawny pipits. Significance of the relationship between d_N and d_S ($d_N<d_S$, $d_N>d_S$) was tested with 10,000 bootstrap replications. This method gives an indication of selection averaged over all sites in the sequence but requires a strong signal to detect selection (Pond & Frost 2005b). In order to explore whether specific codons, rather than the entire sequence, were under positive selection, three codon-based methods to calculate ω were used: two maximum likelihood methods (random effects likelihood (REL, Pond & Frost 2005b), and fast unconstrained Bayesian approximation (FUBAR, Murrell *et al.* 2013)) and the mixed effects

model of evolution (MEME, Murrell *et al.* 2012). Sites with Bayes factor > 50 for REL, posterior probabilities > 0.9 for FUBAR, and *P* values < 0.1 for MEME were considered to support positive selection. Only sites identified by at least two of the three different methods were considered to be under positive selection. Prior to running analyses, the best fitting nucleotide substitution model was determined for each locus using a model selection approach. These tests were all run in Datamonkey (http://datamonkey.org, Pond & Frost 2005a) with sequences from the two pipit species. Gene conversion and recombination were estimated for each locus using methods described by Betrán *et al.* (1997) and Hudson (2007), respectively, implemented in DnaSP 5.10.01 (Librado & Rozas 2009).

To best visualize genetic diversity among Berthelot's pipit populations, islands were pooled into archipelagos in accordance with findings from previous research (Spurgin *et al.* 2014). For tawny pipits we considered two populations – Africa and the Iberian Peninsula. We calculated measures of TLR nucleotide diversity in MEGA 6 (Tamura *et al.* 2013), and tested for differences between the two pipit species, and among the 3 archipelago populations of Berthelot's pipit, using randomization tests (Manly 2007). Tests of Hardy-Weinberg equilibrium were carried out using the web version of GENEPOP (http://genepop.curtin.edu.au/, Raymond & Rousset 1995). Nucleotide diversity comparisons involving Selvagem Grande were not possible because of the low number of haplotypes detected in this population (see results, Table 2.1). In addition to nucleotide diversity, for each TLR locus and population we calculated expected heterozygosity using Arlequin 3.5 (Excoffier & Lischer 2010). Finally, we calculated allelic richness after correcting for differences in sample size between the archipelagos (as a result of there being different numbers of populations per archipelago) using the software HP-RARE 1.0 (Kalinowski 2005).

We limited analyses of genetic differentiation to the Berthelot's pipit, due to the low sample size available for tawny pipits. To obtain levels of genetic differentiation among the three archipelagos we calculated global F_{ST} for each TLR locus using Arlequin 3.5 (Excoffier & Lischer 2010). Significance of F_{ST} values was tested with 10,000 permutations. A previous study of Berthelot's pipits identified a pattern of 'isolation by colonization', whereby neutral genetic structure among population was largely the product of the relative bottleneck severity of populations (Spurgin *et al.* 2014). Here we assessed whether TLR genetic structure among islands was consistent with this pattern. As an index of bottleneck distance, we used a modified version of Garza and Williamson's (2001) *M* ratio, developed to reflect the relative

bottleneck severity of pairs of islands based on microsatellite variation (Spurgin *et al.* 2014). Garza and Williamson's (2001) *M* ratio is calculated by dividing the number of microsatellite alleles found in a population by the range in size in base pairs. This can be used to detect reductions in the effective population size, because when alleles are lost, the number of alleles is expected to be reduced more quickly than the size range. The modified *M* ratio (Spurgin *et al.* 2014) is calculated for pairs of populations as: $M\mathrm{poppair} = -\mathrm{Log}\,(M\mathrm{pop1} \times M\mathrm{pop2})$. For every microsatellite (Spurgin *et al.* 2014) and TLR locus, we tested whether pairwise F_{ST} was related to bottleneck distance with a Mantel test, implemented in the R package Ecodist (Goslee & Urban 2007). We then compared the distribution of correlation coefficients from the Mantel tests for microsatellite and TLR loci using a t-test (Nosil *et al.* 2008). A strong correlation between the pairwise F_{ST} values based on TLR variation and bottleneck distance would indicate that the distribution of TLR variation among populations follows the same pattern of isolation by colonization as the microsatellites. If no correlation is found the patterns of genetic structure may be the product of population specific patterns of selection.

2.4 Results

2.4.1 Divergence time between Berthelot's and tawny pipits

We found nine and two variable sites within the tawny and Berthelot's pipit COI haplotypes, respectively. The mean estimated time to most recent common ancestor of all Berthelot's and tawny pipit haplotypes was 2.34 million years (95% HPD = 1.57-3.13 million years).

2.4.2 TLR variation within and across species

TLR1LA, TLR1LB, TLR3, TLR4 and TLR21 were each sequenced in 78-80 Berthelot's pipits, with 10-53 individuals per archipelago, and 12 tawny pipits (Table 2.1). A summary of variation across the five TLR loci is given in Table 2.2. No SNPs resulted in frame shifts or stop codons. A total of 34 alleles were identified in Berthelot's pipits (Supplementary Table S2.1), with 5-8 haplotypes per locus, which translated as between three and six different amino acid variants at each locus (Table 2.1). In tawny pipits we identified a total of 62 alleles (Supplementary Table S2.2), with between 7-18 haplotypes and 5-11 amino acid variants per locus (Table 2.1).

Table 2.1 Polymorphism at five TLR loci in the three Berthelot's pipit (*A. berthelotii*) archipelago populations (CI = Canary Islands, M = Madeira, S = Selvagem Grande) and the two closest populations of the sister species, the tawny pipit (*A. campestris*, IP = Iberian Peninsula, Af = north Africa). Significant deviations from Hardy-Weinberg equilibrium are underlined.

Locus	Population (N)	# Poly sites	# alleles	π(SD)[1]	# amino acid variants	allelic richness[2]	Hd[3]	Ho[4]	He[5]	HWE p-value[6]
TLR1LA	*A. berthelotii* (78)	6	8	22.7 (3.7)	3					
	CI (53)	6	7	22.7 (4.5)	3	4.06	0.70	0.58	0.70	0.338
	S (10)	1	2	9.9 (5.0)	1	2.00	0.51	0.20	0.48	0.082
	M (15)	3	4	16.5 (4.4)	2	3.94	0.62	0.40	0.60	0.005
	A. campestris (12)	21	15	38.4 (5.7)	9					
	IP (6)	11	8	27.6 (6.8)	6	7.81	0.94	1.00	0.86	1.000
	Af (6)	17	9	44.6 (7.7)	6	7.81	0.94	0.83	0.86	0.428
TLR1LB	*A. berthelotii* (78)	7	8	21.6 (3.3)	3					
	CI (53)	5	6	19.1 (3.4)	3	2.91	0.47	0.56	0.47	0.104
	S (10)	1	2	10.2 (5.1)	2	2.00	0.52	0.10	0.49	0.015
	M (15)	4	5	18.4 (3.9)	2	4.78	0.70	0.60	0.69	0.282
	A. campestris (12)	18	10	44.8 (5.8)	5					
	IP (6)	11	6	39.6 (9.6)	3	5.17	0.68	0.83	0.62	1.000
	Af (6)	10	5	45.0 (7.9)	4	4.33	0.58	0.67	0.53	1.000
TLR3	*A. berthelotii* (80)	5	6	16.0 (2.6)	3					
	CI (53)	4	5	16.0 (2.6)	3	4.31	0.48	0.49	0.47	0.239
	S (12)	1	2	9.6 (4.8)	1	2.00	0.39	0.50	0.37	0.528
	M (15)	2	3	12.8 (4.3)	1	3.00	0.68	0.80	0.66	0.554
	A. campestris (12)	15	12	30.6 (4.0)	8					
	IP (6)	9	6	30.7 (6.3)	5	5.17	0.68	0.83	0.62	1.000
	Af (6)	9	8	28.5 (4.6)	5	6.83	0.85	1.00	0.78	1.000
TLR4	*A. berthelotii* (78)	5	7	29.5 (4.5)	6					
	CI (53)	2	3	25.7 (5.6)	4	3.05	0.62	0.58	0.61	0.528
	S (10)	0	1	0.0 (0.0)	1	1.00	0.00	0.00	0.00	N/A
	M (15)	4	4	30.3 (8.2)	4	4.00	0.76	0.87	0.73	0.369
	A. campestris (12)	22	18	73.1 (7.4)	11					
	IP (6)	15	11	71.8 (7.7)	6	9.32	0.98	0.83	0.90	0.071
	Af (6)	16	9	70.6 (12.8)	7	7.82	0.94	0.83	0.87	0.390
TLR21	*A. berthelotii* (80)	4	5	28.9 (6.2)	3					
	CI (53)	3	4	28.9 (6.2)	2	2.99	0.54	0.55	0.54	0.076
	S (12)	1	2	16.1 (8.0)	1	2.00	0.43	0.42	0.41	1.000
	M (15)	1	2	16.1 (8.0)	1	2.00	0.43	0.33	0.42	0.539
	A. campestris (12)	6	7	32.2 (5.0)	5					
	IP (6)	6	7	32.2 (5.0)	5	5.57	0.82	1.00	0.76	1.000
	Af (6)	1	2	16.1 (8.0)	1	2.00	0.36	0.20	0.32	1.000

[1] Nucleotide diversity x 10^4 (Standard deviation)
[2] Corrected for sampling difference
[3] Haplotype diversity
[4] Observed Heterozygosity
[5] Expected heterozygosity
[6] *P*-value of the Hardy Weinberg equilibrium exact test with 1 million Markov chain steps.

Table 2.2 Variation at the exons encoding extracellular domains of five toll-like receptor genes in Berthelot's pipit (*Anthus berthelotii*).

Locus (exon)	N[1]	Fragment length (bp)	SNPs[2]	Amino acid	Syn:nsyn[3]	H[4]	Amino acid variants	π x10^{-4} (SD)[5]
TLR1LA (2)	78	1011	70: C/T	Leu	4:2	8	3	22.6 (3.7)
			284:C/T	Arg/Cys				
			322:C/T	Tyr				
			657:C/T	Phe/Ser				
			871:C/T	Leu				
			925:C/T	Ala				
TLR1LB (1)	78	977	329:C/T	Cys	5:2	8	3	21.6 (3.3)
			520:C/T	Phe/Ser				
			677:A/G	Pro				
			734:C/T	Leu				
			800:A/G	Pro				
			917:A/G	Thr				
			946:A/T	Leu/Gln				
TLR3 (4)	80	1041	197:C/T	Ser/Leu	3:2	6	3	16.0 (2.6)
			573:C/T	Gly				
			696:A/G	Val				
			748:C/T	Leu				
			814:C/G	Asp/His				
TLR4 (3)	78	661	216:A/G	Asp/Gly	1:4	7	6	29.6 (4.5)
			281:A/G	Lys/Glu				
			301:C/T	Pro				
			303:A/C/T	Lys/Thr/Met				
			321:A/G	Glu/Arg				
TLR21 (1)	80	622	247:C/T	Asp	3:1	5	3	28.9 (6.2)
			554:C/T	Leu				
			579:C/G	Gly/Ala				
			607:C/T	Phe				
Mean		862	5.4		3.2:2.2	6.8		

[1]Number of individuals genotyped
[2]Single nucleotide polymorphisms. Number indicates nucleotide position in the sequence.
[3]Number of synonymous to number of nonsynonymous mutations
[4]Number of haplotypes
[5]Nucleotide diversity (Standard deviation)

The maximum likelihood phylogenies consistently show that alleles of each TLR locus cluster by species, except for the pipit species, which are paraphyletic (Fig. 2.2). Haplotype networks (Supplementary Figs. S2.1-S2.5) show that the two pipit species are very closely related at all TLRs, and that they share haplotypes at TLR1LA (two shared alleles, Supplementary Fig. S2.1) and TLR3 (one shared allele, Supplementary Fig. S2.3). In Berthelot's pipit most haplotypes are separated from one another by a single point mutation. Haplotype networks of TLR1LB, TLR3 and TLR4 (Supplementary Figs. S2.2-S2.4) show a star-like pattern, where most haplotypes are connected to one central haplotype present in all three archipelagos. The networks of TLR1LA,

and TLR21 haplotypes show chain-like patterns in Berthelot's pipits (Supplementary Figs. S2.1 and S2.5). We found no evidence for gene conversion or recombination among the TLR loci.

At the haplotype level we found an excess of synonymous over non-synonymous substitutions in all TLRs except for TLR4 (Table 2.2), with ω values ranging from 0.09 to 0.71 in Berthelot's pipit and from 0.17 to 0.32 when including sequences from tawny pipit (Table 2.3). The codon level selection tests (Table 2.4) identified two amino acid sites in TLR4: (i) codon 252 which produces three different amino acids in tawny pipits but is monomorphic in Berthelot's pipits, and (ii) codon 332 which codes for three and two different amino acids in Berthelot's and tawny pipits, respectively. At TLR3 another codon (315) which codes for two different amino acids in both species showed evidence of positive selection.

All TLR loci were polymorphic in all populations of both species, except TLR4 in the Selvagem Grande Berthelot's pipit population. None of the loci deviated significantly from Hardy-Weinberg equilibrium in any populations except for TLR1LA in Madeira and TLR1LB in Selvagem Grande (Table 2.1). Pairwise nucleotide diversity was significantly higher in tawny compared to Berthelot's pipits ($P < 0.001$; Table 2.1), but not significantly different between the tawny pipit populations sampled in Africa and Iberia ($P = 0.49$), or between Madeira and the Canary Islands populations of Berthelot's pipits ($P = 0.45$). Rarefacted allelic richness was highest in tawny pipits (mean ± s.e. = 6.61 ± 0.84 in the Iberian Peninsula, and 5.76 ± 1.14 in Africa; Fig. 2.3). Berthelot's pipits in Selvagem Grande had the lowest mean allelic richness (1.80 ± 0.15), followed by Madeira (3.54 ± 0.29), and the Canary Islands (3.48 ± 0.08) (Fig. 2.3). Of the Berthelot's pipit populations, the Canary Islands had the most private alleles across all loci (eight), while Madeira had five and Selvagem Grande had none.

Global population differentiation among archipelagos based on TLRs was significant for all five loci. F_{ST} values ranged from 0.08 in TLR21 to 0.38 in TLR3 (Table 2.5). Mantel tests showed that there was a strong correlation between island pairwise F_{ST} and bottleneck distance for TLR1LA ($r = 0.41$, $P = 0.037$), TLR3 ($r = 0.63$, $P = 0.004$) and TLR4 ($r = 0.63$, $P = 0.009$), and non-significant positive relationships for TLR1LB and TLR21 ($r = 0.34$, $P = 0.109$; $r = 0.38$, $P = 0.085$, respectively; Fig. 2.4). The mean (± S.E.) correlation coefficient between island pairwise F_{ST} and bottleneck distance was 0.48 ± 0.06 and 0.54 ± 0.05, for TLRs and microsatellites respectively. These values were not significantly different from one another (Fig. 2.5, $t(9.63) = 0.69$, $P = 0.504$).

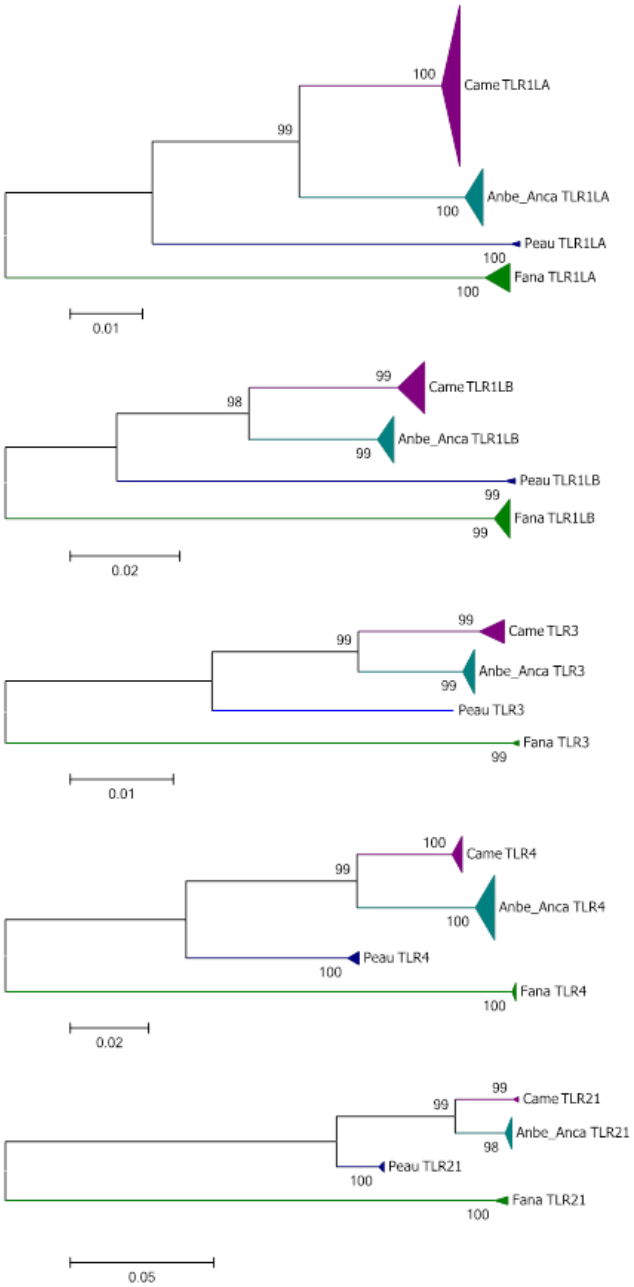


Figure 2.2 Maximum likelihood phylogenetic trees of haplotypes at five TLR loci in five bird species: Came = house finch, *Carpodacus mexicanus*; Peau = New Zealand robin, *Petroica australis rakiura*; Fana = Lesser kestrel, *Falco naumanni*; Anbe = Berthelot's pipit, *Anthus berthelotii,* and Anca = tawny pipit, *Anthus campestris.* Node values represent bootstrap support. Subtrees for each species were collapsed at bootstrap values higher than 98%. Height of the collapsed subtree is proportional to the number of haplotypes in the subtree.

Table 2.3 Mean ω (*P* values for alternative hypothesis of purifying selection, $d_N<d_.$, in brackets) across identified alleles of the five TLR loci in 1) Berthelot's pipit (Aber) and 2) both Berthelot's and tawny pipits (Aber+Acam). Significant values are in bold and underlined.

Locus	ω (Aber)	ω (Aber+Acam)
TLR1LA	**0.15 (0.049)**	**0.23 (0.004)**
TLR1LB	**0.09 (0.024)**	**0.17 (0.001)**
TLR3	0.20 (0.084)	**0.32 (0.022)**
TLR4	0.71 (0.389)	**0.24 (0.014)**
TLR21	0.30 (0.169)	0.22 (0.056)

Table 2.4 Codons in TLR3 and TLR4 identified as being under positive selection across Berthelot's and tawny pipits, using three different methods: REL, FUBAR and MEME. Codon numbers correspond to the chicken mRNA. Codons detected by two or more methods are shown in bold.

Locus	Codons		
	REL	FUBAR	MEME
TLR3	**315**	**315**	-
TLR4	**252**, 303, **332**	**252**, **332**	**252**

2.5 Discussion

Here we show that in the populations of Berthelot's pipits variation at TLRs was probably reduced as a consequence of its colonization history in Macaronesia, which perhaps occurred before the evolutionary split from its sister species, the tawny pipit. In addition, genetic variation at these TLR loci was further reduced by the bottlenecks experienced during the Berthelot's pipit more recent colonization of the Maderian and Selvagens archipelagos from the Canaries. Importantly, the population differentiation observed at the TLR loci in Berthelot's pipits mirrored that found previously with microsatellites, which is consistent with the demographic history of this species. However, we also found evidence for positive selection acting in two of the five TLRs analyzed.

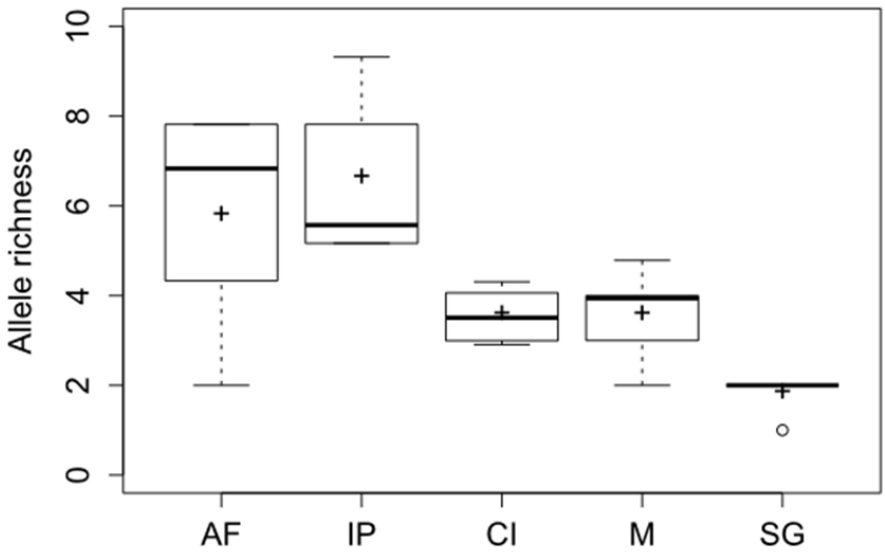


Figure 2.3 Distribution of allelic richness of five TLR loci (after accounting for differences in sample size) across the three archipelago populations of Berthelot's pipit, *Anthus berthelotii* (CI – Canary Islands, M – Madeira and SG – Selvagem Grande) and of tawny pipits, *Anthus campestris*, from the Iberian Peninsula (IP), and from North Africa (AF). Centre lines show the medians and crosses represent the means. Box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by open dots.

Our phylogenetic trees – including sequences from Berthelot's pipit, tawny pipit, house finch, New Zealand robin and lesser kestrel – indicate that within each TLR locus, variation groups according to the taxonomic relationship between the species – a finding consistent with other studies of TLR variation across species (Roach *et al.* 2005; Nakajima *et al.* 2008; Alcaide & Edwards 2011). We found no evidence of gene conversion and recombination among the TLR loci. Thus, contrary to other gene families, such as the MHC, where gene conversion and trans-species evolution are frequent (Ohta 1995; Klein *et al.* 1998; Spurgin *et al.* 2011), vertebrate TLRs seem to evolve independently and mainly by point mutation (Roach *et al.* 2005). However, a few studies have reported that gene conversion occurs in some TLR families that have recently been duplicated (Kruithof *et al.* 2007; Cormican *et al.* 2009; Mikami *et al.* 2012). In our data the sequences from the two pipit species clustered together within the tree and share some alleles (two alleles in TLR1LA and one in TLR3). However this is unsurprising when we consider the mtDNA divergence between the two species (*ca* 5%); given the much lower

nuclear mutation rate (Brown *et al.* 1979) and the relative short size of the TLR loci sequenced (622-1041 bp), we would expect few substitutions in our TLR loci. Thus, it is likely that complete lineage sorting has not yet occurred at these genes.

Table 2.5 Analyses of genetic differentiation expressed by F_{ST} at five TLR loci between three archipelago populations of Berthelot's pipits, *Anthus berthelotii*. CI = Canary Islands, M = Madeira, S = Selvagem Grande. Significance of F_{ST} values is denoted by an asterisk: * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$.

	Population pairs	F_{ST}
TLR1LA	S-M	0.08
	S-CI	0.15**
	M-CI	0.06*
	All populations	0.10***
TLR1LB	S-M	0.04
	S-CI	0.06
	M-CI	0.21***
	All populations	0.14***
TLR3	S-M	0.21**
	S-CI	0.52***
	M-CI	0.28***
	All populations	0.38***
TLR4	S-M	0.17**
	S-CI	0.25***
	M-CI	0.08**
	All populations	0.15***
TLR21	S-M	0.26**
	S-CI	0.02
	M-CI	0.09*
	All populations	0.08**

The effects of genetic drift on TLR variation can be seen at the species and population level. The lower TLR diversity in Berthelot's compared to tawny pipits is consistent with both a founder effect during the initial colonization of the islands, and with consistently lower effective population sizes in the restricted island populations. Indeed, levels of TLR variation in Berthelot's pipit (5-8 alleles per locus) are comparable to those reported for the bottlenecked New Zealand robin (2-5 alleles per locus) (Grueber *et al.* 2012), whereas TLR variation in tawny pipits (2-18 alleles per locus) is more comparable to that of widespread bird species such as lesser kestrel, *Falco naumanni* (2-15 alleles per locus) and house finch, *Carpodacus mexicanus* (2-20 alleles per locus) (Alcaide & Edwards 2011). The secondary colonization of Madeira and Selvagens by Berthelot's pipits from the Canary Islands also involved bottlenecks (Spurgin *et al.*

2014). The extremely low TLR variation on Selvagens is consistent with this, although we did not find such reduced variation in Madeira (see below). Indeed the differentiation at TLR loci among Berthelot's pipit populations are consistent with the pattern of colonization, bottlenecks and subsequent isolation inferred from microsatellite data for this species (Illera *et al.* 2007; Spurgin *et al.* 2014). Most strikingly, we found strong correlations between TLR differentiation and a pairwise bottleneck index which did not differ from those found for microsatellites. These results indicate that the TLR variation observed across populations has largely been shaped by the same pattern of isolation by colonization as inferred from microsatellites (Spurgin *et al.* 2014). Overall, our results concur with other studies that have identified demographic processes as being the predominant force shaping functional genetic variation at immune genes in and among small or bottlenecked populations (Bollmer *et al.* 2011; Girard & Angers 2011; Oliver & Piertney 2012; Sutton *et al.* 2013). Such findings contrast, however, with a recent study comparing patterns of variation at TLR2 haplotypes across populations of bank voles, where isolation by distance at TLR2 was not explained by demographic patterns but possibly by parasite-mediated selection (Tschirren *et al.* 2012).

The question then is whether selection has had any influence on the levels and distribution of genetic variation at the TLRs in Berthelot's pipits. At the whole haplotype level, tests of selection based on relative levels of synonymous and non-synonymous variation (ω) across the Berthelot's and tawny pipits provided no evidence for positive selection operating on the extracellular region of each of the five TLRs. Indeed, values of ω were negative, suggesting that purifying selection has been operating across most of this region of the TLR molecules, purging non-synonymous substitutions that might affect their capacity to bind the highly conserved PAMPs (Medzhitov & Janeway 1997). However, codon-based selection tests may be more appropriate if changes in a few key amino acids dictate which type of PAMP is bound by a specific type of TLR molecule. In line with this logic a number of studies have found specific amino acids under selection in TLRs (Alcaide & Edwards 2011; Areal *et al.* 2011; Fornůsková *et al.* 2013; Grueber *et al.* 2014), including our study which finds evidence that two sites in TLR4 and one at TLR3 are under positive selection in pipits.

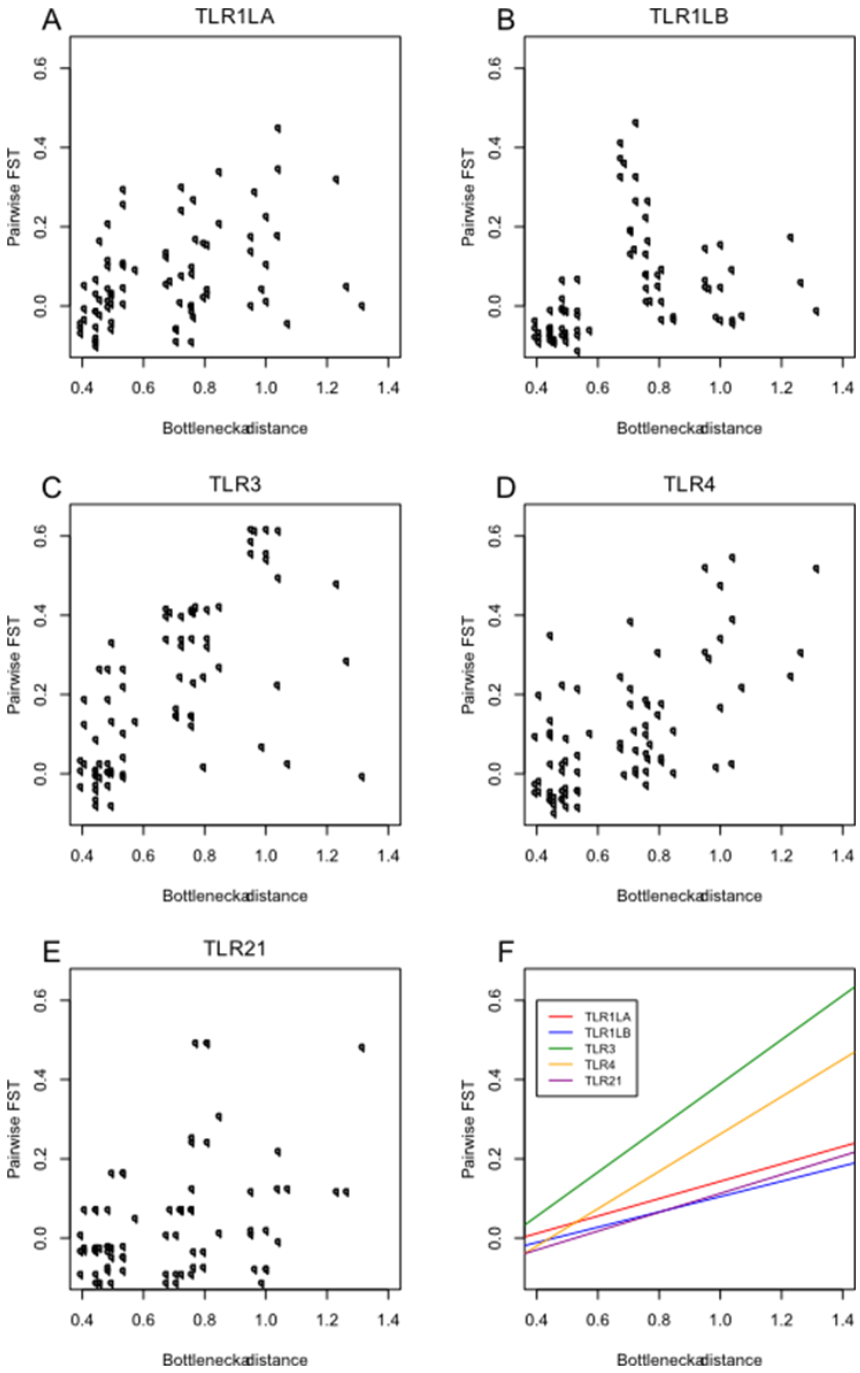


Figure 2.4 Pairwise F_{ST} for each of five TLR loci in relation to pairwise bottleneck distance between the 13 populations of Berthelot's pipits, *Anthus berthelotii*. A-D: scatter plot of pairwise FST in relation to pairwise bottleneck distance; F: Lines fit to the Mantel correlation coefficient between pairwise FST and pairwise bottleneck distance of the five TLRs shown in A-E.

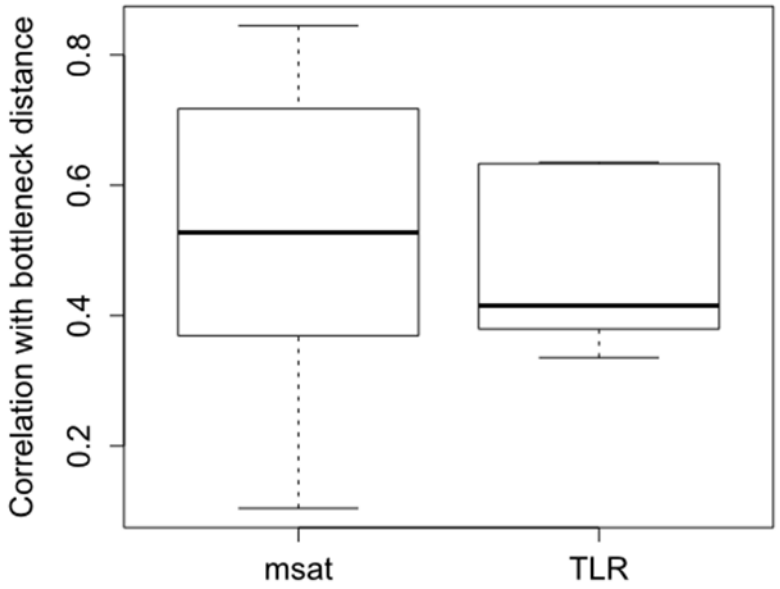


Figure 2.5 Distribution of Mantel correlation coefficients between pairwise F_{ST}s - for 21 microsatellite loci (msat) and five TLRs - and pairwise bottleneck distance between the 13 populations of Berthelot's pipits, *Anthus berthelotii*. Centre lines show the medians. Box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles.

The Madeira population was exceptional in that it had a number of private alleles at TLR loci (one in TLR1LA and two in TLR1LB and TLR4 each), whereas no private alleles were detected in the Madeiran populations at microsatellites or mitochondrial DNA variation (Illera 2007; Spurgin *et al.* 2014). One possibility is that these unique alleles are rare in the Canary Islands and were therefore not detected in our sample from that population. However, given the large sample size used for the Canary Islands (n = 53), and the fact that frequency of these alleles is high in Madeira, this explanation seems unlikely. Another possibility is that these alleles were present in the founding population of Madeira and were subsequently lost in the Canary Islands. Given the severity of the bottleneck after colonization of Madeira (Spurgin *et al.* 2014), this explanation also seems unlikely. This raises the possibility that these unique alleles have evolved *in situ* in Madeira. The haplotype networks appear to support this argument, at least for TLR1LB and TLR4 where the private alleles detected are only one or two base pairs different (and unconnected to any other non-private alleles) from a core central haplotype (Supplementary Figs. S2.2, S2.4) (Posada & Crandall 2001). Indeed, for TLR4 both unique alleles are one non-synonymous base pair change different from the same central haplotype

(Supplementary Fig. S2.4). Furthermore, the four Madeiran TLR4 alleles all translate to different amino acid sequences and exist at relatively high and balanced frequency (Supplementary Table S2.1), perhaps suggesting that new functional variants are being maintained in the population by selection. Based on this evidence, we suggest that the current distribution of functional polymorphisms at some of the TLR loci might be indicative of recent or contemporary selection. The same has been found in other studies where patterns of allele distribution among populations have provided evidence for balancing selection operating at functional loci (Cutrera *et al.* 2010), including TLRs (Ferrer-Admetlla *et al.* 2008). TLR screening of more individuals in these populations would be needed to confirm this.

Our evidence for selection at TLR4 concurs with the results of several multi-species studies which have found evidence of positive selection on this particular locus, both at the haplotype and codon level, (Nakajima *et al.* 2008; Wlasiuk & Nachman 2010; Alcaide & Edwards 2011; Areal *et al.* 2011; Fornůsková *et al.* 2013; Grueber *et al.* 2014). Polymorphisms in TLR4 have been linked to susceptibility and resistance to disease (reviewed in Noreen *et al.* 2012), and to juvenile survival (Grueber *et al.* 2013). The occurrence and distribution of the four functional variants of this locus that we detected in the Madeiran archipelago may, therefore, be the result of intense pathogen-mediated balancing selection within this archipelago. Interestingly, this locus is associated with malaria resistance in humans (Ferwerda *et al.* 2007), and a previous study that screened for pathogens in the Berthelot's pipits found that the Madeiran archipelago has a very high prevalence of two genera of malaria (*Plasmodium* and *Leucocytozoon*) as well as avian pox (Spurgin *et al.* 2012). Individual based infection and fitness investigations are now needed to test the possibility of associations between the TLR4 polymorphisms and specific pathogens within the Madeiran archipelago.

The results of our study on TLR variation offer some interesting comparisons with the patterns of variation previously described for MHC genes in Berthelot's pipits (Spurgin *et al.* 2011). In that study new functional variants appear to have been rapidly generated after the bottlenecks as a result of gene conversion and selection (Spurgin *et al.* 2011). Likewise, our study indicates that, at least in Madeira, new variation at TLR loci may have been recruited and maintained by selection post colonization. However, there was no evidence of gene conversion in Berthelot's pipit TLRs, where point mutations appear to have generated all the variation. Consequently, the populations have not been able to regenerate TLR variation anywhere near

the rate that was inferred for MHC genes. Both TLR and MHC genes have evolved by gene duplication (Klein & Sato 1998; Roach *et al.* 2005), but contrary to the MHC, TLRs appear to have evolved independently since these duplications and are not found in gene complexes (Roach *et al.* 2005). The nucleotide sequence of TLR genes has diverged (68% sequence similarity between the same regions in different loci) to the extent where gene conversion among the duplicated loci is no longer possible. The consequences of this lower variation at TLRs in bottlenecked populations is an interesting area yet to be explored (but see Grueber *et al* 2013).

Overall, our results clearly indicate that demographic history has been the main force behind the evolution of TLRs in the island population of Berthelot's pipit. The multiple bottlenecks this species experienced as it colonized the islands have largely determined the variation within and among populations, while balancing selection appears to have had relatively little effect in preserving variation during these events. There is, however, some evidence that selection has been acting, at least at two of these loci, and the possibility that new functional variants may have been generated and selected for since the bottlenecks. Studies of associations between specific pathogens and TLR polymorphisms, and how these affect individual fitness and innate immunity, are now needed in this and other natural systems to better understand the role of pathogen-mediated selection in shaping functional genetic differences across populations.

2.6 References

Acevedo-Whitehouse K, Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology & Evolution* **21**, 433-438.

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

Akira S (2003) Toll-like receptor signaling. *Journal of Biological Chemistry* **278**, 38105-38108.

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

Areal H, Abrantes J, Esteves PJ (2011) Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *Bmc Evolutionary Biology* **11**, 368.

Basu M, Maji AK, Chakraborty A*, et al.* (2010) Genetic association of Toll-like-receptor 4 and tumor necrosis factor-alpha polymorphisms with *Plasmodium falciparum* blood infection levels. *Infection Genetics and Evolution* **10**, 686-696.

Belvin MP, Anderson KV (1996) A conserved signaling pathway: The *Drosophila* Toll-Dorsal pathway. *Annual Review of Cell and Developmental Biology* **12**, 393-416.

Betrán E, Rozas J, Navarro A, Barbadilla A (1997) The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. *Genetics* **146**, 89-99.

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

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 1967-1971.

Cabe PR (1998) The effects of founding bottlenecks on genetic variation in the European starling (*Sturnus vulgaris*) in North America. *Heredity* **80**, 519-525.

Chen JSC, Wang TY, Tzeng TD, Wang CY, Wang D (2008) Evidence for positive selection in the TLR9 gene of teleosts. *Fish & Shellfish Immunology* **24**, 234-242.

Cormican P, Lloyd AT, Downing T*, et al.* (2009) The avian Toll-Like receptor pathway-Subtle differences amidst general conformity. *Developmental and Comparative Immunology* **33**, 967-973.

Cramp S, Perrins CM (1977-1994) *Handbook of the birds of Europe, the Middle East and Africa. The birds of the western Palearctic* Oxford University Press, Oxford.

Cutrera AP, Lacey EA, Mora MS, Lessa EP (2010) Effects of contrasting demographic histories on selection at major histocompatibility complex loci in two sympatric species of tuco-tucos (Rodentia: Ctenomyidae). *Biological Journal of the Linnean Society* **99**, 260-277.

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

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology* **7**, 214.

Ekblom R, French L, Slate J, Burke T (2010) Evolutionary analysis and expression profiling of zebra finch immune genes. *Genome Biology and Evolution* **2**, 781-790.

England PR, Osler GHR, Woodworth LM*, et al.* (2003) Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. *Conservation Genetics* **4**, 595-604.

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

Ferrer-Admetlla A, Bosch E, Sikora M*, et al.* (2008) Balancing selection is the main force shaping the evolution of innate immunity genes. *Journal of Immunology* **181**, 1315-1322.

Ferwerda B, McCall MBB, Alonso S*, et al.* (2007) TLR4 polymorphisms, infectious diseases, and evolutionary pressure during migration of modern humans. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16645-16650.

Fornůsková A, Vinkler M, Pagès M*, et al.* (2013) Contrasted evolutionary histories of two toll-like receptors (Tlr4 and Tlr7) in wild rodents (murinae). *Bmc Evolutionary Biology* **13**, 194.

Frankham R (1996) Relationship of genetic variation to population size in wildlife. *Conservation Biology* **10**, 1500-1508.

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

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**, 305-318.

Gautschi B, Widmer A, Joshi J, Koella JC (2002) Increased frequency of scale anomalies and loss of genetic variation in serially bottlenecked populations of the dice snake, *Natrix tessellata*. *Conservation Genetics* **3**, 235-245.

Girard P, Angers B (2011) The functional gene diversity in natural populations over postglacial areas: The shaping mechanisms behind genetic composition of longnose dace (*Rhinichthys cataractae*) in northeastern North America. *Journal of Molecular Evolution* **73**, 45-57.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**, 1-19.

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

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

Grueber CE, Wallis GP, King TM, Jamieson IG (2012) Variation at innate immunity toll-like receptor genes in a bottlenecked population of a New Zealand robin. *PloS one* **7**, e45011.

Hall TA (1999) BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium Series* **41**, 95-98.

Hartl DL, Clark AG (2007) *Principles of population genetics*, 4th edn. Sinauer Associates, Sunderland.

Hedrick PW, Thomson G (1983) Evidence for balancing selection at HLA. *Genetics* **104**, 449-456.

Hudson RR (2007) Estimating the recombination parameter of a finite population model without selection (Reprinted). *Genetics Research* **89**, 427-432.

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

Illera JC (2007) Bisbita Caminero *Anthus berthelotii*. In: *Atlas de las aves nidificantes en el archipiélago Canario (1997-2003)* (ed. JA L), pp. 344-347. General de Conservación de la Naturaleza-Sociedad Española de Ornitología, Madrid.

Illera JC, Emerson BC, Richardson DS (2007) Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology* **16**, 4599-4612.

Jensen LF, Hansen MM, Mensberg KL, Loeschcke V (2008) Spatially and temporally fluctuating selection at non-MHC immune genes: evidence from TAP polymorphism in populations of brown trout (*Salmo trutta*, L.). *Heredity* **100**, 79-91.

Jiggins FM, Hurst GDD (2003) The evolution of parasite recognition genes in the innate immune system: Purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. *Journal of Molecular Evolution* **57**, 598-605.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* **5**, 187-189.

Keestra AM, de Zoete MR, Bouwman LI, van Putten JPM (2010) Chicken TLR21 is an innate CpG DNA receptor distinct from mammalian TLR9. *The Journal of Immunology* **185**, 460-467.

Klein J, Sato A (1998) Birth of the major histocompatibility complex. *Scandinavian Journal of Immunology* **47**, 199-209.

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

Kruithof EKO, Satta N, Liu JW, Dunoyer-Geindre S, Fish RJ (2007) Gene conversion limits divergence of mammalian TLR1 and TLR6. *Bmc Evolutionary Biology* **7**, 148.

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

Lohman DJ, Prawiradilaga DM, Meier R (2009) Improved COI barcoding primers for Southeast Asian perching birds (Aves: Passeriformes). *Molecular Ecology Resources* **9**, 37-40.

Manly BF (2007) *Randomization, bootstrap and Monte Carlo methods in biology* Chapman and Hall, London.

Medzhitov R, Janeway CA (1997) Innate immunity: The virtues of a nonclonal system of recognition. *Cell* **91**, 295-298.

Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N (2012) Molecular evolution of vertebrate Toll-like receptors: Evolutionary rate difference between their leucine-rich repeats and their TIR domains. *Gene* **503**, 235-243.

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

Misch EA, Hawn TR (2008) Toll-like receptor polymorphisms and susceptibility to human disease. *Clinical Science* **114**, 347-360.

Murrell B, Moola S, Mabona A*, et al.* (2013) FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution* **30**, 1196-1205.

Murrell B, Wertheim JO, Moola S*, et al.* (2012) Detecting individual sites subject to episodic diversifying selection. *Plos Genetics* **8**, e1002764

Nakajima T, Ohtani H, Satta Y*, et al.* (2008) Natural selection in the TLR-related genes in the course of primate evolution. *Immunogenetics* **60**, 727-735.

Noreen M, Shah MAA, Mall SM*, et al.* (2012) TLR4 polymorphisms and disease susceptibility. *Inflammation Research* **61**, 177-188.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**, 316-336.

Ohta T (1995) Gene conversion vs. point mutation in generating variability at the antigen recognition site of major histocompatibility complex loci. *Journal of Molecular Evolution* **41**, 115-119.

Oliver MK, Piertney SB (2012) Selection Maintains MHC Diversity through a Natural Population Bottleneck. *Molecular Biology and Evolution* **29**, 1713-1720.

Outlaw DC, Voelker G (2006) Phylogenetic tests of hypotheses for the evolution of avian migration: A case study using the Motacillidae. *Auk* **123**, 455-466.

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

Poltorak A, Smirnova I, He XL*, et al.* (1998) Genetic and physical mapping of the LPS locus: Identification of the Toll-4 receptor as a candidate gene in the critical region. *Blood Cells Molecules and Diseases* **24**, 340-355.

Pond SLK, Frost SDW (2005a) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531-2533.

Pond SLK, Frost SDW (2005b) Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**, 1208-1222.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution* **16**, 37-45.

R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A, Drummond A (2009) Tracer v1. 5 Available from http://beast.bio.ed.ac.uk.

Raymond M, Rousset F (1995) GENEPOP (Version-1.2) - Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Richardson DS, Jury FL, Blaakmeer K, Komdeur J, Burke T (2001) Parentage assignment and extra-group paternity in a cooperative breeder: the Seychelles warbler (*Acrocephalus sechellensis*). *Molecular Ecology* **10**, 2263-2273.

Roach JC, Glusman G, Rowen L*, et al.* (2005) The evolution of vertebrate Toll-like receptors. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9577-9582.

Schröder NW, Schumann RR (2005) Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *Lancet Infectious Diseases* **5**, 156-164.

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

Spurgin LG, Illera JC, Jorgensen TH, Dawson DA, Richardson DS (2014) Genetic and phenotypic divergence in an island bird: isolation by distance, by colonisation or by adaptation? *Molecular Ecology* **23**, 1028-1039.

Spurgin LG, Illera JC, Padilla DP, Richardson DS (2012) Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia* **168**, 691-701.

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

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

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics* **68**, 978-989.

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

Takeuchi O, Sato S, Horiuchi T*, et al.* (2002) Cutting edge: Role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *Journal of Immunology* **169**, 10-14.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Teacher AGF, Griffiths DJ (2011) HapStar: automated haplotype network layout and visualization. *Molecular Ecology Resources* **11**, 151-153.

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

Tschirren B, Andersson M, Scherman K, Westerdahl H, Raberg L (2012) Contrasting patterns of diversity and population differentiation at the innate immunity gene toll-like receptor 2 (TLR2) in two sympatric rodent species. *Evolution* **66**, 720-731.

Tschirren B, Raberg L, Westerdahl H (2011) Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents. *Journal of Evolutionary Biology* **24**, 1232-1240.

Turner AK, Begon M, Jackson JA, Paterson S (2012) Evidence for selection at cytokine loci in a natural population of field voles (*Microtus agrestis*). *Molecular Ecology* **21**, 1632-1646.

Voelker G (1999) Dispersal, vicariance, and clocks: Historical biogeography and speciation in a cosmopolitan passerine genus (*Anthus*: motacillidae). *Evolution* **53**, 1536-1552.

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. *Molecular Ecology* **17**, 2321-2328.

Winternitz JC, Wares JP (2013) Duplication and population dynamics shape historic patterns of selection and genetic variation at the major histocompatibility complex in rodents. *Ecology and Evolution* **3**, 1552-1568.

Wlasiuk G, Nachman MW (2010) Adaptation and constraint at toll-like receptors in primates. *Molecular Biology and Evolution* **27**, 2172-2186.

Yoneyama M, Fujita T (2010) Recognition of viral nucleic acids in innate immunity. *Reviews in Medical Virology* **20**, 4-22.

Supplementary Table S2.1 Haplotypes identified at the five TLR loci in Berthelot's pipit, A*nthus berthelotii* populations. CI = Canary Islands (Number of alleles analysed (N) = 106), M = Madeira (N = 30), S = Selvagem Grande (N = 20 for TLR1LA, TLR1LB and TLR4, N = 24 for TLR3 and TLR21).

***A. berthelotii* allele**	**Sequence**	**Amino acid variant**	**N[1]**	**Freq.[2]**	**Freq. in CI**	**Freq. in M**	**Freq. in S**	**Genebank Accession #**
TLR1LA_1	CCCTCC	TLR1LA-1	53	0.340	0.26	0.57	0.40	KJ414322
TLR1LA_2	TCCTCC	TLR1LA-1	54	0.346	0.33	0.23	0.60	KJ414323
TLR1LA_3	CCCCCC	TLR1LA-2	40	0.256	0.35	0.10	0.00	KJ414324
TLR1LA_4	CTCTCC	TLR1LA-3	2	0.013	0.02	0.00	0.00	KJ414325
TLR1LA_5	CCTCCC	TLR1LA-2	1	0.006	0.01	0.00	0.00	KJ414326
TLR1LA_6	TCCTCT	TLR1LA-1	3	0.019	0.00	0.10	0.00	KJ414327
TLR1LA_7	TCCCCC	TLR1LA-2	2	0.013	0.02	0.00	0.00	KJ414328
TLR1LA_8	TCCTTT	TLR1LA-1	1	0.006	0.01	0.00	0.00	KJ414329
TLR1LB_1	CTGCGGT	TLR1LB-1	89	0.571	0.67	0.30	0.45	KJ414330
TLR1LB_2	CCGCGGT	TLR1LB-2	54	0.346	0.28	0.43	0.55	KJ414331
TLR1LB_3	CCGCAGT	TLR1LB-2	3	0.019	0.01	0.07	0.00	KJ414332
TLR1LB_4	TCGCGGT	TLR1LB-2	2	0.013	0.02	0.00	0.00	KJ414333
TLR1LB_5	CCACGGT	TLR1LB-2	4	0.026	0.00	0.13	0.00	KJ414334
TLR1LB_6	CCGCGAT	TLR1LB-2	1	0.006	0.01	0.00	0.00	KJ414335
TLR1LB_7	CCATGGT	TLR1LB-2	2	0.013	0.00	0.07	0.00	KJ414336
TLR1LB_8	CCGCGAA	TLR1LB-3	1	0.006	0.01	0.00	0.00	KJ414337
TLR3_1	CCACG	TLR3-1	31	0.194	0.01	0.40	0.75	KJ414338
TLR3_2	CTACG	TLR3-1	90	0.563	0.71	0.30	0.25	KJ414339
TLR3_3	TTACG	TLR3-2	6	0.038	0.06	0.00	0.00	KJ414340
TLR3_4	CTATG	TLR3-1	13	0.081	0.12	0.00	0.00	KJ414341
TLR3_5	CTACC	TLR3-3	9	0.056	0.08	0.00	0.00	KJ414342
TLR3_6	CTGCG	TLR3-1	11	0.069	0.02	0.30	0.00	KJ414343
TLR4_1	AGCCA	TLR4-1	86	0.538	0.53	0.33	0.00	KJ414344
TLR4_2	AGTAA	TLR4-2	39	0.244	0.29	0.27	0.00	KJ414345
TLR4_3	AGTCA	TLR4-1	16	0.100	0.15	0.00	0.00	KJ414346
TLR4_4	GGCCA	TLR4-3	5	0.031	0.00	0.17	0.00	KJ414347
TLR4_5	AACCA	TLR4-4	7	0.044	0.00	0.23	0.00	KJ414348
TLR4_6	AGCCG	TLR4-5	2	0.013	0.02	0.00	0.00	KJ414349
TLR4_7	AGTTA	TLR4-6	1	0.006	0.01	0.00	0.00	KJ414350
TLR21_1	CCGC	TLR21-1	82	0.513	0.53	0.30	0.71	KJ414351
TLR21_2	CCGT	TLR21-1	73	0.456	0.42	0.70	0.29	KJ414352
TLR21_3	TCGT	TLR21-1	1	0.006	0.01	0.00	0.00	KJ414353
TLR21_4	CCCC	TLR21-2	3	0.019	0.03	0.00	0.00	KJ414354

[1]Number of individuals with the allele across all populations
[2]Overall frequency in the pooled populations

Supplementary Table S2.2 Haplotypes identified at the five TLR loci in tawny pipit, *A. campestris*, populations. IP = Iberian Peninsula (N = 6), Af = north Africa (N = 6). Only variable sites are shown in the sequence.

A. campestris allele	Sequence	Amino acid variant	N[1]	Freq[2]	Freq. in IP	Freq. in Af	Genebank Accession #
TLR1LA_1	GAAGTAGTCCGCTCCCGACCC	TLR1LA_1	1	0.04	0.00	0.08	KJ414356
TLR1LA_3	GAAGTAGTCCGCCCCCGACCC	TLR1LA_2	6	0.25	0.25	0.25	KJ414357
TLR1LA_9	GAAGTAGACCGCCCCCGACCC	TLR1LA_3	1	0.04	0.08	0.00	KJ414358
TLR1LA_10	GAAGTAGTCCGCCTCCGACCC	TLR1LA_4	1	0.04	0.08	0.00	KJ414359
TLR1LA_11	GACGTAGTCCGCCCCCGACCC	TLR1LA_5	2	0.08	0.17	0.00	KJ414360
TLR1LA_12	GCAGTAGTCCGCCCCCGACCC	TLR1LA_2	3	0.08	0.17	0.08	KJ414361
TLR1LA_13	GAAGTAGTCCGCCCCCGCCCC	TLR1LA_6	1	0.04	0.08	0.00	KJ414362
TLR1LA_14	GAAGTAGTCCGTCCCCGACCC	TLR1LA_2	1	0.04	0.08	0.00	KJ414363
TLR1LA_15	GAAGTAGTATGCCCCCTCCTC	TLR1LA_7	1	0.04	0.08	0.00	KJ414364
TLR1LA_16	GACGTAGTCCGCCCCCGATCC	TLR1LA_5	1	0.04	0.00	0.08	KJ414365
TLR1LA_17	GCAATAGTCTGCCCCCGCCCC	TLR1LA_6	1	0.04	0.00	0.08	KJ414366
TLR1LA_18	GAAGCTGTACATCCCCGCCCC	TLR1LA_8	1	0.04	0.00	0.08	KJ414367
TLR1LA_19	AAAGTAGTCCGCCCCCGACCC	TLR1LA_2	2	0.08	0.00	0.17	KJ414368
TLR1LA_20	GCAGTAATCCGCCCCTGCCCC	TLR1LA_6	1	0.04	0.00	0.08	KJ414369
TLR1LA_21	GAAGTAGTATGCCCCCTCCCT	TLR1LA_9	1	0.04	0.00	0.08	KJ414370
TLR1LB_9	GCGCAAGTGCCCTCGGAC	TLR1LB_1	15	0.63	0.58	0.67	KJ414371
TLR1LB_10	CTGCAAGTGCCCTTGGAC	TLR1LB_2	1	0.04	0.08	0.00	KJ414372
TLR1LB_11	GCGTAAGTGCCCTCGGCC	TLR1LB_3	1	0.04	0.08	0.00	KJ414373
TLR1LB_12	GCGCAAGTACCCTCGGAC	TLR1LB_1	1	0.04	0.08	0.00	KJ414374
TLR1LB_13	GCGCAAGTGCCCCCGGAC	TLR1LB_1	1	0.04	0.08	0.00	KJ414375
TLR1LB_14	GCGCGAGTGTCTTCGGCT	TLR1LB_3	1	0.04	0.08	0.00	KJ414376
TLR1LB_15	GCACAAGTGTCCTCGGAC	TLR1LB_4	1	0.04	0.00	0.08	KJ414377
TLR1LB_16	GCGCACGTGCCCTCAGCC	TLR1LB_3	1	0.04	0.00	0.08	KJ414378
TLR1LB_17	GCGCAAGTGTTTTCGGCC	TLR1LB_3	1	0.04	0.00	0.08	KJ414379
TLR1LB_18	GCGCAAAAGCCCTCGTCC	TLR1LB_5	1	0.04	0.00	0.08	KJ414380
TLR3_2	AGTTAGCGGCGCGCC	TLR3_1	12	0.50	0.58	0.42	KJ414381
TLR3_7	GGTTATCAGCGCGCC	TLR3_2	1	0.04	0.08	0.00	KJ414382
TLR3_8	GGTTAGCGGCGCGTC	TLR3_3	1	0.04	0.08	0.00	KJ414383
TLR3_9	AGTTCGCGGCGCGCC	TLR3_1	1	0.04	0.08	0.00	KJ414384
TLR3_10	AGTTAGCGACGCGCC	TLR3_4	1	0.04	0.08	0.00	KJ414385
TLR3_11	AGCTAGCGGCACACC	TLR3_5	2	0.08	0.08	0.08	KJ414386
TLR3_12	AATTAGCGGCGCGCC	TLR3_6	1	0.04	0.00	0.08	KJ414387
TLR3_13	AGTTAGCGGCGTGCC	TLR3_1	1	0.04	0.00	0.08	KJ414388
TLR3_14	AGTAAGCGGCGCGCT	TLR3_1	1	0.04	0.00	0.08	KJ414389
TLR3_15	AGTAAGTGGCGCGCT	TLR3_7	1	0.04	0.00	0.08	KJ414390
TLR3_16	AGTTAGCGGCGCACC	TLR3_5	1	0.04	0.00	0.08	KJ414391
TLR3_17	AGTTAGTGGAGCGCC	TLR3_8	1	0.04	0.00	0.08	KJ414392

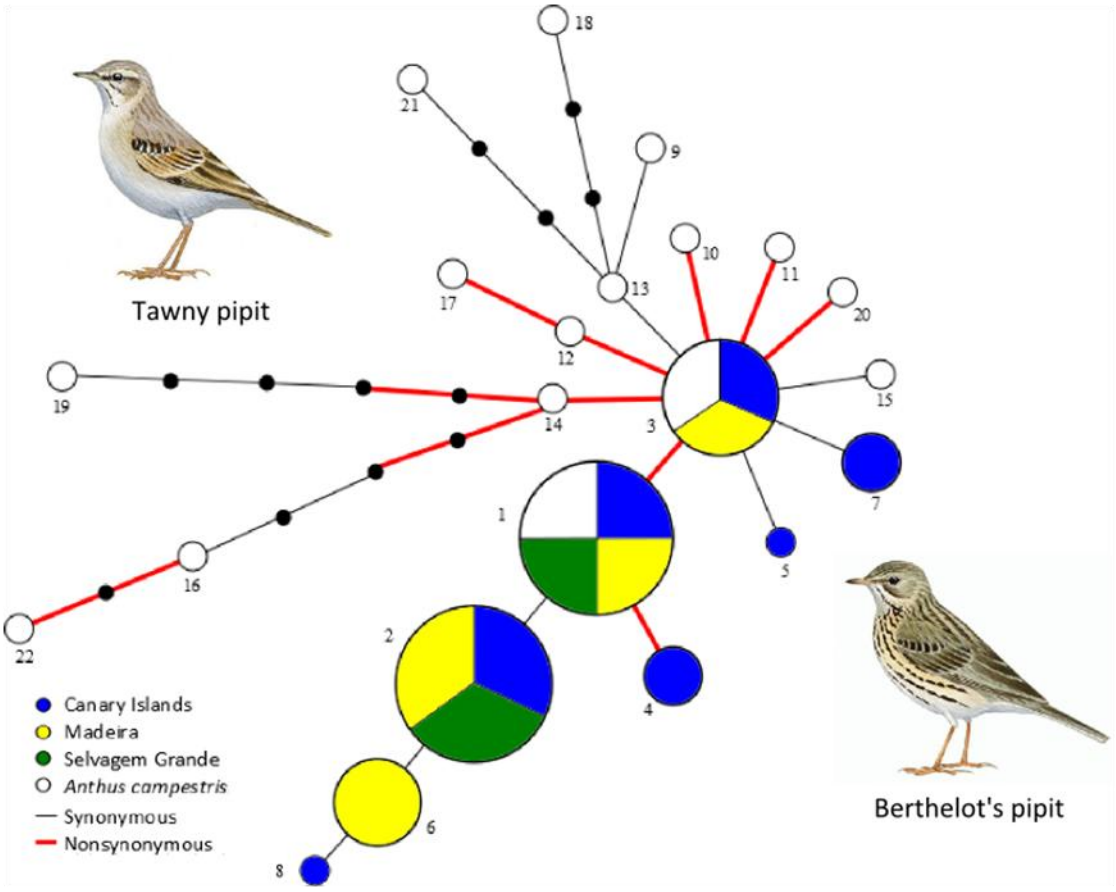
[1]Number of individuals with the allele across all populations
[2]Overall frequency in the pooled populations

Supplementary Table S2.2 (cont.)

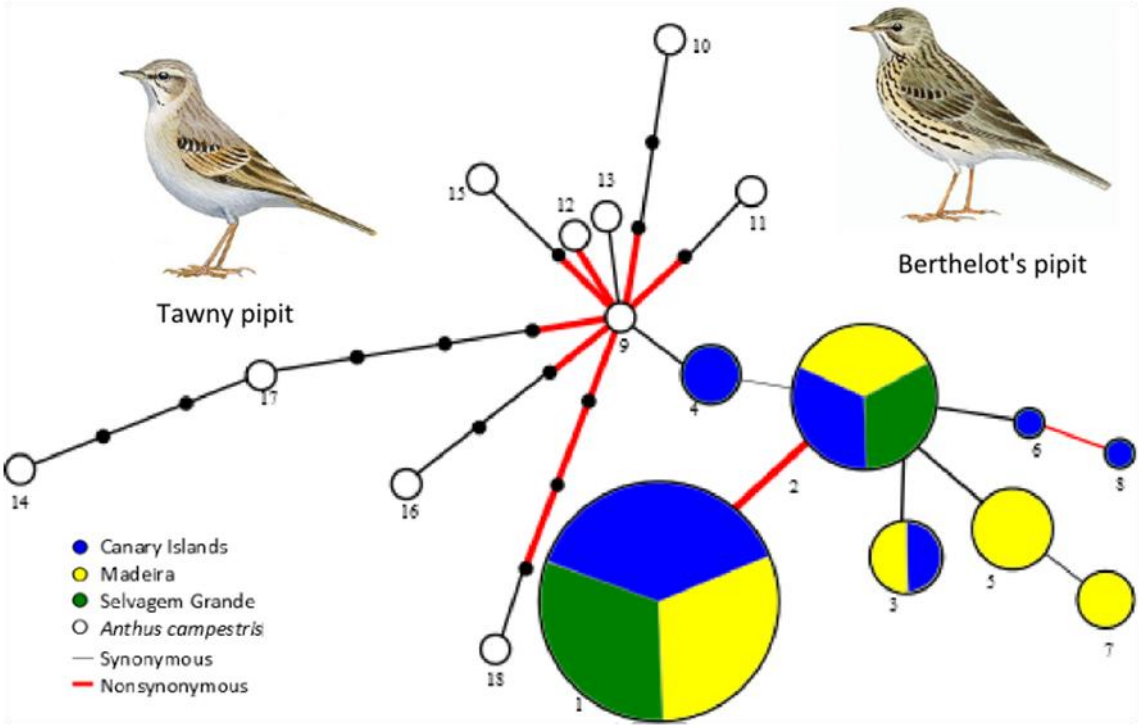
A. campestris allele	Sequence	Amino acid variant	N[1]	Freq.[2]	Freq. in IP	Freq. in Af	Genebank Accession #
TLR4_8	CGAGGACTCCCCCAGGGGACTC	TLR4_1	4	0.17	0.08	0.25	KJ414393
TLR4_9	TAACGGCCTCCCCAAGGGACTC	TLR4_2	1	0.04	0.08	0.00	KJ414394
TLR4_10	CGAGGACTCCTCCAGGGAACTC	TLR4_3	1	0.04	0.08	0.00	KJ414395
TLR4_11	GGACGACCCCCCCAGGGAGCTC	TLR4_4	1	0.04	0.08	0.00	KJ414396
TLR4_12	CAACGACTCCCCCAGGGGACTC	TLR4_5	1	0.04	0.08	0.00	KJ414397
TLR4_13	GAACGACCCCCCCAGGGGACTC	TLR4_5	1	0.04	0.08	0.00	KJ414398
TLR4_14	TGACGAGTCCCGCAGGGGACTC	TLR4_6	2	0.08	0.17	0.00	KJ414399
TLR4_15	CAACGACCCCCCCAGGGGACGC	TLR4_5	2	0.08	0.08	0.08	KJ414400
TLR4_16	CGACGACTCCCCCAGGGAGCTT	TLR4_4	1	0.04	0.08	0.00	KJ414401
TLR4_17	GGACGACCCCCGCAGGGGACTC	TLR4_5	1	0.04	0.08	0.00	KJ414402
TLR4_18	TGACGACTCCCGCAGGGGAATC	TLR4_4	1	0.04	0.08	0.00	KJ414403
TLR4_19	CAACGACTCCCGCAGGGGACTC	TLR4_5	2	0.08	0.00	0.17	KJ414404
TLR4_20	CGAGGACCCTCCCAGAGAGCTC	TLR4_7	1	0.04	0.00	0.08	KJ414405
TLR4_21	CGACGACTCCCCCGGGGGACTC	TLR4_8	1	0.04	0.00	0.08	KJ414406
TLR4_22	TAGCGAGTCCCGCAGGGGACTC	TLR4_9	1	0.04	0.00	0.08	KJ414407
TLR4_23	CAACAACTCCCCCAGGAGACTC	TLR4_10	1	0.04	0.00	0.08	KJ414408
TLR4_24	CGACGACTCCCCCAGGGGACTC	TLR4_5	1	0.04	0.00	0.08	KJ414409
TLR4_25	CGAGGACTCCCCGAGGGAACTC	TLR4_11	1	0.04	0.00	0.08	KJ414410
TLR21_6	CCCCTG	TLR21_1	1	0.04	0.07	0.00	KJ414411
TLR21_7	CCTCCG	TLR21_1	7	0.29	0.36	0.20	KJ414412
TLR21_8	CCCCCG	TLR21_1	12	0.50	0.33	0.80	KJ414413
TLR21_9	CCCTCG	TLR21_2	1	0.04	0.07	0.00	KJ414414
TLR21_10	CTCCTG	TLR21_3	1	0.04	0.07	0.00	KJ414415
TLR21_11	CCCCTA	TLR21_4	1	0.04	0.07	0.00	KJ414416
TLR21_12	GCCCCG	TLR21_5	1	0.04	0.07	0.00	KJ414417
Total	62	38			38	33	

[1]Number of individuals with the allele across all populations
[2]Overall frequency in the pooled populations

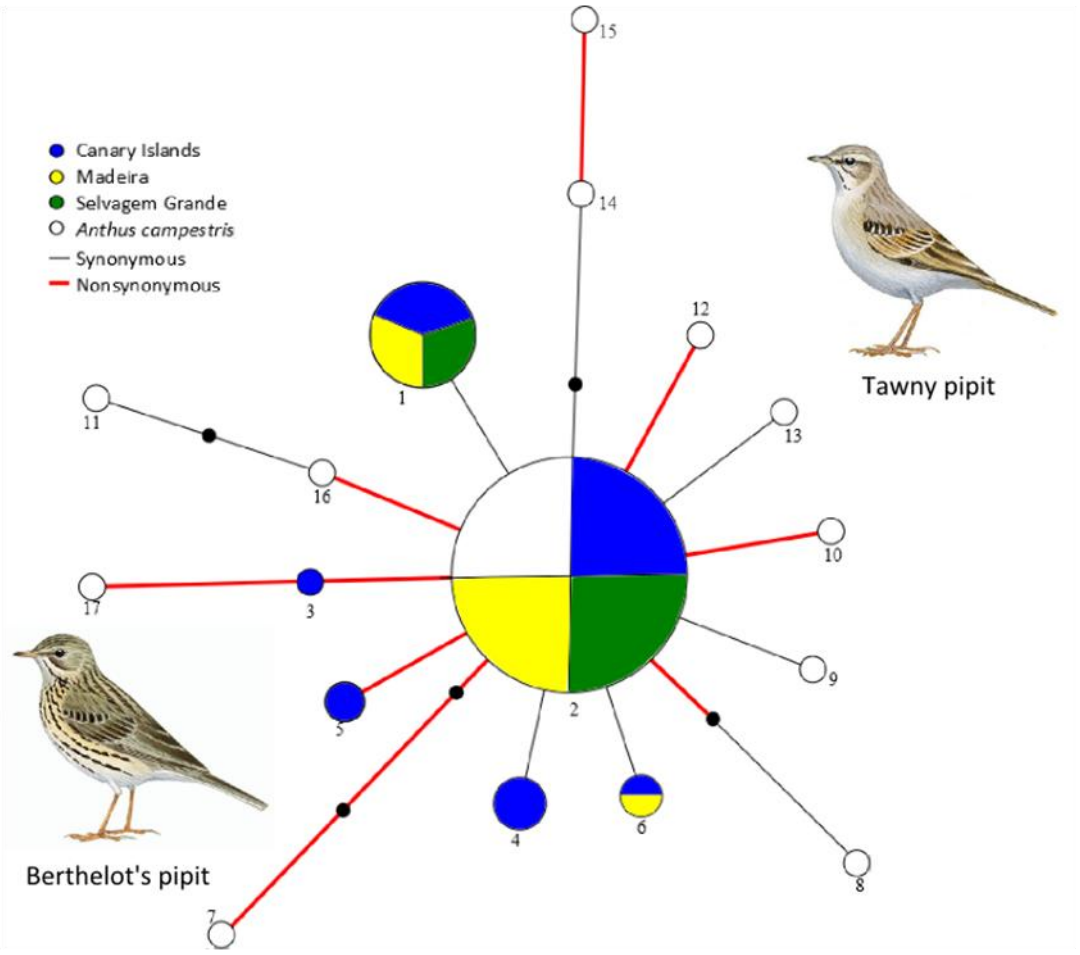
Supplementary Figures S2.1 – S2.5 Network of TLR haplotypes found in populations of Berthelot's pipits (Yellow: Madeira, Blue: Canary Islands, Green: Selvagens) and in tawny pipits (white circles). Each circle represents one haplotype. Connections between circles denote the number of nucleotide substitutions needed to change from one haplotype to another. Nonsynonymous substitutions are marked in red. Haplotype number is denoted beside each circle and size of the circle is proportional to the abundance of the haplotype only in Berthelot's pipits.



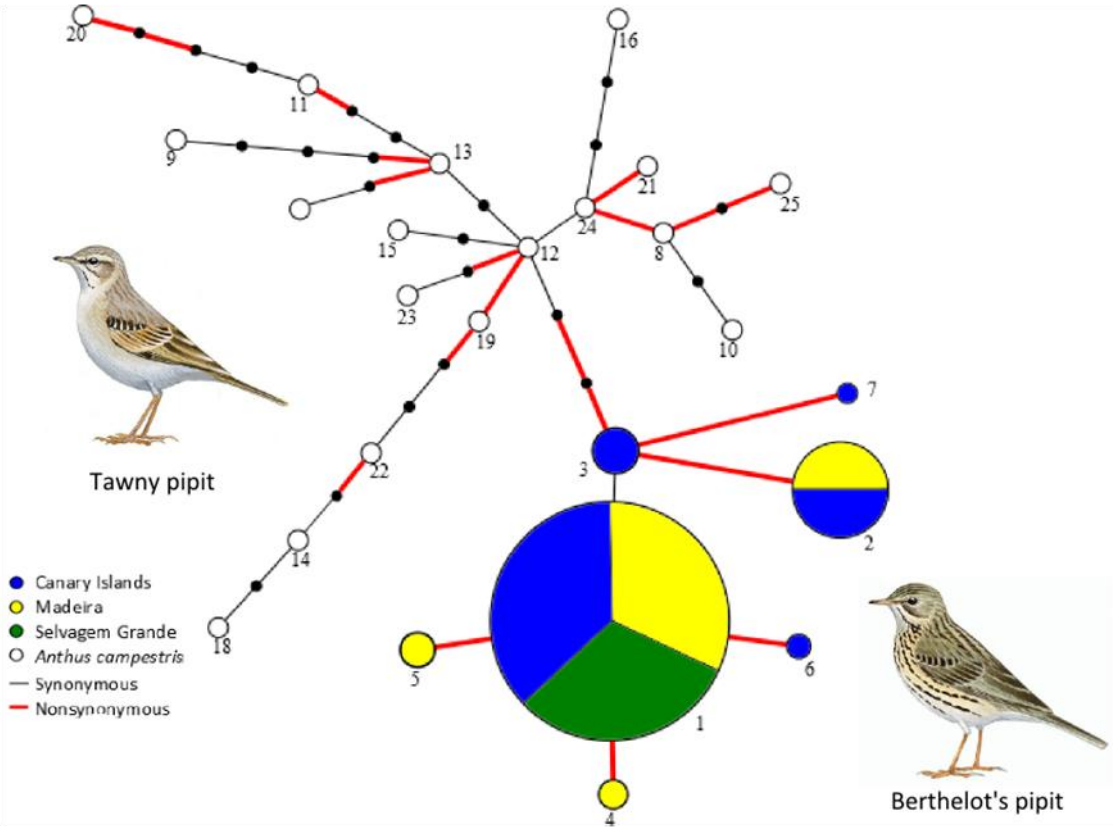
Supplementary Figure S2.1 Haplotype network of TLR1LA



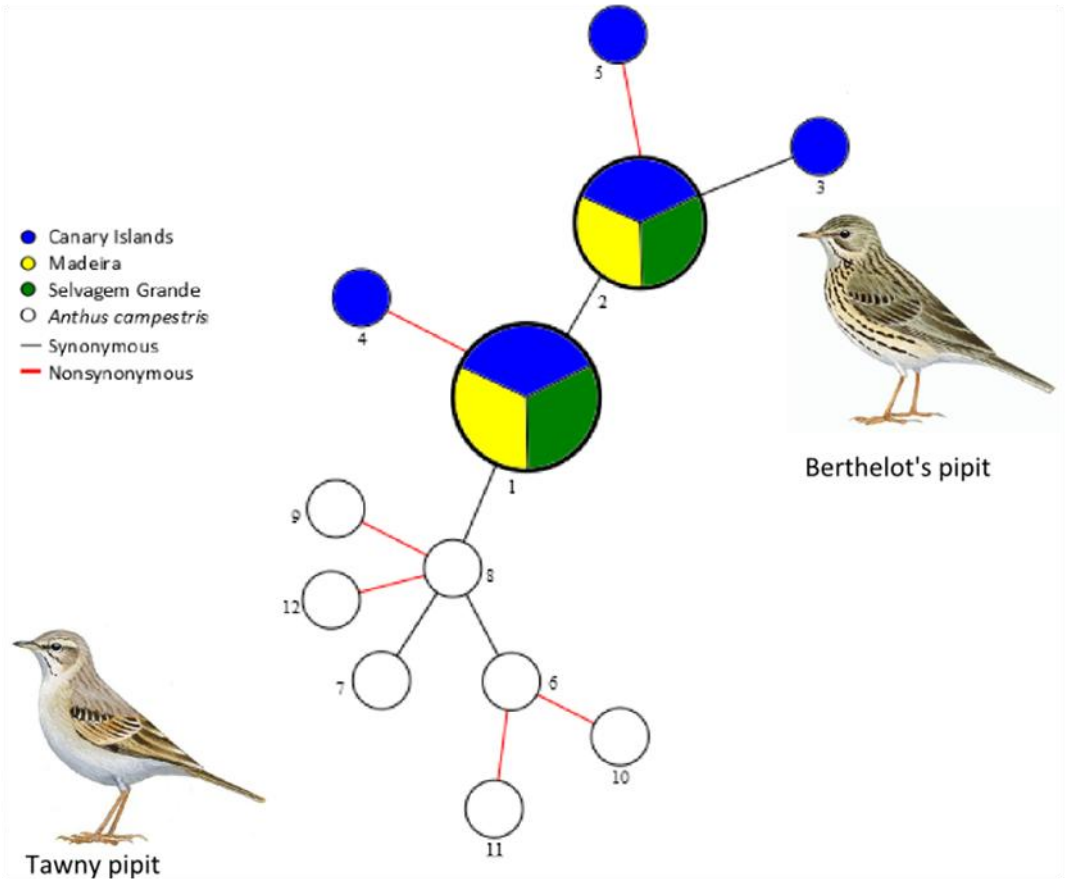
Supplementary Figure S2.2 Haplotype network of TLR1LB



Supplementary Figure S2.3 Haplotype network of TLR3



Supplementary Figure S2.4 Haplotype network of TLR4



Supplementary Figure S2.5 Haplotype network of TLR21

Chapter 3

Predictors of malaria infection in a wild bird population: Landscape level analyses reveal climatic and anthropogenic factors.

A published version of this chapter can be found in *Journal of Animal Ecology*, 83: 1091-1102

Hen and I catching pipits, photo by Karl Phillips

3.1 Abstract

How the environment influences the transmission and prevalence of disease in a population of hosts is a key aspect of disease ecology. The role that environmental factors play in host-pathogen systems has been well studied at large scales, i.e. differences in pathogen pressures among separate populations of hosts, or across land masses. However, despite considerable understanding of how environmental conditions vary at fine spatial scales, the effect of these parameters on host-pathogen dynamics at such scales has been largely overlooked. Here we used a combination of molecular screening and GIS-based analysis to investigate how environmental factors determine the distribution of malaria across the landscape in a population of Berthelot's pipit (*Anthus berthelotii*, Bolle 1862) on the island of Tenerife (Canary Islands, Spain) using spatially explicit models that account for spatial autocorrelation. Minimum temperature of the coldest month was found to be the most important predictor of malaria infection at the landscape scale across this population. Additionally, anthropogenic factors such as distance to artificial water reservoirs and distance to poultry farms were important predictors of malaria. A model including these factors, and the interaction between distance to artificial water reservoirs and minimum temperature, best explained the distribution of malaria infection in this system. These results suggest that levels of malaria infection in this endemic species may be artificially elevated by the impact of humans. Studies such as the one described here improve our understanding of how environmental factors, and their heterogeneity, affect the distribution of pathogens within wild populations. The results demonstrate the importance of measuring fine scale variation - and not just regional effects - in order to understand how environmental variation can influence wildlife diseases. Such understanding is important for predicting the future spread and impact of disease and may help inform disease management programmes as well as the conservation of specific host species.

3.2 Introduction

Understanding how ecological variables influence the prevalence and transmission of disease in a population is a key issue in ecology (Hudson 2002). This is especially important at a time when climate and anthropogenic habitat changes are dramatically affecting the distribution of pathogens and their hosts (Garamszegi 2011), and the number of emerging infectious diseases of wildlife is increasing (Daszak 2000). Establishing how biotic and abiotic factors influence the distribution of disease is important if we are to predict the spatial patterns of future disease threats and effectively manage their impact on biodiversity. Some recent studies have assessed how environmental variables affect pathogen distribution across landscapes in wild populations (Wood *et al.* 2007) and identified relationships have been used to predict where pathogens are likely to survive and/or disease outbreaks to arise (Murray *et al.* 2011). However, such studies have normally been done at coarse scales (but see Eisen & Wright 2001 and Wood *et al*. 2007 for exceptions), thus the effects of fine scale environment variation on the distribution of pathogens has been largely overlooked. Studies at finer spatial scales are now required to provide insight into why pathogens typically have patchy distributions within a landscape.

Haemosporidian parasites, such as species of the genera *Plasmodium*, *Haemoproteous* and *Leucocytozoo*n (hereafter termed malaria for simplicity), are intracellular protozoan blood parasites transmitted by blood sucking invertebrates that occur in every continent apart from Antarctica (Valkiünas 2005). In humans, malaria is a major public health problem with more than 200 million cases and more than one million deaths each year (Chuang & Richie 2012). Malaria also infects other vertebrates including reptiles, turtles, birds and other mammals, reducing their survival and fitness (Martinsen *et al.* 2008).

The vector-transmitted nature of malaria and the need for specific conditions to complete parasite development means that transmission, and therefore patterns of infection, are highly dependent on environmental factors (Guthmann *et al*. 2002). A number of key abiotic variables can affect the distribution of malaria. Temperature has considerable effects (Sehgal *et al.* 2011) because sub-optimal temperatures are associated with slower parasite development (Valkiünas 2005; LaPointe *et al.* 2010) and reductions in vector density (Gilioli & Mariani 2011). Elevation influences malaria through its negative correlation with temperature (Balls *et al.* 2004; LaPointe *et al.* 2010; Lapointe *et al.* 2012). Water availability is important

because of the role it plays in vector larval development (Gilioli & Mariani 2011) and rainfall has been shown to be correlated with malaria prevalence (Galardo *et al.* 2009), as has distance to bodies of water independent of rainfall (Wood *et al.* 2007). In areas where precipitation is low, topography may also play an important role in water accumulation and therefore vector abundance (Balls *et al.* 2004). Topography might also affect vector dispersal and therefore host detection by vectors. Biotic variables, including host life-history traits and demographic factors, may also affect the distribution and transmission of malaria. Malaria prevalence has been shown to be positively associated with host density (Ortego & Cordero 2010), but negatively correlated with non-host species density because of the dilution effect these have on the transmission of the disease (Nah *et al.* 2010).

Anthropogenic activities such as animal husbandry and urbanisation can also affect the prevalence of malaria (Patz *et al*. 2000). One such factor that has been little explored is the influence of livestock on the transmission of malaria. If livestock serve as a reservoir of malaria, their presence might increase prevalence within taxonomically similar local fauna - though this will depend on the host-specificity of the malaria strains involved (Beadell *et al.* 2009). On the other hand, livestock could dilute malaria transmission by providing non-host blood meals to vectors (Nah *et al.* 2010). At another level, livestock farms could increase prevalence by altering the local environment, for example by providing water reservoirs suitable for vector development or by facilitating transmission through aggregations of wild animals attracted to the farms. Urbanisation may also have an impact on malaria prevalence, contingent on the degree of urban adaptation of host species (Bradley & Altizer 2007), and on the extent to which such areas provide suitable habitats for vector development (Guthmann *et al.* 2002).

Assessing and comparing the roles of the environmental variables potentially influencing malaria prevalence is challenging, and their importance may vary between different vector species, strains of malaria, and hosts species. Avian malaria occurs in most bird species (Valkiünas 2005) and can have a heterogeneous spatial distribution even within a single host population (Eisen & Wright 2001; Wood *et al.* 2007; Lachish *et al.* 2011). Infection can have negative effects on individual fitness (Marzal *et al.* 2005; Knowles *et al.* 2010), and is thought to have contributed to the decline and extinction of several bird species (Van Riper *et al.* 1986). Finally, because of their relatively high density, distribution throughout the landscape (often occurring in both pristine and anthropogenic habitats) and the ease with which they can

be sampled, avian species provide excellent models in which to investigate the causes and consequences of malaria transmission in natural systems (Ricklefs *et al.* 2004).

Berthelot's pipit (*Anthus berthelotii*) is a sedentary passerine endemic to the Macaronesian archipelagos. On Tenerife (Canary Islands, Spain, Fig. 3.1) the pipit is numerous and widespread, inhabiting open, semi-arid habitats from sea level up to mountainous habitats at elevations of 3700 m. Importantly it is host to significant levels of a restricted number of avian malaria strains (Illera *et al.* 2008; Spurgin *et al.* 2012) - thus making analyses tractable. This population, therefore, provides an excellent study system with which to assess how ecological and landscape factors influence malaria prevalence in a wild population.

The aim of the present study was to use a combination of molecular disease screening of individuals and fine scale GIS analysis of the habitats in which they were sampled to: (i) determine how the environment modulates the prevalence of malaria, and (ii) assess the relative importance of anthropogenic, natural, biotic and abiotic environmental factors on the presence of malaria in pipits across the landscape.

3.3. Methods

3.3.1. Sampling

To obtain a representative sample of the pipit across its entire range and all environmental gradients on Tenerife, a 1 km^2 grid was laid over a map of the island obtained from Google Earth in ArcGIS version 10 (Esri 2011, Redlands, CA, www.esri.com). Most accessible square kilometres that contained habitat suitable for pipits were visited and whether or not pipits were present was recorded; where present, an attempt was made to catch at least one pipit per km^2 using clap nets baited with *Tenebrio molitor* larvae. Each captured bird was ringed and blood samples were taken by brachial venipuncture and stored in 100% ethanol in screw cap micro-centrifuge tubes at room temperature. We made every effort to ensure our sampling represented as wide an area of the island as possible and that the spatial coverage was not temporally confounded, hence avoiding potential spatial bias in our detection of malaria due to the epidemiology of the disease. The effect of sampling day on the presence of malaria was tested prior to subsequent analyses, and no significant effect was found.

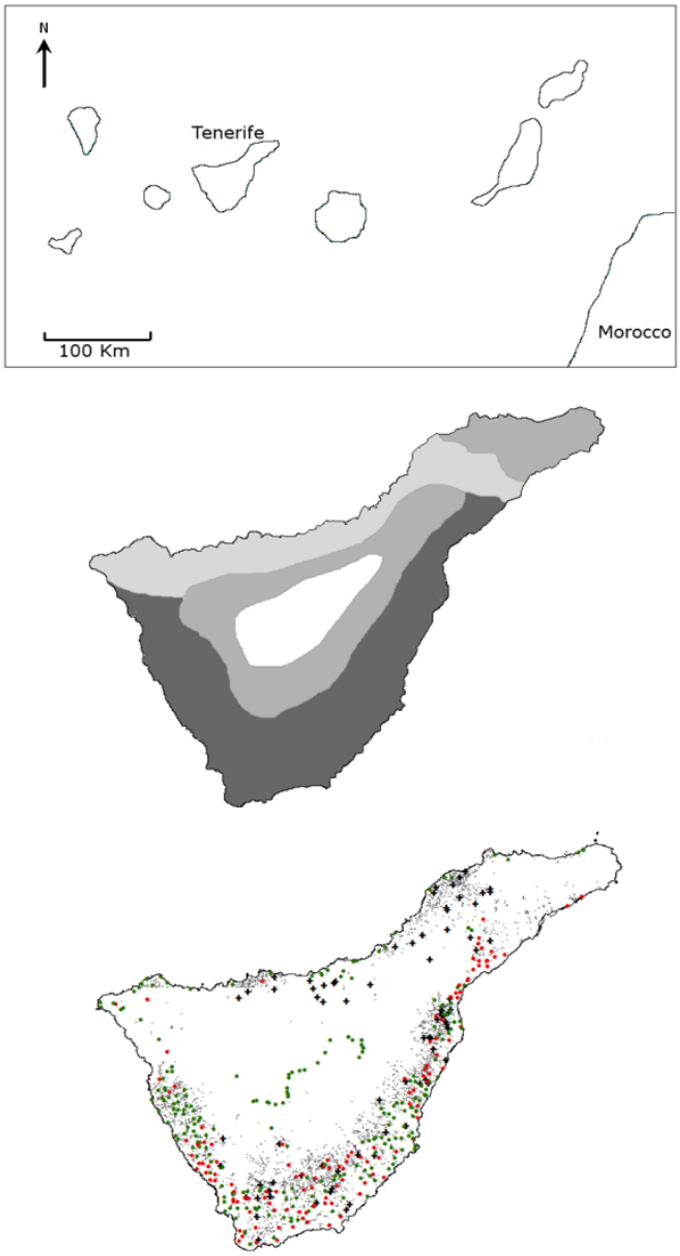


Figure 3.1 Location of Tenerife in the Atlantic Ocean, and Tenerife map insets. The top inset shows the major zonation of habitats categorized by topography and rainfall regime: dark grey = dry areas, generally found from sea level up to 1200 metres above sea level (masl); medium grey = forests at elevations ranging from 1200 to 2400 masl; light grey = wetter areas ranging from sea level to 1200 masl; and white = Teide, a high elevation habitat mainly composed of open areas and some scattered shrubs, from 2400 to 3700 masl. Note that this classification is only for illustrative purposes and a higher resolution subdivision of this habitat classification was used in the analyses (see methods). The lower inset shows the locations where pipits were caught, red dots indicate pipits infected with *Plasmodium* LK6, and green dots show healthy individuals. Crosses indicate locations of poultry farms and grey circles represent the artificial water bodies accounted for in this study.

3.3.2. Molecular procedures

Genomic DNA was extracted from blood using a salt extraction method (Richardson *et al*. 2001). The sex of each bird was determined by polymerase chain reaction (PCR) as described in Griffiths *et al*. (1998). Only DNA samples that successfully amplified the sex specific markers were used in the malaria screening. Avian malaria was screened using the nested PCR method described by Waldenström *et al.* (2004). All samples were screened at least twice and only samples that amplified twice and were verified as malaria through sequencing were taken to be infections. PCR products were sequenced using BigDye terminator reaction kit (Perkin Elmer Inc. Waltham, MA, USA) and products were run on an automated sequencer (ABI PRISM 3700, Applied Biosystems, Carlsbad, USA). Sequences were visually checked and aligned using BIOEDIT version 7.0.9 (Hall 1999) to sequences from the National centre for Biotechnology Information (NCBI) GenBank database and the MalAvi database for avian malaria (Bensch *et al.* 2009).

3.3.3. Environmental variables

Environmental variables were selected on the basis of their hypothesised importance (gleaned from the literature) in explaining malaria prevalence and, based on the most likely associated mechanism (Table 3.1), assigned to one of four different categories as follows: (1) Natural abiotic: minimum temperature of the coldest month (MINTEMP), precipitation (PRECIPITATION), aspect (ASPECT), slope (SLOPE) and altitude (ALTITUDE), (2) Natural biotic: vegetation type (VEGTYPE) and pipit density (DENSITY), (3) Anthropogenic abiotic: distance to nearest artificial water reservoir (DISTWATER), distance to urban site (DIST_URB) and distance to nearest livestock farm (DISTFARM), and (4) Anthropogenic biotic: distance to nearest poultry farm (DISTPOUL). Prior to selecting MINTEMP preliminary analyses were performed with different temperature metrics by exploring their effects using single-predictor models and in full multi-predictor models (see 'statistical analyses' section below) in order to select the one that is most predictive. DISTWATER was investigated as a potential predictor of malaria because in Tenerife rainfall is very scarce and natural permanent bodies of water are almost non-existent. Artificial water reservoirs are, therefore an important source of water for the island population, and are abundant and widespread across the island. In total, we accounted for 14184 artificial water reservoirs with areas ranging from 0.8 to 130084 m^2 (mean=338.7 m^2). DISTFARM was calculated as an alternative predictor to DISTPOUL in order to separate the effects of farms into the abiotic effects derived from the farming infrastructure (such as

presence of standing water and warm environments that promote vector development and survival) from the biotic effects of poultry as potential reservoirs for avian malaria.

3.3.4. GIS analyses

All environmental variable calculations and resampling were carried out in ArcGIS version 10 and R (R Development Core Team 2011). The environmental variables MINTEMP, ALTITUDE, SLOPE, ASPECT, PRECIPITATION, VEGTYPE and DENSITY, were calculated within 50 m, 100 m and 200 m radii buffers around each point where a bird had been caught in order to carry out a sensitivity analysis into the scale-dependence of their effects (see below). In each case an area-weighted mean was calculated within each buffer. Climatic variables (MINTEMP and PRECIPITATION) were obtained from the WorldClim database (Hijmans *et al*. 2005) at a resolution of 30 arc seconds (1 km). Topographic variables (ALTITUDE, SLOPE, ASPECT) at a resolution of 90 m were calculated from digital elevation models obtained from the Shuttle Radar Topography Mission Digital Elevation Database version 4.1 (Consortium for Spatial Information, www.cgiar-csi.org). Vegetation data were obtained from GRAFCAN (Cartográfica de Canarias S.A., www.grafcan.com (Del-Arco *et al*. 2006) and were used to calculate proportional areas of each of five categories of VEGTYPE (forest, grass, shrub, rock associated and urban associated vegetation) within each buffer and each bird was assigned the majority vegetation type within the surrounding buffer.

Distance variables (DISTWATER, DIST_URB, DISTFARM, and DISTPOUL) were calculated by overlaying the layer for pipit location points over polygon layers for: artificial water reservoirs, urban areas, and the position, species and census of livestock farms from the government of Tenerife (http://www.tenerife.es/planes/). For each variable the 'proximity' tool of the analysis extension of ArcGIS 10 was used to calculate the distance to the nearest relevant feature for the variable concerned. Right-skew in all our distance variables was successfully removed using a $\log_{10}$-transformation.

An index of pipit density (DENSITY) was calculated as the number of pipits per square Kilometre, based on our geo-referenced records of pipit presence, using the 'density' tool of the spatial analyst extension in ArcGIS 10, with a neighbourhood size of 2500 m radius around the centre of each square Kilometre sampling cell. This index was used to reflect the size of the subpopulation of pipits found in the same area as the sampled pipit and thus to provide a measure of the local conspecific host population.

Table 3.1 The variables explored in relation to malaria infection in Berthelot's pipits on Tenerife, with a summary of biological mechanisms that explain their potential relevance. Note that the distance variables were log transformed prior to fitting models.

Predictor	Abbreviation	Units (range)/possible values	Mechanisms relevant to malaria infection
Altitude	ALTITUDE	Meters (0-2338)	Parasite rate of development and vector abundance are lower at higher elevations.
Precipitation	PRECIPITATION	Millimeters (254-567)	Water availability is important for vector development at larval stage.
Minimum temperature of the coldest month	MINTEMP	Degrees Celsius (1-14.6)	Parasite rate of development and vector abundance are higher at warmer temperatures.
Aspect	ASPECT	N, NW, W, SW, S, SE, E, NE	Orientation of a surface influences temperature and persistence of temporal puddles that might form after rainfall. It may also affect vector dispersion.
Slope	SLOPE	Degrees (0.2-35)	Drainage is greater on steep terrain, which could affect the availability of water and vector dispersion.
Distance to nearest artificial water reservoir	DISTWATER	Meters (0-5209)	Natural water bodies are very scarce on Tenerife, with artificial water pools providing the main source of still water for vector development.
Distance to nearest poultry farm	DISTPOUL	Meters (27-20853)	Avian malaria can infect poultry. The LK6 strain prevalent in pipits is a generalist strain that could potentially infect poultry.
Distance to nearest livestock farm	DISTFARM	Meters (82-10767)	Livestock farms facilitate atypical aggregations of host birds and can also provide habitats for vectors.
Distance to urbanized sites	DIST_URB	Meters (0-2326)	Urbanized sites might provide small artificial habitats (still water, drains etc) for vector larvae development.
Pipit density	DENSITY	Individuals per km^2 (0.05-1)	Host density affects the transmission dynamics of infectious diseases.
Vegetation type	VEGTYPE	Forest, rock-associated, urban-associated, grass, shrub	Different types of vegetation vary in their provision of suitable habitat for both host and vector.

3.3.5. Statistical analyses

The relative importance of natural abiotic, natural biotic, anthropogenic abiotic, and anthropogenic biotic predictors in influencing prevalence of malaria was assessed using both non-spatial binomial generalized linear models (GLM), and spatial autologistic models (Augustin *et al.* 1996). We implemented a model selection approach (Burnham & Anderson 2001) to compare the relative fit of competing models, or sets of models, using Akaike's information criterion (AIC) as the measure of model fit. Model selection is a valuable alternative to traditional null hypothesis testing, is considered more robust than stepwise approaches (Johnson & Omland 2004; Whittingham *et al.* 2006; Burnham *et al.* 2011; Dochtermann & Jenkins 2011), and has increasingly been used in the last few years, especially in disease ecology (Moore & Borer 2012; Manzoli *et al.* 2013). The *a priori* selection of predictors based on the known ecology of the organisms that are the focus of this study ensure that only a biologically meaningful subset of all the possible predictors (just nine) that could have been tested is used. As such, the testing of all combinations of these nine predictors follows the same rationale as used by Whittingham *et al*. (2006) (see for example Stokke *et al.* 2008; Dochtermann & Jenkins 2011). We performed three sets of modelling procedures, nested within each of our two modelling methods (non-spatial binomial GLMs and spatial autologistic models), one for each buffer radius (50 m, 100 m, and 200 m), hence performing a sensitivity analysis of potential scale-dependent effects of buffer radius on our results. For each of our three model sets, distance based environmental variables (DISTWATER, DIST_URB, DISTFARM, and DISTPOUL) remained invariant. The results obtained at these three sampling scales were very similar (Supplementary Table S3.1), therefore we chose to report only the results using the 100-m radius buffer, since this best approximates the territory size of Berthelot's pipit (Juan Carlos Illera *Pers Comm*.).

In each of our three model sets, nested within our two modelling methods, the same series of modelling steps were repeated. First we compared AICs for single-predictor models. Prior to running multi-predictor models, co-linearity between each pair of predictor variables was evaluated using pairwise bivariate correlations in PASW Statistics version 18 (SPSS Inc. 2009, Chicago, IL, USA www.spss.com). ALTITUDE was highly correlated with PRECIPITATION (Pearson's $r = 0.75$, $p < 0.001$) and with MINTEMP (Pearson's $r = -0.98$, $p < 0.001$). PRECIPITATION was also correlated with MINTEMP (Pearson's $r = -0.80$, $p < 0.001$). Since these three predictors fall into the same natural abiotic category, PRECIPITATION and ALTITUDE

were removed on the basis that MINTEMP had the lowest AIC of the three among single-predictor models (Table 3.2). In line with our rationale for *a priori* selection of predictors, biologically meaningful 2-way interactions of these predictors were also explored for their effects on malaria distribution. The interaction terms we considered are listed in supplementary Table S3.2. First we ran models of the biologically meaningful two-way interactions with their main effects and compared the AIC of each with the corresponding model containing only the two main effects. Only the interactions that reduced the AIC by 2 units were included in the final model selection approach as extra predictors. These interactions were: MINTEMP*DISTWATER, MINTEMP*DISTFARM, MINTEMP*DISTPOUL, DISTWATER*DIST_URB, DISTPOUL*DISTFARM and DISTPOUL*DIST_URB (Table S3.2). We ran all possible combinations of the nine predictors and six interaction terms, recorded the AIC, ΔAIC (the difference between the best model's AIC and that of the model in question), Akaike weights (a measure of the relative explanatory value of the model, compared to all possible ones). We considered models with $\Delta AIC \leq 2$ as having sufficient support (Burnham & Anderson 2001). To estimate the relative importance of predictors we performed model averaging on the models with $\Delta AIC \leq 2$. All modelling calculations were performed using the package MuMIn in R (Barton 2013).

Spatial autocorrelation in model residuals is common in biological processes that are geographically structured and may bias model parameter estimates if it is not accounted for because it violates the assumption of independently and identically distributed errors (Dormann *et al.* 2007). For our binomial generalized linear models (GLM), we checked for spatial autocorrelation in the model residuals by calculating Moran's I coefficients at 1000 m distance classes and generating a correlogram using the package ncf in R (Bjornstad 2012). Autologistic regression modelling (Augustin *et al.* 1996) was implemented to account for the observed spatial autocorrelation by including an autocovariate to assume spatial autocorrelation up to a maximum of 1000 m. This autocovariate was calculated following Dormann *et al.* (2007) using the R package spdep (Bivand 2012). Residual spatial autocorrelation was found to be absent from these autologistic models. The R package fmsb (Nakazawa 2012) was used to calculate the Nagelkerke R^2.

3.4. Results

3.4.1. Pathogen screening

In total 388 Berthelot's pipits were sampled between January and April 2011 (Fig. 3.1). Malaria was detected in 156 out of 388 individuals (40.2%). Of these 156 individuals, 14 (9%) were infected with *Leucocytozoon*, while *Plasmodium* was detected in 148 (95%), with six birds (3.8%) infected with both genera. None of the birds was infected with the genus *Haemaproteus.* Three strains of *Plasmodium* were detected; LK6 and LK5 -first described in the Lesser kestrel (*Falco naumanii*) (Ortego *et al.* 2007) - were detected in 139 and seven individuals, respectively, while KYS9 - first isolated from *Culex pipiens* mosquitoes (Inci *et al.* 2012)- was found in two individuals. Two strains of *Leucocytozoon* were detected; REB11 (previously found in several passerine species in Nigeria, Hellgren *et al.* 2007) in 12 individuals and ANBE1 (previously detected in Berthelot's pipits, Spurgin *et al.* 2012) in two. The two genera of malaria detected (*Plasmodium* and *Leucocytozoon*) are transmitted by different types of vectors with different ecological requirements (See for example van Rooyen *et al.* 2013). Furthermore, the vectors for *Plasmodium* LK5 and LK6 are unknown; therefore, we ran preliminary analyses and found that models which included all strains had a lower fit than models that included only *Plasmodium* LK6. For these reasons, only birds infected with the most common strain, *Plasmodium* LK6, which accounted for 139 out of 156 (89.1%) of all infections, were included as infected in the analyses.

3.4.2. Models and spatial analyses

Single-predictor GLMs showed that minimum temperature of the coldest month (MINTEMP) best predicted malaria infection in pipits, followed by distance to nearest poultry farm (DISTPOUL). The GLMs with 2-way interactions and their main effects showed that the interaction of MINTEMP and DISTPOUL best explained malaria infection in this population, closely followed by the interaction of MINTEMP and distance to nearest water reservoir (DISTWATER) (Table 3.2). For autologistic models the relative importance of single predictors remained unchanged after accounting for spatial autocorrelation. However, for autologistic models fitting 2-way interactions, the interaction of MINTEMP and DISTWATER was best fit, followed by the model with the interaction of MINTEMP and DISTPOUL (Table 3.2). Since the autologistic models consistently resulted in a better fit (lower AIC) than the GLMs, the following sections refer only to the results of the spatial autologistic models.

Table 3.2 Summary of generalized linear (non-spatial) and autologistic (spatial) models predicting presence of *Plasmodium* LK6 infection in Berthelot's pipits on Tenerife, fitting single environmental predictor variables on their own or biologically meaningful 2-way interactions and their main effects (where fitting the interaction lowers AIC by ≥2 compared with the model fitting only main effects). AIC, p-value, Nagelkerke R^2 and rank order of predictors based on AIC are shown.

	Nonspatial					**Spatial**				
Predictor	coefficient	R2	p-value	AIC	Rank	coefficient	R2	p-value	AIC	Rank
MINTEMP	0.201	0.078	<0.001	487.6	1	0.167	0.129	0.001	473.9	1
DISTPOUL	-1.111	0.065	<0.001	491.4	2	-0.920	0.122	0.001	476.1	2
ALTITUDE	-0.001	0.057	0.001	493.8	3	-0.001	0.118	0.005	477.3	3
PRECIPITATION	-0.006	0.052	<0.001	495.2	4	-0.004	0.109	0.006	480.1	4
DISTWATER	-0.475	0.024	0.010	503.3	5	-0.444	0.101	0.021	482.5	5
DISTFARM	-0.664	0.020	0.020	504.7	7	-0.605	0.097	0.043	483.7	6
DENSITY	1.150	0.020	0.018	504.6	6	0.688	0.089	0.179	486.1	7
VEGTYPE	15	0.022	1.000	509.8	10	15	0.106	1.000	486.9	8
DIST_URB	-0.195	0.003	0.362	509.4	9	-0.211	0.086	0.346	487.1	9
SLOPE	-0.034	0.006	0.194	508.5	8	-0.019	0.085	0.480	487.4	10
ASPECT	-0.034	0.072	0.930	503.3	5	0.169	0.126	0.673	488.8	11
2-way interaction										
MINTEMP*DISTWATER	0.590	0.138	0.001	473.2	2	0.550	0.183	0.003	460.6	1
MINTEMP*DISTPOUL	0.476	0.138	0.010	472.9	1	0.445	0.176	0.012	462.9	2
MINTEMP*DISTFARM	0.504	0.108	0.028	482.5	4	0.593	0.165	0.018	466.7	3
DISTPOUL*DISTFARM	-2.188	0.110	0.004	481.9	3	-1.854	0.152	0.015	470.7	4
DISTPOUL*DIST_URB	-1.412	0.088	0.011	488.5	5	-1.261	0.139	0.030	474.8	5
DISTWATER*DIST_URB	-1.046	0.053	0.008	498.1	6	-0.858	0.118	0.038	481.4	6

In the multiple-predictor spatial models, the best fit model contained DISTWATER, DISTPOUL, MINTEMP and the interaction of MINTEMP and DISTWATER. Main effects of DISTPOUL and DISTWATER were negatively associated with malaria presence while the main effect of MINTEMP and the interaction of MINTEMP and DISTWATER had a positive effect (Fig. 3.2). ΔAIC was ≤ 2 in 12 other models, all of which contained MINTEMP, DISTWATER, DISTPOUL and the interaction of DISTWATER and MINTEMP (Table 3.3). Model averaging of the top 13 models showed that DISTPOUL, DISTWATER, MINTEMP and the interaction of DISTWATER and MINTEMP all had a relative importance of 1, while the other predictors and interaction terms had a relative importance lower than 0.37 (Table 3.3).

We investigated the relative importance of each predictor individually within the complete candidate set, by calculating the Akaike weight (w_i) of all possible models where the predictor is present. Models containing MINTEMP had a w_i of 0.332, while models containing DISTWATER had a w_i of 0.214 and models with DISTPOUL had a w_i of 0.223. Models for the remaining predictors had a w_i lower than 0.076 (Table 3.4).

The results of all possible autologistic models representing different categories of environmental variables (biotic, abiotic, anthropogenic and natural) are summarized in table 3.5. The best fit model (lowest AIC) was the abiotic model containing the interaction of DISTWATER and MINTEMP.

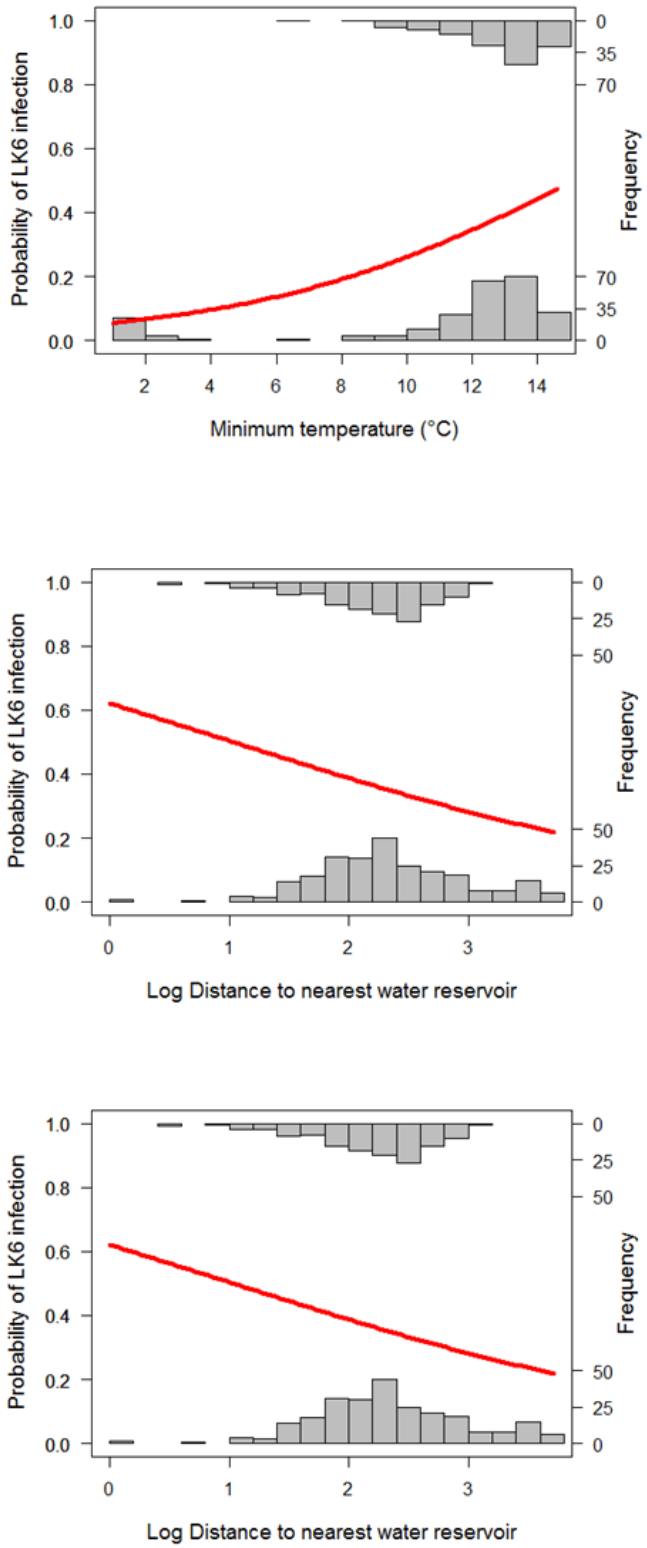


Figure 3.2 Plots of predicted probabilities derived from the logistic regression models of minimum temperature of the coldest month, distance to nearest artificial water reservoir and distance to nearest poultry farm with *Plasmodium* LK6 infection status as the response variable. Histograms show frequency of healthy (lower end) and infected (upper end) individuals at each 100 m distance class.

Table 3.3 Coefficients of predictors with significant p values designated by asterisks (* <0.05, ** <0.01) included in the multi-predictor autologistic regression models with ΔAIC ≤ 2 when compared to the best fit model. Model Nagelkerke R^2 values, averaged coefficients for each variable and their relative importance are shown.

Model rank	(Intercept)	DISTPOUL	DISTWATER	MINTEMP	DIST_URB	DISTFARM	DENSITY	SLOPE	DIST_URB* DISTPOUL	DISTFARM* DISTPOUL	DISTFARM* MINTEMP	DISTPOUL* MINTEMP	DISTWATER* MINTEMP	R2	AIC	ΔAIC
1	17.486 **	-0.694 *	-7.245 **	-1.269 **									0.575 **	0.202	456.60	0.00
2	17.637 **	-0.719 *	-7.530 **	-1.315 **	0.356								0.592 **	0.207	456.96	0.36
3	27.409 **	-3.883	-6.905 **	-2.055 *								0.254	0.547 **	0.206	457.17	0.57
4	27.220 **	-3.811	-7.177 **	-2.074 *	0.349							0.247	0.563 **	0.211	457.59	0.99
5	16.826 **	-0.779 **	-7.177 **	-1.268 **		0.354							0.567 **	0.204	457.84	1.24
6	30.786 *	-0.817 **	-6.214 *	-2.330 *		-5.270					0.433		0.490 *	0.210	457.93	1.33
7	17.808 **	-0.706 *	-7.372 **	-1.318 **			0.401						0.590 **	0.203	458.13	1.53
8	14.317 *	0.290	-7.534 **	-1.319 **	2.250				-0.571				0.594 **	0.209	458.17	1.57
9	17.884 **	-0.669 *	-7.304 **	-1.295 **				-0.018					0.579 **	0.203	458.19	1.59
10	18.017 **	-0.732 *	-7.693 **	-1.373 **	0.372		0.462						0.610 **	0.208	458.34	1.74
11	28.880 **	-4.233	-7.048 **	-2.200 *			0.502					0.280	0.563 **	0.208	458.45	1.85
12	9.549	1.251	-6.928 **	-1.218 **		2.861				-0.767			0.550 **	0.208	458.47	1.87
13	17.147 **	-0.777 *	-7.450 **	-1.309 **	0.318	0.253							0.585 **	0.208	458.59	1.99
Averaged	20.32 *	-1.341	-7.197 **	-1.565 *	0.673	-0.719	0.452	-0.016	-0.57	-0.767	0.433	0.257	0.569 **			
Relative importance		1	1	1	0.37	0.25	0.18	0.06	0.06	0.05	0.07	0.24	1			

Table 3.4 Summed Akaike weights (w_i) of autologistic models that contain each predictor.

Predictor	w_i	Rank
MINTEMP	0.332	1
DISTPOUL	0.223	2
DISTWATER	0.214	3
DISTFARM	0.076	4
DIST_URB	0.044	5
DENSITY	0.038	6
SLOPE	0.034	7
VEGTYPE	0.022	8
ASPECT	0.001	9

3.5. Discussion

By measuring variables at a fine landscape scale in an avian malaria-host system, we found evidence that specific environmental factors influenced the distribution of malaria in a wild population. The minimum temperature of the coldest month (MINTEMP) was the most important predictor of the prevalence of *Plasmodium* LK6 in pipits in Tenerife (birds in locations with a higher minimum temperature were more likely to be infected). Anthropogenic factors, such as distance to artificial bodies of water (DISTWATER) and distance to poultry farms (DISTPOUL) were also shown to be important predictors, both being negatively correlated with presence of malaria. Finally, there was a strong positive effect of the interaction between temperature and distance to artificial water reservoirs on the distribution of malaria in this study system. The model selection approach that confronted models in different categories also suggests that abiotic factors are more important than biotic factors in determining the spatial distribution of malaria in this system.

Natural abiotic factors have been shown to play a major role in the prevalence and transmission of malaria (Van Riper *et al.* 1986; LaPointe *et al.* 2010; Sehgal *et al.* 2011). MINTEMP was the best single predictor of *Plasmodium* LK6 infection in our system, which is

Table 3.5 Summary of separate best-fit multi-predictor autologistic models for the effect of each of four environmental variable categories (biotic, abiotic, natural, and anthropogenic) on the presence of *Plasmodium* LK6 in Berthelot's pipits. We compared all possible models containing variables for each category in turn. Coefficients of predictors with significant p values designated by asterisks (* <0.05, ** <0.01), as well as the AIC and Nagelkerke R^2 of the best fit model for each category are shown.

Model	Rank	(Intercept)	DISTPOUL	DISTWATER	MINTEMP	DISTFARM	DISTFARM* DISTPOUL	DISTWATER * MINTEMP	R2	AIC	ΔAIC
Abiotic	1	14.558 **		-7.020 **	-1.207 **			0.550 **	0.18	460.6	0
Anthropogenic	2	-14.570 *	4.118 *			6.220 *	-1.854 *		0.15	470.7	10.1
Natural	3	-2.932 **			0.167 **				0.13	473.9	13.3
Biotic	4	2.198 *	-0.920 **						0.12	476.1	15.5

consistent with previous studies showing that temperature is positively correlated with the sporogonic development rate of *Plasmodium* parasites (LaPointe *et al.* 2010) and with vector abundance (Van Riper *et al.* 1986). This result is also in accordance with earlier work on pipits in Tenerife indicating a very low prevalence of malaria at high altitudes (> 1600 m, Spurgin *et al*. 2012) with lower minimum temperatures.

Other abiotic natural variables that have been identified as predictors of malaria are the topographic variables of aspect and slope, which affect the presence and persistence of temporary wet habitats that are essential for vector productivity in areas where permanent natural bodies of water are scarce (Balls *et al.* 2004; Githeko *et al.* 2006; Cohen *et al.* 2008), and might also affect vector dispersal. However, in the present study we found no indication that either slope or aspect was important for malaria prevalence. This could be due to the volcanic nature of soils in Tenerife (Fernandez-Caldas *et al.* 1982), which are highly permeable and unlikely to hold water long enough to form temporary puddles of water that would allow for larval development. Furthermore, topography might not be a major factor for the dispersal of the malaria vectors in our study system.

Malaria prevalence has also been shown to be correlated with precipitation levels (Galardo *et al.* 2009; Bomblies 2012), but this is not the case in our study. In Tenerife rainfall is scarce and many artificial water pools have been created for agricultural and other purposes. That DISTWATER is a predictor of malaria infection, suggests that these reservoirs provide suitable habitats for vector larvae development thus facilitating malaria transmission. Previous studies have shown artificial water reservoirs, irrigation canals, and dams are important for the production of malaria vectors (Fillinger *et al*. 2004) and that water bodies are closely associated with malaria in natural systems (Wood *et al.* 2007; Lachish *et al.* 2011). Interestingly, we have also found a positive effect of the interaction between DISTWATER and MINTEMP, suggesting that the positive effect of MINTEMP on malaria infection probability is greater when distance from artificial sources of water is increased. Thus the positive effect of higher minimum temperatures on infection is overwhelmed by the greater influence of DISTWATER as the latter approaches zero. We can only speculate why this may be the case. One possible explanation is that perhaps in warmer areas the vectors of malaria may be better able to survive, disperse and infect birds at a greater distance from the water source they

originate from. Unfortunately, without knowledge of the vectors involved in this system we cannot confirm this assertion.

Various biotic factors may also influence the prevalence of malaria. Vegetation type can have effects on both host and vector (RubioPalis & Zimmerman 1997) but was not found to be an important predictor of malaria infection in our analyses. Berthelot's pipits inhabit open areas and are not found in closed canopy forests, such as the *laurisilva* present in the wetter northern parts of the island. Hence, the low apparent importance of habitat may be a result of focusing on a specific host species, rather than a pattern general to avian malaria. Nevertheless, it is also possible that the vectors of malaria in Tenerife are relatively unconstrained by the vegetation cover and might be equally abundant across areas with different vegetation types. The identification of the vectors of malaria and comprehensive study on their ecology would be needed in order to elucidate this.

Host density is also expected to affect malaria prevalence because it modifies vector-host contact rates. However, while some studies support this prediction (Ortego & Cordero 2010), others, including the present study, fail to find a correlation (Bonneaud *et al*. 2009). It may be that our estimate of pipit density, based on the presence/absence of pipits in a square kilometre grid is not sufficiently accurate. It is possible that pipit density is more highly localized than thought. Furthermore, although adult pipits tend to hold the same breeding territories from year to year (Illera & Diaz 2008), it is unknown whether the spatial structure of pipit density is constant year-round. Patterns of movement and juvenile dispersal, which are not taken into consideration by the present study, could also have important implications for the transmission of infectious diseases such as malaria (Altizer *et al.* 2000). Direct measures of host density may provide a better estimate of the effects that pipit density has on malaria prevalence; however, such measures would be extremely difficult and time consuming to calculate, requiring counts of abundance within each km^2 across the year.

The local density of all vertebrate hosts could potentially have an effect on malaria prevalence. Within host communities, some species act as key hosts harbouring parasitic fauna, thus altering prevalence in other host species (Hellgren *et al*. 2011). The malaria strain detected in the pipits, *Plasmodium* LK6, has also been reported in blackbirds and canaries (Phillips 2009), species that co-occur with pipits in many areas of Tenerife. Unfortunately, we have no data on densities of these two species in order to investigate whether they have an effect on pipit

malaria prevalence. A comprehensive study of the prevalence of avian malaria in the bird community on Tenerife would be needed to understand the role other bird species might play in the prevalence of malaria.

Human activities have been shown to affect vector-borne diseases including malaria (Serandour *et al*. 2007). Factors such as deforestation, animal husbandry, construction and artificial water management modify the ecological balance within which vectors and their parasites develop and transmit disease (Patz *et al.* 2000). Livestock farms have been shown to influence the transmission of infectious diseases by facilitating atypical aggregations of wild birds including infected individuals both of the focal, and other species (Carrete *et al*. 2009). Conversely, such farms have also been shown to dilute the effect of malaria transmission by reducing biting rates on susceptible hosts (Liu *et al.* 2011). In the present study, distance from the nearest poultry farm at which a pipit was caught had a significant negative effect on the probability of malaria infection (Fig. 3.1 shows the locations of poultry farms used in the analyses and *Plasmodium* LK6-infected pipits). That there was no effect of distance to nearest non-poultry livestock farm, suggests that characteristics of poultry farms that differ from those of other livestock farms might be correlated with the transmission of malaria. However, it is difficult to disentangle whether the correlation between distance to nearest poultry farm and malaria is caused by factors - other than the livestock - that are unique to poultry farms (i.e. poultry feed attracting higher densities of birds or poultry attracting ornithophilic vectors) or whether the poultry are themselves reservoirs of the disease. The specific LK6 lineage has not been reported in poultry (though screening for such lineages in poultry has rarely been undertaken), but this strain has been detected in other passerine and non-passerine species (Bensch *et al.* 2009; Phillips 2009), suggesting it is generalist. Moreover, even if the same strain of malaria was detected in the poultry we still wouldn't be able to determine the cause and effect, i.e does the higher prevalence of the disease around poultry farms lead to infection in the fowl, or vice versa. Unfortunately, despite numerous requests, we were unable to obtain permission to sample the fowl within poultry farms.

Other anthropogenic activities, such as urbanisation, can be important predictors of vector-borne diseases (Bradley & Altizer 2007) including malaria (Guthmann *et al*. 2002). For example, mosquito species can quickly adapt to urban environments (Antonio-Nkondjio *et al.* 2011; Kamdem *et al.* 2012) and urbanisation has been shown to increase human malaria prevalence

(Alemu *et al*. 2011). In our analyses, distance to the nearest constructed site was not a good predictor of malaria. It may well be that the broad classification of 'urbanisation' used in this study lacks the resolution to identify the particular characteristics of urbanisation that influence malaria prevalence. Given that urban expansion is happening around the world, further work to understand its impact on wildlife disease prevalence is warranted.

While our models have identified key environmental variables associated with malaria infection, considerable variation (ca. 80%) remains unexplained. This confirms the view that wildlife diseases, such as malaria, are complex and that many different factors, including ones not closely linked to environmental gradients, can influence their spatial distribution within a host population (Hawley & Altizer 2011). It is especially important to note that host related factors, such as individual immunity (and immune variation in the population), host movement patterns and stochastic processes that might influence the epidemiology of malaria were not accounted for in this study. Furthermore, as the specific vectors that transmit avian malaria in Tenerife have not yet been described, we were unable to incorporate an understanding of the ecology of these vectors into our analysis.

One disadvantage of our study is the small sample size in some areas of the island. With larger sample spread across all ecological gradients it is possible that effects of other predictors could have been detected but this does not undermine the importance of the key predictors found.

Assessing the role of the environment in the transmission of pathogens in the wild is crucial to our understanding of disease dynamics and of the causes and consequences of host-pathogen coevolution. The evidence from our study supports previous work which suggests that the prevalence of malaria can vary over small spatial scales (Lachish *et al*. 2011). As with other studies we found that temperature was an important predictor of presence of malaria. We also found that anthropogenic environmental variables, namely proximity to artificial water reservoirs and poultry farms, were also important predictors of malaria in pipits across Tenerife. This may, at least in part, reflect the scale at which the study was performed – when measured across greater scales the influence of locally important predictors of disease may be swamped by regional differences. Given the probable future increase in human population in Tenerife, it is likely that the number of artificial water pools will rise and the intensification of farming will occur, thus increasing the prevalence of avian malaria in the population of Berthelot's pipits.

This study demonstrates the importance of measuring local fine scale variation, and not just regional effects, in order to understand how environmental variation can influence wildlife diseases.

3.6. References

Alemu A, Tsegaye W, Golassa L, Abebe G (2011) Urban malaria and associated risk factors in Jimma town, south-west Ethiopia. *Malaria Journal* **10**, 173.

Altizer SM, Oberhauser KS, Brower LP (2000) Associations between host migration and the prevalence of a protozoan parasite in natural populations of adult monarch butterflies. *Ecological Entomology* **25**, 125-139.

Antonio-Nkondjio C, Fossog BT, Ndo C*, et al.* (2011) *Anopheles gambiae* distribution and insecticide resistance in the cities of Douala and Yaounde (Cameroon): influence of urban agriculture and pollution. *Malaria Journal* **10**, 154.

Augustin NH, Mugglestone MA, Buckland ST (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* **33**, 339-347.

Balls MJ, Bodker R, Thomas CJ*, et al.* (2004) Effect of topography on the risk of malaria infection in the Usambara Mountains, Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **98**, 400-408.

Barton K (2013) MuMIn: Multi-model inference. R package version 1.9.13.

Beadell JS, Covas R, Gebhard C*, et al.* (2009) Host associations and evolutionary relationships of avian blood parasites from West Africa. *International Journal for Parasitology* **39**, 257-266.

Bensch S, Hellgren O, Perez-Tris J (2009) MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. *Molecular Ecology Resources* **9**, 1353-1358.

Bivand R (2012) spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-46.

Bjornstad ON (2012) ncf: spatial nonparametric covariance functions. R package version 1.1-4.

Bomblies A (2012) Modeling the role of rainfall patterns in seasonal malaria transmission. *Climatic Change* **112**, 673-685.

Bonneaud C, Sepil I, Mila B*, et al.* (2009) The prevalence of avian *Plasmodium* is higher in undisturbed tropical forests of Cameroon. *Journal of Tropical Ecology* **25**, 439-447.

Bradley CA, Altizer S (2007) Urbanization and the ecology of wildlife diseases. *Trends in Ecology & Evolution* **22**, 95-102.

Burnham KP, Anderson DR (2001) *Model selection and multi-model inference: a practical information–theoretic approach.* Springer, New York, NY.

Burnham KP, Anderson DR, Huyvaert KP (2011) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* **65**, 23-35.

Carrete M, Serrano D, Illera JC*, et al.* (2009) Goats, birds, and emergent diseases: apparent and hidden effects of exotic species in an island environment. *Ecological Applications* **19**, 840-853.

Chuang I, Richie TL (2012) World Malaria Report 2010: documenting progress towards malaria eradication. *Expert review of vaccines* **11**, 39-41.

Cohen JM, Ernst KC, Lindblade KA*, et al.* (2008) Topography-derived wetness indices are associated with household-level malaria risk in two communities in the western Kenyan highlands. *Malaria Journal* **7**, 40.

Daszak P (2000) Emerging infectious diseases of wildlife - Threats to biodiversity and human health *Science* **287**, 1756-1756.

Del-Arco M, Pérez-de-Paz PL, Acebes JR*, et al.* (2006) Bioclimatology and climatophilous vegetation of Tenerife (Canary Islands). *Annales Botanici Fennici* **43**, 167-192.

Dochtermann NA, Jenkins SH (2011) Developing multiple hypotheses in behavioral ecology. *Behavioral Ecology and Sociobiology* **65**, 37-45.

Dormann CF, McPherson JM, Araújo MB*, et al.* (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**, 609-628.

Eisen RJ, Wright NM (2001) Landscape features associated with infection by a malaria parasite (*Plasmodium mexicanum*) and the importance of multiple scale studies. *Parasitology* **122**, 507-513.

Fernandez-Caldas E, Tejedor-Salguero ML, Quantin P (1982) *Suelos de regiones volcánicas: Tenerife.* Viera y Clavijo, Santa Cruz de Tenerife.

Fillinger U, Sonye G, Killeen GF, Knols BGJ, Becker N (2004) The practical importance of permanent and semipermanent habitats for controlling aquatic stages of *Anopheles gambiae sensu lato* mosquitoes: operational observations from a rural town in western Kenya. *Tropical Medicine & International Health* **9**, 1274-1289.

Galardo AKR, Zimmerman RH, Lounibos LP*, et al.* (2009) Seasonal abundance of anopheline mosquitoes and their association with rainfall and malaria along the Matapi River, Amapi, Brazil. *Medical and Veterinary Entomology* **23**, 335-349.

Garamszegi LZ (2011) Climate change increases the risk of malaria in birds. *Global Change Biology* **17**, 1751-1759.

Gilioli G, Mariani L (2011) Sensitivity of *Anopheles gambiae* population dynamics to meteo-hydrological variability: a mechanistic approach. *Malaria Journal* **10**, 294.

Githeko AK, Ayisi JM, Odada PK*, et al.* (2006) Topography and malaria transmission heterogeneity in western Kenya highlands: prospects for focal vector control. *Malaria Journal* **5**, 107.

Griffiths R, Double MC, Orr K, Dawson RJG (1998) A DNA test to sex most birds. *Molecular Ecology* **7**, 1071-1075.

Guthmann JP, Llanos-Cuentas A, Palacios A, Hall AJ (2002) Environmental factors as determinants of malaria risk. A descriptive study on the northern coast of Peru. *Tropical Medicine & International Health* **7**, 518-525.

Hall TA (1999) BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium Series* **41**, 95-98.

Hawley DM, Altizer SM (2011) Disease ecology meets ecological immunology: understanding the links between organismal immunity and infection dynamics in natural populations. *Functional Ecology* **25**, 48-60.

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

Hellgren O, Waldenström J, Perez-Tris J*, et al.* (2007) Detecting shifts of transmission areas in avian blood parasites - a phylogenetic approach. *Molecular Ecology* **16**, 1281-1290.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.

Hudson PJ (2002) *The ecology of wildlife diseases* Oxford University Press, Oxford.

Illera JC, Diaz M (2008) Site fidelity in the Canary Islands stonechat *Saxicola dacotiae* in relation to spatial and temporal patterns of habitat suitability. *Acta Oecologica-International Journal of Ecology* **34**, 1-8.

Illera JC, Emerson BC, Richardson DS (2008) Genetic characterization, distribution and prevalence of avian pox and avian malaria in the Berthelot's pipit (*Anthus berthelotii*) in Macaronesia. *Parasitology Research* **103**, 1435-1443.

Inci A, Yildirim A, Njabo KY*, et al.* (2012) Detection and molecular characterization of avian *Plasmodium* from mosquitoes in central Turkey. *Veterinary Parasitology* **188**, 179-184.

Johnson JB, Omland KS (2004) Model selection in ecology and evolution. *Trends in Ecology & Evolution* **19**, 101-108.

Kamdem C, Tene Fossog B, Simard F*, et al.* (2012) Anthropogenic habitat disturbance and ecological divergence between incipient species of the malaria mosquito *Anopheles gambiae*. *PloS one* **7**, e39453.

Knowles SCL, Palinauskas V, Sheldon BC (2010) Chronic malaria infections increase family inequalities and reduce parental fitness: experimental evidence from a wild bird population. *Journal of Evolutionary Biology* **23**, 557-569.

Lachish S, Knowles SCL, Alves R, Wood MJ, Sheldon BC (2011) Infection dynamics of endemic malaria in a wild bird population: parasite species-dependent drivers of spatial and temporal variation in transmission rates. *Journal of Animal Ecology* **80**, 1207-1216.

Lapointe DA, Atkinson CT, Samuel MD (2012) Ecology and conservation biology of avian malaria. *Annals of the New York Academy of Sciences* **1249**, 211-226.

LaPointe DA, Goff ML, Atkinson CT (2010) Thermal constraints to the sporogonic development and altitudinal distribution of avian malaria *Plasmodium relictum* in Hawaii. *Journal of Parasitology* **96**, 318-324.

Liu XB, Liu QY, Guo YH*, et al.* (2011) The abundance and host-seeking behavior of culicine species (Diptera: Culicidae) and *Anopheles sinensis* in Yongcheng city, people's Republic of China. *Parasites & Vectors* **4**, 221.

Manzoli DE, Antoniazzi LR, Saravia MJ*, et al.* (2013) Multi-Level Determinants of Parasitic Fly Infection in Forest Passerines. *PloS one* **8**, e67104.

Martinsen ES, Perkins SL, Schall JJ (2008) A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): Evolution of life-history traits and host switches. *Molecular Phylogenetics and Evolution* **47**, 261-273.

Marzal A, de Lope F, Navarro C, Moller AP (2005) Malarial parasites decrease reproductive success: an experimental study in a passerine bird. *Oecologia* **142**, 541-545.

Moore SM, Borer ET (2012) The influence of host diversity and composition on epidemiological patterns at multiple spatial scales. *Ecology* **93**, 1095-1105.

Murray KA, Retallick RWR, Puschendorf R*, et al.* (2011) Assessing spatial patterns of disease risk to biodiversity: implications for the management of the amphibian pathogen, *Batrachochytrium dendrobatidis*. *Journal of Applied Ecology* **48**, 163-173.

Nah K, Kim Y, Lee JM (2010) The dilution effect of the domestic animal population on the transmission of *P. vivax* malaria. *Journal of Theoretical Biology* **266**, 299-306.

Nakazawa M (2012) fmsb: Functions for medical statistics book with some demographic data. R package version 0.3.4.

Ortego J, Calabuig G, Cordero PJ, Aparicio JM (2007) Genetic characterization of avian malaria (Protozoa) in the endangered lesser kestrel, *Falco naumanni*. *Parasitology Research* **101**, 1153-1156.

Ortego J, Cordero PJ (2010) Factors associated with the geographic distribution of leucocytozoa parasitizing nestling eagle owls (*Bubo bubo*): a local spatial-scale analysis. *Conservation Genetics* **11**, 1479-1487.

Patz JA, Graczyk TK, Geller N, Vittor AY (2000) Effects of environmental change on emerging parasitic diseases. *International Journal for Parasitology* **30**, 1395-1405.

Phillips KP (2009) *Prevalence patterns and genetic characterisation of avian malaria in Tenerife: a multi-species approach*, University of East Anglia, Norwich.

R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Richardson DS, Jury FL, Blaakmeer K, Komdeur J, Burke T (2001) Parentage assignment and extra-group paternity in a cooperative breeder: the Seychelles warbler (*Acrocephalus sechellensis*). *Molecular Ecology* **10**, 2263-2273.

Ricklefs RE, Fallon SM, Bermingham E (2004) Evolutionary relationships, cospeciation, and host switching in avian malaria parasites. *Systematic Biology* **53**, 111-119.

RubioPalis Y, Zimmerman RH (1997) Ecoregional classification of malaria vectors in the neotropics. *Journal of Medical Entomology* **34**, 499-510.

Sehgal RNM, Buermann W, Harrigan RJ*, et al.* (2011) Spatially explicit predictions of blood parasites in a widely distributed African rainforest bird. *Proceedings of the Royal Society B-Biological Sciences* **278**, 1025-1033.

Serandour J, Girel J, Boyer S*, et al.* (2007) How human practices have affected vector-borne diseases in the past: a study of malaria transmission in Alpine valleys. *Malaria Journal* **6**, 115.

Spurgin LG, Illera JC, Padilla DP, Richardson DS (2012) Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia* **168**, 691-701.

Stokke BG, Hafstad I, Rudolfsen G*, et al.* (2008) Predictors of resistance to brood parasitism within and among reed warbler populations. *Behavioral Ecology* **19**, 612-620.

Valkiünas G (2005) *Avian malaria parasites and other haemosporidia* CRC Press, New York.

Van Riper C, III, Van Riper SG, Goff ML, Laird M (1986) The epizootiology and ecological significance of malaria in Hawaiian land birds. *Ecological Monographs* **56**, 327-344.

van Rooyen J, Lalubin F, Glaizot O, Christe P (2013) Altitudinal variation in haemosporidian parasite distribution in great tit populations. *Parasites & Vectors* **6**, 139.

Waldenström J, Bensch S, Hasselquist D, Ostman O (2004) A new nested polymerase chain reaction method very efficient in detecting *Plasmodium* and *Haemoproteus* infections from avian blood. *Journal of Parasitology* **90**, 191-194.

Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP (2006) Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology* **75**, 1182-1189.

Wood MJ, Cosgrove CL, Wilkin TA*, et al.* (2007) Within-population variation in prevalence and lineage distribution of avian malaria in blue tits, *Cyanistes caeruleus*. *Molecular Ecology* **16**, 3263-3273.

Supplementary Table S3.1 Summary of single-predictor autologistic models for the 50 m and 200 m radii for the sensitivity analyses.

	50m radius				200m radius			
Model	Predictors	AIC	pval	R2	Predictors	AIC	pval	R2
Altitude	ALT	477.3	0.005	0.118	ALT	477.3	0.005	0.118
Precipitation	PRECIP	482.1	0.001	0.102	PRECIP	482.6	0.023	0.101
Mintemp	MINTEMP	473.8	0.001	0.129	MINTEMP	474.0	0.001	0.129
Aspect	ASPECT	481.1	0.010	0.106	ASPECT	480.6	0.008	0.107
Slope	SLOPE	487.4	0.447	0.085	SLOPE	487.2	0.387	0.085
Water	DISTWATER	482.5	0.002	0.139	DISTWATER	482.5	0.002	0.139
Poultry	DISTPOUL	476.1	0.001	0.123	DISTPOUL	476.1	0.001	0.123
Farm	DISTFARM	483.7	0.008	0.118	DISTFARM	483.7	0.008	0.118
Density	DENSITY	486.2	0.181	0.089	DENSITY	487.0	0.320	0.087
Vegetation	VEGTYPE	486.0	0.982	0.109	VEGTYPE	491.9	0.989	0.097
Construction	DIST_URB	482.9	0.063	0.100	DIST_URB	482.9	0.063	0.099

Supplementary Table S3.2 Biologically meaningful 2-way interactions explored for their possible role on the presence of *Plasmodium* LK6 in Berthelot's pipits. AIC values of autologistic models with the interaction and with only the two main effects are shown. The interactions that reduced the AIC by 2 units are shown in bold and were included in the final model selection approach as extra predictors.

Predictor 1	Predictor 2	AIC only main effects	AIC interaction
MINTEMP	**DISTWATER**	**474.5**	**460.6**
MINTEMP	**DISTFARM**	**475.2**	**466.7**
MINTEMP	DIST_URB	475.9	474.6
MINTEMP	**DISTPOUL**	**468.8**	**462.9**
MINTEMP	SLOPE	475.8	476.2
DISTWATER	**DIST_URB**	**484.5**	**481.4**
DISTWATER	DISTPOUL	476.0	475.6
DISTWATER	DISTFARM	483.1	483.1
DISTWATER	DENSITY	484.1	483.0
DISTPOUL	**DISTFARM**	**477.6**	**470.7**
DISTPOUL	**DIST_URB**	**478.0**	**474.8**
DISTPOUL	DENSITY	476.9	478.8
ASPECT	SLOPE	490.8	501.9
DISTFARM	DENSITY	485.0	486.3
DENSITY	VEGTYPE	486.6	491.4

Chapter 4

454 screening of individual MHC variation in an endemic island passerine

One of the Berthelot's pipits caught in El Teide ready to fly, photo by Karl Phillips

4.1 Abstract

Genes of the major histocompatibility complex (MHC) code for receptors that are central to the adaptive immune response of vertebrates. These genes are therefore an important genetic marker in studies of adaptive genetic variation in the wild. Correct assessment of individual MHC genetic diversity is essential for understanding the evolutionary processes that shape MHC variation. Next generation sequencing (NGS) has increasingly been used in the last decade to genotype the MHC. However, NGS methods are highly prone to sequencing errors, and although several methodologies have been proposed to deal with this, until recently there have been no standard guidelines for the validation of putative MHC alleles. In this study we used the 454 NGS platform to screen MHC class I exon 3 variation in a population of the island endemic Berthelot's pipit (*Anthus berthelotii*). We were able to accurately identify the presence of MHC alleles and thus characterise MHC genotypes across 309 individuals with high levels of repeatability. We were also able to determine alleles that had low amplification efficiencies, whose identification within individuals may thus be less reliable. At the population level, we found lower levels of MHC diversity in Berthelot's pipit than in its widespread continental sister species the tawny pipit (*Anthus Campestris*), and observed trans-species polymorphism. Using the sequence data we identified signatures of gene conversion (including four gene conversion events), and evidence of maintenance of functionally divergent alleles in Berthelot's pipit. We also detected positive selection at 10 codons. The present study therefore shows that we have an efficient method for screening individual MHC variation across large datasets in Berthelot's pipit, and provides data that can be used in future studies investigating spatio-temporal patterns and scales of selection on the MHC.

4.2 Introduction

The vertebrate major histocompatibility complex (MHC) is a family of duplicated genes that code for molecules that detect pathogens and initiate the adaptive immune response (Wakelin 1996; Wakelin & Apanius 1997). There are two main classes of MHC genes: class I codes for molecules that detect intracellular pathogens and present them on the cell surface, whereas class II codes for molecules that recognize extracellular pathogens (Frank 2002). In class I MHC, exon 3 encodes most of the peptide binding region (PBR, Bjorkman *et al.* 1987). The spatial configuration of folds and pockets in the PBR allow each MHC molecule to bind a specific range of peptides (Chelvanayagam 1996). Therefore, amino acid polymorphisms in the PBR should reflect the number of peptides that can be recognized by a cell, and the number and type of pathogens an individual can defend itself against (Potts & Wakeland 1990). However, not all the different variants within MHC genes will generate molecules that bind different sets of pathogens. It has been suggested that MHC variants may be defined into functionally distinct 'supertypes', grouping variants which encode for different sequences of amino acids which, never-the-less, have similar chemical and physical properties and thus similar binding specificities (Sidney et al. 1995).

The interacting effects of various evolutionary processes results in MHC loci displaying the highest levels of genetic variation found in vertebrates (Bodmer *et al.* 1997; Torimiro *et al.* 2006; Mona *et al.* 2008), with some loci, like the human HLA-B, having more than 2,000 alleles (Robinson *et al.* 2013). Variants at the MHC are created by point mutation and gene conversion (Ohta 1995; Edwards & Hedrick 1998; Spurgin *et al.* 2011). Considering the MHC's direct role in initiating immune reactions to pathogens it is not surprising that the high levels of variation, at the individual and population level, are thought to be predominantly maintained by pathogen-mediated balancing selection (PMS; reviewed in Spurgin & Richardson 2010), although sexual selection (reviewed in Edwards & Hedrick 1998) and other mechanisms may also play a role (van Oosterhout 2009). Thus MHC diversity can be extremely important at both the individual and population level.

The characterization of functional MHC alleles and correct assignment of individual genotypes are imperative for understanding patterns of adaptive variation in and among wild populations, for studying host-pathogen co-evolution (Klein *et al.* 1994; Sommer 2005), and potentially, for informing conservation where maximising such variation may be key to population persistence (Ujvari & Belov 2011). However, the presence of multiple gene copies

and the sequence similarity among them (Kelley *et al.* 2005; Cheng *et al.* 2012) prevents the design of locus-specific primers, leading to co-amplification of alleles from multiple loci. Traditionally, MHC genotyping has been done by cloning and subsequent sequencing (e.g. Jarvi *et al.* 2004; Alcaide *et al.* 2008) – a process that is time consuming, especially when applied to ecological-scale datasets. Other methods rely on conformational shifts between different alleles of the MHC which can be separated by gel electrophoresis (Mwenda *et al.* 1997; Baquero *et al.* 2006; Worley *et al.* 2008), but these methods are still time consuming, and, unless they include the direct cloning/sequencing of identified variants, they cannot provide direct sequence information (Promerová *et al.* 2012).

With the introduction of next generation sequencing (NGS) technologies, such as the Roche 454 pyrosequencing platform (Margulies *et al.* 2005), it is now possible to obtain sequences from individual DNA strands, allowing rapid and efficient parallel sequencing of co-amplified alleles. Another advantage of using NGS is the potential for obtaining sequences from a large number of identified individuals in a single run by using 'barcoded' primers. This allows for the subsequent assignment of sequences to individuals during the sequence processing step. Since its introduction in 2005, 454 has been widely used to sequence the MHC in a variety of organisms (e.g. Kloch *et al.* 2010; Sepil *et al.* 2012; Dunn *et al.* 2013). However, NGS techniques are not error free – for example, 454 sequencing is prone to errors, such as insertions, deletions and chimeras generated during the two required amplification steps (Meyerhans *et al.* 1990; Bradley & Hillis 1997), and during the pyrosequencing reaction and base calling (Huse *et al.* 2007; Beuf *et al.* 2012). Consequently, for accurate MHC genotyping it is crucial to distinguish artefacts from real alleles (Galan *et al.* 2010; Sommer *et al.* 2013). Different methods have been proposed to detect sequencing artefacts (Babik *et al.* 2009; Galan *et al.* 2010; Promerová *et al.* 2012; reviewed in Lighten *et al.* 2014a), and these methods generally rely on thresholds of number of reads that a genuine allele is expected to be represented by. However, a problem that these approaches do not address is that alleles differ in their amplification efficiency, meaning that some alleles will be systematically missed from individuals where they are present (Sommer *et al.* 2013). This 'allelic dropout' can inflate homozygosities and deflate individual MHC diversity if it is not accounted for (Sommer *et al.* 2013).

Berthelot's pipit (*Anthus berthelotii*) is a sedentary passerine endemic to 12 islands across the Canary, Madieran and Salvagens archipelagos in the Macaronesian region (Cramp & Perrins

1977, Fig. 4.1). Previous work has shown that the population bottlenecks that occurred during the colonization of each archipelago (Illera *et al.* 2007; Spurgin *et al.* 2014) substantially reduced MHC variation in this species (Spurgin *et al.* 2011). However MHC variation has, at least partially, been regenerated largely by gene conversion (Spurgin *et al.* 2011). Interestingly, the pipit populations are exposed to consistent but spatially varying pathogen pressures both within (Gonzalez-Quevedo *et al.* 2014) and among populations (Spurgin *et al.* 2012). Thus these populations provide an excellent system in which to test different evolutionary hypothesis on the role of pathogen-mediated selection in shaping the patterns of MHC variation at various spatio-temporal scales. Population-level variation at the MHC of Berthelot's pipit has been assessed (Spurgin *et al.* 2011), but individual level screening within populations is needed if we are to investigate the factors that drive MHC variation at different scales. With that as our overall aim, here we test the utility of the methods outlined by Sommer *et al.* (2013) to individually sequence MHC class I exon 3 variation in 310 Berthelot's pipits from Tenerife and from 10 tawny pipits (*A. campestris*), a geographically widespread species that is the closest relative of Berthelot's pipit (Voelker 1999). Using the data generated we then compare the levels of MHC variation found in the two species, and test for signatures of selection.

4.3 Methods

4.3.1 Study species and sampling

We sampled Berthelot's pipits on Tenerife, in the Canary Islands, from January to April 2011. To obtain a representative sample across the pipits range on Tenerife, a 1 km^2 grid was laid over a map of the island obtained from Google Earth in ArcGIS version 10 (Esri 2011, Redlands, CA, www.esri.com). Most accessible square kilometres that contained habitat suitable for pipits were visited and, where present, an attempt was made to catch at least one pipit per km^2 using clap nets baited with *Tenebrio molitor* larvae. Each captured bird was fitted with a unique metal ring from the Spanish Environment Ministry and a ca. 25 µl blood sample was taken by brachial venipuncture and stored in absolute ethanol in a 2 ml screw cap micro-centrifuge tube at room temperature. The final sample of 388 birds included 30 birds caught on the mountain of El Teide on Tenerife (2,500 m above sea level), a separate population which is isolated from the lowlands by dense pine and laurel forest which the pipit does not inhabit (Illera 2007). Ten tawny pipits (two from Morocco, three from Mauritania and five from the Iberian Peninsula (provided by J.C. Illera) were also screened.

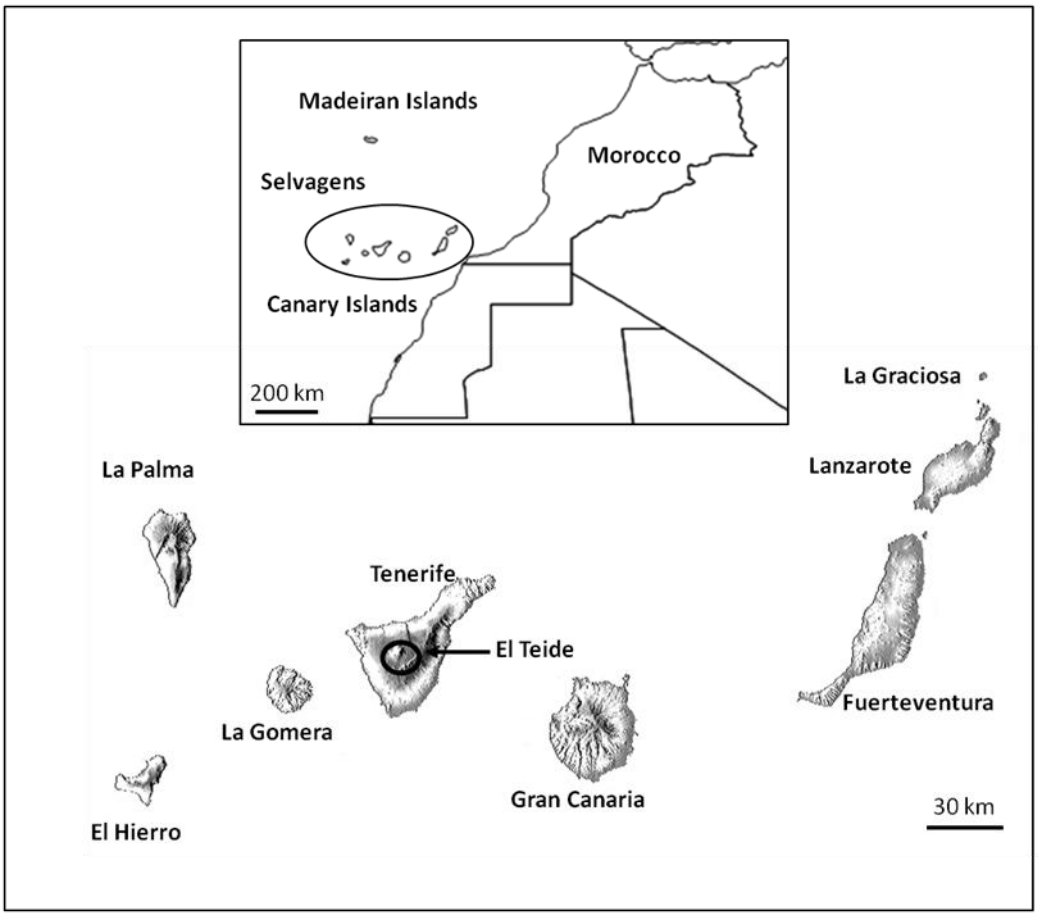


Figure 4.1 Distribution of Berthelot's pipits (*Anthus berthelotii*) in the Macaronesian archipelagos of Madeira, Selvagens and the Canary Islands (Inset), and detail of the nine populations from the Canary Islands.

4.3.2 MHC genotyping

DNA was extracted from blood using a salt method (Richardson *et al.* 2001). Pooling large numbers of individuals in a single run can result in highly uneven representation of individuals (S. Paterson Pers. Comm.), thus samples were pooled in eight pools of 80 samples each. In total, 310 Berthelot's pipits (including 26 individuals from El Teide) out of the 388 originally sampled, and 10 tawny pipits were genotyped at the MHC. We screened each of the 320 samples twice, and each replicate of a given sample was amplified using different combinations of barcoded primers (thus screening a total of 640 amplicons). For the library preparation we used forward and reverse fusion primers consisting of the 454 adaptor (forward adaptor, 5'-CGTATCGCCTCCCTCGCGCCA-3' and reverse adaptor, 5'-

CTATGCGCCTTGCCAGCCCGC-3'), followed by a key sequence (TCAG), a 10-bp multiplex identifier (MID), and the MHC class I exon 3 specific primers (DG2, 5'-TTGCGCTCYAGCTCYTTCTGCT-3' and GENDG, 5'-TCCCCACAGGTCTCCACAC-3'; Spurgin *et al* (2011), Fig. 4.2). The primers used differed only in the MID sequence. MID sequences were obtained from the 10-base extended MID set from Roche Diagnostics (454 Life Sciences Corp. 2009). We chose nine MIDs with at least three base pair differences between them (MID numbers 1, 2, 3, 4, 5, 7, 8, 10 and 11). By using these MIDs in the forward and reverse primers we had 81 possible combinations with which to barcode individuals within the pools of 80 samples. We pooled samples so that the two replicates of a sample were present in different pools and had a different forward and reverse fusion primer combination. Two strategies were implemented in order to reduce the formation of chimeras during the PCR (Lenz & Becker 2008; Holcomb *et al.* 2014). First, each of the 640 amplicons screened was generated by two independent PCRs that were then pooled in equimolar amounts. Independent amplification reactions have been proposed to eliminate PCR bias and random artefacts, because they would only occur in one of the replicate reactions (Kanagawa 2003). The second approach was the reduction of the number of cycles to the minimum number that provided a clear, well defined amplicon when visualised in an agarose gel (27 cycles). This was done because in the later stages of PCR dNTP and primer concentrations are reduced, therefore the incomplete amplicons work as 'primers' by hybridising to the wrong templates due to sequence homology. PCRs were performed in 25 µl volumes containing 0.5 µM of each fusion primer, 12.5 µl of 2X Roche FastStart master mix and *ca.* 60 ng of DNA. Thermocycling consisted of an initial denaturation at 96°C for 4 mins, followed by 27 cycles of 94°C for 30 s, 61°C for 30 s, and 72°C for 60 s, with a final extension of 72°C for 10 mins. PCR products were run on 1.5% agarose gels, and amplicons were cut from the gels using a sterile scalpel and purified using a gel purification kit (Qiagen). Purified products were quantified using the PicoGreen dsDNA assay kit (Life technologies) and pooled in order to get 1.25 ng per sample for each pool of 80 samples. Emulsion PCR and 454 sequencing were conducted at the Centre for Genomic Research at the University of Liverpool. Sequencing was run on a PicoTitre plate (each pool in one eighth of the plate) on a GS FLX Titanium system.

4.3.3 Bioinformatics

High-quality sequences (Phred quality score > 20 at more than 95% bases) with complete forward and reverse MID and primer sequences were assigned to amplicons based on MID

combinations. We then followed the workflow outlined by Sommer *et al.* (2013) to assign reads to putative MHC alleles and to identify artefacts based on the assumptions that: 1) artefacts are less frequent than their source allele, and 2) artefacts should have a lower intra-amplicon frequency than any true allele. The workflow consists of three main steps that classify variants into 'putative alleles', 'putative artefacts', and 'unclassified variants'. In the first step, applied to each amplicon independently, reads were assigned to groups of identical sequences, hereafter referred to as 'clusters'. The most frequent cluster within each amplicon was classified as a putative allele. Sequences represented by only one read were discarded as sequencing artefacts. Clusters were then classified as 'chimera', '1-2 bp difference' or '>2 bp difference' by comparing their sequences to more frequent clusters within the amplicon (full details in Sommer *et al.* 2013). Chimeras were identified using a Python script that checked, within each amplicon, whether a given sequence could, at any point along its length, be formed by joining together the forward section of a more frequent sequence with the reverse section of another more frequent sequence (further details in Appendix 4.1).

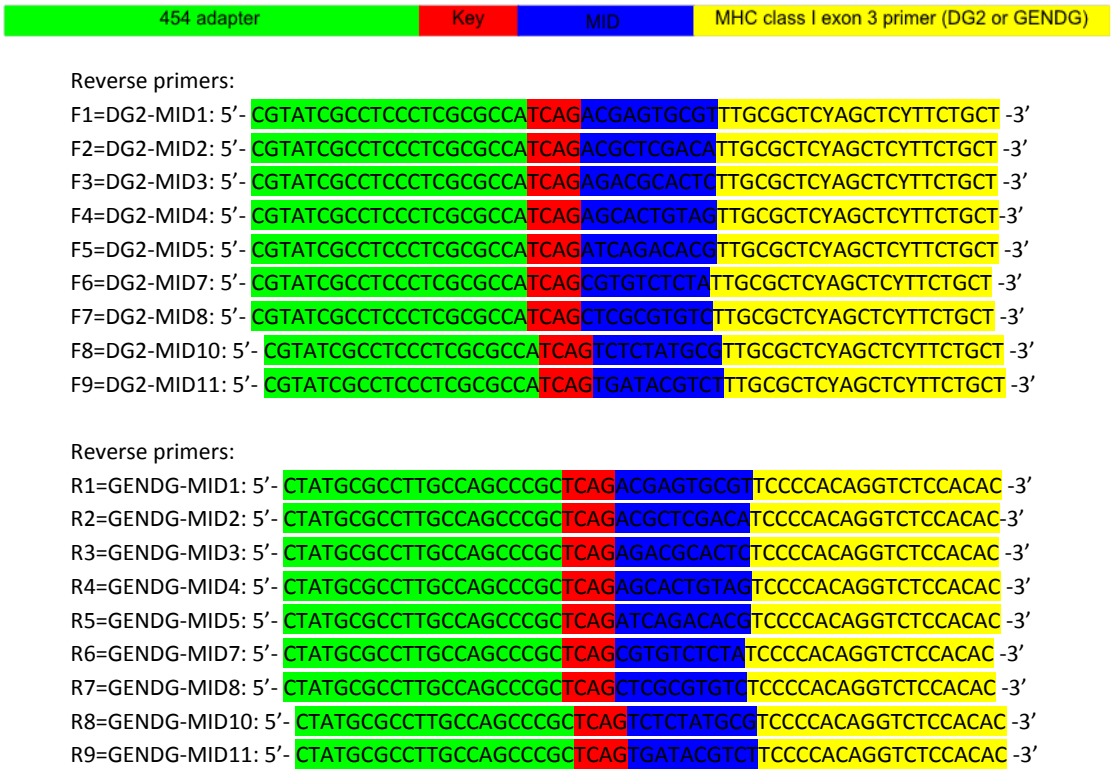


Figure 4.2 Fusion primers used for preparation of MHC class I exon 3 amplicons. Nine forward and nine reverse primers were used in combination to identify each of 80 amplicons.

The second step of the pipeline compares both replicates of a given individual. Clusters listed as 'chimera' or '1-2 bp difference' in step one were classified as putative artefacts if they were absent from the replicate amplicon. Clusters were also classified as putative artefacts if they had been identified as chimeras in both replicates. Clusters classified as '>2 bp difference' were only classified as putative artefacts if, across the whole data set, they were unique to a single amplicon. All other clusters were retained for further checking in step three.

In step three, clusters retained on the '1-2 bp difference' list from step two (i.e. those that were observed in both replicate amplicons) were classified as putative alleles if they had a higher intra-amplicon frequency than any entry in the putative artefact list, and as unclassified variants if this criterion was not met. The same process was used on clusters in the '>2 bp difference' list when the sequence was also present in the sample's replicate. Other clusters in the '>2bp difference' list were labelled as unclassified variants if present as a putative allele in another individual, but as putative artefacts if not. Chimeras retained from step two were classified as putative alleles if the same sequence was present as a putative allele in another individual; otherwise, retained chimeras were labelled as unclassified variants.

After completing the pipeline, we further checked the lists of putative alleles, putative artefacts and unclassified variants for Berthelot's and tawny pipits separately. For Berthelot's pipits we discarded sequences classified as putative alleles that occurred in only one individual, but pulled out clusters from the unclassified variants list that matched sequences in the list of putative alleles. We assessed the frequency of each unclassified variant in the amplified samples (the number of individuals they occurred in out of the total of 310), and further inspected for the presence of any MHC alleles that had been described in the previous, population-level characterisation of MHC variation in Berthelot's pipits (Spurgin *et al.* 2011). We also inspected the sequences in the putative artefacts list by pulling out sequences that matched a putative allele already identified in another individual, and that were also present in both amplicons of a bird with an intra-amplicon frequency higher than the least frequent entry to the list of putative alleles.

For tawny pipits we proceeded differently due to the small sample size. If a variant was classified as a putative allele in both replicates of only one sample it was kept on the putative allele list. If the variant was classified as an allele in only one amplicon but was found in other amplicons as an unclassified variant, it was also treated as a putative allele. Finally, for

unclassified variants, we pulled out, and classified as alleles, sequences that were present in both replicates of at least two birds. For both species, repeatability of genotyping was calculated as the percentage of shared alleles between the two replicate amplicons of each individual.

4.3.4 Allele amplification efficiency

We followed Sommer *et al's* (2013) rationale, which assumes that the amplification efficiency of an allele is independent of the genotype and similar among PCR products with the same conditions, to estimate the relative amplification efficiency of each allele. We used scripts provided in Sommer *et al.* (2013) to perform this calculation in R (R Development Core Team 2011). We standardised allele amplification efficiencies relative to ANBE11 and ANCA17 for Berthelot's and tawny pipits, respectively. The choice of standardising alleles is arbitrary, as our use of degenerate primers means we cannot know which allele is the 'best' amplifier in our data set. We also calculated a variant of Galan's T1 (Galan *et al.* 2010), which uses the lowest amplification efficiency to estimate the minimum number of reads per amplicon necessary to reach a coverage of 99.9% for a genotype with a given number of alleles (in our case, 12 – the maximum number of alleles observed in an individual) and with a minimum number of 2 reads per allele. This calculation was done using the R function 'T1.min.efficiency.replicated' provided in Sommer *et al.* (2013). Any sample that had a number of reads lower than the T1 threshold was discarded and not used in downstream analyses.

4.3.5 MHC sequence analyses

Using the software DnaSP 5.10.01 (Librado & Rozas 2009), we calculated the number of nucleotide differences and nucleotide diversity among sequences for the set of alleles identified in each species. In order to investigate the mutation to recombination ratio among MHC alleles we calculated the recombination (R_m) and mutation (θ) parameters (Hudson & Kaplan 1985) and obtained the 95% confidence interval using a coalescent approach with 10,000 replications. To compare allele diversity between the two species we calculated pairwise nucleotide distance using the Nei-Gojobori/Jukes-Cantor method (Nei & Gojobori 1986) in MEGA 6 (Tamura *et al.* 2011), and assessed differences using a Mann-Whitney U test.

To explore phylogenetic relationships among Berthelot's and tawny pipit MHC class I alleles we built a neighbour net with Jukes-Cantor distance between all pairs of alleles using the software SplitsTree 4.13.1 (Huson & Bryant 2006). We explored the presence of gene conversion tracts

in the MHC class I alleles identified using the following methods: 3Seq (Boni *et al.* 2007), GENECONV (Padidam *et al.* 1999), MaxChi (Smith 1992), Chimaera (Posada & Crandall 2001) and SiScan (Gibbs *et al.* 2000), all implemented in the software RDP4 (Martin *et al.* 2010). The highest acceptable *P*-value was set to 0.05, and 100 permutations were performed for all methods. Tracts identified by at least two methods were considered true recombination events.

The number of putative functionally different MHC class I alleles was estimated based on the amino acid sequences. The codons involved in the peptide binding region (PBR) of the MHC class I of Berthelot's pipits have been identified previously (Spurgin *et al.* 2011) based on the sites known to code for the PBR in humans (Brown *et al.* 1993). Using the Berthelot's pipit sequences, the rate of non-synonymous (d_N) and synonymous (d_S) substitutions per site was calculated in MEGA 6 (Tamura *et al.* 2011) using the Nei-Gojobori/Jukes-Cantor method (Nei & Gojobori 1986) for: 1) the full exon sequence, 2) the non-peptide binding region (non-PBR), and 3) the peptide binding region (PBR). Differences between d_N and d_S were assessed with Mann-Whitney U tests. We did not perform these analyses on tawny pipits because the small sample size for this species means that we have likely underestimated the number of alleles.

In order to identify codon-specific signatures of positive selection at the MHC class I across the pipit species, four codon-based methods to detect selection based on the d_N and d_S were implemented in the webserver Datamonkey (http://datamonkey.org, Pond & Frost 2005a). The fixed effects likelihood (FEL), random effects likelihood (REL) (Pond & Frost 2005b), and fast unconstrained Bayesian approximation (FUBAR) (Murrell *et al.* 2013) were used to detect codons under pervasive selection. In addition, the mixed effects model of evolution (MEME) (Murrell *et al.* 2012) was used to detect codons under episodic diversifying selection. Sites with Bayes factor > 50 for REL, posterior probabilities > 0.9 for FUBAR and *P*-values < 0.1 for FEL and MEME were considered to have enough support for positive selection. Only sites that were detected to be under positive selection by at least two different methods were considered to be candidates of evolution under positive selection. Prior to running analyses, the best fitting nucleotide substitution model was determined using a model selection approach. All sequences identified from both Berthelot's and tawny pipits were used in this analysis.

We explored whether MHC alleles found in Berthelot's pipits could be clustered into 'supertypes' (Doytchinova & Flower 2005) according to the antigen-binding characteristics of

either the amino acids in the PBR (fifteen amino acids), or the amino acids that were detected as positively selected sites (PSS) by our analysis. Five descriptors were obtained for each amino acid (Sandberg *et al.* 1998): z1 (hydrophobicity), z2 (steric bulk), z3 (polarity), z4 and z5 (electronic effects). Amino acid descriptors were arranged in a matrix where each row represented one unique PBR or PSS sequence and the columns represented the five descriptors for each amino acid in the region being analyzed. We performed a *k*-means clustering algorithm to identify the most likely number of clusters of alleles based on the amino acid descriptors using the function 'find.clusters' in the 'adegenet' package (Jombart 2008; Jombart *et al.* 2010) in R. The algorithm was run four times for different numbers of clusters from one to the total number of unique sequences, and for each run a mean Bayesian information criterion (BIC) was obtained. The most likely number of clusters in the data is the one with the lowest BIC. After identifying the optimal number of clusters, a discriminant analysis of principal components (DAPC) is used to identify the alleles in each cluster (Jombart *et al.* 2010).

4.4 Results

4.4.1 MHC allele identification

We obtained a total of 1,019,897 high quality sequences ranging from 53 to 4,308 reads per amplicon (mean ± SD = 1436 ± 652). Of these, 919,046 reads had complete forward and reverse MIDs and correct primer sequences, leaving 45 - 4,049 reads per amplicon (1427 ± 647) and 744 - 6,263 reads per individual (2855 ± 954).

At the end of the bioinformatics processing, 41 clusters were classified as putative alleles, of which 31 were assigned to Berthelot's pipits and 10 to tawny pipits. The list of unclassified variants contained the greatest number of entries, followed by putative artefacts and putative alleles (Table 4.1). We discarded 11 of the 31 putative alleles detected in Berthelot's pipits because they were classified as alleles in only one individual. Since these variants were not confirmed across individuals, we believe these must be either artefacts or very low frequency alleles. We determined that the more conservative approach would be to discard them, as even if they were true alleles, at such low frequencies (*ca.* 0.3%) they would have little consequence in downstream analyses. We identified 588 sequences as unclassified variants, of which 134 were present in more than two samples. Among these sequences we found two that occurred in both amplicons of more than 80% of birds and were considerably more

frequent in the 310 Berthelot's pipit samples than all the other unclassified variants (Fig. 4.3), but had a very small number of reads in most amplicons (mean ± S.E.: ANBE3 = 24.5 ± 0.8; ANBE31 = 18.0 ± 0.6). These two variants – ANBE3 and ANBE31 – also matched sequences identified by Spurgin *et al.*'s (2011) previous population level study on MHC in Berthelot's pipit. We treated ANBE3 and ANBE31 as 'low efficiency alleles' (see amplification efficiency results below) but because we could not be certain of their absence/presence in all individuals we recommend they are excluded from future individual based MHC disease association studies in Berthelot's pipit. However, we retained these alleles when assessing sequence-level selection, as their amplification efficiency should not bias such analyses.

Table 4.1 Number of clusters classified as chimera, 1-2 bp difference, >2 bp difference, putative artefact, putative allele and unclassified variant in each of the three bioinformatics steps to classify reads from Berthelot's pipits MHC class I exon 3 sequences. In step 1, variants are classified as putative alleles if they are the most common sequence in an amplicon, and as artefacts if they are singletons. The remaining clusters are classified as chimeras, 1-2 bp difference or >2 bp difference compared to the most similar cluster. Variants in each of these three classes are subject to steps 2 and 3 and further classified as putative artefact, putative allele or unclassified variant. Note that alleles can be identified in different bioinformatic steps in each individual.

	Step 1	**Step 2**			**Step 3**		
Variant class		chimera	1-2 bp diff	>2 bp diff	chimera	1-2 bp diff	>2 bp diff
chimera	2,023						
1-2 bp difference	24,631						
>2 bp difference	14,583						
Putative artefact	17,408	1,903	21,960	3,863			5,800
Putative allele	640				29	51	611
Unclassified					91	2,620	4,309

After processing the unclassified variants we were left with 22 alleles identified for the Tenerife population of Berthelot's pipits. Of these, seven (named ANBE43-ANBE49) had not been previously identified in this species, and have been deposited in GenBank (Accession numbers KM593305 – KM593311). We also detected seven alleles (ANBE1, ANBE6, ANBE7, ANBE9, ANBE13, ANBE31 and ANBE38) that Spurgin *et al.* (2011) found in other populations of

Berthelot's pipits but not in Tenerife. Among these, ANBE31 had previously been found only on Lanzarote (another Canary Island), and ANBE38 had only been found on Selvagem Grande (one of the islands of the Selvagens archipelago). We failed to find five alleles that had been previously reported from an earlier, smaller sample of birds (30 individuals sampled in 2006) from each of the Tenerife and Teide populations (Spurgin *et al.* 2011): ANBE12 had been detected on El Teide, and Fuerteventura (one of the Canary Islands); ANBE24 had been detected previously in the low lands of Tenerife and in other islands of the Canary archipelago; ANBE19 and ANBE39 had been detected in the low lands of Tenerife and on El Teide, and ANBE41 was restricted to the population of El Teide.

Intra-amplicon frequencies (proportion of reads within an amplicon) of the 22 alleles identified in Berthelot's pipits varied from a low of 0.013 for ANBE31 (S.E. = 0.0003) to a high of 0.192 for ANBE10 (S.E. = 0.003, Fig. 4.4a). Population allele frequencies ranged from 0.99 for ANBE7 to 0.01 for ANBE28 (Fig. 4.5). The number of alleles per individual ranged from 4 to 12, with a mode and median of 8 alleles (Fig. 4.6), suggesting the potential presence of six loci.

After processing unclassified variants for the tawny pipit we were left with 28 clusters, of which 11 were classified as putative alleles and 17 as unclassified variants. One allele, ANBE9, was shared between Berthelot's and tawny pipits. The other 27 alleles, unique to the tawny pipit, were named ANCA1 to ANCA27 and their sequences have been deposited in GenBank (Accession numbers KM593312 – KM593338). Intra-amplicon frequencies for these 28 alleles ranged from a low of 0.015 (S.E. = 0.003) for ANCA20 to a high of 0.142 (S.E. = 0.039) for ANCA1 (Fig. 4.4b). The number of alleles per individual in tawny pipits ranged from 6 to 11 (mean = 7.4, median = 7, Fig. 4.6), which also suggests the presence of six loci.

The repeatability (i.e. the mean percentage of alleles shared between the two replicates of the same sample) of our genotyping was 96.1%. (S.E. = 5.45). The lowest repeatability (44.4%) was obtained for a sample with an amplicon with only 45 reads. Fourteen samples had repeatabilities lower than 80% and 240 samples had a repeatability of 100%.

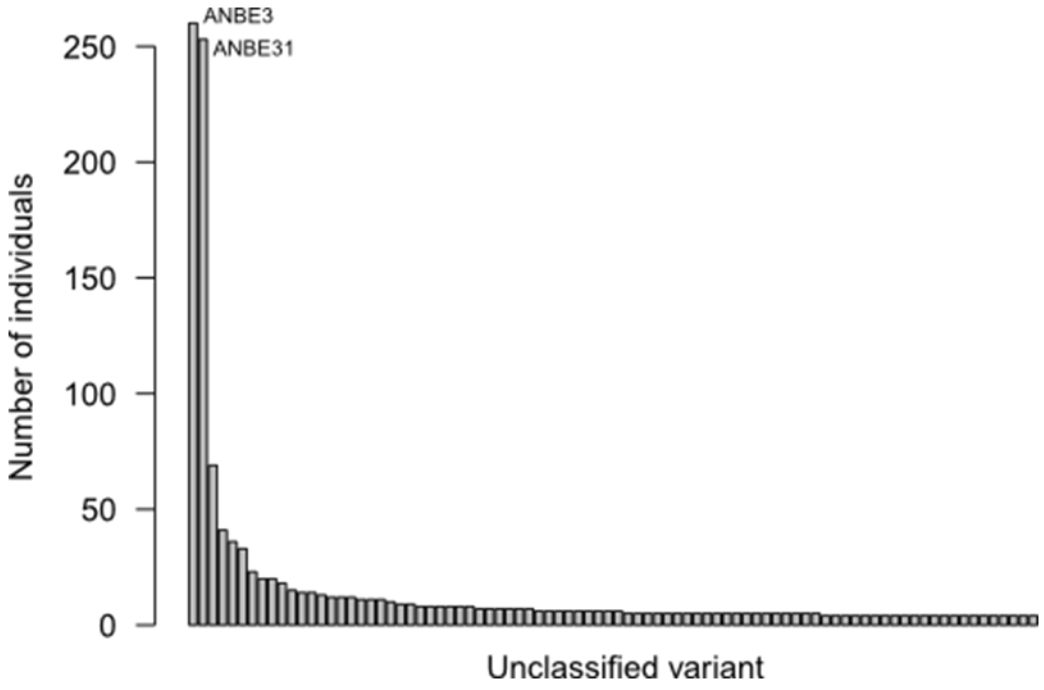


Figure 4.3 Frequencies of unclassified variants identified after the bioinformatics processing of MHC class I exon 3 sequences obtained from 310 Berthelot's pipits (*Anthus berthelotii*) and 10 tawny pipits (*A. campestris*). For simplicity only the unclassified variants found in four or more samples are shown. The two most common unclassified variants matched two alleles previously described in Berthelot's pipits (Spurgin *et al.* 2011), ANBE3 and ANBE31, detected in both replicates of 260 and 253 Berthelot's pipit samples, respectively.

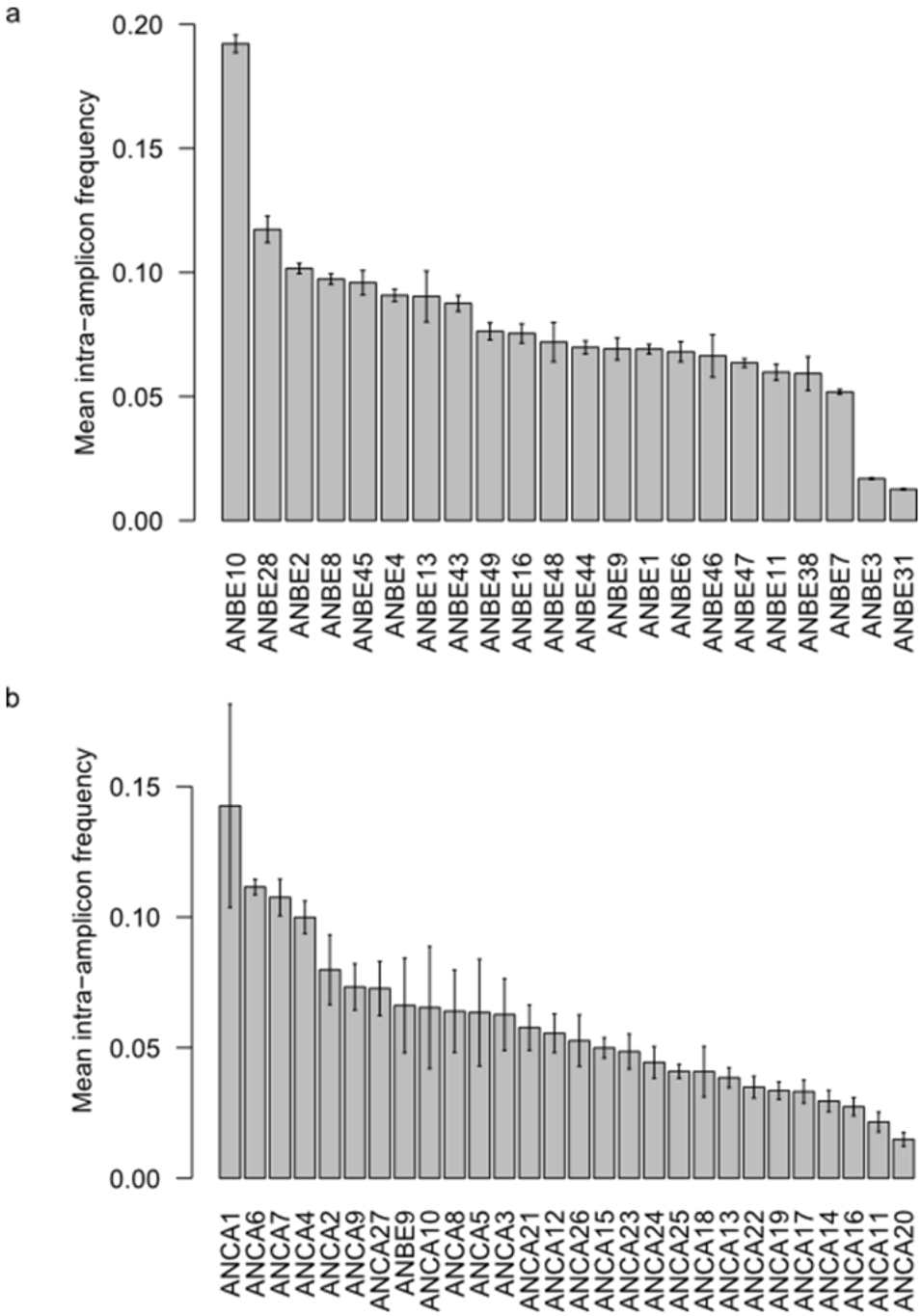


Figure 4.4 Mean intra-amplicon frequencies for each of the MHC class I exon 3 alleles identified in a) Berthelot's pipits (*Anthus berthelotii*)) and b) tawny pipits (*A. campestris*). Error bars are standard errors.

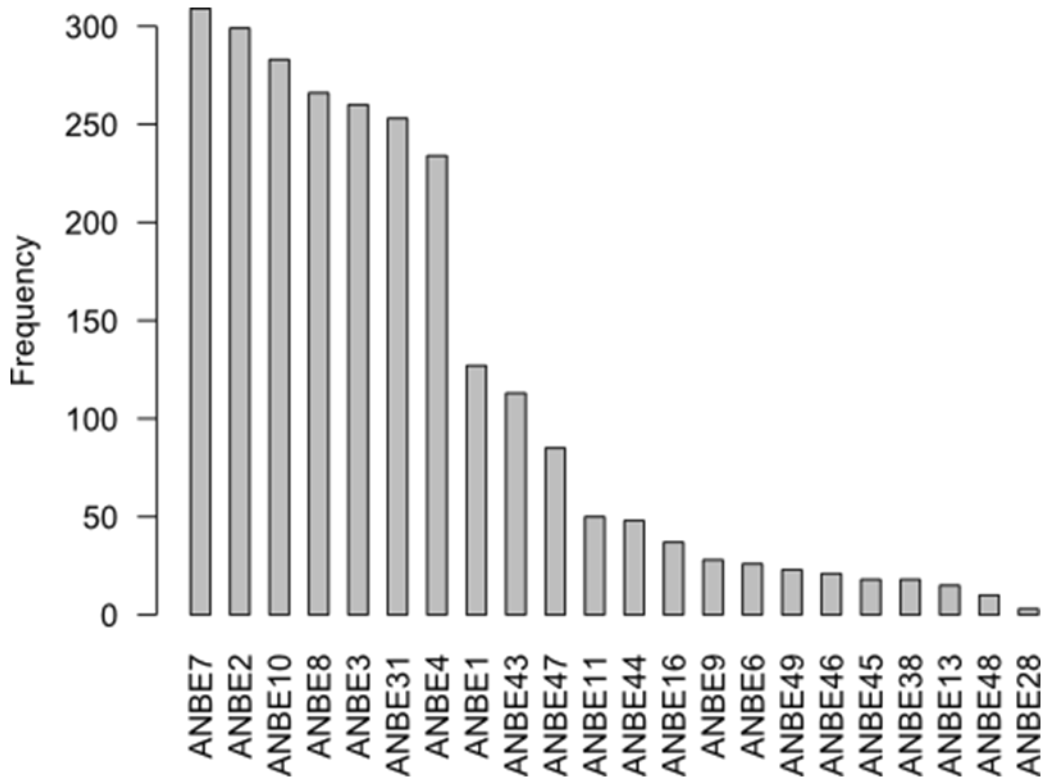


Fig. 4.5 Frequency of Berthelot's pipit (*Anthus berthelotii*) MHC class I exon 3 alleles identified in 310 individuals in the population of Tenerife.

4.4.2 Allele amplification efficiencies

In Berthelot's pipits, the lowest amplification efficiency was obtained for allele ANBE31 (0.2, i.e. five times lower than the reference ANBE11). The highest amplification efficiency obtained was for ANBE10 (3.0) (Fig. 4.7a). The modified Galan's T1 threshold showed that 139 reads (range 132-144) were needed to reach coverage of 99.9% of a genotype with twelve alleles. Given this value, we excluded the one sample (identified above) that had only 45 reads in one of its replicates from all downstream analyses. In tawny pipits the lowest amplification efficiency obtained was for ANCA20 (0.33) and the highest amplification efficiency for ANCA1 (3.9) (Fig. 4.7b). The modified Galan's T1 threshold for the lowest amplification efficiency in tawny pipit alleles was 291. All tawny pipit amplicons had more than 291 reads.

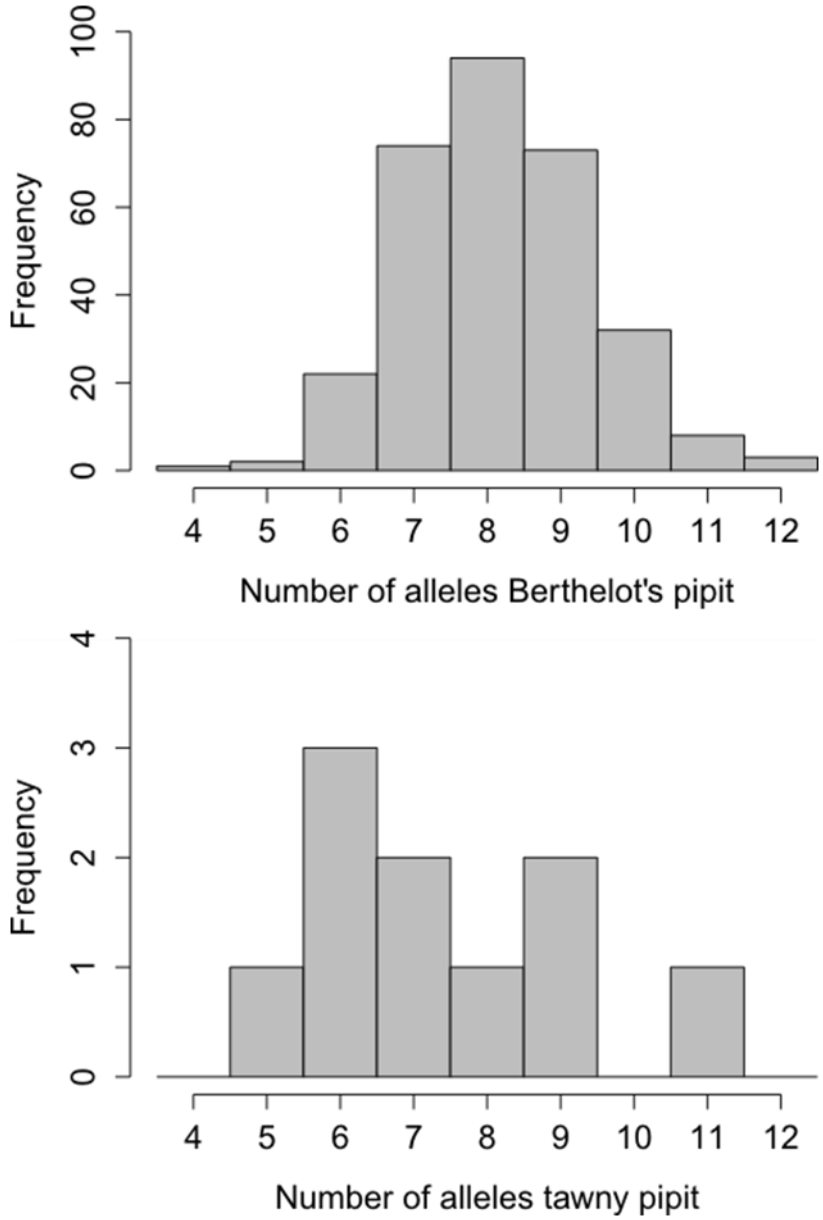


Figure 4.6 Number of Berthelot's pipit, *Anthus berthelotii* (upper panel) or tawny pipit, *A. campestris* (lower panel) MHC class I exon 3 alleles (at the nucleotide level) per individual in the sampled population.

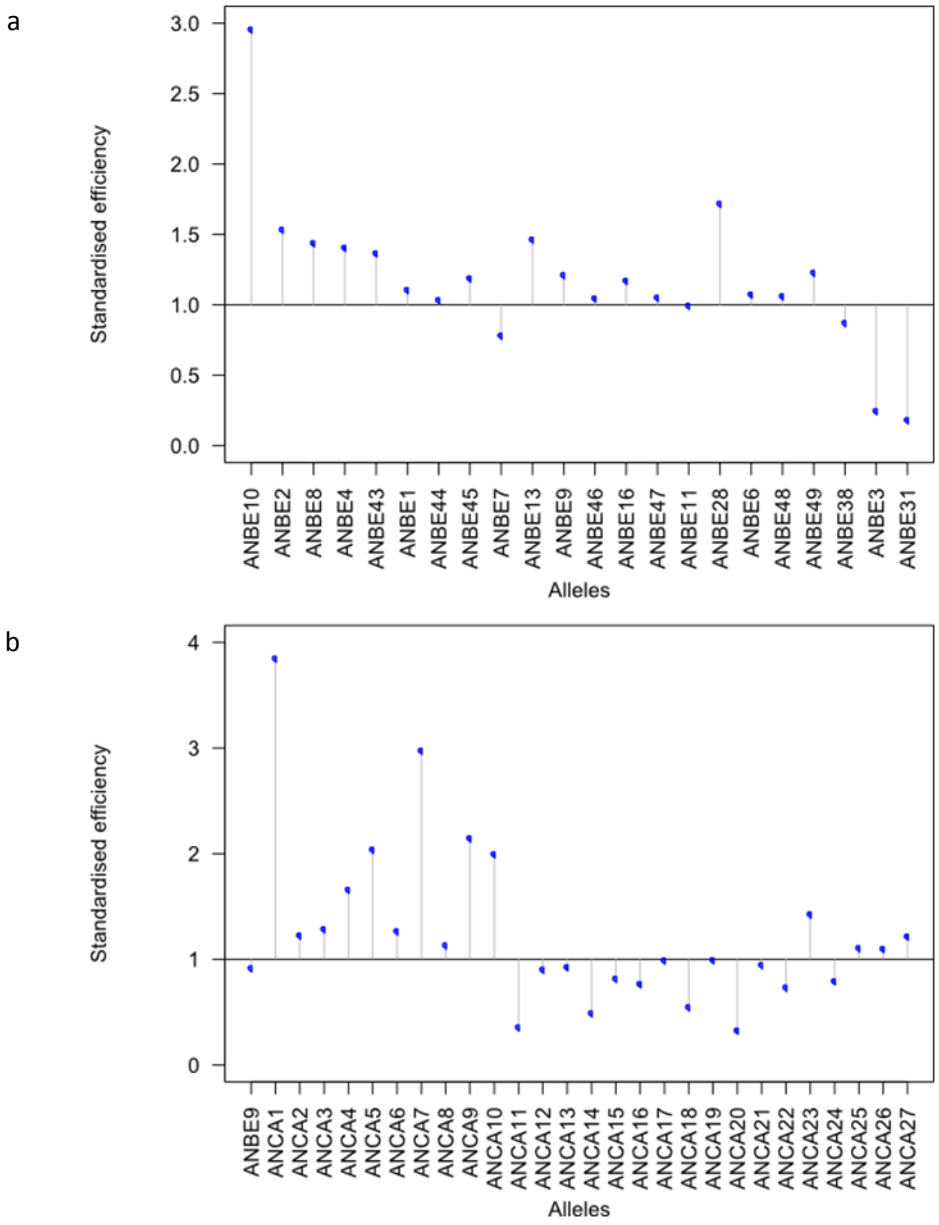


Figure 4.7 Standardised amplification efficiency of MHC class I exon 3 alleles identified in a) 310 Berthelot's pipits from Tenerife, and b) 10 tawny pipits. The horizontal line represents the amplification efficiency of 1.0, obtained for the reference allele ANBE11 (Berthelot's pipits) or ANCA17 (tawny pipits).

4.4.3 MHC sequence analyses

Descriptors of sequence variation within MHC class I exon 3 in Berthelot's and tawny pipits are summarised in Table 4.2. Seven alleles (ANBE3, ANBE31, ANCA11, ANCA12, ANCA13, ANCA17 and ANCA25) had an insertion of three nucleotides, resulting in a protein with one amino acid insertion but no disruption of the reading frame. Pairwise nucleotide distance of MHC alleles was significantly higher in tawny pipit than in Berthelot's pipit ($P < 0.001$, Table 4.2).

Table 4.2 Summary of nucleotide sequence variation of MHC class I exon 3 sequences described in this study in Berthelot's pipits, *Anthus berthelotii*, and tawny pipits, *A. campestris.*

Descriptor	Berthelot's pipit	Tawny pipit
Number of alleles	20	28
Number of variable sites	84	103
Number of mutations	105	136
$\pi \pm SE$[1]	0.11 ± 0.01	0.14 ± 0.01
$k \pm SE$[2]	27.09 ± 0.86	32.84 ± 0.81
Rm (95 % CI)[3]	32.5 (25.0 – 40.0)	66.7 (52.0 – 83.1)
Θ (95 % CI)[4]	23.0 (19.2 – 27)	32.9 (27.2 – 38.5)

[1]Nuclotide diversity ± standard deviation

[2]Average number of nucleotide differences ± standard deviation

[3]Recombination rate (lower and upper 95% confidence limits)

[4]Mutation rate (lower and upper 95% confidence limits)

The neighbour net of Berthelot's and tawny pipits MHC class I exon 3 alleles revealed a total of 17 lineages (Fig. 4.8). The lineage partitions were chosen based on the previous phylogeny described for Berthelot's pipit MHC class I alleles (Spurgin *et al.* 2011). Berthelot's pipit alleles were included in eleven lineages, while tawny pipit alleles were grouped in 13 lineages, of which seven were shared with Berthelot's pipit and six were unique to tawny pipits (Fig. 4.8). The presence of boxes in the net means that there can be several paths between any two alleles, typical of sequence datasets with gene conversion and recombination events (Bryant & Moulton 2004). Four gene conversion events in MHC class I alleles of Berthelot's and tawny pipits were identified by at least two methods (Table 4.3).

Table 4.3 Tracts of gene conversion identified by at least two recombination-detection methods in MHC class I alleles of Berthelot's pipits (*Anthus* berthelotii, ANBE) and tawny pipits (*A. campestris*, ANCA). Position corresponds to nucleotides that limit the gene conversion tract.

Recombinant allele	Major parent	Minor parent	Positions of breakpoints	Methods (*P* value)
ANBE4	ANCA18	ANCA5	35 – 179	MaxChi (0.014) 3Seq (0.039)
ANCA16	ANCA13	ANBE6	48 – 216	SiScan (0.037) 3Seq (0.001)
ANCA17	ANCA15	ANBE44	170 – 238	Chimaera (0.005) 3Seq (0.003)
ANCA12	ANCA18	ANCA5	12 – 197	MaxChi (0.014) 3Seq (0.039)

The 22 Berthelot's pipit MHC class I alleles translated as 20 different amino acid sequences, containing 42 variable amino acid sites and 63 amino acid changes (14 positions had more than two amino acids) (Fig. 4.9). Considering only the 15 PBR sites, 15 unique PBR sequences were detected, encoded by 12 variable amino acids (80%, Fig. 4.10), while the 65 non-PBR sites harboured 30 variable amino acids (46%). The 28 MHC class I alleles found in tawny pipits translated as 24 unique amino acid sequences containing 47 variable amino acid sites, with 86 amino acid changes (18 positions had more than two amino acids) (Fig. 4.9). Considering only the PBR sites, there were 20 unique PBR sequences, with 12 variable amino acids (80%), while the 65 non-PBR sites had 36 variable amino acids (55%).

Over the full exon 3 sequence of Berthelot's pipit MHC class I, the rate of synonymous substitutions (d_S) was significantly higher than the rate of non-synonymous substitutions (d_N) ($P < 0.001$, Fig. 4.11), and the ratio of d_N/d_S for the full exon was 0.75. In both the PBR and non-PBR regions, d_N was not significantly different from d_S ($P = 0.45$ and 0.52, respectively). For the PBR $d_N/d_S = 1.07$ and for the non-PBR $d_N/d_S = 0.97$. However, both d_S and d_N were significantly higher at the PBR compared to the non-PBR ($P < 0.001$, Fig. 4.11). When performing selection tests with sequences identified in both pipit species, ten codons were identified as having evidence of positive selection. Four of these (5, 19, 61 and 62) match the PBR (Fig. 4.9).

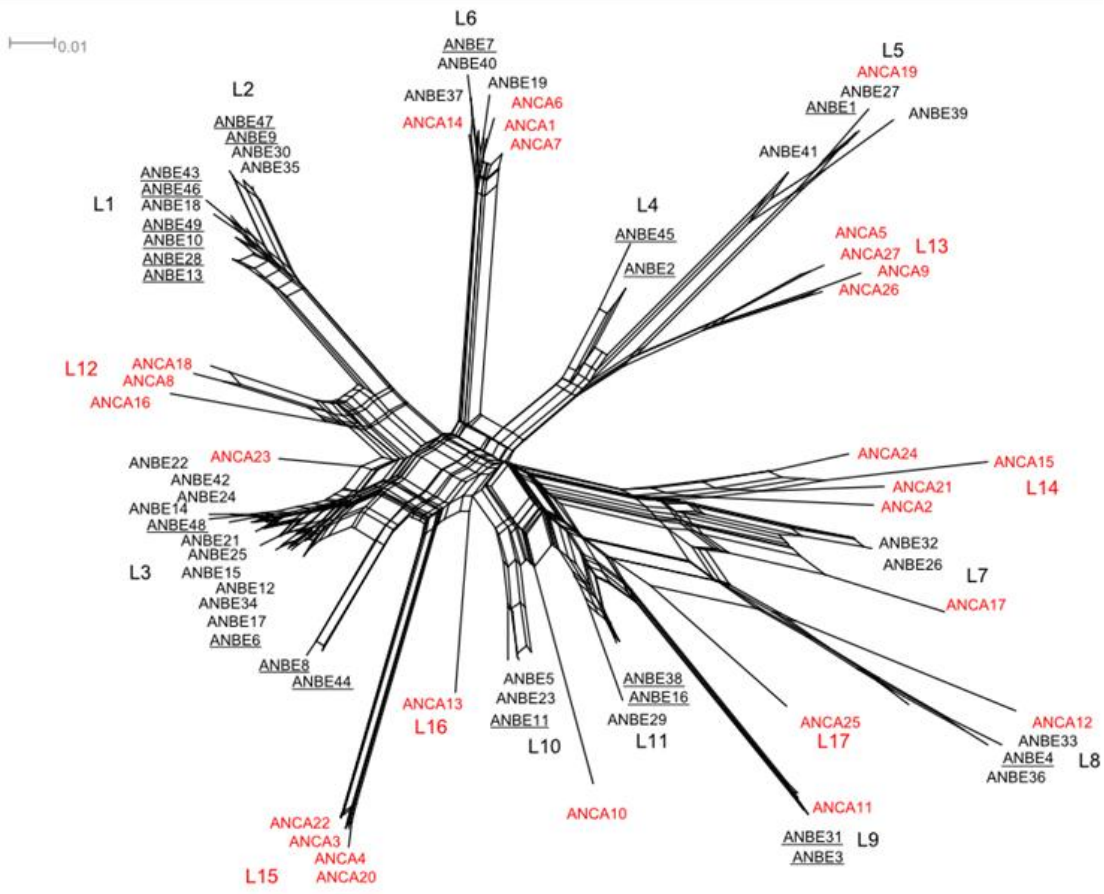


Figure 4.8 Neighbour net of 49 Berthelot's pipit (*Anthus berthelotii*, ANBE) and 27 tawny pipit (*A. campestris*, ANCA) MHC class I exon 3 alleles using Jukes-Cantor distance, including alleles previously identified in Berthelot's pipits (Spurgin *et al.* 2011). ANBE alleles that we found in the present study are underlined. Alleles and lineages found in Berthelot's pipits are shown in black font and alleles and lineages unique to tawny pipits are shown in red. Labels L1-L17 correspond to lineages. Length along lines is proportional to genetic distance between any two alleles.

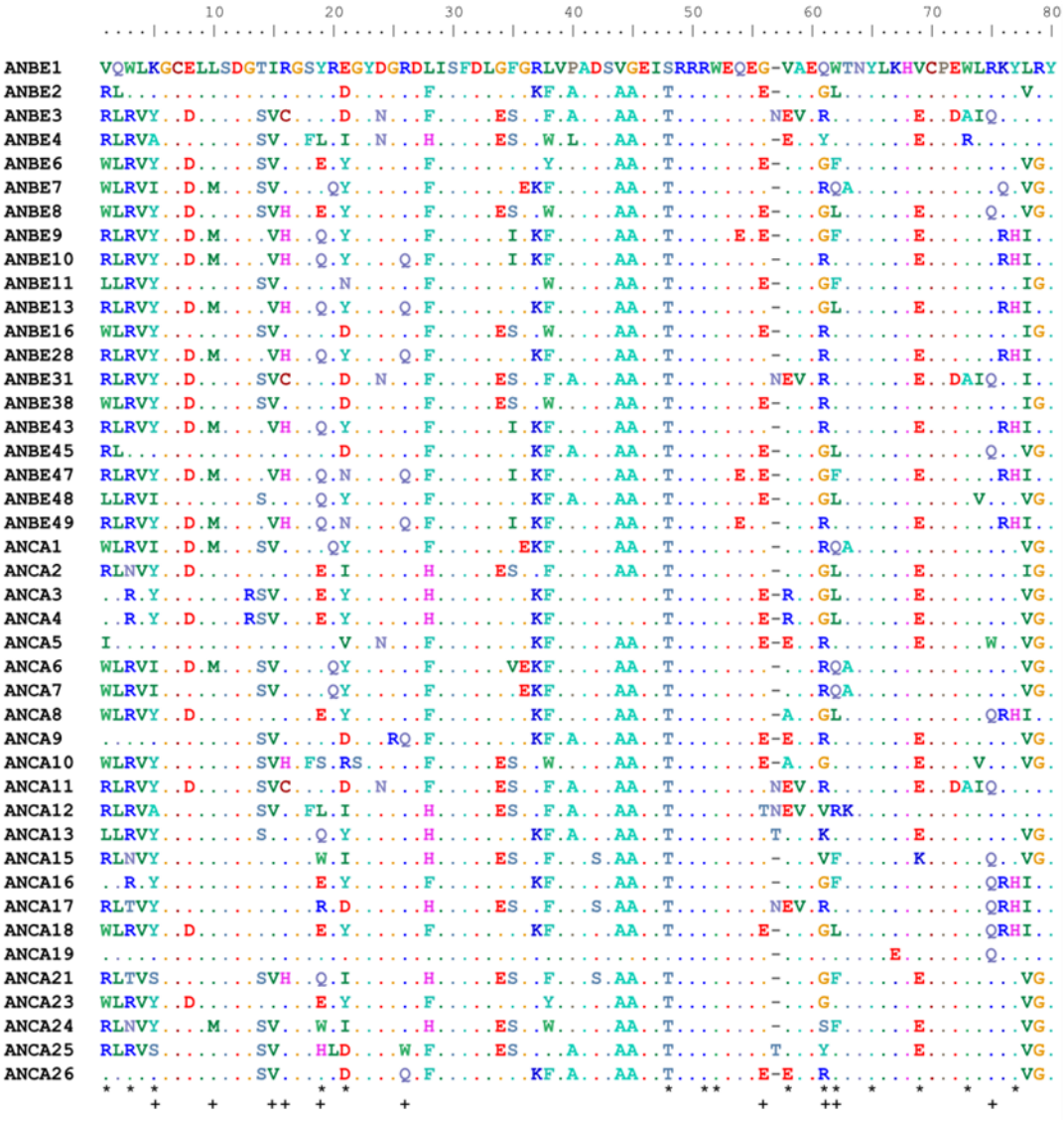


Figure 4.9 Amino acid alignment of the 20 Berthelot's pipit (*Anthus berthelotii*) and 23 tawny pipit (*A. campestris*) functional MHC class I exon 3 alleles identified in this study. Dots in the alignment denote consensus amino acids with ANBE1. The amino acids comprising the peptide binding region are shown with asterisks. Amino acids under positive selection using Berthelot's and tawny pipit alleles are indicated with a plus sign.

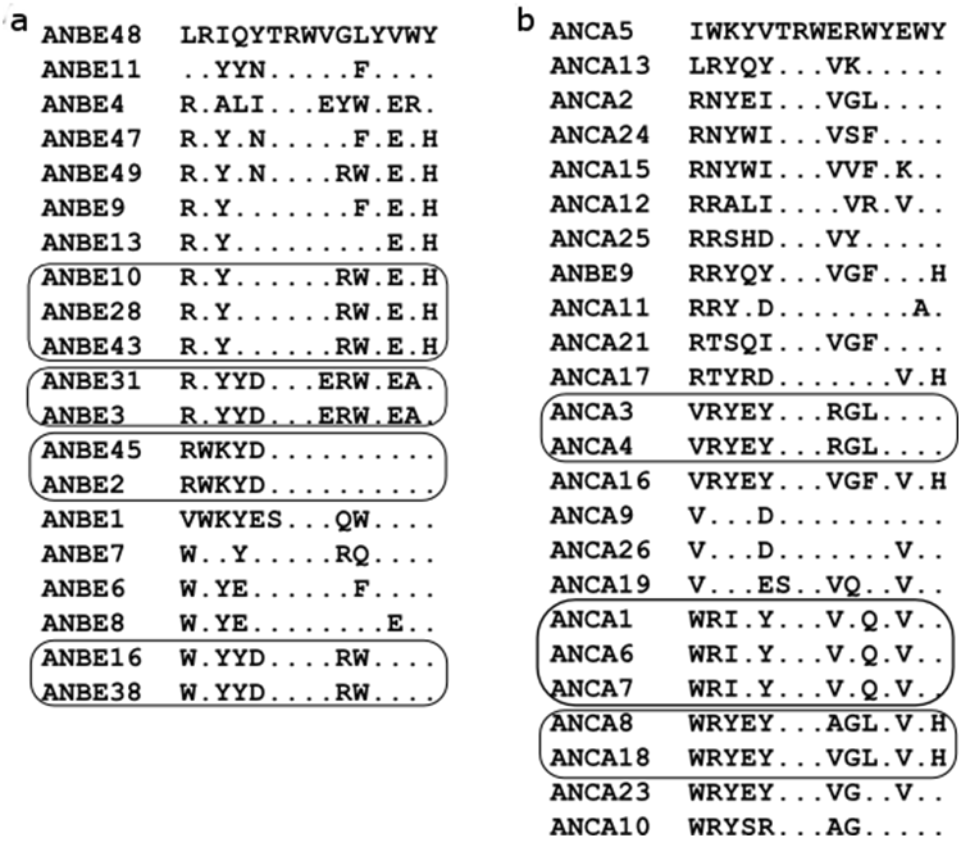


Figure 4.10 Alignment of amino acids of peptide binding region (PBR) sequences of a) Berthelot's pipit (*Anthus berthelotii*) or b) tawny pipit (*Anthus campestris*) MHC class I exon 3 alleles. Squares enclose similar PBR sequences. Consensus with first sequence is denoted by dots. Order of amino acids corresponds to the position of PBR amino acids in the full exon, as shown in figure 4.9.

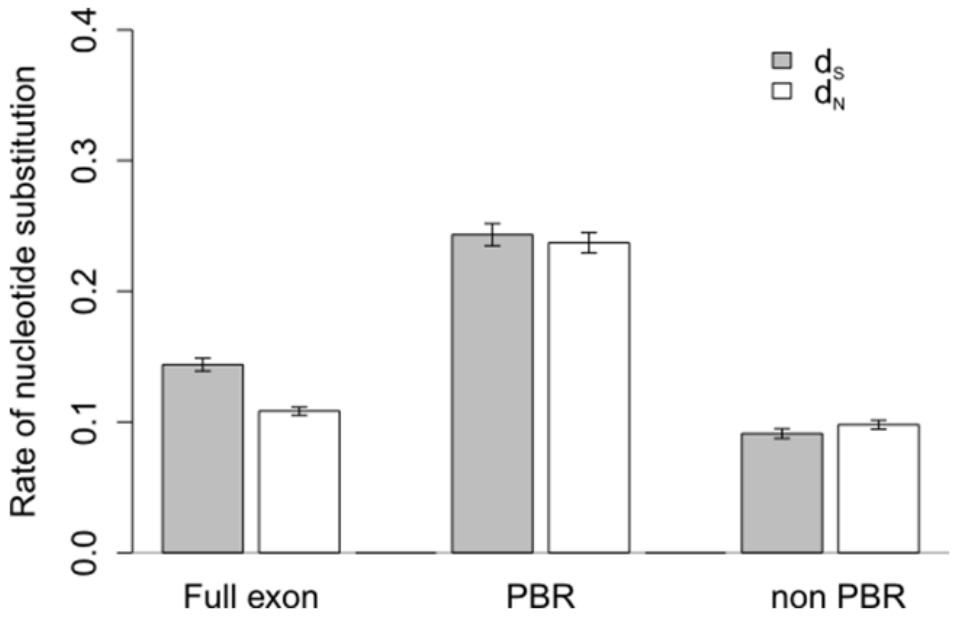


Figure 4.11 Rates of non-synonymous (d_N) and synonymous (d_S) substitutions in the full exon, peptide binding (PBR) and non-PBR regions MHC class I exon 3 alleles of Berthelot's pipits in the population of Tenerife. Error bars represent 95 % confidence

The number of supertypes of Berthelot's pipit MHC class I alleles that could be identified with the *k*-means clustering algorithm was 20, and therefore equal to the number of unique amino acid sequences identified. This was also the case if only alleles that differed at sites within the PBR (15 PBR unique sequences) and the PSS (11 PSS unique sequences) were included, suggesting that each allele has different antigen binding properties. For this reason, we didn't perform the discriminant analysis of principal components (DAPC) to identify the alleles in each cluster (Jombart *et al.* 2010).

4.5 Discussion

We used 454 pyrosequencing to screen MHC class I exon 3 variation in individual Berthelot's pipits from across the population of Tenerife and, for comparison, in its sister species the tawny pipit. The variant/artefact identification procedure including replication of all samples allowed the successful genotyping of 309/310 Berthelot's pipits, with high genotype repeatability (96%). Validation procedures identified two alleles with low amplification efficiencies. The resulting sequence data were used to characterise MHC variation across these

two closely related species. We found significantly higher MHC sequence variation in the outbred tawny pipit than in the previously bottlenecked Berthelot's pipit and found evidence for maintenance of divergent MHC supertypes in Berthelot's pipit. We also found evidence for gene conversion, an increased recombination to mutation ratio, and positive selection at specific codons within the exon 3 sequence.

454 pyrosequencing is a valuable tool for parallel sequencing of multilocus genes, such as those of the MHC, in a large number of samples. However, this method is prone to errors generated from two PCR amplifications and the pyrosequencing reaction (Meyerhans *et al.* 1990; Huse *et al.* 2007). In our study, artefacts and unclassified variants clearly outnumbered putative alleles. This concurs with other studies that used 454 for genotyping the MHC, which also assigned more reads to artefacts than to putative alleles (Zagalska-Neubauer *et al.* 2010; Sepil *et al.* 2012; Sommer *et al.* 2013).

For accurate reconstruction of individual MHC genotypes and testing of evolutionary hypotheses, it is essential to be able to reliably detect the majority of alleles that exist in a population, even if some of them do not amplify well in the study species. That MHC alleles differ in amplification efficiency has been known for some time (Babik 2010; Kiemnec-Tyburczy *et al.* 2010; Zagalska-Neubauer *et al.* 2010), but only recently has a methodology been proposed to calculate amplification efficiencies for each allele (Sommer *et al.* 2013). Of the 22 MHC alleles we identified in Berthelot's pipit, two had low amplification efficiencies that put them at high risk of allelic dropout. These two alleles never amplified well enough to make the initial list of putative alleles, and were only identified by post-processing examination of the list of unclassified variants. Sommer *et al.* (2013) report 3/64 of the alleles they identified as having low amplification efficiency. This, together with our results, supports their hypothesis that low amplification is common but that it only affects a few alleles. However, it is important to note that some alleles might not amplify at all for a given set of primers, and that some instances of allelic dropout will thus not be accounted for. It is crucial that studies on the MHC acknowledge the variable amplification efficiency of alleles (regardless of the sequencing method used), and that they use the minimum amplification efficiency to obtain a minimum number of reads required per amplicon for reliable genotyping. When assessing MHC-disease associations, we advocate excluding samples that do not pass the minimum-read criterion, as well as excluding alleles with low amplification efficiency. Failure to do so incurs a high risk of

incomplete data, i.e. missing the presence of these specific alleles within individuals, leading to inaccurate conclusions. We suggest that the calculation of amplification efficiencies be implemented and reported in all future MHC studies.

Initial cloning of MHC class I exon 3 from six Berthelot's pipits (five from Tenerife, one from Madeira) suggested the presence of a minimum of four loci (Spurgin *et al.* 2011). Our results suggest that there are at least six MHC class I loci in Berthelot's pipits. When cloning is used to characterize the MHC amplified in a few individuals from a few populations the number of alleles is likely to be underestimated, and even more so if there are alleles with low amplification efficiency.

Our results differ to some degree from the earlier study that sequenced the MHC in the populations of Berthelot's pipit (Spurgin *et al.* 2011). We found seven alleles that had not been described for Tenerife, but had been found in other islands. The smaller sample size of the earlier study (only 30 individuals from each of two populations) might have been the reason why four of these were not detected (ANBE6, ANBE9, ANBE13 and ANBE38), because in the present study these were found at frequencies lower than 0.1. However, the other three alleles (ANBE1, ANBE7 and ANBE31) were found at frequencies of 0.4, 0.9 and 0.8, respectively in the present study, suggesting they are common in Tenerife. We also detected seven alleles that hadn't been described in any of the populations from the earlier study (ANBE43-ANBE49), and thus may be unique to the population of Tenerife. The frequencies of five of these alleles was lower than 0.1, thus not previously finding them was probably due to the small sample size of that earlier study. However, the other two alleles, ANBE43 and ANBE47, were found at frequencies of 0.4 and 0.3, respectively, and it is unclear why these were not detected in the earlier study. We also failed to find five alleles that had been previously reported in Tenerife (Spurgin *et al.* 2011): ANBE12 and ANBE39 from El Teide, ANBE41 and ANBE24 from the lowlands and ANBE19 from both El Teide and lowlands. In that earlier study these alleles were not identified as putative ancestral alleles, but rather as recombinant alleles that were derived through gene conversion. These five recombinant alleles did not co-occur with their ancestral alleles on Tenerife, thus rejecting the possibility that they could have been PCR artefacts derived from the putative ancestral alleles (Spurgin *et al.* 2011). That we failed to detect alleles ANBE19, ANBE24 and ANBE41 could be due to the possibility that they are rare in lowland Tenerife. However, this is unlikely given that they were earlier detected in a much smaller

sample (30 individuals), and given our sampling scheme that facilitated capture of variation throughout the population. However, in this study we could have reasonably missed ANBE12 and ANBE39 in the 26 birds we sampled from El Teide if they are rare in this population. Since the earlier study did not assess individual-level variation, it is not possible to know the frequency of these alleles in the earlier sample.

We cannot discard the possibility that the differences between the two studies lie in their methodological differences. The earlier study assessed MHC class I variation at the population level using population-specific tags, rather than individual-specific tags. Another difference is that here we sequenced only from the forward end of the amplicon, while the earlier study sequenced amplicons from both ends with subsequent assembly of sequences by read overlap. This is because, at the time of the earlier study, the average 454 read length was shorter than the total length of the amplicon. However, both studies used the same template-specific primer sequences and the same PCR conditions. The earlier study also used a different bioinformatics strategy for validating alleles. It is difficult to determine in which of the two studies errors were generated, given that both used different methodologies. It is also possible that the differences may be real and due to differences in the subset of individuals sampled in each case. For example, allele frequencies might have changed from 2006, when the samples of the earlier study (Spurgin *et al.* 2011) were collected, to 2011. Given that Berthelot's pipit generation time is *ca.* three years (Garcia-Del-Rey & Cresswell 2007), approximately two generations have passed between the two samplings. In support of this hypothesis, rapid changes in MHC allele frequencies (between successive cohorts) have been reported in another passerine (Westerdahl *et al.* 2004). This pattern might result from fluctuating pathogen-mediated selection on these alleles. However, we cannot assess this with our data and further investigation involving sampling across different years is needed in order to confirm this.

It might be necessary to employ our individual-level MHC genotyping to screen the same samples that were used in the earlier study in order to clarify this. It is also important to note that other NGS technologies might provide better alternatives for accurately sequencing the MHC. For example, Illumina has proven to be highly repeatable and accurate for genotyping duplicated loci (Lighten *et al.* 2014b), and might therefore be the platform of choice in future studies that require sequencing of the MHC. However, there may equally be problems with

other NGS methods, and only future comparative studies between methods will resolve such issues.

Maintaining and restoring MHC diversity is especially important in isolated, bottlenecked populations where low genetic diversity might have implications for disease resistance and population survival (Bollmer *et al.* 2011; Yasukochi *et al.* 2012; Niskanen *et al.* 2014). Isolated species with populations that have undergone bottlenecks generally show lower MHC genetic diversity than outbred species (reviewed in Radwan *et al.* 2010). In the present study 22 alleles were detected in the 309 individuals sampled in the Berthelot's pipits population on Tenerife, compared to the 28 alleles detected in just 10 tawny pipits. This suggests that the Berthelots' pipit population on Tenerife has much lower levels of genetic variation at these MHC loci than its continental sister species. This is confirmed by the fact that levels of nucleotide variation were lower in Berthelot's pipit alleles than in tawny pipit alleles (nucleotide diversity = 0.11 ± 0.01 and 0.14 ± 0.01, respectively). However If we consider the whole range of Berthelot's pipits, a total of 49 MHC class I alleles have been identified across the 13 populations, indicating that MHC variation can be regenerated in a bottlenecked population, in this case by gene conversion (see below). The same pattern of higher genetic diversity in tawny pipit compared to Berthelot's pipit has been described for the Toll-like receptor loci (TLR) of the innate immune system (Chapter 2).

Interestingly, we found that each Berthelot's pipit MHC class I allele represented one supertype, suggesting that each allele has unique binding properties. In isolated island populations, such as the Berthelot's pipit, alleles may be lost during bottleneck events, but the maintenance of supertypes through these events may occur if alleles of different supertypes segregate at different loci and thus, PMS results in the maintenance of these functionally divergent alleles at different loci (van Oosterhout 2013). Our finding concurs with previous evidence for the maintenance of divergent supertypes in other systems (Huchard *et al.* 2008; Ellison *et al.* 2012). This process has important evolutionary implications in bottlenecked populations, allowing the maintenance of divergent MHC alleles that can potentially detect a broad range of pathogens.

Our neighbour net of the phylogenetic relationships among Berthelot's and tawny pipit MHC class I alleles revealed that these species share allele lineages and do not separate according to species. Seven out of the 17 MHC lineages found were shared between Berthelot's and tawny

pipits and one allele (ANBE9) was found to be exactly the same across the 240 base pairs we screened in both species. Given the low sample size for the tawny pipit it is possible that we missed amplifying some lineages and thus our estimate of lineage sharing must be treated as approximate. Furthermore, when colonizing the Canary Islands, some MHC lineages might have been absent from the founding population, hence explaining why some lineages found in tawny pipits are not represented in Berthelot's pipits. Despite these possibilities we still found considerable lineage sharing between these two species. Tawny and Berthelot's pipits diverged only *ca.* 2.5 m.y. ago (Chapter 2; Voelker 1999); the lineage sharing could therefore be attributed to incomplete lineage sorting (Klein *et al.* 1993; Klein *et al.* 1998). Another possibility is that trans-species persistence of MHC lineages has been promoted by balancing selection, whereby allele lineages that confer selective advantage persist over evolutionary time with little change (Klein *et al.* 1993). Such trans-species persistence has been reported in many MHC studies (e.g. Graser *et al.* 1996; Kikkawa *et al.* 2009; Jaratlerdsiri *et al.* 2014). Interestingly, we have previously found that the two pipit species also share a small number of alleles at Toll-like receptors (three out of 94 alleles across five loci, Chapter 2). That we see similar patterns in the two immune gene families may suggest a greater role for incomplete lineage sorting than balancing selection in the two species sharing alleles, but testing this hypothesis would require both a larger sample of tawny pipits and assessment of other loci.

Pathogen-mediated balancing selection on the PBR is thought to be the main force maintaining variation at MHC (Westerdahl *et al.* 2005; Evans & Neff 2009; Spurgin & Richardson 2010). In Berthelot's pipits MHC class I alleles we found no evidence of an elevated d_N/d_S in the PBR compared to the non-PBR which is the classic indication of selection at MHC (Hughes & Nei 1988; Bernatchez & Landry 2003). However, we found that both d_N and d_S were higher at the PBR than at the non PBR (Fig. 4.11), which is to be expected when gene conversion is the main source of variation (Ohta 1995). Gene conversion results in the transfer of sections of DNA containing synonymous and non-synonymous changes between alleles within or across loci. When this process involves sites in the PBR, the new molecular conformation of amino acids encoded for may be advantageous if it creates a new allele that allows the better binding of peptides from pathogens (Ohta 1995). On the other hand, if such events occur within the non-PBR they are likely to be selected against, because this region is functionally constrained due to its role in molecule integrity (Klein *et al.* 1993). Several previous studies have argued that gene conversion is one of the main mutational forces generating MHC allelic variation in

vertebrates (reviewed in Hogstrand & Bohme 1999; Ohta 1999), a hypothesis that has been directly supported by earlier data from Berthelot's pipit (Spurgin *et al.* 2011). In line with this earlier study, we also detected four gene conversion events and an elevated recombination to mutation ratio across the MHC alleles of both pipit species (1.4 for Berthelot's pipit and 2.0 for tawny pipit alleles). Within the exon amplified we found evidence of historical balancing selection at ten specific amino acid sites, of which four corresponded to the estimated PBR. Three of these sites have previously been identified to be under positive selection in other bird species (Sutton *et al.* 2013), which suggests that these sites are important determinants of the binding properties of the PBR in avian species.

Overall, our results suggest that Berthelot's pipit population on Tenerife has reduced allelic diversity at the MHC compared to its closest sister species. Nevertheless, the allele lineages that persisted, or were generated after the colonization of Tenerife, display divergent antigen binding properties. These divergent alleles might be sufficient to successfully initiate an adequate immune response to the local pathogens that threaten this population. This mechanism can have significant implications for the survival and establishment of populations that colonize new areas possibly containing novel diseases. Berthelot's pipits in Tenerife have a high incidence of malaria (Gonzalez-Quevedo *et al.* 2014) and the MHC is likely to play a role in the epidemiology of this disease in the population given that it has previously been linked to malaria resistance (Hill *et al.* 1991; Westerdahl *et al.* 2005; Bonneaud *et al.* 2006; Westerdahl *et al.* 2013). This screening methodology and the data outlined in the present study can now be used to assess the association of MHC alleles and disease susceptibility/resistance, and to investigate what causes temporal changes in MHC allele frequency within populations to further understand how pathogen mediated selection might shape the variation of the MHC at the population level.

4.6 References

454 Life Sciences Corp. (2009) Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Extended MID Set *454 Sequencing Technical Bulletin* **005-2009**, 1-7.

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

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

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

Baquero JE, Miranda S, Murillo O*, et al.* (2006) Reference strand conformational analysis (RSCA) is a valuable tool in identifying MHC-DRB sequences in three species of *Aotus* monkeys. *Immunogenetics* **58**, 590-597.

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

Beuf KD, Schrijver JD, Thas O*, et al.* (2012) Improved base-calling and quality scores for 454 sequencing based on a Hurdle Poisson model. *BMC bioinformatics* **13**, 303-303.

Bjorkman PJ, Saper MA, Samraoui B*, et al.* (1987) Structure of the human class-I histocompatibility antigen, HLA-A2. *Nature* **329**, 506-512.

Bodmer JG, Marsh SGE, Albert ED*, et al.* (1997) Nomenclature for factors of the HLA system, 1996. *Tissue Antigens* **49**, 297-321.

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

Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* **176**, 1035-1047.

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

Bradley RD, Hillis DM (1997) Recombinant DNA sequences generated by PCR amplification. *Molecular Biology and Evolution* **14**, 592-593.

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**, 3444-3455.

Brown JH, Jardetzky TS, Gorga JC*, et al.* (1993) 3-Dimensional structure of the human class II-histocompatinility antigen HLA-DR1. *Nature* **364**, 33-39.

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

Chelvanayagam G (1996) A roadmap for HLA-A, HLA-B, and HLA-C peptide binding specificities. *Immunogenetics* **45**, 15-26.

Cheng Y, Stuart A, Morris K*, et al.* (2012) Antigen-presenting genes and genomic copy number variations in the Tasmanian devil MHC. *Bmc Genomics* **13**, 87.

Cramp S, Perrins CM (1977-1994) *Handbook of the birds of Europe, the Middle East and Africa. The birds of the western Palearctic* Oxford University Press, Oxford.

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

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

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

Ellison A, Allainguillaume J, Girdwood S*, et al.* (2012) Maintaining functional major histocompatibility complex diversity under inbreeding: the case of a selfing vertebrate. *Proceedings of the Royal Society B-Biological Sciences* **279**, 5004-5013.

Evans ML, Neff BD (2009) Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology* **18**, 4716-4729.

Frank SA (2002) *Immunology and evolution of infectious disease* Princeton University Press, Princeton.

Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *Bmc Genomics* **11**, 15.

Garcia-Del-Rey E, Cresswell W (2007) The breeding biology of the endemic Berthelot's Pipit, *Anthus berthelotii*, in a harsh oceanic island environment (Tenerife, Canary Islands). *Ostrich* **78**, 583-589.

Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**, 573-582.

Gonzalez-Quevedo C, Davies RG, Richardson DS (2014) Predictors of malaria infection in a wild bird population: landscape-level analyses reveal climatic and anthropogenic factors. *Journal of Animal Ecology* **83**, 1091–1102.

Graser R, O'Huigin C, Vincek V, Meyer A, Klein J (1996) Trans-species polymorphism of class II Mhc loci in *Danio* fishes. *Immunogenetics* **44**, 36-48.

Hill AVS, Allsopp CEM, Kwiatkowski D*, et al.* (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature* **352**, 595-600.

Hogstrand K, Bohme J (1999) Gene conversion can create new MHC alleles. *Immunological Reviews* **167**, 305-317.

Holcomb CL, Rastrou M, Williams TC*, et al.* (2014) Next-generation sequencing can reveal in vitro-generated PCR crossover products: some artifactual sequences correspond to HLA alleles in the IMGT/HLA database. *Tissue Antigens* **83**, 32-40.

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

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

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

Huse SM, Huber JA, Morrison HG, Sogin ML, Mark Welch D (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8**, R143.

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

Illera JC (2007) Bisbita Caminero *Anthus berthelotii*. In: *Atlas de las aves nidificantes en el archipiélago Canario (1997-2003)* (ed. JA L), pp. 344-347. General de Conservación de la Naturaleza-Sociedad Española de Ornitología, Madrid.

Illera JC, Emerson BC, Richardson DS (2007) Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology* **16**, 4599-4612.

Jaratlerdsiri W, Isberg SR, Higgins DP*, et al.* (2014) Evolution of MHC class I in the Order Crocodylia. *Immunogenetics* **66**, 53-65.

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**, 2157-2168.

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

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

Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering* **96**, 317-323.

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

Kiemnec-Tyburczy KM, Richmond JQ, Savage AE, Zamudio KR (2010) Selection, trans-species polymorphism, and locus identification of major histocompatibility complex class II beta alleles of New World ranid frogs. *Immunogenetics* **62**, 741-751.

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**, 341-352.

Klein J, O'Huigin C, Mims CA*, et al.* (1994) MHC polymorphism and parasites (and discussion). *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **346**, 351-358.

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

Klein J, Satta Y, Ohuigin C, Takahata N (1993) The molecular descent of the major histocompatibility complex. *Annual Review of Immunology* **11**, 269-295.

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

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.

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

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

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

Margulies M, Egholm M, Altman WE*, et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.

Martin DP, Lemey P, Lott M*, et al.* (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462-2463.

Meyerhans A, Vartanian JP, Wainhobson S (1990) DNA recombinatiion during PCR. *Nucleic Acids Research* **18**, 1687-1691.

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

Murrell B, Moola S, Mabona A*, et al.* (2013) FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution* **30**, 1196-1205.

Murrell B, Wertheim JO, Moola S*, et al.* (2012) Detecting individual sites subject to episodic diversifying selection. *Plos Genetics* **8**, e1002764

Mwenda JM, Hashiba K, Bambra CS, Shotake T (1997) Analysis of primate major histocompatibility complex (MHC)-DQA1 locus by PCR-single strand conformation polymorphism (SSCP). *Cellular and Molecular Biology* **43**, 471-475.

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

Niskanen AK, Kennedy LJ, Ruokonen M*, et al.* (2014) Balancing selection and heterozygote advantage in major histocompatibility complex loci of the bottlenecked Finnish wolf population. *Molecular Ecology* **23**, 875-889.

Ohta T (1995) Gene conversion vs. point mutation in generating variability at the antigen recognition site of major histocompatibility complex loci. *Journal of Molecular Evolution* **41**, 115-119.

Ohta T (1999) Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. *Immunological Reviews* **167**, 319-325.

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218-225.

Pond SLK, Frost SDW (2005a) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531-2533.

Pond SLK, Frost SDW (2005b) Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**, 1208-1222.

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

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

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

R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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

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

Richardson DS, Jury FL, Blaakmeer K, Komdeur J, Burke T (2001) Parentage assignment and extra-group paternity in a cooperative breeder: the Seychelles warbler (*Acrocephalus sechellensis*). *Molecular Ecology* **10**, 2263-2273.

Robinson J, Halliwell JA, McWilliam H*, et al.* (2013) The IMGT/HLA database. *Nucleic Acids Research* **41**, D1222-D1227.

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

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

Sidney J, Delguercio MF, Southwood S*, et al.* (1995) Several HLA alleles share overlapping peptide specificities. *Journal of Immunology* **154**, 247-259.

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

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

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

Spurgin LG, Illera JC, Jorgensen TH, Dawson DA, Richardson DS (2014) Genetic and phenotypic divergence in an island bird: isolation by distance, by colonisation or by adaptation? *Molecular Ecology* **23**, 1028-1039.

Spurgin LG, Illera JC, Padilla DP, Richardson DS (2012) Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia* **168**, 691-701.

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

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

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

Tamura K, Peterson D, Peterson N*, et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**, 2731-2739.

Torimiro JN, Carr JK, Wolfe ND*, et al.* (2006) HLA class I diversity among rural rainforest inhabitants in Cameroon: identification of A*2612-B*4407 haplotype. *Tissue Antigens* **67**, 30-37.

Ujvari B, Belov K (2011) Major Histocompatibility Complex (MHC) Markers in Conservation Biology. *International Journal of Molecular Sciences* **12**, 5168-5186.

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

van Oosterhout C (2013) Maintenance of major histocompatibility supertype variation in selfing vertebrate is no evidence for overdominant selection. *Proceedings of the Royal Society B-Biological Sciences* **280**, 20122501.

Voelker G (1999) Dispersal, vicariance, and clocks: Historical biogeography and speciation in a cosmopolitan passerine genus (*Anthus*: motacillidae). *Evolution* **53**, 1536-1552.

Wakelin D (1996) *Immunity to Parasites: How Parasitic Infections are Controlled*, Second edn. Cambridge University Press, Cambridge.

Wakelin D, Apanius V (1997) *Immune defence: genetic control* Oxford University Press, Oxford.

Westerdahl H, Hansson B, Bensch S, Hasselquist D (2004) Between-year variation of MHC allele frequencies in great reed warblers: selection or drift? *Journal of Evolutionary Biology* **17**, 485-492.

Westerdahl H, Stjernman M, Raberg L, Lannefors M, Nilsson J-A (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, Waldenstrom J, Hansson B*, et al.* (2005) Associations between malaria and MHC genes in a migratory songbird. *Proceedings of the Royal Society B-Biological Sciences* **272**, 1511-1518.

Worley K, Gillingham M, Jensen P*, et al.* (2008) Single locus typing of MHC class I and class IIB loci in a population of red jungle fowl. *Immunogenetics* **60**, 233-247.

Yasukochi Y, Kurosaki T, Yoneda M, Koike H, Satta Y (2012) MHC class II DQB diversity in the Japanese black bear, *Ursus thibetanus japonicus*. *Bmc Evolutionary Biology* **12**, 230.

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.

Appendix 4.1 Description of workflow used to identify chimeras within amplicons.

To test whether a given sequence Z could be a chimera of sequences X and Y, we listed the base pair positions at which Z differed from each of the putative parent sequences. If the first difference from one parent occurred after the last difference from the other (or vice versa), sequence Z was listed as a chimera. This is because such a pattern means that sequence Z is identical to one parent sequence before a given point and identical to the other parent sequence after that point.

In accordance with the rationale of Sommer *et* al's (2013) workflow, we only tested whether a given sequence could be a chimera of two more frequent sequences within the amplicon of interest. For computational efficiency, we restricted the list of putative parent sequences for chimeras to the top twenty most frequent sequences in an amplicon.

Chapter 5

Landscape scale assessment of genetic structure within a wild population reveals the role of an anthropogenic factor in shaping immune gene variation

Road that goes through El Teide Plateau. Photo by Karl Phillips

5.1 Abstract

Understanding the spatial scale at which selection acts upon adaptive genetic variation in natural populations is fundamental to our understanding of evolutionary ecology and has important ramifications for conservation. The environmental factors to which individuals of a population are exposed can vary at fine spatial scales, potentially generating localised patterns of adaptation. Here, we compared patterns of neutral and major histocompatibility complex (MHC) variation within an island population of Berthelot's pipit (*Anthus berthelotii*) to assess whether landscape-level differences in pathogen-mediated selection generate fine-scale spatial structuring in these adaptive immune genes. Specifically, we tested for spatial associations between the distribution of avian malaria, or the factors previously shown to influence that distribution, and MHC variation within resident individuals. Although we found no overall genetic structure across the population for either neutral or MHC loci, we did find localised associations with MHC variation. One MHC class I allele (ANBE48) was directly associated with malaria infection risk, while the presence of the ANBE48 and ANBE38 alleles within individuals correlated (positively and negatively respectively) with distance to the nearest poultry farm, a factor previously shown to be an important determinant of malaria infection in this pipit population. This study highlights the importance of considering small spatial scales when studying the patterns and processes involved in the evolution of variation within adaptive loci such as the MHC.

5.2 Introduction

Understanding the spatial scale at which selection acts upon adaptive genetic variation in natural populations provides information on the degree of local adaptation of populations, and thus, potentially, on the beginnings of speciation (Chave 2013). Furthermore, assessment of the spatial scale of evolutionary processes provides information on epidemiology of wildlife diseases, mechanisms of fluctuating selection, and patterns of dispersal (DeAngelis & Mooij 2005), and should provide background information for delineating conservation strategies. When different groups of individuals evolve in different environments, each becomes adapted to the local conditions. There is a large amount of empirical evidence for such local adaptation (e.g. Gomez-Mestre & Tejedo 2003; Laugen *et al.* 2003; Kawecki & Ebert 2004; Vincent *et al.* 2013), but studies have generally been carried out at coarse scales, with well delineated populations, where differences in environment are conspicuous and limited gene flow does not counteract the effects of selection (Lenormand 2002). However, the environment and its associated selection pressures vary at fine spatial scales (Hargeby *et al.* 2004; Wood *et al.* 2007; Soto-Centeno *et al.* 2013). Selection and adaptation at such fine scales has been increasingly studied in recent years (reviewed in Richardson *et al.* 2014), and there is growing evidence that fine-scale evolutionary divergence is more common than previously thought (Svensson & Sinervo 2004; Ray & King 2006; Mila *et al.* 2010; Richardson & Urban 2013).

How loci that play a role in individual disease resistance/resilience vary across space and in relation to environmental factors is of particular interest when studying local adaptation (Kawecki & Ebert 2004). Pathogens can be strong selective agents in wild host populations (Haldane 1949; Fumagalli *et al.* 2011), and their distribution is highly dependent on environmental factors (Ostfeld *et al.* 2005). For example, both climatic (Harvell *et al.* 2002; LaPointe *et al.* 2010; Becker *et al.* 2012) and anthropogenic factors (Patz *et al.* 2000; Bradley & Altizer 2007; Beadell *et al.* 2009) have been shown to be important determinants of the distribution of pathogenic disease. Furthermore, adaptation of host immune genes to local parasite assemblages appears to be widespread (Evans *et al.* 2010; Tonteri *et al.* 2010; Eizaguirre *et al.* 2012; Lenz *et al.* 2013). Assessment of the scale of selection pressures exerted by pathogens on immune genes is important for understanding the patterns of disease epidemiology and transmission in the landscape.

The genes of the major histocompatibility complex (MHC), with their extraordinary levels of variation and their key role in the vertebrate acquired immune response, have become a classic model for investigating spatio-temporal patterns of selection (reviewed in Hughes & Yeager 1998; Bernatchez & Landry 2003; Spurgin & Richardson 2010). These loci produce cell surface receptors that bind specific peptides derived from intracellular pathogens (antigens) via the peptide binding region (PBR, Wakelin 1996; Wakelin & Apanius 1997; Roitt *et al.* 2001). High genetic variation at the MHC is thought to be driven largely by pathogen-mediated selection (PMS), through three main, non-mutually exclusive, mechanisms (reviewed in Spurgin & Richardson 2010): (i) heterozygote advantage (Doherty & Zinkernagel 1975), (ii) rare allele advantage (Takahata & Nei 1990), and (iii) fluctuating selection (Hill 1991). Sexual selection (reviewed in Edwards & Hedrick 1998; Richardson *et al.* 2005) and other mechanisms may also play a role (van Oosterhout 2009). In an effort to understand the relative role of these different selection mechanisms on the MHC, many studies have investigated among-population structure of MHC genes (reviewed in Bernatchez & Landry 2003; Babik *et al.* 2008; Biedrzycka & Radwan 2008). Nevertheless, despite it being clear that the distribution of pathogens within an environment can vary greatly at small spatial scales (Eisen & Wright 2001; Wood *et al.* 2007), studies which assess the causes and consequences of selection on the MHC at fine spatial scales, within a single population, are lacking. Such studies may provide considerable understanding into how different mechanisms and selective factors act within a population to maintain overall levels of variation, insight which can be obscured if we only focus on coarser patterns of variation.

Fine-scale genetic structure can result from two processes: differential selection pressures (extrinsic factors) that result from fine-scale environmental variation, and endogenous biotic processes (intrinsic factors) particular to the studied species (Legendre & Legendre 2012). These endogenous processes include dispersal patterns (in relation to landscape features), kin structure, mating system and genetic drift (Legendre & Legendre 2012; Wagner & Fortin 2013; Richardson *et al.* 2014). Differentiating between these external and intrinsic processes is vital in order to draw conclusions about the causes and consequences of fine-scale genetic structure and spatial autocorrelation. Several approaches have been proposed to do this (Wagner & Fortin 2005; Dray *et al.* 2006; Jombart *et al.* 2009). One approach, that uses principal components of neighbour matrices (PCNM) to reveal patterns of genetic structure that are not accounted for by environmental gradients (Borcard & Legendre 2002) has been

used successfully in several recent studies (Manel *et al.* 2010; Garroway *et al.* 2013; Pavlova *et al.* 2013). PCNM analysis allows the modelling of spatial patterns, incorporating variation dependent on geographic position (see Dray *et al.* 2006; Legendre & Legendre 2012 for details). Including these derived PCNMs in subsequent models investigating genetic structure accounts for spatial autocorrelation-related processes.

In the present study, we used the PCNM approach to investigate if and why fine-scale genetic structure at neutral markers and MHC class I loci exists within the population of Berthelot's pipit (*Anthus berthelotii*) on Tenerife, in the Canary Islands. This population is isolated from other conspecific populations, and within Tenerife birds are widespread and abundant across the landscape. Importantly, it exhibits a high and spatially varying prevalence of avian malaria (Spurgin *et al.* 2012; Gonzalez-Quevedo *et al.* 2014) which has already been shown to be associated with fine-scale variation in key environmental factors (Gonzalez-Quevedo *et al.* 2014). Although Berthelot's pipit is territorial and sedentary (JC Illera, Pers. Comm.), the population on Tenerife is likely to be interconnected because much of the habitat they use is continuous. However, no studies have investigated how much dispersal and gene flow occurs and whether all the Berthelot's pipits inhabiting Tenerife are part of a large panmictic population. Pipits found on the top of the mountain of El Teide (above 2000 m asl) may be isolated from the rest of the island by the band of pine forest extending from 1600 to 2000 m asl, which the pipit does not inhabit (Garcia-Del-Rey & Cresswell 2007). This Berthelot's pipit population provides an excellent study system in which to test hypotheses related to the spatial scale of PMS effects on MHC variation. Specifically, we sought to i) assess neutral genetic structure in the population, ii) estimate the fine-scale genetic structure of the MHC, and iii) test for associations between the spatial distribution of MHC alleles and both malaria infection risk and other spatially variable environmental factors.

5.3 Methods

5.3.1 Study species and sampling

Berthelot's pipit (*Anthus berthelotii*) is a sedentary passerine endemic to the Macaronesian archipelagos (Fig. 4.1). To obtain a representative sample of the pipit across its entire range and all environmental gradients on Tenerife, a 1 km^2 grid was laid over a map of the island obtained from Google Earth in ArcGIS version 10 (Esri 2011, Redlands, CA, www.esri.com). The majority of accessible square kilometres that contained habitat suitable for pipits were visited

and, where pipits were present, an attempt was made to catch at least one per km^2 using clap nets baited with *Tenebrio molitor* larvae. The GPS coordinates of all visited sites were recorded. Each captured bird was ringed and a blood sample was taken by brachial venipuncture and stored in absolute ethanol in screw-cap micro-centrifuge tubes at room temperature. DNA was extracted using a salt extraction method following Richardson *et al.* (2001).

5.3.2 Genotyping

A total of 388 pipits were genotyped at 21 microsatellite markers (Martínez *et al.* 1999; Dawson *et al.* 2010b; Dawson *et al.* 2012; Dawson *et al.* 2013; Table 5.1). PCRs were set up in four multiplexes, determined by primer compatibility using Multiplex Manager (Holleley & Geerts 2009). Multiplex reactions and details of microsatellite markers are summarized in Table 5.1. PCRs were performed in a DNA Engine Tetrad2 thermal cycler (MJ Research) in 2 µl reactions following Kenta *et al.* (2008): approximately 25 ng of DNA were added to each reaction tube and the liquid evaporated; 1 µl of QIAGEN Multiplex Mastermix (Qiagen) and 1 µl of primer mix (see table 5.1 for primer concentrations) were then added to the dried DNA, with the mixture overlaid by a drop of mineral oil to avoid evaporation during thermal cycling. The thermal profile was as follows: initial activation of 15 mins at 95°C , followed by 40 cycles of 30 seconds at 94°C, 90 seconds at 56°C, and 60 seconds at 60°C, with a final extension of 30 mins at 60°C. PCR reactions were diluted 1:30 with water and 1 µl of the dilution was added to 9 µl of a solution containing Hi-Di Formamide (Life technologies) and ROX 500 (GeneScan) size standard prepared following manufacturer's instructions. PCR products were ran on an ABI3730XL (Life technologies) and allele sizes visualized with the software GENEMAPPER 4.0 (Applied Biosystems).

Pipits were genotyped at the MHC class I by sequencing the exon 3, which codes for the peptide binding region (PBR), using 454 sequencing. Due to cost restrictions, we were only able to run one 454 plate. In order to avoid sacrificing coverage per sample (considering that one 454 run yields *ca.* 1 million reads), samples were pooled in eight pools of 80 samples each. Therefore, 310 pipits out of the 388 were genotyped at the MHC. Ten individuals of the sister species of Berthelot's pipit, the tawny pipit (*Anthus camprestris*) were included in this 454 run for assessment of MHC trans-species evolution (Chapter 4). For detailed methods of the 454 amplification procedure and the bioinformatics analyses performed for the MHC genotyping

see chapter 4. One sample had very poor coverage in the MHC screening and was discarded; therefore analyses are based on a sample of 309 pipits. As is the case for most avian species (Miller & Lambert 2004; Balakrishnan *et al.* 2010; Wutzler *et al.* 2012), because there are linked duplicated loci within the class I MHC, alleles cannot be assigned to specific loci for the pipit (however for simplicity, all variants identified are termed 'alleles' hereafter). Therefore, we characterise all the class I alleles an individual carries irrespective of locus, and use the total number as a measure that reflects individual heterozygosity across the MHC loci (hereafter termed 'MHC diversity').

Table 5.1 Summary of the 21 markers used for assessing neutral genetic variation in 388 Berthelot's pipits (*Anthus berthelotii*) from Tenerife.

Marker	Dye	Min size	Max size	Multiplex reaction	Final concentration in primer mix	Reference
TG01-024	6-FAM	401	403	1	0.2 µM	Dawson *et al.* (2010)
TG01-147	HEX	258	260	3	0.1 µM	Dawson *et al.* (2010)
TG02-088	6-FAM	260	262	1	0.2 µM	Dawson *et al.* (2010)
TG03-002	6-FAM	125	135	1	0.2 µM	Dawson *et al.* (2010)
TG03-098	HEX	233	235	4	0.4 µM	Dawson *et al.* (2010)
TG04-004	HEX	165	169	1	0.2 µM	Dawson *et al.* (2010)
TG05-053	6-FAM	198	202	1	1.0 µM	Dawson *et al.* (2010)
TG06-009	6-FAM	119	123	3	0.1 µM	Dawson *et al.* (2010)
TG09-014	6-FAM	285	293	1	0.2 µM	Dawson *et al.* (2010)
TG13-009	HEX	194	196	2	0.2 µM	Dawson *et al.* (2010)
CAM-2	6-FAM	351	375	3	0.6 µM	Dawson *et al.* (2013)
CAM-4	6-FAM	275	289	2	0.2 µM	Dawson *et al.* (2013)
CAM-8	6-FAM	220	222	4	0.4 µM	Dawson *et al.* (2013)
CAM-13	HEX	209	211	3	0.2 µM	Dawson *et al.* (2013)
CAM-18	HEX	342	348	2	0.2 µM	Dawson *et al.* (2013)
CAM-23	6-FAM	124	148	2	0.2 µM	Dawson *et al.* (2013)
PDO46	DY-549	158	186	2	0.2 µM	Dawson *et al.* (2012)
PDO47	6-FAM	163	199	3	0.2 µM	Dawson *et al.* (2012)
PDOµ5	DY-549	230	244	2	0.2 µM	Dawson *et al.* (2012)
PPi2	HEX	238	258	2	0.2 µM	Martínez *et al.* (1999)
PCA7	HEX	112	116	3	0.1 µM	Dawson *et al*. (2010)

5.3.3 Genetic variation and overall population structure

The software Micro-checker 2.2.3 (Van Oosterhout *et al.* 2004) was used to check for microsatellite null alleles and scoring errors due to stuttering and allele drop out. An exact test

of Hardy-Weinberg equilibrium was performed for each microsatellite locus using the web version of Genepop 4.0.10 (Raymond & Rousset 1995; Rousset 2008). Allele frequencies, observed heterozygosity and expected heterozygosity for each microsatellite locus were calculated using the software CERVUS 3.0.3 (Marshall *et al.* 1998). We estimated individual diversity at microsatellites by calculating homozygosity by loci (HL), a microsatellite heterozygosity-derived measure that weighs the contribution of each locus to the homozygosity value depending on its allelic variability, using the Excel macro Cernicalin (Aparicio *et al.* 2006).

We divided Tenerife into four populations, chosen according to climatic and topographic differences across the island: the south with dry conditions, the north with wet conditions, the west with narrower coastlines and high cliffs, and El Teide with high altitude conditions. We assessed overall (coarse-scale) patterns of genetic structure within the population of pipits on Tenerife. First, an analysis of molecular variance (AMOVA) and microsatellite and MHC F_{ST} calculation (Weir & Cockerham 1984) was performed in Arlequin 3.1 (Excoffier & Lischer 2010) based on the four pre-defined populations. Significance of F_{ST} was evaluated using 50,000 permutations. Second, the programme Structure 2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2003) was used to infer the number of genetic groups (*K*) with individual genotype-based clustering methods using microsatellite data. We used the admixture model and correlated allele frequencies with 100,000 Markov chain steps and a burn-in of 10,000 steps, and performed four independent runs for each value of *K* from 1 to 4. Structure harvester (Earl & vonHoldt 2012) was used to visualize the results. Third, we tested whether pairwise genetic distance, based on microsatellites or MHC, correlates with pairwise geographic distance. Microsatellite genetic distance was obtained by calculating pairwise relatedness using the Queller-Goodnight moment estimator (Queller & Goodnight 1989) implemented in the software COANCESTRY (Wang 2011). For the MHC we calculated a pairwise nucleotide distance matrix for all pairs of MHC alleles by calculating p-distance in MEGA 6.0 (Tamura *et al.* 2013). MHC genetic distance between individuals was calculated as the mean p-distance between all alleles present in the individuals in each dyad. A Mantel test was then performed using the R package Ecodist (Goslee & Urban 2007) with 1,000 permutations and 500 bootstrap iterations to assess correlation of pairwise geographic distance with pairwise microsatellite or MHC genetic distance.

5.3.4 Landscape genetics analyses

All landscape analyses were performed in R 3.1.0 (R Development Core Team 2011) unless stated otherwise. The methods outlined above assess genetic structure at large scales. In order to assess how MHC and microsatellite allele distribution varied across the landscape at a fine scale we performed a spatial principal components analysis (sPCA, Jombart *et al.* 2008) implemented in the R package 'adegenet' (Jombart 2008). The sPCA assesses spatial patterns of genetic variability by finding synthetic components (eigenvectors) that maximize the product of the variance in the data and Moran's I (Jombart *et al.* 2008), the latter being a measure of the spatial dependency (or autocorrelation) associated with that gradient of variation. Each eigenvector captures either positive or negative autocorrelation, and is hence referred to as either 'global' or 'local', respectively (Jombart *et al.* 2008). Local structures (negative autocorrelation) result from greater genetic differences among neighbours, reflecting repulsion processes such as dispersal for inbreeding avoidance. Global structures (positive autocorrelation) arise where there are discrete clusters of genetic similarity or spatial gradients reflecting either isolation by distance and/or isolation by adaptation. For the sPCA we defined neighbouring sites by building a connection network using the minimum distance that would keep all points (individual birds) in the network connected. We mapped the scores at each sampling point for the most important eigenvectors as a means of visualising the spatial genetic structures.

5.3.5 Environmental variables

In a previous study of the spatial distribution of avian malaria in Berthelot's pipits in Tenerife, one strain of malaria, *Plasmodium* LK6, was found to infect 36% of the pipit population (Gonzalez-Quevedo *et al.* 2014). Infection was best predicted by minimum temperature of the coldest month, distance to artificial water sources and distance to poultry farms, in decreasing order of importance (Gonzalez-Quevedo *et al.* 2014). Based on this model, we calculated the predicted probability of an individual being infected with malaria based on the location inhabited (hereafter referred to as 'malaria risk') for each bird. We also directly assessed the separate effects of the following environmental variables on the MHC: minimum temperature of the coldest month, slope, density of pipits, distance to artificial water sources, distance to poultry farms, distance to other livestock farms, and distance to urbanized sites. These variables were chosen on the basis of known effects on disease distribution (Harvell *et al.*

2002; Bradley & Altizer 2007; LaPointe *et al.* 2010; Gonzalez-Quevedo *et al.* 2014) and their potential effects on fine-scale structure at the MHC.

All environmental variable calculations and resampling were carried out in ArcGIS 10 and R (R Development Core Team 2011). Minimum temperature of the coldest month was obtained from the WorldClim database (Hijmans *et al*. 2005) at a resolution of 30 arc seconds (1 km). Slope at a resolution of 90 m was calculated from digital elevation models obtained from the Shuttle Radar Topography Mission Digital Elevation Database 4.1 (Consortium for Spatial Information, www.cgiar-csi.org). Each of these variables was measured within 100 m radius of each sampled bird. For details on methodology see Gonzalez-Quevedo *et al* (2014). Distance variables were calculated by overlaying the layer for pipit location points over polygon layers for artificial water reservoirs, urban areas, and the position, species and census of livestock farms from the government of Tenerife (http://www.tenerife.es/planes/). For each variable the 'proximity' tool of the analysis extension of ArcGIS 10 was used to calculate the distance to the nearest relevant feature for the variable concerned. Right-skew in all our distance variables was successfully removed using a $\log_{10}$-transformation. An index of pipit density was calculated as the number of pipits per square Kilometre, based on our geo-referenced records of pipit presence from the visited sites, using the 'density' tool of the spatial analyst extension in ArcGIS 10, with a neighbourhood size of 2500 m radius around the centre of each square Kilometre sampling cell. This index was used to reflect the size of the subpopulation of pipits found in the same area as the sampled pipit and thus to provide a measure of the local conspecific host population.

5.3.6 Models of malaria risk

We assessed whether individual malaria risk and malaria infection status was associated with genetic diversity at microsatellites or the MHC using a general linear models (LM) for malaria risk, or generalised linear model (GLM) for malaria infection. Malaria risk was logit-transformed prior to fitting LMs. We investigated whether there was an association between individual microsatellite diversity with malaria risk with an LM of malaria risk as the response variable and HL as explanatory variable. We then tested the association between malaria risk and MHC class I diversity by running an LM of malaria risk as the response variable and MHC diversity as an explanatory variable. We also assessed which MHC class I alleles best explained malaria risk using a multi-predictor model of malaria risk as response variable and including all

MHC alleles as explanatory variables. We checked for spatial autocorrelation in model residuals by building correlograms with a 1,000 m distance increment and resampling 1,000 times at each distance class, implemented in the R package ncf (Bjornstad 2012). Where residual spatial autocorrelation was present, we tried accounting for it by using simultaneous autoregressive (SAR) models (Kissling & Carl 2008) implemented in the R package spdep (Bivand 2012), specifying an appropriate neighbourhood size within which autocorrelation is accounted for.

5.3.7 Models of MHC variation in the landscape

We used environmental and spatial predictors, respectively, to assess the extent to which extrinsic and intrinsic factors explain MHC variation, both in terms of diversity and specific alleles. We selected our spatial predictors from a set of principal components of neighbour matrices (PCNMs, Dray *et al.* 2006), computed in the R package PCNM (Borcard *et al.* 2011; Legendre *et al.* 2013) as follows: First, we computed a pairwise geographic distance matrix between all individuals sampled. Second, we chose a threshold (t) value of 5 km to construct a truncated distance matrix; if pairwise distance between two samples was higher than this threshold the value in the truncated distance matrix was assigned the value 4t (20 km), and if the pairwise distance was lower than the threshold, the values remained invariant. Third, we performed a principal coordinates analysis (PCoA) of the truncated distance matrix to obtain PCNM base functions associated with the spatial distribution of sampling sites (in our case, individual birds). Selection of the subset of PCNMs used in models (above) was based on a redundancy analysis (RDA) using individual MHC genotypes as multivariate response and the PCNMs selected from the previous step as predictors. A forward selection procedure was used to select PCNMs that were significantly correlated with the distribution of MHC alleles across samples (pipits). To minimise type I error rates, forward selection employed a double stopping criterion described in Blanchet *et al.* (2008) to test for significance of PCNMs: if a PCNM increased the α above 0.05 or raised the adjusted R^2 above that of the global model (with all PCNMs included) it was not retained. The PCNMs identified by the RDA forward selection as being significantly correlated with the MHC allele distribution were included as additional predictors in the general linear models (LM) of MHC diversity as response variable, and in generalized linear models (GLMs) of each MHC allele as response variable. PCNMs have been used as predictor variables in environmental models to account for variation in a response variable that would otherwise be retained as spatially autocorrelated residual variation

(Borcard & Legendre 2002; Borcard *et al.* 2011). Here we use PCNMs to characterise the spatial variation in our genetic data that is not explained by the environmental gradients we measure, and is therefore attributable to isolation by distance/adaptation. In doing so, we simultaneously deal with the statistical issue of spatial autocorrelation in model residuals biasing environmental parameter estimates and erroneously inflating their levels of certainty (Legendre 1993). Nevertheless, we checked the success of the models that included PCNMs in removing spatial autocorrelation in model residuals using correlograms as detailed in the previous section. Spatial autocorrelation in residuals of these models was not significant, thus showing that any spatial autocorrelation in models residuals was successfully accounted for.

We explored the associations between MHC diversity and the environmental variables using multi-predictor models of MHC diversity as the response variable including the previously selected PCNMs as additional predictors to the full model of MHC diversity with all environmental variables as predictors. In order to test whether separate environmental variables were associated with specific MHC class I alleles we performed allele-by-allele multi-predictor generalized linear models of each MHC allele as response variable fitting all possible combinations of environmental variables and previously selected PCNMs as predictors. For this purpose we used a model selection approach (Burnham & Anderson 2001) and compared the relative fit of models using the Akaike information criterion (AIC). We ranked all resulting models for each MHC allele according to their AIC and considered models with ΔAIC (the difference between the best model's AIC and that of the model in question) ≤ 2 as having sufficient support (Burnham & Anderson 2001). To estimate the relative importance of predictors we performed model averaging on the models with $\Delta AIC \leq 2$. All model selection calculations were performed using the R package MuMIn (Barton 2013). Model selection is a valuable alternative to traditional null hypothesis testing (Johnson & Omland 2004; Whittingham *et al.* 2006; Burnham *et al.* 2011; Dochtermann & Jenkins 2011), and is being used increasingly in studies of disease ecology (Moore & Borer 2012; Manzoli *et al.* 2013). On all occasions when multi-predictor models were built we checked our final models by comparing them with a series of single-predictor models to ensure consistency of results.

Table 5.2 Summary of 19 microsatellite marker diversity in Berthelot's pipits (*Anthus berthelotii*) of Tenerife. k - number of alleles, H_{Obs} - Observed heterozygosity, H_{Exp} - Expected heterozygosity. Markers TG05-053 and CAM2 showed evidence of null alleles and were therefore excluded from analyses.

Locus	k	H_{Obs}	H_{Exp}
TG01-024	5	0.064	0.099
TG01-147	2	0.466	0.479
TG02-088	2	0.023	0.023
TG03-002	8	0.335	0.373
TG03-098	3	0.341	0.629
TG04-004	5	0.382	0.387
TG06-009	3	0.373	0.385
TG09-014	8	0.561	0.782
TG13-009	4	0.057	0.061
CAM4	4	0.687	0.704
CAM8	2	0.452	0.484
CAM13	3	0.244	0.264
CAM18	4	0.611	0.655
CAM23	13	0.760	0.784
PDO46	6	0.366	0.353
PDO47	17	0.860	0.894
PDOμ5	10	0.576	0.583
PPi2	13	0.811	0.852
PCA7	2	0.163	0.163

5.4 Results

5.4.1 Neutral and MHC genetic diversity

A total of 21 microsatellite markers were screened in 388 individuals. Two microsatellite markers showed evidence of homozygote excess and presence of null alleles and were therefore excluded from further analyses. The other 19 microsatellites did not deviate from Hardy-Weinberg equilibrium (Table 5.2). The number of alleles per locus ranged from 2 to 17 and observed heterozygosity ranged from 0.023 to 0.894 (Table 5.2). The MHC genotyping results are described in detail in chapter 4. A total of 22 MHC class I alleles were identified in the population of pipits on Tenerife. Two of these alleles, which had frequencies of 0.84 and 0.82, were found to have very low amplification efficiencies and are therefore likely to suffer from allelic dropout. This could therefore mean that they are present in most, if not all of the individuals but missed by the screening process in some cases. To avoid this uncertainty

confounding the spatial analysis these two alleles were excluded from the analysis. After removing these two alleles, pipits in Tenerife each had between three and ten MHC class I alleles (Fig. 5.1). Allele ANBE7 was also excluded from the spatial analysis because it was fixed in the population.

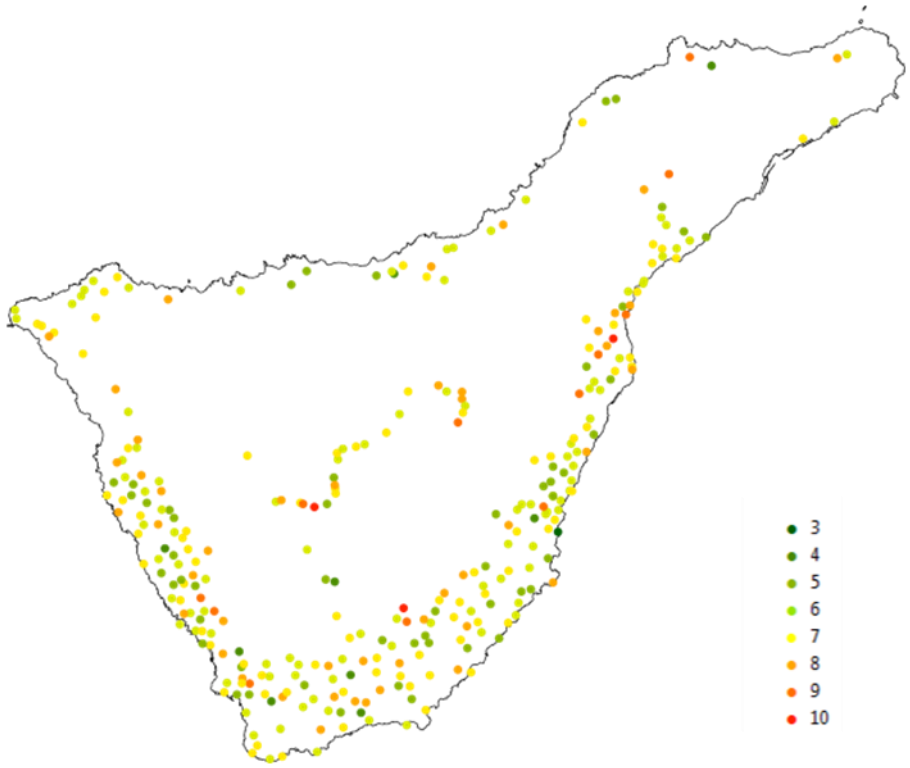


Figure 5.1 Number of MHC class I alleles identified in individual Berthelot's pipits (*Anthus berthelotii*) across Tenerife. Each dot represents an individual.

5.4.2 Overall population genetic structure

Microsatellite F_{ST} revealed very low levels of differentiation among pipits inhabiting the four pre-defined zones of Tenerife (F_{ST} = 0.008, *P* = 0.003). For the MHC the levels of differentiation were even lower, and non-significant (F_{ST} = 0.001, *P* = 0.770). The STRUCTURE analysis indicated that K =1 was the most likely number of genetic clusters, i.e. all individuals had approximately the same probability of belonging to any of the simulated populations. Pairwise genetic distance based on microsatellites or the MHC was not significantly associated with geographic distance (microsatellites: r = -0.015, *P* = 0.339; MHC: r = 0.038, *P* = 0.104). Therefore both analyses based on neutral and MHC genetic variation indicate that the pipits on Tenerife exist as one single population with little or no genetic structure.

5.4.3 Landscape genetics analyses

The sPCA of microsatellite genotypes showed that there was no evidence of global structure (*P* = 0.437). While the test for local structure was not formally significant (*P* = 0.055), the plot of the eigenvector with the largest negative eigenvalue suggested short-scale spatial structure indicative of some dissimilarity among neighbours (Fig. 5.2). The sPCA of MHC class I genotypes also did not reveal significant patterns of global (*P* = 0.587) or local (*P* = 0.732) structure. Furthermore, when visually scrutinised, none of the global (positive eigenvalues) or local axis (negative eigenvalues) revealed any obvious spatial structuring in the distribution of MHC alleles. So, overall there was no evidence for the apparent presence of spatial genetic clusters and no evidence for a genetic cline across the landscape.

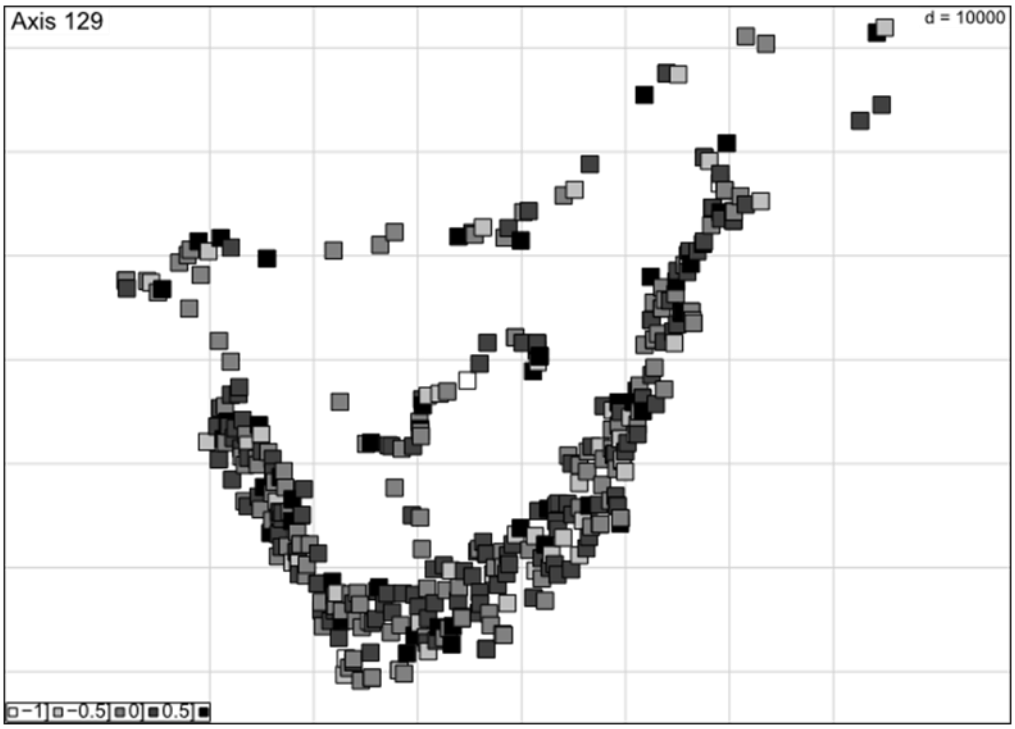


Figure 5.2 Scores from the largest local eigenvector (Axis 129, P = 0.055) of spatial principal component analysis of microsatellite allele frequencies of Berthelot's pipits (*Anthus berthelotii*) in Tenerife. Each sampled location has a score that ranges from -1 (white squares) to 0.5 (black squares). Significance of local eigenvectors reveals negative spatial autocorrelation, hence differentiation between neighbouring sites.

Table 5.3 Summary of general linear models predicting malaria risk (LK6_Prob) in Berthelot's pipits (*Anthus berthelotii*) in Tenerife fitting (a) number of MHC alleles, (b) each MHC allele in single predictor models, and (c) all MHC alleles in a multi predictor model. Significance of predictors is designated with an asterisk (P <0.05).

	Response	Predictor	Coefficient	R^2
a				
	LK6_Prob	Nalleles	-0.009	0.001
b				
	LK6_Prob	ANBE1	0.023	0.001
	LK6_Prob	ANBE2	0.037	-0.002
	LK6_Prob	ANBE4	-0.039	0.006
	LK6_Prob	ANBE6	-0.024	-0.002
	LK6_Prob	ANBE7	-0.188	<0.001
	LK6_Prob	ANBE8	0.016	-0.002
	LK6_Prob	ANBE9	-0.012	-0.003
	LK6_Prob	ANBE10	0.030	-0.001
	LK6_Prob	ANBE11	0.038	0.003
	LK6_Prob	ANBE13	-0.012	-0.003
	LK6_Prob	ANBE16	-0.032	<0.001
	LK6_Prob	ANBE28	0.117	0.001
	LK6_Prob	ANBE38	0.019	-0.002
	LK6_Prob	ANBE43	-0.005	-0.003
	LK6_Prob	ANBE44	-0.034	0.002
	LK6_Prob	ANBE45	0.044	<0.001
	LK6_Prob	ANBE46	-0.066	0.006
	LK6_Prob	ANBE47	-0.028	0.002
	LK6_Prob	ANBE48	-0.113*	0.010
	LK6_Prob	ANBE49	-0.004	-0.003
c				
	LK6_Prob	ANBE10	0.022	0.015
		ANBE2	0.037	
		ANBE8	<0.001	
		ANBE4	-0.030	
		ANBE43	-0.003	
		ANBE1	0.014	
		ANBE45	0.046	
		ANBE7	-0.358	
		ANBE13	-0.016	
		ANBE9	-0.010	
		ANBE16	-0.033	
		ANBE47	-0.029	
		ANBE11	0.041	
		ANBE28	0.088	
		ANBE6	-0.011	
		ANBE48	-0.142*	
		ANBE49	-0.010	
		ANBE38	0.003	
		ANBE44	-0.035	
		ANBE46	-0.075	

5.4.4 Models of malaria risk

Malaria risk was not associated with either microsatellite diversity (P = 0.327, R^2 = 0.001), nor with MHC class I diversity (P = 0.277, R^2 = 0.001, Table 5.3a). However, when testing the association between all MHC class I alleles and malaria risk in a multi-predictor model, the presence of the ANBE48 allele did significantly correlate with malaria risk, although the explanatory power was low (P = 0.019; R^2 of multi-predictor model with all alleles as predictors = 0.015, R^2 of single-predictor model with ANBE48 as predictor = 0.010; Table 5.3b and c). None of the alleles or MHC diversity had a significant relationship with malaria infection status. Explanatory power of these models was less than 0.009.

Table 5.4 Summary of general linear models predicting MHC diversity ($N_{alleles}$) per individual in Berthelot's pipits (*Anthus berthelotii*) in Tenerife fitting (a) environmental variables and PCNMs as single predictors, (b) all environmental variables in a multi predictor model and (c) all environmental variables and PCNMs in a multi predictor model. Significance of predictors is designated by asterisks (* <0.05, ** <0.01, ***<0.001).

Response	Predictor	coefficient	R2
$N_{alleles}$	Poultry	-0.513*	0.034
	Water	0.924	
	Temperature	0.191	
	Slope	0.054**	
	Farm	0.251	
	Pipit density	-0.114	
	Urbanization	0.030	
	PCNM3	-1.91×10^{-5}	
	PCNM4	2.61×10^{-5}	
	PCNM9	-3.67×10^{-5}	
	PCNM13	-4.98×10^{-5}	
	PCNM23	2.41×10^{-5}	
	PCNM24	2.66×10^{-5}	
	PCNM87	-1.74×10^{-5}	

5.4.5 Models of MHC variation in the landscape

The PCoA performed on spatial coordinates of samples identified 176 PCNMs of which 89 had significant Moran's I values, the first six being positive Moran's I values, while the remaining 83 were negative Moran's I values. The overall redundancy analysis (RDA) indicated that there

was a weak association between these PCNMs and the MHC allele distribution (adjusted R^2 = 0.022, *P* = 0.030). The forward selection retained seven PCNMs: numbers 3, 4, 9, 13, 23, 24 and 87 (adjusted R^2 = 0.028, *P* = 0.005), which modelled spatial variation at different scales (Supplementary Fig. S5.1). PCNMs 3 and 4 are positive and reflect relatively large-scale spatial structures, while the other five PCNMs are negative and reflect intermediate- to fine-scale spatial structures.

In the multi-predictor model of MHC diversity as the response variable, 'slope' and 'distance to poultry farm' were the only significant environmental predictors (*P* = 0.002 and 0.014, respectively; R^2 of full model = 0.064, Table 5.4). In combination, the PCNMs accounted for an additional 3% of variance in the response and residual spatial autocorrelation was successfully removed. Nevertheless, overall explanatory power remained low and none of the PCNMs were associated with MHC diversity.

Model selection and model averaging results of the multi-predictor GLMs investigating the presence/absence of MHC alleles (response variables) in relation to environmental variables and significant PCNMs as predictors showed different patterns for different alleles. The highest reduction in AIC compared to the null model was obtained for the model predicting ANBE48 (ΔAIC = 18.5, Table 5.5). Model averaging shows that distance to poultry farms has a relative importance of 1 in explaining ANBE48 distribution, while the other environmental variables had a relative importance of 0.1. The best model for ANBE48 showed the highest explanatory power among all alleles (R^2 = 0.310) and included a positive association with distance to poultry farms and PCNMs 3 and 23 (Table 5.5, Fig. 5.3). The next highest explanatory power (R^2 = 0.210) was obtained for the best model of ANBE38 which (interestingly) included a negative association with distance to poultry farms, as well as a positive correlation with slope, and PCNMs 3 and 4 (Table 5.5, Fig. 5.3). The explanatory power of the multi-predictor models for the other alleles was relatively low ($R^2 \leq 0.17$, Supplementary Table S5.1), indicating that association of environmental and spatial predictors on other alleles are lower than those identified for ANBE38 and ANBE48.

Single-predictor models broadly supported the results of multi-predictor models in this study, confirming, for example, the relative importance of poultry farms in explaining ANBE48 and ANBE38 distribution (R^2 = 0.154 and 0.051, respectively), and the relatively low explanatory

power of many of the individual predictor-response relationships investigated (Supplementary Table S5.2).

Table 5.5 Summary of model selection performed on multi-predictor generalized linear models (with ΔAIC ≤ 2 when compared to the best fit model) of Berthelot's pipits (*Anthus berthelotii*) MHC class I alleles ANBE38 and ANBE48 in Tenerife as response variables. Significant coefficients are underlined, level of significance is designated by asterisks (* <0.05, ** <0.01, ***<0.001), and relative importance in explaining variation in the presence of each allele after model averaging is shown in brackets.

	ANBE38	ANBE48
Poultry	-2.03** (1.00)	5.27** (1.00)
Water	0.98 (0.75)	-0.24 (0.08)
Slope	0.12* (1.00)	0.02 (0.07)
Temperature	-0.01 (0.04)	-0.06 (0.09)
Farm	1.30 (0.38)	0.52(0.09)
Urbanization	-0.72 (0.31)	0.50 (0.09)
Pipit density	-0.04 (0.04)	1.02 (0.08)
PCNM3	-8.63×10^{-5} (0.20)	4.20×10^{-4}* (1.00)
PCNM4	2.61×10^{-4}** (1.00)	3.36×10^{-5}* (0.04)
PCNM9	2.83×10^{-5} (0.04)	1.79×10^{-4} (0.12)
PCNM13	3.34×10^{-4}* (1.00)	-9.87×10^{-6} (0.04)
PCNM23	-1.82×10^{-4} (0.20)	-1.78×10^{-4}* (1.00)
PCNM24	-7.86×10^{-5} (0.05)	2.05×10^{-4} (0.10)
PCNM87	-7.33×10^{-6} (0.04)	5.53×10^{-4} (0.44)
AIC Null[1]	139.30	90.30
AIC best[2]	125.36	71.78
Δ AIC	13.94	18.52
R^2 best[3]	0.21	0.31

[1]AIC of the model with only the intercept
[2]AIC of the model with the lowest AIC compared to all other possible models
[3]Explanatory power of the model with the lowest AIC

5.5 Discussion

In this study we assessed fine-scale structure at neutral markers and at MHC loci in relation to environmental factors within a population of Berthelot's pipit. We found no evidence of consistent population-wide genetic structure at either the neutral or functional markers. Nevertheless, we did find a weak association between malaria risk and one specific MHC allele (ANBE48). Furthermore when taking into account spatial processes independent of

environmental gradients, we found stronger, opposing, associations between two MHC alleles (ANBE48 and ANBE38) and an anthropogenic environmental variable (distance to poultry farms) already known to be important in disease transmission in the pipit.

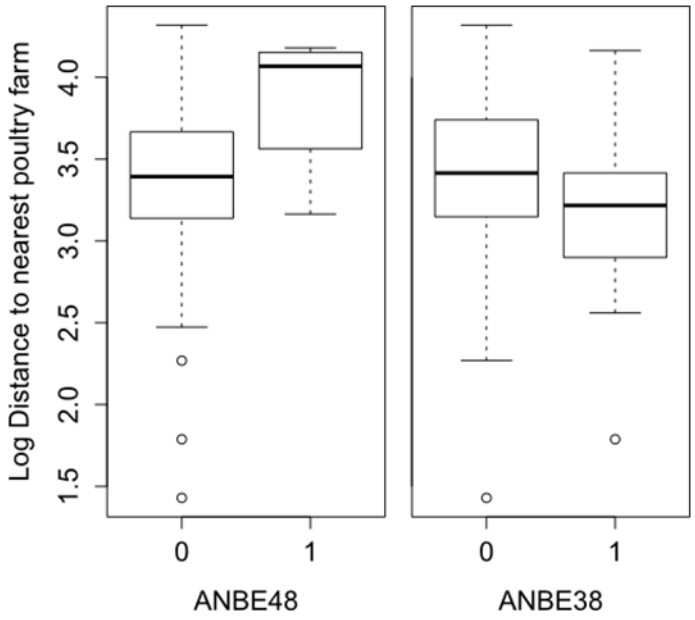


Figure 5.3 Distribution of Log (Distance to nearest poultry farms) across individual Berthelot's pipits carrying (1) or not carrying (0) alleles ANBE38 and ANBE48. Centre lines show the medians. Box limits indicate the 25th and 75th percentiles and outliers are represented by open dots.

The STRUCTURE and F_{ST} analyses based on neutral markers and four pre-defined populations revealed that there was no evidence of separate populations: Berthelot's pipits in Tenerife are part of one panmictic population. The Mantel test which found no correlation between MHC genetic distance and geographic distance identified the same pattern of panmixia for these adaptive loci. These results were further confirmed by the spatial PCA analyses, which indicated that there is no overall pattern of spatial genetic structure at either type of marker. This study therefore indicates that neither climatic differences nor apparent barriers to dispersal across Tenerife impede gene flow between different areas of the island. The opposite has been found in other systems where intra-population spatial clines and global (positively

spatially autocorrelated) structures in allele frequencies of adaptive loci have been described at relatively small spatial scales (Garroway *et al.* 2013; reviewed in Richardson *et al.* 2014).

The lack of local or global structure at the MHC was further confirmed by the PCNM analyses. The significant PCNMs retained after the RDA performed on the MHC genotypes modelled spatial variation at different scales: two of them were positive and reflected relatively large-scale spatial structures, while the other five PCNMs were negative and reflected intermediate- to fine-scale spatial structures. This suggests that spatial variation at the MHC decomposes into a complex set of spatial structures, suggesting different scales of influence on different alleles or groups of alleles. This can be expected if a few specific alleles are under selective pressure, while others are evolving under neutrality, which might occur if only some MHC alleles confer resistance or susceptibility to the particular pathogens to which Berthelot's pipits are currently exposed in Tenerife. This pattern of structure at specific MHC alleles is swamped when performing analyses with individual genotypes, and single allele effects cannot be revealed. Therefore, analyses performed on individual alleles are needed in any fine-scale genetic structure analysis.

Individual MHC alleles produce molecules which are able to bind subsets of specific pathogen derived peptides and thus trigger the appropriate immune responses to those pathogens (Wakelin 1996). Thus MHC characteristics can be linked with pathogens in two ways. First, specific MHC alleles can confer resistance or susceptibility to a specific pathogen, and under this scenario we would expect a correlation between allele presence/absence and disease (e.g. Meyer-Lucht & Sommer 2005; Bonneaud *et al.* 2006; Schad *et al.* 2012; Zhang & He 2013). Second, individuals with greater allelic diversity may be better at responding to both individual pathogens and to the diversity of pathogens in the environment; if this is the case we expect MHC diversity to be negatively associated with disease (Westerdahl *et al.* 2005; e.g.Kloch *et al.* 2010; but see Radwan *et al.* 2012). In the present study we found no association of malaria risk with MHC diversity of an individual. On the other hand, we did find that allele ANBE48 was significantly associated with a reduced malaria risk, although this relationship was too weak (R^2 = 0.010 for the single predictor model, Table 5.3b) to be able to draw definitive conclusions. However, while our explanatory power was relatively low in the present study, our results do concur with other studies that have found MHC alleles that confer resistance/susceptibility to

disease (Collins *et al.* 1977; Croisetiere *et al.* 2008; Xu *et al.* 2010; Biedrzycka *et al.* 2011; Kloch *et al.* 2013), including malaria (Hill *et al.* 1991; Bonneaud *et al.* 2006; Westerdahl *et al.* 2013).

The effect of different climatic variables on spatial distribution of immune genes has been previously documented (Tonteri *et al.* 2010), and spatially variable selection on specific MHC alleles has been reported at large scales (Landry & Bernatchez 2001; Ekblom *et al.* 2007). Interestingly, in Berthelot's pipits we found that distribution of allele ANBE48 was positively associated with distance to poultry farms, a variable previously found to have a negative association with malaria infection in pipits in Tenerife (Gonzalez-Quevedo *et al.* 2014). In short, the closer an individual pipit was to a poultry farm, where malaria transmission has been shown to be higher, the less likely it was to be carrying the ANBE48 allele (Fig. 5.3). Distance to a poultry farm explained 15% of the variation in the distribution of ANBE48, and was the most important variable in the best model for this allele. Another allele, ANBE38, was negatively associated with distance to poultry farms (Fig. 5.3), although the amount of variation in ANBE38 explained by this variable was not as large ($R^2 = 0.051$) as for ANBE48. That we found no effect of microsatellite or MHC diversity suggests that this result is not explained by genome-wide diversity but is directly associated with the alleles identified. It is important to note that PCNM23 had a strong association with ANBE48 ($R^2 = 0.105$, $P < 0.01$), but this PCNM does not explain the distribution of any other MHC alleles (Table S5.2). This therefore indicates that spatial variation in ANBE48 is not attributable to isolation by distance, as otherwise other alleles would show very similar spatial associations. An alternative explanation is that a spatial gradient in a currently unmeasured environmental factor may be exerting a selective pressure on this allele.

Given that, 1) ANBE48 only explained 1% variation in an index of malaria risk, 2) ANBE38 was not significantly associated with malaria risk, and, 3) both ANBE48 and ANBE38 were much more significantly associated with distance to poultry farms, we hypothesize that these alleles are linked with the incidence of diseases other than malaria. The relationship that poultry farms have with the transmission of other avian diseases that may interact with the MHC, and possibly affect the survival of pipits, has yet to be explored. However, we do know that other pathogens exist within this population (Spurgin *et al.* 2012), such as avian pox and avian malaria of the genus *Leucozytozoon*. Therefore, assessment of the interaction of these pathogens with the MHC and of their association with poultry farms might give some insight

into the mechanisms behind the association we found of ANBE48 and ANBE38 with distance to poultry farms.

Previous work has provided considerable evidence that selection has shaped MHC class I variation in Berthelot's pipit (Spurgin *et al.* 2011; Chapter 4). In particular, selection appears to be focused on the key peptide binding region (PBR; Chapter 4), which comprises 15 amino acids within the exon 3 (Brown *et al.* 1993). Among the alleles found in the Berthelot's pipits in Tenerife, allele ANBE48 has a unique PBR, different from that of allele ANBE38 at nine amino acids, suggesting these two alleles have very different binding properties (see chapter 4) and supporting our finding of opposing effects associated with these alleles. In line with this reasoning, the most logical explanation we can put forward for the results that we find is that the ANBE48 allele makes an individual susceptible to a pathogen that exists (or is at higher levels) around poultry farms, while ANBE38 is a non-susceptible alternative allele. Whatever the specific pathogen, we hypothesise that birds that live close to poultry farms have a higher risk of contracting a pathogenic disease and die if they carry ANBE48. Another possible explanation is that birds living close to poultry farms which carry ANBE48 have a higher probability of being sick, and are not active, and that, these birds are thus not caught using our method.

Understanding the mechanisms that drive fine-scale genetic structure at adaptive loci is vital in evolutionary research (Richardson *et al.* 2014). To our knowledge, this is the first study to show an effect of an environmental variable on MHC variation at the intra-population level. The fact that this environmental variable (the presence of poultry farms) has an anthropogenic source has considerable implications for understanding evolution in the context of global change and human impact on disease transmission in wild populations. Moreover, this study highlights the importance of considering fine spatial scales when assessing patterns of selection at adaptive loci. Key patterns and associations may be overlooked when we lump together within-population variation to assess differences at greater scales, potentially undermining our understanding of the factors and mechanisms that drive the evolution of the loci and species in question. Furthermore, understanding the scales, speed and causes of local adaptation within a species can have important implications for conservation, particularly when populations are challenged by new factors induced by environmental changes, either because of restorative translocations or because of habitat disturbance.

5.6 References

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

Babik W, Pabijan M, Radwan J (2008) Contrasting patterns of variation in MHC loci in the Alpine newt. *Molecular Ecology* **17**, 2339-2355.

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

Barton K (2013) MuMIn: Multi-model inference. R package version 1.9.13.

Beadell JS, Covas R, Gebhard C*, et al.* (2009) Host associations and evolutionary relationships of avian blood parasites from West Africa. *International Journal for Parasitology* **39**, 257-266.

Becker CG, Rodriguez D, Longo AV, Talaba AL, Zamudio KR (2012) Disease risk in temperate amphibian populations Is higher at closed-canopy sites. *PloS one* **7**.

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

Biedrzycka A, Kloch A, Buczek M, Radwan J (2011) Major histocompatibility complex DRB genes and blood parasite loads in fragmented populations of the spotted suslik *Spermophilus suslicus*. *Mammalian Biology* **76**, 672-677.

Biedrzycka A, Radwan J (2008) Population fragmentation and major histocompatibility complex variation in the spotted suslik, *Spermophilus suslicus*. *Molecular Ecology* **17**, 4801-4811.

Bivand R (2012) spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-46.

Bjornstad ON (2012) ncf: spatial nonparametric covariance functions. R package version 1.1-4.

Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. *Ecology* **89**, 2623-2632.

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

Borcard D, Gillet F, Legendre P (2011) *Numerical ecology with R* Springer, London.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* **153**, 51-68.

Bradley CA, Altizer S (2007) Urbanization and the ecology of wildlife diseases. *Trends in Ecology & Evolution* **22**, 95-102.

Brown JH, Jardetzky TS, Gorga JC*, et al.* (1993) 3-Dimensional structure of the human class II-histocompatinility antigen HLA-DR1. *Nature* **364**, 33-39.

Burnham KP, Anderson DR (2001) *Model selection and multi-model inference: a practical information–theoretic approach.* Springer, New York, NY.

Burnham KP, Anderson DR, Huyvaert KP (2011) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* **65**, 23-35.

Chave J (2013) The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecology Letters* **16**, 4-16.

Collins WM, Briles WE, Zsigray RM*, et al.* (1977) The B locus (MHC) in the chicken: Association with the fate of RSV-induced tumors. *Immunogenetics* **5**, 333-343.

Croisetiere S, Tarte PD, Bernatchez L, Belhumeur P (2008) Identification of MHC class II beta resistance/susceptibility alleles to *Aeromonas salmonicida* in brook charr (*Salvelinus fontinalis*). *Molecular Immunology* **45**, 3107-3116.

Dawson DA, Ball AD, Spurgin LG*, et al.* (2013) High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species. *Bmc Genomics* **14**.

Dawson DA, Horsburgh GJ, Krupa AP*, et al.* (2012) Microsatellite resources for Passeridae species: a predicted microsatellite map of the house sparrow *Passer domesticus*. *Molecular Ecology Resources* **12**, 501-523.

Dawson DA, Horsburgh GJ, Kupper C*, et al.* (2010a) New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility - as demonstrated for birds. *Molecular Ecology Resources* **10**, 475-494.

Dawson DA, Horsburgh GJ, Kupper C*, et al.* (2010b) New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility - as demonstrated for birds. *Molecular Ecology Resources* **10**, 475-494.

DeAngelis DL, Mooij WM (2005) Individual-based modeling of ecological and evolutionary processes. In: *Annual Review of Ecology Evolution and Systematics*, pp. 147-168.

Dochtermann NA, Jenkins SH (2011) Developing multiple hypotheses in behavioral ecology. *Behavioral Ecology and Sociobiology* **65**, 37-45.

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

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling* **196**, 483-493.

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

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

Eisen RJ, Wright NM (2001) Landscape features associated with infection by a malaria parasite (*Plasmodium mexicanum*) and the importance of multiple scale studies. *Parasitology* **122**, 507-513.

Eizaguirre C, Lenz TL, Kalbe M, Milinski M (2012) Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecology Letters* **15**, 723-731.

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

Evans ML, Neff BD, Heath DD (2010) MHC-mediated local adaptation in reciprocally translocated Chinook salmon. *Conservation Genetics* **11**, 2333-2342.

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

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Fumagalli M, Sironi M, Pozzoli U*, et al.* (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *Plos Genetics* **7**, e1002355.

Garcia-Del-Rey E, Cresswell W (2007) The breeding biology of the endemic Berthelot's Pipit, *Anthus berthelotii*, in a harsh oceanic island environment (Tenerife, Canary Islands). *Ostrich* **78**, 583-589.

Garroway CJ, Radersma R, Sepil I*, et al.* (2013) Fine-scale genetic structure in a wild bird population: the role of limited dispersal and environmentally based selection as causal factors. *Evolution* **67**, 3488-3500.

Gomez-Mestre I, Tejedo M (2003) Local adaptation of an anuran amphibian to osmotically stressful environments. *Evolution* **57**, 1889-1899.

Gonzalez-Quevedo C, Davies RG, Richardson DS (2014) Predictors of malaria infection in a wild bird population: landscape-level analyses reveal climatic and anthropogenic factors. *Journal of Animal Ecology* **83**, 1091–1102.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**, 1-19.

Haldane JBS (1949) Disease and evolution *La Ricerca Scientifica Suplemento* **19**, 1-11.

Hargeby A, Johansson J, Ahnesjo J (2004) Habitat-specific pigmentation in a freshwater isopod: Adaptive evolution over a small spatiotemporal scale. *Evolution* **58**, 81-94.

Harvell CD, Mitchell CE, Ward JR*, et al.* (2002) Climate warming and disease risks for terrestrial and marine biota. *Science* **296**, 2158-2162.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.

Hill AVS (1991) HLA associations with malaria in Africa: some implications for MHC evolution. In: *Molecular Evolution of the Major Histocompatibility Complex* (eds. Klein J, Klein D), pp. 403-420. Springer Berlin Heidelberg.

Hill AVS, Allsopp CEM, Kwiatkowski D*, et al.* (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature* **352**, 595-600.

Holleley CE, Geerts PG (2009) Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *Biotechniques* **46**, 511-517.

Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics* **32**, 415-435.

Johnson JB, Omland KS (2004) Model selection in ecology and evolution. *Trends in Ecology & Evolution* **19**, 101-108.

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

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* **101**, 92-103.

Jombart T, Dray S, Dufour A-B (2009) Finding essential scales of spatial variation in ecological data: a multivariate approach. *Ecography* **32**, 161-168.

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

Kenta T, Gratten J, Haigh NS*, et al.* (2008) Multiplex SNP-SCALE: a cost-effective medium-throughput single nucleotide polymorphism genotyping method. *Molecular Ecology Resources* **8**, 1230-1238.

Kissling WD, Carl G (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography* **17**, 59-71.

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

Kloch A, Baran K, Buczek M, Konarzewski M, Radwan J (2013) MHC influences infection with parasites and winter survival in the root vole *Microtus oeconomus*. *Evolutionary Ecology* **27**, 635-653.

Landry C, Bernatchez L (2001) Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). *Molecular Ecology* **10**, 2525-2539.

LaPointe DA, Goff ML, Atkinson CT (2010) Thermal constraints to the sporogonic development and altitudinal distribution of avian malaria *Plasmodium relictum* in Hawaii. *Journal of Parasitology* **96**, 318-324.

Laugen AT, Laurila A, Rasanen K, Merila J (2003) Latitudinal countergradient variation in the common frog (*Rana temporaria*) development rates - evidence for local adaptation. *Journal of Evolutionary Biology* **16**, 996-1005.

Legendre P (1993) Spatial autocorrelation - trouble or new paradigm *Ecology* **74**, 1659-1673.

Legendre P, Borcard D, Blanchet G, Dray S (2013) PCNM: MEM spatial eigenfunction and principal coordinate analyses. R package version 2.1-2/r109. Available from http://R-Forge.R-project.org/projects/sedar/.

Legendre P, Legendre L (2012) *Numerical ecology* Elsevier, Amsterdam, The Netherlands.

Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* **17**, 183-189.

Lenz TL, Eizaguirre C, Rotter B, Kalbe M, Milinski M (2013) Exploring local immunological adaptation of two stickleback ecotypes by experimental infection and transcriptome-wide digital gene expression analysis. *Molecular Ecology* **22**, 774-786.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology* **19**, 3824-3835.

Manzoli DE, Antoniazzi LR, Saravia MJ*, et al.* (2013) Multi-level determinants of parasitic fly infection in forest passerines. *PloS one* **8**, e67104.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* **7**, 639-655.

Martínez JG, Soler JJ, Soler M, Moller AP, Burke T (1999) Comparative population structure and gene flow of a brood parasite, the great spotted cuckoo (*Clamator glandarius*), and its primary host, the magpie (*Pica pica*). *Evolution* **53**, 269-278.

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

Mila B, Warren BH, Heeb P, Thebaud C (2010) The geographic scale of diversification on islands: genetic and morphological divergence at a very small spatial scale in the Mascarene grey white-eye (Aves: *Zosterops borbonicus*). *Bmc Evolutionary Biology* **10**.

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

Moore SM, Borer ET (2012) The influence of host diversity and composition on epidemiological patterns at multiple spatial scales. *Ecology* **93**, 1095-1105.

Ostfeld RS, Glass GE, Keesing F (2005) Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution* **20**, 328-336.

Patz JA, Graczyk TK, Geller N, Vittor AY (2000) Effects of environmental change on emerging parasitic diseases. *International Journal for Parasitology* **30**, 1395-1405.

Pavlova A, Amos JN, Joseph L*, et al.* (2013) Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution* **67**, 3412-3428.

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

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution* **43**, 258-275.

R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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

Ray JM, King RB (2006) The temporal and spatial scale of microevolution: fine-scale colour pattern variation in the Lake Erie watersnake, *Nerodia sipedon insularum*. *Evolutionary Ecology Research* **8**, 915-928.

Raymond M, Rousset F (1995) GENEPOP (Version-1.2) - Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Richardson DS, Jury FL, Blaakmeer K, Komdeur J, Burke T (2001) Parentage assignment and extra-group paternity in a cooperative breeder: the Seychelles warbler (*Acrocephalus sechellensis*). *Molecular Ecology* **10**, 2263-2273.

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-Biological Sciences* **272**, 759-767.

Richardson JL, Urban MC (2013) Strong selection barriers explain microgeographic adaptation in wild salamander populations. *Evolution* **67**, 1729-1740.

Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution* **29**, 165-176.

Roitt IM, Brostoff J, Male DK (2001) *Immunology* Mosby, Philadelphia.

Rousset F (2008) GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Schad J, Dechmann DKN, Voigt CC, Sommer S (2012) Evidence for the 'Good Genes' Model: Association of MHC Class II DRB Alleles with Ectoparasitism and Reproductive State in the Neotropical Lesser Bulldog Bat, *Noctilio albiventris*. *PloS one* **7**, e37101.

Soto-Centeno JA, Barrow LN, Allen JM, Reed DL (2013) Reevaluation of a classic phylogeographic barrier: new techniques reveal the influence of microgeographic climate variation on population divergence. *Ecology and Evolution* **3**, 1603-1613.

Spurgin LG, Illera JC, Padilla DP, Richardson DS (2012) Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia* **168**, 691-701.

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

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

Svensson EI, Sinervo B (2004) Spatial scale and temporal component of selection in side-blotched lizards. *American Naturalist* **163**, 726-734.

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

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Tonteri A, Vasemagi A, Lumme J, Primmer CR (2010) Beyond MHC: signals of elevated selection pressure on Atlantic salmon (*Salmo salar*) immune-relevant loci. *Molecular Ecology* **19**, 1273-1282.

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

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

Vincent B, Dionne M, Kent MP, Lien S, Bernatchez L (2013) Landscape genomics in Atlantic salmon (*Salmo salar*): searching for gene-environment interactions driving local adaptation. *Evolution* **67**, 3469-3487.

Wagner HH, Fortin M-J (2013) A conceptual framework for the spatial analysis of landscape genetic data. *Conservation Genetics* **14**, 253-261.

Wagner HH, Fortin MJ (2005) Spatial analysis of landscapes: Concepts and statistics. *Ecology* **86**, 1975-1987.

Wakelin D (1996) *Immunity to Parasites: How Parasitic Infections are Controlled*, Second edn. Cambridge University Press, Cambridge.

Wakelin D, Apanius V (1997) *Immune defence: genetic control* Oxford University Press, Oxford.

Wang J (2011) COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources* **11**, 141-145.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution* **38**, 1358-1370.

Westerdahl H, Stjernman M, Raberg L, Lannefors M, Nilsson J-A (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, Waldenstrom J, Hansson B*, et al.* (2005) Associations between malaria and MHC genes in a migratory songbird. *Proceedings of the Royal Society B-Biological Sciences* **272**, 1511-1518.

Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP (2006) Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology* **75**, 1182-1189.

Wood MJ, Cosgrove CL, Wilkin TA*, et al.* (2007) Within-population variation in prevalence and lineage distribution of avian malaria in blue tits, *Cyanistes caeruleus*. *Molecular Ecology* **16**, 3263-3273.

Wutzler R, Foerster K, Kempenaers B (2012) MHC class I variation in a natural blue tit population (*Cyanistes caeruleus*). *Genetica* **140**, 349-364.

Xu T-j, Chen S-l, Zhang Y-x (2010) MHC class II alpha gene polymorphism and its association with resistance/susceptibility to *Vibrio anguillarum* in Japanese flounder (*Paralichthys olivaceus*). *Developmental and Comparative Immunology* **34**, 1042-1050.

Zhang M, He H (2013) Parasite-mediated selection of major histocompatibility complex variability in wild brandt's voles (*Lasiopodomys brandtii*) from Inner Mongolia, China. *Bmc Evolutionary Biology* **13**.

Supplementary Table S5.1 Summary of model selection performed on multi-predictor generalized linear models (with ΔAIC ≤ 2 when compared to the best fit model) of MHC class I alleles of Berthelot's pipits (*Anthus berthelotii*) in Tenerife as response variables. Only alleles with the best model $R^2 \leq 0.17$ are shown. Models for ANBE38 and ANBE48 can be found in Table 5.2. Significant coefficients are underlined, level of significance is designated by asterisks (* <0.05, ** <0.01, ***<0.001), and relative importance in explaining variation in the presence of each allele after model averaging is shown in brackets.

	ANBE1	ANBE2	ANBE4	ANBE6	ANBE8	ANBE9	ANBE10	ANBE11
Poultry	-0.79* (1.00)	-0.11 (0.03)	-0.27 (0.08)	0.24 (0.04)	-1.08* (1.00)	-0.11 (0.02)	-1.44* (0.94)	-0.67 (0.70)
Water	0.12 (0.07)	0.01 (0.03)	0.51 (0.91)	0.33 (0.06)	-0.73* (1.00)	0.27 (0.14)	-0.47 (0.20)	0.44 (0.48)
Slope	0.07* (1.00)	-0.03 (0.07)	-0.06 (0.91)	0.06 (0.56)	0.10* (1.00)	0.047 (0.26)	0.05 (0.18)	-0.05 (0.38)
Temperature	0.01 (0.06)	0.05 (0.04)	-0.06 (0.22)	-0.09 (0.48)	-0.01 (0.06)	-0.03 (0.08)	0.03 (0.05)	-0.05 (0.09)
Farm	-0.09 (0.07)	-0.23 (0.03)	-0.45 (0.14)	-0.05 (0.05)	1.21* (1.00)	0.19 (0.07)	0.94 (0.46)	0.45 (0.05)
Urbanization	0.01 (0.06)	0.58 (0.19)	0.50 (0.63)	-0.32 (0.10)	0.32 (0.09)	-0.04 (0.02)	0.48 (0.19)	0.16 (0.02)
Pipit density	-0.44 (0.16)	-0.19 (0.03)	0.08 (0.03)	0.41 (0.03)	0.35 (0.07)	0.32 (0.06)	0.03 (0.03)	0.57 (0.04)
PCNM3	2.32×10^{-6} (0.06)	3.66×10^{-5} (0.03)	1.59×10^{-5} (0.04)	4.09×10^{-5} (0.09)	-4.97×10^{-6} (0.06)	7.38×10^{-5} (0.44)	-1.17×10^{-5} (0.03)	1.95×10^{-5} (0.02)
PCNM4	8.81×10^{-5}* (1.00)	-3.42×10^{-5} (0.03)	9.69×10^{-5} *(1.00)	-7.67×10^{-5} (0.25)	3.32×10^{-5} (0.08)	-5.87×10^{-5} (0.20)	1.16×10^{-4} (0.91)	-5.15×10^{-5} (0.19)
PCNM9	2.34×10^{-5} (0.07)	5.65×10^{-5} (0.03)	2.67×10^{-5} (0.08)	-4.01×10^{-5} (0.05)	9.75×10^{-5} (0.82)	-1.19×10^{-5} (0.02)	-7.98×10^{-5} (0.19)	6.57×10^{-5} (0.29)
PCNM13	-1.59×10^{-5} (0.07)	-1.65×10^{-5} (0.03)	-1.88×10^{-4}** (1.00)	-6.99×10^{-5} (0.12)	1.30×10^{-5} (0.06)	-3.04×10^{-5} (0.06)	1.03×10^{-4} (0.19)	2.72×10^{-5} (0.02)
PCNM23	-2.37×10^{-5} (0.07)	1.99×10^{-4} (0.69)	-2.02×10^{-5} (0.03)	1.51×10^{-4} (0.57)	-3.96×10^{-5} (0.07)	-3.81×10^{-6} (0.02)	1.69×10^{-5} (0.03)	-8.31×10^{-5} (0.08)
PCNM24	8.52×10^{-5} (0.20)	1.93×10^{-4} (0.22)	-1.58×10^{-4} (1.00)	-2.08×10^{-4} (0.84)	1.77×10^{-4} (0.85)	-1.56×10^{-4} (0.53)	-1.49×10^{-4} (0.18)	-1.22×10^{-4} (0.45)
PCNM87	-5.24×10^{-4}** (1.00)	4.71×10^{-4} (0.25)	-1.88×10^{-4} (0.15)	1.05×10^{-5} (0.01)	-1.66×10^{-4} (0.08)	-6.98×10^{-4}** (1.00)	-4.14×10^{-5} (0.02)	1.83×10^{-4} (0.14)
AIC Null[1]	419.80	97.00	346.70	180.50	255.00	189.50	185.20	272.20
AIC best[2]	401.98	96.77	329.15	178.54	247.35	184.45	182.21	272.00
Δ AIC	17.82	0.23	17.55	1.96	7.65	5.05	2.99	0.20
R^2 best[3]	0.11	0.03	0.13	0.06	0.11	0.06	0.05	0.02

[1]AIC of the model with only the intercept
[2]AIC of the model with the lowest AIC compared to all other possible models
[3]Explanatory power of the model with the lowest AIC

Supplementary Table S5.1 (Cont.)

	ANBE13	ANBE16	ANBE28	ANBE43	ANBE44	ANBE45	ANBE46	ANBE47	ANBE49
Poultry	1.57 (0.72)	-0.07 (0.02)	-0.72 (0.04)	-0.30 (0.16)	0.58 (0.47)	0.56 (0.12)	0.61 (0.17)	-0.11 (0.03)	-0.04 (0.03)
Water	-0.04 (0.04)	-0.53 (0.50)	-0.25 (0.03)	-0.37 (0.57)	-0.10 (0.02)	0.96* (0.82)	0.35 (0.10)	0.36 (0.26)	0.18 (0.07)
Slope	0.04 (0.05)	0.08* (1.00)	-0.25 (0.25)	0.02 (0.1)	-0.05 (0.31)	-0.04 (0.04)	3.33 X 10-4 (0.01)	0.03 (0.15)	0.03 (0.04)
Temperature	-0.06 (0.09)	-0.11 (0.72)	0.04 (0.03)	0.01 (0.03)	-0.05 (0.16)	0.07 (0.08)	-0.06 (0.10)	-0.08 (0.77)	0.08 (0.14)
Farm	0.63 (0.10)	0.57 (0.13)	-1.16 (0.11)	0.02 (0.02)	0.34 (0.08)	-0.40 (0.04)	0.76 (0.31)	-0.35 (0.10)	0.95 (0.69)
Urbanization	-0.45 (0.06)	-0.49 (0.30)	-0.84 (0.11)	0.48 (0.68)	-0.19 (0.03)	-1.25* (1.00)	0.44 (0.14)	-0.09 (0.03)	0.51 (0.14)
Pipit density	-0.15 (0.04)	-0.77 (0.06)	0.34 (0.03)	-0.18 (0.03)	-0.83 (0.19)	2.85* (1.00)	-1.32 (0.29)	-0.80 (0.27)	0.24 (0.03)
PCNM3	-7.43×10^{-5} (0.06)	7.31×10^{-6} (0.02)	3.21×10^{-5} (0.03)	-1.22×10^{-4}*** (1.00)	-3.16×10^{-5} (0.08)	-8.01×10^{-6} (0.04)	5.06×10^{-5} (0.12)	9.92×10^{-7} (0.03)	-5.57 X 10-5 (0.12)
PCNM4	-2.45×10^{-4}** (1.00)	4.82×10^{-5} (0.14)	-1.69×10^{-4} (0.14)	-3.19×10^{-5} (0.15)	-3.50×10^{-5} (0.05)	-8.00×10^{-5} (0.16)	5.12×10^{-5} (0.08)	3.54×10^{-5} (0.15)	-5.28 X 10-5 (0.08)
PCNM9	-1.88×10^{-4} (0.94)	-6.00×10^{-5} (0.11)	3.32×10^{-5} (0.03)	-2.86×10^{-5} (0.06)	-1.26×10^{-4}* (1.00)	1.99×10^{-4}* (1.00)	-3.49×10^{-5} (0.03)	-1.65×10^{-4}** (1.00)	-2.47 X 10-5 (0.03)
PCNM13	-1.82×10^{-4} (0.31)	-6.15×10^{-5} (0.03)	4.52×10^{-5} (0.03)	3.37×10^{-5} (0.06)	-9.01×10^{-5} (0.33)	4.64×10^{-5} (0.04)	-1.18×10^{-4} (0.39)	-2.09×10^{-5} (0.03)	1.76 X 10-5 (0.03)
PCNM23	1.77×10^{-5} (0.04)	2.34×10^{-4}** (1.00)	-5.14×10^{-6} (0.03)	-1.12×10^{-4} (0.79)	1.80×10^{-4}* (1.00)	-2.03×10^{-4} (0.61)	4.87×10^{-6} (0.01)	4.52×10^{-5} (0.04)	6.37 X 10-6 (0.03)
PCNM24	1.02×10^{-5} (0.04)	1.83×10^{-4} (0.77)	4.90×10^{-4} (0.28)	9.58×10^{-5} (0.47)	-5.74×10^{-6} (0.02)	8.70×10^{-5} (0.04)	-2.08×10^{-4} (0.75)	2.30×10^{-4}** (1.00)	1.97 X 10-5 (0.03)
PCNM87	-2.86×10^{-4} (0.10)	1.65×10^{-4} (0.05)	4.91×10^{-4} (0.11)	4.32×10^{-4}** (1.00)	-6.54×10^{-5} (0.02)	-4.06×10^{-4} (0.21)	2.30×10^{-5} (0.01)	2.68×10^{-4} (0.81)	2.54 X 10-4 (0.17)

[1]AIC of the model with only the intercept
[2]AIC of the model with the lowest AIC compared to all other possible models
[3]Explanatory power of the model with the lowest AIC

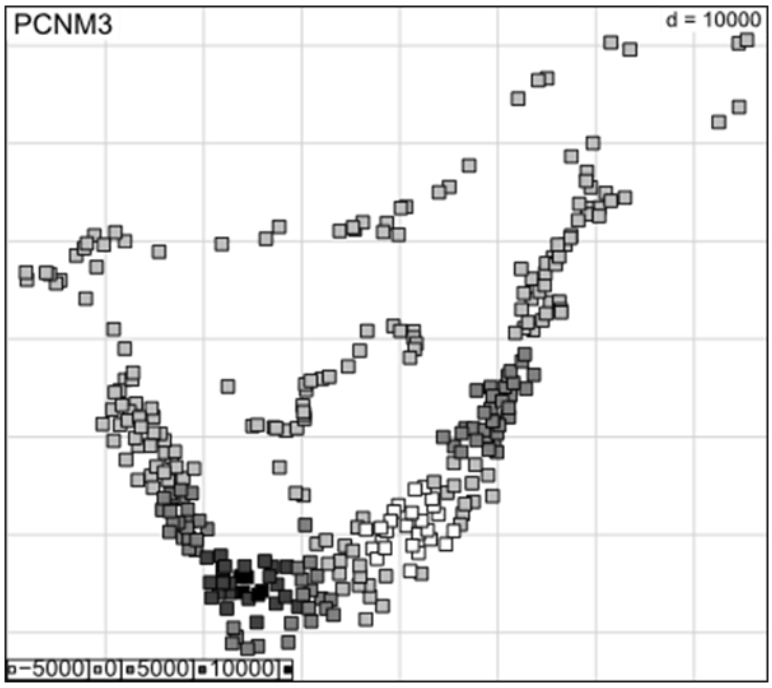
Supplementary Table S5.2 Summary of generalized linear models predicting presence of MHC class I aleles in Berthelot's pipits (*Anthus berthelotii*) in Tenerife, fitting single environmental predictor variables. AIC and Nagelkerke R^2 are shown. AIC values that result in a $\Delta AIC \geq 2$ compared to the null model are underlined. Significance of predictors is denoted by an asterisk: * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$.

		ANBE1	ANBE2	ANBE4	ANBE6	ANBE8	ANBE9	ANBE10	ANBE11	ANBE13	ANBE16
Null	AIC	419.80	97.00	346.70	180.50	255.00	189.50	185.20	272.20	122.00	228.40
Poultry	Coefficient	- 0.434	- 0.314	0.463	0.447	- 0.807*	0.033	- 0.922	- 0.474	0.437	0.379
	R^2	0.012	0.002	0.010	0.007	0.026	3.79×10^{-5}	0.027	0.010	0.005	0.005
	AIC	419.10	98.80	346.30	181.60	252.50	191.80	183.40	272.40	123.50	229.50
Water	Coefficient	- 0.071	0.036	0.544*	0.318	- 0.419	0.379	- 0.430	0.256	0.218	- 0.054
	R^2	0.001	6.10×10^{-5}	0.030	0.006	0.014	0.009	0.012	0.005	0.002	2.00 X 10-4
	AIC	421.60	99.00	342.50	181.50	254.60	190.50	185.60	273.30	123.80	230.40
Slope	Coefficient	0.058*	- 0.030	- 0.029	0.071	0.056	0.049	0.022	-0.051	0.030	0.077*
	R^2	0.021	0.003	0.005	0.025	0.011	0.011	0.002	0.009	0.003	0.031
	AIC	416.90	98.70	347.70	179.00	255.10	190.30	187.00	272.50	123.70	225.30
Temperature	Coefficient	0.018	0.040	- 0.087	- 0.068	- 0.007	- 0.045	0.028	- 0.024	0.021	- 0.074
	R^2	1.00 X 10-3	0.002	0.020	0.012	1.00 X 10-4	0.005	0.002	0.001	0.001	0.016
	AIC	421.50	98.80	345.00	180.80	256.90	191.10	186.90	274.00	123.90	227.80
Farm	Coefficient	- 0.544	- 0.263	0.214	0.186	0.152	0.372	0.087	0.146	0.562	0.212
	R^2	0.015	0.001	2.00 X 10-3	0.001	0.001	0.004	2.00 X 10-4	0.001	0.008	0.001
	AIC	418.30	98.90	348.30	182.30	256.80	191.30	187.20	274.10	123.30	230.20
Urbanization	Coefficient	- 0.128	0.530	0.532*	- 0.078	0.091	0.185	0.244	0.130	- 0.180	- 0.236
	R^2	0.001	0.010	0.002	3.00 X 10-4	5.00 X 10-4	0.001	0.003	0.001	0.001	0.003
	AIC	421.50	98.20	344.70	182.40	256.90	191.60	186.80	274.10	123.90	230.00
Pipit density	Coefficient	- 0.397	- 0.100	- 0.796	- 0.449	0.547	0.166	0.030	0.156	1.194	- 1.105
	R^2	0.003	6.90×10^{-5}	0.010	0.002	0.003	2.00 X 10-4	8.37 X 10-6	3.00 X 10-4	0.01	0.012
	AIC	421.20	99.00	346.90	182.20	256.40	191.80	187.20	274.20	123.00	228.40
PCNM3	Coefficient	1.03×10^{-5}	3.56×10^{-5}	2.24×10^{-5}	3.21×10^{-5}	2.15 X 10-6	7.09×10^{-5}	5.05 X 10-6	2.46×10^{-5}	-2.36×10^{-5}	- 2.65 X 10-6
	R^2	5.00 X 10-4	0.002	0.002	0.003	1.40×10^{-5}	0.014	6.40×10^{-5}	0.002	0.001	2.02×10^{-5}
	AIC	421.70	98.80	348.30	182.10	256.90	189.90	187.20	273.90	123.90	230.40
PCNM4	Coefficient	6.08 X 10-5	$- 3.52 \times 10^{-5}$	8.75×10^{-5}*	$- 5.23 \times 10^{-5}$	$- 1.22 \times 10^{-5}$	$- 5.95 \times 10^{-5}$	5.95×10^{-5}	$- 5.54 \times 10^{-5}$	-2.57 X 10-4**	5.88×10^{-5}
	R^2	0.015	0.002	0.020	0.006	4.00 X 10-4	0.008	0.008	0.009	0.107	0.009
	AIC	418.30	98.80	343.40	181.60	256.90	190.70	186.10	272.70	113.20	229.00
PCNM9	Coefficient	2.76×10^{-5}	0.04	2.15×10^{-5}	$- 1.74 \times 10^{-5}$	9.49×10^{-5}	$- 1.20 \times 10^{-5}$	$- 6.69 \times 10^{-5}$	6.01×10^{-5}	-2.12 X 10-4*	- 5.50 X 10-6
	R^2	0.002	0.002	0.001	4.57 X 10-4	0.017	2.23 X 10-4	0.007	0.007	0.055	5.19×10^{-5}
	AIC	421.30	98.78	348.50	182.40	254.00	191.80	186.30	273.01	118.50	230.40
PCNM13	Coefficient	1.02×10^{-5}	5.21×10^{-5}	- 1.71 X 10-4**	$- 6.79 \times 10^{-5}$	7.22×10^{-5}	$- 1.17 \times 10^{-5}$	1.46 X 10-4*	3.69×10^{-5}	-1.34 X 10-4	- 1.05 X 10-4
	R^2	0.002	0.002	0.040	0.006	0.008	1.00 X 10-4	0.029	0.002	0.019	0.016
	AIC	421.70	98.80	339.60	181.70	255.60	191.80	183.20	273.90	122.00	227.90
PCNM23	Coefficient	-1.71×10^{-5}	1.93 X 10-4	2.50 X 10-6	1.49 X 10-4	$- 7.83 \times 10^{-5}$	- 3.63 X 10-6	1.71 X 10-6	$- 6.09 \times 10^{-5}$	5.09×10^{-5}	2.36 X 10-4**
	R^2	3.45 X 10-4	0.027	6.30 X 10-6	0.017	0.005	8.70 X 10-6	1.90 X 10-6	0.003	0.001	0.048
	AIC	421.70	96.77	348.70	180.10	256.10	191.80	187.20	273.70	123.90	222.50
PCNM24	Coefficient	6.88×10^{-5}	1.80 X 10-4	- 1.29 X 10-4	- 2.00 X 10-4	1.29 X 10-4	- 1.42 X 10-4	$- 8.39 \times 10^{-5}$	$- 8.58 \times 10^{-5}$	6.39 X 10-6	1.14 X 10-4
	R^2	0.005	0.010	0.010	0.025	0.012	0.012	0.004	0.006	1.89×10^{-5}	0.008
	AIC	420.70	97.70	345.80	179.10	254.80	190.10	186.60	273.20	124.00	229.10
PCNM87	Coefficient	- 5.00 X 10-4**	4.20 X 10-4	1.71 X 10-4	- 3.60 X 10-6	- 1.43 X 10-4	- 6.95 X 10-4**	$- 2.66 \times 10^{-5}$	1.80×10^{-5}	-3.04 X 10-4	1.58 X 10-4
	R^2	0.04	0.010	0.005	1.45 X 10-6	0.003	0.051	8.13×10^{-5}	0.005	0.008	0.003
	AIC	411.40	97.80	347.70	182.50	256.40	184.50	187.20	273.30	123.20	229.90

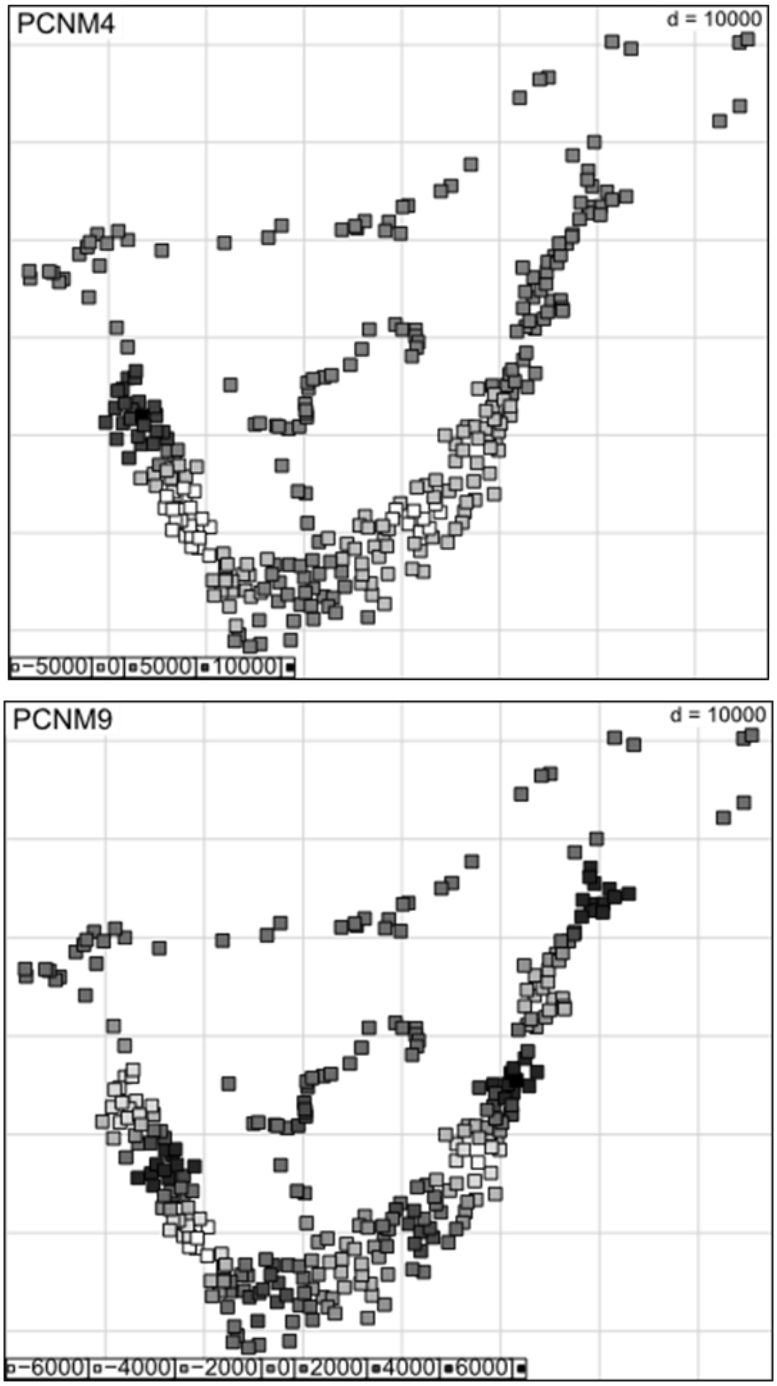
Supplementary Table S5.2 (Cont.)

		ANBE28	ANBE38	ANBE43	ANBE44	ANBE45	ANBE46	ANBE47	ANBE48	ANBE49
Null	AIC	35.80	139.30	407.80	268.90	139.30	155.50	363.60	90.30	165.70
Poultry	Coefficient	- 0.719	- 1.221*	-0.213	0.737*	0.339	0.749	-0.059	2.954**	0.332
	R^2	0.010	0.051	0.003	0.022	0.003	0.016	2.00×10^{-4}	0.154	0.003
	AIC	37.40	135.70	409.20	266.90	140.90	155.50	365.50	80.20	167.30
Water	Coefficient	- 0.250	0.138	-0.246	0.069	- 0.046	0.403	0.324	0.339	0.331
	R^2	0.002	0.001	0.007	4.00×10^{-4}	1.00×10^{-4}	0.009	0.011	0.005	0.007
	AIC	37.70	141.20	408.20	270.80	141.30	156.30	363.30	91.90	166.90
Slope	Coefficient	- 0.240	0.067	0.008	-0.035	- 0.023	0.008	0.007	0.051	0.016
	R^2	0.047	0.019	5.00×10^{-4}	0.005	0.001	2.00×10^{-4}	3.00×10^{-4}	0.009	0.001
	AIC	36.30	139.10	409.70	270.00	141.10	157.40	365.50	91.60	167.60
Temperature	Coefficient	0.036	- 0.004	4.10×10^{-4}	- 0.033	0.106	- 0.061	-0.049	- 0.100	- 0.002
	R^2	0.001	2.38×10^{-5}	6.03×10^{-7}	0.003	0.012	0.009	0.009	0.021	8.74×10^{-6}
	AIC	37.70	141.30	409.80	270.30	139.90	156.40	363.70	90.20	167.70
Farm	Coefficient	- 1.042	- 0.316	-0.072	0.255	- 0.777	0.737	-0.027	1.413	0.843
	R^2	0.016	0.002	3.00×10^{-4}	0.002	0.014	0.014	3.51×10^{-5}	0.045	0.020
	AIC	37.30	141.00	409.70	270.50	139.70	155.70	365.60	88.80	165.20
Urbanization	Coefficient	- 0.758	- 0.829	0.243	- 0.045	- 1.174**	0.515	0.08	0.855	0.582
	R^2	0.015	0.028	0.005	1.00×10^{-4}	0.058	0.01	4.00×10^{-4}	0.022	0.013
	AIC	37.30	138.10	408.70	270.90	134.80	156.20	365.50	90.60	166.00
Pipit density	Coefficient	0.341	- 0.106	-0.012	- 0.656	2.154	- 1.555	-0.443	- 0.716	- 0.317
	R^2	5.00×10^{-4}	8.88×10^{-5}	2.27×10^{-6}	0.005	0.035	0.019	0.003	0.003	0.001
	AIC	37.80	141.30	409.80	270.00	137.31	155.10	365.00	92.00	167.60
PCNM3	Coefficient	3.21×10^{-5}	$- 3.73 \times 10^{-5}$	-1.22×10^{-4}***	$- 3.02 \times 10^{-5}$	1.57×10^{-5}	3.790	-8.16×10^{-6}	1.00×10^{-4}	$- 4.60 \times 10^{-5}$
	R^2	0.001	0.003	0.058	0.003	5.00×10^{-4}	0.003	2.00×10^{-4}	0.020	0.005
	AIC	37.70	140.90	396.40	270.40	141.20	157.05	365.50	90.70	167.10
PCNM4	Coefficient	$- 1.70 \times 10^{-4}$	1.18×10^{-4}	-3.75×10^{-5}	$- 9.93 \times 10^{-6}$	$- 7.49 \times 10^{-5}$	4.33×10^{-5}	4.19×10^{-5}	7.03×10^{-5}	$- 5.24 \times 10^{-5}$
	R^2	0.032	0.029	0.006	3.00×10^{-4}	0.011	0.004	0.006	0.008	0.006
	AIC	36.70	138.00	408.50	270.80	140.10	157.00	364.20	91.60	167.00
PCNM9	Coefficient	3.32×10^{-5}	4.99×10^{-5}	-2.87×10^{-5}	$- 1.11 \times 10^{-4}$*	1.94×10^{-4}*	$- 2.02 \times 10^{-5}$	-1.28×10^{-4}**	8.66×10^{-5}	$- 9.00 \times 10^{-6}$
	R^2	0.001	0.003	0.002	0.023	0.044	6.00×10^{-4}	0.039	0.008	1.16×10^{-4}
	AIC	37.70	140.90	409.30	266.60	136.40	157.40	357.10	91.70	167.70
PCNM13	Coefficient	4.52×10^{-5}	2.00×10^{-4}*	3.06×10^{-5}	$- 1.05 \times 10^{-4}$	1.18×10^{-5}	$- 1.03 \times 10^{-4}$	2.21×10^{-8}	$- 2.68 \times 10^{-4}$**	3.64×10^{-5}
	R^2	0.001	0.044	0.002	0.017	1.00×10^{-4}	0.012	9.04×10^{-10}	0.080	0.001
	AIC	37.70	136.30	409.40	267.80	141.30	156.00	365.60	85.70	167.60
PCNM23	Coefficient	$- 5.13 \times 10^{-6}$	$- 1.04 \times 10^{-5}$	-1.12×10^{-4}	2.00×10^{-4}*	$- 1.50 \times 10^{-4}$	$- 1.12 \times 10^{-7}$	4.95×10^{-5}	$- 3.20 \times 10^{-4}$**	6.42×10^{-6}
	R^2	8.97×10^{-6}	6.10×10^{-5}	0.013	0.035	0.015	7.37×10^{-9}	0.002	0.105	2.53×10^{-5}
	AIC	37.80	141.30	406.80	264.50	139.60	157.50	365.04	84.10	167.70
PCNM24	Coefficient	5.00×10^{-4}	$- 9.60 \times 10^{-5}$	8.77×10^{-5}	$- 2.24 \times 10^{-5}$	7.68×10^{-5}	$- 1.97 \times 10^{-4}$	2.02×10^{-4}**	$- 9.80 \times 10^{-6}$	1.05×10^{-5}
	R^2	0.059	0.005	0.008	4.00×10^{-4}	0.003	0.022	0.034	3.96×10^{-5}	5.98×10^{-5}
	AIC	35.90	140.70	408.00	270.80	140.90	154.80	358.10	92.30	167.70
PCNM87	Coefficient	4.70×10^{-4}	$- 2.21 \times 10^{-5}$	3.92×10^{-4}*	$- 7.15 \times 10^{-5}$	$- 2.60 \times 10^{-4}$	2.77×10^{-5}	2.70×10^{-4}	4.49×10^{-4}	2.60×10^{-4}
	R^2	0.015	4.77×10^{-5}	0.030	0.001	0.006	8.02×10^{-5}	0.013	0.018	0.008
	AIC	37.30	141.30	403.00	270.80	140.60	157.50	362.70	90.80	166.80

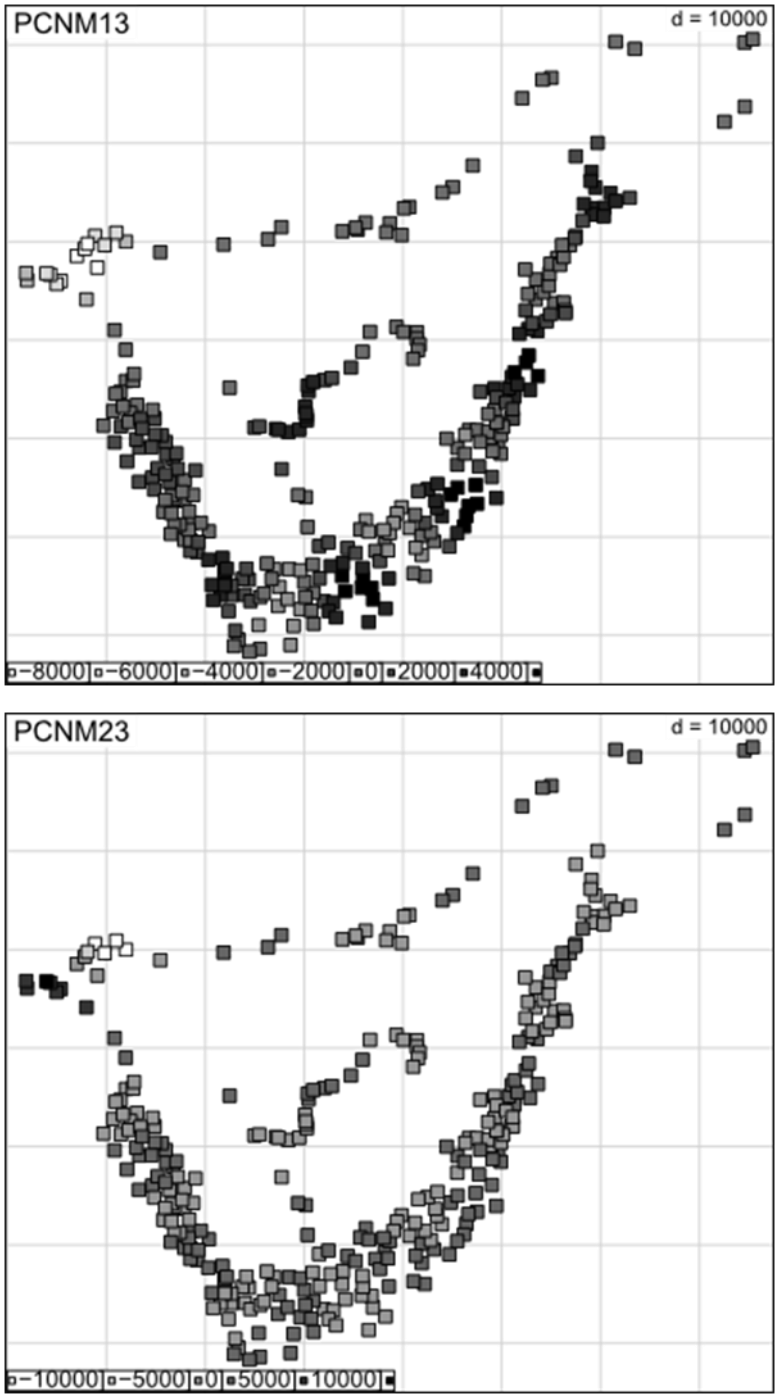
Supplementary Figure S5.1 Scores at each sampled location from the seven PCNMs significantly associated with the spatial distribution of MHC class I alleles in Berthelot's pipits (*Anthus berthelotii*) in Tenerife, as determined by forward selection within a redundancy analysis (RDA). PCNMs are eigenvectors computed from the spatial coordinate positions of samples (see methods) and are used to characterise the spatial variation in allele frequencies (genetic structure). PCNMs 3 and 4 have positive eigenvalues, hence indicating positively autocorrelated (global) spatial structures. PCNMs 9, 13, 23, 24 and 87 have negative eigenvalues indicating negatively autocorrelated (local) spatial structures. Selection of negative PCNMs may indicate neighbour dissimilarities resulting from e.g. repulsion processes such as dispersal for inbreeding avoidance, hence may show shorter-distance spatial structuring than positive PCNMs (as indicated here). Each sampled location has a score that ranges from lowest values in white to highest values in black.



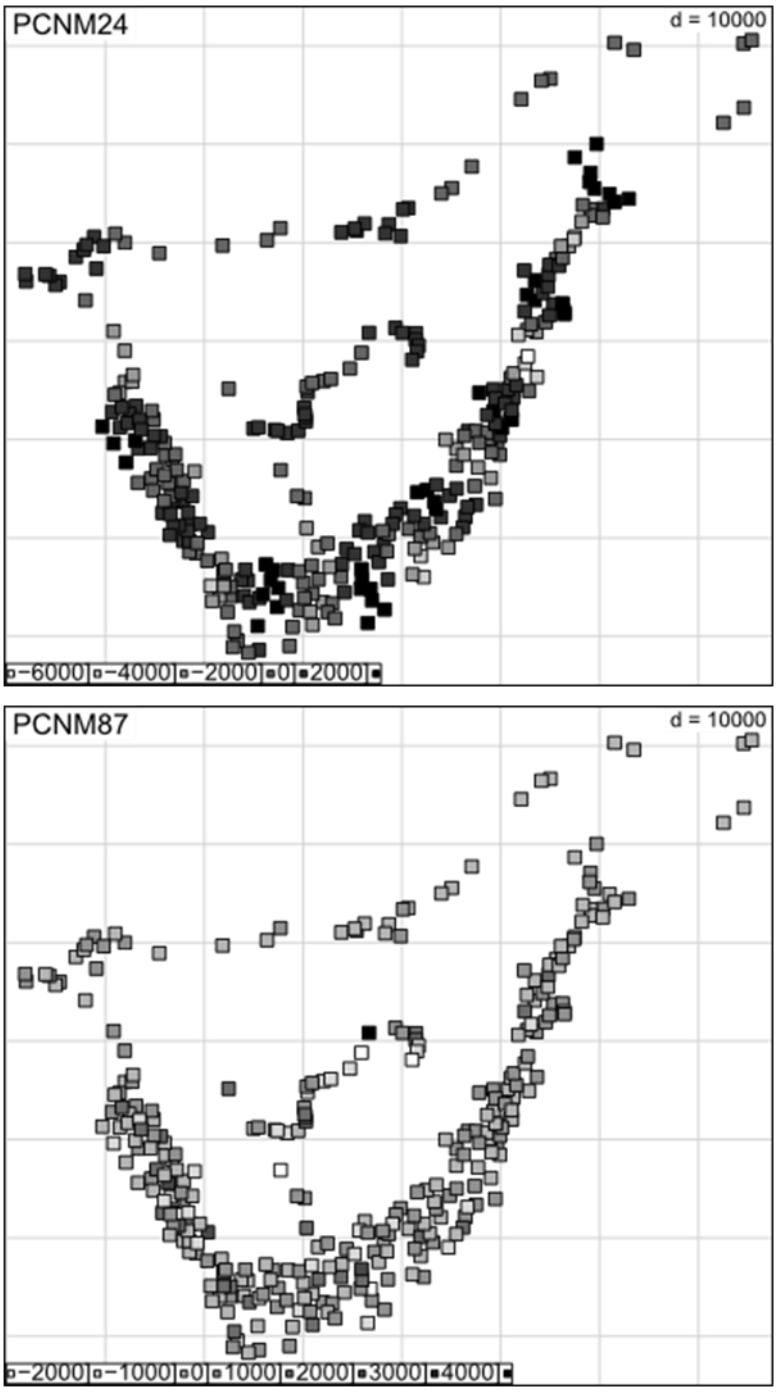
Supplementary Figure S5.1 (cont.)



Supplementary Figure S5.1 (cont.)



Supplementary Figure S5.1 (cont.)



Chapter 6

General Discussion

El Teide, raising 3700 m above sea level viewed from the plateau, photo by Karl Phillips

6.1 General discussion

In this thesis I have emphasised the importance of exploring evolutionary processes in a spatially explicit manner. This is because evolutionary change depends not only on the genetic variation available in a population, but also on the environmental characteristics to which individuals carrying such genetic variation are exposed, and how these differ across time and space. I explored how different sets of genes relevant to the immune system evolve in response to demographic and selective factors at different spatial scales in populations of Berthelot's pipit. In this final chapter, I discuss my findings collectively and propose ideas for future research.

6.1.1 Comparative evolution of MHC and TLRs

Elucidating the roles of demographic and selective factors in shaping genetic variation is of utmost importance for understanding how genetic variation evolves in populations (Lande 1976). It is especially important to understand how these evolutionary processes shape variation at functional loci, particularly in bottlenecked populations, where genetic diversity is reduced. Functional loci directly involved in individual survival, such as those of the immune system, play an important role in wild populations (Anderson & May 1979). In populations that have colonised new habitats variation at these loci may play a key role because it is likely that individuals will encounter pathogens to which they haven't been previously exposed. In such circumstances, having variation that can initiate an appropriate immune response to these novel pathogens is essential for survival and, consequently, the successful establishment of the newly founded population (O'Brien & Evermann 1988). However, the evolution of different loci depends on their specific structure and function, and genes involved in different aspects of the immune response might evolve differently. In this thesis I examined variation at two sets of loci, the Toll-Like Receptors (TLRs) involved in the innate immune response (Chapter 2) and the Major Histocompatibility Complex (MHC) involved in the acquired immune response (Chapters 4 and 5) in Berthelot's pipit, a recently bottlenecked coloniser of new populations across the north Atlantic archipelagos of Canary Islands, Selvagens and Madeira. The results of this study indicate various similarities and differences in the way that these two sets of genes I investigated have evolved in Berthelot's pipit.

The first difference I detected was in how genetic variation is generated at these two loci. In Chapter 2 I found that genetic variation at TLRs was generated mainly by point mutation with

no evidence of gene conversion. In contrast, in Chapter 4, I found that variants within the MHC are mainly generated by gene conversion, resulting in a higher recombination rate compared to the mutation rate. MHC loci have evolved by gene duplication, therefore, an individual possesses more than one MHC locus and alleles have highly similar sequences (Bach 1976). This facilitates the occurrence of gene conversion events (Ohta 1995). Although TLRs also evolved by gene duplication (Roach *et al.* 2005), they have diverged greatly to the extent of being found in separate chromosomes, and sequence similarity is not as high as between alleles at different MHC loci (Chapter 2). However, a few studies have reported that gene conversion occurs in some TLR families that have recently been duplicated (Kruithof *et al.* 2007; Cormican *et al.* 2009; Mikami *et al.* 2012). Except for TLR1LA and TLR1LB, the TLRs that I screened are members of different families and this might be the reason why I didn't find evidence of gene conversion. Why the structure of these gene families is so different, and whether this is a cause or consequence of gene conversion, is not known, but the consequences are that gene conversion events are less common between TLR loci than between MHC loci. It might be that gene conversion events are more functionally constraining at TLRs than at the MHC; thus even if gene conversion events are equally common in the two gene families, probably the ones that occur at TLRs are more often selected against than the ones occurring at the MHC. My results concur with studies that have reported gene conversion as an important source of variation at the MHC (Hogstrand & Bohme 1999; Miller & Lambert 2004a; Spurgin *et al.* 2011), and point mutation as the main source of genetic variation at TLRs (Roach *et al.* 2005; Barreiro *et al.* 2009; Alcaide & Edwards 2011). Importantly, gene conversion results in the copying across of portions of DNA from one variant to another, while point mutation results in a change of only one nucleotide. Thus, a change is more likely to be functionally significant if it occurs by gene conversion. In accordance with this, In Chapter 2 I found that in populations of Berthelot's pipits, the TLR with most polymorphic sites had seven variable sites out of 997 bp (TLR1LB), while at the MHC there were 84 variable sites out of 240 bp (Spurgin *et al.* 2011; Chapter 4). Furthermore, nucleotide diversity between MHC alleles (0.11 ± 0.01, Chapter 4) was two orders of magnitude greater than nucleotide diversity at TLRs (0.29×10^{-2} for the most variable TLR4, Chapter 2).

Once generated, genetic variation at immune loci is subject to evolutionary forces. When first dispersing across its entire range, Berthelot's pipit populations went through bottlenecks. Also, across the island populations of Berthelot's pipits, especially across archipelagos, there is very

little gene flow (Illera *et al.* 2007; Spurgin *et al.* 2014). Therefore, the main evolutionary forces promoting evolutionary change in these populations are genetic drift and selection. Genetic drift has previously been shown to be a strong evolutionary force shaping patterns of variation at immune loci in small populations (Cramer *et al.* 1988; Eimes *et al.* 2011; Ejsmond & Radwan 2011), and can outweigh the effects of selection (Biedrzycka & Radwan 2008; Bouzat 2010; Grueber *et al.* 2013). By investigating variation at neutral microsatellite markers, I was able to determine the effects of past demographic changes across both large (Spurgin *et al.* 2014; Chapter 2) and fine scales (Chapter 5). This then allowed me to assess the role of demographic processes in shaping variation at the functional TLRs and MHC loci, respectively. In Chapter 2 I found that TLR variation among populations has been shaped mainly by the bottlenecks involved in the colonisation of the Selvagens and Madeira archipelagos by pipits from the Canary Islands. TLR variation appeared to be much reduced as a result of the bottlenecks. Furthermore, genetic variation at these loci has remained low after the colonisation events, thus the main evolutionary force shaping such variation has been genetic drift. I was able to contrast this pattern with a previous study that assessed population-level variation at the MHC in Berthelot's pipit (Spurgin *et al.* 2011). This study found, as I did for TLRs, that MHC variation had been reduced during the bottlenecks but, unlike TLRs, the generation, and subsequent selection, of new variants by gene conversion has restored genetic variation at the MHC relatively quickly in an evolutionary time scale. This reflects once more the differences between the two gene families in how they respond to evolutionary forces. My findings concur with other studies that have shown a reduction in adaptive genetic variation in bottlenecked populations (Hedrick & Parker 1998; Grueber *et al.* 2012; Sutton *et al.* 2013) and strong effects of genetic drift in shaping variation post-bottlenecks (Miller & Lambert 2004b; Miller *et al.* 2010; Sutton *et al.* 2011; Grueber *et al.* 2013; but see Oliver & Piertney 2012).

Selection is also an important evolutionary force acting on immune gene variation in natural populations. If a newly generated variant does not result in a functional change in the molecule - i.e. a synonymous change that doesn't change the amino acid, or if a change does alter the amino acid composition but the resultant molecule has the same binding properties as the original molecule - then it is selectively neutral (but see Shields *et al.* 1988; Bhardwaj 2014 for evidence of selection on synonymous substitutions). On the other hand, if the genetic variant results in the molecule having a novel and advantageous function - i.e. binding a novel pathogen derived antigen - then it is selected for, and its frequency will increase in the

population (Fisher 1930; Hedrick & Thomson 1983). The mechanism of selection also depends on the function of the molecule. TLRs recognize and bind conserved molecules of pathogens, known as pathogen associated molecular patterns (PAMPs), such as lipopolysacharides, DNA and RNA fragments, and lipids (Poltorak *et al.* 1998; Lien *et al.* 1999; Takeuchi *et al.* 2002; Keestra *et al.* 2010). Molecules of the MHC, on the other hand, bind peptides that are generally variable among pathogen species/strains (Roudier *et al.* 1991). Thus, TLRs are more constrained in their function than the MHC molecules. In line with this difference in functional constrain between the two gene families I found more evidence of codons under selection at the MHC than at TLRs: in Chapter 2 I found evidence of positive selection at three codons at two TLR loci out of a total of 1,437 codons across the five loci, while in Chapter 4 I detected ten codons out of 80 with evidence of positive selection in the exon 3 of the MHC class I. My findings concur with other studies that have identified evidence of selection at specific codons both at TLRs (Alcaide & Edwards 2011; Areal *et al.* 2011; Fornůsková *et al.* 2013; Grueber *et al.* 2014) and at the MHC (Aguilar & Garza 2007; Fraser *et al.* 2011; Sutton *et al.* 2013; Scherman *et al.* 2014).

Obviously pathogens are the main underlying selective pressure acting on immune genes (reviewed in Spurgin & Richardson 2010). However, TLRs and MHC loci are probably associated with different sets of pathogens. This and the fact that different species or populations may be challenged by different sets of pathogens means that in Berthelot's pipit pathogen-mediated selection might not be of the same strength at the different gene families, or at different loci within these families, and this may differ between populations. Furthermore, selection at the MHC has been linked to other, non-pathogen related mechanisms, i.e. kin recognition and mate choice (Wedekind 1994; Brown 1998; Reusch *et al.* 2001; Penn 2002; Brouwer *et al.* 2010). Thus, selective pressures might be stronger at the MHC than at TLRs. Nevertheless, it is unknown whether TLRs may play a role in mechanisms other than the direct immune response. TLRs haven't been as extensively studied as the MHC, but with more studies on TLR molecular function and on TLR variation in wild populations, the role of selective pressures other than pathogens in shaping variation at these loci may be uncovered.

6.1.2 The importance of considering spatial scale in evolutionary processes

The observed effects of pathogens on immune gene evolution may vary depending on the spatial scale at which they are measured. Across populations, pathogen pressures might vary

because of differences in biogeographic and environmental features (reviewed in Ostfeld *et al.* 2005). For example, the islands/locations that support populations of Berthelot's pipits have different environmental characteristics, and differences have been previously found in the prevalence of pathogens such as avian malaria and pox among these populations (Spurgin *et al.* 2012). Such variation in pathogen-related selective pressures among populations might result in spatially varying selection at immune loci. Although no associations of disease with TLRs were assessed in the present study, in Chapter 2 I reported evidence of four non-synonymous alleles of TLR4 being maintained at a high frequency in Madeira – the archipelago with highest levels of both pox and malaria infection – and found only one TLR4 allele in Selvagens – the population with lowest levels of infection. This might be the result of high levels of pathogen-mediated selection acting specifically on this locus in Madeira, although more information linking specific pathogens with TLR4 are needed to clarify this. Within host populations the distribution of pathogens has been shown to vary because of fine scale environmental variation (Eisen & Wright 2001; Wood *et al.* 2007; Lachish *et al.* 2011). In line with this previous evidence, in Chapter 3 I found that within a single population of Berthelot's pipit, the distribution of avian malaria was highly dependent on environmental features that vary at small spatial scales. Factors important for malaria infection in this population were temperature of the coldest month, distance to artificial water sources and distance to poultry farms. Such fine-scale variation in pathogen distribution might be reflected in the spatial patterns of immune gene variation at the same scale. In fact, in Chapter 5 I report a significant and opposing relationship between two MHC class I alleles and distance to poultry farms. Since the association of these alleles directly with malaria was much weaker (only weakly significant for one and non significant for the other) it is likely that other pathogens linked to poultry farms might be associated with these alleles. In order to compare the microevolutionary processes at both sets of immune genes, an intra-population screening of TLRs would have to be done, as was done in Chapter 5 for the MHC.

Throughout this thesis I emphasise the importance of considering spatial scale when assessing evolutionary processes. Studies that address evolution have generally done so at large scales, because patterns are easier to infer when there are conspicuous differences among the populations being considered (Eizaguirre & Lenz 2010; Fraser *et al.* 2011; Savolainen *et al.* 2013). Uncovering the role of spatial scale in driving evolution across a wide range of scales is important (Richardson *et al.* 2014). There has been increasing interest in studying evolutionary

patterns at microgeographic scales in the last few years (Kavanagh *et al.* 2010; Willi & Hoffmann 2012; Richardson & Urban 2013), and such studies suggest that microgeographic adaptation is a widespread phenomenon. In the case of pathogen distribution, for example, when looking at finer scales I was able to detect the effect of environmental variables that would not have been picked up if only large scales had been considered (Chapter 3). Furthermore, fine scale patterns in genetic diversity have been recently described (Garroway *et al.* 2013; Richardson & Urban 2013). The patterns I described in Chapter 5, where I detected an association of two MHC alleles with an environmental variable measured at a fine scale, would not have been possible if I had focused on genetic variation at a large scale. This thesis therefore can serve as an example of the importance for studies on spatially explicit evolutionary processes to assess adaptive divergence at a different set of scales, varying from across populations to within populations.

6.1.3 Directions for future research

The research presented in this thesis provides some insight into the different spatial scales at which pathogens exert selective pressures on immune genes. However, additional research is needed in order to fully understand the evolutionary processes that operate at these loci. Screening more individuals at TLRs from each population is needed in order to confirm that the unique variants found in Madeira (Chapter 2) are not present in the Canary Islands, the source population. Furthermore, assessment of fine spatial scale variation at TLRs within Tenerife would be a valuable contribution to this study because it would allow comparison with the fine-scale patterns identified at the MHC in Chapter 5. Associations of individual TLR genotypes with specific pathogens have not been done in wild populations. I consider that such studies are greatly needed to contribute to the growing understanding of TLR evolution.

Further description of selective pressures on immune loci in Berthelot's pipit populations would also help understand the patterns described in this thesis both at TLRs and at the MHC. Screening of other pathogens like bacteria, viruses and intestinal parasites, and assessment of associations between infection and immune loci would provide further understanding on pathogen-mediated selection. Additionally, individual-level MHC screening in all Berthelot's pipit populations would contribute largely to the evidence of associations between poultry farms and two MHC alleles that I found in Chapter 5, and would probably help detect other patterns of MHC-pathogen associations, making use of the difference in pathogen fauna that

have been reported across populations (Spurgin *et al.* 2012). Doing this would also help to understand why the results I reported in Chapter 4 differ from the population-level screening of the MHC done previously. We were unable to do this for the present thesis due to time and cost restrictions.

Temporal patterns in ecological processes have long been interesting to evolutionary ecologists (Levin 1992), and temporal fluctuations of pathogenic pressures have previously been described (Montgomery & Montgomery 1989; Oesterholt *et al.* 2006; Cosgrove *et al.* 2008; Lachish *et al.* 2011). In Berthelot's pipit populations pathogen richness and prevalence have been consistent over a period of three years (Spurgin *et al.* 2012). It is likely however that such pathogens fluctuate at longer time scales, as new pathogens arrive on the islands or as climatic fluctuations that occur at longer time scales affect abundance of local pathogens. Therefore, assessment of pathogen pressures across populations across multiple years, at longer time scales, would greatly help in understanding how selective pressures fluctuate over time. Equally important would be to assess genetic diversity at both TLR and MHC across populations across time. Accomplishing this would require investment of a large amount of time and human and economic resources, but in the long run, this would give important insight into the evolutionary processes involving pathogen and immune gene evolution.

In conclusion, the populations of Berthelot's pipit provide an incredibly useful study system in which to test different macro and micro-evolutionary processes because of its presence in several isolated and recently bottlenecked populations. This thesis provided an interesting example of how spatial scales can be incorporated into evolutionary studies to address questions about these micro and macro-evolutionary processes whilst providing novel information about multiple gene families within natural populations.

6.2 References

Aguilar A, Garza JC (2007) Patterns of historical balancing selection on the salmonid major histocompatibility complex class II beta gene. *Journal of Molecular Evolution* **65**, 34-43.

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

Anderson RM, May RM (1979) Population biology of infectious-diseases. *Nature* **280**, 361-367.

Areal H, Abrantes J, Esteves PJ (2011) Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *Bmc Evolutionary Biology* **11**, 368.

Bach FH (1976) Genetics of transplantation - Major Histocompatibility Complex. *Annual Review of Genetics* **10**, 319-339.

Barreiro LB, Ben-Ali M, Quach H*, et al.* (2009) Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *Plos Genetics* **5**, e1000562.

Bhardwaj A (2014) Investigating the role of site specific synonymous variation in disease association studies. *Mitochondrion* **16**, 83-88.

Biedrzycka A, Radwan J (2008) Population fragmentation and major histocompatibility complex variation in the spotted suslik, *Spermophilus suslicus*. *Molecular Ecology* **17**, 4801-4811.

Bouzat JL (2010) Conservation genetics of population bottlenecks: the role of chance, selection, and history. *Conservation Genetics* **11**, 463-478.

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**, 3444-3455.

Brown JL (1998) The new heterozygosity theory of mate choice and the MHC. *Genetica* **104**, 215-221.

Cormican P, Lloyd AT, Downing T*, et al.* (2009) The avian Toll-Like receptor pathway-Subtle differences amidst general conformity. *Developmental and Comparative Immunology* **33**, 967-973.

Cosgrove CL, Wood MJ, Day KP, Sheldon BC (2008) Seasonal variation in *Plasmodium* prevalence in a population of blue tits *Cyanistes caeruleus*. *Journal of Animal Ecology* **77**, 540-548.

Cramer DV, Chakravarti A, Arenas O, Humprieres J, Mowery PA (1988) Genetic diversity within and between natural populations of *Rattus norvegicus*. *Journal of Heredity* **79**, 319-324.

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

Eisen RJ, Wright NM (2001) Landscape features associated with infection by a malaria parasite (*Plasmodium mexicanum*) and the importance of multiple scale studies. *Parasitology* **122**, 507-513.

Eizaguirre C, Lenz TL (2010) Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *Journal of Fish Biology* **77**, 2023-2047.

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

Fisher RA (1930) *The genetical theory of natural selection* Clarendon Press, Oxford.

Fornůsková A, Vinkler M, Pagès M*, et al.* (2013) Contrasted evolutionary histories of two toll-like receptors (Tlr4 and Tlr7) in wild rodents (murinae). *Bmc Evolutionary Biology* **13**, 194.

Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB (2011) Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity* **106**, 404-420.

Garroway CJ, Radersma R, Sepil I*, et al.* (2013) Fine-scale genetic structure in a wild bird population: the role of limited dispersal and environmentally based selection as causal factors. *Evolution* **67**, 3488-3500.

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

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

Grueber CE, Wallis GP, King TM, Jamieson IG (2012) Variation at innate immunity toll-like receptor genes in a bottlenecked population of a New Zealand robin. *PloS one* **7**, e45011.

Hedrick PW, Parker KM (1998) MHC variation in the endangered Gila topminnow. *Evolution* **52**, 194-199.

Hedrick PW, Thomson G (1983) Evidence for balancing selection at HLA. *Genetics* **104**, 449-456.

Hogstrand K, Bohme J (1999) Gene conversion can create new MHC alleles. *Immunological Reviews* **167**, 305-317.

Illera JC, Emerson BC, Richardson DS (2007) Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology* **16**, 4599-4612.

Kavanagh KD, Haugen TO, Gregersen F, Jernvall J, Vollestad LA (2010) Contemporary temperature-driven divergence in a Nordic freshwater fish under conditions commonly thought to hinder adaptation. *Bmc Evolutionary Biology* **10**, 350.

Keestra AM, de Zoete MR, Bouwman LI, van Putten JPM (2010) Chicken TLR21 is an innate CpG DNA receptor distinct from mammalian TLR9. *The Journal of Immunology* **185**, 460-467.

Kruithof EKO, Satta N, Liu JW, Dunoyer-Geindre S, Fish RJ (2007) Gene conversion limits divergence of mammalian TLR1 and TLR6. *Bmc Evolutionary Biology* **7**, 148.

Lachish S, Knowles SCL, Alves R, Wood MJ, Sheldon BC (2011) Infection dynamics of endemic malaria in a wild bird population: parasite species-dependent drivers of spatial and temporal variation in transmission rates. *Journal of Animal Ecology* **80**, 1207-1216.

Lande R (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314-334.

Levin SA (1992) The problem of pattern and scale in ecology. *Ecology* **73**, 1943-1967.

Lien E, Sellati TJ, Yoshimura A*, et al.* (1999) Toll-like receptor 2 functions as a pattern recognition receptor for diverse bacterial products. *Journal of Biological Chemistry* **274**, 33419-33425.

Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N (2012) Molecular evolution of vertebrate Toll-like receptors: Evolutionary rate difference between their leucine-rich repeats and their TIR domains. *Gene* **503**, 235-243.

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

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

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

Montgomery SSJ, Montgomery WI (1989) Spatial and temporal variation in the infracommunity structure of helminths of *Apodemus sylvaticus* (rodentia, muridae). *Parasitology* **98**, 145-150.

O'Brien SJ, Evermann JF (1988) Interactive influence of infectious disease and genetic diversity in natural populations. *Trends in Ecology & Evolution* **3**, 254-259.

Oesterholt MJAM, Bousema JT, Mwerinde OK*, et al.* (2006) Spatial and temporal variation in malaria transmission in a low endemicity area in northern Tanzania. *Malaria Journal* **5**, 98.

Ohta T (1995) Gene conversion vs. point mutation in generating variability at the antigen recognition site of major histocompatibility complex loci. *Journal of Molecular Evolution* **41**, 115-119.

Ostfeld RS, Glass GE, Keesing F (2005) Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution* **20**, 328-336.

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

Poltorak A, Smirnova I, He XL*, et al.* (1998) Genetic and physical mapping of the LPS locus: Identification of the Toll-4 receptor as a candidate gene in the critical region. *Blood Cells Molecules and Diseases* **24**, 340-355.

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

Richardson JL, Urban MC (2013) Strong selection barriers explain microgeographic adaptation in wild salamander populations. *Evolution* **67**, 1729-1740.

Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution* **29**, 165-176.

Roach JC, Glusman G, Rowen L*, et al.* (2005) The evolution of vertebrate Toll-like receptors. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9577-9582.

Roudier J, Albani S, Carson DA (1991) Immune response to peptides from the third hypervariable region of the beta chain of MHC class II molecules. Implications for the immune response to foreign antigens. *Seminars in cancer biology* **2**, 283-285.

Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**, 807-820.

Scherman K, Raberg L, Westerdahl H (2014) Positive Selection on MHC Class II DRB and DQB Genes in the Bank Vole (*Myodes glareolus*). *Journal of Molecular Evolution* **78**, 293-305.

Shields DC, Sharp PM, Higgins DG, Wright F (1988) Silent sites in *Drosophila* genes are not neutral - Evidence of selection among synonymous codons. *Molecular Biology and Evolution* **5**, 704-716.

Spurgin LG, Illera JC, Jorgensen TH, Dawson DA, Richardson DS (2014) Genetic and phenotypic divergence in an island bird: isolation by distance, by colonisation or by adaptation? *Molecular Ecology* **23**, 1028-1039.

Spurgin LG, Illera JC, Padilla DP, Richardson DS (2012) Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia* **168**, 691-701.

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

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

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

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

Takeuchi O, Sato S, Horiuchi T*, et al.* (2002) Cutting edge: Role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *Journal of Immunology* **169**, 10-14.

Wedekind C (1994) Mate choice and maternal selection for specific parasite resistances before, during and after fertilization. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **346**, 303-311.

Willi Y, Hoffmann AA (2012) Microgeographic adaptation linked to forest fragmentation and habitat quality in the tropical fruit fly *Drosophila birchii*. *Oikos* **121**, 1627-1637.

Wood MJ, Cosgrove CL, Wilkin TA*, et al.* (2007) Within-population variation in prevalence and lineage distribution of avian malaria in blue tits, *Cyanistes caeruleus*. *Molecular Ecology* **16**, 3263-3273.