The role of pathogens in shaping genetic variation within and among populations in an island bird

Photo by Philip Lamb

Claire Armstrong

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

School of Biological Sciences

September 2018

Abstract

This thesis aimed to investigate the role of pathogen-mediated selection in shaping patterns of genetic variation in Berthelot's pipit *Anthus berthelotii*, a passerine endemic to the Macaronesian archipelagos of the Canary Islands, Madeira, and Selvagens. I used restriction-site associated DNA sequencing (RAD-seq) to investigate patterns of neutral diversity among Berthelot's pipit populations, and in its sister species, the tawny pipit *Anthus campestris*, finding a loss of genome-wide diversity associated with colonisation history and genetic bottlenecks. I performed genome-wide association studies (GWAS) to identify genomic regions associated with malaria infection and bill length. I detected signatures of divergent selection around potential candidate genes related to immunity and metabolism, suggesting that these traits play roles in divergence and adaptation in this species. I then characterised genetic variation in Berthelot's and tawny pipits at avian β-defensins (AvBDs), a key gene family of the innate immune system. Allelic richness decreased with increasing numbers of bottlenecks. However, some AvBDs showed elevated nucleotide diversity compared with genome-wide trends. We found no evidence of local adaptation or balancing selection in Berthelot's pipits, suggesting that AvBD variation in this species is predominantly driven by genetic drift. Finally, I investigated whether malaria was driving fine-scale patterns of variation at SNPs identified in the earlier GWAS, and at the toll-like receptor 4 (TLR4) locus, a pathogen recognition receptor of the innate immune system, in two island populations. Both islands showed potential associations between malaria infection and TLR4 variation. In contrast, there was no association between a GWAS SNP and malaria risk in Tenerife, whereas in Porto Santo, the opposite trend to the original GWAS was found, potentially indicating local adaptation and population divergence. Overall, the evidence suggests that although functional variation in Berthelot's pipits is strongly influenced by demographic history, adaptation to pathogens may be occurring within and between populations.

Acknowledgements

I would first like to thank my supervisors David Richardson, Richard Davies and Lewis Spurgin. This work would not have been possible without their teaching and support over the past four years. Thank you for helping me through the hard times, for celebrating my achievements, and for encouraging me to have confidence in my abilities.

I would like to thank Gavin Horsburgh, Helen Hipperson, Clemens Küpper and Terry Burke from the NERC Biomolecular Analysis Facility at the University of Sheffield, and Matt Clark and Lawrence Percival-Alwyn at the Earlham Institute, for all of their training, help and contributions to my first publication. I am grateful for the support of Deborah Dawson, Rachel Tucker and Andy Krupa, who facilitated my time in Sheffield.

My fieldwork was made possible by a Heredity fieldwork grant from the Genetics Society, and by the Madeiran Natural Parks department, who generously offered us accommodation on Porto Santo. Thank you to Juan Carlos Illera for providing fieldwork training during our time on Madeira. I am very appreciative of the lab work carried out by Dom Hill, Molly Dunne and Maria-Elena Mannarelli, which greatly helped with analysing my fieldwork samples.

Thank you to everyone at UEA who has made the last four years highly enjoyable. Special thanks to Natasha Senior, David Litchfield, Kris Sales and Jessie Gardner for being great friends and housemates. A massive thank you to Stephanie Hirsch, Sean McGregor and Bethan Edmunds for always being there through ups and downs. I would like to thank Virginia, Simon and Louise Lamb for being so welcoming, it's been lovely getting to know you all.

Thank you Mum and Dad for everything you've done for me over the years, and for always believing in me. You've helped so much to get me to where I am today, and I will always be grateful for that. Thank you Nick for being my big brother, Bekki for becoming part of the Armstrong family, and both of you for making me an auntie to Zak.

Finally, I would like to thank Philip Lamb for your love and support over the past four years. I am so grateful for your patience and encouragement, and for always being there when things have been tough. Thank you for always trying to make me laugh, for your excellent pipit herding skills, and for being an all-round wonderful person.

Author contributions

One of the chapters in this thesis has been published, and two manuscripts are currently in preparation. All chapters have involved collaboration. I am the lead author on all manuscripts, and have made the largest contribution to all. Chapter 2 consists of two original chapters that were combined into one large paper.

Chapter 2. Armstrong C, Richardson DS, Hipperson H, Horsburgh GJ, Küpper C, Percival-Alwyn L, Clark M, Burke T, Spurgin LG (2018) *Evolution Letters* 2: 22-36

CA performed most of the lab work and bioinformatics, analysed the data and drafted the manuscript.

Chapter 3. Armstrong C, González-Quevedo C, Spurgin LG, Richardson DS (In prep.)

CA performed most of the lab work, analysed the data and drafted the manuscript.

Chapter 4. Armstrong C, Davies RG, González-Quevedo C, Dunne M, Spurgin LG, Richardson DS (In prep.)

CA collected the samples, performed part of the lab work, analysed the data and drafted the manuscript.

Contents

List of tables and figures

Chapter 1. General introduction

Berthelot’s pipit. Photo by Philip Lamb

1.1. Evolutionary drivers of genetic variation

Evolution can be defined as a change in allele frequencies in a population over generations. As such, the evolutionary dynamics of wild populations are influenced by forces that change allele frequencies: mutation, gene flow, genetic drift, and selection. The combination of these forces has the ability to increase or decrease variation within a population, and to shape patterns of between-population divergence or cohesion (Hartl & Clark 2007). Mutation is the fundamental mechanism for generating genetic novelty, giving rise to changes in the nucleotide sequence. However, the fate of the genetic variation produced through mutation is dependent upon the strength and direction of other evolutionary forces. In any finite population the frequency of alleles will fluctuate between generations due to the effects of random genetic drift (Hartl & Clark 2007). The alleles in the gametes passed from parents to offspring represent only a subsample of all the available alleles in the parental generation, resulting in stochastic changes in allele frequency. Genetic drift can increase an allele to fixation or cause it to be lost from the gene pool, but large populations are less prone to dramatic shifts in allele frequencies through drift than smaller populations (Wright 1931). As drift is a random process, the trajectory of allele frequency changes is likely to differ between populations, causing increasing divergence over time. Conversely, gene flow, the movement of genes between populations, has the power to decrease the degree of genetic variation found between populations, by homogenising the gene pools across populations. In addition, gene flow can increase within-population genetic variation by re-establishing alleles previously lost to drift or selection, or introducing new alleles that have arisen elsewhere via mutation (Slatkin 1987). Finally, natural selection is a key driver of genetic variation. In any given environment, there will be differential success in survival and reproduction between individuals, partially due to genetic differences. Organisms with genotypes that are well-adapted to the environment will produce relatively more offspring, causing beneficial alleles to spread throughout the population over time (Hartl & Clark 2007). The effect that natural selection has upon patterns of genetic variation depends on the type of selection. Directional selection can be either positive or purifying, acting to increase the frequency of an advantageous allele, or remove a maladapted allele, respectively. Alternatively, genes can be under balancing selection, a general term for any mechanism or combination of mechanisms by which multiple alleles at a locus are maintained at intermediate frequencies within a population.

Individuals within populations are exposed to many different selective pressures within their environment, such as stressful climatic conditions, intra- and inter-specific competition for resources, predation, and pathogens (here defined as infectious diseases and parasites). Pathogens pose a major challenge to wild populations, often imposing significant fitness costs

on their hosts through reduced reproductive output and increased mortality risk (Acevedo-Whitehouse & Cunningham 2006). This creates an evolutionary arms race between host and pathogen, where selection upon host mutations that confer resistance drives selection in the pathogen for alternative means to infect the host, and vice versa (Dawkins & Krebs 1979). Elevated levels of genetic diversity have been observed at loci involved in pathogen recognition within many different host populations of plants (Tian *et al.* 2002), invertebrates (Bérénos *et al.* 2011), and vertebrates (Piertney & Oliver 2006; Ferrer-Admetlla *et al.* 2008; Fumagalli *et al.* 2009), suggesting that balancing selection may be an important force in host-pathogen coevolution, maintaining a diverse pool of alleles capable of detecting and responding to a variety of pathogens.

1.2. Defining neutral variation and demography

The patterns of genetic variation found at functional loci such as the genes of the immune system may be influenced by all of the evolutionary forces outlined above. To determine the degree of within- and between-population variation that is due to natural selection within any organism, a baseline measurement of neutral variation is first required (Kohn *et al.* 2006). Patterns of immune gene variation that mirror neutral variation indicate that stochastic and/or demographic processes are the main determinant of functional diversity within the organism (eg. Miller *et al.* 2010; Marsden *et al.* 2012), whereas divergence from the pattern of neutral variation is suggestive of selection upon these specific genes (eg. Westerdahl *et al.* 2004; Cammen *et al.* 2011). Genetic markers such as microsatellite and mitochondrial DNA (mtDNA) have often been used to quantify historical and contemporary patterns of colonisation, genetic drift, and gene flow in plants (Reusch *et al.* 2000; Coyer *et al.* 2003; Olsen *et al.* 2004), invertebrates (Shaw *et al.* 1999; Duran *et al.* 2004; Figuerola *et al.* 2005), and vertebrates (Zhivotovsky *et al.* 2003; Martínez *et al.* 2006; Pinho *et al.* 2007). However, these aproaches have their limitations. Compared with nuclear DNA, mtDNA is (for the most part; see White *et al.* 2008) inherited uniparentally, and undergoes recombination at a lower rate (Piganeau *et al.* 2004; Tsaousis *et al.* 2005). The mutation rate of mtDNA is much higher (Brown *et al.* 1979), due to damage from reactive oxygen species and limited DNA repair (Bogenhagen 1999). Population history determined through mtDNA can result in limited or misleading interpretations (Palumbi & Baker 1994; Alves *et al.* 2006; Vallo *et al.* 2013). Elevated mutation rates are also found in microsatellites, at 10^{-2}–10^{-6} mutations per locus per generation (Ellegren 2000) versus 10^{-9} in nonrepetitive DNA (Li 1997). These differences in evolution may cause issues when comparing neutral variation at microsatellites and mtDNA with variation at functional loci (Miller *et al.* 2010).

With the introduction of next-generation sequencing (NGS) technology, it has become possible to study evolutionary processes by sequencing thousands of markers across the genome, or indeed sequencing the whole genome (Stapley *et al.* 2010; Jones *et al.* 2012; Hu *et al.* 2017). Despite falling costs of sequencing technologies, trade-offs between sample sizes and number of markers sequenced may occur, prohibiting whole-genome sequencing on large numbers of samples. Reduced-representation methods, which sample a smaller proportion of the genome while still resulting in thousands of markers, provide relatively cheap ways to sequence hundreds of individuals. One such method is restriction site associated DNA sequencing (RAD-seq). In this method, DNA is cut into fragments by restriction enzymes, followed by ligation of adapters to the sheared ends for the unique identification of each individual. Illumina sequencing of PCR-amplified and size-selected fragments can then generate thousands of reads from multiple individuals on a single sequencing run (Davey & Blaxter 2010). RAD markers were originally developed for use in microarray genotyping (Miller *et al.* 2007) and subsequently adapted for NGS technologies (Baird *et al.* 2008), massively increasing the breadth and resolution of single nucleotide polymorphism (SNP) discovery across the genome.

Compared to studies of demographic history based on neutral markers such as microsatellites, which only sample a very small section of the genome, the sequencing of thousands of genome-wide SNPs greatly improves the precision with which parameters such as gene flow, colonisation and divergence times and routes, and bottlenecks can be detected (Narum *et al.* 2013). RAD-seq has proven substantially more effective at detecting cryptic population structure in highly dispersive fish species than microsatellite or mtDNA approaches (Catchen *et al.* 2013; Corander *et al.* 2013; Hess *et al.* 2013), and has been utilised for assessing fine-scale demographic processes in a range of taxa such as invertebrates (Emerson *et al.* 2010; Lozier 2014; Alam *et al.* 2017), vertebrates (Stervander *et al.* 2015; Trucchi *et al.* 2016; Humble *et al.* 2018), plants (Andrew *et al.* 2013; Lexer *et al.* 2014) and yeast (Cromie *et al.* 2013; Hyma & Fay 2013).

A key advantage of RAD-seq for demographic studies is that it does not require prior knowledge of an organism's genome, or development of species- or population-specific primers, making it ideal for use in nonmodel organisms (Davey *et al.* 2011; Reitzel *et al.* 2013). An additional benefit of RAD-seq over mtDNA and microsatellites for studying pathogen-mediated selection is that RAD loci are subject to similar evolutionary mechanisms as immune loci. The use of SNPs in RAD-seq is likely to give more biologically meaningful measures of neutral variation, enabling greater accuracy in the quantification of selection (Brito & Edwards 2009). As the cost of whole-genome sequencing falls, the use of RAD-seq is likely to decline; however, it remains a cost-effective method for generating genomic data from large sample sizes.

1.3. The genomic basis of adaptation

In addition to improving demographic inferences, improvements in sequencing technologies have likewise facilitated our ability to locate genomic regions associated with phenotypic variation in ecologically important traits. The genome-wide association study (GWAS) approach, which detects statistical associations between genetic variants and phenotypes, has been widely used to identify potential candidate genes influencing disease resistance or susceptibility in humans (Price *et al.* 2015; Sud *et al.* 2017; Visscher *et al.* 2017) and productivity traits in agricultural species (Goddard & Hayes 2009; Huang & Han 2014). The increased availability of reference genomes for wild species has enabled the detection of candidate genes for adaptive traits in natural populations (Hecht *et al.* 2013; Johnston *et al.* 2014; Lind *et al.* 2017).

Signatures of divergent selection between populations can be identified through the use of genome scans. First proposed by Lewontin & Krakauer (1973; with numerous further developments, eg. Beaumont & Balding 2004; Foll & Gaggiotti 2008; Fariello *et al.* 2013), genome scans use the fixation index F_{ST} as a measure of allele frequency divergence between populations, giving a baseline level of F_{ST} expected under neutral divergence from genetic drift. Genetic outliers showing significantly greater F_{ST} than expected through drift alone are indicative of divergent selection. Mapping these regions to a reference genome can reveal potential candidates underlying local adaptation. In humans, genome scans have found evidence of recent local adaptation related to agricultural practices, climate, and disease (Voight *et al.* 2006; Pickrell *et al.* 2009; Kita & Fraser 2016). By sampling populations exposed to different environments, genomic regions undergoing local adaptation have been identified related to environmental conditions such as altitude (Bonin *et al.* 2006; Guo *et al.* 2016), diet availability (Bosse *et al.* 2017), latitude (Pujolar *et al.* 2014), salinity (Dalongeville *et al.* 2018) and precipitation (Jones *et al.* 2013).

The utility of RAD-seq for GWAS and genome scan approaches has been questioned, owing to the relatively low density of genetic markers compared with other NGS methods (Lowry *et al.* 2017). In response, Mckinney *et al.* (2017) argue that the presence of many RAD-seq studies with positive findings, combined with the ease of use for studying non-model species, demonstrate that RAD-seq is an advantageous tool for investigating the genomics of adaptation. Reference genomes from closely related species can be utilised to map high numbers of RAD loci from species without genomic resources (Bourgeois *et al.* 2013), to enable the identification of potential candidate loci. RAD-seq has been used to successfully detect genetic associations with adaptive traits such as disease resistance (Liu *et al.* 2015; Arafa *et al.* 2017; Aslam *et al.* 2018), growth rate (Yu *et al.* 2018), beak morphology (Chaves *et al.* 2016) and fire tolerance (He *et al.* 2018). Genome scans on RAD-seq loci have revealed divergent selection and local

adaptation to environmental variables such as pollution levels (Laporte *et al.* 2016), salinity (Guo *et al.* 2016), urbanisation (Theodorou *et al.* 2018), and temperature (Babin *et al.* 2017; Leydet *et al.* 2018). While it may be likely that some outlier regions may go undetected when using RAD loci for genome scans and GWAS, this is outweighed by the benefits of RAD-seq, in terms of both cost and readiness of use in non-model species, to facilitate the discovery of adaptive evolution (Catchen *et al.* 2017).

1.4. Pathogen-mediated balancing selection

Three main mechanisms have been proposed to explain how pathogens may drive the maintenance of genetic variation by balancing selection. In the heterozygote advantage model, genetic diversity is maintained because individuals that are heterozygous at a given locus have higher fitness than individuals that are homozygous (Dobzhansky 1955). In respect to immune genes, if multiple alleles are able to recognise and respond to different pathogen strains, or provide a stronger response to a single pathogen, then being heterozygous increases the ability of an individual to defend itself against infection (Doherty & Zinkernagel 1975). The classical concept of heterozygote advantage is due to overdominance, where heterozygote fitness is higher than either form of homozygote. However, the advantage caused by dominance, whereby heterozygotes have equal fitness to homozygote dominant individuals, but higher fitness than homozygote recessive individuals, giving an overall higher fitness on average (Penn *et al.* 2002) is also considered a type of heterozygote advantage.

In laboratory tests of pathogen-mediated selection, overdominance was detected for mice strains coinfected with *Salmonella enterica* and Theiler's murine encephalomyelitis virus, with superior survival in heterozygotes relative to both homozygotes (McClelland *et al.* 2003). Converseley, in mice exposed to multiple strains of *Salmonella*, heterozygote advantage in the F2 generation was due to dominance, with heterozygotes outperforming only one of the parental homozygote lines (Penn *et al.* 2002). However, the artificial conditions of laboratory experiments involving inbred strains of mice may not be ideal for inferring evolutionary processes in wild populations. Comparisons of natural populations exposed to a disease outbreak have demonstrated heterozygote advantage in red junglefowl *Gallus gallus* due to the masking of a susceptible allele in heterozygotes (Worley *et al.* 2010). House finch *Carpodacus mexicanus* individuals with high MHC diversity were more resistant to *Mycoplasma gallisepticum* infection, with balancing selection observed only in populations that had experienced *M. gallisepticum* outbreaks (Hawley & Fleischer 2012). In humans, heterozygote advantage was found at the MHC for delaying the onset of AIDS in HIV patients; homozygosity

at one or more of the three loci sampled was linked to rapid progression of AIDS and death (Carrington *et al.* 1999). Heterozygote advantage at a polymorphism in CYBA (a regulatory gene of the immune system) has been detected for tuberculosis susceptibility, with lower rates of tuberculosis found in heterozygotes (Liu *et al.* 2016).

Balancing selection may also be explained by the rare-allele advantage hypothesis, a negative frequency-dependent model, where the selective advantage of an allele is dependent on its relative frequency within a population (Ayala & Campbell 1974). In cyclical coevolution between host and pathogen, selection drives increased resistance in the host, and increased infectivity in the pathogen. When an immune gene allele is common, there is selection upon the pathogen for increased infectivity against hosts possessing this allele. As pathogen infectivity increases, the fitness of these hosts will decrease. Having a rare allele that pathogens have not evolved against gives hosts a selective advantage and will increase the frequency of the rare allele, thereby continuing the cycle of host-pathogen coevolution (Seger 1988). Rare-allele advantage can promote the maintenance of genetic variation by constantly selecting for rare alleles, thus sustaining multiple alleles at low frequencies in the population (Slade & McCallum 1992).

Evidence of negative frequency-dependent selection has been demonstrated in HIV patients, with highest viral load in individuals with common alleles (Trachtenberg *et al.* 2003). Similar results have been found for the wild Soay sheep *Ovis aries* population, where rare MHC alleles are significantly associated with increased survivorship of lambs and yearlings (Paterson *et al.* 1998), and in Malagasy mouse lemur *Microcebus murinus*, where individals with rare alleles are more likely to be uninfected or have low parasite loads (Schad *et al.* 2005). In other studies, significant associations between specific MHC alleles and susceptibility to pathogens have been found, without evidence of heterozygote advantage, in yellow-necked mouse *Apodemus flavicollis* (Meyer-Lucht & Sommer 2005), giant pandas *Ailuropoda melanoleuca* (Zhang *et al.* 2014), rabbits *Oryctolagus cuniculus* (Schwensow *et al.* 2017), and guppies *Poecilia reticulata* (Phillips *et al.* 2018). A comparison of two populations of house sparrow *Passer domesticus* located several hundred kilometres apart identified two population-specific alleles that each conferred protection against a single strain of malaria present in both populations (Bonneaud *et al.* 2006). This suggests that selection at these loci is driven by rare allele advantage, with each population independently involved in a coevolutionary cycle with the malaria strain (Spurgin & Richardson 2010). An alternative means to detect cyclical coevolution between host and pathogen is through time-shift experiments, where a host population from a particular time period is exposed to pathogen populations from past, contemporary, and future time periods (Gaba & Ebert 2009). Rare-allele advantage has been detected in this manner between plants and fungi (Thrall *et al.* 2012), invertebrates and bacteria (Decaestecker *et al.* 2007), bacteria and

bacteriophages (Gómez & Buckling 2011; Koskella 2013), and between host-defensive and pathogenic bacteria (Ford *et al.* 2017). In line with negative frequency-dependence, pathogens showed highest infectivity against contemporary hosts. These experiments found little evidence of escalating arms races characterised by selective sweeps; instead, fluctuations in frequencies of several host and pathogen genotypes maintained genetic diversity at immune loci.

Thirdly, the fluctuating selection hypothesis proposes that genetic diversity at immune genes is maintained by spatial and temporal heterogeneity of pathogen communities (Hill 1991). Across a host meta-population, different subpopulations will be subjected to different types and frequencies of pathogens. This will be partly due to abiotic and biotic factors that determine whether an environment is suitable for a particular pathogen (Craig *et al.* 1999), but also due to stochastic and demographic processes acting upon the pathogen population and governing patterns of local extinction and colonisation (Hanski 1991). This can result in episodes of directional selection and local adaptation at the subpopulation level, whilst the overall pattern at the meta-population level is of balancing selection (Hedrick 2002).

Evidence of fluctuating selection acting on immune genes has often been identified through comparisons of neutral and adaptive genetic variation, with greater population divergence in immune genes relative to neutral variation being suggestive of divergent selection (Charbonnel & Pemberton 2005; Ekblom *et al.* 2007; Evans *et al.* 2010; Li *et al.* 2016). Temporal fluctuations in immune loci variation have been found at a variety of scales. For example, within-population divergence was found in brown trout *Salmo trutta* populations sampled 12 years apart (Jensen *et al.* 2008). However, no temporal selection was detected for populations of brown hare *Lepus europaeus* over a period of eight generations, suggesting that this period may be too short to find signatures of selection, or that selection pressures maintained relatively constant over this time period (Campos *et al.* 2011). Conversely, a survey period of eight years was long enough to reveal temporal fluctuations in MHC allele frequencies of great reed warbler *Acrocephalus arundinaceus* (Westerdahl *et al.* 2004). At a much smaller scale, within-year fluctuations of malaria resistance genes were found in *Anopheles gambiae* mosquitoes, linked to seasonal rainfall changes (Rottschaefer *et al.* 2011). Though spatial and/or temporal variation in immune loci has been demonstrated in the above studies, none of these monitored changes in the pathogen community present in the host population. Characterising spatial and temporal changes in pathogen density and diversity, and relating it to host genetic variation (Hill *et al.* 1991; Mackinnon *et al.* 2016; Osborne *et al.* 2017; Wang *et al.* 2017), provides much stronger evidence for the role of fluctuating selection dynamics in shaping immune gene variation. Pathogen heterogeneity in relation to altitude (Jones *et al.* 2014; Kosch *et al.* 2016), rainfall (Froeschke & Sommer 2014), temperature (Dionne *et al.* 2007), habitat type (Eizaguirre *et al.*

2012), and agriculture (González-Quevedo *et al.* 2016) has been shown to drive patterns of local adaptation in host immune genes.

Several studies have indicated that a combination of models of pathogen-mediated balancing selection may act within the same population or between locations. Both heterozygote advantage at the subpopulation level, and spatial fluctuation with local adaptation has been demonstrated in three-spined stickleback *Gasterosteus aculeatus* (Wegner *et al.* 2003), Chinook salmon *Oncorhynchus tshawytscha* (Evans & Neff 2009), leopard frog *Lithobates yavapaiensis* (Savage & Zamudio 2011), and water vole *Arvicola terrestris* (Oliver *et al* 2009a; Oliver *et al* 2009b). In a bottlenecked population of wolves *Canis lupus*, both heterozygotes and carriers of a specific MHC allele had lower nematode burdens (Niskanen *et al.* 2014). Rare-allele advantage and fluctuating selection were recorded for brandt's vole *Lasiopodomys brandtii* (Zhang & He 2013) and rufous-collared sparrows *Zonotrichia capensis* (Jones *et al.* 2015), and all three models were implicated in populations of badger *Meles meles* (Sin *et al.* 2014). Furthermore, modelling has demonstrated that heterozygote advantage alone is not enough to maintain the high levels of polymorphism observed at immune loci (De Boer *et al.* 2004; Ding & Goudet 2005; Ejsmond & Radwan 2015). It is likely that in many populations, different mechanisms of pathogen-mediated selection are interacting to shape immunogenetic variation, though they may go undetected due to inherent difficulties in separating them (reviewed in Spurgin & Richardson 2010). For example, any heterozygote advantage detected could be due to fitness benefits from rare alleles, which, due to their low frequencies, would be more commonly found in heterozygotes than homozygotes (Lachance 2008). Separating the effects of rare-allele advantage and fluctuating selection is also problematic, as both mechanisms could present the same patterns. Ideally, long-term studies with spatial and temporal variation of immune genes, neutral markers, and pathogen load would be required to determine whether selection is directional in response to fluctuations in the environment or pathogen regime, or whether alllele frequencies are subject to frequency-dependent selection (Spurgin & Richardson 2010).

1.5. Toll-like receptors

Despite the heavy focus on MHC genes in vertebrate studies of pathogen-mediated selection (Sommer 2005; Piertney & Oliver 2006; Eizaguirre & Lenz 2010; Bordes & Morand 2011; Kosch *et al.* 2016; Wang *et al.* 2017), the adaptive immunity conferred by these genes comprises only one component of a larger vertebrate immune response. The various genes of the innate immune system also play a major role in host response to infection (Kumar *et al.* 2011); in humans, the proportion of phenotypic variance in immune response attributable to MHC genes

is exceeded by that of non-MHC genes (Jepson *et al.* 1997). As it can take up to one week for the adaptive immune system to be ready to combat a new infection, the innate immune system provides an invaluable first line of defence for a rapid response to pathogens (Vinkler & Albrecht 2009). As such, the innate immune system is an important aspect of host evolution of pathogen resistance in natural populations (Acevedo-Whitehouse & Cunningham 2006; Iwasaki & Medzhitov 2015).

Toll-like receptor (TLR) genes are excellent candidates for studying pathogen-mediated selection. The TLR molecules they encode are transmembrane receptors that regulate the innate immune response through pathogen recognition and binding (Medzhitov *et al.* 1997). TLRs recognise highly conserved molecular motifs of pathogens such as bacteria, viruses, fungi and protozoa known as pathogen-associated molecular patterns (PAMPs; Janeway & Medzhitov 2002; Netea *et al.* 2012). Pathogen recognition occurs at the extracellular domain of the TLR, within which PAMPs are bound to a diverse array of three-dimensional horseshoe-like structures formed of leucine-rich repeat (LRR) motifs (Bella *et al.* 2008). TLRs can be classified as either viral or non-viral, based on functional roles. Viral TLRs respond to viral PAMPs such as single- and double-stranded RNA and unmethylated CpG DNA, and are expressed within the cell. Non-viral TLRs recognise features such as lipopolysaccharides, lipopeptides, flagellins and bacterial ribosomal RNA, and are expressed on the surface of the cell (Barton 2007; Wlasiuk & Nachman 2010).

Signatures of positive selection on TLRs have been demonstrated in inter-species comparisons in mammalian and avian lineages. By determining the ratio of nonsynonymous nucleotide substitutions at nonsynonymous sites to synonymous substitutions at synonymous sites (dN/dS), it is possible to infer whether selection is positive (dN/dS > 1), neutral (dN/dS = 1), or purifying (dN/dS < 1). An analysis of TLR sequences in 15 primate species revealed signatures of positive selection for only six of ten TLR loci, with viral TLRs showing stronger purifying selection than non-viral TLRs (Wlasiuk & Nachman 2010). The authors predict this is due to the use of nucleic acids for viral recognition, with strong evolutionary constraints required to avoid triggering an autoimmune response. However, this approach measured selection across the whole TLR gene sequence; analysis of the same genes in mammals found positive selection within the extracellular domain for every gene, with similar trends for both viral and non-viral TLRs (Areal *et al.* 2011). In rodent TLR2, overall selection across the gene was predominantly purifying, however significant dN/dS peaks were evident within the extracellular domain (Tschirren *et al.* 2011b). Similar results were obtained across 10 genes of the avian TLR gene family for up to 23 bird species per gene, with signs of episodic positive selection in an average of 4.5% of codons. In addition, when comparing the three TLR domains, the extracellular domain

consistently portrayed higher average dN/dS, and greater numbers of codons under positive selection (Grueber *et al.* 2014). Strong positive selection was detected in the majority of avian TLR genes across 63 species representing all major clades, with over a quarter of positively selected sites predicted to have large effects on physicochemical properties of the resulting proteins (Velová *et al.* 2018). Positive selection was found at a number of TLR genes in the pig family (Suidae), at amino acid sites in close proximity to functionally relevant sites (Darfour-Oduro *et al.* 2015). While purifying selection has dominated the evolution of newt *Lissotriton* spp. TLRs, positive selection has been found between two diverging lineages (Babik *et al.* 2014).

Recent studies have investigated TLR variation at the population level, to explore the role of pathogen-mediated selection in driving TLR evolution. Low to moderate levels of TLR polymorphisms have been discovered in wild populations of koala *Phascolarctos cinereus* (Cui *et al.* 2015), lesser kestrel *Falco naumanni* and house finch *Haemorhous mexicanus* (Alcaide & Edwards 2011), suggesting the potential for balancing selection. Though unlikely to reach the extreme levels of diversity observed at the MHC, TLR variation may be widespread, as polymorphisms were discovered in a highly bottlenecked population of New Zealand's Stewart Island robin *Petroica australis* that originated from just 12 founders, though diversity was low (Grueber *et al.* 2012). Comparable levels of TLR diversity were also found in the endangered African penguin *Spheniscus demersus*, which has experienced recent declines in population size (Dalton *et al.* 2016). Moderate variation at TLR4 was found in populations of the mouse subspecies *Mus musculus domesticus* and *M. m. musculus*, though with contrasting patterns of diversity between the two subspecies. While *M. m. musculus* was characterised by a number of haplotypes at intermediate frequency, *M. m. domesticus* was largely dominated by a single haplotype (Fornůsková *et al.* 2014). Spatial variation in TLR2 diversity across sympatric populations of yellow-necked mouse *Apodemus flavicollis* and bank vole *Myodes glareolus* likewise revealed contradictory selection processes (Tschirren *et al.* 2011a); in bank voles, TLR2 showed evidence of balancing selection through local adaptation, whereas yellow-necked mice appeared to have undergone a selective sweep. This may relate to the observations of episodic positive selection across avian lineages by Grueber *et al.* (2014), with an asynchrony in the timing of selective sweeps between bank voles and yellow-necked mice (Tschirren *et al.* 2011a). Drift was found to be the predominant force driving contemporary TLR variation in the bottlenecked Seychelles warbler *Acrocephalus sechellensis*, with patterns of variation suggestive of historical balancing selection (Gilroy *et al.* 2017). Conversely, balancing selection was suggested to be maintaining TLR4 variants associated with parasite burden in a population of water voles *Arvicola amphibius* through a bottleneck (Gavan *et al.* 2015).

1.6. Antimicrobial peptides and β-defensins

Antimicrobial peptides (AMPs) are an ancient lineage of effector molecules common to all life (Hancock & Diamond 2000). In plants and animals, AMPs provide innate, broad-scale protection against bacteria, viruses, protozoa and fungi (Zasloff 2002). AMPs are composed of three sections: a signal sequence, a propiece, and a mature peptide, which cleaves off from the rest of the protein and kills microbes through disruption of the cellular membrane (Yeaman & Yount 2003). Positive selection on the nucleotide sequence encoding the mature peptide has been discovered throughout the evolutionary history of AMPs, where adaptation to fight new or modified pathogens has been inferred from elevated rates of amino acid substitutions (Tennessen 2005). This has resulted in a highly diverse selection of AMP genes across major groups such as invertebrates (Bulmer & Crozier 2004), primates (Maxwell *et al.* 2003), amphibians (Duda *et al.* 2002), fish (Padhi & Verghese 2007), and birds (Hellgren & Ekblom 2010).

It has been hypothesised that balancing selection may be less important for maintaining variation at generalist compared to specialist defence genes (Tiffin *et al.* 2004). For a pathogen to avoid detection by a specialist defence gene, it may only require a small sequence change, whereas for a generalist that targets the cell membrane, several coordinated changes may be necessary, allowing enough evolutionary time for an allele to undergo a selective sweep to fixation (Schröder 1999). Support for this view was demonstrated at the frog AMP locus Ranatuerin2, where a recent positive selective sweep has resulted in no allelic variation in *Rana pipiens*, but high divergence between *R. pipiens* and the closely related *Rana chiricahuensis* (Tennessen & Blouin 2007). By contrast, however, a population genetic study of the Brevinin1.1 AMP locus in *R. pipiens* revealed evidence of strong balancing selection and geographic variation in allelc diversity (Tennessen & Blouin 2008).

In vertebrates, the defensins are a diverse family of cationic AMPs, providing antimicrobial defence as well as signalling roles and anti-inflammatory responses (Van Dijk *et al.* 2008). Defensins are grouped into α-, β- and θ-defensin gene families. β-defensins are present in all vertebrates, while α-defensins are found in mammals, and θ-defensis are only present in some primates (Liu *et al.* 1997; Tang *et al.* 1999). Gene duplication, and subsequent divergence through positive selection on the mature peptide, has been a common feature throughout the evolution of defensins (Semple *et al.* 2006; Hellgren & Ekblom 2010; Cheng *et al.* 2014; Jones *et al.* 2017; Tang *et al.* 2018). Evidence of balancing selection has been found in the β-defensins: analysis of five β-defensins shown to be under positive selection in primates revealed high levels of allelic diversity in human populations, with 16 out of 20 polymorphic sites within the mature peptide resulting in amino acid substitutions. This balancing selection appears to be due to

ecological differences in selective pressures, with strong divergence between hunter-gatherer and agricultural populations (Hollox & Armour 2008). Further support for balancing selection within this gene family has been found in waterfowl (Anatidae) avian β-defensins (AvBDs), both within and across species (Chapman *et al.* 2016).

Contrary to gene families such as the MHC and TLRs whose products trigger a cascade of immune responses upon detection of a pathogen, it is the β-defensins themselves that kill or inhibit pathogens. Thus there is a direct link between the sequence of β-defensin genes and the ability of their proteins to combat pathogens. This makes them a highly attractive target for the study of host-pathogen evolution, as the effects of genetic variation at β-defensins are not dependent upon the variation present at other loci further down an immune response pathway (Hellgren & Sheldon 2011). This same property also makes it possible to investigate *in vitro* the link between allelic variation and antimicrobial efficacy, to confirm that changes to the amino acid sequence translate into functional differences (Klüver *et al.* 2005; Higgs *et al.* 2007; Yang *et al.* 2017). Natural allelic variation at β-defensins has been found in a number of species (eg. Hellgren 2015; Gilroy *et al.* 2016; Schmitt *et al.* 2017b), with evidence for functional variation in antimicrobial activity (Hellgren *et al.* 2010; Cadwell *et al.* 2017; Schmitt *et al.* 2017a). With evidence of both balancing selection and functional allelic variation, the β-defensins are an ideal, yet relatively unexplored, family for studying pathogen-mediated selection.

1.7. Pathogen-mediated selection and adaptive variation in Berthelot's pipit

Berthelot's pipit *Anthus berthelotii* is a small, sedentary passerine endemic to the Macaronesian islands comprising the Canary Islands, the Selvagens, and the Madeiran archipelago, and is both locally abundant and widespread throughout these islands (Martín & Lorenzo 2001; Oliveira & Menezes 2004; Illera *et al.* 2007). An analysis of 21 microsatellite loci detected genetic bottlenecks in both the Selvagens and the Madeiran islands. From this, alongside approximate Bayesian computation analyses, it has been inferred that Berthelot's pipit underwent two simultaneous colonisation events from the Canary Islands to Madeira and Selvagens ca. 8000 years ago, with genetic and morphological variation supporting a model of isolation-by-colonisation, shaped by founder events (Spurgin *et al.* 2014). Berthelot's pipit is classified into two subspecies: *A. berthelotii madeirensis*, found in the Madeiran archipelago, is characterised by longer bill length than *A. berthelotii berthelotii*, which inhabits the Canary Islands and Selvagens (Martín & Lorenzo 2001; Oliveira & Menezes 2004). Bill morphology is an ecologically important trait associated with local adaptation and speciation (Grant 1986; Lamichhaney *et al.* 2016; Bosse *et al.* 2017). However, as both divergent selection and founder effects can drive

phenotypic differentiation between populations (Clegg *et al.* 2002; Kolbe *et al.* 2012), it is not known if bill length differences in Berthelot's pipits are driven by selection.

Pathogens appear to exert substantial selection pressure on Berthelot's pipit. Screening of birds from each of the islands in their range detected two strains of *Plasmodium*, four strains of *Leucocytozoon*, and one strain of avian pox (Spurgin *et al.* 2012). Infected individuals had significantly greater body mass, suggesting that these pathogens inflict severe fitness costs, whereby only the highest quality individuals are capable of surviving. The same study identified spatial variation in pathogen communities, which could allow for pathogen-mediated selection maintained by local adaptation to particular pathogen regimes. In line with island biogeography theory (MacArthur & Wilson 1967), pathogen prevalence was strongly associated with island size and isolation, with smaller, more isolated islands harbouring fewer pathogens. Within Tenerife, the largest island inhabited by Berthelot's pipit, subpopulations from coastal areas and an alpine plateau showed strong spatial differentiation in pathogen densities (Spurgin *et al.* 2012). Fine-scale mapping of *Plasmodium* infections was carried out by catching birds in every 1 km^2 of accessible pipit habitat (González-Quevedo *et al.* 2014). The key predictor of *Plasmodium* variation was minimum temperature in the coldest month, in line with environmental constraints of its vector species; in addition, anthropogenic factors were revealed to have strong effects. Distance to artificial water reservoirs played a role in predicting *Plasmodium* distributions, due to increased habitat for vector development. Proximity to poultry farms was also a significant predictor, however it is unclear whether chickens are acting as a reservoir from which the local pipit population is becoming infected, or whether factors such as increased density of wild birds around poultry farms are causing this relationship. In a further study of the same population, the presence of two MHC alleles were influenced by the proximity to poultry farms, with one allele also showing a significant association with malaria risk, suggesting that pathogens may be driving fine-scale adaptive evolution in this species (González-Quevedo *et al.* 2016).

Preliminary work on avian TLRs in Berthelot's pipit has detected polymorphisms at five TLR genes, with significant differentiation in genetic variation between archipelagos (González-Quevedo *et al.* 2015). This closely mirrored inferred demographic patterns (Illera *et al.* 2007; Spurgin *et al.* 2014), suggesting strong influences of founder effects; however, signatures of positive selection within the extracellular domain were identified in TLR3 and TLR4. Furthermore, two private TLR alleles were discovered in the Madeiran population, at relatively high equal frequencies, each resulting in nonsynonymous sequence changes. No private alleles were found for either microsatellites or mtDNA in this population (Illera *et al.* 2007; Spurgin *et*

al. 2014), suggesting that these TLR alleles have evolved *in situ* as a result of recent selection (González-Quevedo *et al.* 2015).

1.8. Thesis aims

In this thesis, I use a combination of genomic approaches and candidate genes to investigate the processes driving adaptive genetic variation in wild populations and the role of pathogens in shaping genetic diversity of immune loci. In Chapter 2, I use RAD-seq to first assess fine-scale patterns of neutral diversity and population structure within Berthelot's pipits and in comparison to the mainland sister species, the tawny pipit *Anthus campestris*. I then detect genomic regions associated with malaria infection and bill length throughout Canary Islands populations of Berthelot's pipits. I also perform a genome scan to identify signatures of divergent selection between the archipelagos. In Chapter 3, I design primers to sequence the mature peptides of AvBD loci, to characterise population-level AvBD diversity in Berthelot's and tawny pipits. I assess whether balancing selection has maintained AvBD diversity through colonisation events, using the RAD data as a measure of neutral diversity. In Chapter 4, I determine the environmental drivers of malaria risk on Porto Santo, the population with the highest recorded blood parasite prevalence throughout Berthelot's pipits. I then compare fine-scale genetic associations with malaria infection and risk on Porto Santo and Tenerife, both at TLR4 and at two RAD SNPs identified in Chapter 1 as showing significant associations with malaria infection in the Canary Islands. Finally, in Chapter 5, I discuss the findings of the previous Chapters and their combined significance, and suggest future research directions.

1.9. References

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

Alam, M.M.M., Westfall, K.M. & Pálsson, S. (2017). Historical demography and genetic differentiation of the giant freshwater prawn *Macrobrachium rosenbergii* in Bangladesh based on mitochondrial and ddRAD sequence variation. *Ecology and Evolution*, 7, 4326–4335.

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

Alves, P.C., Harris, D.J., Melo-Ferreira, J., Branco, M., Suchentrunk, F. & Boursot, P. *et al.* (2006). Hares on thin ice: Introgression of mitochondrial DNA in hares and its implications for recent phylogenetic analyses. *Molecular Phylogenetics and Evolution*, 40, 640–641.

Andrew, R.L., Kane, N.C., Baute, G.J., Grassa, C.J. & Rieseberg, L.H. (2013). Recent nonhybrid origin of sunflower ecotypes in a novel habitat. *Molecular Ecology*, 22, 799–813.

Arafa, R.A., Rakha, M.T., Soliman, N.E.K., Moussa, O.M., Kamel, S.M. & Shirasawa, K. (2017). Rapid identification of candidate genes for resistance to tomato late blight disease using next-generation sequencing technologies. *PLoS ONE*, 12, e0189951.

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

Aslam, M.L., Carraro, R., Bestin, A., Cariou, S., Sonesson, A.K. & Bruant, J.S. *et al.* (2018). Genetics of resistance to photobacteriosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *BMC Genetics*, 19, 43.

Ayala, F.J. & Campbell, A. (1974). Frequency-dependent selection. *Annual Review of Ecology and Systematics*, 5, 115–138.

Babik, W., Dudek, K., Fijarczyk, A., Pabijan, M., Stuglik, M. & Szkotak, R. *et al.* (2014). Constraint and adaptation in newt Toll-like receptor genes. *Genome Biology and Evolution*, 7, 81–95.

Babin, C., Gagnaire, P.A., Pavey, S.A. & Bernatchez, L. (2017). RAD-Seq reveals patterns of additive polygenic variation caused by spatially-varying selection in the American eel (*Anguilla rostrata*). *Genome Biology and Evolution*, 9, 2974–2986.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L. & Lewis, Z.A. *et al.* (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3, e3376.

Barton, G.M. (2007). Viral recognition by Toll-like receptors. *Seminars in Immunology*, 19, 33–40.

Beaumont, M.A. & Balding, D.J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, 13, 969–980.

Bella, J., Hindle, K.L., McEwan, P.A. & Lovell, S.C. (2008). The leucine-rich repeat structure. *Cellular and Molecular Life Sciences*, 65, 2307–2333.

Bérénos, C., Wegner, K.M. & Schmid-Hempel, P. (2011). Antagonistic coevolution with parasites maintains host genetic diversity: an experimental test. *Proceedings of the Royal Society B: Biological Sciences*, 278, 218–224.

Bogenhagen, D.F. (1999). Repair of mtDNA in vertebrates. *American Journal of Human Genetics*, 64, 1276–1281.

Bonin, A., Taberlet, P., Miaud, C. & Pompanon, F. (2006). Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, 23, 773–783.

Bonneaud, C., Pérez-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.

Bordes, F. & Morand, S. (2011). The impact of multiple infections on wild animal hosts: a review. *Infection Ecology & Epidemiology*, 1, 7346.

Bosse, M., Spurgin, L.G., Laine, V.N., Cole, E.F., Firth, J.A. & Gienapp, P. *et al.* (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, 358, 365–368.

Bourgeois, Y.X.C., Lhuillier, E., Cézard, T., Bertrand, J.A.M., Delahaie, B. & Cornuault, J. *et al.* (2013). Mass production of SNP markers in a nonmodel passerine bird through RAD

sequencing and contig mapping to the zebra finch genome. *Molecular Ecology Resources*, 13, 899–907.

Brito, P.H. & Edwards, S.V. (2009). Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, 135, 439–455.

Brown, W.M., George, M. & Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, 76, 1967–971.

Bulmer, M.S. & Crozier, R.H. (2004). Duplication and diversifying selection among termite antifungal peptides. *Molecular Biology and Evolution*, 21, 2256–2264.

Cadwell, K., Niranji, S.S., Armstrong, V.L., Mowbray, C.A., Bailey, R. & Watson, K.A. *et al.* (2017). AvBD1 nucleotide polymorphisms, peptide antimicrobial activities and microbial colonisation of the broiler chicken gut. *BMC Genomics*, 18, 637.

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

Campos, J.L., De Bellocq, J.G., Schaschl, H. & Suchentrunk, F. (2011). MHC class II DQA gene variation across cohorts of brown hares (*Lepus europaeus*) from eastern Austria: Testing for different selection hypotheses. *Mammalian Biology*, 76, 251–257.

Carrington, M., Nelson, G.W., Martin, M.P., Kissner, T., Vlahov, D. & Goedert, J.J. *et al.* (1999). HLA and HIV-1: heterozygote advantage and *B*35-Cw*04* disadvantage. *Science*, 283, 1748–1752.

Catchen, J.M., Bassham, S., Wilson, T., Currey, M.C., O'Brien, C. & Yeates, Q. *et al.* (2013). The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology*, 22, 2864–2883.

Catchen, J.M., Hohenlohe, P.A., Bernatchez, L., Funk, W.C., Andrews, K.R. & Allendorf, F.W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, 17, 362–365.

Chapman, J.R., Hellgren, O., Helin, A.S., Kraus, R.H.S., Cromie, R.L. & Waldenström, J. (2016). The evolution of innate immune genes: Purifying and balancing selection on β-defensins in waterfowl. *Molecular Biology and Evolution*, 33, 3075–3087.

Charbonnel, N. & Pemberton, J. (2005). A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. *Heredity*, 95, 377–388.

Chaves, J.A., Cooper, E.A., Hendry, A.P., Podos, J., De León, L.F. & Raeymaekers, J.A.M. *et al.* (2016). Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Molecular Ecology*, 25, 5282–5295.

Cheng, D.-Q., Li, Y. & Huang, J.-F. (2014). Molecular Evolution of the Primate α-/θ-Defensin Multigene Family. *PLoS ONE*, 9, e97425.

Clegg, S.M., Degnan, S.M., Moritz, C., Estoup, A., Kikkawa, J. & Owens, I.P.F. (2002). Microevolution in island forms: the roles of drift and directional selection in morphological divergence of a passerine bird. *Evolution*, 56, 2090–2099.

Corander, J., Majander, K.K., Cheng, L. & Merilä, J. (2013). High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Molecular Ecology*, 22, 2931–2940.

Coyer, J.A., Peters, A.F., Stam, W.T. & Olsen, J.L. (2003). Post-ice age recolonization and differentiation of *Fucus serratus* L. (Phaeophyceae; Fucaceae) populations in Northern Europe. *Molecular Ecology*, 12, 1817–1829.

Craig, M.H., Le Sueur, D. & Snow, R.W. (1999). A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today*, 15, 105–111.

Cromie, G.A., Hyma, K.E., Ludlow, C.L., Garmendia-Torres, C., Gilbert, T.L. & May, P. *et al.* (2013). Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3: Genes, Genomes, Genetics*, 3, 2163–2171.

Cui, J., Frankham, G.J., Johnson, R.N., Polkinghorne, A., Timms, P. & O'Meally, D. *et al.* (2015). SNP marker discovery in koala TLR genes. *PLoS ONE*, 10, e0121068.

Dalongeville, A., Benestan, L., Mouillot, D., Lobreaux, S. & Manel, S. (2018). Combining six genome scan methods to detect candidate genes to salinity in the Mediterranean striped red mullet (*Mullus surmuletus*). *BMC Genomics*, 19, 217.

Dalton, D.L., Vermaak, E., Roelofse, M. & Kotze, A. (2016). Diversity in the toll-like receptor genes of the African penguin (*Spheniscus demersus*). *PLoS ONE*, 11, e0163331.

Darfour-Oduro, K.A., Megens, H.-J., Roca, A.L., Groenen, M.A.M. & Schook, L.B. (2015). Adaptive evolution of Toll-like receptors (TLRs) in the family Suidae. *PLoS ONE*, 10, e0124069.

Davey, J.W. & Blaxter, M.L. (2010). RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, 9, 416–423.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. & Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, 499–510.

Dawkins, R. & Krebs, J.R. (1979). Arms races between and within species. *Proceedings of the Royal Society B: Biological Sciences*, 205, 489–511.

De Boer, R.J., Borghans, J.A.M., Boven, M. van, Keşmir, C. & Weissing, F.J. (2004). Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics*, 55, 725–731.

Decaestecker, E., Gaba, S., Raeymaekers, J.A.M., Stoks, R., Van Kerckhoven, L. & Ebert, D. *et al.* (2007). Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature*, 450, 870–873.

Ding, A. & Goudet, J. (2005). Heterozygote advantage and the maintenance of polymorphism for multilocus traits. *Theoretical Population Biology*, 68, 157–166.

Dionne, M., Miller, K.M., Dodson, J.J., Caron, F. & Bernatchez, L. (2007). Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution*, 61, 2154–2164.

Dobzhansky, T. (1955). A review of some fundamental concepts and problems of population genetics. *Cold Spring Harbor Symposium Quantitative Biology*, 20, 1–15.

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

Duda, T.F., Vanhoye, D. & Nicolas, P. (2002). Roles of diversifying selection and coordinated evolution in the evolution of amphibian antimicrobial peptides. *Molecular Biology and Evolution*, 19, 858–864.

Duran, S., Pascual, M., Estoup, A. & Turon, X. (2004). Strong population structure in the marine sponge *Crambe crambe* (Poecilosclerida) as revealed by microsatellite markers. *Molecular Ecology*, 13, 511–522.

Eizaguirre, C. & Lenz, T.L. (2010). Major histocompatability complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *Journal of Fish Biology*, 77, 2023–2047.

Eizaguirre, C., Lenz, T.L., 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.

Ejsmond, M.J. & Radwan, J. (2015). Red Queen Processes Drive Positive Selection on Major Histocompatibility Complex (MHC) Genes. *PLoS Computational Biology*, 11, e1004627.

Ekblom, R., Saether, S.A., Jacobsson, P., Fiske, P., Sahlman, T. & Grahn, M. *et al.* (2007). Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology*, 16, 1439–1451.

Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, 16, 551–558.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A. & Bradshaw, W.E. *et al.* (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, 107, 16196–16200.

Evans, M.L. & Neff, B.D. (2009). Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, 18, 4716–4729.

Evans, M.L., Neff, B.D. & Heath, D.D. (2010). MHC genetic structure and divergence across populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Heredity*, 104, 449–459.

Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M. & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, 193, 929–941.

Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marquès-Bonet, T., Ramírez-Soriano, A. & Muntasell, A. *et al.* (2008). Balancing selection is the main force shaping the evolution of innate immunity genes. *The Journal of Immunology*, 181, 1315–1322.

Figuerola, J., Green, A.J. & Michot, T.C. (2005). Invertebrate eggs can fly: evidence of waterfowl-mediated gene flow in aquatic invertebrates. *The American Naturalist*, 165, 274–280.

Foll, M. & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977–993.

Ford, S.A., Williams, D., Paterson, S. & King, K.C. (2017). Co-evolutionary dynamics between a defensive microbe and a pathogen driven by fluctuating selection. *Molecular Ecology*, 26, 1778–1789.

Fornůsková, A., Bryja, J., Vinkler, M., Macholán, M. & Piálek, J. (2014). Contrasting patterns of polymorphism and selection in bacterial-sensing toll-like receptor 4 in two house mouse subspecies. *Ecology and Evolution*, 4, 2931–2944.

Froeschke, G. & Sommer, S. (2014). Role of selection versus neutral processes determining genetic variation in a small mammal along a climatic gradient in southern Africa. *Evolutionary Ecology*, 28, 1169–1190.

Fumagalli, M., Cagliani, R., Pozzoli, U., Riva, S., Comi, G.P. & Menozzi, G. *et al.* (2009). Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Research*, 19, 199–212.

Gaba, S. & Ebert, D. (2009). Time-shift experiments as a tool to study antagonistic coevolution. *Trends in Ecology and Evolution*, 24, 226–232.

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

Gilroy, D., Oosterhout, C., Komdeur, J. & Richardson, D.S. (2016). Avian β-defensin variation in bottlenecked populations: the Seychelles warbler and other congeners. *Conservation Genetics*, 17, 661–674.

Gilroy, D.L., Phillips, K.P., Richardson, D.S. & Oosterhout, C. van. (2017). Toll-like receptor variation in the bottlenecked population of the Seychelles warbler: computer simulations see the 'ghost of selection past' and quantify the 'drift debt'. *Journal of Evolutionary Biology*, 30, 1276–1287.

Goddard, M.E. & Hayes, B.J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10, 381–391.

González-Quevedo, C., Davies, R.G. & Richardson, D.S. (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.

González-Quevedo, C., Davies, R.G., Phillips, K.P., Spurgin, L.G. & Richardson, D.S. (2016). Landscape-scale variation in an anthropogenic factor shapes immune gene variation within a wild population. *Molecular Ecology*, 25, 4234–4246.

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

Gómez, P. & Buckling, A. (2011). Bacteria-phage antagonistic coevolution in soil. *Science*, 332, 106–109.

Grant, P.R. (1986). *Ecology and Evolution of Darwin's Finches*. Princeton University Press, Princeton, NJ.

Grueber, C.E., Wallis, G.P. & Jamieson, I.G. (2014). Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. *PLoS ONE*, 9, e89632.

Grueber, C.E., Wallis, G.P., King, T.M. & Jamieson, I.G. (2012). Variation at innate immunity toll-like receptor genes in a bottlenecked population of a New Zealand robin. *PLoS ONE*, 7, e45011.

Guo, B., Lu, D., Liao, W.B. & Merilä, J. (2016). Genomewide scan for adaptive differentiation along altitudinal gradient in the Andrew's toad *Bufo andrewsi*. *Molecular Ecology*, 25, 3884–3900.

Hancock, R.E. & Diamond, G. (2000). The role of cationic antimicrobial peptides in innate host defences. *Trends in Microbiology*, 8, 402–410.

Hanski, I. (1991). Single-Species Metapopulation Dynamics: Concepts, Models and Observations. *Biological Journal of the Linnean Society*, 42, 17–38.

Hartl, D.L. & Clark, A.G. (2007). *Principles of Population Genetics*. 4th edn. Sinauer Associates, Sunderland, MA.

Hawley, D.M. & Fleischer, R.C. (2012). Contrasting epidemic histories reveal pathogen-mediated balancing selection on class II MHC diversity in a wild songbird. *PLoS ONE*, 7, e30222.

He, T., Lamont, B.B., Enright, N.J., D'Agui, H.M. & Stock, W. (2018). Environmental drivers and genomic architecture of trait differentiation in fire-adapted *Banksia attenuata* ecotypes. *Journal of Integrative Plant Biology*, doi:10.1111/jipb.12697.

Hecht, B.C., Campbell, N.R., Holecek, D.E. & Narum, S.R. (2013). Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, 22, 3061–3076.

Hedrick, P.W. (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution*, 56, 1902–1908.

Hellgren, O. (2015). Allelic variation at innate immune genes (avian β-defensins), within a natural population of great tits. *Journal of Avian Biology*, 46, 113–118.

Hellgren, O. & Ekblom, R. (2010). Evolution of a cluster of innate immune genes (β-defensins) along the ancestral lines of chicken and zebra finch. *Immunome Research*, 6, 3.

Hellgren, O. & Sheldon, B.C. (2011). Locus-specific protocol for nine different innate immune genes (antimicrobial peptides: β-defensins) across passerine bird species reveals within-species coding variation and a case of trans-species polymorphisms. *Molecular Ecology Resources*, 11, 686–692.

Hellgren, O., Sheldon, B.C. & Buckling, A. (2010). *In vitro* tests of natural allelic variation of innate immune genes (avian β-defensins) reveal functional differences in microbial inhibition. *Journal of Evolutionary Biology*, 23, 2726–2730.

Hess, J.E., Campbell, N.R., Close, D.A., Docker, M.F. & Narum, S.R. (2013). Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, 22, 2898–2916.

Higgs, R., Lynn, D.J., Cahalane, S., Alaña, I., Hewage, C.M. & James, T. *et al.* (2007). Modification of chicken avian β-defensin-8 at positively selected amino acid sites enhances specific antimicrobial activity. *Immunogenetics*, 59, 573–580.

Hill, A.V., Allsopp, C.E., Kwiatkowski, D., Anstey, N.M., Twumasi, P. & Rowe, P.A. *et al.* (1991). Common west African HLA antigens are associated with protection from severe malaria. *Nature*, 352, 595–600.

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

Hollox, E.J. & Armour, J.A.L. (2008). Directional and balancing selection in human beta-defensins. *BMC Evolutionary Biology*, 8, 113.

Hu, H., Petousi, N., Glusman, G., Yu, Y., Bohlender, R. & Tashi, T. *et al.* (2017). Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLoS Genetics*, 13, e1006675.

Huang, X. & Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology*, 65, 531–551.

Humble, E., Dasmahapatra, K.K., Martinez-Barrio, A., Gregório, I., Forcada, J. & Polikeit, A.-C. *et al.* (2018). RAD sequencing and a hybrid Antarctic fur seal genome assembly reveal rapidly decaying linkage disequilibrium, global population structure and evidence for inbreeding. *G3: Genes, Genomes, Genetics*, 8, 2709–2722.

Hyma, K.E. & Fay, J.C. (2013). Mixing of vineyard and oak-tree ecotypes of *Saccharomyces cerevisiae* in North American vineyards. *Molecular Ecology*, 22, 2917–2930.

Illera, J.C., Emerson, B.C. & Richardson, D.S. (2007). Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology*, 16, 4599–4612.

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

Janeway, C.A. & Medzhitov, R. (2002). Innate immune recognition. *Annual Review of Immunology*, 20, 197–216.

Jensen, L.F., Hansen, M.M., Mensberg, K.-L.D. & 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.

Jepson, A., Banya, W., Sisay-joof, F., Hassan-King, M., Nunes, C. & Bennett, S. *et al.* (1997). Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens. *Infection and Immunity*, 65, 872–876.

Johnston, S.E., Orell, P., Pritchard, V.L., Kent, M.P., Lien, S. & Niemelä, E. *et al.* (2014). Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, 23, 3452–3468.

Jones, E.A., Cheng, Y., O'Meally, D. & Belov, K. (2017). Characterization of the antimicrobial peptide family defensins in the Tasmanian devil (*Sarcophilus harrisii*), koala (*Phascolarctos cinereus*), and tammar wallaby (*Macropus eugenii*). *Immunogenetics*, 69, 133–143.

Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E. & Johnson, J. *et al.* (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484, 55–61.

Jones, M.R., Cheviron, Z. a & Carling, M.D. (2014). Variation in positively selected major histocompatibility complex class I loci in rufous-collared sparrows (*Zonotrichia capensis*). *Immunogenetics*, 66, 693.

Jones, M.R., Cheviron, Z.A. & Carling, M.D. (2015). Spatially variable coevolution between a haemosporidian parasite and the MHC of a widely distributed passerine. *Ecology and Evolution*, 5, 1045–1060.

Jones, M.R., Forester, B.R., Teufel, A.I., Adams, R.V., Anstett, D.N. & Goodrich, B.A. *et al.* (2013). Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*, 67, 3455–68.

Kita, R. & Fraser, H.B. (2016). Local adaptation of sun-exposure-dependent gene expression regulation in human skin. *PLoS Genetics*, 12, e1006382.

Klüver, E., Schulz-Maronde, S., Scheid, S., Meyer, B., Forssmann, W.G. & Adermann, K. (2005). Structure-activity relation of human β-defensin 3: Influence of disulfide bonds and cysteine substitution on antimicrobial activity and cytotoxicity. *Biochemistry*, 44, 9804–9816.

Kohn, M.H., Murphy, W.J., Ostrander, E.A. & Wayne, R.K. (2006). Genomics and conservation genetics. *Trends in Ecology and Evolution*, 21, 629–637.

Kolbe, J.J., Leal, M., Schoener, T.W., Spiller, D.A. & Losos, J.B. (2012). Founder effects persist despite adaptive differentiation: a field experiment with lizards. *Science*, 335, 1086–1089.

Kosch, T.A., Bataille, A., Didinger, C., Eimes, J.A., Rodríguez-Brenes, S. & Ryan, M.J. *et al.* (2016). Major histocompatibility complex selection dynamics in pathogen-infected túngara frog (*Physalaemus pustulosus*) populations. *Biology Letters*, 12, 20160345.

Koskella, B. (2013). Phage-mediated selection on microbiota of a long-lived host. *Current Biology*, 23, 1256–1260.

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

Lachance, J. (2008). A fundamental relationship between genotype frequencies and fitnesses. *Genetics*, 180, 1087–1093.

Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almén, M.S. & Webster, M.T. *et al.* (2016). A beak size locus in Darwin's finches facilitates character displacement during a drought. *Science*, 352, 470–474.

Laporte, M., Pavey, S.A., Rougeux, C., Pierron, F., Lauzent, M. & Budzinski, H. *et al.* (2016). RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular Ecology*, 25, 219–237.

Lewontin, R.C. & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 74, 175–195.

Lexer, C., Wüest, R.O., Mangili, S., Heuertz, M., Stölting, K.N. & Pearman, P.B. *et al.* (2014). Genomics of the divergence continuum in an African plant biodiversity hotspot, I: Drivers of population divergence in *Restio capensis* (Restionaceae). *Molecular Ecology*, 23, 4373–4386.

Leydet, K.P., Grupstra, C.G., Coma, R., Ribes, M. & Hellberg, M.E. (2018). Host-targeted RAD-Seq reveals genetic changes in the coral *Oculina patagonica* associated with range expansion along the Spanish Mediterranean coast. *Molecular Ecology*, 27, 2529–2543.

Li, J., Shen, H., Wang, H., Zhao, M., Luo, Z. & Wu, H. (2016). Diversifying selection is the dominant factor affecting the geographical variation of MHC class II genes in the Omei tree frog. *Journal of Zoology*, 300, 197–204.

Li, W.H. (1997). *Molecular Evolution*. Sinauer Associates Incorporated, Sunderland, MA.

Lind, B.M., Friedline, C.J., Wegrzyn, J.L., Maloney, P.E., Vogler, D.R. & Neale, D.B. *et al.* (2017). Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Molecular Ecology*, 26, 3168–3185.

Liu, L., Zhao, C., Heng, H.H. & Ganz, T. (1997). The human β-defensin-1 and α-defensins are encoded by adjacent genes: Two peptide families with differing disulfide topology share a common ancestry. *Genomics*, 43, 316–320.

Liu, Q., Wu, S., Xue, M., Sandford, A.J., Wu, J. & Wang, Y. *et al.* (2016). Heterozygote advantage of the rs3794624 polymorphism in CYBA for resistance to tuberculosis in two Chinese populations. *Scientific Reports*, 6, 38213.

Liu, S., Vallejo, R.L., Palti, Y., Gao, G., Marancik, D.P. & Hernandez, A.G. *et al.* (2015). Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Frontiers in Genetics*, 6, 298.

Lowry, D.B., Hoban, S., Kelley, J.L., Lotterhos, K.E., Reed, L.K. & Antolin, M.F. *et al.* (2017). Breaking RAD: An evaluation of the utility of restriction site associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, 17, 142–152.

Lozier, J.D. (2014). Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Molecular Ecology*, 23, 788–801.

MacArthur, R.H. & Wilson, E.O. (1967). *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ.

Mackinnon, M.J., Ndila, C., Uyoga, S., Macharia, A., Snow, R.W. & Band, G. *et al.* (2016). Environmental correlation analysis for genes associated with protection against malaria. *Molecular Biology and Evolution*, 33, 1188–1204.

Marsden, C.D., Woodroffe, R., Mills, M.G.L., McNutt, J.W., Creel, S. & Groom, R. *et al.* (2012). Spatial and temporal patterns of neutral and adaptive genetic variation in the endangered African wild dog (*Lycaon pictus*). *Molecular Ecology*, 21, 1379–1393.

Martín, A. & Lorenzo, J.A. (2001). *Aves del Archipiélago Canario*. Francisco Lemus Editor, La Laguna.

Martínez, P., González, E.G., Castilho, R. & Zardoya, R. (2006). Genetic diversity and historical demography of Atlantic bigeye tuna (*Thunnus obesus*). *Molecular Phylogenetics and Evolution*, 39, 404–416.

Maxwell, A.I., Morrison, G.M. & Dorin, J.R. (2003). Rapid sequence divergence in mammalian beta-defensins by adaptive evolution. *Molecular Immunology*, 40, 413–421.

McClelland, E.E., Penn, D.J. & Potts, W.K. (2003). Major histocompatibility complex heterozygote superiority during coinfection. *Infection and Immunity*, 71, 2079–2086.

Mckinney, G.J., Larson, W.A., Seeb, L.W. & Seeb, J.E. (2017). RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry *et al*. (2016). *Molecular Ecology Resources*, 17, 356–361.

Medzhitov, R., Preston-Hurlburt, P. & Janeway, C.A. (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.

Miller, H.C., Allendorf, F.W. & Daugherty, C.H. (2010). Genetic diversity and differentiation at MHC genes in island populations of tuatara (*Sphenodon* spp.). *Molecular Ecology*, 19, 3894–3908.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A. & Johnson, E.A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17, 240–248.

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R. & Hohenlohe, P.A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22, 2841–2847.

Netea, M.G., Wijmenga, C. & O'Neill, L.A. (2012). Genetic variation in Toll-like receptors and disease susceptibility. *Nature Immunology*, 13, 535–542.

Niskanen, A.K., Kennedy, L.J., Ruokonen, M., Kojola, I., Lohi, H. & Isomursu, 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.

Oliveira, P. & Menezes, D. (2004). *Birds of the Archipelago of Madeira*. Serviço do Parque Natural de Madeira, Madeira.

Oliver, M.K., Lambin, X., Cornulier, T. & Piertney, S.B. (2009a). Spatio-temporal variation in the strength and mode of selection acting on major histocompatibility complex diversity in water vole (*Arvicola terrestris*) metapopulations. *Molecular Ecology*, 18, 80–92.

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

Olsen, J.L., Stam, W.T., Coyer, J.A., Reusch, T.B.H., Billingham, M. & Boström, C. *et al.* (2004). North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina* L. *Molecular Ecology*, 13, 1923–1941.

Osborne, M.J., Pilger, T.J., Lusk, J.D. & Turner, T.F. (2017). Spatio-temporal variation in parasite communities maintains diversity at the major histocompatibility complex class IIβ in the endangered Rio Grande silvery minnow. *Molecular Ecology*, 26, 471–489.

Padhi, A. & Verghese, B. (2007). Evidence for positive Darwinian selection on the hepcidin gene of Perciform and Pleuronectiform fishes. *Molecular Diversity*, 11, 119–130.

Palumbi, S.R. & Baker, C.S. (1994). Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology & Evolution*, 11, 426–435.

Paterson, S., Wilson, K. & Pemberton, J.M. (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population. *Proceedings of the National Academy of Sciences*, 95, 3714–3719.

Penn, D.J., Damjanovich, K. & Potts, W.K. (2002). MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences*, 2002, 11260–11264.

Phillips, K.P., Cable, J., Mohammed, R.S., Herdegen-Radwan, M., Raubic, J. & Przesmycka, K.J. *et al.* (2018). Immunogenetic novelty confers a selective advantage in host–pathogen coevolution. *Proceedings of the National Academy of Sciences*, doi: 10.1073/pnas.1708597115.

Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z. & Absher, D. *et al.* (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Research*, 19, 826–837.

Piertney, S.B. & Oliver, M.K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96, 7–21.

Piganeau, G., Gardner, M. & Eyre-Walker, A. (2004). A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution*, 21, 2319–2325.

Pinho, C., Harris, D.J. & Ferrand, N. (2007). Contrasting patterns of population subdivision and historical demography in three western Mediterranean lizard species inferred from mitochondrial DNA variation. *Molecular Ecology*, 16, 1191–1205.

Price, A.L., Spencer, C.C.A. & Donnelly, P. (2015). Progress and promise in understanding the genetic basis of common diseases. *Proceedings of the Royal Society B: Biological Sciences*, 282, 20151684.

Pujolar, J.M., Jacobsen, M.W., Als, T.D., Frydenberg, J., Munch, K. & Jónsson, B. *et al.* (2014). Genome-wide single-generation signatures of local selection in the panmictic European eel. *Molecular Ecology*, 23, 2514–2528.

Reitzel, A.M., Herrera, S., Layden, M.J., Martindale, M.Q. & Shank, T.M. (2013). Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology*, 22, 2953–2970.

Reusch, T.B.H., Stam, W.T. & Olsen, J.L. (2000). A microsatellite-based estimation of clonal diversity and population subdivision in *Zostera marina*, a marine flowering plant. *Molecular Ecology*, 9, 127–140.

Rottschaefer, S.M., Riehle, M.M., Coulibaly, B., Sacko, M., Niaré, O. & Morlais, I. *et al.* (2011). Exceptional diversity, maintenance of polymorphism, and recent directional selection on the APL1 malaria resistance genes of *Anopheles gambiae*. *PLoS Biology*, 9, e1000600.

Savage, A.E. & Zamudio, K.R. (2011). MHC genotypes associate with resistance to a frog-killing fungus. *Proceedings of the National Academy of Sciences*, 108, 16705–16710.

Schad, J., Ganzhorn, J.U. & Sommer, S. (2005). Parasite burden and constitution of major histocompatibility complex in the Malagasy mouse lemur, *Microcebus murinus*. *Evolution*, 59, 439–450.

Schmitt, C., Garant, D., Bélisle, M. & Pelletier, F. (2017a). Linking innate immunogenetic variation with phenotypic traits in a wild population of tree swallows, *Tachycineta bicolor*. *Biological Journal of the Linnean Society*, 121, 685–697.

Schmitt, C., Garant, D., Doyon, K., Bousquet, N., Gaudreau, L. & Bélisle, M. *et al.* (2017b). Patterns of diversity and spatial variability of β-defensin innate immune genes in a declining wild population of tree swallows. *Journal of Heredity*, 108, 262–269.

Schröder, J.M. (1999). Epithelial peptide antibiotics. *Biochemical Pharmacology*, 57, 121–134.

Schwensow, N., Mazzoni, C.J., Marmesat, E., Fickel, J., Peacock, D. & Kovaliski, J. *et al.* (2017). High adaptive variability and virus-driven selection on major histocompatibility complex (MHC) genes in invasive wild rabbits in Australia. *Biological Invasions*, 19, 1255–1271.

Seger, J. (1988). Dynamics of some simple host-parasite models with more than two genotypes in each species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 319, 541–555.

Semple, C.A., Gautier, P., Taylor, K. & Dorin, J.R. (2006). The changing of the guard: molecular diversity and rapid evolution of beta-defensins. *Molecular Diversity*, 10, 575–584.

Shaw, P.W., Pierce, G.J. & Boyle, P.R. (1999). Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. *Molecular Ecology*, 8, 407–417.

Sin, Y.W., Annavi, G., Dugdale, H.L., Newman, C., Burke, T. & MacDonald, D.W. (2014). Pathogen burden, co-infection and major histocompatibility complex variability in the European badger (*Meles meles*). *Molecular Ecology*, 23, 5072–5088.

Slade, R.W. & McCallum, H.I. (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.

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

Spurgin, L.G. & Richardson, D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277, 979–988.

Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A. & Richardson, D.S. (2014). Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Molecular Ecology*, 23, 1028–1039.

Spurgin, L.G., Illera, J.C., Padilla, D.P. & Richardson, D.S. (2012). Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia*, 168, 691–701.

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J. & Ekblom, R. *et al.* (2010). Adaptation genomics: the next generation. *Trends in Ecology and Evolution*, 25, 705–712.

Stervander, M., Illera, J.C., Kvist, L., Barbosa, P., Keehnen, N.P. & Pruisscher, P. *et al.* (2015). Disentangling the complex evolutionary history of the Western Palearctic blue tits (*Cyanistes* spp.) – phylogenomic analyses suggest radiation by multiple colonization events and subsequent isolation. *Molecular Ecology*, 24, 2477–2494.

Sud, A., Kinnersley, B. & Houlston, R.S. (2017). Genome-wide association studies of cancer: current insights and future perspectives. *Nature Reviews Cancer*, 17, 692–704.

Tang, K.Y., Wang, X., Wan, Q.H. & Fang, S.G. (2018). A crucial role of paralogous β-defensin genes in the Chinese alligator innate immune system revealed by the first determination of a Crocodilia defensin cluster. *Developmental and Comparative Immunology*, 81, 193–203.

Tang, Y.Q., Yuan, J., Österapay, G., Österapay, K., Tran, D. & Miller, C.J. *et al.* (1999). A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated α-defensins. *Science*, 286, 498–502.

Tennessen, J.A. (2005). Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. *Journal of Evolutionary Biology*, 18, 1387–1394.

Tennessen, J.A. & Blouin, M.S. (2007). Selection for antimicrobial peptide diversity in frogs leads to gene duplication and low allelic variation. *Journal of Molecular Evolution*, 65, 605–615.

Tennessen, J.A. & Blouin, M.S. (2008). Balancing selection at a frog antimicrobial peptide locus: fluctuating immune effector alleles? *Molecular Biology and Evolution*, 25, 2669–2680.

Theodorou, P., Radzevičiūtė, R., Kahnt, B., Soro, A., Grosse, I. & Paxton, R.J. (2018). Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (*Bombus lapidarius* L.). *Proceedings of the Royal Society B: Biological Sciences*, 285, 20172806.

Thrall, P.H., Laine, A.L., Ravensdale, M., Nemri, A., Dodds, P.N. & Barrett, L.G. *et al.* (2012). Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. *Ecology Letters*, 15, 425–435.

Tian, D., Araki, H., Stahl, E., Bergelson, J. & Kreitman, M. (2002). Signature of balancing selection in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 99, 11525–11530.

Tiffin, P., Hacker, R. & Gaut, B.S. (2004). Population genetic evidence for rapid changes in intraspecific diversity and allelic cycling of a specialist defense gene in Zea. *Genetics*, 168, 425–434.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T. & Hayes, E. *et al.* (2003). Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, 9, 928–35.

Trucchi, E., Facon, B., Gratton, P., Mori, E., Stenseth, N.C. & Jentoft, S. (2016). Long live the alien: is high genetic diversity a pivotal aspect of crested porcupine (*Hystrix cristata*) long-lasting and successful invasion? *Molecular Ecology*, 25, 3527–3539.

Tsaousis, a.D., Martin, D.P., Ladoukakis, E.D., Posada, D. & Zouros, E. (2005). Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution*, 22, 925–933.

Tschirren, B., Andersson, M., Scherman, K., Westerdahl, H. & Råberg, L. (2011a). 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., Råberg, L. & Westerdahl, H. (2011b). 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.

Vallo, P., Benda, P., Červený, J. & Koubek, P. (2013). Conflicting mitochondrial and nuclear paraphyly small-sized West African house bats (Vespertilionidae). *Zoologica Scripta*, 42, 1–12.

Van Dijk, A., Veldhuizen, E.J.A. & Haagsman, H.P. (2008). Avian defensins. *Veterinary Immunology and Immunopathology*, 124, 1–18.

Velová, H., Gutowska-Ding, M.W., Burt, D.W. & Vinkler, M. (2018). Toll-like receptor evolution in birds: gene duplication, pseudogenisation and diversifying selection. *Molecular Biology and Evolution*, 35, 2170–2184.

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

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I. & Brown, M.A. *et al.* (2017). 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*, 101, 5–22.

Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, 4, 0446–0458.

Wang, S., Liu, C., Wilson, A.B., Zhao, N., Li, X. & Zhu, W. *et al.* (2017). Pathogen richness and abundance predict patterns of adaptive major histocompatibility complex variation in insular amphibians. *Molecular Ecology*, 26, 4671–4685.

Wegner, K.M., Reusch, T.B.H. & Kalbe, M. (2003). Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology*, 16, 224–232.

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.

White, D.J., Wolff, J.N., Pierson, M. & Gemmell, N.J. (2008). Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, 17, 4925–4942.

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

Worley, K., Collet, J., Spurgin, L.G., Cornwallis, C., Pizzari, T. & Richardson, D.S. (2010). MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, 19, 3064–3075.

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

Yang, M., Zhang, C., Zhang, M.Z. & Zhang, S. (2017). Novel synthetic analogues of avian β-defensin-12: the role of charge, hydrophobicity, and disulfide bridges in biological functions. *BMC Microbiology*, 17, 43.

Yeaman, M.R. & Yount, N.Y. (2003). Mechanisms of antimicrobial peptide action and resistance. *Pharmacological Reviews*, 55, 27–55.

Yu, H., You, X., Li, J., Zhang, X., Zhang, S. & Jiang, S. *et al.* (2018). A genome-wide association study on growth traits in orange-spotted grouper (*Epinephelus coioides*) with RAD-seq genotyping. *Science China Life Sciences*, 61, 934–946.

Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, 415, 389–395.

Zhang, L., Wu, Q., Hu, Y., Wu, H. & Wei, F. (2014). Major histocompatibility complex alleles associated with parasite susceptibility in wild giant pandas. *Heredity*, 114, 85–93.

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, 149.

Zhivotovsky, L.A., Rosenberg, N.A. & Feldman, M.W. (2003). Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *American Journal of Human Genetics*, 72, 1171–1186.

Chapter 2. Genomic associations with bill length and disease reveal drift and selection across island bird populations

Catching pipits on Porto Santo. Photo by Philip Lamb

2.1. Abstract

Island species provide excellent models for investigating how selection and drift operate in wild populations, and for determining how these processes act to influence local adaptation and speciation. Here, we examine the role of selection and drift in shaping genomic and phenotypic variation across recently separated populations of Berthelot's pipit (*Anthus berthelotii*), a passerine bird endemic to three archipelagos in the Atlantic. We first characterised genetic diversity and population structuring, which supported previous inferences of a history of recent colonisations and bottlenecks. We then tested for regions of the genome associated with the ecologically important traits of malaria infection and bill length, both of which vary substantially across populations in this species. We identified a SNP associated with variation in bill length among individuals, islands and archipelagos; patterns of variation at this SNP suggest that both phenotypic and genotypic variation in bill length is largely shaped by founder effects. Malaria was associated with SNPs near/within genes involved in the immune response, but this relationship was not consistent among archipelagos, supporting the view that disease resistance is complex and rapidly evolving. Although we found little evidence for divergent selection at candidate loci for bill length and malarial infection, genome scan analyses pointed to several genes related to immunity and metabolism as having important roles in divergence and adaptation. Our findings highlight the utility and challenges involved with combining association mapping and population genetic analysis in non-equilibrium populations, to disentangle the effects of selection and drift on shaping genotypes and phenotypes.

2.2. Introduction

The process of speciation is a continuum, beginning with small genetic differences between lineages that accumulate to produce reproductive isolation (Coyne & Orr 2004). Studying the traits that diverge in the early stages of speciation will increase our understanding of the factors and forces that give rise to the huge array of species we see today. Phenotypic divergence between populations may be driven by natural selection acting upon adaptive traits, by genetic drift mediated by demographic changes, or by a combination of the two. To determine whether variation in traits across populations is adaptive, it is first necessary to understand the demographic history of the diverging populations in order to separate the relative contribution of drift. Molecular markers have been used since the 1960s to characterise genetic diversity within and among populations (Lewontin & Hubby 1966; Kreitman 1983), identify demographic patterns (Jorde *et al.* 1997; Garris *et al.* 2005), and infer adaptation (Hughes & Nei 1988; Zhang

et al. 1998), and therefore have been instrumental in the study of reproductive isolation and speciation.

Within the last decade, next-generation sequencing has greatly enhanced our ability to study evolutionary patterns and processes in wild populations (Stapley *et al.* 2010). Using techniques such as restriction site-associated DNA sequencing (RAD-seq; Miller *et al.* 2007; Baird *et al.* 2008), it is possible to sequence thousands of loci without any prior knowledge of the target species' genome (Davey *et al.* 2011). The high marker density provides enough power to identify subtle population differentiation and calculate accurate genome-wide estimates of variation (Luikart *et al.* 2003; Corander *et al.* 2013; Benestan *et al.* 2015; Fernández *et al.* 2016). It has also facilitated the identification of individual loci or regions of the genome directly under natural selection. "Top-down" approaches, such as genome-wide association studies and quantitative trait locus mapping, enable the discovery of regions of the genome associated with ecologically important traits (e.g. Jones *et al.* 2012; Küpper *et al.* 2015; Liu *et al.* 2015; Pavey *et al.* 2015; Comeault *et al.* 2016). Alternatively, using "bottom-up" (or genome-scanning) approaches, it is possible to statistically detect loci under divergent selection between populations, and to use this information to identify candidate phenotypes previously not known to be under selection (Turner *et al.* 2010; Fumagalli *et al.* 2015; Kardos *et al.* 2015; Manel *et al.* 2016). Combining trait mapping and genome scans is a particularly powerful approach, as it potentially allows us to simultaneously identify novel candidate traits under selection, determine the genetic basis of traits, and investigate how those traits have evolved over ecological and evolutionary timescales (Stinchcombe & Hoekstra 2008; Gagnaire & Gaggiotti 2016; Josephs *et al.* 2017). A number of studies have now combined top-down and bottom-up approaches to study adaptation in the wild (Nadeau *et al.* 2012; Gompert *et al.* 2013; Johnston *et al.* 2014; Brachi *et al.* 2015; Bourgeois *et al.* 2017), but few have been carried out in bottlenecked or isolated populations (Hohenlohe *et al.* 2010).

Oceanic island systems provide excellent models in which to study adaptive evolution. Populations on different islands will face a unique set of selection pressures, driven by variation in abiotic factors, or differences in ecological communities between islands that have arisen due to independent histories of colonisation and extinction events (Carlquist 1974). The sea provides a natural barrier to the migration of terrestrial fauna and flora, enabling populations to independently evolve, with limited scope for gene flow to counteract local adaptation (Grant 1986).

Here we combine association mapping and genome-scanning approaches to identify the key traits involved in adaptive evolution and population divergence in Berthelot's pipit *Anthus berthelotii*, a small passerine endemic to three volcanic island archipelagos in the Atlantic Ocean.

The recent colonisation of this species across a set of ecologically distinct islands makes it an interesting model for examining adaptive evolution in a natural setting. Within the last 2.5 million years Berthelot's pipits colonised the Canary Islands and diverged from their sister species, the tawny pipit *Anthus campestris* (Voelker 1999). Then, around 8,000 years ago, Berthelot's pipits from the Canary Islands reached the Madeiran and Selvagens islands in two independent, northward colonisations, with a subsequent absence of gene flow (Illera *et al.* 2007; Spurgin *et al.* 2014). The recent colonisation of the latter two archipelagos has resulted in clear genetic and phenotypic signatures of founder effects (Spurgin *et al.* 2014; González-Quevedo *et al.* 2015b). This nested pattern of population structure enables comparisons of genetic and phenotypic divergence at a range of spatial scales (Spurgin *et al.* 2014; González-Quevedo *et al.* 2015a), and allows us to examine how natural selection and genetic drift interact in bottlenecked populations (Spurgin *et al.* 2011).

Importantly, in Berthelot's pipits there is also inter-island variation in selection pressures and potential signatures of adaptation (reviewed in Illera *et al.* 2016). Two traits of particular interest are pathogen resistance and bill morphology. The different pipit populations have a high degree of variation in avian malaria prevalence, and this variation is consistent over time and constrained by island biogeographic features, suggesting that avian malaria may be an important agent of divergent selection in this system (Illera *et al.* 2008; Spurgin *et al.* 2012; González-Quevedo *et al.* 2014). Bill morphology also differs markedly among populations, with pipits from the Madeiran archipelago classified as a separate subspecies, *A. berthelotii madeirensis*, based on longer bill lengths (Martín & Lorenzo 2001; Oliveira & Menezes 2004). Although overall patterns of body size in Berthelot's pipits have likely been shaped by founder effects (Spurgin *et al.* 2014), the difference in bill length could be adaptive. Importantly, both pathogen resistance and bill morphology are key traits in local adaptation, divergence and speciation in other taxa (Amadon 1950; Grant 1986; Ekblom *et al.* 2007; Eizaguirre *et al.* 2012; Lenz *et al.* 2013; Lamichhaney *et al.* 2016; Bosse *et al.* 2017), and as a result there is a great deal of interest in identifying their genetic basis in wild populations. Indeed, characterising the molecular basis of these traits increases our understanding of the mechanisms driving adaptive evolution, and could contribute to improving biodiversity conservation efforts (Smith *et al.* 2009).

Our specific aims were as follows. First, we used RAD-seq data to re-examine the population history of Berthelot's pipit at greater resolution than previously achieved with microsatellites (Spurgin *et al.* 2014). Second, we use association mapping to identify candidate SNPs associated with bill length and malaria infection among the "founder" Berthelot's pipit populations in the Canary Islands. We then explore how selection and drift have interacted to shape variation at

candidate SNPs by comparing variation among founder and derived bottlenecked populations. Finally, we performed genome scans to detect regions of the genome under divergent selection between populations and examine how drift and selection shape genome diversity in bottlenecked populations.

2.3. Methods

2.3.1. Sample collection and sequencing

Berthelot's pipits were sampled on 12 islands across their geographical range (Fig. 2.1A), as reported by Illera *et al.* (2007) and Spurgin *et al.* (2012). Pipits caught on the alpine plateau of El Teide volcano (> 2,000 m above sea level) were classified as a separate population (Spurgin *et al.* 2014). Birds were caught in spring traps baited with *Tenebrio molitor* larvae, and each bird was fitted with either a unique numbered aluminium ring issued from the Spanish or Portuguese ministries, or with a coloured plastic ring. Blood samples were obtained by brachial venipuncture and stored at room temperature in 800 µl of 100% ethanol in screw-cap microcentrifuge tubes. Bill length and tarsus length were measured with callipers to the nearest 0.1 mm. The age of the birds was determined as adult or juvenile based on feather moult pattern (Cramp 1988). Further information on each bird was derived from the blood samples as detailed in Illera *et al.* (2008) and Spurgin *et al.* (2012). In brief, DNA was extracted using a salt extraction protocol described in Richardson *et al.* (2001), and molecularly sexed (Griffiths *et al.* 1998). Haemosporidian blood parasites such as *Plasmodium* were detected using a nested PCR approach (Waldenström *et al.* 2004), and PCR products from samples with infections were Sanger sequenced to characterise the strains present across Berthelot's pipit populations.

2.3.2. Berthelot's pipit genome

A draft Berthelot's pipit genome was prepared using a sample collected from Porto Santo, Madeira. This bird was selected due to its low level of heterozygosity, calculated from the RAD data (see below). The genome was used in this study to increase the number of RAD loci successfully mapped to the zebra finch genome.

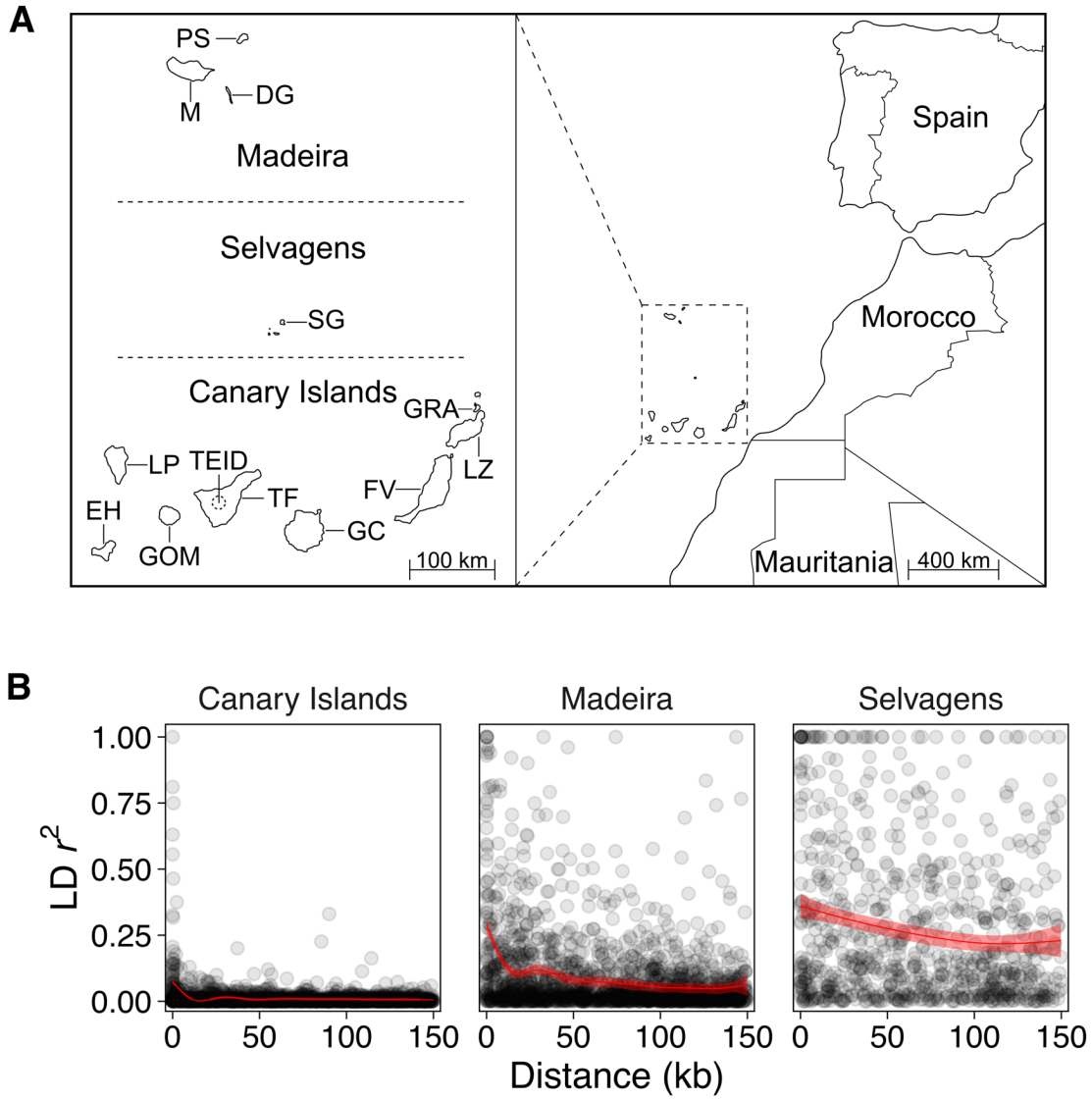


Figure 2.1 Sampling locations, genetic diversity, and linkage disequilibrium in Berthelot's and tawny pipits. A) Map of Berthelot's pipit and tawny pipit sampling locations. B) The relationship between linkage disequilibrium and base-pair distance for SNPs in the Canary Islands, Madeira and Selvagens archipelagos. The red line shows a general additive model, with the shaded band indicating 95% confidence intervals.

Sample preparation, sequencing, and genome assembly were performed at the Earlham Institute in Norwich. 10 µg of blood was extracted with the Agencourt Genfind V2 Blood & Serum Genomic DNA Isolation Kit (Beckman Coulter). Following extraction, DNA concentration was measured using the Qubit dsDNA HS Assay Kit (Life Technologies). Five hundred nanograms of genomic DNA was sheared into ~400-bp fragments using a Covaris S2 acoustic sonicator (Covaris Inc.), with the following shearing parameters: duty cycle = 10%, intensity = 5, cycles per burst = 200, time = 45 seconds. DNA fragments were then cleaned using 0.6x AMPure XP beads (Beckman Coulter) and eluted into 40 ul Tris-HCl pH 8.0. A further cleaning with 3.0x AMPure XP

beads without elution was followed by end repair, A-tailing and ligation of Illumina PCR-free TruSeq adapters (Illumina) using KAPA Library Preparation Kit reagents (Kapa Biosystems). An additional 0.7x AMPure XP bead clean up removed adapter dimers and fragments shorter than ~150 bp. The library was then eluted off the AMPure beads. The library was QC checked with the Bioanalyzer DNA HS assay system (Agilent Technologies Inc.) and quantified by both Qubit dsDNA HS Assay Kit and qPCR using a KAPA Library Quantification Kit on a StepOnePlus Real-Time PCR System (Life Technologies). Illumina sequencing was performed using paired-end sequencing (2 x 125 bp), with a 1% PhiX spike, on an Illumina HiSeq 2500 sequencer in rapid-run mode. Following sequencing, read quality was assessed with FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). The draft genome was assembled using DISCOVAR *de novo* (Weisenfeld *et al.* 2014), and assembly statistics were calculated with the "abyss-fac" utility in ABySS (Simpson *et al.* 2009). The completeness of the genome was assessed using CEGMA (Parra *et al.* 2007) and BUSCO (Simão *et al.* 2015), which searched the genome for 248 highly conserved core eukaryotic genes and 3023 vertebrate-specific single copy orthologs, respectively.

2.3.3. RAD library preparation

We chose 20 samples per population (22 from the lowland Tenerife population) for RAD-seq, with efforts made to reduce the probability of sampling relatives by maximising geographical coverage within each island, and to equalise the number of males and females where possible. In addition, 16 samples from tawny pipits were included; these were collected in Spain (*n* = 11), Mauritania (*n* = 4) and Morocco (*n* = 1; Fig. 2.1A) following the same catching, measuring and sampling protocols as Berthelot's pipit.

DNA was extracted from samples using a salt extraction protocol (Richardson *et al.* 2001), using a sufficient quantity of blood to gain a DNA concentration > 50 ng/µl. The quality of the DNA extractions was visually assessed under UV after electrophoresis on a 1.2% agarose gel, and DNA concentration was quantified with Quantifluor dsDNA dye (Promega) on a FLUOstar OPTIMA microplate reader (BMG Labtech Ltd). Library preparation was carried out at the NERC Biomolecular Analysis Facility in Sheffield, following a modified version of the ddRAD protocol by DaCosta & Sorenson (2014). Samples were allocated to one of three libraries, with populations evenly spread across the libraries to reduce bias from library-specific effects. 1 µg of DNA was digested with high fidelity SbfI and EcoRI restriction enzymes (New England Biolabs). P1 and P2 sequencing adapters were ligated onto the digested fragments. These adapters include the Illumina TruSeq amplification and sequencing primer sequences, a unique eight bp

barcode sequence, and overhangs to ligate to the sticky ends produced by SbfI and EcoRI, respectively (details of primer design in DaCosta & Sorenson 2014). The DNA concentration of each sample was then quantified on a StepOnePlus Real-Time PCR System (Applied Biosystems), and samples were pooled into 12 pools of equimolar amounts. Pools were size-selected for fragments between 300 and 450 bp by cutting from a 2% low melt agarose gel with size standards added to each pool, and purified using a MinElute Gel Extraction Kit (Qiagen). Pools were PCR amplified with Phusion High-Fidelity PCR Master Mix (Finnzymes) and purified with AMPure XP beads (Agencourt), then quantified as above. The 12 pools were then mixed in equimolar amounts into a final pool of at least 2 nM concentration, which was sent to Edinburgh Genomics for sequencing on an Illumina HiSeq2500.

2.3.4. RAD library bioinformatics

The raw RAD sequencing data was first quality-checked using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/), then processed using the ddRAD pipeline created by DaCosta & Sorenson (2014). This pipeline first demultiplexed the reads, assigning them to samples based on the 8-bp barcode sequence. Following demultiplexing, samples were grouped into five datasets, as outlined in Table 2.1. These were created to maximize the number of loci at different levels of clustering within our dataset.

The following was performed on each dataset independently. Reads were sorted by quality and clustered with UCLUST in USEARCH version 5 (Edgar 2010) at an identity of 85%. In each cluster, the sequence with the highest quality score was then mapped to the zebra finch genome (*Taeniopygia guttata* version 3.2.4; Warren *et al.* 2010) with BLASTN version 2.2.31 (Altschul *et al.* 1990), and clusters sharing a single BLAST hit were combined. The reads in each cluster were aligned in MUSCLE version 3.8.31 (Edgar 2004). Genotypes were scored using the script *RADGenotypes.py* (see DaCosta & Sorenson 2014 for full details). The "All Pipits" dataset containing all Berthelot's pipit and tawny pipit samples was then filtered to contain all SNPs from RAD loci that were successfully genotyped in 100% of individuals, removing loci that contained SNPs with > 2 alleles. Filtering for the other four datasets accepted loci that could not be confidently genotyped in ≤ 3 samples (these ambiguous genotypes were then treated as missing data in downstream analysis), and allowed up to 10% of samples with missing genotypes. Allele frequencies were calculated with the R package *adegenet* (Jombart 2008), and biallelic SNPs with a minor allele frequency ≥ 3% were retained. Further filtering resulted in datasets containing one SNP, per RAD locus, that had the highest minor allele frequency within that locus.

The draft Berthelot's pipit genome was used to improve mapping ability for loci across the four Berthelot's pipit datasets that failed to be conclusively mapped to the zebra finch genome. Where possible, the sequences for unmapped loci were taken from the same individual that was used to construct the genome, otherwise the consensus sequence was used. Unmapped sequences were then searched for in the pipit genome using BLASTN version 2.6.0 (Altschul *et al.* 1990). When a single hit to the pipit genome was obtained, the region of the genome spanning 1000 bp in each direction from the RAD locus was recorded. This extended sequence was mapped against the zebra finch genome, using the *runBLAST.py* script from the ddRAD pipeline (DaCosta & Sorenson 2014) and BLASTN version 2.3.0 (Altschul *et al.* 1990). Positions of loci successfully mapped in this way were updated.

Table 2.1 Summary of RAD-Seq datasets: sample locations, locus filtering, and analyses performed on each dataset. Read depth = median number of reads per sample per RAD locus. Filtering steps: A = loci unambiguously genotyped in 100% of samples, excluding loci with multiallelic SNPs; B = loci with ≤ 3 ambiguous genotypes, and up to 10% missing/ambiguous genotypes; C = remove SNPs with minor allele frequency of < 3%, then select one SNP per locus with the highest minor allele frequency; D = remove SNPs out of Hardy-Weinberg equilibrium in more than two populations, and SNPs mapped to the Z chromosome.

	Dataset					
	All pipits	Berthelot's	Berthelot's HWE	Canary Islands	Madeira	Selvagens
n Samples	278	262	262	182	60	20
n Loci	1826	2745	2598	3330	2983	1918
Read depth	184	198	198	187	195	176
Density of loci in zebra finch genome (%)	0.016	0.025	0.024	0.030	0.027	0.017
Sample locations						
Canary Islands	X	X	X	X		
Madeira	X	X	X		X	
Selvagens	X	X	X			X
Tawny pipits	X					
Filtering steps						
Filtering A	X					
Filtering B		X	X	X	X	X
Filtering C		X	X	X	X	X
Filtering D			X			
Analyses						
Nucleotide diversity	X					
Linkage disequilibrium				X	X	X
Admixture			X			
BSLMM/LMM		X		X		
EigenGWAS		X				

A final filtering step was applied to the "Berthelot's" dataset to make a sixth dataset, "Berthelot's HWE" (Table 2.1) that removes SNPs that could have confounding effects on estimates of population structure. First we removed SNPs located on the Z chromosome, which is evolving under different selective forces to the autosomes (Li & Merilä 2010). We then calculated *P* values for deviations from Hardy Weinberg Equilibrium (HWE) in PLINK 1.9 (Chang *et al.* 2015), and removed SNPs that were out of HWE at $P < 0.05$ in three or more populations.

2.3.5. Population genetics analysis

We first compared the overall levels of genetic diversity among the different Berthelot's pipit archipelagos and the tawny pipit, using the "All Pipits" dataset. We used the approach outlined by Lozier (2014) to calculate per-SNP nucleotide diversity (π_{SNP}) at 1,826 loci present in all samples, and the overall nucleotide diversity per locus (π_{RAD}) as the sum of the nucleotide diversity at every site along a RAD locus, including invariable sites, divided by the length of the locus (excluding loci < 50 bp long, $n = 1{,}722$). Nucleotide diversity was calculated in VCFtools version 0.1.14015 (Danecek *et al.* 2011). To account for differences in sample size between groups, nucleotide diversity for each SNP was averaged across 100 runs using subsampled datasets of 16 samples, equal to the smallest sample size (tawny pipits).

Linkage disequilibrium was calculated in PLINK 1.9 (Chang *et al.* 2015). We used the separate archipelago datasets ("Canary Islands", "Madeira" and "Selvagens", rather than the "Berthelot's" dataset, in order to maximize the number of loci available for analysis. The r^2 values obtained were compared to physical distance between loci, excluding pairs of SNPs situated on different chromosomes.

Population structure was examined using the "Berthelot's HWE" dataset. Genetic population clustering was identified with ADMIXTURE (Alexander *et al.* 2009), for $K = 1$ to $K = 13$ putative clusters. We calculated five-fold cross-validation error to determine the optimal number of clusters (Alexander & Lange 2011). PLINK 1.9 (Chang *et al.* 2015) was used to calculate the mean F_{ST} between each pair of populations, which we correlated with pairwise geographic distance using a one-tailed partial Mantel test, using the Ecodist package in R (Goslee & Urban 2007). To control for archipelago-level effects, we included "archipelago comparison" (e.g. Canary Islands vs. Canary Islands, Canary Islands vs. Madeira, etc.) as a categorical variable in the partial Mantel tests.

2.3.6. Genetic associations with traits

We used two genome-wide association study (GWAS) analyses to identify loci associated with bill length and malaria infection. A central issue with GWAS is accounting for population structure (Marchini *et al.* 2004) and we used two approaches to deal with this. First, we restricted association analyses to the "Canary Islands" dataset, in which population structure is limited (see Results). Secondly, we used kinship/relatedness matrices to statistically control for any population structure among the Canary Islands. We first used a univariate linear mixed-model (LMM) implemented in the software package GEMMA (Zhou & Stephens 2012). This model assumes that every genetic variant affects the phenotype, which would suit a highly polygenic genetic architecture, and accounts for population structure with a centred kinship matrix calculated in GEMMA.

In our second analysis, we performed association mapping with a Bayesian sparse linear mixed-model (BSLMM; Zhou *et al.* 2013), again using GEMMA and the "Canary Islands" dataset. BSLMM combines linear mixed-models with sparse regression models, giving the benefits of each when the underlying genetic architecture of the trait (many genes of small effect vs. few genes of large effect) is unknown (Zhou *et al.* 2013). Population structure is controlled for with a relatedness matrix as a covariate of the model. The output includes estimates of the hyperparameters PVE (the proportion of the phenotypic variance explained), PGE (the proportion of PVE that can be explained by SNPs with a nonzero effect on phenotypic variation), and the number of SNPs that explain the PVE. Each SNP is assigned a posterior inclusion probability (PIP), which is the proportion of times the SNP has a nonzero effect on phenotypic variation. We used a threshold of PIP > 0.1 to identify candidate SNPs associated with phenotypes (Chaves *et al.* 2016). Each analysis was run for 20 million iterations with a burn-in of 5 million. This was repeated 10 times, and the results were averaged across runs. We subsequently repeated the BSLMM analysis using the "Berthelot's" dataset, to discover whether the increase in power gained from a larger sample size improved the ability to detect loci associated with either bill length or malaria infection.

To check whether SNPs associated with bill length also covary with overall body size, we also ran a BSLMM with tarsus length as a candidate trait. Tarsus length is strongest predictor of overall body size that can be accurately measured from a live bird (Senar & Pascual 1997). Bill and tarsus length were only weakly correlated in our dataset (r = 0.13, P = 0.04).

To identify genes located near outlier SNPs, we downloaded Ensembl gene predictions for zebra finch (ftp://ftp.ensembl.org/pub/release-90/fasta/taeniopygia_guttata/pep/), and viewed

regions of interest using the University of Santa Cruz Genome Browser (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

We examined how selection and drift shaped variation at candidate SNPs, as follows. For the most significant SNPs identified in the BSLMM analyses (PIP > 0.1), we calculated allele frequencies for each island population and compared this against population-level trait variation (mean bill length and malaria prevalence), using Pearson correlations. We also calculated pairwise F_{ST} individually for BSLMM outliers, and compared outlier F_{ST} and genome-wide F_{ST}, using partial Mantel tests. To better visualise how variation at outlier SNPs was partitioned among populations, we also tested for isolation-by distance at outlier SNPs, using the same approach described above for genome-wide F_{ST}.

2.3.7. EigenGWAS analysis

We used EigenGWAS (Chen *et al.* 2016), implemented in the program GEAR (https://github.com/gc5k/GEAR/wiki), to separate loci under divergent selection from the genome-wide effects of drift between the archipelagos, using the "Berthelot's" dataset. This analysis uses eigenvector decomposition on the genotype data to characterise the underlying population genetic structuring, without the need to define discrete populations. This then enables the identification of SNPs under selection across gradients of population structure. EigenGWAS provides adjusted *P* values that use the genomic inflation factor λ_{GC} (Devlin & Roeder 1999) to correct for population stratification (i.e. drift), to avoid interpreting ancestry-informative markers as loci under selection. Genes in the zebra finch genome located near to outlier SNPs were identified as above. We then calculated F_{ST} in PLINK 1.9 (Chang *et al.* 2015), grouping samples according to eigenvector clustering, to examine the relationship between EigenGWAS significance and genetic divergence.

We found an apparent overlap between the most significant SNP identified in the BSLMM for bill length, and an outlier SNP identified in the EigenGWAS analysis of EV1 (see Results). To test whether this overlap was significant, we used a randomisation approach. We created 1000 randomised datasets, in which we randomly selected a SNP, and we calculated the distance between the randomly selected SNP and the closest, genome-wide significant EigenGWAS SNP. We repeated this process 1000 times, and therefore generated a distribution of distances. We compared the distance of our significant BSLMM SNP and the closest EigenGWAS SNP against this distribution of distances in order to generate a one-tailed *P* value.

2.4. Results

2.4.1. RAD libraries and genome summary

The three RAD libraries produced 277,627,890 (first), 202,466,438 (second), and 184,660,891 (third) raw reads with GC contents of 52%, 51% and 50%, respectively. Following filtering, we produced six datasets containing between 1,826 and 3,330 polymorphic RAD loci per dataset, at a median read depth of between 176 and 198 reads per sample per locus (Table 2.1).

We created a draft Berthelot's pipit genome which we used to improve our ability to map RAD loci to the zebra finch genome. The pipit genome sequencing run produced 143,415,538 paired reads, which were assembled in DISCOVAR *de novo* (Weisenfeld *et al.* 2014) to create a draft genome of 1,153,192,274 bp, comprising 350,587 contigs, with a contig N50 of 355,835 bp. The completeness of the genome, assessed through CEGMA (Parra *et al.* 2007) and BUSCO (Simão *et al.* 2015), is given in Supplementary Table S2.1.

2.4.2. Genetic diversity and population structure

We used two measures of nucleotide diversity, π_{SNP} and π_{RAD}, to compare genetic diversity between the three separate Berthelot's pipit archipelagos and tawny pipits (Table 2.2, Supplementary Fig. S2.1). In line with expectations based on population sizes, tawny pipits had the highest levels of diversity (π_{SNP} = 0.104, π_{RAD} = 0.0049), followed by Berthelot's pipit populations in the Canary Islands (π_{SNP} = 0.012, π_{RAD} = 0.0005), Madeira (π_{SNP} = 0.008, π_{RAD} = 0.0004), and Selvagens (π_{SNP} = 0.006, π_{RAD} = 0.0003). There was moderate divergence between Deserta Grande and the rest of the Madeiran archipelago (see below), however, exclusion of Deserta Grande had little impact on measures of diversity (Table 2.2).

This trend was reflected throughout Berthelot's pipit archipelagos for linkage disequilibrium (LD; Fig. 2.1B). Baseline levels of LD were lowest in the Canary Islands and highest in Selvagens, with the rate of decay sharpest in the Canary Islands and shallowest in Selvagens.

Population admixture analysis (Fig. 2.2A) first separated Madeira from the Canary Islands and Selvagens (K = 2) and then into the three archipelagos (K = 3). Further population structuring beyond the archipelago level was seen with Deserta Grande diverging from the other islands in the Madeiran archipelago (K = 4), and an east–west gradient in admixture across the Canary Islands (K = 5; Fig. 2.2A). The most likely number of clusters determined by cross-validation error was four (Fig. 2.2B), however similar cross-validation errors were found for K = 3 (0.434), K = 4 (0.431) and K = 5 (0.433). ADMIXTURE plots for K = 7 to K = 13 are shown in Supplementary Fig.

S2.2. We found a moderate signal of isolation by distance within the Canary Island archipelago (Mantel test, r = 0.37, P = 0.002; Supplementary Fig. S2.3), although levels of structure within the archipelago were generally low (F_{ST} < 0.03).

Table 2.2 Nucleotide diversity across groups of Berthelot's pipits and tawny pipits. Mean and bootstrapped 95% confidence intervals for π_{SNP} (per-SNP heterozygosity) and π_{RAD} (nucleotide diversity per RAD locus). In addition to archipelago-level grouping in Berthelot's pipit, this analysis was repeated in the Madeiran archipelago with Deserta Grande (DG) analysed separately, in line with population structuring identified in ADMIXTURE.

Dataset	Mean π_{SNP}	95% CI π_{SNP}	Mean π_{RAD}	95% CI π_{RAD}
Tawny pipits	0.1043	0.1019-0.1066	0.00488	0.00466-0.00510
Canary Islands	0.0118	0.0107-0.0128	0.00053	0.00047-0.00059
Madeira	0.0079	0.0069-0.0090	0.00036	0.00030-0.00043
Madeira excl. DG	0.0077	0.0066-0.0087	0.00035	0.00029-0.00041
DG	0.0078	0.0067-0.0089	0.00036	0.00028-0.00042
Selvagens	0.0058	0.0048-0.0067	0.00027	0.00022-0.00032

2.4.3. Association studies

Both LMMs and BSLMMs were used to find regions of the genome associated with malaria and bill length in the "Canary Islands" dataset, while accounting for population structure. The LMM analysis failed to find any SNPs significantly associated with either malaria or bill length when using a Bonferroni-corrected significance threshold of $P < 1.5 \times 10^{-5}$; however, Bonferroni correction is often overly conservative, inflating the risk of type II errors (Johnson *et al.* 2010). There was strong concordance between LMM P values and BSLMM PIPs for malaria (Pearson's correlation between -log LMM P values and log BSLMM PIPs, r = 0.91, P < 0.0001) and bill length (r = 0.91, P < 0.0001).

In the BSLMM analysis for malaria, a median of 68.4% of phenotypic variation was explained by the genotype (95% CI 11.2–99.8%), of which 22.5% was explained by SNPs with non-zero effects, but the credible intervals on this estimate were very high (95% CI 0.0–91.3%). Five SNPs were found with a PIP > 0.1, after controlling for population structure (Fig. 2.3A). Of these outlier SNPs, the top two also had the smallest P values in the LMM (Table 2.3). The strongest association was found for a SNP on chromosome 10 (5239s1), approximately 2,000 bp from the *IL-16* gene. The next strongest association was with a SNP on chromosome 20 (7259s1) situated within an intronic region of *RIMS4*. Both of these SNPs had "A" and "T" alleles. The "AA" and "TT" homozygotes were associated with the highest and lowest levels of malaria, respectively,

while heterozygotes had intermediate malaria prevalence (Fig. 2.3B). We then examined how population-level allele frequencies at these SNPs were related to population-level prevalence of malaria. Within the Canary Islands, there was a clear correlation between malaria prevalence and allele frequency for SNP 5239s1 (r = -0.76, P = 0.02); however, this pattern was not observed when the populations from the bottlenecked northern archipelagos were included (r = -0.35, P = 0.25; Fig. 2.3C). Despite showing an individual-level association between malaria prevalence and genotype (Fig. 2.3B), the population-level allele frequency at SNP 7259s1 was not significantly correlated with malaria infection in either the Canary Islands (r = -0.59; P = 0.09) or across all populations (r = -0.50; P = 0.09; Fig. 2.3C). Estimates of F_{ST} revealed that patterns of structure at both of these SNPs was correlated with genome-wide F_{ST} across all Berthelot's pipit populations (partial Mantel tests, 5239s1: r = 0.38, P = 0.0070; 7259s1: r = 0.51; P = 0.0060; Fig. 2.4A). Outlier F_{ST} values were correlated with pairwise geographic distance (5239s1: r = 0.30, P = 0.00800; 7259s1: r = 0.36; P = 0.01000; Fig. 2.4B).

The BSLMM for bill length found that genotype explained a median of 67.9% of phenotypic variation (95% CI 14.0–99.7%), with 14.3% explained by SNPs with non-zero effects (although again the latter had wide credible intervals; 0.0–85.6%). The strongest association with bill length was for a SNP on chromosome 5 (Fig. 2.5A; Table 2.3). At this SNP (8914s1), individual bill length decreases with increasing number of copies of the "G" allele, and heterozygotes had intermediate bill length (Fig. 2.5B). This region was not associated with tarsus length (Supplementary Fig. S2.4A), and there were no SNPs that showed strong associations for both bill length and tarsus length (Supplementary Fig. S2.4B). At the population level, mean bill length decreases with increasing frequency of the "G" allele at the most significant BSLMM SNP, both in the Canary Islands (r = -0.69, P = 0.04), and across all Berthelot's pipit populations (r = -0.82, P = 0.0006; Fig. 2.5C). After controlling for archipelago-level effects, patterns of F_{ST} at this SNP were strongly correlated with genome-wide F_{ST} (r = 0.69, P = 0.0002; Fig. 2.4A), and with geographic distance (r = 0.48, P = 0.00080; Fig. 2.4B).

The BSLMM analyses outlined in the Methods were repeated on the "Berthelot's" dataset, to test for additional power in detecting SNPs associated with bill length and malaria infection with a larger sample size. For bill length, a median of 77.6% of phenotypic variation was explained by the genotype (95% CI 55.3–98.2%), of which 14.8% was explained by SNPs of non-zero effects (95% CI 0.0–91.4%). Associations with malaria infection found a median of 84.0% of variation in phenotype explained by genotype (95% CI 43.2–99.9%), with 23.8% of this explained by SNPs of non-zero effects (95% CI 0.0–84.8%). A number of additional outlier SNPs were identified (Table 2.4, Fig. 2.6).

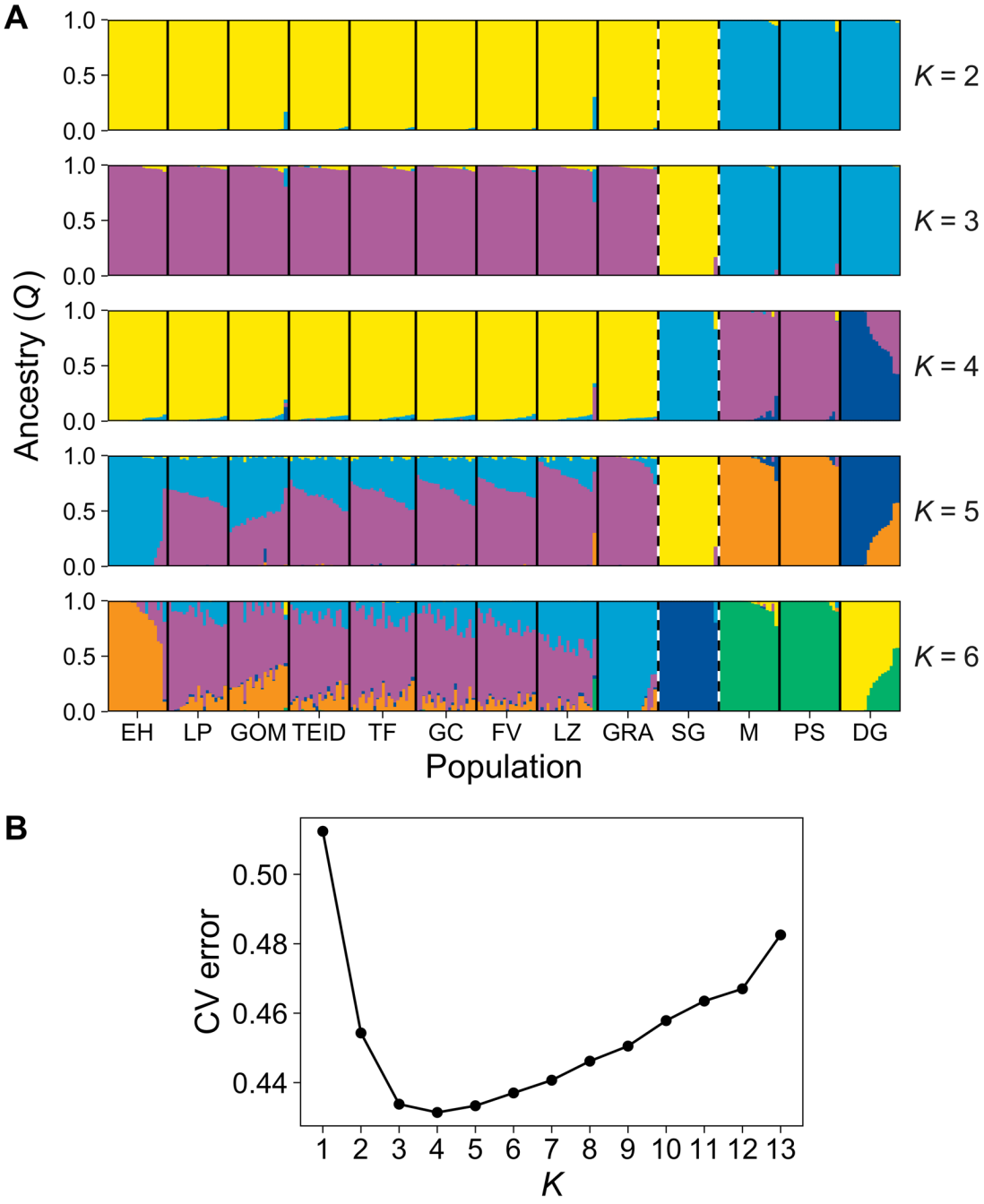


Figure 2.2 Admixture analysis across populations of Berthelot's pipits. A) Admixture for *K* = 2 to *K* = 6 clusters. Vertical bars represent individual pipits, and are coloured by their assignment to each one of *K* clusters. Solid black lines indicate separate populations; dashed black lines indicate separate archipelagos. Populations from left to right: Canary Islands – El Hierro (EH), La Palma (LP), Gomera (GOM), Teide (TEID), Tenerife (TF), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ), Graciosa (GRA); Selvagens – Selvagem Grande (SG); Madeira – Madeira (M), Porto Santo (PS), Deserta Grande (DG). B) Cross-validation (CV) error for *K* = 1 to *K* = 13 clusters, calculated by Admixture analysis of Berthelot's pipit populations.

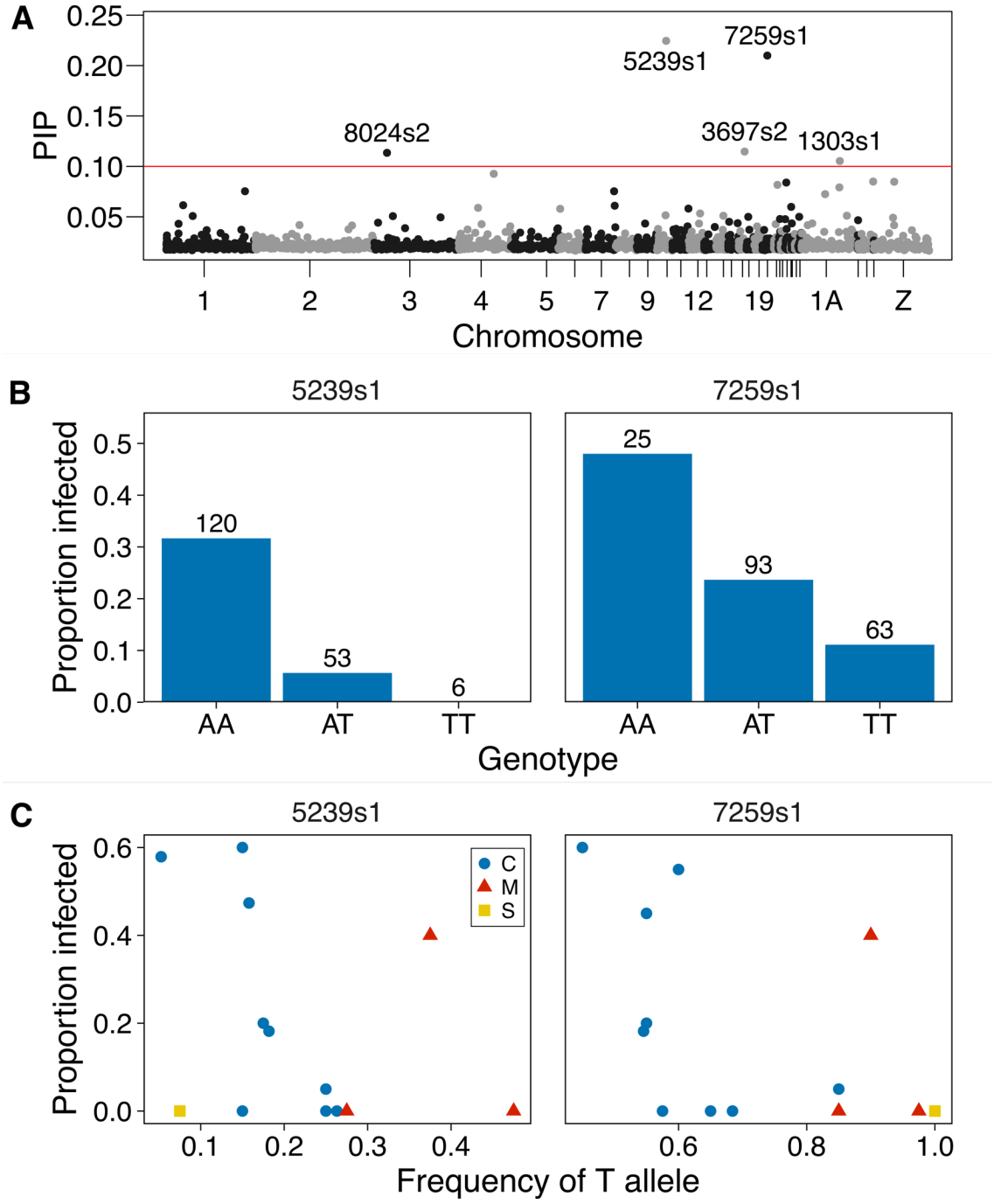


Figure 2.3 Genetic associations with malaria infection. A) Manhattan plot of BSLMM analysis of malaria infection in the Canary Islands. Red line indicates the threshold of posterior inclusion probability (PIP) = 0.1. For the top two outlier SNPs: B) the association between genotype and the proportion of individuals infected with malaria in the Canary Islands. Numbers above bars indicate the number of individuals with each genotype; and C) per-population minor allele frequency and the proportion of infected individuals, coloured by archipelago (C = Canary Islands, M = Madeira, S = Selvagens).

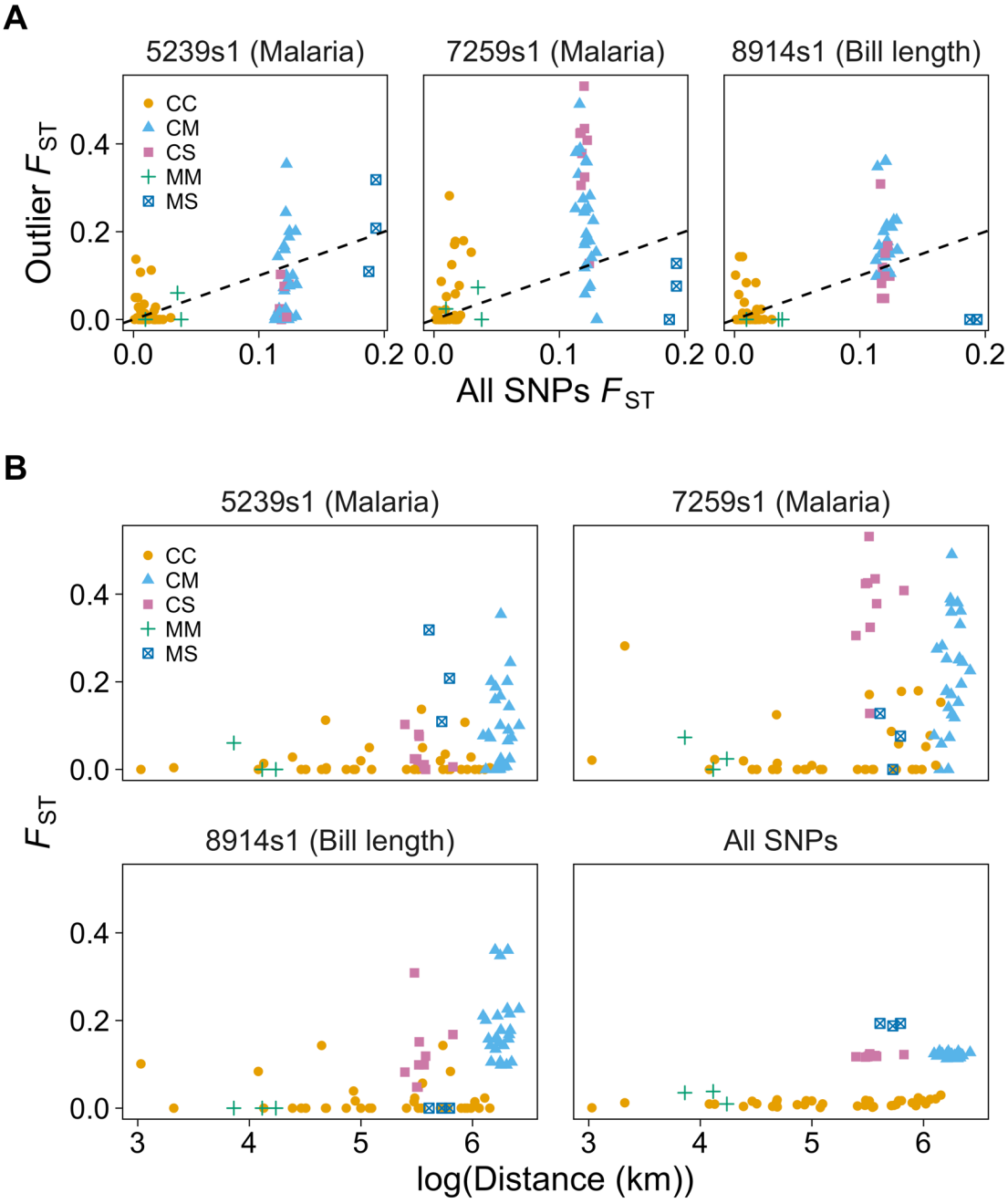


Figure 2.4 The relationship between BSLMM outlier SNP pairwise F_{ST} and A) pairwise genome-wide F_{ST}; and B) pairwise geographic distance between all pairs of Berthelot's pipit populations. Point colours and symbols represent pairwise comparisons between the different archipelagos, eg. CC = Canary Islands vs. Canary Islands, CM = Canary Islands vs. Madeira, etc. Dashed line in A) indicates a 1:1 relationship.

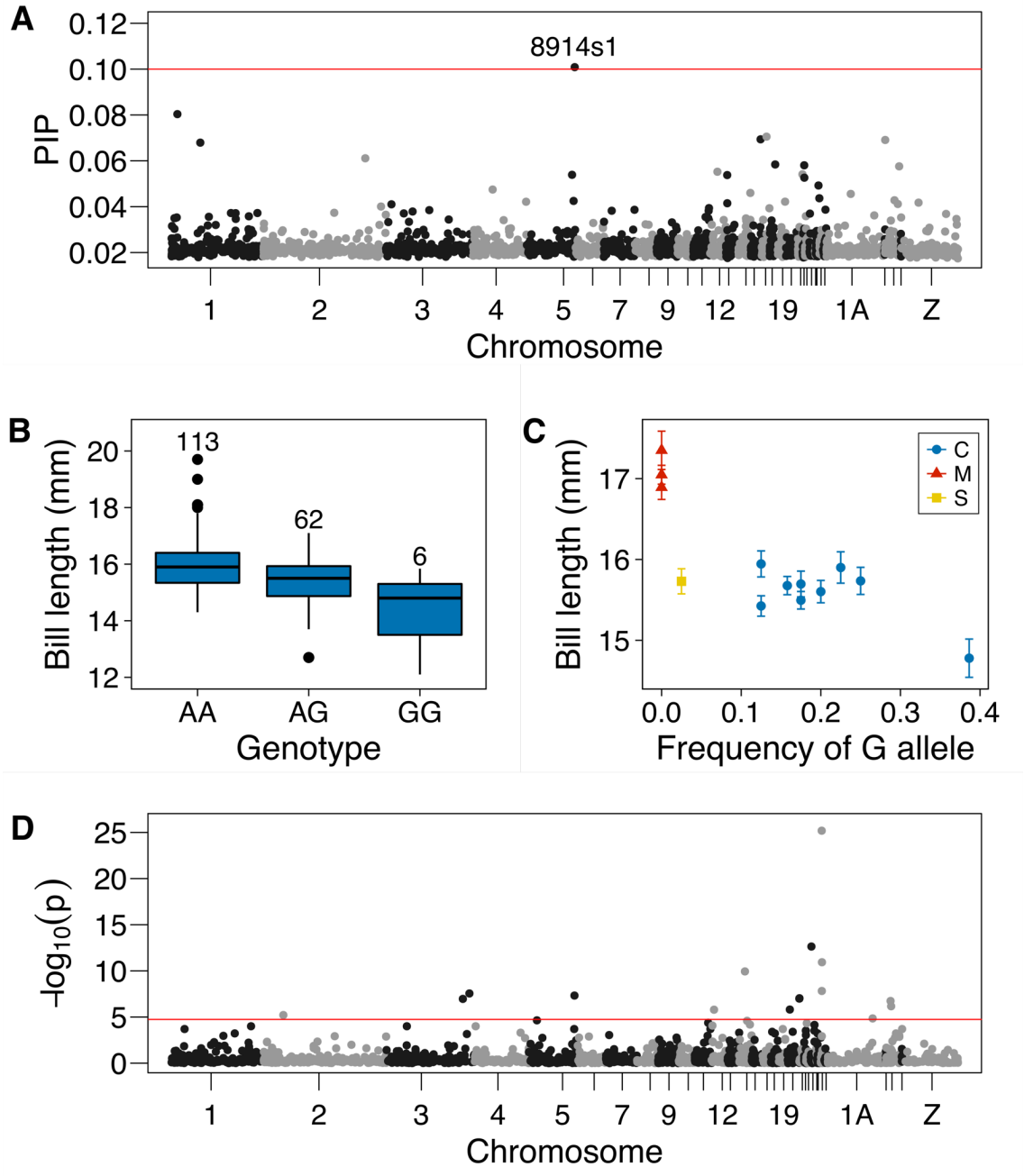


Figure 2.5 Genetic associations with bill length and selection across archipelagos. A) Manhattan plot of BSLMM analysis of bill length in the Canary Islands. Red line indicates the threshold of posterior inclusion probability (PIP) = 0.1. For the top outlier SNP: B) the association between genotype and bill length in the Canary Islands. Numbers above box plots indicate the number of individuals with each genotype; and C) per-population minor allele frequency and mean bill length, coloured by archipelago (C = Canary Islands, M = Madeira, S = Selvagens). D) Manhattan plot of EigenGWAS eigenvector 1. Red line indicates the Bonferroni-corrected *P* value threshold.

Table 2.3 Outlier SNPs with PIP > 0.1 in BSLMM analyses for bill length and malaria. The *P* value from the LMM analyses is shown. The nearest gene within 10,000 bp of the SNP is identified.

Phenotype	SNP	Locus ID	PIP	LMM *P*	Position	Gene	Distance (bp)
Bill length	8914s1	4494	0.10	0.0005	Chr5:60950937	-	-
Malaria	5239s1	2739	0.22	0.0001	Chr10:12048280	*IL-16*	2188
	7259s1	3870	0.21	0.0002	Chr20:6483195	*RIMS4*	In gene
	3697s2	4893	0.11	0.002	Chr17:8691633	-	-
	8024s2	4199	0.11	0.002	Chr3:18288583	*CNIH4*	1964
	1303s1	632	0.11	0.0008	Chr1A:50095389	*CACNA1I*	In gene

Table 2.4 Outlier SNPs with PIP > 0.1 in BSLMM analyses for bill length and malaria in the "Berthelot's" dataset. The *P* value from the LMM analyses is shown. The nearest gene within 10,000 bp of the SNP is identified.

Phenotype	SNP	Locus ID	PIP	LMM *P*	Position	Gene	Distance (bp)
Bill length	3916s2	1990	0.40	0.0001	Chr28:3343405	-	-
	176s3	86	0.24	0.0007	Chr21:5714527	*PLEKHN1*	4828
	4903s3	2562	0.17	0.001	Chr15:1987982	-	-
	1350s3	656	0.17	0.002	Chr23:1959411	-	-
	6668s3	3542	0.13	0.003	Chr7:38943610	-	-
	3384s3	1679	0.13	0.0005	-	-	-
	8914s1	4494	0.12	0.001	Chr5:60950937	-	-
	4605s3	2390	0.12	0.003	Chr4A:18361942	-	-
	8588s1	4410	0.11	0.006	Chr3:35547677	*FEZ2*	In gene
	6170s2	3269	0.10	0.005	Chr2:58237321	-	-
Malaria	7259s1	3870	0.49	0.00006	Chr20:6483195	*RIMS4*	In gene
	1303s1	632	0.34	0.0002	Chr1A:50095389	*CACNA1I*	In gene
	370s2	194	0.21	0.002	Chr11:20923524	*CDH11*	In gene
	207s3	108	0.19	0.002	Chr24:2861103	*TBCEL*	In gene
	3431s2	1710	0.18	0.004	Chr21:4658523	-	-
	13s1	5	0.18	0.0008	Chr1:22008821	*NLGN4X*	In gene
	5239s1	2739	0.17	0.002	Chr10:12048280	*IL-16*	2188
	5812s3	3077	0.15	0.001	Chr15:1405244	-	-
	3159s1	1563	0.14	0.002	Chr26:88595	*ALX3*	6137
	8626s5	4425	0.13	0.007	Chr7:37287107	-	-
	3087s3	1531	0.12	0.01	ChrZ:49829841	*KIF2A*	In gene
	8222s3	4253	0.12	0.02	Chr19:10536989	*MSI2*	In gene
	3054s3	1514	0.11	0.01	-	-	-
	3135s1	1548	0.11	0.007	Chr4:62637967	*ZFYVE28*	In gene
	565s1	283	0.10	0.007	Chr5:58806769	*ATG14*	In gene
	2824s1	1378	0.10	0.003	Chr4:47092989	*SLC30A9*	In gene
	8827s1	4480	0.10	0.01	-	-	-
	6877s4	3657	0.10	0.03	Chr27:173314	*UBTF*	4966

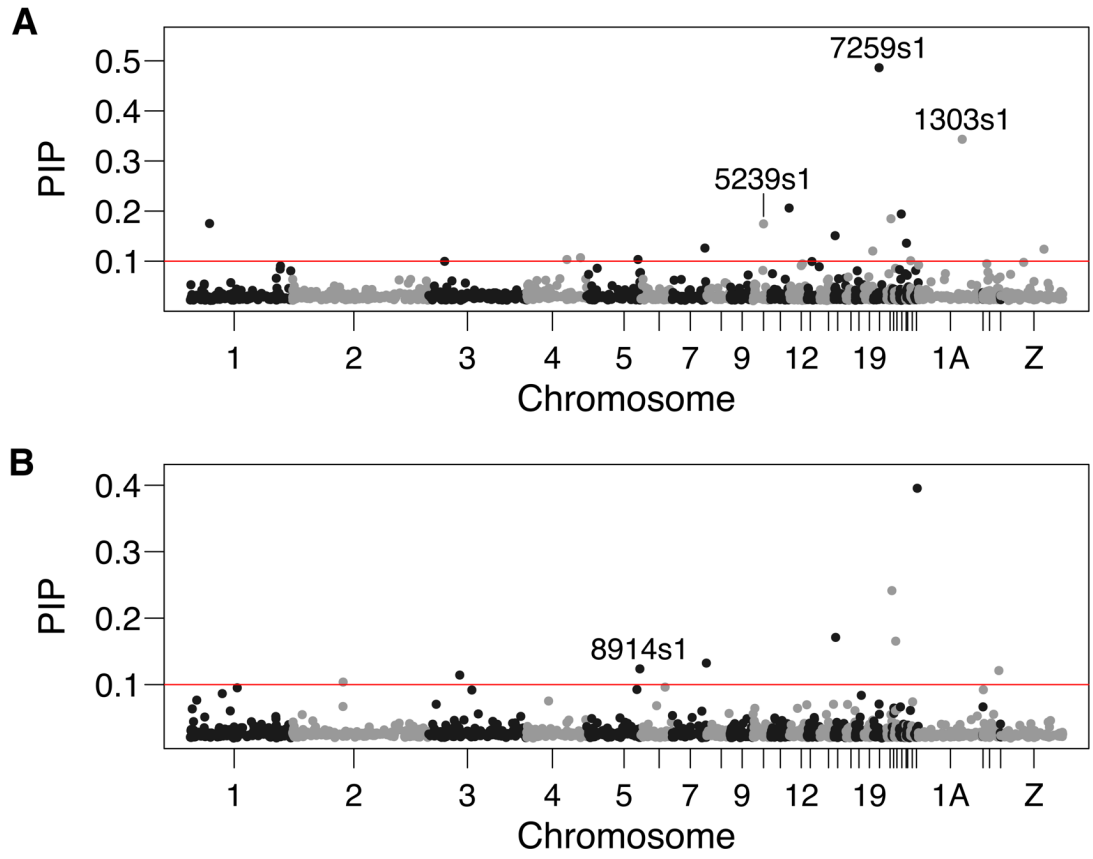


Figure 2.6 Genetic associations with malaria and bill length in the "Berthelot's" dataset. Manhattan plots of BSLMM analyses for A) malaria infection and B) bill length across all Berthelot's pipit populations. Red line indicates the threshold of posterior inclusion probability (PIP) = 0.1.

2.4.4. Differential selection across archipelagos

We used EigenGWAS to detect loci putatively under divergent selection between the archipelagos in the "Berthelot's" dataset. The first eigenvector (EV1) separated the Madeiran archipelago from the Canary Islands and Selvagens, and the second (EV2) separated Selvagens from Madeira and the Canary Islands (Fig. 2.7A). The first two eigenvectors captured a large proportion of the genetic variation, with eigenvalues of 31.1 and 14.6, respectively. Both EV1 and EV2 had considerable genomic inflation factors (26.5 and 9.7, respectively), suggesting high levels of population structuring (a genomic inflation factor ≤ 1 indicates no population structuring; Hinrichs *et al.* 2009), so we used the adjusted *P* values calculated by EigenGWAS that use genomic control to account for population structure.

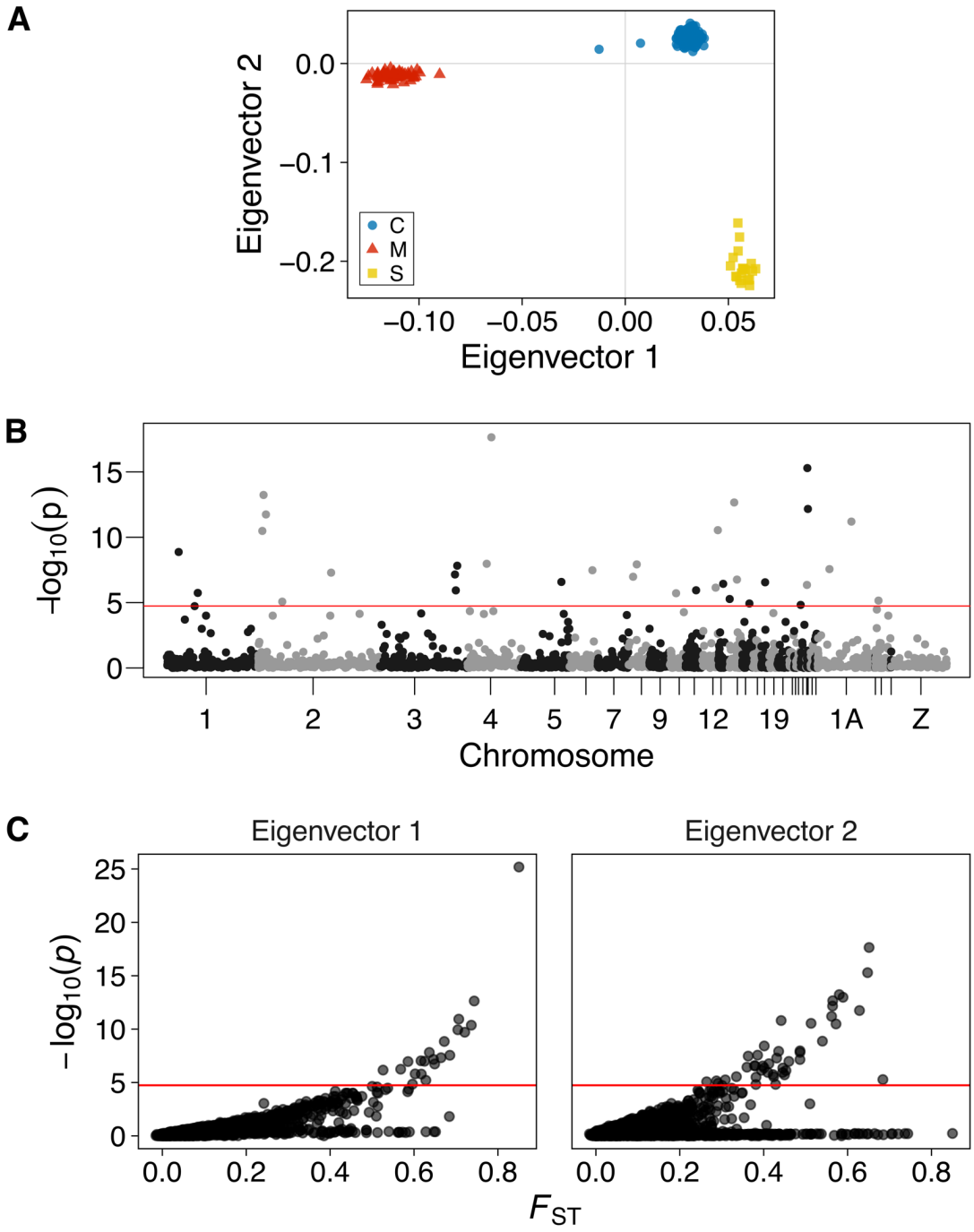


Figure 2.7 Selection across archipelagos. A) EigenGWAS eigenvalues along eigenvectors (EV) 1 and 2. Points represent individual samples, coloured by archipelago. B) Manhattan plot of EigenGWAS EV2. Red line indicates the Bonferroni-corrected *P* value threshold. C) The relationship between EigenGWAS *P* value and F_{ST}. In the left-hand plot, per-SNP F_{ST} was calculated by grouping individuals according to clustering along Eigenvector 1 (Canary Islands and Selvagens vs. Madeira), and plotted against each SNP's respective EigenGWAS *P* values. Likewise, the right-hand plot groups samples in line with Eigenvector 2 clustering: Canary Islands and Madeira vs. Selvagens.

Table 2.5 Outlier SNPs from Eigenvector 1 of EigenGWAS analysis. The nearest gene within 10,000 bp of the SNP is identified.

SNP	Locus ID	*P*	Position	Gene	Distance (bp)
203s1	105	6×10^{-26}	Chr27:1229392	*MAP3K14*	In gene
7826s2	4142	2×10^{-13}	Chr24:1583716	*DLAT*	In gene
7622s1	4065	1×10^{-11}	Chr27:1534267	*GPATCH8*	965
5483s1	2876	4×10^{-11}	-	-	-
8418s1	4348	1×10^{-10}	Chr14:7479033	*SNX29*	In gene
2982s2	1469	2×10^{-10}	-	-	-
2041s1	997	1×10^{-9}	-	-	-
174s1	85	2×10^{-8}	Chr27:1376142	*MYO1D*	In gene
5752s1	3039	3×10^{-8}	Chr3:105320550	-	-
8322s2	4296	5×10^{-8}	Chr5:58018667	*AP5M1*	2868
7216s3	3844	7×10^{-8}	-	-	-
4131s1	2115	9×10^{-8}	Chr20:15447166	*PREX1*	In gene
2473s1	1217	1×10^{-7}	Chr20:15293405	*CSE1L*	In gene
6932s2	3689	1×10^{-7}	Chr3:96985953	-	-
7180s1	3825	2×10^{-7}	Chr4A:5912527	-	-
1728s2	849	6×10^{-7}	-	-	-
6765s1	3602	7×10^{-7}	Chr4A:6655419	*VSIG4*	In gene
4138s2	2120	2×10^{-6}	Chr20:3241763	*GSS*	4
1723s1	846	2×10^{-6}	Chr12:3914968	-	-
1541s1	760	6×10^{-6}	Chr2:25304006	*VPS50*	In gene
2273s1	1117	1×10^{-5}	Chr1A:56816580	-	-

Using a Bonferroni-corrected $P < 1.8 \times 10^{-5}$, we detected signatures of selection in EV1 at 21 SNPs, 16 of which could be mapped to the zebra finch genome (Fig. 2.5D; Table 2.5). A larger number of outliers was found along EV2, with 44 outlier SNPs (33 mapped) between Selvagens and the other archipelagos (Fig. 2.7B; Table 2.6). A comparison of EigenGWAS *P* values and F_{ST} between eigenvector clusters shows that highly significant EigenGWAS SNPs always had high F_{ST}, whereas a substantial number of SNPs with high F_{ST} were nonsignificant in the EigenGWAS analysis (Fig. 2.7C).

Table 2.6 Outlier SNPs from Eigenvector 2 of EigenGWAS analysis. The nearest gene within 10,000 bp of the SNP is identified.

SNP	Locus ID	*P*	Position	Gene	Distance (bp)
1797s2	876	2×10^{-18}	Chr4:31559634	*KLHL2*	4598
5631s2	2968	5×10^{-16}	Chr26:543144	-	-
6982s1	3720	6×10^{-14}	Chr2:7119859	*GALNT11*	1932
4533s3	2343	1×10^{-13}	-	-	-
3167s2	1566	2×10^{-13}	Chr14:5894626	-	-
2736s1	1338	7×10^{-13}	Chr26:1268116	*SRGAP2*	In gene
1561s1	769	2×10^{-12}	Chr2:10208736	-	-
2882s2	1410	6×10^{-12}	Chr1A:43132973	-	-
134s2	63	2×10^{-11}	-	-	-
6166s3	3266	3×10^{-11}	Chr12:20586177	-	-
2242s2	1103	3×10^{-11}	Chr2:5580203	*CSRNP1*	3640
8568s1	4404	1×10^{-9}	Chr1:15057899	*PDHA1*	5737
4367s1	2248	4×10^{-9}	-	-	-
1161s1	563	1×10^{-8}	Chr4:25601155	*CENPC*	9325
2378s1	1165	1×10^{-8}	Chr8:11503072	-	-
1024s1	499	1×10^{-8}	Chr3:98902195	*YWHAQ*	7233
4207s1	2163	3×10^{-8}	Chr1A:15105641	*CELSRI*	In gene
7258s2	3869	3×10^{-8}	Chr6:29212571	-	-
1182s1	575	5×10^{-8}	Chr2:93593389	*ELP2*	In gene
6566s1	3483	7×10^{-8}	Chr3:95993765	-	-
7247s1	3863	1×10^{-7}	Chr8:6825755	-	-
7509s6	3997	2×10^{-7}	Chr14:9421638	*ZNF598*	2844
5744s3	3034	3×10^{-7}	Chr5:51616700	-	-
980s3	477	3×10^{-7}	Chr18:4947869	*SHISA6*	In gene
3618s1	1820	3×10^{-7}	-	-	-
8383s2	4328	4×10^{-7}	Chr13:6606756	*SLC7A1*	1796
3705s2	1867	4×10^{-7}	-	-	-
3722s1	1877	4×10^{-7}	Chr25:1124724	*VPS45*	262
319s1	166	7×10^{-7}	Chr12:17793212	-	-
3519s2	1766	8×10^{-7}	-	-	-
3776s1	1910	1×10^{-6}	Chr11:13227522	-	-
7560s2	4029	1×10^{-6}	Chr3:97076767	-	-
3764s1	1902	2×10^{-6}	Chr1:39521212	*CLYBL*	In gene
5439s5	2850	2×10^{-6}	Chr10:8024182	*WDR72*	8811
9140s1	4529	3×10^{-6}	-	-	-
8261s3	4270	3×10^{-6}	-	-	-
6760s2	3598	3×10^{-6}	-	-	-
2480s1	1223	5×10^{-6}	Chr13:14913074	*SLIT3*	In gene
87s1	41	6×10^{-6}	-	-	-
543s1	273	7×10^{-6}	Chr4A:4170899	*DACH2*	In gene
5753s2	3040	8×10^{-6}	Chr2:31080579	*CNTNAP2*	In gene
647s1	318	1×10^{-5}	Chr15:9293017	*MED13L*	In gene
1609s2	790	2×10^{-5}	Chr24:718434	-	-
194s1	99	2×10^{-5}	-	-	-

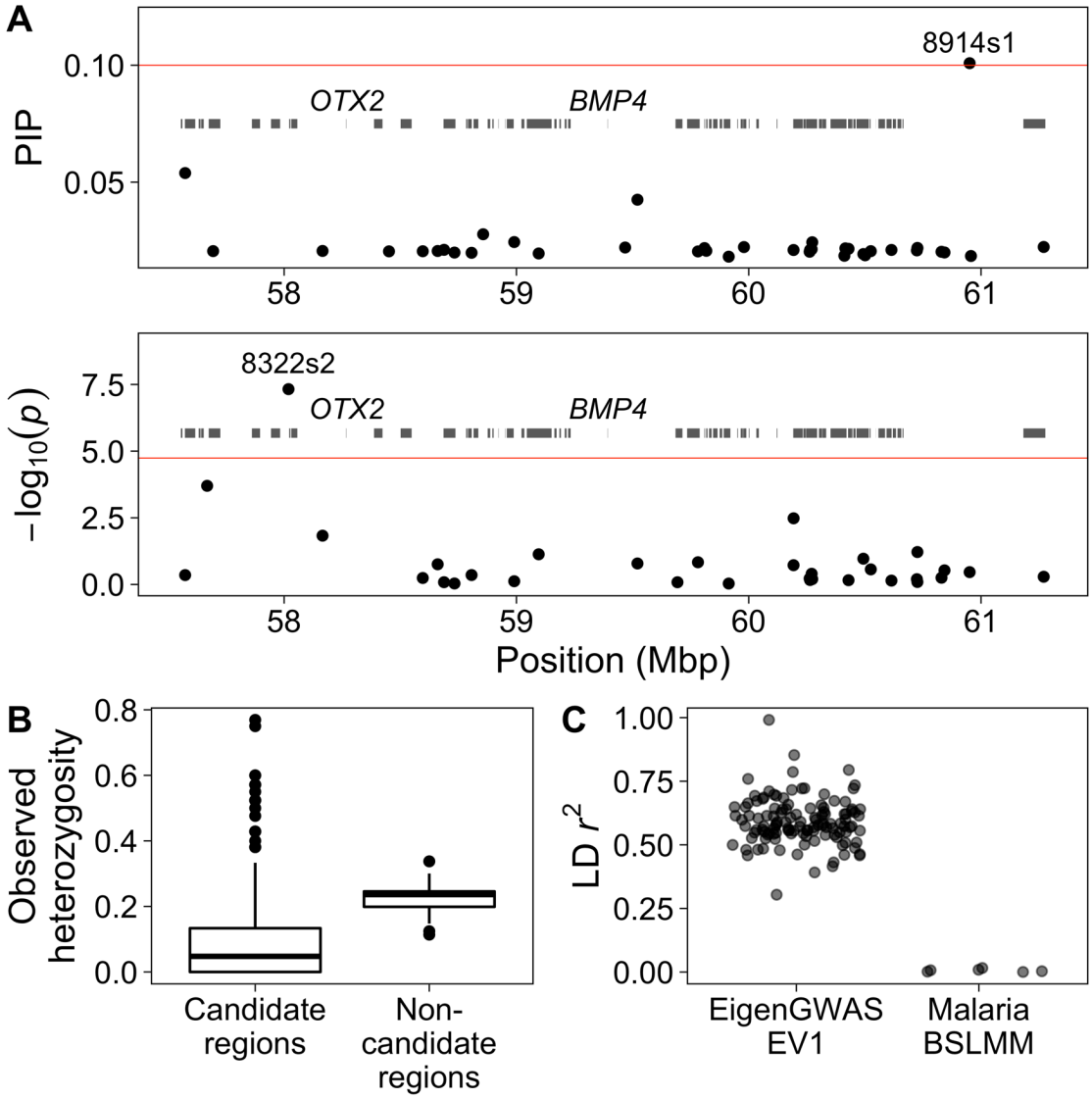


Figure 2.8 Outlier SNPs from BLSMM and EigenGWAS analyses. A) Zoomed in region of interest on Chromosome 5. Top panel: chromosome locations and posterior inclusion probabilities (PIP) for SNPs in the BSLMM bill length analysis in the "Canary Islands" dataset. Red line indicates PIP threshold of 0.1. Bottom panel: SNP chromosome locations and EigenGWAS Eigenvector 1 *P* values. Red line indicates Bonferroni-corrected *P* value threshold. Grey rectangles show location of genes within this region, with *OTX2* and *BMP4* labelled. B) Individual observed heterozygosity averaged across SNPs taken from EigenGWAS outliers (candidate regions – see main text) versus the rest of the genome (non-candidate regions). C) Linkage disequilibrium between all pairs of outlier SNPs from EigenGWAS Eigenvector 1 and malaria BSLMM analysis. Measures of LD for both sets of SNPs were calculated using the "Berthelot's" dataset.

None of the candidate SNPs identified in the BSLMM analyses were identified as outliers in the EigenGWAS analysis. An outlier SNP in the EigenGWAS analysis of EV1 was located close to the region on chromosome 5 that showed an association with bill length (Fig. 2.5). However, closer inspection of the region revealed that the most significant SNP was 2,932,270 bp from the most significant BSLMM SNP (Fig. 2.8A), which was not closer than expected by chance

(randomisation test, P = 0.1). A second SNP in the region also had an elevated (but not significant) PIP value in the BSLMM analysis and was much closer to the significant EigenGWAS candidate SNP (444,885 bp; Fig. 2.8A; P < 0.001). The regions identified in the BSLMM for association with malaria did not show signs of divergent selection between the archipelagos.

2.5. Discussion

Using RAD sequencing, we examined fine-scale population structure among Berthelot's pipit populations, finding within-archipelago genetic structuring between island populations, as well as confirming the previously detected patterns of among-archipelago variation. After controlling for demographic history, we identified candidate SNPs associated with the ecologically important traits of bill length and malaria resistance. Examining allele frequency variation at candidate SNPs revealed that: i) population-level candidate SNP variation was related to population-level trait variation for bill length both within the founder Canary Islands archipelago and throughout all archipelagos, but only within the Canary Islands for malaria; and, ii) candidate SNP variation was correlated with genome-wide variation. Finally, although we found little evidence for selection on bill length or malaria resistance, we found signatures of divergent selection between the archipelagos across the genome, including at genes involved in metabolism and immune function.

Comparisons of genome-wide diversity in Berthelot's and tawny pipits provided support for previous inferences of colonisation history from microsatellites and mitochondrial DNA (Illera *et al.* 2007; Spurgin *et al.* 2014). The past colonisation history and associated bottlenecks are reflected in linkage disequilibrium within the archipelagos, indicating a larger, more outbred population in the Canary Islands, compared with smaller, more inbred populations in Madeira and Selvagens. This is in concordance with previous estimations of effective population size using approximate Bayesian Computation, which predicted contemporary effective population sizes of ca. 4000, 3000 and 400 individuals for the Canary Islands, Madeira, and Selvagens, respectively (Spurgin *et al.* 2014). Using population admixture analysis, we were able to further describe population structure: Deserta Grande diverged from the rest of the Madeiran archipelago, possibly suggesting a further bottleneck on Deserta Grande. We also detected a weak east–west gradient in population structure within the Canary Islands, which was supported by the finding of isolation-by-distance throughout this archipelago.

We next aimed to identify the genetic regions associated with bill length and pathogen resistance — two traits of major evolutionary importance (Smith *et al.* 1995; Daszak 2000; Grant & Grant 2006). The high level of LD relative to outbred populations (e.g. Kardos *et al.* 2016) has

enabled us to detect associations with these two polygenic traits in Berthelot's pipit, with a modest number of markers. The strongest association with bill length was found at a SNP in a region on chromosome 5. This SNP (8914s1) is not located within a gene, which is perhaps not surprising, given i) the density of our marker set, and ii) the fact that many SNPs associated with polygenic traits are located in regulatory regions (Maurano *et al.* 2012). The wider region surrounding SNP 8914s1 and a significant EigenGWAS SNP (see below) contains, among others, the genes *BMP4* and *OTX2*. *BMP4* is a clear candidate for involvement in bill morphology, as studies in Darwin's finches, chickens and ducks have experimentally demonstrated its role in beak development (Abzhanov *et al.* 2004; Wu *et al.* 2004). *OTX2* plays a crucial role in craniofacial development across jawed vertebrates (Kimura *et al.* 1997), and mutations in this gene and structural variants within the wider genomic region have been linked with craniofacial abnormalities in mice and humans (Hide *et al.* 2002; Zielinski *et al.* 2014). While *BMP4* and *OTX2* are good candidates, we stress that further work is required to identify which specific genes affect variation in bill length. Further, bill length is almost certainly a polygenic trait (e.g. Abzhanov *et al.* 2006; Lamichhaney *et al.* 2015, 2016), and many causative loci have likely gone undetected in this study.

There are now several examples linking variation at avian bills, at both the genetic and phenotypic levels, which have shown natural selection to be the primary driver of bill shape variation (Lamichhaney *et al.* 2015; Bosse *et al.* 2017). Here we revealed a striking relationship between genetic variation at our candidate bill-length SNP 8914s1, and population-level variation in bill length; the long-bill allele identified using birds from the Canary Islands is at near fixation in the Madeiran Islands, where birds have the longest bills (Fig. 2.5C). We then found that variation at SNP 8914s1 was correlated with genome-wide variation among populations (Fig. 2.4), with the lowest levels of diversity in the bottlenecked Madeiran and Selvagens populations (Fig. 2.5C). Combined with the finding that SNP 8914s1 was not identified in the EigenGWAS analysis, this suggests that founder effects are likely to have played a major role in shaping bill length variation among Berthelot's pipit populations. This is consistent with our previous phenotypic work on this species (Spurgin *et al.* 2014), but only now have we been able to show how founder effects simultaneously affect phenotypes and underlying genotypes — indeed, very few studies have done so.

Despite a clear role for founder effects in shaping bill length variation, we did find that genetic structuring at SNP 8914s1 was higher than the genome-wide average, particularly across the Canarian and Madeiran archipelagos (Fig. 2.4). We also detected an EigenGWAS outlier SNP (8322s2) putatively under divergent selection within the wider *BMP4*/*OTX2* bill-length candidate region (Fig. 2.8A). Furthermore, we might not expect to be able to identify genotype–phenotype

associations at SNPs under strongest divergent selection, as these are expected to have low variability within populations (indeed this is the case in our data — Fig. 2.8B) and detecting rare genetic variants associated with phenotypes is very difficult (Li & Leal 2008). Thus, while the majority of observed bill length variation among archipelagos can be explained by neutral forces, we cannot rule out that selection has occurred for longer bills in Madeira (or shorter bills in southerly populations). Further research is required to identify whether selection has played any role in shaping beak morphology, and the underlying genetic variants associated with this key ecological trait.

Malaria can act as a strong selective force (Warner 1968; Van Riper *et al.* 1986; Ortego *et al.* 2008; Knowles *et al.* 2010), and drive the evolution of increased tolerance to infection (Atkinson *et al.* 2013). Genotypic associations with malaria infection in individual Berthelot's pipits revealed two outlier SNPs with high levels of significance (Table 2.3). The first of these was on chromosome 10, approximately 2,000 bp from *IL-16*. This gene encodes a pro-inflammatory cytokine that has been implicated in susceptibility to various inflammatory disorders (Gao *et al.* 2009; Wu *et al.* 2011), and interacts with other cytokines that are associated with malaria infection (Kern *et al.* 1989; Mathy *et al.* 2000; Lyke *et al.* 2004). The second outlier SNP was found on chromosome 20, within *RIMS4*. Research on this gene is lacking, though it has been found to be overexpressed in breast cancer tumour cells (Abba *et al.* 2005).

Examining allele frequency variation at the malaria resistance candidate SNPs revealed a complex set of relationships. Despite finding a clear relationship of malaria infection with both individual- and population-level SNP variation in the Canary Islands, this pattern did not hold when the Madeiran and Selvagens archipelagos were included (Fig. 2.3). Further, our EigenGWAS analysis did not detect signatures of divergent selection between archipelagos in the regions surrounding the malaria candidates *RIMS4* or *IL-16*. In contrast to bill length, which seems more likely to have a conserved genetic architecture, resistance to malaria is probably a rapidly evolving trait, and candidate-gene studies have shown that different alleles can be associated with malaria infection among different populations (Bonneaud *et al.* 2006). Variation at both malaria candidate SNPs was correlated with genome-wide variation (Fig. 2.4) and likely to be shaped by founder effects. However, our study suggests that, because the genetic basis of malaria resistance is likely to vary among populations (as expected with rapid host–pathogen coevolution, Slade & McCallum 1992), there is unlikely to be a simple relationship between resistance to malaria infection and population demography in wild populations.

Any analyses that requires high genotyping accuracy at individual loci, such as our BSLMM and EigenGWAS analyses, are susceptible to erroneous results due to PCR duplicates (Andrews *et al.* 2016). However, in our case PCR duplicates are likely to make our analyses more conservative.

For the BSLMM analyses, PCR duplicates will be random with respect to bill length or malaria infection status, so if anything they would obscure relationships between genotype and phenotype. A greater possibility is that PCR duplicates may generate false signatures of divergent selection. However, the high levels of LD between all of our candidate SNPs from the EigenGWAS analysis suggest a highly correlated, non-random pattern of evolution. It is therefore unlikely that PCR duplicates have introduced a systematic bias into our results.

The EigenGWAS analysis detected putative signatures of selection among the archipelagos at a number of locations around the genome (Table 2.5). The significant EigenGWAS SNPs were all in high LD with one another (Fig. 2.8C), suggesting that these sites have evolved non-independently. Also important to note is that many of these SNPs are not situated close to candidate genes. Some of these could be situated close to features such as trans-regulatory elements or structural variants (Yvert *et al.* 2003; Manolio *et al.* 2009), or could be false positives that have arisen as a result of population bottlenecks (Akey *et al.* 2004). Our analyses suggest that the EigenGWAS analysis was more conservative than a simple F_{ST}-based approach to detecting selection, with numerous F_{ST} outliers having non-significant EigenGWAS *P* values after employing a genomic control (Fig. 2.7). This is in contrast to other studies which have shown almost perfect correlations between F_{ST} and EigenGWAS *P* values (Chen *et al.* 2016; Bosse *et al.* 2017), and suggests that EigenGWAS may be more conservative when levels of drift are high. However, it is unlikely that all false positives are accounted for, and further research is now needed to understand how EigenGWAS performs in bottlenecked populations.

Confounding effects of bottlenecks aside, a substantial number of our SNPs identified in the EigenGWAS analyses are likely to be "true" positives, suggesting that traits other than bill length and malaria resistance may have been under selection among archipelagos. We identified SNPs located within loci involved in immune function (e.g. *PREX1*, Welch *et al.* 2005; *MAP3K14*, Thu & Richmond 2010), or with potential to enable adaptation to climate (e.g. *DLAT*, Blier & Guderley 1993; *SNX29*, Sung *et al.* 2016). These findings are reflected in other studies that utilise genome scans to detect signatures of selection. Adaptation to toxic food sources in an isolated population of *Drosophila yakuba* was detected in a genome scan between island and mainland populations (Yassin *et al.* 2016). Divergent selection for pathogen-resistance candidate genes has been found in *Daphnia* (Bourgeois *et al.* 2017) and across a bank vole range expansion (White *et al.* 2013), and pathogens were the strongest selective pressure identified in human evolutionary history (Fumagalli *et al.* 2011). Likewise, signatures of divergent selection have been detected for genes related to metabolic processes along latitudinal gradients in numerous species (Sezgin *et al.* 2004; Hancock *et al.* 2011; Pujolar *et al.* 2014). This suggests that there may be general patterns in the types of genes that show divergence between populations, with

similar processes of adaptation across species. Future studies investigating links between genotypes under selection and ecologically important phenotypic traits will provide further support for the role of adaptation in driving patterns of biodiversity.

2.6. References

Abba, M.C., Hu, Y., Sun, H., Drake, J.A., Gaddis, S. & Baggerly, K. *et al.* (2005). Gene expression signature of estrogen receptor alpha status in breast cancer. *BMC Genomics*, 6, 37.

Abzhanov, A., Kuo, W.P., Hartmann, C., Grant, B.R., Grant, P.R. & Tabin, C.J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, 442, 563–567.

Abzhanov, A., Protas, M., Grant, B.R., Grant, P.R. & Tabin, C.J. (2004). *Bmp4* and morphological variation of beaks in Darwin's finches. *Science*, 305, 1462–1465.

Akey, J.M., Eberle, M.A., Rieder, M.J., Carlson, C.S., Shriver, M.D. & Nickerson, D.A. *et al.* (2004). Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biology*, 2, e286.

Alexander, D.H. & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, 12, 246.

Alexander, D.H., Novembre, J. & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403–410.

Amadon, D. (1950). The Hawaiian honeycreepers (Aves, Drepaniidae). *Bulletin of the American Museum of Natural History*, 95, 153–262.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. & Hohenlohe, P.A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17, 81–92.

Atkinson, C.T., Saili, K.S., Utzurrum, R.B. & Jarvi, S.I. (2013). Experimental evidence for evolved tolerance to avian malaria in a wild population of low elevation Hawai'i 'Amakihi (*Hemignathus virens*). *EcoHealth*, 10, 366–375.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L. & Lewis, Z.A. *et al.* (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3, e3376.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R. & Bernatchez, L. (2015). RAD-genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species; the American lobster (*Homarus americanus*). *Molecular Ecology*, 24, 3299–3315.

Blier, P.U. & Guderley, H.E. (1993). Effects of pH and temperature on the kinetics of pyruvate oxidation by muscle mitochondria from rainbow trout (*Oncorhynchus mykiss*). *Physiological Zoology*, 66, 474–489.

Bonneaud, C., Pérez-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.

Bosse, M., Spurgin, L.G., Laine, V.N., Cole, E.F., Firth, J.A. & Gienapp, P. *et al.* (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, 358, 365–368.

Bourgeois, Y., Roulin, A.C., Müller, K. & Ebert, D. (2017). Parasitism drives host genome evolution: Insights from the *Pasteuria ramosa–Daphnia magna* system. *Evolution*, 71, 1106–1113.

Brachi, B., Meyer, C.G., Villoutreix, R., Platt, A., Morton, T.C. & Roux, F. *et al.* (2015). Coselected genes determine adaptive variation in herbivore resistance throughout the native range of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 112, 4032–4037.

Carlquist, S. (1974). *Island Biology*. Columbia University Press, New York, NY.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7.

Chaves, J.A., Cooper, E.A., Hendry, A.P., Podos, J., De León, L.F. & Raeymaekers, J.A.M. *et al.* (2016). Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Molecular Ecology*, 25, 5282–5295.

Chen, G.-B., Lee, S.H., Zhu, Z.-X., Benyamin, B. & Robinson, M.R. (2016). EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Heredity*, 117, 51–61.

Comeault, A.A., Carvalho, C.F., Dennis, S., Soria-Carrasco, V. & Nosil, P. (2016). Color phenotypes are under similar genetic control in two distantly related species of *Timema* stick insect. *Evolution*, 70, 1283–1296.

Corander, J., Majander, K.K., Cheng, L. & Merilä, J. (2013). High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Molecular Ecology*, 22, 2931–2940.

Coyne, J.A. & Orr, H.A. (2004). *Speciation*. Sinauer, Sunderland, MA.

Cramp, S. (1988). *The Birds of the Western Palearctic, Volume 5*. Oxford University Press, Oxford, UK.

DaCosta, J.M. & Sorenson, M.D. (2014). Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS ONE*, 9, e106713.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E. & DePristo, M.A. *et al.* (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.

Daszak, P. (2000). Emerging infectious diseases of wildlife - threats to biodiversity and human health. *Science*, 287, 443–449.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. & Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, 499–510.

Devlin, B. & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, 55, 997–1004.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460–2461.

Eizaguirre, C., Lenz, T.L., 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, S.A., Jacobsson, P., Fiske, P., Sahlman, T. & Grahn, M. *et al.* (2007). Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology*, 16, 1439–1451.

Fernández, R., Schubert, M., Vargas, A., Brownlow, A., Víkingsson, G. & Siebert, U. *et al.* (2016). A genomewide catalogue of single nucleotide polymorphisms in white-beaked and Atlantic white-sided dolphins. *Molecular Ecology Resources*, 16, 266–276.

Fumagalli, M., Moltke, I., Grarup, N., Racimo, F., Bjerregaard, P. & Jorgensen, M.E. *et al.* (2015). Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science*, 349, 1343–1347.

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L. & Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, 7, e1002355.

Gagnaire, P.A. & Gaggiotti, O.E. (2016). Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology*, 62, 603–616.

Gao, L.B., Rao, L., Wang, Y.Y., Liang, W.B., Li, C. & Xue, H. *et al.* (2009). The association of *interleukin-16* polymorphisms with IL-16 serum levels and risk of colorectal and gastric cancer. *Carcinogenesis*, 30, 295–299.

Garris, A.J., Tai, T.H., Coburn, J., Kresovich, S. & McCouch, S. (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics*, 169, 1631–1638.

Gompert, Z., Lucas, L.K., Nice, C.C. & Buerkle, C.A. (2013). Genome divergence and the genetic architecture of barriers to gene flow between *Lycaeides idas* and *L. melissa*. *Evolution*, 67, 2498–2514.

González-Quevedo, C., Davies, R.G. & Richardson, D.S. (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.

González-Quevedo, C., Phillips, K.P., Spurgin, L.G. & Richardson, D.S. (2015a). 454 screening of individual MHC variation in an endemic island passerine. *Immunogenetics*, 67, 149–162.

González-Quevedo, C., Spurgin, L.G., Illera, J.C. & Richardson, D.S. (2015b). Drift, not selection, shapes toll-like receptor variation among oceanic island populations. *Molecular Ecology*, 24, 5852–5863.

Goslee, S.C. & Urban, D.L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22, 1–19.

Grant, P.R. (1986). *Ecology and Evolution of Darwin's Finches*. Princeton University Press, Princeton, New Jersey.

Grant, P.R. & Grant, B.R. (2006). Evolution of character displacement in Darwin's finches. *Science*, 313, 224–226.

Griffiths, R., Double, M.C., Orr, K. & Dawson, R.J.G. (1998). A DNA test to sex most birds. *Molecular Ecology*, 7, 1071–1075.

Hancock, A.M., Witonsky, D.B., Alkorta-Aranburu, G., Beall, C.M., Gebremedhin, A. & Sukernik, R. *et al.* (2011). Adaptations to climate-mediated selective pressures in humans. *PLoS Genetics*, 7, e1001375.

Hide, T., Hatakeyama, J., Kimura-Yoshida, C., Tian, E., Takeda, N. & Ushio, Y. *et al.* (2002). Genetic modifiers of otocephalic phenotypes in *Otx2* heterozygous mutant mice. *Development*, 129, 4347–4357.

Hinrichs, A.L., Larkin, E.K. & Suarez, B.K. (2009). Population stratification and patterns of linkage disequilibrium. *Genetic Epidemiology*, 33, S88–S92.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. & Cresko, W.A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, 6, e1000862.

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

Illera, J.C., Emerson, B.C. & Richardson, D.S. (2007). Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology*, 16, 4599–4612.

Illera, J.C., Emerson, B.C. & Richardson, D.S. (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.

Illera, J.C., Spurgin, L.G., Rodriguez-Exposito, E., Nogales, M. & Rando, J.C. (2016). What are we learning about speciation and extinction from the Canary Islands? *Ardeola*, 63, 15–33.

Johnson, R.C., Nelson, G.W., Troyer, J.L., Lautenberger, J.A., Kessing, B.D. & Winkler, C.A. *et al.* (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*, 11, 724.

Johnston, S.E., Orell, P., Pritchard, V.L., Kent, M.P., Lien, S. & Niemelä, E. *et al.* (2014). Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, 23, 3452–3468.

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

Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E. & Johnson, J. *et al.* (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484, 55–61.

Jorde, L.B., Rogers, A.R., Bamshad, M., Watkins, W.S., Krakowiak, P. & Sung, S. *et al.* (1997). Microsatellite diversity and the demographic history of modern humans. *Evolution*, 94, 3100–3103.

Josephs, E.B., Stinchcombe, J.R. & Wright, S.I. (2017). What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytologist*, 214, 21–33.

Kardos, M., Husby, A., McFarlane, S.E., Qvarnström, A. & Ellegren, H. (2016). Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resources*, 16, 727–741.

Kardos, M., Luikart, G., Bunch, R., Dewey, S., Edwards, W. & McWilliam, S. *et al.* (2015). Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Molecular Ecology*, 24, 5616–5632.

Kern, P., Hemmer, C.J., Damme, J.V., Gruss, H.-J. & Dietrich, M. (1989). Elevated tumor necrosis factor alpha and interleukin-6 serum levels as markers for complicated *Plasmodium falciparum* malaria. *The American Journal of Medicine*, 87, 139–143.

Kimura, C., Takeda, N., Suzuki, M., Oshimura, M., Aizawa, S. & Matsuo, I. (1997). Cis-acting elements conserved between mouse and pufferfish *Otx2* genes govern the expression in mesencephalic neural crest cells. *Development*, 124, 3929–3941.

Knowles, S.C.L., Palinauskas, V. & Sheldon, B.C. (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.

Kreitman, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature*, 304, 412–417.

Küpper, C., Stocks, M., Risse, J.E., Remedios, N. dos, Farrell, L.L. & McRae, S.B. *et al.* (2015). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, 48, 79–83.

Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M. & Martinez-Barrio, A. *et al.* (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518, 371–375.

Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almén, M.S. & Webster, M.T. *et al.* (2016). A beak size locus in Darwin's finches facilitates character displacement during a drought. *Science*, 352, 470–474.

Lenz, T.L., 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.

Lewontin, R.C. & Hubby, J.L. (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, 54, 595–609.

Li, B. & Leal, S.M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *American Journal of Human Genetics*, 83, 311–321.

Li, M.-H. & Merilä, J. (2010). Sex-specific population structure, natural selection, and linkage disequilibrium in a wild bird population as revealed by genome-wide microsatellite analyses. *BMC Evolutionary Biology*, 10, 66.

Liu, S., Vallejo, R.L., Palti, Y., Gao, G., Marancik, D.P. & Hernandez, A.G. *et al.* (2015). Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Frontiers in Genetics*, 6, 298.

Lozier, J.D. (2014). Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Molecular Ecology*, 23, 788–801.

Luikart, G., England, P.R., Tallmon, D., Jordan, S. & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4, 981–994.

Lyke, K.E., Burges, R., Cissoko, Y., Sangare, L., Dao, M. & Diarra, I. *et al.* (2004). Serum levels of the proinflammatory cytokines Interleukin-1 Beta (IL-1β), IL-6, IL-8, IL-10, Tumor Necrosis Factor Alpha, and IL-12 (p70) in Malian children with severe *Plasmodium*

falciparum malaria and matched uncomplicated malaria or healthy cont. *Infection and Immunity*, 72, 5630–5637.

Manel, S., Perrier, C., Pratlong, M., Abi-Rached, L., Paganini, J. & Pontarotti, P. *et al.* (2016). Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular Ecology*, 25, 170–184.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A. & Hunter, D.J. *et al.* (2009). Finding the missing heritability of complex diseases. *Nature*, 461, 747–753.

Marchini, J., Cardon, L.R., Phillips, M.S. & Donnelly, P. (2004). The effects of human population structure on large genetic association studies. *Nature Genetics*, 36, 512–517.

Martín, A. & Lorenzo, J.A. (2001). *Aves del Archipiélago Canario*. Francisco Lemus Editor, La Laguna.

Mathy, N.L., Scheuer, W., Lanzendörfer, M., Honold, K., Ambrosius, D. & Norley, S. *et al.* (2000). Interleukin-16 stimulates the expression and production of pro-inflammatory cytokines by human monocytes. *Immunology*, 100, 63–69.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E. & Wang, H. *et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 337, 1190–1195.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A. & Johnson, E.A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17, 240–248.

Nadeau, N.J., Whibley, A., Jones, R.T., Davey, J.W., Dasmahapatra, K.K. & Baxter, S.W. *et al.* (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 343–353.

Oliveira, P. & Menezes, D. (2004). *Birds of the Archipelago of Madeira*. Serviço do Parque Natural de Madeira, Madeira.

Ortego, J., Cordero, P.J., Aparicio, J.M. & Calabuig, G. (2008). Consequences of chronic infections with three different avian malaria lineages on reproductive performance of Lesser Kestrels (*Falco naumanni*). *Journal of Ornithology*, 149, 337–343.

Parra, G., Bradnam, K. & Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, 23, 1061–1067.

Pavey, S.A., Normandeau, E., Castonguay, M. & Normandeau, E. (2015). RAD sequencing highlights polygenic discrimination of habitat ecotypes in the Panmictic American Eel. *Current Biology*, 25, 1666–1671.

Pujolar, J.M., Jacobsen, M.W., Als, T.D., Frydenberg, J., Munch, K. & Jónsson, B. *et al.* (2014). Genome-wide single-generation signatures of local selection in the panmictic European eel. *Molecular Ecology*, 23, 2514–2528.

Richardson, D.S., Jury, F.L., 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.

Senar, J.C. & Pascual, J. (1997). Keel and tarsus length may provide a good predictor of avian body size. *Ardea*, 85, 269–274.

Sezgin, E., Duvernell, D.D., Matzkin, L.M., Duan, Y., Zhu, C.T. & Verrelli, B.C. *et al.* (2004). Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*. *Genetics*, 168, 923–931.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210–3212.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J. & Birol, İ. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research*, 19, 1117–1123.

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

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

Smith, T.B., Freed, L.A., Lepson, J.K. & Carothers, J.H. (1995). Evolutionary consequences of extinctions in populations of a Hawaiian honeycreeper. *Conservation Biology*, 9, 107–113.

Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A. & Richardson, D.S. (2014). Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Molecular Ecology*, 23, 1028–1039.

Spurgin, L.G., Illera, J.C., Padilla, D.P. & Richardson, D.S. (2012). Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia*, 168, 691–701.

Spurgin, L.G., Oosterhout, C. van, Illera, J.C., Bridgett, S., Gharbi, K. & Emerson, B.C. *et al.* (2011). Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, 20, 5213–5225.

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J. & Ekblom, R. *et al.* (2010). Adaptation genomics: the next generation. *Trends in Ecology and Evolution*, 25, 705–712.

Stinchcombe, J.R. & Hoekstra, H.E. (2008). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, 100, 158–170.

Sung, Y.J., Pérusse, L., Sarzynski, M.A., Fornage, M., Sidney, S. & Sternfeld, B. *et al.* (2016). Genome-wide association studies suggest sex-specific loci associated with abdominal and visceral fat. *International Journal of Obesity*, 40, 662–674.

Thu, Y.M. & Richmond, A. (2010). NF-κB inducing kinase: A key regulator in the immune system and in cancer. *Cytokine and Growth Factor Reviews*, 21, 213–226.

Turner, T.L., Bourne, E.C., Von Wettberg, E.J., Hu, T.T. & Nuzhdin, S.V. (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics*, 42, 260–263.

Van Riper, C., Van Riper, S.G., Goff, L.M. & Laird, M. (1986). The epizootiology and ecological significance of malaria in Hawaiian land birds. *Ecological Monographs*, 56, 327–344.

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

Waldenström, J., Bensch, S., Hasselquist, D. & Östman, Ö. (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.

Warner, R.E. (1968). The role of introduced diseases in the extinction of the endemic Hawaiian avifauna. *The Condor*, 70, 101–120.

Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W. & Künstner, A. *et al.* (2010). The genome of a songbird. *Nature*, 464, 757–762.

Weisenfeld, N.I., Yin, S., Sharpe, T., Lau, B., Hegarty, R. & Holmes, L. *et al.* (2014). Comprehensive variation discovery in single human genomes. *Nature Genetics*, 46, 1350–1355.

Welch, H.C.E., Condliffe, A.M., Milne, L.J., Ferguson, G.J., Hill, K. & Webb, L.M.C. *et al.* (2005). P-Rex1 regulates neutrophil function. *Current Biology*, 15, 1867–1873.

White, T.A., Perkins, S.E., Heckel, G. & Searle, J.B. (2013). Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology*, 22, 2971–2985.

Wu, J., Wang, Y., Zhang, Y. & Li, L. (2011). Association between *interleukin-16* polymorphisms and risk of coronary artery disease. *DNA and Cell Biology*, 30, 305–308.

Wu, P., Jiang, T.-X., Suksaweang, S., Widelitz, R.B. & Chuong, C.-M. (2004). Molecular shaping of the beak. *Science*, 305, 1465–1466.

Yassin, A., Debat, V., Bastide, H., Gidaszewski, N., David, J.R. & Pool, J.E. (2016). Recurrent specialization on a toxic fruit in an island *Drosophila* population. *Proceedings of the National Academy of Sciences*, 113, 4771–4776.

Yvert, G., Brem, R.B., Whittle, J., Akey, J.M., Foss, E. & Smith, E.N. *et al.* (2003). Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genetics*, 35, 57–64.

Zhang, J., Rosenberg, H.F. & Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proceedings of the National Academy of Sciences*, 95, 3708–3713.

Zhou, X. & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44, 821–824.

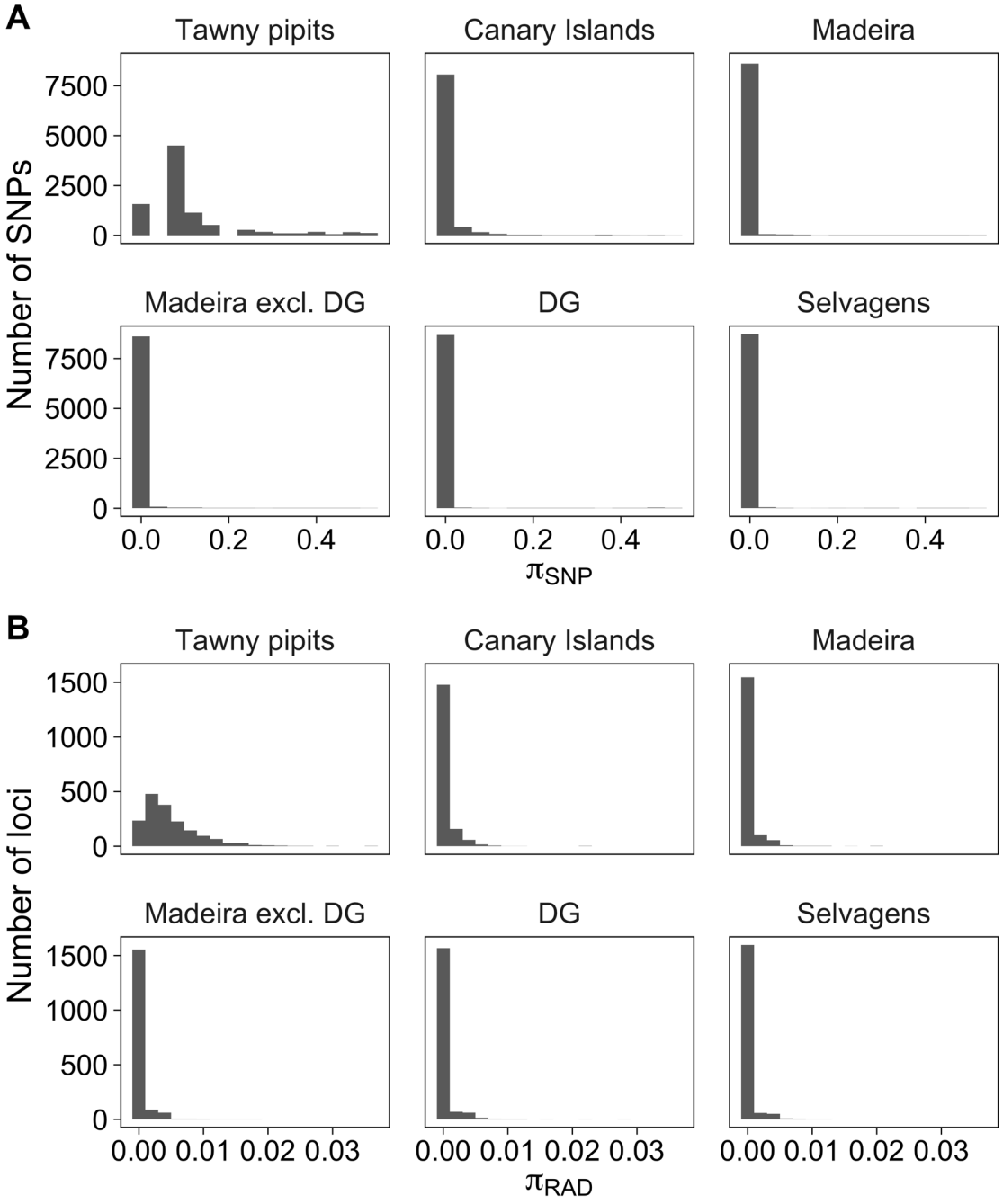
Zhou, X., Carbonetto, P. & Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics*, 9, e1003264.

Zielinski, D., Markus, B., Sheikh, M., Gymrek, M., Chu, C. & Zaks, M. *et al.* (2014). OTX2 duplication is implicated in hemifacial microsomia. *PLoS ONE*, 9, e96788.

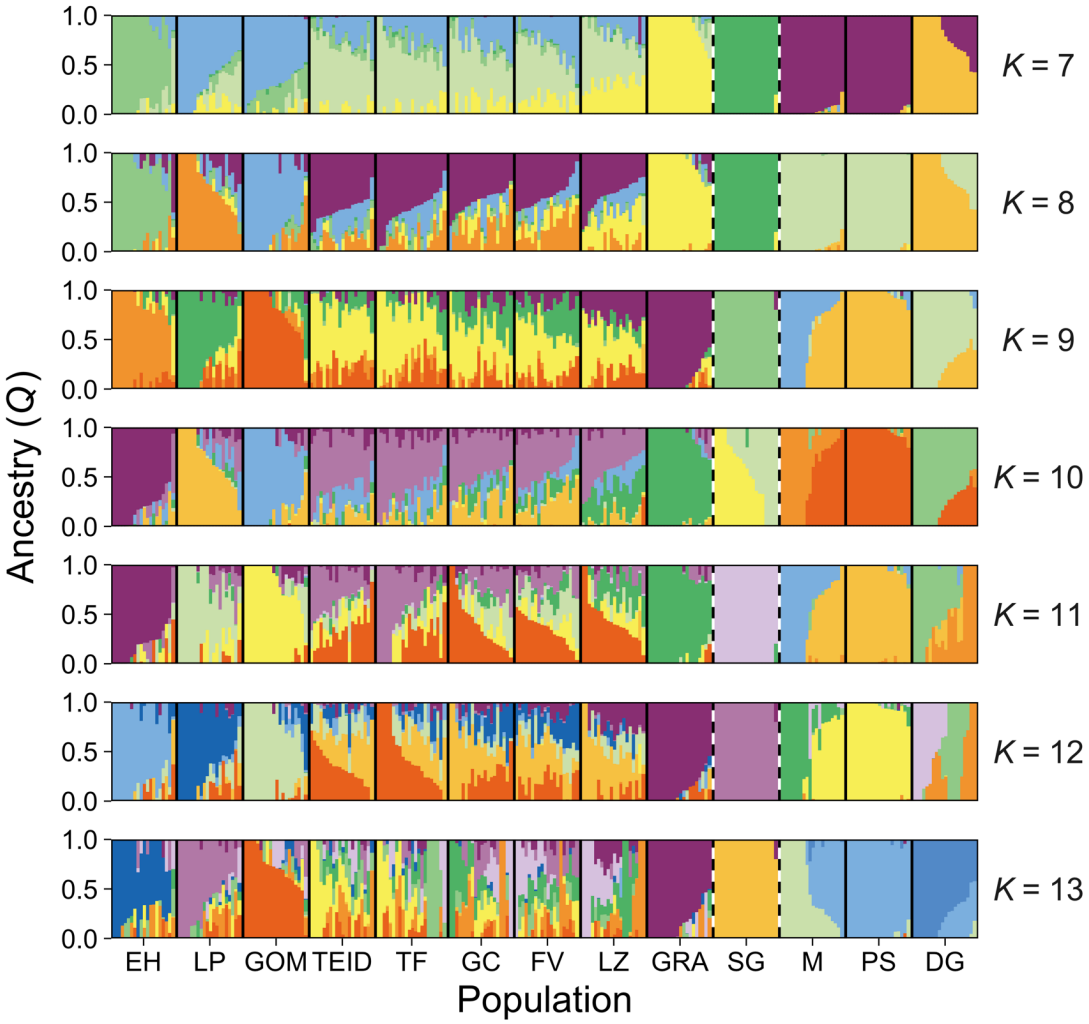
2.7. Supplementary Material

Supplementary Table S2.1 CEGMA and BUSCO results for the Berthelot's pipit genome assembly.

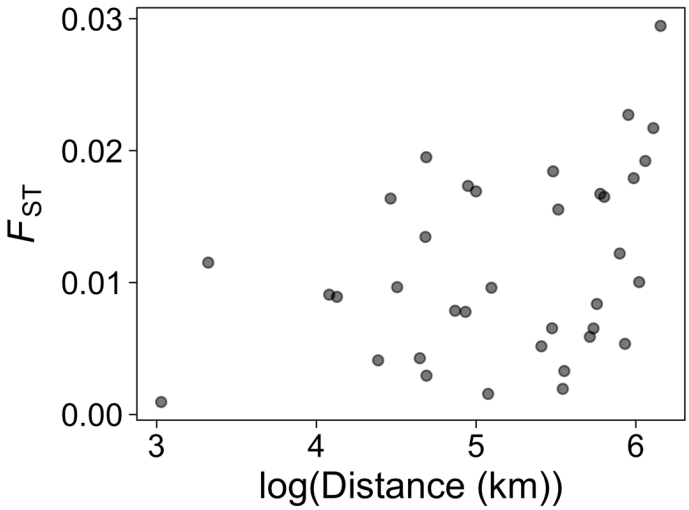
CEGMA		BUSCO vertebrate genes			
Complete	Partial	Complete single-copy	Complete duplicated	Fragmented	Missing
152 (61%)	224 (90%)	1944 (64%)	21 (0.6%)	476 (15%)	603 (19%)



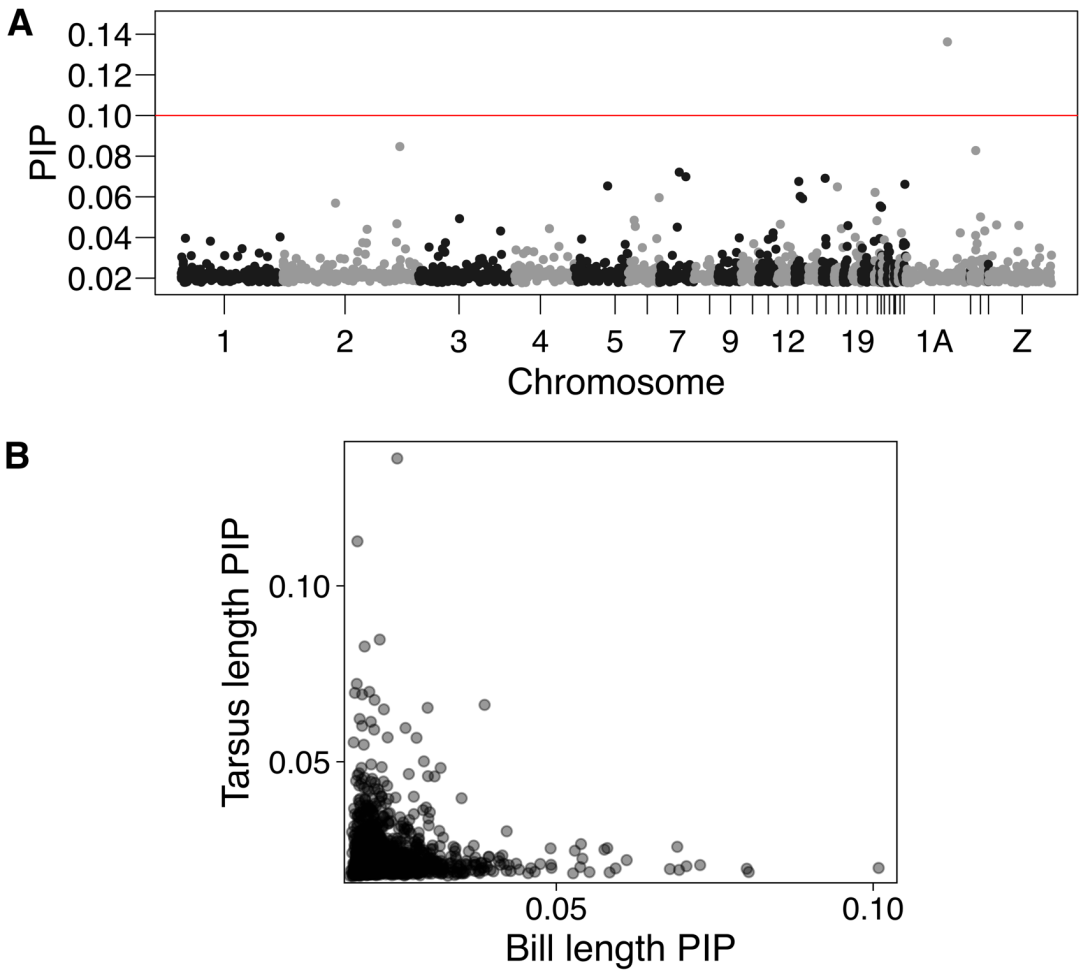
Supplementary Figure S2.1 Nucleotide diversity in Tawny pipits and Berthelot's pipit archipelagos. Histograms showing the distributions of π_{SNP} (top two rows) π_{RAD} (bottom two rows) in each group. As Admixture analysis demonstrated moderate population structuring within the Madeiran archipelago (Fig. 2.2A), π_{SNP} and π_{RAD} have been calculated for all islands in the Madeiran archipelago, and additionally separating the more divergent population, Deserta Grande (DG), from Madeira and Porto Santo.



Supplementary Figure S2.2 Admixture analysis at *K* = 7 to *K* = 13 clusters for Berthelot's pipit. Vertical bars represent individual pipits, and are coloured by their assignment to each one of *K* clusters. Solid black lines indicate separate populations; dashed black lines indicate separate archipelagos. Populations from left to right: Canary Islands – El Hierro (EH), La Palma (LP), Gomera (GOM), Teide (TEID), Tenerife (TF), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ), Graciosa (GRA); Selvagens – Selvagem Grande (SG); Madeira – Madeira (M), Porto Santo (PS), Deserta Grande (DG).



Supplementary Figure S2.3 Pairwise genetic distance in relation to geographical distance across Berthelot's pipit populations in the Canary Islands.



Supplementary Figure S2.4 Genetic associations with tarsus length. A) Manhattan plot of BSLMM analysis of tarsus length in the Canary Islands. Red line indicates the threshold of posterior inclusion probability (PIP) = 0.1. B) Tarsus length PIP values plotted against bill length PIP values demonstrate a lack of overlap between outlier SNPs.

Chapter 3. Genetic drift outweighs selection to shape AvBD diversity in a bottlenecked island bird

Berthelot's pipit. Photo by Philip Lamb

3.1. Abstract

Population bottlenecks are often associated with the loss of genetic diversity by drift. Low immunogenetic diversity may have negative effects on population persistence, though processes such as pathogen-mediated balancing selection may act to maintain genetic diversity at these loci in the face of bottlenecks. We investigated variation at a family of innate immune genes, the avian β-defensins, in the Berthelot's pipit *Anthus berthelotii*. This endemic island bird has undergone successive bottlenecks throughout its colonisation history, resulting in reductions in neutral genetic diversity. Comparisons between archipelago populations with different severities of bottlenecks, and with the widespread mainland sister species, the tawny pipit *Anthus campestris*, demonstrate a greater loss of allelic and amino acid variation with increasing numbers of bottlenecks. We did not find direct evidence of balancing selection at the AvBD loci in the Berthelot's pipit. However, elevated levels of nucleotide diversity existed at some AvBD loci relative to that observed at genome-wide RAD sequencing loci, probably as a result of past balancing selection on AvBD loci in the ancestral mainland population. In support of this, we found evidence of balancing selection in the tawny pipits at AvBD2. This balancing selection may explain the relatively even frequency of the two functional AvBD2 alleles in the Canary Islands populations of Berthelot's pipits, which may, in turn, explain why these alleles were not lost during the severe bottlenecks associated with the colonisation of the Maderian and Selvagens archipelagos. F_{ST} was often higher at AvBD loci than at RAD loci; while this could suggest population divergence driven by local adaptation, we found no evidence of divergent selection between the archipelagos. Instead, these findings suggest that drift is the predominant force driving AvBD variation in this species, leading to low functional variation at these immunogenetic loci. This may have deleterious consequences for the survival of Berthelot's pipit populations if they are challenged by novel pathogens in the future.

3.2. Introduction

Genetic variation is fundamental to adaptive evolution, providing the raw material on which selection can act. However, genetic drift, which is especially strong in small populations, can erode genetic diversity at adaptive loci (Wright 1931; Bollmer *et al.* 2011; Grueber *et al.* 2013). Populations that have experienced bottlenecks may, therefore, suffer reductions in genetic diversity, which may have detrimental impacts on population persistence and adaptive potential (Lande 1988; Frankham *et al.* 1999; Reed & Frankham 2003). Genetic diversity can also be lost due to the actions of purifying or strong directional selection (Jiggins & Hurst 2003; Winternitz & Wares 2013). However, if loci are subject to balancing selection— where the maintenance of

multiple genetic variants at intermediate frequencies is favoured—genetic diversity can potentially be preserved through a bottleneck (Aguilar *et al.* 2004; Gos *et al.* 2012; Oliver & Piertney 2012). Understanding how genetic drift and selection interact to shape genetic diversity in small populations is therefore important for our understanding of evolution, and also of onsiderable conservation interest.

Pathogen-mediated balancing selection is thought to be responsible for maintaining the high levels of variation observed at many genes of the immune system, because elevated immunogenetic variation allows hosts to defend themselves against a diverse and rapidly-evolving pathogen community (Gaudieri *et al.* 2000; Waterhouse *et al.* 2007; Spurgin & Richardson 2010; Alcaide *et al.* 2014). Disentangling the processes that maintain immunogenetic variation is of importance, as it has direct implications for understanding host-pathogen coevolution, wildlife disease and epidemiology. Furthermore, immune genes represent an interesting model for investigating the more general question of how selection may maintain genetic variation in natural populations (Turner *et al.* 2012; Arbanasić *et al.* 2014; Caseys *et al.* 2015), including populations that have undergone bottlenecks (Aguilar *et al.* 2004; Oliver & Piertney 2012; Grueber *et al.* 2015). Understanding the mechanisms that maintain such variation also has important conservation implications for declining populations with high disease risk (Acevedo-Whitehouse & Cunningham 2006).

Pathogen-mediated selection has been widely studied using the major histocompatibility complex (MHC), a group of genes which play an essential role in pathogen recognition in the adaptive immune system (Sommer 2005; Piertney & Oliver 2006; Eizaguirre & Lenz 2010; Spurgin & Richardson 2010). However, characterising patterns of variation, and the underlying evolutionary forces maintaining diversity at the MHC, is difficult. This is because MHC loci are characterized by extraordinarily high levels of inter- and intra-locus levels of diversity, shaped by multiple, complex and interacting mutational and evolutionary processes other than selection, including gene duplication (Jeffery & Bangham 2000; Hess & Edwards 2002), gene conversion (Ohta 1995; Spurgin *et al.* 2011), epistasis, and mutational load (Van Oosterhout 2009). In addition, MHC-based sexual selection may act on top of pathogen-mediated selection (Richardson *et al.* 2005; Milinski 2006; Eizaguirre *et al.* 2009; Brouwer *et al.* 2010) and lead to escalated levels of MHC diversity (Ejsmond *et al.* 2014). Therefore, MHC genes may be exceptional and not representative of the processes that happen across other immune genes.

Attention is increasingly focused on understanding the drivers of variation in genes involved in innate immunity—the first line of defence against infection in plants and animals, and an evolutionarily important component of host immunity (Akira *et al.* 2006; Jones & Dangl 2006). Antimicrobial peptides are an ancient element of the innate immune system, found in all known

species (Hancock & Diamond 2000). These short (12–50 amino acids) cationic peptides are able to directly kill microbes by disrupting phospholipid membranes and intracellular structures (Zasloff 2002; Yeaman & Yount 2003). In plants, animals and fungi, antimicrobial peptides provide broad defence against bacteria, viruses, fungi and protozoa (Zasloff 2002; Ageitos *et al.* 2017), and play additional roles in inflammation, chemo-attraction of leukocytes and wound healing (Brown & Hancock 2006; Lai & Gallo 2009; Mangoni *et al.* 2016). Defensins are a specific type of antimicrobial peptide characterised by a structure of six (animals) or eight (plants) disulphide-linked cysteine residues (Garcia-Olmedo *et al.* 1998; Lehrer & Ganz 2002; Froy & Gurevitz 2003), which in vertebrates are grouped into the α-, β-, and θ-defensin gene families. Lineage-specific gene duplication followed by divergence of paralogs through positive selection on the mature peptide has been prevalent throughout vertebrate defensin evolution (Semple *et al.* 2006; Hellgren & Ekblom 2010; Cheng *et al.* 2014; Jones *et al.* 2017; Tang *et al.* 2018). Substitutions which change the net charge or hydrophobicity of the mature peptide can alter antimicrobial activity (Antcheva *et al.* 2004; Klüver *et al.* 2005; Higgs *et al.* 2007; Yang *et al.* 2017), suggesting that constantly evolving pathogen communities may be driving selection for diverse defensin repertoires.

Gene duplications in the avian β-defensins (AvBDs) have given rise to various numbers of loci across different avian lineages: 14 AvBDs exist in the chicken *Gallus gallus* (Xiao *et al.* 2004), with 16, 19, 22, and 30 AvBDs in the golden pheasant *Chrysolophus pictus*, mallard *Anas platyrhynchos*, zebra finch *Taeniopygia guttata*, and hwamei *Garrulax canorus*, respectively (Hellgren & Ekblom 2010; Huang *et al.* 2013; Chen *et al.* 2015). Within individual AvBDs purifying selection appears to maintain a conserved overall protein structure, but positive and balancing selection have been detected at particular amino acid positions within the mature peptide domain (Hellgren & Ekblom 2010; Cheng *et al.* 2015; Chapman *et al.* 2016), the region of the defensin which directly interacts with pathogens. Characterisation of AvBDs within populations has detected allelic variation at a number of loci (eg. Hellgren *et al.* 2010; Hellgren & Sheldon 2011; Hellgren 2015; Chapman *et al.* 2016; Gilroy *et al.* 2016; Schmitt *et al.* 2017b). Furthermore, antimicrobial activity has been shown to vary between amino acid variants of AvBD loci found in wild populations (Hellgren *et al.* 2010; Cadwell *et al.* 2017; Schmitt *et al.* 2017a). Trans-species polymorphisms of amino acid sequences have been found in AvBD loci (Hellgren & Sheldon 2011; Chapman *et al.* 2016), suggesting that functional variation can be maintained over large temporal scales. Together, these findings indicate that selection may be acting to increase within-AvBD diversity, in addition to creating diversity through gene duplication. It is therefore possible that selection may counteract genetic drift to maintain adaptive variation at AvBD loci through a population bottleneck.

Berthelot's pipit *Anthus berthelotii* is a passerine bird endemic to the Macaronesian archipelagos of the Canary Islands, Madeira and Selvagens, in the north Atlantic. This species colonised the Canary Islands from Africa, where it diverged from its mainland sister species, the tawny pipit *Anthus campestris* ca. 2.5 million years ago (Voelker 1999), accompanied by reductions in genetic diversity (Armstrong *et al.* 2018). More severe bottlenecks associated losses of genome-wide diversity occurred during the separate northward colonisations of Berthelot's pipits to Madeira and Selvagens from the Canary Islands, ca. 8500 years ago (Spurgin *et al.* 2014; Armstrong *et al.* 2018). Diversity at both the MHC and toll-like receptors (TLRs; pathogen-recognition genes of the innate immune system) was found to decrease significantly during these bottlenecks, although new MHC variants appear to have been rapidly regenerated through gene conversion (Spurgin *et al.* 2011), and some functional variation has been maintained at TLR loci (González-Quevedo *et al.* 2015). Pathogen community composition and infection rates are spatially varied throughout this species (Illera *et al.* 2008; Spurgin *et al.* 2012; González-Quevedo *et al.* 2014), potentially facilitating balancing selection through spatial variation in pathogen-mediated selection. Therefore, this is an excellent system in which to investigate the forces acting upon, and the resulting patterns of, functional genetic diversity across bottlenecked populations.

In this study, our aims were as follows: 1) to characterise genetic variation at AvBD loci in Berthelot's pipit and its sister species, the tawny pipit, and look for potential signatures of balancing selection maintaining genetic variation; 2) to compare genetic diversity at AvBDs with genome-wide measures of neutral variation to assess whether adaptive variation at AvBD loci is maintained through bottlenecks, and; 3) to test for signatures of divergent selection at AvBD loci between the different archipelagos inhabited by Berthelot's pipit, indicative of divergent local adaptation acting upon AvBDs. We will discuss these results in the context of the evolution of AvBDs, and of the general processes acting to retain genetic variation in small and bottlenecked populations.

3.3. Methods

3.3.1. Sample collection

Berthelot's pipits were sampled on 12 islands throughout their range (Fig. 3.1), across three archipelagos: the Canary Islands (Graciosa, Lanzarote, Fuerteventura, Gran Canaria, Tenerife, La Gomera, La Palma, El Hierro); Selvagens (Selvagem Grande); and Madeira (Madeira, Porto Santo, Deserta Grande). As reported by Illera *et al.* (2007), Spurgin *et al.* (2012) and González-Quevedo *et al.* (2014), samples were collected in April 2005 (Selvagens), January–March 2006 (Canary

Islands), September 2006 (Madeiran archipelago), January–April 2009 (all islands), and January–April 2011 (Tenerife). Berthelot's pipits collected on Tenerife from the alpine plateau of El Teide at altitudes > 2000 m were characterised as a separate population to coastal lowland pipits. Birds were caught in spring traps baited with *Tenebrio molitor* larvae. Blood samples (ca. 50 µl) were collected by brachial venipuncture and stored in 100% ethanol in screw-top eppendorf tubes at room temperature. Tawny pipit samples were obtained from Spain, Morocco and Mauritania following the same procedure. DNA was extracted from blood samples using a salt extraction technique (Richardson *et al.* 2001). We confirmed the quality of the extractions by visualising the genomic DNA after electrophoresis on 1% agarose gels containing ethidium bromide, and confirmed that the DNA was amplifiable by performing a sexing PCR using the primers and conditions described in Griffiths *et al.* (1998).

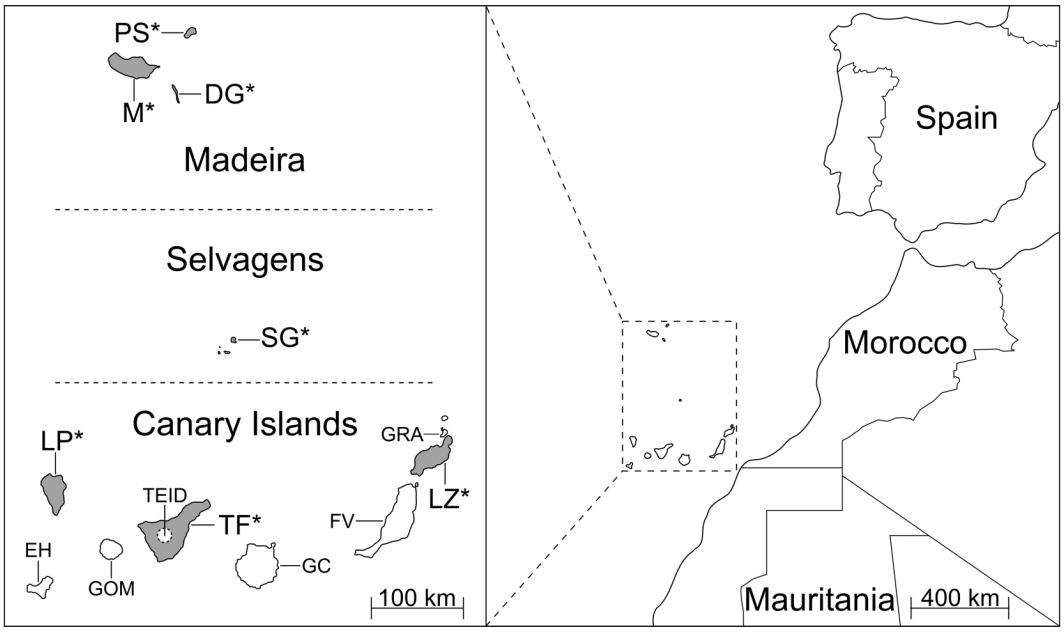


Figure 3.1 Map of Berthelot's pipit and tawny pipit sample locations. Berthelot's pipits were sampled from the Canary Islands (EH = El Hierro, LP = La Palma, GOM = La Gomera, TF = Tenerife, TEID = El Teide, GC = Gran Canaria, FV = Fuerteventura, LZ = Lanzarote, GRA = Graciosa), Selvagens (SG = Selvagem Grande) and Madeira (M = Madeira, PS = Porto Santo, DG = Deserta Grande). All populations were included in the preliminary study; islands denoted with * and shaded grey were used for further sequencing. Tawny pipits were sampled in Spain, Morocco and Mauritania.

3.3.2. Preliminary AvBD study

A preliminary study was carried out to screen AvBD loci for the presence of polymorphisms. We used primers described by Hellgren & Sheldon (2011) to amplify AvBD2, 4 and 7–13 in 15

(AvBD2, 10–13) or 20 (AvBD4, 7–9) Berthelot's pipits. To cover the geographic range of the species, 1–4 samples were selected per population (Supplementary Table S3.1). In addition, we amplified each locus in 12 tawny pipit samples (from Spain *n* = 6; Mauritania *n* = 4; Morocco *n* = 2). DNA was amplified using 5 µl TopTaq Master Mix (Quiagen), 2 µl H_2O, 0.5 µl each of the forward and reverse primers at 10 µM, and 1 µl genomic DNA (ca. 30 ng/µl). Annealing temperatures were either 55°C (AvBD10 Berthelot's pipits; AvBD13 tawny pipits), 61°C (AvBD7–8 and 11), 63°C (AvBD13 Berthelot's pipits; AvBD10 tawny pipits) or 65°C (AvBD4 and 9). The following thermal profile was used: 94°C for 5 minutes; 35 cycles of 94°C for 30 seconds, annealing with the above temperatures for 30 seconds, 72°C for 45 seconds; 72°C for 10 minutes. 5 µl PCR product was cleaned with 0.1 µl ExoI, 0.2 µl FastAP (Thermo Scientific) and 4.7 µl H_2O at 37°C for 15 minutes and 85°C for 15 minutes, then sequenced with Sanger sequencing. SNPs were called through visual assessment of chromatograms in FinchTV (www.geospiza.com/finchtv), and accepted if they were confirmed across multiple PCRs in multiple individuals.

3.3.3. AvBD primer design and additional sequencing

A number of the primers designed by Hellgren & Sheldon (2011) are positioned within the coding sequence for the AvBD mature peptide, preventing sequencing of the full gene. We therefore designed new primers to capture the complete mature peptide for loci which initial tests showed variation existed in Berthelot's pipits (AvBD8, 10, 11), and for loci which failed to amplify with the initial primers (AvBD2, 12; see Results). For each of our selected AvBD loci, we downloaded zebra finch mRNA sequences published in Hellgren & Ekblom (2010) from Genbank, and used the amino acid sequence alignments in Hellgren & Ekblom (2010) to trim sequences to the mature peptide. We then searched for these sequences in the NCBI BLAST database of whole-genome shotgun contigs, to obtain sequences spanning either side of the mature peptide region from a range of passerines with the best matches to the zebra finch sequence (Supplementary Table S3.2). Alignments for each locus were created in BioEdit v7.2.5 (Hall 1999). We visually inspected aligned sequences to find regions which showed high levels of sequence conservation across passerines. Primers were then designed within these regions, using the sequence from zebra finch or Atlantic canary, which had the lowest divergence times from Berthelot's pipit (determined using www.timetree.org; Supplementary Table S3.2). We used a combination of PerlPrimer v1.1.21 (Marshall 2004) and Primer3 (Rozen & Skaletsky 1999) to design primers with compatible melting temperatures and low dimerisation. We tested for primer specificity with NCBI Primer-BLAST (Ye *et al.* 2012) and accepted primer pairs that gave only the correct BLAST hit per species. Primers were tested on DNA samples that had previously successfully amplified

AvBD loci with existing primers (Hellgren & Sheldon 2011), using the same PCR conditions and sequencing protocols as above. For AvBD10 and 11, it was possible to sequence the entire mature peptide using just the forward primer. Final primer pairs, annealing temperatures and sequencing directions for each locus are given in Table 3.1.

Table 3.1 Details of new AvBD primer pairs. Primers designed by Hellgren & Sheldon (2011) are indicated by * following the primer name. Annealing temperatures (T_a °C) are given for Berthelot's pipit (first) and tawny pipit (second). Amplicon length (AL) includes primer sequences. Mature peptide length (MPL) is the length of the amplicon trimmed to the mature peptide sequence. Sequencing direction (Dir) required to achieve 100% coverage of the mature peptide is either forward (F) or forward and reverse (F&R).

Locus	Primer	Sequence	T_a °C	AL (bp)	MPL (bp)	Dir
AvBD2	AvBD2F2mat*	TCA TTG CAG GTT TGT CTT CG	56, 55	153	105	F&R
	AvBD2R1C	TGC AGA TGA GCC TTG GT				
AvBD8	AvBD8F2*	TTA CTC TAC CCA GGT TTC CTG C	56, 55	169	114	F&R
	AvBD8R1C	ATC CAT CAC ATC CTG GAC TC				
AvBD10	AvBD10F1C	TAC CTC TCA TCA TTT CAG GCA G	56, 57	213	114	F
	AvBD10R2C	CCC TGG GAT GTT TTG TCA CC				
AvBD11	AvBD11F2C	CCA TGA GAA TGG ACT GAT CCT	61, 59	273	117	F
	AvBD11R1C	TGA GTT CAG CCA TTT CCC				
AvBD12	AvBD12F2C	AAT TTG CTG CTC CCT TGT CC	58, 62	188	114	F&R
	AvBD12R2C	GGC TCC TCT GGA AAC CTT CA				

For sequencing at each of the AvBD loci, we selected 15 Berthelot's pipits per population from the following islands: Tenerife, Lanzarote and La Palma (Canary Islands); Madeira, Porto Santo and Deserta Grande (Madeiran archipelago); and Selvagem Grande (Selvagens archipelago). Samples were selected from different geographical locations within each population to capture island-level variation, and to minimise the chance of sequencing related individuals. We also included 15 tawny pipits (Spain *n* = 9; Mauritania *n* = 4; Morocco *n* = 2). All samples were sequenced at AvBD2, 8, and 10–12. We also used the 12 tawny pipit sequences obtained in the preliminary study for AvBD7 and 13, and sequenced additional samples to give a total of 15 tawny pipit samples at these loci. Where forward and reverse sequences overlapped, SNPs were verified by sequencing both the forward and the reverse directions. Sequencing was repeated across multiple PCRs to confirm SNPs when only one homozygote was present, and when the minor allele was present in only one sample. Accuracy of SNP calling was assessed by independent calling of a subset of sequences by an additional person.

3.3.4. RAD-seq libraries

We used two previously constructed restriction-site associated DNA sequencing (RAD-seq) datasets (https://doi.org/10.5061/dryad.9642b; Armstrong *et al.* 2018) for comparisons between AvBD and genome-wide variation. The "All pipits" RAD dataset consisted of 20 Berthelot's pipits per island and 16 tawny pipits. This dataset used RAD loci that were present in 100% of individuals, with loci removed if they they contained multiallelic SNPs. Here, we used a subset of this dataset that excluded RAD loci that were monomorphic in Berthelot's pipits, and kept loci that were 100–125 bp long (thus comparable to the length of the mature AvBD peptides: see Table 3.1), resulting in 850 loci with 5433 SNPs across Berthelot's and tawny pipits. In addition, we removed samples that were not from the eight populations sequenced at AvBD loci. This enabled us to perform meaningful comparisons of nucleotide diversity and F_{ST} (see below) between polymorphic AvBD and RAD loci.

The "Berthelot's" RAD dataset contained 20 Berthelot's pipits per island, with loci that were successfully genotyped in ≥ 90% of individuals. Further filtering produced a dataset with one SNP per locus (the SNP with the highest minor allele frequency of the SNPs in that locus), totalling 2745 SNPs.

3.3.5. Analysis

AvBD haplotypes were phased in DnaSP v6 (Librado & Rozas 2009). Samples with a phasing certainty < 90% were excluded from further analysis (see Table 3.2 for final sample sizes). We used PopART v.17 (Leigh & Bryant 2015) to make median-joining haplotype networks for each AvBD. We constructed a phylogenetic tree showing the evolutionary relationships between AvBD haplotypes in MEGA v7.1 (Kumar *et al.* 2016), with one *Taeniopygia guttata* sequence per AvBD locus included, obtained from the NCBI BLAST database. The most suitable model of sequence evolution was determined by MEGA to be the Kimura 2-parameter model (Kimura 1980) with invariant sites allowed, and this model was used to build the phylogeny. We retained all sites for analysis, to account for codon insertions in some AvBD loci. A total of 120 sites were analysed.

Table 3.2 Genetic variation at AvBD loci in Berthelot's pipit archipelagos (C = Canary Islands, M = Madeira, S = Selvagens) and in tawny pipits. *S* = number of variable sites; *h* = number of haplotypes; π = nucleotide diversity × 10^3; AA = number of amino acid variants; H_d = haplotype diversity; H_o/H_e = observed/expected heterozygosity (Hardy-Weinberg Equilibrium test); F_o/F_e = observed/expected homozygosity (Ewens-Watterson test); *D* = Tajima's *D*. H_o, F_o, and *D* values followed by * indicate a significant deviation from H_e, F_e, or zero, respectively, at $P < 0.05$.

Locus	Population (*n*)	*S*	*h*	π	AA	H_d	H_o	H_e	F_o	F_e	*D*
AvBD2	Berthelot's pipits (105)	1	2	2.9	2						
	C (45)	1	2	4.5	2	0.47	0.53	0.48	0.53	0.81	1.65
	M (45)	1	2	0.6	2	0.07	0.07	0.07	0.94	0.81	-0.78
	S (15)	1	2	1.2	2	0.13	0.13	0.13	0.88	0.76	-0.76
	Tawny pipits (14)	3	3	13.0	2	0.67	0.57	0.67	0.36*	0.59	1.85
AvBD4	Berthelot's pipits (17)	0	1	0.0	1						
	Tawny pipits (11)	0	1	0.0	1						
AvBD7	Berthelot's pipits (20)	0	1	0.0	1						
	Tawny pipits (14)	2	3	2.2	2	0.20	0.21	0.20	0.80	0.59	-1.24
AvBD8	Berthelot's pipits (105)	5	4	2.5	4						
	C (45)	5	4	5.6	4	0.30	0.31	0.30	0.70	0.56	-0.76
	M (45)	0	1	0.0	1						
	S (15)	0	1	0.0	1						
	Tawny pipits (14)	9	10	10.7	5	0.64	0.57	0.63	0.39	0.18	-1.50
AvBD9	Berthelot's pipits (20)	0	1	0.0	1						
	Tawny pipits (12)	0	1	0.0	1						
AvBD10	Berthelot's pipits (104)	3	5	4.2	1						
	C (44)	3	5	6.8	1	0.53	0.52	0.53	0.47	0.48	0.54
	M (45)	0	1	0.0	1						
	S (15)	2	3	6.1	1	0.52	0.33*	0.51	0.50	0.59	0.79
	Tawny pipits (13)	6	6	11.3	3	0.81	0.77	0.81	0.22	0.32	-0.53
AvBD11	Berthelot's pipits (105)	2	3	2.4	2						
	C (45)	2	3	3.7	2	0.42	0.29*	0.42	0.58	0.67	0.18
	M (45)	1	2	0.4	1	0.04	0.04	0.04	0.96	0.81	-0.91
	S (15)	1	2	3.2	1	0.37	0.20	0.37	0.64	0.76	0.73
	Tawny pipits (9)	8	8	18.4	2	0.89	0.22*	0.89	0.16	0.20	-0.25
AvBD12	Berthelot's pipits (105)	2	3	0.7	2						
	C (45)	2	3	1.5	2	0.17	0.18	0.17	0.83	0.67	-0.89
	M (45)	0	1	0.0	1						
	S (15)	0	1	0.0	1						
	Tawny pipits (15)	1	2	1.1	1	0.13	0.00*	0.13	0.88	0.76	-0.76
AvBD13	Berthelot's pipits (15)	0	1	0.0	1						
	Tawny pipits (15)	1	2	0.9	1	0.07	0.07	0.07	0.94	0.76	-1.15

We calculated a range of summary statistics from each AvBD locus. Haplotype diversity and Tajima's *D* were calculated in DnaSP. Tajima's *D* detects deviations from neutral evolution, with values around zero signifying neutral evolution, whereas positive or negative values indicate balancing or purifying selection, respectively (Tajima 1989). Observed and expected heterozygosity, and deviations from Hardy-Weinberg equilibrium, were calculated at both the population- and archipelago-level in Arlequin (Excoffier & Lischer 2010) using an exact test with 1,000,000 steps in the Markov chain and a burn-in of 100,000 steps.

We used the Ewens-Watterson test in Arlequin to detect deviations from mutation-drift equilibrium in each Berthelot's pipit archipelago and in tawny pipits. This test uses the number of alleles within a sample to simulate the expected homozygosity under neutrality. An observed homozygosity significantly lower than the simulated expected homozygosity is indicative of balancing selection (Watterson 1978). This was carried out at the archipelago level to increase sample size; however, low-level population structuring within archipelagos (Armstrong *et al.* 2018) may weaken the ability to detect balancing selection. We averaged the simulated expected homozygosity and *P* values across 20 repeats of the test.

We calculated nucleotide diversity (π) at AvBD and RAD loci using VCFtools (Danecek *et al.* 2011). We used π_{LOCUS} as a measure of per-locus nucleotide diversity, by averaging π across every site in each AvBD locus, including invariable sites. In addition, we calculated the average nucleotide diversity at nonsynonymous SNPs (π_{NS_SNP}) per AvBD locus. To enable comparisons with genome-wide variation, we calculated π_{LOCUS} for each RAD locus as above. However, for π_{NS_SNP} we included all SNPs within each RAD locus, as the majority of RAD loci are not expected to be within coding regions and therefore cannot be classified as synonymous or nonsynonymous.

Pairwise F_{ST} was calculated in PLINK 1.9 (Chang *et al.* 2015) between each pair of Berthelot's pipit populations, and separately between Berthelot's and tawny pipits, at every SNP in the AvBD loci and in the "All pipits" RAD dataset. The mean F_{ST} for SNPs within each locus (AvBD) or across all SNPs (RAD) was taken. We then used one-tailed partial Mantel tests with 100,000 permutations, performed using the Ecodist package in R (Goslee & Urban 2007), to correlate AvBD F_{ST} with genome-wide RAD F_{ST} between Berthelot's pipit populations. We included "archipelago comparison" (eg. Canary Islands vs. Canary Islands, Canary Islands vs. Selvagens etc.) as a categorical variable in the partial Mantel test, to control for within- vs. between-archipelago effects on F_{ST}.

We tested for signatures of divergent selection at the AvBD loci across Berthelot's pipit archipelagos by including AvBD SNPs into an EigenGWAS analysis (Chen *et al.* 2016) performed on the "Berthelot's" RAD dataset. This was a repeat of the EigenGWAS analysis performed by

Armstrong *et al.* (2018), on a subset of the "Berthelot's" dataset for individuals that had both AvBD and RAD sequences (Canary Islands *n* = 38, Selvagens *n* = 13, Madeira *n* = 37), with AvBD SNPs added to the dataset. This allowed us to test the hypothesis that, controlling for genome-wide variation, AvBD SNPs are likely to be outliers under divergent selection.

3.4. Results

We performed preliminary sequencing to detect AvBD polymorphisms. Using the primer sequences designed by Hellgren & Sheldon (2011), we sequenced Berthelot's pipits from each Macaronesian population, and tawny pipits, at AvBD4, 7–11, and 13 (AvBD2 and 12 failed to amplify; Supplementary Table S3.1). We found variation in Berthelot's pipits and tawny pipits in AvBD8, 10 and 11. AvBD7 and 13 were monomorphic in Berthelot's pipits and polymorphic in tawny pipits. AvBD4 and 9 were monomorphic in both species.

We designed new primers to amplify the entire mature peptide regions of AvBD2, 8 and 10–12 (Table 3.1), which we then used to sequence 105 Berthelot's pipits across seven populations, and 15 tawny pipits. Haplotype networks (Fig. 3.2) indicated that all AvBD loci share at least one haplotype between Berthelot's pipits and tawny pipits. We found two shared haplotypes at AvBD12. In all loci except AvBD10, the most common haplotype in Berthelot's pipits was also found in tawny pipits. The majority of haplotypes differed by just one mutation. Using a maximum-likelihood phylogenetic tree, AvBD haplotypes consistently clustered by locus (Supplementary Fig. S3.1), supporting evolutionary independence between these genes.

Genetic diversity at each of the AvBD loci is summarised in Table 3.2. We found 1–5 alleles per locus in Berthelot's pipits, and 2–10 alleles per locus in tawny pipits. These alleles encoded 1–4 amino acid variants in Berthelot's pipits, and 1–5 in tawny pipits. AvBD10 was monomorphic in Madeira, and AvBD8 and 12 were monomorphic in both Madeira and Selvagens. The Canary Islands and Selvagens were out of Hardy-Weinberg equilibrium at AvBD11 and 10, respectively, both with an excess of homozygosity, though each individual island within the Canary Islands was in Hardy Weinberg equilibrium at AvBD11 (Table 3.3). In the tawny pipits, both AvBD11 and 12 were out of Hardy Weinberg equilibrium, with significantly less heterozygosity than expected. We did not detect any significant deviations from neutrality in the Ewens-Watterson test within Berthelot's pipit archipelagos, though at AvBD2 in the Canary Islands, observed homozygosity (0.53) was considerably lower than expected homozygosity (0.81). This is suggestive of balancing selection, though the difference was not statistically significant. In the tawny pipits, however, the same trend was found at AvBD2, with a significant deviation between observed (0.36) and expected (0.59) homozygosity. Tajima's *D* values in Berthelot's pipit

archipelagos and in tawny pipits ranged from -1.5 to 1.85. The highest Tajima's *D* values were found for AvBD2, in the Canary Islands (1.65) and tawny pipits (1.85), which may indicate balancing selection at this locus; however, no Tajima's *D* tests showed significant differences from zero (all $P > 0.05$).

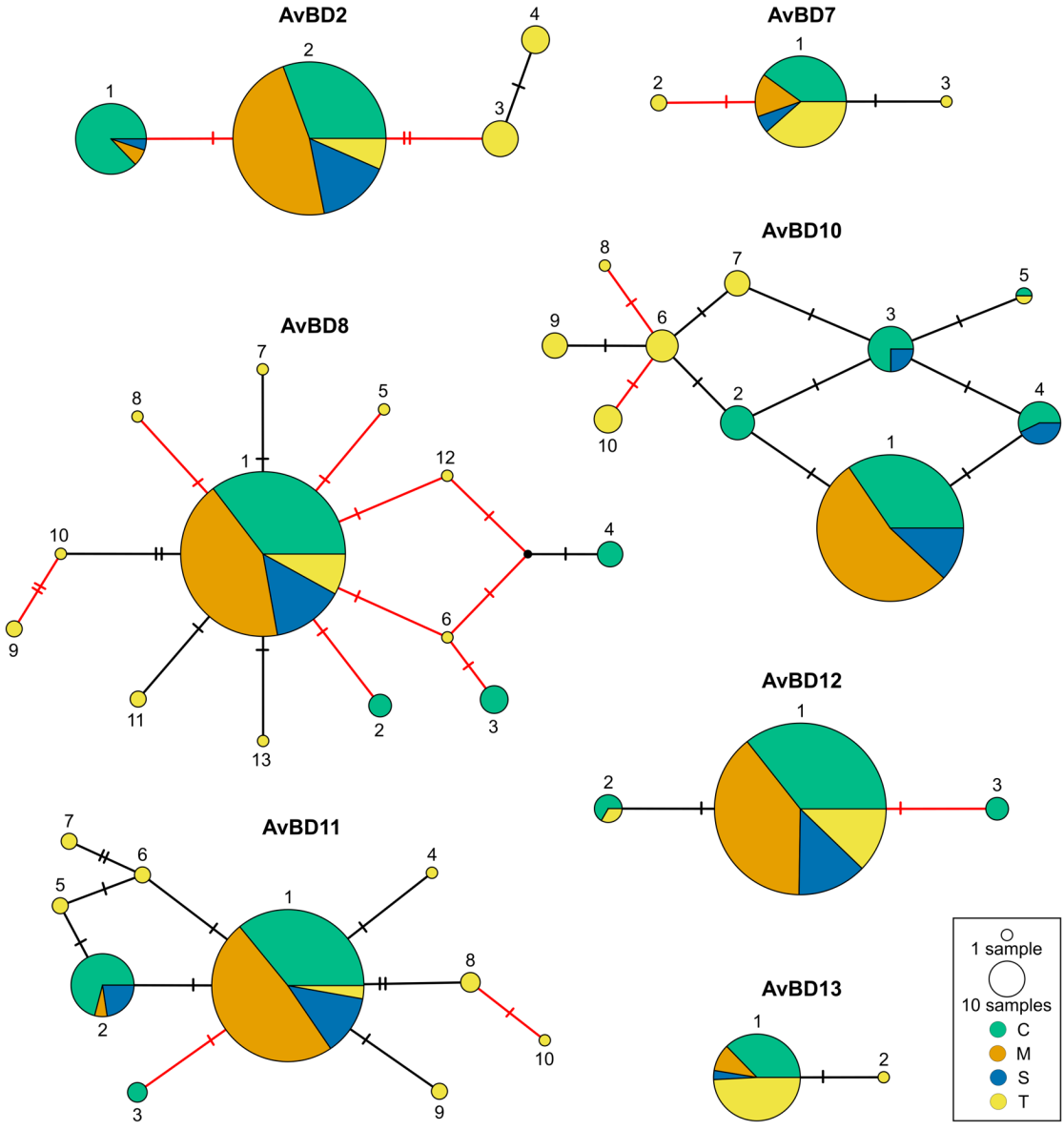


Figure 3.2 AvBD haplotype networks. Coloured circles represent haplotypes, shaded by archipelago. Synonymous mutations are indicated by black lines joining haplotypes; red lines show nonsynonymous mutations. Number of hatches on lines represents the number of nucleotide differences between haplotypes. Black circles represent inferred alleles. Size of circles represents sample size. C = Canary Islands, M = Madeira, S = Selvagens, T = tawny pipits.

Table 3.3 AvBD observed (H_o) and expected (H_e) heterozygosity per population. Values followed by * indicate a significant deviation from Hardy-Weinberg equilibrium at $P < 0.05$. Monomorphic populations are excluded from the table.

Locus	Archipelago	Population	H_o	H_e
AvBD2	Canary Islands	Lanzarote	0.47	0.48
		Tenerife	0.47	0.37
		La Palma	0.67	0.51
	Madeira	Madeira	0.13	0.13
		Porto Santo	0.07	0.07
	Selvagens	Selvagem Grande	0.13	0.13
AvBD8	Canary Islands	Lanzarote	0.33	0.30
		Tenerife	0.33	0.36
		La Palma	0.27	0.25
AvBD10	Canary Islands	Lanzarote	0.67	0.59
		Tenerife	0.67	0.64
		La Palma	0.21	0.33
	Selvagens	Selvagem Grande	0.33*	0.51
AvBD11	Canary Islands	Lanzarote	0.27	0.40
		Tenerife	0.40	0.51
		La Palma	0.20	0.37
	Madeira	Madeira	0.07	0.07
		Porto Santo	0.07	0.07
	Selvagens	Selvagem Grande	0.20	0.37
AvBD12	Canary Islands	Lanzarote	0.13	0.13
		Tenerife	0.33	0.30
		La Palma	0.07	0.07

Patterns of allelic and amino acid richness followed a general trend of highest richness in the tawny pipits, intermediate levels in the Canary Islands, and lowest in Madeira and Selvagens (Fig. 3.3), as expected based on population size and colonisation history (Spurgin *et al.* 2014). AvBD2 was the only locus to maintain amino acid variation in Madeira and Selvagens, with the same two amino acid variants found across all Berthelot's pipit archipelagos (Fig. 3.2). These two variants were found at high frequency in the Canary Islands (0.38 and 0.62 for AvBD2_1 and AvBD2_2, respectively), whereas AvBD2_1 was present at low frequency in Madeira (0.03) and Selvagens (0.07). Allele frequencies per island are shown in Table 3.4.

We calculated nucleotide diversity at AvBD and RAD loci averaged across the whole locus (π_{LOCUS}) and averaged across nonsynonymous (AvBD) or all (RAD) SNPs (π_{NS_SNP}) for each AvBD locus and for loci in the "All pipits" RAD dataset. In line with allelic and amino acid richness, highest levels of π_{LOCUS} and π_{NS_SNP} were found in tawny pipits, and lowest in Madeira and Selvagens (Fig. 3.4). Compared with genome-wide patterns in the "All pipits" dataset, AvBD2, 8, 10 and 11 had

relatively high π_{LOCUS} in tawny pipits and the Canary Islands. The same was found for AvBD10 and 11 in Selvagens (Fig. 3.4A). Results were more mixed for nonsynonymous SNPs; high π_{NS_SNP} relative to RAD loci across all groups of pipits was found only for AvBD2 (Fig. 3.4B; Table 3.5). AvBDs in Madeira and Selvagens were often monomorphic; however, where variation was present, nucleotide diversity in AvBD loci tended to be relatively high, with levels of π_{LOCUS} and π_{NS_SNP} at AvBD loci often falling within the top 5% of π_{LOCUS} and π_{NS_SNP} values found in RAD loci (Table 3.5).

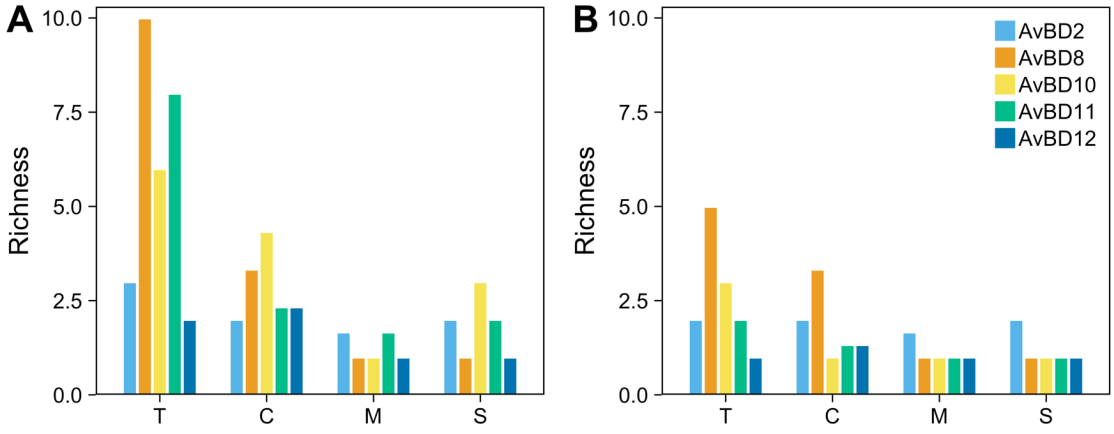


Figure 3.3 A) Allele and B) amino acid richness across five AvBD loci in Berthelot's and tawny pipits. Populations are ordered by estimated/effective population size: tawny pipits (T) estimated global population size > 4,000,000 (BirdLife International, 2016); effective population sizes estimated by Spurgin *et al*. (2014), Canary Islands (C) 4080, Madeira (M) 2930, Selvagens (S) 428. In the Canary Islands and Madeira, richness was calculated for each individual island population, with the mean richness for each archipelago shown here.

Table 3.4 AvBD2 allele frequencies in populations of Berthelot's pipits and in tawny pipits. As shown in Fig. 3.2, AvBD2_1, AvBD2_2 and AvBD2_3 encode different amino acid variants, whereas AvBD2_3 and AvBD2_4 differ by one synonymous SNP.

Archipelago	Island	AvBD2_1	AvBD2_2	AvBD2_3	AvBD2_4
Canary Islands	Lanzarote	0.37	0.63	0.00	0.00
	Tenerife	0.23	0.77	0.00	0.00
	La Palma	0.53	0.47	0.00	0.00
Madeira	Madeira	0.07	0.93	0.00	0.00
	Porto Santo	0.03	0.97	0.00	0.00
	Deserta Grande	0.00	1.00	0.00	0.00
Selvagens	Selvagem Grande	0.07	0.93	0.00	0.00
Tawny pipits		0.00	0.43	0.36	0.21

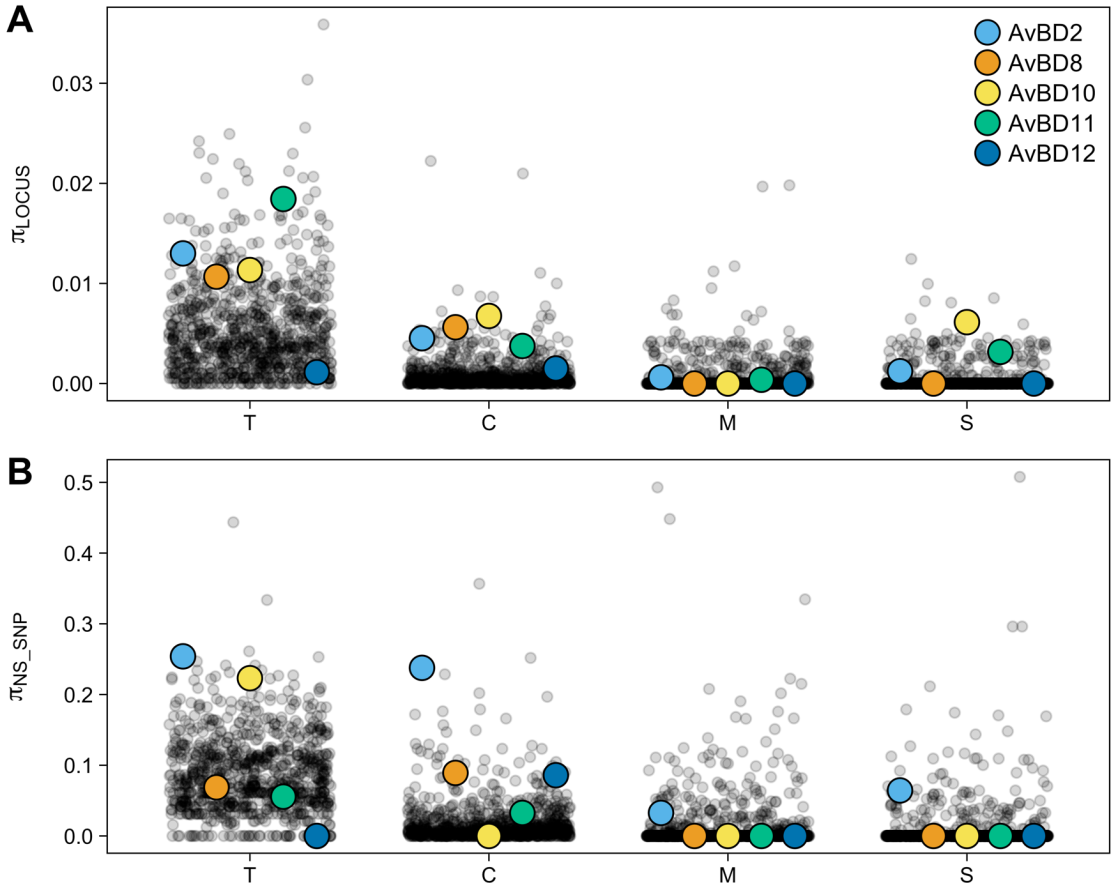


Figure 3.4 Nucleotide diversity (π) across Berthelot's pipits and tawny pipits at AvBD loci (coloured points) and RAD loci (black points). π was calculated as A) π_{LOCUS}, the mean of π across all sites along the AvBD or RAD locus; and B) π_{NS_SNP}, the mean of π at nonsynonymous SNPs (AvBD) or all SNPs (RAD) per locus. T = tawny pipits, C = Canary Islands, M = Madeira, S = Selvagens.

Table 3.5 AvBD nucleotide diversity (π) as a percentile of overall RAD nucleotide diversity per population. Nucleotide diversity is measured as π_{LOCUS} (π over the whole locus) or π_{NS_SNP} (π averaged across nonsynonymous AvBD SNPs/each SNP per RAD locus).

Locus	π	Tawny pipits	Canary Islands	Madeira	Selvagens
AvBD2	π_{LOCUS}	93.3	96.1	79.6	87.9
	π_{NS_SNP}	99.6	99.8	88.4	95.8
AvBD8	π_{LOCUS}	88.7	97.9	0.0	0.0
	π_{NS_SNP}	42.0	95.8	0.0	0.0
AvBD10	π_{LOCUS}	90.1	98.8	0.0	99.3
	π_{NS_SNP}	98.4	0.0	0.0	0.0
AvBD11	π_{LOCUS}	97.9	93.9	77.5	93.6
	π_{NS_SNP}	29.8	83.6	0.0	0.0
AvBD12	π_{LOCUS}	17.6	83.2	0.0	0.0
	π_{NS_SNP}	0.0	95.6	0.0	0.0

Within Berthelot's pipits, we found significant correlations between pairwise F_{ST} at AvBD loci and pairwise F_{ST} in the "All pipits" RAD dataset at AvBD2, 10 and 11, and at all AvBDs combined (Fig. 3.5). However, this result merely indicated that the slope of the relationship was significantly different from zero. The majority of pairwise population comparisons had higher F_{ST} in AvBDs than in RAD loci, indicating greater population structuring at AvBD loci relative to the genome average. The high F_{ST} at AvBD loci between Berthelot's pipit islands resulted from a loss of alleles in bottlenecked populations (AvBD8 and 12), changes in allele frequencies between populations (AvBD2), or both processes (AvBD10 and 11). F_{ST} between Berthelot's pipits and tawny pipits was considerably higher than within Berthelot's pipits at both AvBD and RAD loci, reflecting species-level divergence (Table 3.6; Fig. 3.5). The exception was AvBD12 with an F_{ST} of only 0.01 between Berthelot's pipits and tawny pipits. Only three haplotypes were found for this locus, with two of these shared between the two species (Fig. 3.2).

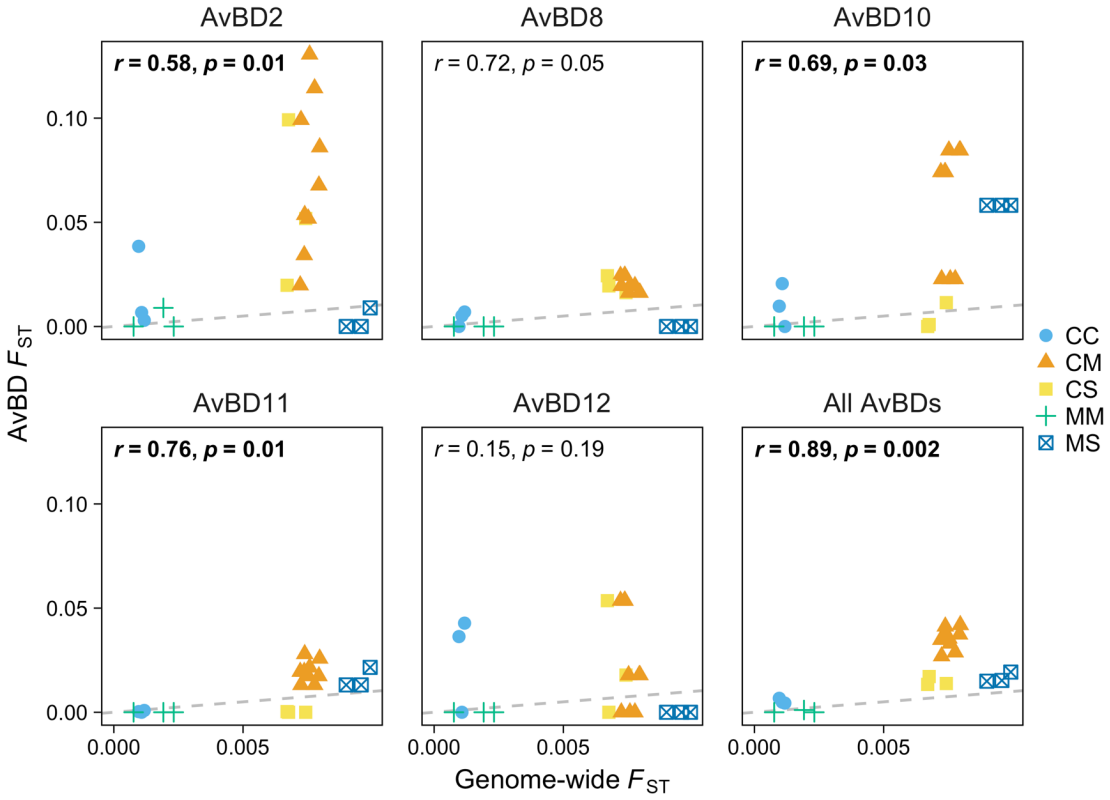


Figure 3.5 The relationship between pairwise F_{ST} at AvBD SNPs and genome-wide pairwise F_{ST} calculated from RAD SNPs in the 'All pipits' dataset. Points represent each pairwise comparison between populations, with colours and symbols indicating whether the comparison was within an archipelago (eg. CC = F_{ST} between two Canary Island populations) or between archipelagos (eg. CM = F_{ST} between a Canary Islands population and a Madeiran population). One-tailed partial Mantel tests were performed to test for correlations between AvBD and RAD pairwise F_{ST}. The correlation coefficient r and corresponding P value are shown in each plot; bold text indicates $P < 0.05$. Grey dashed line indicates the location of a 1:1 relationship.

Table 3.6 F_{ST} between Berthelot's pipits and tawny pipits at AvBD and RAD SNPs in the "All pipits" dataset. Berthelot's pipits were treated as a single population, to quantify divergence at the species-level.

Locus	F_{ST}
AvBD2	0.58
AvBD8	0.13
AvBD10	0.43
AvBD11	0.38
AvBD12	0.01
All AvBDs	0.31
RAD	0.20

We previously used EigenGWAS to detect outlier loci showing signatures of divergent selection between the archipelagos, using the "Berthelot's" RAD dataset (Armstrong *et al.* 2018). This analysis identifies loci that strongly contribute to population structuring, but is unable to detect uniform selection occurring throughout all populations. Here, we have added the AvBD SNPs which show variation in Berthelot's pipits to the "Berthelot's" dataset, to examine whether signatures of divergent selection can be detected at these SNPs. As there is a large degree of population structuring between the archipelagos, we used *P* values that have been corrected for genetic structure using genomic control (Devlin & Roeder 1999). The first eigenvector separates out Madeira from the Canary Islands and Selvagens, while the second eigenvector separates Selvagens from the Canary Islands and Madeira. We found no evidence of divergent selection along either eigenvector at any AvBD SNPs (Fig. 3.6). The lowest *P* value for AvBD SNPs was found in AvBD2 on the second eigenvector ($P = 0.07$; Table 3.7). While this was in the 8th percentile of *P* values for eigenvector 2, suggesting an above-average degree of structuring at this SNP between Selvagens and the other archipelagos, it did not reach statistical significance even before Bonferroni correction for multiple tests.

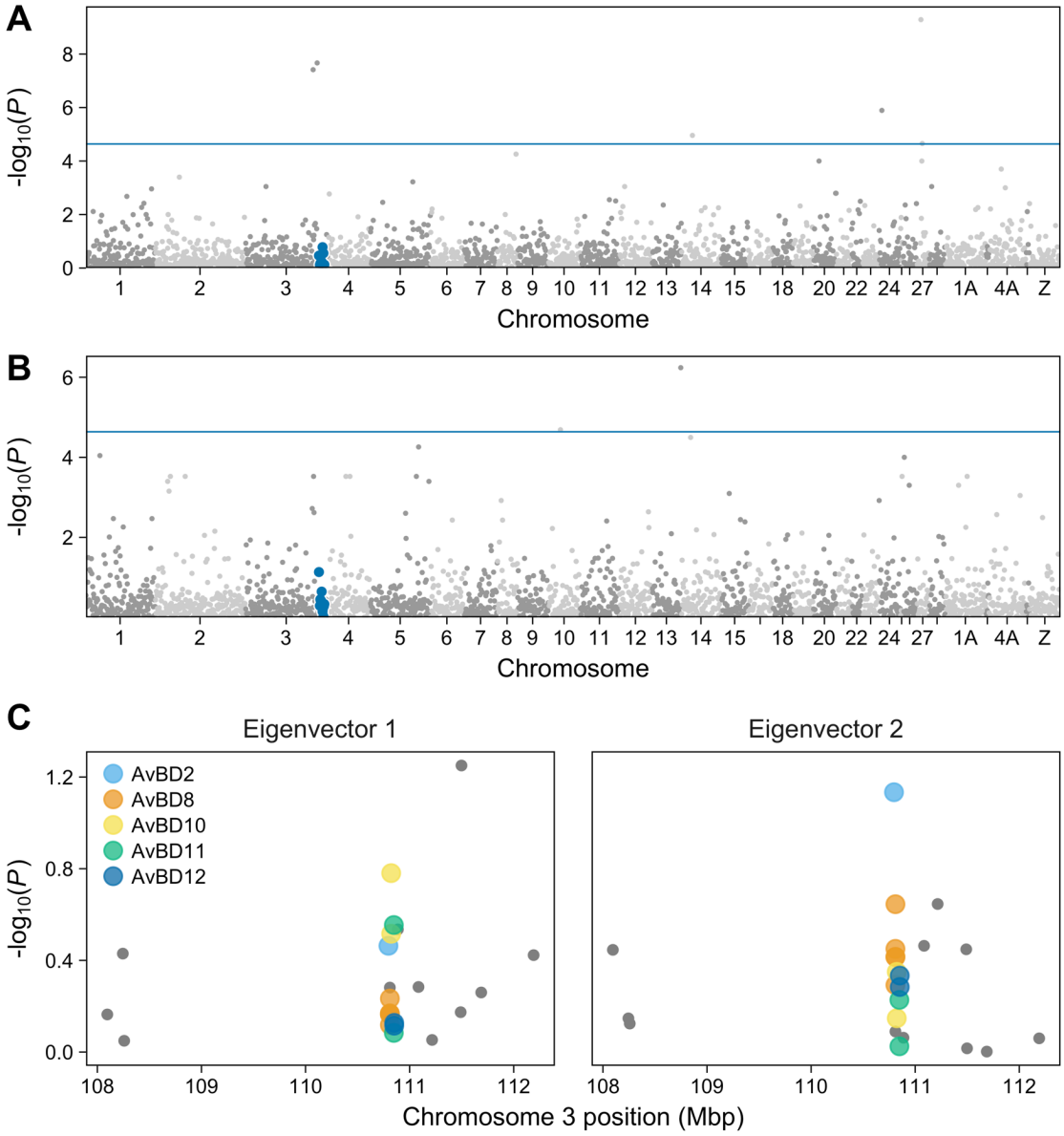


Figure 3.6 Manhattan plots of EigenGWAS analyses in Berthelot's pipits. Results are shown for eigenvector 1, separating Madeira from the Canary Islands and Selvagens (A), and eigenvector 2, separating Selvagens from the Canary Islands and Madeira (B). Each point represents a SNP ordered by chromosome position, from the "Berthelot's" RAD dataset (light and dark grey) or AvBDs (blue). The y-axis shows *P* values with genomic control applied to correct for underlying population structure. Blue line indicates the Bonferroni-corrected significance level. C) Close-up of region of Chromosome 3 containing AvBD SNPs for each eigenvector, coloured by AvBD locus.

Table 3.7 AvBD SNPs from the first and second eigenvector (EV1 & EV2) in an EigenGWAS analysis of Berthelot's pipit populations. For each SNP, we have calculated the percentile of the *P* value relative to all SNPs in each eigenvector. The Bonferroni-corrected significance threshold of 2.3×10^{-5} was met for SNPs below the 0.3 percentile for EV1, and below the 0.1 percentile for EV2. The genomic position of each SNP refers to the equivalent position in the zebra finch genome version 3.2.4 (Warren *et al.* 2010).

EV	SNP	Position	*P*	Percentile
EV1	AvBD2_01	Chr3:110796494	0.34	35.8
	AvBD8_03	Chr3:110810213	0.76	80.1
	AvBD8_04	Chr3:110810218	0.68	69.3
	AvBD8_05	Chr3:110810227	0.68	70.1
	AvBD8_07	Chr3:110810237	0.68	70.1
	AvBD8_10	Chr3:110810269	0.58	58.1
	AvBD10_07	Chr3:110822894	0.30	32.7
	AvBD10_02	Chr3:110822945	0.17	19.0
	AvBD11_01	Chr3:110848163	0.83	87.0
	AvBD11_09	Chr3:110848275	0.28	30.7
	AvBD12_01	Chr3:110851828	0.77	80.9
	AvBD12_02	Chr3:110851863	0.75	77.9
EV2	AvBD2_01	Chr3:110796494	0.07	8.9
	AvBD8_03	Chr3:110810213	0.51	52.2
	AvBD8_04	Chr3:110810218	0.35	35.5
	AvBD8_05	Chr3:110810227	0.38	38.5
	AvBD8_07	Chr3:110810237	0.38	38.5
	AvBD8_10	Chr3:110810269	0.23	22.1
	AvBD10_07	Chr3:110822894	0.45	44.5
	AvBD10_02	Chr3:110822945	0.71	75.0
	AvBD11_01	Chr3:110848163	0.59	60.9
	AvBD11_09	Chr3:110848275	0.94	95.3
	AvBD12_01	Chr3:110851828	0.52	53.3
	AvBD12_02	Chr3:110851863	0.46	46.5

3.5. Discussion

Our results indicate that genetic variation at AvBDs in Berthelot's pipits is overwhelmingly shaped by drift. We found lower nucleotide diversity and allelic/amino acid richness at AvBD loci in the Canary Islands relative to the tawny pipits, reflecting the initial colonisation of the Canary Islands by the mainland ancestor (Illera *et al.* 2007). Further reductions in richness and nucleotide diversity at AvBD loci were found in Madeira and Selvagens, which were each independently colonised from the Canary Islands much later (Spurgin *et al.* 2014). This pattern is in line with previous findings in Berthelot's pipits, which demonstrate reductions in genetic diversity at neutral loci (using both microsatellites and RAD loci) across the colonisation events

(Illera *et al.* 2007; Spurgin *et al.* 2014; Armstrong *et al.* 2018). We found no evidence of balancing selection at AvBD loci in Berthelot's pipits with either Tajima's *D* or the Ewens-Watterson test, which are used to detect deviations from neutral evolution (Watterson 1978; Tajima 1989). We assessed patterns of pairwise F_{ST} between Berthelot's pipit populations, and between Berthelot's pipits and tawny pipits, at AvBD and RAD loci. F_{ST} between the two species was higher at both AvBD and RAD loci than F_{ST} between Berthelot's pipit populations, as expected given the much longer divergence time between species (ca. 2.5 MYA; Voelker 1999) than between archipelagos (ca. 8,500 YA; Spurgin *et al.* 2014). Using partial Mantel tests, F_{ST} at most AvBD loci showed a significant correlation with F_{ST} across RAD loci, indicating that, for the most part, AvBD variation is following the same overall genome-wide trends. One locus, AvBD2, retained functional variation across all archipelagos, which could potentially be driven by ongoing balancing seelction in the Canary Islands.

Despite the overall effects of drift arising from the colonisation history of Berthelot's pipits, we found higher levels of nucleotide diversity at several AvBD loci relative to genome-wide RAD loci in the Canary Islands. This was shown both across whole loci (AvBD2, 8, 10, 11), and at nonsynonymous SNPs (AvBD2, 8, 11, 12), which could potentially result in functional differences in AvBD antimicrobial activity (Hellgren *et al.* 2010). These patterns were not observed in Madeira and Selvagens, where patterns of AvBD diversity were similar to that of the neutral RAD loci. This suggests that either contemporary or historical balancing selection may have had a role in shaping AvBD variation in Berthelot's pipits. Balancing selection prior to a population bottleneck can buffer against the loss of genetic variation during a bottleneck, resulting in signatures of past selection that are not entirely eroded by the effect of drift (Gavan *et al.* 2015; Gilroy *et al.* 2017). Although we do not know the genetic diversity of the ancestral population in Berthelot's pipits, current diversity in the mainland sister species, the tawny pipit, may reflect this to some extent. In the tawny pipit we found that several AvBDs showed elevated nucleotide diversity relative to genome-wide patterns. It is therefore possible that pathogen-mediated balancing selection, which has been detected in AvBDs and other antimicrobial peptides in other species (Tennessen & Blouin 2008; Halldórsdóttir & Árnason 2015; Chapman *et al.* 2016; Unckless *et al.* 2016), may be responsible for elevated genetic variation at AvBD loci in the tawny pipit, and that signatures of such past selection remain in the Berthelot's pipit populations.

We also found statistical evidence of balancing selection at one AvBD locus in tawny pipits. Using a Ewens-Watterson test, we found significantly lower homozygosity than expected under neutral evolution at AvBD2 in the tawny pipits. In addition, AvBD2 in the tawny pipits had the highest Tajima's *D* found in this study, although this was not significantly different from zero. Finally, it is worth noting that AvBD2 was the only locus to retain functional variation throughout

Berthelot's pipit populations, with the same two amino acid variants found in each archipelago. These findings should be interpreted with caution: although positive Tajima's *D* scores can indicate balancing selection, they can also be associated with decreasing population size (Tajima 1989), and estimates of Tajima's *D* calculated from small sample sizes (as is the case here for tawny pipits, *n* = 15) are prone to error (Subramanian 2016). Nonetheless, this combination of signatures of balancing selection in the tawny pipit and maintenance of functional variation across multiple bottlenecks in the Berthelot's pipit, is consistent with past selection shaping contemporary genetic diversity in a system characterised by high levels of genetic drift (Gilroy *et al.* 2017). We also cannot rule out the possibility that balancing selection in the Canary Islands at AvBD2 has gone undetected due to low statistical power. The two amino acid variants found across all Berthelot's pipit archipelagos were both found at high frequency in the Canary Islands (Table 3.4). One of these alleles was not detected in the tawny pipits, though due to the small sample size of this species, we cannot tell whether it exists at low frequencies, whether it has been lost since the divergence of Berthelot's and tawny pipits, or whether this allele originated in the Canary Islands prior to the colonisation of Madeira and Selvagens. Regardless, it is now present at high frequency in the Canary Islands, and it may be this factor that enabled it to be retained through subsequent bottlenecks in Madeira and Selvagens. This corresponds with a study of TLR4 dynamics in a bottlenecked population of water voles, where two alleles initially found at near equal frequency were still present following a severe bottleneck. Simulations of allele frequency dynamics suggested that loss of an allele became increasingly likely with greater differences between pre-bottleneck allele frequencies (Gavan *et al.* 2015).

Multiple empirical studies have found that if strong enough, balancing selection can counteract drift, and maintain immunogenetic variation through bottlenecks (Aguilar *et al.* 2004; Jarvi *et al.* 2004; Munguia-Vega *et al.* 2007; Gos *et al.* 2012; Oliver & Piertney 2012). However, drift is often reported to be the major evolutionary force influencing immune gene diversity in bottlenecked populations (Miller *et al.* 2010; Grueber *et al.* 2013; Sutton *et al.* 2013; González-Quevedo *et al.* 2015; Quéméré *et al.* 2015). While a lack of evidence of balancing selection in Berthelot's pipits could be attributed to strong genetic drift overwhelming balancing selection, it is also possible that pathogen-mediated balancing selection is inherently weaker on island populations relative to the mainland. Island biogeography theory predicts that smaller, more isolated islands will have less immigration and higher extinction, leading to lower species richness (MacArthur & Wilson 1967). Pathogen communities on islands have been shown to follow such biogeographic trends, with higher pathogen diversity on larger, less isolated islands (Orrock *et al.* 2011; Svensson-Coelho & Ricklefs 2011; Pérez-Rodríguez *et al.* 2013; Jean *et al.* 2016). This is also evident across Berthelot's pipit populations, with fewer pathogen strains found in smaller, more remote islands (Spurgin *et al.* 2012). Under pathogen-mediated balancing selection, host

populations are predicted to maintain high immunogenetic diversity to protect against diverse pathogen communities (Spurgin & Richardson 2010). If pathogen richness is low, then selection may be directional instead of balancing, favouring the most effective allele against the local pathogen regime (Hartmann *et al.* 2014). Thus it may be that in island populations, strong drift is accompanied by relaxed balancing selection; however, this hypothesis requires further investigation.

In line with drift being a major driver of AvBD diversity, pairwise F_{ST} between Berthelot's pipit populations was, for most AvBD loci, significantly correlated with genome-wide F_{ST} calculated from RAD loci. However, we often found higher F_{ST} at AvBD loci compared to RAD loci. Elevated levels of structure at AvBD loci could stem from differences between datasets, where individual candidate loci are being compared to RAD loci spread throughout the genome. However, we have taken steps to minimise the inherent sampling differences between the two datasets: we retained only RAD loci that were polymorphic in Berthelot's pipits, and that were 100–125 bp long (approximately the length of the AvBD mature peptide region). On closer inspection, especially high F_{ST} in AvBDs relative to RAD loci was often seen in comparisons between Canary Islands and Madeira, or Canary Islands and Selvagens. This could be suggestive of divergent selection for local pathogen regimes, as local adaptation is predicted to cause elevated F_{ST} at sites under selection relative to neutrally evolving loci (Charlesworth *et al.* 1997). This has been observed at MHC and TLR loci, where elevated F_{ST} relative to neutral loci is thought to have arisen from local adaptation to spatially divergent pathogen communities (Ekblom *et al.* 2007; Miller *et al.* 2010; Tschirren *et al.* 2011). We used an EigenGWAS analysis to test for divergent selection between Berthelot's pipit archipelagos at AvBD SNPs, though this did not find significant signatures of divergent selection in any AvBD loci, even before correcting for multiple testing.

Low genetic diversity can increase the vulnerability of a population to infection and epidemics (King & Lively 2012), and hinder the evolution of pathogen resistance (Barrett & Schluter 2008; Morris *et al.* 2015). Emerging infectious diseases can have dramatic impacts on survival and population persistence, especially in hosts which have not coevolved with the pathogen (Berger *et al.* 1998; Hawkins *et al.* 2006). A number of endemic Hawaiian bird species are now extinct as a result of the introduction of avian malaria (Warner 1968; Van Riper *et al.* 1986), though resistance or tolerance has rapidly evolved, likely as a result of standing immunogenetic variation (Woodworth *et al.* 2005; Sorci 2013). The low diversity at AvBD loci seen in Berthelot's pipits is similar to that found at AvBD and TLR loci in other bottlenecked island endemics (Grueber *et al.* 2013; Dalton *et al.* 2016; Gilroy *et al.* 2016). The low standing variation at AvBDs in Berthelot's pipits could increase the risk of local extinctions if faced with a novel pathogen.

However, as AvBDs gain broad antimicrobial activity by exploiting a fundamental feature common to many microbial pathogens (Zasloff 2002), low diversity at AvBDs might not be as detrimental as immune loci that detect more specific pathogen antigens or patterns, such as the MHC or TLRs. In the Canary Islands, we found two amino acid variants at high frequency at AvBD2; the same trend was also observed at this locus in an outbred population of great tits *Parus major*, where two functional alleles were present at near-equal frequencies (Hellgren 2015). We found no evidence of local adaptation at AvBD loci, even though it is likely that Berthelot's and tawny pipits will be encountering different microbial communities (Martiny *et al.* 2006). Instead, at nearly every AvBD locus, the two species shared the same major allele. Low spatial structuring of AvBD variation has been found in tree swallows *Tachycineta bicolor* (Schmitt *et al.* 2017b), and across global distributions of mallards (Chapman *et al.* 2016). In addition, sharing of alleles across large phylogenetic distances has been found in a range of AvBDs and other antimicrobial peptides (Hellgren & Sheldon 2011; Halldórsdóttir & Árnason 2015; Chapman *et al.* 2016; Unckless & Lazzaro 2016). Together, these suggest that the broad activity of antimicrobial peptides reduces the need for local adaptation to specific pathogen communities. Sufficient pathogen defence by AvBDs might therefore be achieved with a low number of highly effective alleles per locus. If so, low AvBD diversity may not pose a particularly high risk to population persistence.

Frequent gene duplication is a common evolutionary feature of many antimicrobial peptide gene families (Maxwell *et al.* 2003; Lee *et al.* 2005; Tennessen & Blouin 2007; Tang *et al.* 2018), including the AvBDs (Hellgren & Ekblom 2010; Chen *et al.* 2015). This appears to be an important mechanism for generating functional diversity at these genes: whilst individual loci may be under purifying selection to maintain a conserved structure and function (Chapman *et al.* 2016), relaxed selection pressure on duplicates facilitates rapid divergence (Duda *et al.* 2002; Semple *et al.* 2006; Hellgren & Ekblom 2010). It may therefore be possible that variation between AvBD genes is more important than variation within genes for maintaining diverse antimicrobial activity (Tennessen & Blouin 2007). Our phylogenetic tree and haplotype networks suggest that the AvBDs studied here are evolving independently of each other, with new haplotypes arising through point mutations. Mechanisms such as gene conversion, which have been shown to facilitate the generation of MHC diversity in Berthelot's pipits (Spurgin *et al.* 2011), do not appear to play a role in shaping genetic diversity at AvBD (this study) or TLR (González-Quevedo *et al.* 2015) loci in Berthelot's pipits.

Characterising immunogenetic variation throughout a range of gene families has important implications for predicting the adaptive potential of wild populations (Acevedo-Whitehouse & Cunningham 2006). Here, we have designed primers to sequence the entire mature peptide at

five AvBD loci which showed sequence variation in both Berthelot's pipits and tawny pipits. Though these have only been tested in two closely related pipit species, the primers were positioned within regions that showed high conservation across divergent passerine species (up to 66 MY estimated divergence time from Berthelot's pipit, Supplementary Table S3.2), so will likely be of utility for other passerine study systems.

In conclusion, our results suggest that genetic drift as a result of population bottlenecks is the key driver of AvBD variation in the Berthelot's pipit. This is in line with previous findings in this species, where immunogenetic variation at TLRs and the MHC have also decreased during bottlenecks. Successive colonisation events and population bottlenecks in Berthelot's pipits appear to have markedly reduced functional allelic variation in the Berthelot's pipit, compared to the tawny pipit, although there is some evidence for historical balancing selection at these loci. It is not yet clear whether low genetic diversity at AvBD loci has implications for the adaptability and long-term persistence of bottlenecked populations and species. Future research will reveal whether high intra-locus AvBD diversity is important to evolutionary potential and population persistence, or whether instead the broad antimicrobial activity of these genes, the widespread gene duplication events, and the diversity of other antimicrobial peptides, may reduce the need for high levels of within-locus variation.

3.6. References

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

Ageitos, J.M., Sánchez-Pérez, A., Calo-Mata, P. & Villa, T.G. (2017). Antimicrobial peptides (AMPs): Ancient compounds that represent novel weapons in the fight against bacteria. *Biochemical Pharmacology*, 133, 117–138.

Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D. & Wayne, R.K. (2004). High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences*, 101, 3490–3494.

Akira, S., Uematsu, S. & Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell*, 124, 783–801.

Alcaide, M., Muñoz, J., Martínez-de la Puente, J., Soriguer, R. & Figuerola, J. (2014). Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot *Fulica atra* (Aves: Rallidae). *Ecology and Evolution*, 4, 688–698.

Antcheva, N., Boniotto, M., Zelezetsky, I., Pacor, S., Falzacappa, M.V.V. & Crovella, S. *et al.* (2004). Effects of positively selected sequence variations in human and *Macaca fascicularis* β-defensins 2 on antimicrobial activity. *Antimicrobial Agents and Chemotherapy*, 48, 685–688.

Arbanasić, H., Duras, M., Podnar, M., Gomerčić, T., Ćurković, S. & Galov, A. (2014). Major histocompatibility complex class II variation in bottlenose dolphin from Adriatic Sea: inferences about the extent of balancing selection. *Marine Biology*, 161, 2407–2422.

Armstrong, C., Richardson, D.S., Hipperson, H., Horsburgh, G.J., Küpper, C. & Percival-Alwyn, L. *et al.* (2018). Genomic associations with bill length and disease reveal drift and selection across island bird populations. *Evolution Letters*, 2, 22–36.

Barrett, R.D.H. & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, 23, 38–44.

Berger, L., Speare, R., Daszak, P., Green, D.E., Cunningham, A.A. & Goggin, C.L. *et al.* (1998). Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proceedings of the National Academy of Sciences*, 95, 9031–9036.

Bollmer, J.L., Hull, J.M., Ernest, H.B., Sarasola, J.H. & Parker, P.G. (2011). Reduced MHC and neutral variation in the Galápagos hawk, an island endemic. *BMC Evolutionary Biology*, 11, 143.

Brouwer, L., Barr, I., Pol, M. van de, Burke, T., Komdeur, J. & Richardson, D.S. (2010). MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, 19, 3444–3455.

Brown, K.L. & Hancock, R.E.W. (2006). Cationic host defense (antimicrobial) peptides.

Cadwell, K., Niranji, S.S., Armstrong, V.L., Mowbray, C.A., Bailey, R. & Watson, K.A. *et al.* (2017). AvBD1 nucleotide polymorphisms, peptide antimicrobial activities and microbial colonisation of the broiler chicken gut. *BMC Genomics*, 18, 637.

Caseys, C., Stölting, K.N., Barbará, T., González-Martínez, S.C. & Lexer, C. (2015). Patterns of genetic diversity and differentiation in resistance gene clusters of two hybridizing European *Populus* species. *Tree Genetics & Genomes*, 11, 81.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7.

Chapman, J.R., Hellgren, O., Helin, A.S., Kraus, R.H.S., Cromie, R.L. & Waldenström, J. (2016). The evolution of innate immune genes: Purifying and balancing selection on β-defensins in waterfowl. *Molecular Biology and Evolution*, 33, 3075–3087.

Charlesworth, B., Nordborg, M. & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, 70, 155–174.

Chen, G.-B., Lee, S.H., Zhu, Z.-X., Benyamin, B. & Robinson, M.R. (2016). EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Heredity*, 117, 51–61.

Chen, H., Ma, M.Y., Sun, L., Fang, S.G. & Wan, Q.H. (2015). Genomic structure and evolution of beta-defensin genes in the golden pheasant and hwamei. *Science Bulletin*, 60, 679–690.

Cheng, D.-Q., Li, Y. & Huang, J.-F. (2014). Molecular Evolution of the Primate α-/θ-Defensin Multigene Family. *PLoS ONE*, 9, e97425.

Cheng, Y., Prickett, M.D., Gutowska, W., Kuo, R., Belov, K. & Burt, D.W. (2015). Evolution of the avian β-defensin and cathelicidin genes. *BMC Evolutionary Biology*, 15, 188.

Dalton, D.L., Vermaak, E., Smit-Robinson, H.A. & Kotze, A. (2016). Lack of diversity at innate immunity Toll-like receptor genes in the Critically Endangered White-winged Flufftail (*Sarothrura ayresi*). *Scientific Reports*, 6, 36757.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E. & DePristo, M.A. *et al.* (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.

Devlin, B. & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, 55, 997–1004.

Duda, T.F., Vanhoye, D. & Nicolas, P. (2002). Roles of diversifying selection and coordinated evolution in the evolution of amphibian antimicrobial peptides. *Molecular Biology and Evolution*, 19, 858–864.

Eizaguirre, C. & Lenz, T.L. (2010). Major histocompatability complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *Journal of Fish Biology*, 77, 2023–2047.

Eizaguirre, C., Yeates, S.E., Lenz, T.L., Kalbe, M. & Milinski, M. (2009). MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology*, 18, 3316–3329.

Ejsmond, M.J., Radwan, J. & Wilson, A.B. (2014). Sexual selection and the evolutionary dynamics of the major histocompatibility complex. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20141662.

Ekblom, R., Saether, S.A., Jacobsson, P., Fiske, P., Sahlman, T. & Grahn, M. *et al.* (2007). Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology*, 16, 1439–1451.

Excoffier, L. & Lischer, H.E.L. (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.

Frankham, R., Lees, K., Montgomery, M.E., England, P.R., Lowe, P.R. & Briscoe, D.A. (1999). Do population bottlenecks reduce evolutionary potential? *Animal Conservation*, 2, 255–260.

Froy, O. & Gurevitz, M. (2003). Arthropod and mollusk defensins – evolution by exon-shuffling. *Trends in Genetics*, 19, 684–687.

Garcia-Olmedo, F., Molina, A., Alamillo, J.M. & Rodriguez-Palenzuela, P. (1998). Plant defense peptides. *Biopolymers*, 47, 479–491.

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

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

Gilroy, D., Oosterhout, C., Komdeur, J. & Richardson, D.S. (2016). Avian β-defensin variation in bottlenecked populations: the Seychelles warbler and other congeners. *Conservation Genetics*, 17, 661–674.

Gilroy, D.L., Phillips, K.P., Richardson, D.S. & Oosterhout, C. van. (2017). Toll-like receptor variation in the bottlenecked population of the Seychelles warbler: computer

simulations see the 'ghost of selection past' and quantify the 'drift debt'. *Journal of Evolutionary Biology*, 30, 1276–1287.

González-Quevedo, C., Davies, R.G. & Richardson, D.S. (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.

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

Gos, G., Slote, T. & Wright, S.I. (2012). Signatures of balancing selection are maintained at disease resistance loci following mating system evolution and a population bottleneck in the genus *Capsella*. *BMC Evolutionary Biology*, 12, 152.

Goslee, S.C. & Urban, D.L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22, 1–19.

Griffiths, R., Double, M.C., Orr, K. & Dawson, R.J.G. (1998). A DNA test to sex most birds. *Molecular Ecology*, 7, 1071–1075.

Grueber, C.E., Knafler, G.J., King, T.M., Senior, A.M., Grosser, S. & Robertson, B. *et al.* (2015). Toll-like receptor diversity in 10 threatened bird species: relationship with microsatellite heterozygosity. *Conservation Genetics*, 16, 595–611.

Grueber, C.E., Wallis, G.P. & Jamieson, I.G. (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.

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

Halldórsdóttir, K. & Árnason, E. (2015). Trans-species polymorphism at antimicrobial innate immunity cathelicidin genes of Atlantic cod and related species. *PeerJ*, 3, e976.

Hancock, R.E. & Diamond, G. (2000). The role of cationic antimicrobial peptides in innate host defences. *Trends in Microbiology*, 8, 402–410.

Hartmann, S.A., Schaefer, H.M. & Segelbacher, G. (2014). Genetic depletion at adaptive but not neutral loci in an endangered bird species. *Molecular Ecology*, 5712–5725.

Hawkins, C.E., Baars, C., Hesterman, H., Hocking, G.J., Jones, M.E. & Lazenby, B. *et al.* (2006). Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. *Biological Conservation*, 131, 307–324.

Hellgren, O. (2015). Allelic variation at innate immune genes (avian β-defensins), within a natural population of great tits. *Journal of Avian Biology*, 46, 113–118.

Hellgren, O. & Ekblom, R. (2010). Evolution of a cluster of innate immune genes (β-defensins) along the ancestral lines of chicken and zebra finch. *Immunome Research*, 6, 3.

Hellgren, O. & Sheldon, B.C. (2011). Locus-specific protocol for nine different innate immune genes (antimicrobial peptides: β-defensins) across passerine bird species reveals within-species coding variation and a case of trans-species polymorphisms. *Molecular Ecology Resources*, 11, 686–692.

Hellgren, O., Sheldon, B.C. & Buckling, A. (2010). *In vitro* tests of natural allelic variation of innate immune genes (avian β-defensins) reveal functional differences in microbial inhibition. *Journal of Evolutionary Biology*, 23, 2726–2730.

Hess, C.M. & Edwards, S. (2002). The evolution of the major histocompatibility complex in birds. *BioScience*, 52, 423–431.

Higgs, R., Lynn, D.J., Cahalane, S., Alaña, I., Hewage, C.M. & James, T. *et al.* (2007). Modification of chicken avian β-defensin-8 at positively selected amino acid sites enhances specific antimicrobial activity. *Immunogenetics*, 59, 573–580.

Huang, Y., Li, Y., Burt, D.W., Chen, H., Zhang, Y. & Qian, W. *et al.* (2013). The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nature Genetics*, 45, 776–783.

Illera, J.C., Emerson, B.C. & Richardson, D.S. (2007). Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology*, 16, 4599–4612.

Illera, J.C., Emerson, B.C. & Richardson, D.S. (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.

Jarvi, S.I., Tarr, C.L., Mcintosh, C.E., Atkinson, C.T. & Fleischer, R.C. (2004). Natural selection of the major histocompatibility complex (*Mhc*) in Hawaiian honeycreepers (Drepanidinae). *Molecular Ecology*, 13, 2157–2168.

Jean, K., Burnside, W.R., Carlson, L., Smith, K. & Guégan, J.F. (2016). An equilibrium theory signature in the island biogeography of human parasites and pathogens. *Global Ecology and Biogeography*, 25, 107–116.

Jeffery, K.J. & Bangham, C.R. (2000). Do infectious diseases drive MHC diversity?

Jiggins, F.M. & Hurst, G.D.D. (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.

Jones, E.A., Cheng, Y., O'Meally, D. & Belov, K. (2017). Characterization of the antimicrobial peptide family defensins in the Tasmanian devil (*Sarcophilus harrisii*), koala (*Phascolarctos cinereus*), and tammar wallaby (*Macropus eugenii*). *Immunogenetics*, 69, 133–143.

Jones, J.D. & Dangl, J.L. (2006). The plant immune system. *Nature*, 444, 323–329.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16, 111–120.

King, K.C. & Lively, C.M. (2012). Does genetic diversity limit disease spread in natural host populations? *Heredity*, 109, 199–203.

Klüver, E., Schulz-Maronde, S., Scheid, S., Meyer, B., Forssmann, W.G. & Adermann, K. (2005). Structure-activity relation of human β-defensin 3: Influence of disulfide bonds and cysteine substitution on antimicrobial activity and cytotoxicity. *Biochemistry*, 44, 9804–9816.

Kumar, S., Stecher, G. & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870–1874.

Lai, Y. & Gallo, R.L. (2009). AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends in Immunology*, 30, 131–141.

Lande, R. (1988). Genetics and demography in biological conservation. *Science*, 241, 1455–1460.

Lee, W.H., Li, Y., Lai, R., Li, S., Zhang, Y. & Wang, W. (2005). Variety of antimicrobial peptides in the *Bombina maxima* toad and evidence of their rapid diversification. *European Journal of Immunology*, 35, 1220–1229.

Lehrer, R.I. & Ganz, T. (2002). Defensins of vertebrate animals. *Current Opinion in Immunology*, 14, 96–102.

Leigh, J.W. & Bryant, D. (2015). PopART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6, 1110–1116.

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

MacArthur, R.H. & Wilson, E.O. (1967). *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ.

Mangoni, M.L., Mcdermott, A.M. & Zasloff, M. (2016). Antimicrobial peptides and wound healing: biological and therapeutic considerations. *Experimental Dermatology*, 25, 167–173.

Marshall, O.J. (2004). PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics*, 20, 2471–2472.

Martiny, J.B., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A. & Green, J.L. *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4, 102–112.

Maxwell, A.I., Morrison, G.M. & Dorin, J.R. (2003). Rapid sequence divergence in mammalian beta-defensins by adaptive evolution. *Molecular Immunology*, 40, 413–421.

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

Miller, H.C., Allendorf, F.W. & Daugherty, C.H. (2010). Genetic diversity and differentiation at MHC genes in island populations of tuatara (*Sphenodon* spp.). *Molecular Ecology*, 19, 3894–3908.

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

Munguia-Vega, A., Esquer-Garrigos, Y., Rojas-Bracho, L., Vazquez-Juarez, R., Castro-Prieto, A. & Flores-Ramirez, S. (2007). Genetic drift vs. natural selection in a long-term small isolated population: major histocompatibility complex class II variation in the Gulf of California endemic porpoise (*Phocoena sinus*). *Molecular Ecology*, 16, 4051–4065.

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, M.K. & Piertney, S.B. (2012). Selection maintains MHC diversity through a natural population bottleneck. *Molecular Biology and Evolution*, 29, 1713–1720.

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

Orrock, J.L., Allan, B.F. & Drost, C.A. (2011). Biogeographic and ecological regulation of disease: prevalence of Sin Nombre Virus in island mice is related to island area, precipitation, and predator richness. *The American Naturalist*, 177, 691–697.

Pérez-Rodríguez, A., Ramírez, Á., Richardson, D.S. & Pérez-Tris, J. (2013). Evolution of parasite island syndromes without long-term host population isolation: parasite dynamics in Macaronesian blackcaps *Sylvia atricapilla*. *Global Ecology and Biogeography*, 22, 1272–1281.

Piertney, S.B. & Oliver, M.K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96, 7–21.

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

Reed, D.H. & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology*, 17, 230–237.

Richardson, D.S., Jury, F.L., 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, D.S., Komdeur, J., Burke, T. & Schantz, T. von. (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.

Rozen, S. & Skaletsky, H. (1999). Primer3 on the WWW for General Users and for Biologist Programmers. In: *Bioinformatics methods and protocols* (eds. Misener, S. & Krawetz, S.A.). Humana Press, Totowa, NJ, pp. 365–386.

Schmitt, C., Garant, D., Bélisle, M. & Pelletier, F. (2017a). Linking innate immunogenetic variation with phenotypic traits in a wild population of tree swallows, *Tachycineta bicolor*. *Biological Journal of the Linnean Society*, 121, 685–697.

Schmitt, C., Garant, D., Doyon, K., Bousquet, N., Gaudreau, L. & Bélisle, M. *et al.* (2017b). Patterns of diversity and spatial variability of β-defensin innate immune genes in a declining wild population of tree swallows. *Journal of Heredity*, 108, 262–269.

Semple, C.A., Gautier, P., Taylor, K. & Dorin, J.R. (2006). The changing of the guard: molecular diversity and rapid evolution of beta-defensins. *Molecular Diversity*, 10, 575–584.

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

Sorci, G. (2013). Immunity, resistance and tolerance in bird-parasite interactions. *Parasite Immunology*, 35, 350–361.

Spurgin, L.G. & Richardson, D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277, 979–988.

Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A. & Richardson, D.S. (2014). Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Molecular Ecology*, 23, 1028–1039.

Spurgin, L.G., Illera, J.C., Padilla, D.P. & Richardson, D.S. (2012). Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia*, 168, 691–701.

Spurgin, L.G., Oosterhout, C. van, Illera, J.C., Bridgett, S., Gharbi, K. & Emerson, B.C. *et al.* (2011). Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, 20, 5213–5225.

Subramanian, S. (2016). The effects of sample size on population genomic analyses – implications for the tests of neutrality. *BMC Genomics*, 17, 123.

Sutton, J.T., Robertson, B.C., Grueber, C.E., Stanton, J.A.L. & Jamieson, I.G. (2013). Characterization of MHC class II B polymorphism in bottlenecked New Zealand saddlebacks reveals low levels of genetic diversity. *Immunogenetics*, 65, 619–633.

Svensson-Coelho, M. & Ricklefs, R.E. (2011). Host phylogeography and beta diversity in avian haemosporidian (Plasmodiidae) assemblages of the Lesser Antilles. *Journal of Animal Ecology*, 80, 938–946.

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

Tang, K.Y., Wang, X., Wan, Q.H. & Fang, S.G. (2018). A crucial role of paralogous β-defensin genes in the Chinese alligator innate immune system revealed by the first determination of a Crocodilia defensin cluster. *Developmental and Comparative Immunology*, 81, 193–203.

Tennessen, J.A. & Blouin, M.S. (2007). Selection for antimicrobial peptide diversity in frogs leads to gene duplication and low allelic variation. *Journal of Molecular Evolution*, 65, 605–615.

Tennessen, J.A. & Blouin, M.S. (2008). Balancing selection at a frog antimicrobial peptide locus: fluctuating immune effector alleles? *Molecular Biology and Evolution*, 25, 2669–2680.

Tschirren, B., Andersson, M., Scherman, K., Westerdahl, H. & Råberg, L. (2011). 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.

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

Unckless, R.L. & Lazzaro, B.P. (2016). The potential for adaptive maintenance of diversity in insect antimicrobial peptides. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20150291.

Unckless, R.L.R., Howick, V.V.M. & Lazzaro, B.P.B. (2016). Convergent balancing selection on an antimicrobial peptide in *Drosophila*. *Current Biology*, 26, 257–262.

Van Riper, C., Van Riper, S.G., Goff, L.M. & Laird, M. (1986). The epizootiology and ecological significance of malaria in Hawaiian land birds. *Ecological Monographs*, 56, 327–344.

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

Warner, R.E. (1968). The role of introduced diseases in the extinction of the endemic Hawaiian avifauna. *The Condor*, 70, 101–120.

Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W. & Künstner, A. *et al.* (2010). The genome of a songbird. *Nature*, 464, 757–762.

Waterhouse, R.M., Kriventseva, E.V., Meister, S., Xi, Z., Alvarez, K.S. & Bartholomay, L.C. *et al.* (2007). Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science*, 316, 1738–1743.

Watterson, G.A. (1978). The homozygosity test of neutrality. *Genetics*, 88, 405–417.

Winternitz, J.C. & Wares, J.P. (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.

Woodworth, B.L., Atkinson, C.T., LaPointe, D.A., Hart, P.J., Spiegel, C.S. & Tweed, E.J. *et al.* (2005). Host population persistence in the face of introduced vector-borne diseases: Hawaii amakihi and avian malaria. *Proceedings of the National Academy of Sciences*, 102, 1531–1536.

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

Xiao, Y., Hughes, A.L., Ando, J., Matsuda, Y., Cheng, J.-F. & Skinner-Noble, D. *et al.* (2004). A genome-wide screen identifies a single beta-defensin gene cluster in the chicken: implications for the origin and evolution of mammalian defensins. *BMC Genomics*, 5, 56.

Yang, M., Zhang, C., Zhang, M.Z. & Zhang, S. (2017). Novel synthetic analogues of avian β-defensin-12: the role of charge, hydrophobicity, and disulfide bridges in biological functions. *BMC Microbiology*, 17, 43.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134.

Yeaman, M.R. & Yount, N.Y. (2003). Mechanisms of antimicrobial peptide action and resistance. *Pharmacological Reviews*, 55, 27–55.

Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, 415, 389–395.

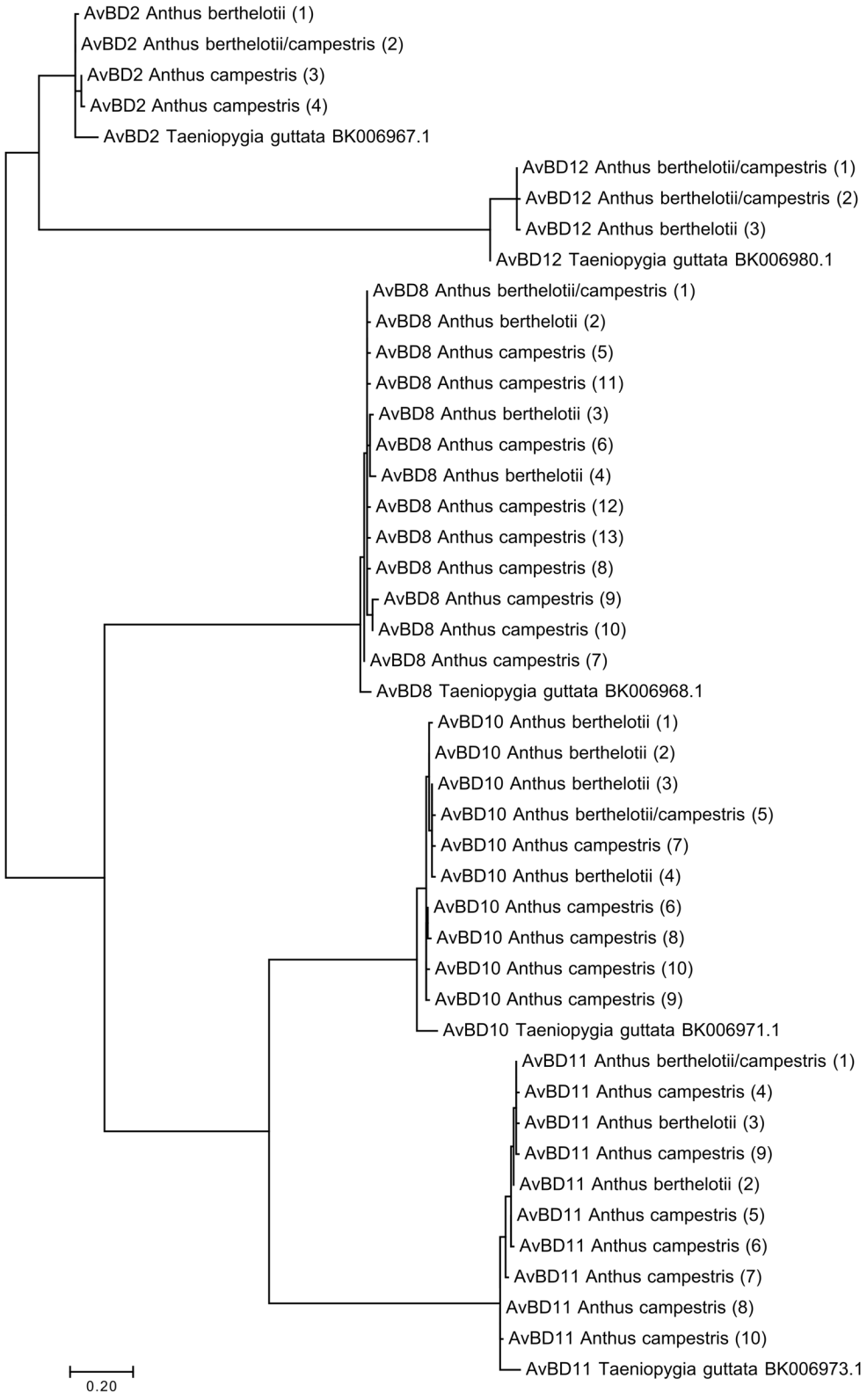
3.7. Supplementary Material

Supplementary Table S3.1 Number of samples per population of Berthelot's and tawny pipits successfully sequenced in the AvBD preliminary study.

Archipelago/species	Population	AvBD4	AvBD7	AvBD8	AvBD9	AvBD10	AvBD11	AvBD13
Canary Islands	Graciosa		1	1	1	1	1	1
	Lanzarote	1	1	1	1	1	1	1
	Fuerteventura	1	1	1	1	1		1
	Gran Canaria		2	2	2	1	1	1
	Tenerife	3	3	3	3	3	3	3
	El Teide	1	1	1	1	1	1	1
	La Gomera	1	1	1	1	1	1	1
	La Palma	2	2	2	2	1	1	1
	El Hierro	1	1	1	1	1	1	1
Selvagens	Selvagem Grande	2	2	2	1	1		1
Madeira	Madeira	2	2	2	2	1	1	1
	Porto Santo	2	2	1	2	1	1	1
	Deserta Grande	1	1	2	2	1	1	1
Berthelot's pipit		17	20	20	20	15	13	15
Tawny pipit		11	12	12	12	12	12	12

Supplementary Table S3.2 Species and Genbank accession numbers of whole-genome shotgun contigs used to design AvBD primers. Divergence times from Berthelot's pipits (DT) in millions of years are estimated from TimeTree (www.timetree.org).

Species	Common Name	DT	Loci	GenBank Accession
Serinus canaria	Atlantic canary	35	AvBD2, 8	CAVT010003542.1
			AvBD10, 11, 12	CAVT010003543.1
Taeniopygia guttata	Zebra finch	35	AvBD2, 8, 10	ABQF01025082.1
			AvBD11, 12	ABQF01025086.1
Geospiza fortis	Medium ground finch	38	AvBD2	AKZB01061284.1
			AvBD8	AKZB01061286.1
			AvBD10, 11, 12	AKZB01061287.1
Zonotrichia albicollis	White-throated sparrow	38	AvBD2, 8	ARWJ01020695.1
			AvBD10	ARWJ01020696.1
			AvBD11, 12	ARWJ01020699.1
Ficedula albicollis	Collared flycatcher	42	AvBD2, 8	AGTO02002612.1
			AvBD10, 11, 12	AGTO02002613.1
Corvus brachyrhynchos	American crow	44	AvBD2	JMFN01028189.1
			AvBD8	JMFN01028191.1
			AvBD10	JMFN01028192.1
			AvBD11, 12	JMFN01028195.1
Corvus cornix cornix	Hooded crow	44	AvBD2, 8	JPSR01026974.1
			AvBD10, 11, 12	JPSR01026975.1
Pseudopodoces humilis	Ground tit	44	AvBD2, 8, 10	ANZD01015336.1
			AvBD11, 12	ANZD01015337.1
Manacus vitellinus	Golden-collared manakin	66	AvBD2	JMFM02004670.1
			AvBD8, 10	JMFM02011779.1
			AvBD11	JMFM02011780.1
			AvBD12	JMFM02011781.1



Supplementary Figure S3.1 Maximum likelihood phylogenetic tree of the relationship between AvBD loci in Berthelot's and tawny pipits. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Species names are followed by the haplotype number in brackets (*Anthus*) or GenBank accession number (*Taeniopygia guttata*). Haplotypes recorded as "*Anthus berthelotii/campestris*" were detected in both Berthelot's (*A. berthelotii*) and tawny (*A. campestris*) pipits.

Chapter 4. Adaptive landscape genetics and malaria across replicate island populations

Testing the limits of our trusty hire car

4.1. Abstract

Environmental conditions play a major role in shaping the spatial distributions of pathogens, which in turn can drive local adaptation and divergence in host genetic diversity. Malaria is a strong selective force, impacting survival and fitness of hosts, with geographic distributions largely determined by habitat suitability for its insect vectors. Here, we have tested whether patterns of fine-scale local adaptation to malaria are replicated across discrete, ecologically divergent island populations of Berthelot's pipits (*Anthus berthelotii*). We sequenced TLR4, an innate immunity candidate gene that is potentially under positive selection in Berthelot's pipits, and two SNPs previously identified as being associated with malaria infection in a genome-wide association study (GWAS) in Berthelot's pipits in the Canary Islands. We determined the environmental predictors of malaria infection, using these to estimate variation in malaria risk on Porto Santo, while finding some congruence with previously identified environmental risk factors on Tenerife. We found evidence of associations between malaria infection/risk and TLR4 variants, with the same trends found in both islands. By contrast, one of the GWAS SNPs showed significant associations with malaria risk in Porto Santo, but in the opposite direction to that found in the GWAS. Together, these findings suggest the potential for malaria-driven local adaptation within and between divergent populations of a bottlenecked island bird.

4.2. Introduction

Spatial variation in the environment is a key force driving local adaptation and population divergence (Hereford 2009). There is increasing evidence that fine-scale changes in environmental conditions can result in variation in selection pressures over small geographic distances (Garroway *et al.* 2013; Richardson *et al.* 2014; Langin *et al.* 2015), and that fine-scale adaptation can persist despite the homogenising effects of gene flow (e.g. Lenormand 2002). Adaptation to spatially heterogeneous environmental conditions can also facilitate balancing selection, where genetic variation is maintained within and between populations due to differential selection on genetic variants (Levene 1953; Schmidt *et al.* 2000; Bockelmann *et al.* 2003).

Pathogens are major drivers of adaptive evolution, exerting strong selective pressures on their hosts (Fumagalli *et al.* 2011). High levels of genetic variation can be found at many genes involved in immune-related processes, and this variation is maintained by pathogen-mediated balancing selection (Bernatchez & Landry 2003; Ferrer-Admetlla *et al.* 2008). The three main, non-mutually exclusive mechanisms by which pathogens are thought to drive spatial heterogeneity in immune gene diversity are heterozygote advantage, rare-allele advantage and

fluctuating selection (Meyer & Thomson 2001; Piertney & Oliver 2006; Spurgin & Richardson 2010). Heterozygotes may gain a selective advantage over homozygotes if having multiple alleles at a locus confers resistance to a broader range of pathogens, or a better response to a single pathogen (Doherty & Zinkernagel 1975). Selection for heterozygotes in regions with high pathogen abundances can lead to increases in heterozygote frequency relative to uninfected areas (Flint *et al.* 1998; Moormann *et al.* 2003). Under the rare-allele advantage model, host and pathogen are in a coevolutionary arms race, causing cyclical fluctuations in allele frequencies (Takahata & Nei 1990; Slade & McCallum 1992). Under this model, spatially separated host populations may show different genetic responses to the same pathogen, as they are involved in independent coevolutionary cycles (Bonneaud *et al.* 2006). Finally, the fluctuating selection hypothesis predicts that spatial and temporal heterogeneity in pathogen communities, driven by a combination of environmental conditions and stochastic processes, will result in local adaptation of host immune genes to the local pathogen selection regimes (Hill *et al.* 1991), with diversity being maintained over broad spatial scales. All three of these mechanisms have the potential to vary over spatial scales; however, only recently has spatial scale been recognised as an important factor in understanding pathogen-mediated selection (Tschirren *et al.* 2011; Tack *et al.* 2014; Larson *et al.* 2016).

Haemosporidians in the genera *Plasmodium*, *Haemoproteus* and *Leukocytozoon* are protozoan parasites that infect the red blood cells of mammals, reptiles and birds (Martinsen *et al.* 2008). These parasites can have strong demographic effects on their hosts: malaria (*Plasmodium* spp.) caused nearly half a million human deaths in 2016 (World Health Organization 2017), and has been implicated in the decline and extinction of a number of Hawaiian bird species (Warner 1968; Van Riper *et al.* 1986; Foster *et al.* 2004). Infection by haemosporidians has been associated with increased mortality (Kilpatrick & LaPointe 2006; Marzal *et al.* 2008; Guggisberg *et al.* 2018), decreased body condition (Marzal *et al.* 2008; Otero *et al.* 2018), and reductions in fitness in natural populations (Marzal *et al.* 2005; Ortego *et al.* 2008; Knowles *et al.* 2010). The strong selective pressure exerted by haemosporidians has driven the evolution of increased host resistance and tolerance (Hill *et al.* 1991; Tishkoff *et al.* 2001; Atkinson *et al.* 2013; Westerdahl *et al.* 2013), with evidence of local adaptation to spatially heterogeneous selection pressures (Gilbert *et al.* 1998; Piel *et al.* 2010; Loiseau *et al.* 2011).

Haemosporidian parasites are dependent on vector transmission to complete their lifecycle, and as a result their spatial distributions are constrained by the availability of suitable habitat for their insect vectors. Climatic conditions such as temperature and rainfall play important roles in determining haemosporidian and vector distrubutions and abundances (Craig *et al.* 1999; Paaijmans *et al.* 2009; Blanford *et al.* 2013; Jones *et al.* 2013; Otero *et al.* 2018). Landscape

features such as altitude and standing water likewise influence vector prevalence, with high altitudes associated with decreased *Plasmodium* infection prevalence (Bødker *et al.* 2003; Niebuhr *et al.* 2016; Illera *et al.* 2017; Padilla *et al.* 2017). Aquatic larval development of vectors makes availability of standing water an important determinant of haemosporidian distributions, and topographic features that increase surface water persistance may promote increased vector abundance (González-Quevedo *et al.* 2014; Ferraguti *et al.* 2016, 2018; Ganser *et al.* 2016). In addition, anthropogenic factors such as habitat degradation, agriculture and urbanization can all influence haemosporidian dynamics (Yanoviak *et al.* 2006; González-Quevedo *et al.* 2014; Roiz *et al.* 2015; Reinoso-Pérez *et al.* 2016; Turcotte *et al.* 2018). Together, these factors have the potential to shape fine-scale spatial structuring in pathogen selection pressures and host immunogenetic variation.

Studying pathogen-mediated selection has largely involved a candidate gene approach, where variation at genes of known, or predicted, function relevant to host immunity is investigated in relation to pathogen infection (Bernatchez & Landry 2003; Acevedo-Whitehouse & Cunningham 2006; Netea *et al.* 2012). Many studies have focused on the major histocompatibility complex (MHC), a group of genes that play a key role in pathogen recognition in the adaptive immune system. Both overall MHC diversity and specific MHC alleles have been associated with malaria infection probability and immune response (Jepson *et al.* 1997b; Westerdahl *et al.* 2005; Jones *et al.* 2015), indicating that this gene family is an important component of host immunity to malaria (Hill *et al.* 1991). However, a greater proportion of phenotypic variance in malaria response has been attributed to non-MHC than MHC genes (Jepson *et al.* 1997a). For example, the innate immune system, which acts as a first line of defence against infection, is under strong selection from pathogens (Vinkler & Albrecht 2009). Within this, the toll-like receptors (TLRs) are a family of pattern-recognition receptors which trigger signalling cascades to launch an immune response upon detection of pathogen-associated molecular patterns (Medzhitov *et al.* 1997; Janeway & Medzhitov 2002). TLR polymorphisms have been linked to susceptibility and resistance to malaria (Mockenhaupt *et al.* 2006; Mockenhaupt *et al*. 2006b; Ferwerda *et al.* 2007) and other infectious diseases (Lorenz *et al.* 2002; Tschirren *et al.* 2013; Bateson *et al.* 2016). Evidence of pathogen-mediated balancing selection acting on TLRs has been found in humans and wild animals (Bonneaud *et al.* 2006; Ferrer-Admetlla *et al.* 2008; Gavan *et al.* 2015). TLR genes therefore represent important candidates for investigating the role of pathogens in maintaining host genetic variation. An alternative approach to studying pathogen-mediated selection is the use of genome-wide association studies (GWAS). By comparing genotypes at single-nucleotide polymorphisms (SNPs) throughout the genome between infected and noninfected individuals, it is possible to detect genetic variants that are associated with increased resistance or susceptibility to pathogen infection. In addition to identifying

associations with known immune loci such as the MHC, TLRs and interleukins (Fellay *et al.* 2007; Wong *et al.* 2010; He *et al.* 2015), GWAS approaches enable the detection of novel candidates (Thye *et al.* 2010; Fu *et al.* 2012; Ravenhall *et al.* 2018) for further study of the evolutionary dynamics between host and pathogen.

Islands are excellent environments for investigating pathogen-mediated selection. In line with island biogeographical theory (MacArthur & Wilson 1967), pathogen diversity and abundance tends to be lower on islands compared with the mainland, and on smaller, more remote islands (Moro *et al.* 2003; Pérez-Rodríguez *et al.* 2013; Clark *et al.* 2014; but see Illera *et al.* 2015; Jean *et al.* 2016), simplifying the study of host–pathogen interactions. Pathogen communities on each island are shaped by chance colonisation and extinction events, which can result in distinct pathogen assemblages and selection pressures between islands (Fallon *et al.* 2005; Olsson-Pons *et al.* 2015; Wang *et al.* 2017). Limited gene flow into and out of islands allows for stable communities of hosts and pathogens (Spurgin *et al.* 2012), which may facilitate strong coevolutionary relationships.

Berthelot's pipit (*Anthus berthelotii*) is a small sedentary passerine endemic to three Macaronesian archipelagos in the Atlantic Ocean (Chapter 2: Fig. 2.1A). Following the colonisation of the Madeiran archipelago from the Canary Islands ca. 8000 years ago, there has been a lack of gene flow between the archipelagos (Illera *et al.* 2007; Spurgin *et al.* 2014), potentially facilitating local adaptation and divergent selection (Armstrong *et al.* 2018). Levels of haemosporidian infections show high spatial variability in this species, both between and within islands, making it a highly suitable model for investigating the role of spatial scale in pathogen-mediated selection. The highest prevalence of *Plasmodium* and *Leucocytozoon* infection has been found on the island of Porto Santo, while, interestingly, no infection was detected on neighbouring islands within the Madeiran archipelago (Illera *et al.* 2008; Spurgin *et al.* 2012). Aside from Porto Santo, *Leucocytozoons* have only been detected at low frequency in two other populations (Tenerife and Gran Canaria), and no evidence of *Haemoproteus* has been found in this species (Illera *et al.* 2008; Spurgin *et al.* 2012; González-Quevedo *et al.* 2014). On Tenerife, moderate levels of haemosporidians have been found in lowland pipit populations, whereas none have been detected at high altitude (Illera *et al.* 2008; Spurgin *et al.* 2012; González-Quevedo *et al.* 2014). Furthermore, prevalence of malaria on Tenerife is influenced by a combination of climatic and anthropogenic effects (González-Quevedo *et al.* 2014), and associations between the distribution of MHC variants and environmental predictors of malaria infection have been detected (González-Quevedo *et al.* 2016).

Here, we have investigated the association between fine-scale patterns of genetic variation and malaria in Berthelot's pipits across two divergent populations in Tenerife and Porto Santo. These

two islands, situated on different archipelagos, show high genetic divergence at neutral loci, with limited to no gene flow (Spurgin *et al.* 2014; Armstrong *et al.* 2018). Despite a sharp decline in genetic diversity associated with the initial colonisation of Madeira (Chapter 3; González-Quevedo *et al.* 2015; Armstrong *et al.* 2018), higher levels of TLR4 allelic and amino acid richness exist in Madeira compared to the Canary Islands, and there is other evidence of positive selection at this locus (González-Quevedo *et al.* 2015). TLR4 polymorphisms have been associated with increased malaria risk in humans (Mockenhaupt *et al.* 2006; Ferwerda *et al.* 2007; Iwalokun *et al.* 2015), and may play a potential role in mediating malaria infection in Berthelot's pipits. In addition to TLR4, we have investigated associations between malaria risk and two SNPs previously identified in a GWAS of malaria infection across Berthelot's pipit populations (Armstrong *et al.* 2018). Both SNPs showed significant associations between genotype and infection status throughout the Canary Islands, but this relationship was not consistent across the other archipelagos (Armstrong *et al.* 2018).

Our previous test for associations between malaria infection and genotype in Berthelot's pipits (Armstrong *et al.* 2018) did not take into account the strong environmental drivers of malaria prevalence. Using only infection status, it would be unclear whether a bird was uninfected due to possessing a resistant genotype, or if they are simply not yet infected. Here, in addition to exploring genetic associations with malaria infection status in Tenerife and Porto Santo, we have used in-depth sampling of Porto Santo to model the environmental predictors of malaria infection. We have applied the predicted values of this model as a measure of individual malaria risk based on the local environmental conditions surrounding each bird, to investigate the associations with genotype at TLR4 and previously detected candidate malaria SNPs (Armstrong *et al.* 2018).

By testing the genetic associations with malaria infection and risk within two divergent island populations, we aimed to investigate the spatial scale of local adaptation in the presence of gene flow. Our aims were as follows: 1) test for associations between malaria infection status and genetic variation in Tenerife and Porto Santo; 2) determine the environmental predictors of malaria risk on Porto Santo; 3) compare genetic associations with malaria risk in Porto Santo and Tenerife, utilising the above measures of malaria risk for Porto Santo, and those previously calculated for Tenerife (González-Quevedo *et al.* 2014, 2016).

4.3. Methods

4.3.1. Sample collection

Berthelot's pipits were sampled over several field seasons on Tenerife (Feb–April 2006, Jan–Aug 2009, April–May 2010, Jan–May 2011) and Porto Santo (Sept 2006, March 2009, April–June 2016). For samples collected prior to 2010, 30–60 birds were sampled widely across each island per season (Illera *et al.* 2007; Spurgin *et al.* 2012). In 2010, ≥ 30 birds were caught from each of four potential geographical subpopulations in the west, east and south of Tenerife, and from the volcanic plateau of Teide in the centre of Tenerife (Spurgin *et al.* 2012). For samples collected in Tenerife in 2011, the island was divided into 1 km^2 grid cells. Attempts were made to catch one pipit per 1 km^2 within suitable habitat, with a total of 388 birds sampled (González-Quevedo *et al.* 2014). Porto Santo was intensively sampled in 2016, with the aim of surveying all areas of suitable and accessible pipit habitat throughout the island. We attempted to catch every pipit that was encountered, resulting in a sample size of 129 birds for Porto Santo in 2016, and a total of 780 birds across all years and islands.

Birds were caught in spring traps baited with *Tenebrio molitor* larvae. Each bird was fitted with either a coloured plastic ring or with a numbered aluminium ring issued by the Spanish or Portuguese authorities as appropriate for the location, to avoid resampling a previously encountered individual. A blood sample (ca. 50 µl) was collected from each bird by brachial venipuncture and stored in 800 µl 100% ethanol in a screw top Eppendorf tube at room temperature. Birds were classified as juvenile (EURING age code 3; born this calendar year) or adult (EURING age codes 4–6; born before this calendar year), based on feather moult pattern (Cramp 1988), with the exception of samples collected in 2009, where age codes 3 and 5 were classified as juvenile.

4.3.2. Molecular methods

DNA was extracted from each sample using a salt extraction protocol detailed in Richardson *et al.* (2001). Sexing PCRs (Griffiths *et al.* 1998) were performed to determine the sex of the bird, and ensure that DNA extractions were successful.

We used a nested PCR approach that detects *Plasmodium* and *Haemoproteus* to characterise haemosporidian infection status. We amplified each sample first with the primers *HAEMF* and *HAEMR2*, followed by a second amplification with *HAEMNF* and *HAEMNR2* (Waldenström *et al.* 2004), with positive and negative controls included in each PCR plate. Samples that successfully amplified at least twice were classified as infected. The strain of *Plasmodium* was determined

by Sanger sequencing; *Haemoproteus* has not been detected in this species. We have focused on *Plasmodium* in this study as it is the most widespread and abundant haemosporidian family in Berthelot's pipits (Illera *et al.* 2008; Spurgin *et al.* 2012). In addition, the vector species of *Plasmodium* and *Leucocytozoon* have different ecological niches: the mosquito (Culicidae) vectors of *Plasmodium* require hotter thermal ranges to the blackfly (Simulidae) vectors of *Leucocytozoons*, resulting in *Leucocytozoon* distributions at higher altitudes and colder climates than tolerated by *Plasmodium* vectors (Haas *et al.* 2012; Imura *et al.* 2012; Harrigan *et al.* 2014).

A previous study genotyped 23–30 Berthelot's pipits from all 13 Berthelot's pipit populations at TLR4 (González-Quevedo *et al.* 2015). The primers *PauTLR4F* and *PauTLR4R* (Grueber *et al.* 2012) were used to amplify a 660-bp region located within the leucine-rich repeat domain of TLR4. Five SNPs were found within this region: TLR4_1 (905 bp, nonsynonymous), TLR4_2 (970 bp, nonsynonymous), TLR4_3 (990 bp, synonymous), TLR4_4 (992 bp, nonsynonymous and triallelic), and TLR4_5 (1010 bp, nonsynonymous). Base-pair positions are stated according to the zebra finch TLR4 protein coding region, Genbank accession FJ695612. Further Sanger sequencing of TLR4 was performed on all samples collected from Porto Santo in 2016. Following Sanger sequencing, TLR4 SNPs were called through visual inspection of the chromatograms in FinchTV (www.geospiza.com/finchtv). We used LGC Genomics' proprietary KASP™ genotyping technology (www.lgcgroup.com) to genotype all additional samples from Tenerife and Porto Santo at each TLR4 SNP, except for TLR4_5 which was excluded due to a very low minor allele frequency of < 0.05. Assay design and genotyping was performed by LGC Genomics, Hertfordshire.

A genome-wide association study performed on restriction site-associated DNA sequencing (RAD-seq) data in Berthelot's pipits by Armstrong *et al.* (2018) detected two SNPs (5239s1, Chr10:12048280; 7259s1, Chr20:6483195; SNP locations refer to equivalent position on zebra finch genome version 3.2.4; Warren *et al.* 2010) that showed significant associations with *Plasmodium* strain LK6 (Ortego *et al.* 2007) infection. These were sequenced in 20 birds from Porto Santo and 42 from Tenerife; KASP™ assays were likewise designed for these two SNPs, and all remaining birds were genotyped at these sites.

Genotypes AT, CT and TT at the triallelic TLR4_4 SNP were coded as missing data (Tenerife n = 12; Porto Santo n = 1), to effectively treat this SNP as biallelic. We used DnaSP v6 (Librado & Rozas 2009) to phase the TLR4 SNPs into haplotypes. To aid phasing, we included TLR4 sequences from all Berthelot's pipit populations, and each phased TLR4 haplotype previously detected in Berthelot's pipits (González-Quevedo *et al.* 2015). Samples with < 90% phasing certainty were excluded from models that included TLR4 haplotypes as predictors. We translated the phased TLR4 sequences originating from Sanger sequencing into protein

haplotypes. This gave us the amino acid residues at each of the codons containing a SNP, from which we were able to infer the amino acids at each SNP for samples that were genotyped with KASP™ genotyping.

We used PLINK 1.9 (Chang *et al.* 2015) to calculate linkage disequilibrium (LD) between each pair of SNPs, and test for deviations from Hardy Weinberg Equilibrium with Hardy Weinberg exact tests. Where frequencies ≤ 0.05 were found for SNP minor alleles or TLR4 protein haplotypes, these variants were not included as predictors in genetic analyses.

4.3.3. Porto Santo GIS analyses

A previous study demonstrated the importance of environmental and anthropogenic factors in shaping malaria risk in Tenerife (González-Quevedo *et al.* 2014). Predictors were selected based on the potential for influencing the abundance of avian malaria or of its mosquito vectors: the minimum temperature of the coldest month (MINTEMP), annual precipitation (PRECIPITATION), altitude (ALTITUDE), aspect (ASPECT), slope (SLOPE), pipit density (DENSITY), vegetation type (VEGTYPE), distance to nearest poultry farm (DISTPOUL), distance to nearest livestock farm (DISTFARM), distance to nearest artificial water reservoir (DISTWATER), and distance to urban site (DIST_URB). For analyses on Tenerife, we have used the values calculated by González-Quevedo *et al.* (2014) for all environmental predictors. Values for Porto Santo were calculated as outlined below.

All GIS analyses were performed in QGIS v2.18 (QGIS Development Team 2017). MINTEMP and PRECIPITATION were obtained from WorldClim global climate data v2 (Fick & Hijmans 2017) at a resolution of 30 arc-seconds (approximately 1 km^2). ALTITUDE was obtained from the Shuttle Radar Topography Mission (SRTM) 3 Arc-Second Global elevation data (srtm.csi.cgiar.org) at a resolution of approximately 90 m^2. SLOPE and ASPECT were calculated from the SRTM data. VEGTYPE was characterised using the CORINE Land Cover inventory (CLC 2012; www.eea.europa.eu). We combined the CLC land cover classes into six vegetation categories: arable (non-irrigated arable land; vineyards; complex cultivation patterns; land principally occupied by agriculture with significant areas of natural vegetation), urban-associated (discontinuous urban fabric; port areas; airports), forest (coniferous forest), rock-associated (beaches, dunes and sand; bare rock; sparsely vegetated areas), grass (natural grassland; pastures), and shrub (moors and heathland; transitional woodland-shrub). The CLC class 'sport and leisure' was characterised as either grass or urban-associated, depending on the land use (golf course or resort complex, respectively). Values of MINTEMP, PRECIPITATION, ALTITUDE, SLOPE and ASPECT at each sample location were calculated by applying a 100 m buffer around

each sample location, and taking the average value of each variable within the buffer. In the case of VEGTYPE, the sample was assigned the category with the largest area. ASPECT was classified as one of eight categories: N, NW, W, SW, S, SE, E and NE.

The location of water sources within Porto Santo was determined in two ways. First, during sample collection, we recorded the coordinates of all standing water sources encountered. These were verified through visual inspection of Google satellite imagery, from which polygons were manually drawn. Second, we used the OpenStreetMap data-filtering tool Overpass Turbo (overpass-turbo.eu) to extract polygons of water sources using the query 'natural=water'. We combined the polygons obtained from both methods to calculate DISTWATER for each sample location. The presence of livestock was dependent on visual encounters during sample collection, as farming census data were not publically available. The type of livestock was used to differentiate between factors related to livestock farming that might cause aggregations of birds (DISTFARM), and the potential effect of poultry as a reservoir of avian malaria (DISTPOUL). DIST_URB was calculated as the distance to the nearest urban-associated area, as classified by VEGTYPE.

DENSITY was calculated as follows. First, a 1 km^2 grid was overlaid on Porto Santo, and a 1 km radius buffer was drawn around the centroid of each grid cell. Any part of the buffer that extended into the ocean was removed. This was converted into a measure of pipits per km^2 for each grid cell by dividing the area of the buffer by the number of pipits within that area. The average density within a 100 m buffer around each sample, weighted by the area of each 1 km^2 grid cell occupied within the sample buffer, was used as the DENSITY measure.

4.3.4. Environmental predictors of malaria risk

We modelled the relationship between environmental variables and malaria infection status to gain a measure of malaria risk for each sample on Porto Santo, using a model selection and model averaging approach. Model selection tests all combinations of variables as predictors of the response variable and calculates the Akaike Information Criterion (AIC; Akaike 1973) as a measure of fit of each model, to determine which models have the strongest relative likelihood of approximating reality. Model averaging is then applied to a set of models with the lowest AIC (and therefore the highest likelihood) to calculate weighted averages of parameter estimates and the relative importance of model predictors (Burnham & Anderson 2002). Throughout this study we report AICc, a modification of AIC that includes a correction for sample size, which is recommended over AIC when sample sizes are small (Hurvich & Tsai 1989).

We first used variance inflation factors (VIFs) to test for collinearity between environmental variables. This method treats each variable in turn as the response variable in a regression with all other variables as predictors; the VIF is related to the R^2 from this regression, giving a measure of dependency between environmental variables. VIFs were calculated using the R package car (Fox & Weisberg 2011). Zuur *et al.* (2010) recommend using a threshold of VIF > 3 to indicate that collinearity is likely to be problematic. As our set of environmental variables included categorical variables, this analysis reported generalized VIFs (GVIFs; Fox & Monette 1992). We used the transformation $(GVIF^{1/2df})^2$, where *df* is the degrees of freedom associated with the variable, to calculate a value equivalent to a standard VIF (Fox & Weisberg 2011). When including all variables, we found VIFs > 3, hence collinearity, for every variable except ASPECT. To decide which variables to remove, we performed single-predictor binomial generalized linear models (GLMs) of each variable against malaria infection status (Table 4.1). Variables with VIFs > 3 and high AICc scores were sequentially removed and VIFs recalculated until all remaining variables had VIFs < 3. The remaining variables were VEGTYPE, ALTITUDE, DISTWATER, DENSITY, ASPECT and DISTPOUL.

Table 4.1. Single-predictor binomial generalized linear models of the environmental predictors of *Plasmodium* strain LK6 infection in adult Berthelot's pipits on Porto Santo. Environmental variables are ordered by increasing AICc scores. Parameter estimates, McFadden pseudo R^2 and *P*-values are shown. Estimates are not included for VEGTYPE and ASPECT as these are categorical variables.

Variable	Estimate	R^2	*P*-value	AICc
ALTITUDE	-0.0160	0.129	<0.001	100.0
DISTWATER	-0.0013	0.118	<0.001	101.2
DIST_URB	-0.0014	0.113	<0.001	101.8
VEGTYPE	—	0.157	0.170	103.4
MINTEMP	2.5188	0.080	0.005	105.4
SLOPE	-0.1569	0.069	0.008	106.6
PRECIPITATION	-0.0308	0.041	0.035	109.7
DISTPOUL	-0.0006	0.018	0.152	112.2
DISTFARM	-0.0005	0.010	0.284	113.1
DENSITY	0.0747	0.003	0.586	113.9
ASPECT	—	0.094	0.271	117.2

Interactions between pairs of environmental predictors may have biologically meaningful influences on malaria risk. We selected pairs of predictors that could potentially interact to affect malaria risk (Table 4.2) and tested whether the inclusion of an interaction term resulted in an improvement in model fit. We performed a binomial GLM of each pair of predictors as main effects and malaria infection as the response, with and without an interaction term. The interaction DENSITY*DISTWATER gave the largest improvement in AICc, and was therefore included in model selection (Table 4.2).

Table 4.2. The effect of including biologically relevant interaction terms between environmental variables for predicting *Plasmodium* LK6 infection in adult Berthelot's pipits on Porto Santo. Binomial generalized linear models were performed with each pair of variables as main effects only, or including an interaction term. Where the addition of an interaction resulted in a change in AICc (ΔAICc) < -7, that interaction term was included in model selection.

	Main effect AICc	Interaction AICc	ΔAICc	Retained
DENSITY*DISTWATER	103.3	90.8	-12.5	✓
DISTPOUL*DISTWATER	103.1	99.6	-3.5	
SLOPE*ALTITUDE	100.2	97.1	-3.1	
ALTITUDE*DISTWATER	92.3	91.2	-1.1	
ALTITUDE*DISTPOUL	101.2	101.3	0.1	
DENSITY*DISTPOUL	114.3	114.7	0.4	
ASPECT*SLOPE	116.4	118.3	1.9	
VEGTYPE*DENSITY	105.6	107.7	2.2	

We performed model selection and model averaging following Grueber *et al.* (2011) using the R package MuMIn (Bartoń 2018) to obtain a set of the best-supported models for explaining occurrence of malaria infection. This took a binomial GLM consisting of the six environmental variables and one interaction outlined above as predictors of malaria infection, and ran all possible combinations of this model. Prior to analysis, we used the R package arm (Gelman & Su 2018) to standardize the input variables to a mean of zero and a standard deviation of 0.5. This allows meaningful comparisons of the relative strength of parameter estimates (Gelman 2008; Grueber *et al.* 2011). The model selection process calculated ΔAICc, the difference in AICc between each model and the 'best' model (the model with the lowest AICc). It also reported the Akaike weight for each model, which quantifies the likelihood of a model having the best explanatory power within a set of models (Burnham & Anderson 2002). Using the R package DescTools (Signorell 2018), we calculated the McFadden adjusted pseudo R^2 (the likelihood of a model relative to an intercept model) for each model. This measure assesses the goodness-of-fit of a logistic regression, representing the improvement in likelihood of the full model over an

intercept-only model, adjusted to account for the number of predictors in the model (McFadden 1974). Values of McFadden R^2 between 0.2 and 0.4 represent a strong fit, equivalent to a linear regression R^2 of 0.7–0.9 (Louviere *et al.* 2000).

To determine which models would be included in model averaging, we used a threshold of ΔAICc ≤ 7. This threshold is recommended to retain models with sufficient support without dismissing models which still provide some explanatory power (Burnham *et al.* 2011). We applied model averaging over this set of models to calculate weighted averages of parameter estimates. We used the zero method of model averaging, which applied an estimate of zero for models in which the predictor is absent, to avoid biasing results towards predictors with low explanatory power (Burnham & Anderson 2002; Lukacs *et al.* 2007). In addition, model averaging calculated the relative importance of each predictor, an estimate of the probability of a predictor appearing in the best model, calculated as the sum of the Akaike weights for models which include that predictor.

We tested for spatial autocorrelation in model residuals since it is well known that spatial autocorrelation may lead to spurious associations between predictor and response variables (Lennon 2000; Dormann *et al.* 2007). We created Moran's I correlograms at distance class intervals of 750 m and 1000 m using the R package ncf (Bjornstad 2018), with 1000 permutations to test the significance of Moran's I at each interval. After correcting for multiple testing using the Holm correction (a less conservative alternative to Bonferroni correction; Holm 1979; Legendre & Legendre 2012), there was no evidence of spatial autocorrelation in the model residuals (all adjusted *P* values > 0.05). Correcting for spatial autocorrelation was therefore not required for the estimation of malaria risk in Porto Santo.

We used the predicted values of the best model identified by model selection as a malaria risk score between 0 and 1 for each sample location. This was a measure of the probability of an individual at that location being infected with malaria, as a result of the environmental conditions. An earlier study used model selection to determine that malaria infection in Tenerife was best explained by DISTPOUL, DISTWATER, MINTEMP and DISTWATER*MINTEMP (González-Quevedo *et al.* 2014). As significant spatial autocorrelation was present in Tenerife, an autocovariate term was included in all model combinations during model selection, to account for autocorrelation up to 1000 m (see González-Quevedo *et al.* 2014 for autocovariate methods). Hence, the predicted values from an autologistic model containing these predictors, interaction and autocovariate were used as our estimate of malaria risk for Tenerife. Malaria risk was logit-transformed prior to use in models.

4.3.5. Genetic associations with malaria infection

Genetic variation was classified in three ways: 1) SNP genotype, encoded as 1 for heterozygotes, and 0 or 2 for each of the homozygotes; 2) presence (1) or absence (0) of each TLR4 protein haplotype; 3) SNP heterozygosity, with heterozygotes encoded as 1 and homozygotes as 0. Each model described below was performed three times, using each of the three classes of genetic variants as model predictors.

We first tested for associations between genetic variation and malaria infection status. We ran separate models for each island as the lack of gene flow between Berthelot's pipits on Tenerife and Porto Santo, along with environmental differences in malaria risk patterns between the islands, might obscure genetic associations if both islands were included in the same model. We tested for genetic associations with malaria infection either across all years using binomial generalized linear mixed models (GLMMs) with sampling year as a random effect, or within Tenerife (2011) or Porto Santo (2016) with binomial GLMs. Testing across all years gave the advantage of a larger sample size; however, temporal fluctuations in selection pressures could interfere with genetic associations with malaria, so single-year models were also performed.

We then tested the genetic variables as predictors of malaria risk in general linear models (LMs) for Tenerife (2011) and Porto Santo (2016). As malaria risk was derived from spatially varying environmental predictors, genetic variation alone was unlikely to account for all spatial autocorrelation in malaria risk. We therefore included distance-based Moran's eigenvector maps (dbMEMs) as spatial predictors of malaria risk. dbMEMs are used to identify gradients of spatial variation (spatial structure) in a response variable, across multiple potential scales from broad to fine, and are calculated by eigenvector decomposition of distance matrices based on the spatial co-ordinates of samples (Borcard & Legendre 2002; Dray *et al.* 2006). Hence, the use of dbMEMs as predictors of malaria risk accounts for spatially-autocorrelated variation in malaria prevalence that would otherwise be explained by environmental conditions, transmission dynamics, and/or unmeasured genetic gradients. We calculated dbMEMs for each island using the R packages adespatial (Dray *et al.* 2018) and vegan (Oksanen *et al.* 2018). We took all dbMEMs with positive eigenvalues, representing positively autocorrelated spatial variation, and tested for associations between each dbMEM and malaria risk using LMs. Starting with best-fit models fitting genetic predictors of malaria risk, we sequentially added dbMEMs as predictors, fitting those with the highest R^2 first, until additional dbMEMs no longer improved AICc. Using dbMEMs in this way, we checked that this was the minimum number of dbMEMs required to control for spatial autocorrelation in model residuals. The inclusion of dbMEMs resulted in VIFs < 3, indicating that any collinearity between dbMEMs and genetic variants was acceptably low. We performed hierarchical partitioning using the 'lmg' method in the R package

relaimpo (Grömping 2006) to calculate the proportion of variance in malaria risk explained by genetic variants.

4.4. Results

4.4.1. Sequencing

Overall, avian malaria was detected in 126 out of 190 individuals (66.3%) from Porto Santo and 189 out of 590 individuals (32.0%) from Tenerife (Table 4.3). All infected samples had one of four strains of *Plasmodium.* Between 2006 and 2010, only LK6 (Ortego et al. 2007) was detected in Berthelot's pipits (Illera *et al.* 2008; Spurgin *et al.* 2012). Sampling of Tenerife in 2011 found *Plasmodium* in 148 out of 388 individuals (38.1%). The majority of infections were LK6 (139 samples; 93.9%). Two additional strains were found at low frequency: LK5 (Ortego et al. 2007) and KYS9 (Inci *et al.* 2012) were present in seven (4.7%) and two (1.4%) individuals, respectively (González-Quevedo *et al.* 2014). Out of 129 samples collected from Porto Santo in 2016, 97 (75.2%) were infected. Of these, 91 (93.8%) had LK6, four (4.1%) had LK5, and a previously undocumented strain, PS1530, was found in two (2.1%) individuals.

Excluding sampling years where no juveniles were caught, the prevalence of malaria was significantly higher in adults than in juveniles, both in Porto Santo (test of equal proportions χ^2 = 25.6, $P < 0.001$) and Tenerife (χ^2 = 3.1, P = 0.039). As juveniles were present in much lower numbers than adults (Table 4.3), we removed juveniles from further analysis. In previous studies, environmental models of malaria risk in Tenerife had greater explanatory power when considering only individuals with the LK6 strains of *Plasmodium* (González-Quevedo *et al.* 2014). Preliminary analysis in Porto Santo found the same to be true; all further analyses therefore exclude those rare individuals that were infected with other strains of *Plasmodium* (see Table 4.3 for final sample sizes).

Sanger sequencing of TLR4 in the 2016 Porto Santo samples only detected SNPs which have been previously documented in Berthelot's pipits (González-Quevedo *et al.* 2015). The low-frequency T allele in the triallelic TLR4_4 SNP was present in 12 samples from Tenerife and one from Porto Santo; these genotypes were treated as missing, resulting in only biallelic alleles at this SNP. SNP allele frequencies differed between the two islands, with different minor alleles found at TLR4_3 and TLR4_4 (Fig. 4.1A). The SNPs TLR4_1 and TLR4_2 were monomorphic in Tenerife. SNP minor allele frequencies < 0.05 were found in Porto Santo for 7259s1 and TLR4_1. Phasing of the TLR4 SNPs produced five nucleotide haplotypes (Table 4.4), all of which had been previously detected in Berthelot's pipits (González-Quevedo *et al.* 2015). The five haplotypes translated into four

protein haplotypes (denoted with the prefix 'TLR4_P'; Table 4.4; Fig. 4.1B). TLR4_P2 was translated from two haplotypes differing at the synonymous SNP TLR4_3. TLR4_P3 and TLR4_P4 were absent from Tenerife, and TLR4_P3 was present at low frequency (< 0.05) in Porto Santo. As amino acid substitutions could potentially alter TLR4 function (Schröder & Schumann 2005), here we have classified TLR4 as protein haplotypes instead of nucleotide haplotypes.

We tested for deviations from Hardy Weinberg equilibrium at each SNP for each combination of island and year, and in each island across all years. SNPs 5239s1 (Tenerife 2011 and all years), 7259s1 (Tenerife 2006 and 2009), and TLR4_1 (Porto Santo 2009) showed significant deviations from Hardy Weinberg equilibrium at $P < 0.05$; however, following Holm correction for multiple comparisons, all adjusted P values were > 0.05. Linkage disequilibrium between TLR4_3 and TLR4_4 (excluding the low-frequency T allele) was high (Porto Santo $R^2 = 0.99$; Tenerife $R^2 = 0.62$). We found moderate LD in Porto Santo between TLR4_2 and TLR4_3 ($R^2 = 0.39$), and between TLR4_2 and TLR4_4 ($R^2 = 0.37$). All other combinations of SNPs had low levels of LD ($R^2 < 0.1$).

4.4.2. Malaria risk models

Model selection of the environmental predictors of malaria infection found 17 models with ΔAICc ≤ 7 relative to the 'best' (lowest AICc) model (Table 4.5). The best model contained VEGTYPE, ALTITUDE, DISTWATER, DENSITY and DISTWATER*DENSITY. ALTITUDE and DISTWATER were negatively associated with malaria infection, whereas DENSITY was positively associated (Fig. 4.2). A post-hoc Tukey test of VEGTYPE as a predictor of malaria infection (performed using the R package multcomp) found that rock-associated habitat had a significantly negative effect on malaria infection relative to arable ($P = 0.015$) and grass habitats ($P = 0.039$; Fig. 4.2). These four predictors, and the interaction term DISTWATER*DENSITY, had relative importances of 0.63–0.99 in model averaging across models with ΔAICc ≤ 7 relative to the best model (Table 4.6). We used the predicted values from the model containing these predictors and interaction as our estimate for malaria risk for each sample location (Fig. 4.3). DISTPOUL and SLOPE had low relative importance in model averaging (0.33 and 0.29, respectively), and ASPECT did not feature within the top model set. The malaria risk model for Porto Santo had a McFadden adjusted R^2 of 0.25 (values of 0.2–0.4 represent a strong fit, equivalent to a linear regression R^2 of 0.7–0.9; Louviere et al. 2000), whereas the adjusted R^2 for the Tenerife malaria risk model was 0.10 (González-Quevedo *et al.* 2014). Distributions of malaria risk differed markedly between the islands, with a tendency for higher malaria risk in Porto Santo (Fig. 4.4).

Table 4.3. *Plasmodium* prevalence in Berthelot's pipits from Porto Santo and Tenerife across sampling years. Sample sizes are given for all birds (*n*), adults (*n* adults) and juveniles (*n* juveniles). The proportion of birds infected with any strain of *Plasmodium* is listed for all birds, adults and juveniles. The different strains of *Plasmodium* found in Berthelot's pipits (LK6, LK5, KYS9, PS1530) are displayed as proportions of the infected samples. *n* final shows the sample size of the final dataset used in analyses of malaria risk and genetic associations after filtering out juveniles and samples infected with strains other than LK6.

Island	Year	*n*	Infected	*n* adults	Infected adults	*n* juveniles	Infected juveniles	LK6	LK5	KYS9	PS1530	*n* final
Porto Santo		190	0.66	142	0.77	48	0.35	0.95	0.03	0.00	0.02	136
	2006	31	0.65	10	0.90	21	0.52	1.00	0.00	0.00	0.00	10
	2009	30	0.30	16	0.38	14	0.21	1.00	0.00	0.00	0.00	16
	2016	129	0.75	116	0.81	13	0.23	0.94	0.04	0.00	0.02	110
Tenerife		590	0.32	528	0.34	62	0.16	0.95	0.04	0.01	0.00	519
	2006	51	0.08	51	0.08	0	0.00	1.00	0.00	0.00	0.00	51
	2009	56	0.20	25	0.16	31	0.23	1.00	0.00	0.00	0.00	25
	2010	96	0.27	65	0.35	31	0.10	1.00	0.00	0.00	0.00	65
	2011	387	0.38	387	0.38	0	0.00	0.94	0.05	0.01	0.00	378

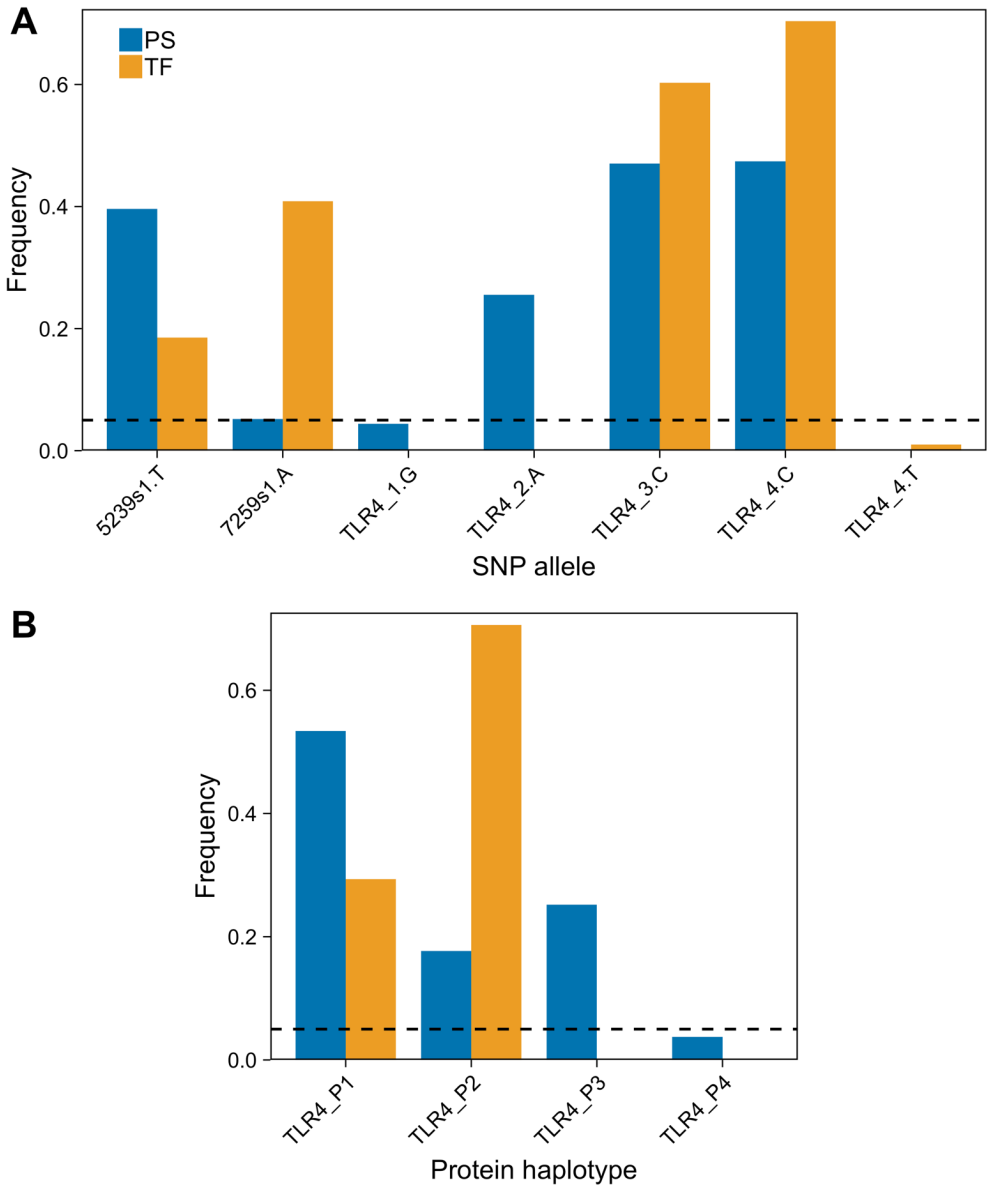


Figure 4.1. Genetic variant frequencies in adult Berthelot's pipits on Porto Santo (PS; blue) and Tenerife (TF; orange). A) Allele frequencies per SNP. In each instance, the allele which is the minor allele in Porto Santo is represented. In addition, the low-frequency T allele of the triallelic SNP TLR4_4 is also shown. B) Frequencies of TLR4 protein alleles. In both plots, the black dashed line indicates a frequency threshold of 0.05; variants below this threshold were excluded from models of genetic associations with malaria risk and infection status.

Table 4.4. TLR4 nucleotide and protein haplotypes in Berthelot's pipits on Tenerife and Porto Santo.

Haplotype	Sequence	Protein haplotype	Amino acid sequence
1	AGTA	TLR4_P1	DGPK
2	AGCC	TLR4_P2	DGPT
3	AGTC	TLR4_P2	DGPT
4	AACC	TLR4_P3	DDPT
5	GGCC	TLR4_P4	GGPT

Table 4.5. Model selection of the environmental predictors of *Plasmodium* LK6 infection in adult Berthelot's pipits on Porto Santo. Parameter estimates, AICc, ΔAICc, Akaike weight (Weight) and McFadden adjusted pseudo R^2 are given for models within a threshold of ΔAICc ≤ 7.

Model	(Intercept)	VEGTYPE	ALTITUDE	DENSITY	DISTPOUL	DISTWATER	SLOPE	DENSITY* DISTWATER	AICc	ΔAICc	Weight	Adjusted R^2
1	2.62	+	-1.91	0.63		-1.54		-5.33	84.9	0.0	0.29	0.25
2	2.61	+	-1.75	0.40	-0.82	-1.39		-5.85	86.4	1.5	0.14	0.24
3	1.97		-1.32	1.01		-2.47		-4.17	86.4	1.5	0.13	0.22
4	2.63	+	-2.00	0.74		-1.56	0.31	-5.28	87.2	2.3	0.09	0.23
5	1.94		-1.23	0.79	-0.69	-2.18		-4.74	87.8	2.9	0.07	0.21
6	1.94		-1.08	0.83		-2.38	-0.52	-4.25	88.1	3.2	0.06	0.21
7	2.62	+	-1.96	0.59	-1.00	-1.34	0.70	-5.89	88.5	3.6	0.05	0.22
8	1.77			0.60		-2.36	-1.21	-4.79	89.0	4.1	0.04	0.20
9	1.93		-1.06	0.68	-0.62	-2.14	-0.42	-4.71	89.7	4.8	0.03	0.20
10	1.76			0.47	-0.65	-2.09	-1.05	-5.22	90.4	5.5	0.02	0.19
11	1.98	+		0.65		-1.81		-5.45	90.5	5.6	0.02	0.19
12	1.74			0.94		-2.43		-4.62	90.8	5.9	0.01	0.18
13	2.01	+		0.35	-1.09	-1.43		-6.20	91.0	6.1	0.01	0.19
14	1.72			0.66	-0.99	-2.04		-5.45	91.1	6.2	0.01	0.18
15	2.32	+	-2.45						91.3	6.4	0.01	0.18
16	2.06	+		0.36		-1.93	-0.98	-5.61	91.4	6.6	0.01	0.19
17	2.32	+	-2.12			-1.35			91.5	6.6	0.01	0.18

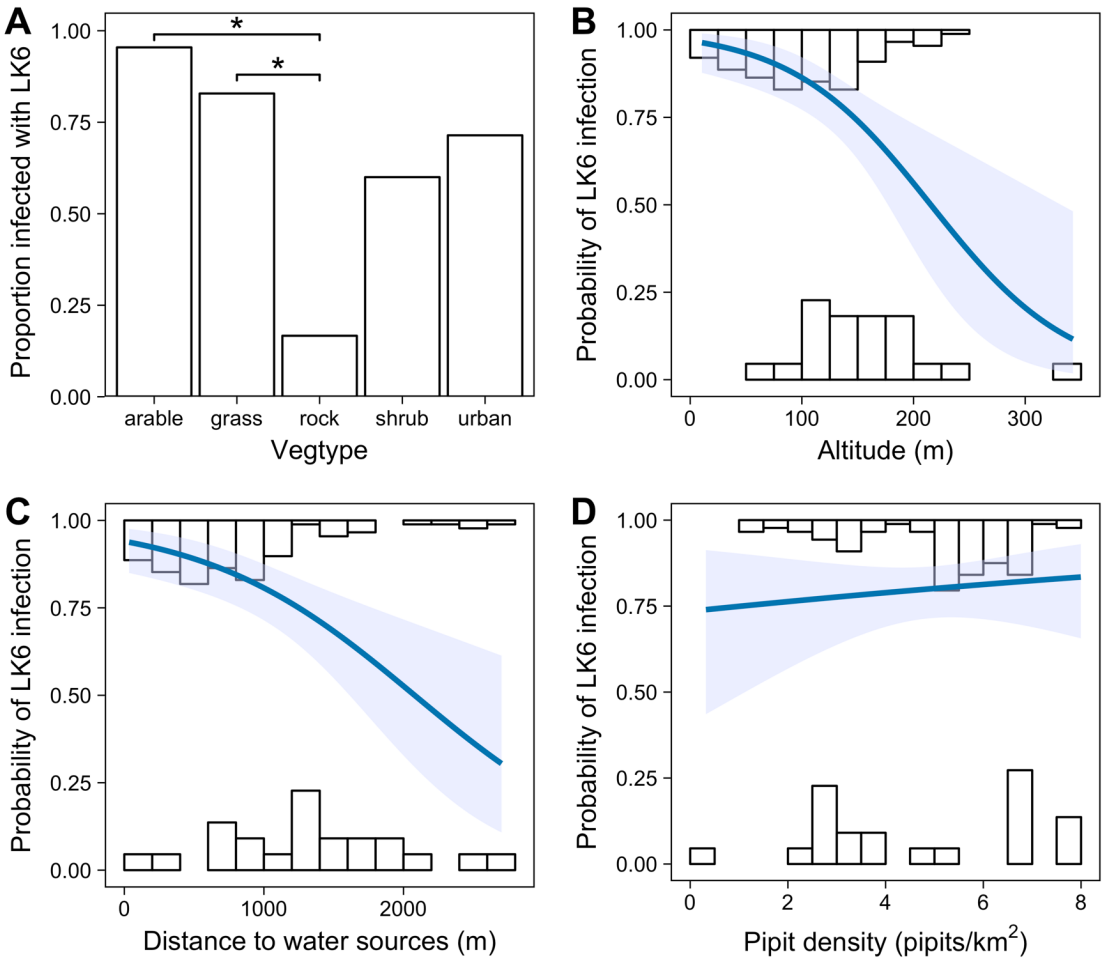


Figure 4.2. Environmental predictors of *Plasmodium* strain LK6 risk in adult Berthelot's pipits on Porto Santo. A) The proportion of samples infected with LK6 per category of VEGTYPE. Post-hoc Tukey tests found a significantly lower probability of LK6 infection from birds on rock-associated versus arable or grass habitats at $P < 0.05$ (*). Plots B to D show logistic regression models of the effect of B) altitude, C) distance to water sources, and D) pipit density on the probability of LK6 infection. Histograms indicate the frequency of uninfected (lower) and infected (upper) individuals at each 50 m altitude class (B), 200 m distance class (C), and 0.5 pipits/km^2 class. Shaded area represents 95% confidence intervals.

Table 4.6. Environmental predictors of malaria risk in adult Berthelot's pipits on Porto Santo. Parameter estimates (and 95% confidence intervals) are presented for the 'best' model with lowest AICc as determined by model selection, and for model averaging across 17 models with ΔAICc ≤ 7 relative to the best model. Averaged parameter estimates were calculated using the zero method of model averaging (Burnham & Anderson 2002). Relative importance for each predictor/interaction was calculated as the sum of Akaike weights over models in which the predictor/interaction occurs. Within VEGTYPE, coefficients of arable, rock, shrub and urban were calculated relative to grass as the reference category. Adjusted R^2 for the best model was calculated as McFadden's adjusted pseudo R^2.

	Best model	Model averaging	Relative importance
(Intercept)	2.62 (1.62, 4.04)	2.32 (1.10, 3.55)	
DISTWATER	-1.54 (-3.64, 0.67)	-1.79 (-3.83, 0.24)	0.99
DENSITY	0.63 (-0.59, 1.91)	0.66 (-0.64, 1.97)	0.98
DENSITY*DISTWATER	-5.33 (-9.01, -2.19)	-5.02 (-8.72, -1.32)	0.98
ALTITUDE	-1.91 (-3.85, -0.54)	-1.47 (-3.33, 0.39)	0.87
VEGTYPE	+	+	0.63
arable	-0.56 (-2.97, 2.58)	-0.31 (-2.42, 1.80)	
rock	-3.77 (-7.55, -1.10)	-2.42 (-6.91, 2.06)	
shrub	-3.37 (-7.28, -0.00)	-1.95 (-6.00, 2.11)	
urban	-1.59 (-3.85, 0.80)	-0.94 (-3.26, 1.38)	
DISTPOUL		-0.26 (-1.47, 0.94)	0.33
SLOPE		-0.05 (-1.27, 1.16)	0.29
Adjusted R^2	0.25		

4.4.3. Genetic associations with malaria infection

As we found high levels of LD between TLR4 SNPs, we calculated VIFs for models of genetic associations with malaria infection and risk. All VIF scores were < 3 in Tenerife. In Porto Santo, TLR4_3 and TLR4_4 had elevated VIF scores (> 50), however after removing the synonymous TLR4_3 SNP, all VIF scores were < 3.

Parameter estimates and statistical significance of genetic variants as predictors of malaria infection are summarised in Supplementary Table S4.1. In Porto Santo, across all sampling years, increasing frequency of the T allele in SNP 5239s1 was significantly associated with increased malaria infection (estimate = 0.75, SE = 0.34, P = 0.026; Fig. 4.5A). This effect was no longer significant when looking only at 2016 (P = 0.099), although the direction of the result remained consistent, with the highest probability of malaria prevalence in TT genotype individuals. This may be an effect of a reduction in power given that the 2016 sample set was smaller than the all years sample (n = 110 vs n = 136). No other SNP was significantly associated with malaria infection in either 2016 or across all years. There were no significant associations between malaria infection and either SNP heterozygosity or TLR4 protein haplotypes in Porto Santo.

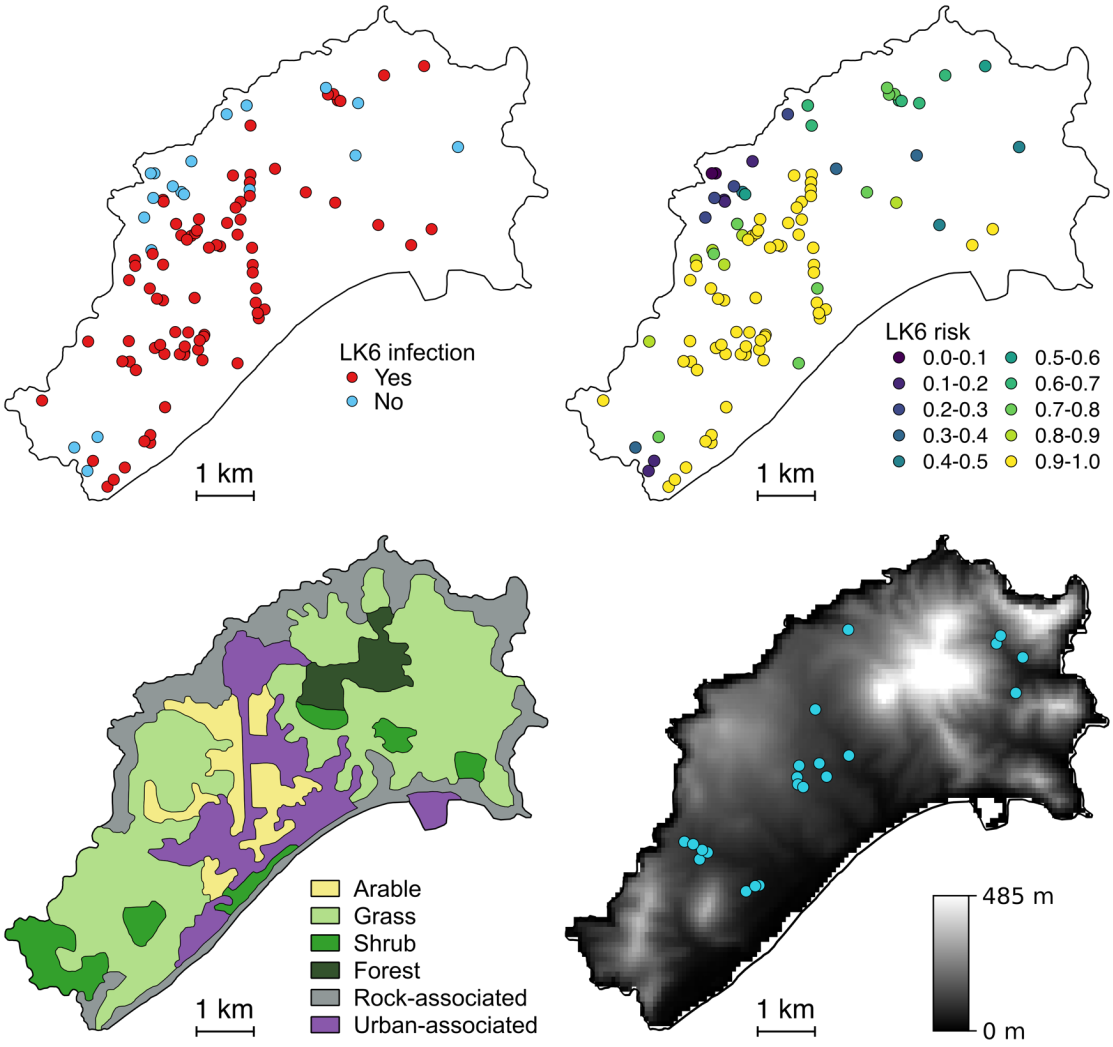


Figure 4.3. Spatial distribution of *Plasmodium* strain LK6 and environmental risk factors in adult Berthelot's pipits on Porto Santo. Points indicate location of (A) individual infection status and (B) infection risk in adult Berthelot's pipits. C) VEGTYPE categories, adapted from the CORINE Land Cover inventory (CLC 2012; www.eea.europa.eu). D) Altitude on Porto Santo was calculated from Shuttle Radar Topography Mission global elevation data (SRTM 90m; https://srtm.csi.cgiar.org). Overlaid blue points indicate locations of standing water sources.

There was a significant negative effect of TLR4_P1 presence on malaria infection on Tenerife in 2011 (estimate = -0.46, SE = 0.23, $P = 0.041$; Fig. 4.5B). We found the same trend across all years, although the association between malaria infection and TLR4_P1 presence was not significant ($P = 0.091$), potentially due to a larger proportion of samples originating from the malaria-free high-altitude El Teide site in earlier sampling years. The TLR4_P2 haplotype was not significantly associated with malaria infection in 2011 or across all years. We found no significant associations with malaria infection for SNP genotypes or SNP heterozygosity in Tenerife.

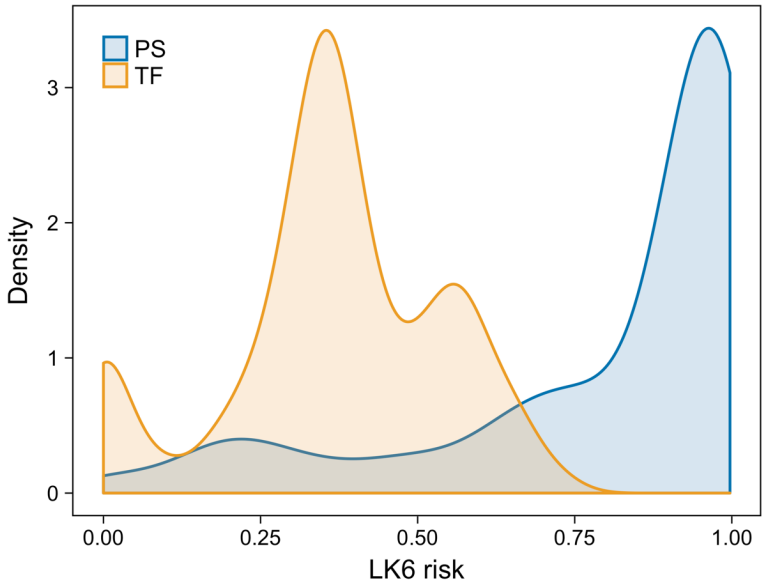


Figure 4.4. The distribution of *Plasmodium* strain LK6 infection risk in adult Berthelot's pipits on Porto Santo (PS; blue) and Tenerife (TF; orange).

4.4.4. Genetic associations with malaria risk

We then tested for associations between genetic variants and malaria risk on Porto Santo (2016) and Tenerife (2011). The following results are summarised in Supplementary Table S4.2. On Porto Santo, increasing numbers of T alleles at SNP 5239s1 (estimate = 0.69, SE = 0.27, $P = 0.011$), and A alleles at SNP TLR4_2 (estimate = 1.03, SE = 0.42, $P = 0.016$), were significantly associated with increased malaria risk. However, the residuals of this model were highly spatially autocorrelated. To control for this, we included seven dbMEMs which had high R^2 in single-predictor models of malaria risk (Supplementary Fig. S4.1), chosen from a set of dbMEMs which gave the lowest AICc in a multipredictor model of malaria risk. After controlling for autocorrelation, SNP 5239s1 was still significantly associated with malaria risk (estimate = 0.38, SE = 0.17, $P = 0.030$; Fig. 4.6), in contrast to TLR4_2 ($P = 0.423$). We did not find a significant association between TLR4_4 and malaria risk, either before or after controlling for autocorrelation. Hierarchical partitioning of the above models (Table 4.7) found that SNP 5239s1 explained 5.2% of the variance in malaria risk before controlling for autocorrelation, and 3.3% of the variance after the addition of dbMEMs. Despite having a nonsignificant association with malaria risk in the model containing dbMEMs, TLR4_2 explained a greater proportion of the variance in malaria risk compared to 5239s1, both before (7.4%) and after (4.3%) controlling for autocorrelation.

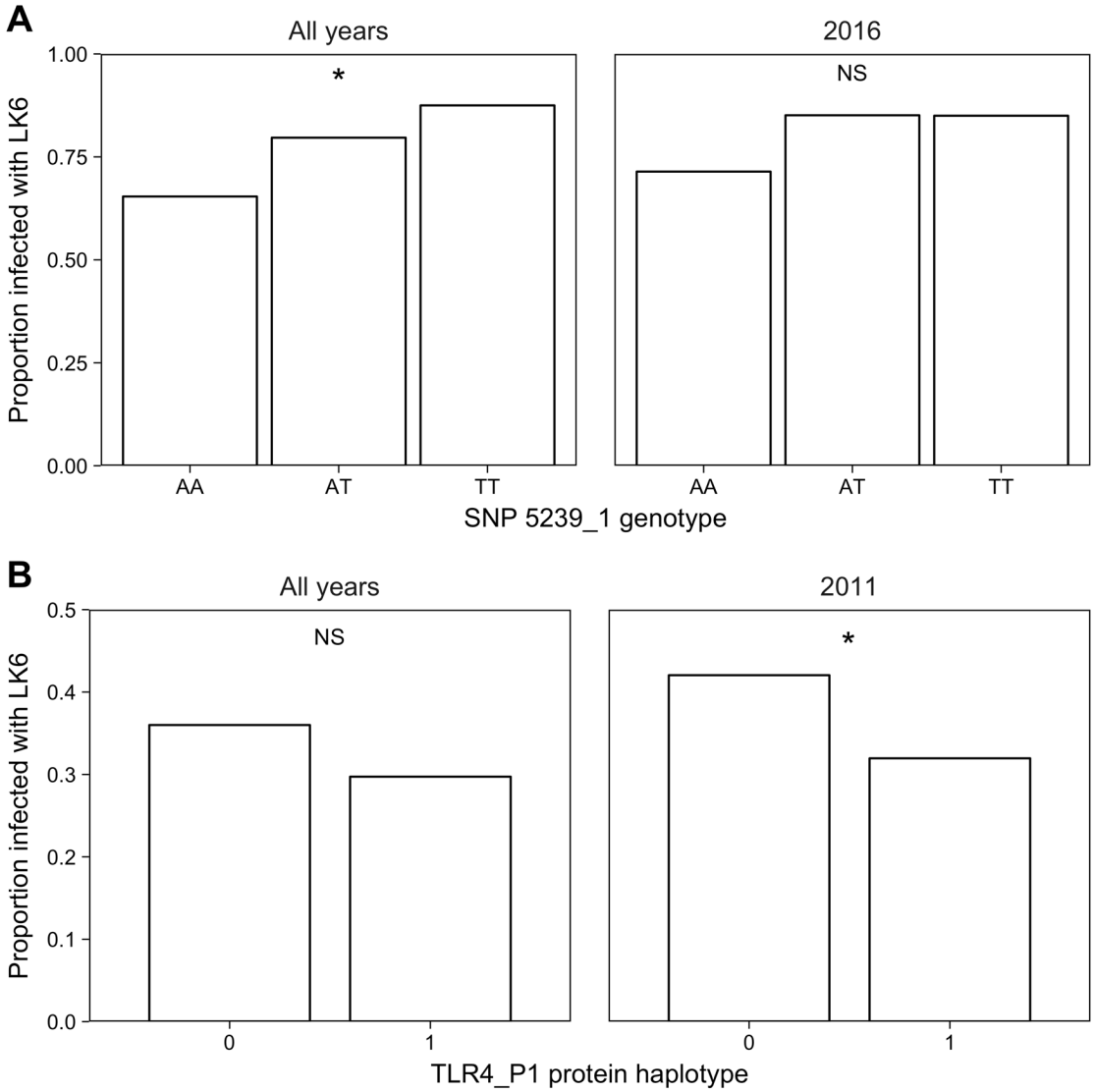


Figure 4.5. The association between *Plasmodium* strain LK6 infection status and genetic variants in adult Berthelot's pipits. A) SNP 5239s1 genotype was significantly associated with LK6 infection in Porto Santo across all sampling years, but was nonsignificant in 2016. B) Protein haplotype TLR4_P1 was significantly associated with LK6 infection in Tenerife in 2011 but not across all sampling years. 0 = absent, 1 = present. * denotes significance at $P < 0.05$; NS = not significant.

Before taking into account autocorrelation, there were no significant associations between SNP heterozygosity and malaria risk on Porto Santo; however, after including seven dbMEMs to remove spatial autocorrelation in model residuals, heterozygosity at SNP 5239s1 was significantly associated with increased malaria risk (estimate = 0.74, SE = 0.25, $P = 0.004$). Heterozygosity at other SNPs was not significantly associated with malaria risk.

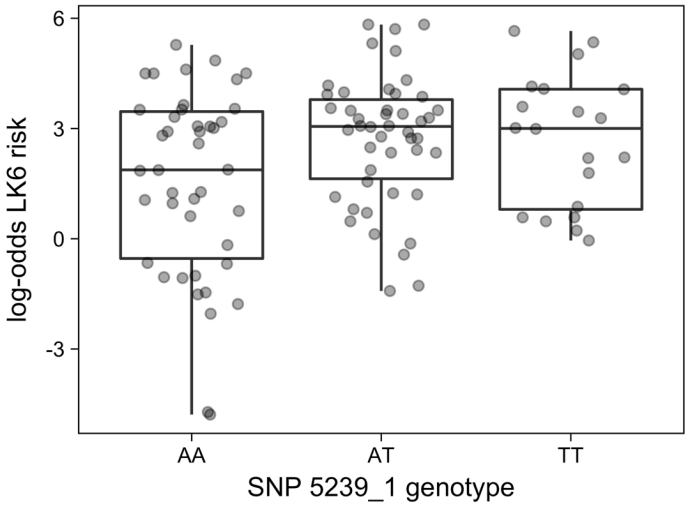


Figure 4.6. The association between *Plasmodium* strain LK6 infection risk and SNP 5239s1 in adult Berthelot's pipits on Porto Santo. LK6 risk was logit-transformed prior to model fitting; a log-odds risk score of 0 is equivalent to an infection probability of 0.5.

Table 4.7. Hierarchical partitioning of variance for predictors of malaria risk in Berthelot's pipits on Porto Santo. R^2 values were calculated from a general linear model with genotypes at SNPs 5239s1, TLR4_2 and TLR4_4 as predictors of malaria risk, with and without the inclusion of dbMEMs as predictors to control for autocorrelation.

Predictor	R^2	R^2 (with dbMEMs)
5239s1	0.052	0.033
TLR4_2	0.074	0.043
TLR4_4	0.039	0.027
dbMEMs		0.589

Prior to controlling for autocorrelation, the presence of protein haplotype TLR4_P1 was significantly associated with reduced malaria risk on Porto Santo (estimate = -1.32, SE = 0.62, P = 0.037). This model likewise had significant levels of spatial autocorrelation in model residuals, and required the same seven dbMEMs to control for it. Following the addition of dbMEMs, there was no longer a significant effect of TLR4_P1 presence (P = 0.318). No other TLR4 protein haplotypes showed significant associations with malaria risk.

To investigate the loss of significance of associations between TLR4_P1 or TLR4_2 and malaria risk on Porto Santo, we ran a binomial glm (TLR4_P1) and lm (TLR4_2) with the seven dbMEMs as predictors for the two genetic variants. dbMEM1 was significantly associated with both TLR4_P1 (estimate = -1.18, SE = 0.39, P = 0.003) and TLR4_2 (estimate = 0.16, SE = 0.06, P = 0.005).

Without dbMEMs, there were no significant associations between SNP heterozygosity and malaria risk on Porto Santo; however, after including seven dbMEMs to remove spatial autocorrelation in model residuals, heterozygosity at 5239s1 was significantly associated with increased malaria risk (estimate = 0.74, SE = 0.25, *P* = 0.004).

We did not find any significant associations between SNP genotype, SNP heterozygosity, or TLR4 protein haplotypes and malaria risk on Tenerife. We were unable to remove spatial autocorrelation in model residuals through the addition of dbMEMs as model predictors.

4.5. Discussion

We used previously identified candidate SNPs linked to malaria infection across populations (from a GWAS analysis performed on RAD-seq SNPs; Armstrong *et al.* 2018), and TLR4 SNPs (González-Quevedo *et al.* 2015) to investigate the relationship between potentially adaptive genetic variation and avian malaria within two island populations of Berthelot's pipits. In addition to testing for associations with infection status, we calculated the malaria risk at each sampling location, predicted by modelling fine-scale environmental drivers of malaria infection. We found evidence of associations between actual malaria infection and SNP 5239s1 in Porto Santo, and TLR4 protein haplotype 1 in Tenerife. Furthermore, the SNPs 5239s1 and TLR4_2 showed associations with malaria risk in Porto Santo, whereas in Tenerife, where malaria risk was lower, we did not find any genetic associations with malaria risk.

We have previously used RAD-seq SNPs to detect genetic variants that were significantly associated with LK6 infection in a GWAS of Berthelot's pipits in the Canary Islands (Armstrong *et al.* 2018). The strongest association was found for SNP 5239s1, which was situated ca. 2000 bp from interleukin-16 (IL-16), a proinflammatory cytokine. Though direct associations between IL-16 and malaria infection have not been found, IL-16 moderates the expression of other cytokines that are associated with malaria infection (Kern *et al.* 1989; Mathy *et al.* 2000; Lyke *et al.* 2004). In the present study, SNP 5239s1 was a significant predictor of malaria on Porto Santo, with the lowest infection and risk found in samples with the AA genotype. This was the opposite relationship to that found in the Canary Islands (Armstrong *et al.* 2018), where increased incidence of the T allele was associated with reduced infection. This may be indicative of pathogen-mediated balancing selection, which can arise from heterozygote advantage (Doherty & Zinkernagel 1975), rare-allele advantage (Takahata & Nei 1990; Slade & McCallum 1992), and local adaptation to fluctuating pathogen selection pressures (Hill *et al.* 1991). When controlling for spatial autocorrelation, we found a significant association between SNP 5239s1 heterozygosity and malaria risk on Porto Santo, although contrary to the heterozygote

advantage model of pathogen-mediated balancing selection, heterozygotes were associated with greater malaria risk than homozygotes (an effect which was largely driven by the decline in risk found with AA genotypes). Berthelot's pipit populations on the Madeiran and Canary Islands archipelagos have been isolated from each other for ca. 8000 years (Spurgin *et al.* 2014). Different populations may therefore be undergoing independent coevolutionary cycles with the same malaria strain, with alternative alleles conferring an advantage between divergent populations. In line with our findings, population-specific relationships between MHC alleles and susceptibility to a single malaria strain have been detected in other species, such as in house sparrows (Bonneaud *et al.* 2006). Alternatively, undetected genetic and phenotypic differences within the LK6 strain could potentially drive local adaptation between the archipelagos, with different alleles favoured in different populations (Alcaide *et al.* 2008; Loiseau *et al.* 2009). Here, we have used a single genetic marker, the mitochondrial cytochrome b locus, to classify the malaria strain. Several genes on the *Plasmodium* genome with relevance to infection success have shown greater genetic variation than at cytochrome b (Jarvi *et al.* 2008; Hellgren *et al.* 2013; Lauron *et al.* 2014). It is therefore possible that Berthelot's pipits on separate archipelagos could be adapting to different malaria strains within LK6, although this remains to be tested.

We did not find evidence of associations between SNP 5239s1 and malaria infection or risk on Tenerife, despite this population being included in the previous GWAS (Armstrong *et al.* 2018). SNPs that are related to individual-level variation in parasite burden do not necessarily show the same associations at the landscape scale (Wenzel *et al.* 2016). It is possible that with the comparatively low malaria risk in Tenerife, gene flow is overriding landscape-scale associations between SNP 5239s1 and malaria risk (Lenormand 2002). The previous GWAS result could have been driven by other populations such as Lanzarote and Fuerteventura, where malaria infection rates were higher (Illera *et al.* 2008; Spurgin *et al.* 2012).

Polymorphisms in genes of the immune system can alter the effectiveness of their proteins for detecting and responding to pathogens (Lazarus *et al.* 2002; Sommer 2005). The TLR4 SNPs sequenced here are situated within the ligand-binding region, which plays a key role in TLR pathogen recognition (Werling *et al.* 2009). Evidence of positive selection in birds or mammals has been detected at each of the codons in which the Berthelot's pipit TLR4 SNPs are found (Vinkler & Albrecht 2009; Wlasiuk & Nachman 2010; Areal *et al.* 2011; Králová *et al.* 2018), suggesting that variation at these sites may be important for the evolution of pathogen recognition. On Tenerife, we found that the presence of the TLR4 protein haplotype TLR4_P1 was associated with decreased malaria infection prevalence in 2011, but not across all sampling years. In earlier years, approximately half of samples were collected from the high-altitude (> 2000 m above sea level) plateau of El Teide. Avian malaria has not been found in Berthelot's

pipits in this location (Illera *et al.* 2008; Spurgin *et al.* 2012; González-Quevedo *et al.* 2014), although a survey of passerine communities on Tenerife found malaria at low frequency in high-altitude habitats (Padilla *et al.* 2017). The relationship seen in 2011 between TLR4_P1 and infection may be masked in other sampling years by the increase in uninfected individuals from areas of low malaria abundance. We did not find a relationship between TLR4_P1 presence and malaria risk, potentially due to the limitations of the Tenerife malaria risk model outlined above. On Porto Santo, both TLR4_P1 and the SNP TLR4_2 were significantly associated with malaria risk, although these relationships were no longer significant after including dbMEMs to remove autocorrelation in model residuals. Both of these genetic variants showed significant associations with dbMEM1, which itself explained 22% of the variance in malaria risk, making it difficult to disentangle the real effects of these variants from any spurious associations arising from residual autocorrelation.

In this study we focused on the *Plasmodium* strain LK6. This strain has been found in a number of passerines across the Macaronesian islands (Illera *et al.* 2015; Padilla *et al.* 2017) and northwestern Africa (Mata *et al.* 2015), and in falcons in Spain and the Canary Islands (Ortego *et al.* 2007; Gutiérrez-López *et al.* 2015), suggesting that LK6 is a generalist parasite with large host phylogenetic diversity. The benefit of increased transmission through multiple hosts is predicted to trade off against virulence (Leggett *et al.* 2013). This could mean that the severity of infection by LK6, and hence the strength of selection against it, is relatively lower than more host-specific pathogens, potentially making it harder to detect genetic associations with this strain. Increased mortality during the acute stage of infection has been documented in other avian malaria strains (Atkinson *et al.* 2000; Lachish *et al.* 2011; Sepil *et al.* 2013). Although reductions in fitness traits have been found with chronic LK6 infections (Ortego *et al.* 2008), there is a lack of research into the mortality effects of acute LK6 infection.

We modelled the environmental predictors of malaria distributions in Porto Santo to understand fine-scale spatial differences in malaria risk. Higher altitudes and increased distance to water were associated with decreased probability of malaria infection, whereas a positive association was found between pipit density and infection. Distance to water and pipit density negatively interacted, resulting in high malaria prevalence in areas close to water with large aggregations of pipits. Previous work in Tenerife found the main drivers of avian malaria distributions were minimum temperature of the coldest month, distance to artificial water sources, and distance to poultry farms (González-Quevedo *et al.* 2014). In both Tenerife and Porto Santo, high collinearity was found between altitude, minimum temperature and precipitation, with strong environmental links between the three variables. The same climatic processes are therefore likely to be influencing malaria distributions on each island. This is perhaps not surprising as

malaria vector distributions are constrained by thermal requirements, with decreased malaria prevalence often reported at high altitudes (Eggert *et al.* 2008; Niebuhr *et al.* 2016) and low temperatures (Craig *et al.* 1999; Blanford *et al.* 2013; Loiseau *et al.* 2013).

Distance to water sources was an important predictor of malaria distributions in both Porto Santo and Tenerife (González-Quevedo *et al.* 2014). Standing water is essential for larval development of mosquitoes, resulting in higher malaria and vector abundance in close proximity to water sources (Ganser *et al.* 2016; Illera *et al.* 2017; Ferraguti *et al.* 2018). In the present study, distance to urban areas was removed prior to model selection due to high collinearity with distance to water sources. Therefore we cannot rule out the importance of additional sources of standing water that may be associated with urban environments. Other studies have found links between urbanisation and increased malaria or vector abundance (Alemu *et al.* 2011; Li *et al.* 2014; Reinoso-Pérez *et al.* 2016), although this appears to vary between vector species, with some favouring more natural habitats (Johnson *et al.* 2008; Ferraguti *et al.* 2016).

Pipit density was positively associated with malaria risk on Porto Santo, although compared to other predictors of malaria risk, the model-averaged parameter estimate was relatively small. There was, however, a strong negative interaction between distance to water sources and pipit density on this island. With the scarcity of standing water available throughout Porto Santo, these water sources may attract higher densities of pipits than found in areas far from water. Higher densities of hosts are expected to increase disease transmission rates (Begon *et al.* 2002; Greer *et al.* 2008; Raghwani *et al.* 2011). Standing water is necessary for breeding of malaria vectors, and female mosquitoes have been shown to aggregate around water sources (Le Menach *et al.* 2005). The potential attraction of standing water for both hosts and vectors is likely to explain the elevated levels of malaria found at high densities in close proximity to water sources on Porto Santo. Pipit density itself was not an important predictor of malaria distributions in Tenerife (González-Quevedo *et al.* 2014), though as noted by the authors, defining density based on the presence or absence of a caught bird in each 1 km^2 (as they did) may not be the most appropriate measure. In the present study, we calculated density based on the number of sampled birds within a 1 km radius of each sampling location. While this measure is unlikely to accurately reflect the true density of pipits at each site, it may function as an indicator of relative density, as we would expect to catch a greater number of birds in areas with higher densities.

Vegetation type was associated with malaria prevalence on Porto Santo. The highest abundance of malaria was found in arable and grassland habitats, with significantly lower malaria in rock-associated habitats. However, this should be interpreted with caution due to small sample sizes, as only six pipits were caught on rock-associated habitats (one of which was infected). Of the

predictors that were included in the model of malaria risk, vegetation type had the lowest relative importance in model averaging (0.63). Despite this, there does appear to be an important role of vegetation type for explaining the spatial distributions of malaria: calculating malaria risk using a model without vegetation type resulted in an inability to control for spatial autocorrelation using dbMEMs (results not shown). While not an important predictor of malaria infection in Tenerife (González-Quevedo *et al.* 2014), differences in malaria and vector abundances between vegetation types have been found elsewhere (Rubio-Palis & Zimmerman 1997; Clark *et al.* 2016; Ferreira Junior *et al.* 2017).

Contrary to findings from Tenerife (González-Quevedo *et al.* 2014), distance to poultry was not an important predictor of malaria prevalence on Porto Santo. As farm census data was not publicly available for Porto Santo, poultry locations were identified using visual observations, and we likely missed a number of occurrences. However, the difference in the effects of poultry between the two islands is perhaps unsurprising due to the nature of poultry farming on the two islands: in contrast to the commercial-scale poultry farming on Tenerife, poultry were only ever found in small numbers on Porto Santo (often < 5 birds per sighting). Potential disease reservoir effects, or the increased aggregation of wild birds around poultry farms, that may be driving increased malaria abundance on Tenerife (González-Quevedo *et al.* 2014) are less likely at the small scale of poultry farming witnessed on Porto Santo.

The malaria risk model in Porto Santo had a higher McFadden pseudo R^2 (R^2 = 0.25) than the risk model in Tenerife (R^2 = 0.10) for explaining malaria distributions in Berthelot's pipits. The ability to detect significant ecological relationships with environmental gradients can be highly scale-dependent (Lomolino & Creighton 1996; Steffan-Dewenter *et al.* 2002; Holland *et al.* 2005). The smaller area of Porto Santo compared to Tenerife (< 50 km^2 vs. > 2000 km^2) might be a more suitable spatial scale for assessing the climatic and landscape drivers of malaria distributions. Alternatively, differences between the islands could be driven by vector community composition. Habitat suitability and niche breadth differs between mosquito species (Mughini-Gras *et al.* 2014); it is possible that vector species in Tenerife could have lower habitat specificity than in Porto Santo and therefore be less constrained by environmental conditions. However, without knowledge of the vector species on the two islands, at present we are unable to test this theory.

By testing for associations with malaria infection and risk at TLR4 and novel malaria-associated SNPs in divergent populations, we have revealed contrasting patterns of malaria risk and potential local adaptation between the two islands, in addition to genetic associations with environmentally-driven fine-scale spatial variation in malaria risk at the landscape scale within Porto Santo. A lack of genetic associations with malaria risk in Tenerife may indicate the

importance of spatial scales for assessing local adaptation across landscapes, where fine-scale associations may be obscured over larger areas. Understanding the processes of local adaptation and the environmental drivers of infectious disease will be of additional importance for conservation efforts, as future climatic fluctuations alter the prevalence of disease.

4.6. References

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

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In: *Second international symposium on information theory* (eds. Petrov, B.N. & Csaki, F.). Akademiai Kiado, Budapest, pp. 267–281.

Alcaide, M., Edwards, S.V., Negro, J.J., Serrano, D. & Tella, J.L. (2008). Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (*Falco naumanni*). *Molecular Ecology*, 17, 2652–2665.

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.

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

Armstrong, C., Richardson, D.S., Hipperson, H., Horsburgh, G.J., Küpper, C. & Percival-Alwyn, L. *et al.* (2018). Genomic associations with bill length and disease reveal drift and selection across island bird populations. *Evolution Letters*, 2, 22–36.

Atkinson, C.T., Dusek, R.J., Woods, K.L. & Iko, W.M. (2000). Pathogenicity of avian malaria in experimentally-infected Hawaii amakihi. *Journal of Wildlife Diseases*, 36, 197–201.

Atkinson, C.T., Saili, K.S., Utzurrum, R.B. & Jarvi, S.I. (2013). Experimental evidence for evolved tolerance to avian malaria in a wild population of low elevation Hawai'i 'Amakihi (*Hemignathus virens*). *EcoHealth*, 10, 366–375.

Bartoń, K. (2018). MuMIn: Multi-Model Inference. R package version 1.40.4. https://CRAN.R-project.org/package=MuMIn.

Bateson, Z.W., Hammerly, S.C., Johnson, J.A., Morrow, M.E., Whittingham, L.A. & Dunn, P.O. (2016). Specific alleles at immune genes, rather than genome-wide heterozygosity, are related to immunity and survival in the critically endangered Attwater's prairie-chicken. *Molecular Ecology*, 25, 4730–4744.

Begon, M., Bennett, M., Bowers, R.G., French, N.P., Hazel, S.M. & Turner, J. (2002). A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiology and Infection*, 129, 147–153.

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.

Bjornstad, O.N. (2018). ncf: Spatial Covariance Functions. R package version 1.2-4. https://CRAN.R-project.org/package=ncf.

Blanford, J.I., Blanford, S., Crane, R.G., Mann, M.E., Paaijmans, K.P. & Schreiber, K.V. *et al.* (2013). Implications of temperature variation for malaria parasite development across Africa. *Scientific Reports*, 3, 1300.

Bockelmann, A.C., Reusch, B.H., Bijlsma, R. & Bakker, J.P. (2003). Habitat differentiation vs. isolation-by-distance: the genetic population structure of *Elymus athericus* in European salt marshes. *Molecular Ecology*, 12, 505–515.

Bonneaud, C., Pérez-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. & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, 153, 51–68.

Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic-Approach*. 2nd edn. Springer, Berlin.

Burnham, K.P., Anderson, D.R. & Huyvaert, K.P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65, 23–35.

Bødker, R., Akida, J., Shayo, D., Kisinza, W., Msangeni, H.A. & Pedersen, E.M. *et al.* (2003). Relationship between altitude and intensity of malaria transmission in the Usambara Mountains, Tanzania. *Journal of Medical Entomology*, 40, 706–717.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7.

Clark, N.J., Clegg, S.M. & Lima, M.R. (2014). A review of global diversity in avian haemosporidians (*Plasmodium* and *Haemoproteus*: Haemosporida): new insights from molecular data. *International Journal for Parasitology*, 44, 329–338.

Clark, N.J., Wells, K., Dimitrov, D. & Clegg, S.M. (2016). Co-infections and environmental conditions drive the distributions of blood parasites in wild birds. *Journal of Animal Ecology*, 85, 1461–1470.

Craig, M.H., Le Sueur, D. & Snow, R.W. (1999). A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today*, 15, 105–111.

Cramp, S. (1988). *The Birds of the Western Palearctic, Volume 5*. Oxford University Press, Oxford, UK.

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

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J. & Carl, G. *et al.* (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609–628.

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S. & Guenard, G. *et al.* (2018). adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.2-0. https://CRAN.R-project.org/package=adespatial.

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

Eggert, L.S., Terwilliger, L.A., Woodworth, B.L., Hart, P.J., Palmer, D. & Fleischer, R.C. (2008). Genetic structure along an elevational gradient in Hawaiian honeycreepers reveals contrasting evolutionary responses to avian malaria. *BMC Evolutionary Biology*, 8, 315.

Fallon, S.M., Bermingham, E. & Ricklefs, R.E. (2005). Host specialization and geographic localization of avian malaria parasites: a regional analysis in the Lesser Antilles. *The American Naturalist*, 165, 466–480.

Fellay, J., Shianna, K.V., Ge, D., Colombo, S., Ledergerber, B. & Weale, M. *et al.* (2007). A whole-genome association study of major determinants for host control of HIV-1. *Science*, 317, 944–947.

Ferraguti, M., Martínez-de la Puente, J., Bensch, S., Roiz, D., Ruiz, S. & Viana, D.S. *et al.* (2018). Ecological determinants of avian malaria infections: An integrative analysis at landscape, mosquito and vertebrate community levels. *Journal of Animal Ecology*, 87, 727–740.

Ferraguti, M., Martínez-De La Puente, J., Roiz, D., Ruiz, S., Soriguer, R. & Figuerola, J. (2016). Effects of landscape anthropization on mosquito community composition and abundance. *Scientific Reports*, 6, 29002.

Ferreira Junior, F.C., Rodrigues, R.A., Ellis, V.A., Leite, L.O., Borges, M.A. & Braga, E.M. (2017). Habitat modification and seasonality influence avian haemosporidian parasite distributions in southeastern Brazil. *PLoS ONE*, 12, e0178791.

Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marquès-Bonet, T., Ramírez-Soriano, A. & Muntasell, A. *et al.* (2008). Balancing selection is the main force shaping the evolution of innate immunity genes. *The Journal of Immunology*, 181, 1315–1322.

Ferwerda, B., McCall, M.B.B., Alonso, S., Giamarellos-Bourboulis, E.J., Mouktaroudi, M. & Izagirre, N. *et al.* (2007). TLR4 polymorphisms, infectious diseases, and evolutionary pressure during migration of modern humans. *Proceedings of the National Academy of Sciences*, 104, 16645–16650.

Fick, S.E. & Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302–4315.

Flint, J., Harding, R.M., Boyce, A.J. & Clegg, J.B. (1998). The population genetics of the haemoglobinopathies. *Bailliere's Clinical Haematology*, 11, 1–51.

Foster, J.T., Tweed, E.J., Camp, R.J., Woodworth, B.L., Adler, C.D. & Telfer, T. (2004). Long-term population changes of native and introduced birds in the Alaka'i Swamp, Kaua'i. *Conservation Biology*, 18, 716–725.

Fox, J. & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 87, 178–183.

Fox, J. & Weisberg, S. (2011). *An R Companion to Applied Regression*. Second. Sage, Thousand Oaks, CA.

Fu, W.-X., Liu, Y., Lu, X., Niu, X.-Y., Ding, X.-D. & Liu, J.-F. *et al.* (2012). A genome-wide association study identifies two novel promising candidate genes affecting *Escherichia coli* F4ab/F4ac susceptibility in swine. *PLoS ONE*, 7, e32127.

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L. & Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, 7, e1002355.

Ganser, C., Gregory, A.J., Mcnew, L.B., Hunt, L.A., Sandercock, B.K. & Wisely, S.M. (2016). Fine-scale distribution modeling of avian malaria vectors in north-central Kansas. *Journal of Vector Ecology*, 41, 114–122.

Garroway, C.J., Radersma, R., Sepil, I., Santure, A.W., De Cauwer, I. & Slate, J. *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.

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

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27, 2865–2873.

Gelman, A. & Su, Y.-S. (2018). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. R package version 1.10-1. https://CRAN.R-project.org/package=arm.

Gilbert, S.C., Plebanski, M., Gupta, S., Morris, J., Cox, M. & Aidoo, M. *et al.* (1998). Association of malaria parasite population structure, HLA, and immunological antagonism. *Science*, 279, 1173–1177.

González-Quevedo, C., Davies, R.G. & Richardson, D.S. (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.

González-Quevedo, C., Davies, R.G., Phillips, K.P., Spurgin, L.G. & Richardson, D.S. (2016). Landscape-scale variation in an anthropogenic factor shapes immune gene variation within a wild population. *Molecular Ecology*, 25, 4234–4246.

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

Greer, A.L., Briggs, C.J. & Collins, J.P. (2008). Testing a key assumption of host-pathogen theory: density and disease transmission. *Oikos*, 117, 1667–1673.

Griffiths, R., Double, M.C., Orr, K. & Dawson, R.J.G. (1998). A DNA test to sex most birds. *Molecular Ecology*, 7, 1071–1075.

Grömping, U. (2006). Relative importance for linear regression in R: the package relaimpo. *Journal of Statistical Software*, 17, 1–27.

Grueber, C.E., Nakagawa, S., Laws, R.J. & Jamieson, I.G. (2011). Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology*, 24, 699–711.

Grueber, C.E., Wallis, G.P., King, T.M. & Jamieson, I.G. (2012). Variation at innate immunity toll-like receptor genes in a bottlenecked population of a New Zealand robin. *PLoS ONE*, 7, e45011.

Guggisberg, A.M., Sayler, K.A., Wisely, S.M. & Odom John, A.R. (2018). Natural history of *Plasmodium odocoilei* malaria infection in farmed white-tailed deer. *mSphere*, 3, e00067–18.

Gutiérrez-López, R., Gangoso, L., Martínez-De La Puente, J., Fric, J., López-López, P. & Mailleux, M. *et al.* (2015). Low prevalence of blood parasites in a long-distance migratory raptor: the importance of host habitat. *Parasites and Vectors*, 8, 189.

Haas, M., Lukáň, M., Kisková, J. & Hrehová, Z. (2012). Occurrence of blood parasites and intensity of infection in *Prunella modularis* in the montane and subalpine zone in the Slovak Carpathians. *Acta Parasitologica*, 57, 221–227.

Harrigan, R.J., Sedano, R., Chasar, A.C., Chaves, J.A., Nguyen, J.T. & Whitaker, A. *et al.* (2014). New host and lineage diversity of avian haemosporidia in the northern Andes. *Evolutionary Applications*, 7, 799–811.

He, D., Tao, S., Guo, S., Li, M., Wu, J. & Huang, H. *et al.* (2015). Interaction of TLR-IFN and HLA polymorphisms on susceptibility of chronic HBV infection in Southwest Han Chinese. *Liver International*, 35, 1941–1949.

Hellgren, O., Kutzer, M., Bensch, S., Valkiunas, G. & Palinauskas, V. (2013). Identification and characterization of the merozoite surface protein 1 (*msp1*) gene in a host-generalist avian malaria parasite, *Plasmodium relictum* (lineages SGS1 and GRW4) with the use of blood transcriptome. *Malaria Journal*, 12, 381.

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist*, 173, 579–588.

Hill, A.V., Allsopp, C.E., Kwiatkowski, D., Anstey, N.M., Twumasi, P. & Rowe, P.A. *et al.* (1991). Common west African HLA antigens are associated with protection from severe malaria. *Nature*, 352, 595–600.

Holland, J.D., Fahrig, L. & Cappuccino, N. (2005). Body size affects the spatial scale of habitat-beetle interactions. *Oikos*, 110, 101–108.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.

Hurvich, C.M. & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76, 297–307.

Illera, J.C., Emerson, B.C. & Richardson, D.S. (2007). Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology*, 16, 4599–4612.

Illera, J.C., Emerson, B.C. & Richardson, D.S. (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.

Illera, J.C., Fernández-Álvarez, Á., Hernández-Flores, C.N. & Foronda, P. (2015). Unforeseen biogeographical patterns in a multiple parasite system in Macaronesia. *Journal of Biogeography*, 42, 1858–1870.

Illera, J.C., López, G., García-Padilla, L. & Moreno, Á. (2017). Factors governing the prevalence and richness of avian haemosporidian communities within and between temperate mountains. *PLoS ONE*, 12, e0184587.

Imura, T., Suzuki, Y., Ejiri, H., Sato, Y., Ishida, K. & Sumiyama, D. *et al.* (2012). Prevalence of avian haematozoa in wild birds in a high-altitude forest in Japan. *Veterinary Parasitology*, 183, 244–248.

Inci, A., Yildirim, A., Njabo, K.Y., Duzlu, O., Biskin, Z. & Ciloglu, A. (2012). Detection and molecular characterization of avian *Plasmodium* from mosquitoes in central Turkey. *Veterinary Parasitology*, 188, 179–184.

Iwalokun, B.A., Oluwadun, A., Iwalokun, S.O. & Agomo, P. (2015). Toll-like receptor (TLR4) Asp299Gly and Thr399Ile polymorphisms in relation to clinical falciparum malaria

among Nigerian children: a multisite cross-sectional immunogenetic study in Lagos. *Genes and Environment*, 37, 3.

Janeway, C.A. & Medzhitov, R. (2002). Innate immune recognition. *Annual Review of Immunology*, 20, 197–216.

Jarvi, S.I., Farias, M.E.M. & Atkinson, C.T. (2008). Genetic characterization of Hawaiian isolates of *Plasmodium relictum* reveals mixed-genotype infections. *Biology Direct*, 3, 25.

Jean, K., Burnside, W.R., Carlson, L., Smith, K. & Guégan, J.F. (2016). An equilibrium theory signature in the island biogeography of human parasites and pathogens. *Global Ecology and Biogeography*, 25, 107–116.

Jepson, A., Banya, W., Sisay-joof, F., Hassan-King, M., Nunes, C. & Bennett, S. *et al.* (1997a). Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens. *Infection and Immunity*, 65, 872–876.

Jepson, A., Sisay-Joof, F., Banya, W., Hassan-King, M., Frodsham, A. & Bennett, S. *et al.* (1997b). Genetic linkage of mild malaria to the major histocompatibility complex in Gambian children: study of affected sibling pairs. *BMJ*, 315, 96–97.

Johnson, M.F., Gómez, A. & Pinedo-Vasquez, M. (2008). Land use and mosquito diversity in the Peruvian Amazon. *Journal of Medical Entomology*, 45, 1023–1030.

Jones, M.R., Cheviron, Z.A. & Carling, M.D. (2013). Spatial patterns of avian malaria prevalence in *Zonotrichia capensis* on the western slope of the Peruvian Andes. *The Journal of Parasitology*, 99, 903–905.

Jones, M.R., Cheviron, Z.A. & Carling, M.D. (2015). Spatially variable coevolution between a haemosporidian parasite and the MHC of a widely distributed passerine. *Ecology and Evolution*, 5, 1045–1060.

Kern, P., Hemmer, C.J., Damme, J.V., Gruss, H.-J. & Dietrich, M. (1989). Elevated tumor necrosis factor alpha and interleukin-6 serum levels as markers for complicated *Plasmodium falciparum* malaria. *The American Journal of Medicine*, 87, 139–143.

Kilpatrick, A.M. & LaPointe, D.A. (2006). Effects of chronic avian malaria (*Plasmodium relictum*) infection on reproductive success of Hawaii Amakihi (*Hemignathus virens*). *The Auk*, 123, 764–774.

Knowles, S.C.L., Palinauskas, V. & Sheldon, B.C. (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.

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

Lachish, S., Knowles, S.C.L., Alves, R., Wood, M.J. & Sheldon, B.C. (2011). Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. *Journal of Animal Ecology*, 80, 1196–1206.

Langin, K.M., Sillett, T.S., Funk, W.C., Morrison, S.A., Desrosiers, M.A. & Ghalambor, C.K. (2015). Islands within an island: repeated adaptive divergence in a single population. *Evolution*, 69, 653–665.

Larson, W.A., Lisi, P.J., Seeb, J.E., Seeb, L.W. & Schindler, D.E. (2016). Major histocompatibility complex diversity is positively associated with stream water temperatures in proximate populations of sockeye salmon. *Journal of Evolutionary Biology*, 29, 1846–1859.

Lauron, E.J., Oakgrove, K.S., Tell, L.A., Biskar, K., Roy, S.W. & Sehgal, R.N. (2014). Transcriptome sequencing and analysis of *Plasmodium gallinaceum* reveals polymorphisms and selection on the apical membrane antigen-1. *Malaria Journal*, 13, 382.

Lazarus, R., Vercelli, D., Palmer, L.J., Klimecki, W.J., Silverman, E.K. & Richter, B. *et al.* (2002). Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. *Immunological Reviews*, 190, 9–25.

Le Menach, A., McKenzie, F.E., Flahault, A. & Smith, D.L. (2005). The unexpected importance of mosquito oviposition behaviour for malaria: non-productive larval habitats can be sources for malaria transmission. *Malaria Journal*, 4, 23.

Legendre, P. & Legendre, L. (2012). *Numerical Ecology*. 3rd edn. Elsevier, Amsterdam.

Leggett, H.C., Buckling, A., Long, G.H. & Boots, M. (2013). Generalism and the evolution of parasite virulence. *Trends in Ecology and Evolution*, 28, 592–596.

Lennon, J.J. (2000). Red-shifts and red herrings in geographical ecology. *Ecography*, 23, 101–113.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology and Evolution*, 17, 183–189.

Levene, H. (1953). Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, 87, 331–333.

Li, Y., Kamara, F., Zhou, G., Puthiyakunnon, S., Li, C. & Liu, Y. *et al.* (2014). Urbanization increases *Aedes albopictus* larval habitats and accelerates mosquito development and survivorship. *PLoS Neglected Tropical Diseases*, 8, e3301.

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

Loiseau, C., Harrigan, R.J., Bichet, C., Julliard, R., Garnier, S. & Lendvai, Á.Z. *et al.* (2013). Predictions of avian *Plasmodium* expansion under climate change. *Scientific Reports*, 3, 1126.

Loiseau, C., Richard, M., Garnier, S., Chastel, O., Julliard, R. & Zoorob, R. *et al.* (2009). Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Molecular Ecology*, 18, 1331–1340.

Loiseau, C., Zoorob, R., Robert, A., Chastel, O., Julliard, R. & Sorci, G. (2011). *Plasmodium relictum* infection and MHC diversity in the house sparrow (*Passer domesticus*). *Proceedings of the Royal Society B: Biological Sciences*, 278, 1264–1272.

Lomolino, M.V. & Creighton, J.C. (1996). Habitat selection, breeding success and conservation of the endangered American burying beetle *Nicrophorus americanus*. *Biological Conservation*, 77, 235–241.

Lorenz, E., Mira, J.P., Frees, K.L. & Schwartz, D.A. (2002). Relevance of mutations in the TLR4 receptor in patients with gram-negative septic shock. *Archives of Internal Medicine*, 162, 1028–1032.

Louviere, J.J., Hensher, D.A. & Swait, J.D. (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge University Press, Cambridge, UK.

Lukacs, P.M., Thompson, W.L., Kendall, W.L., Gould, W.R., Doherty, P.F. & Burnham, K.P. *et al.* (2007). Concerns regarding a call for pluralism of information theory and hypothesis testing. *Journal of Applied Ecology*, 44, 456–460.

Lyke, K.E., Burges, R., Cissoko, Y., Sangare, L., Dao, M. & Diarra, I. *et al.* (2004). Serum levels of the proinflammatory cytokines Interleukin-1 Beta (IL-1β), IL-6, IL-8, IL-10, Tumor Necrosis Factor Alpha, and IL-12 (p70) in Malian children with severe *Plasmodium falciparum* malaria and matched uncomplicated malaria or healthy cont. *Infection and Immunity*, 72, 5630–5637.

MacArthur, R.H. & Wilson, E.O. (1967). *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ.

Martinsen, E.S., Perkins, S.L. & Schall, J.J. (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., Bensch, S., Reviriego, M., Balbontin, J. & De Lope, F. (2008). Effects of malaria double infection in birds: one plus one is not two. *Journal of Evolutionary Biology*, 21, 979–987.

Marzal, A., De Lope, F., Navarro, C. & Møller, A.P. (2005). Malarial parasites decrease reproductive success: an experimental study in a passerine bird. *Oecologia*, 142, 541–545.

Mata, V.A., Silva, L.P. da, Lopes, R.J. & Drovetski, S.V. (2015). The Strait of Gibraltar poses an effective barrier to host-specialised but not to host-generalised lineages of avian Haemosporidia. *International Journal for Parasitology*, 45, 711–719.

Mathy, N.L., Scheuer, W., Lanzendörfer, M., Honold, K., Ambrosius, D. & Norley, S. *et al.* (2000). Interleukin-16 stimulates the expression and production of pro-inflammatory cytokines by human monocytes. *Immunology*, 100, 63–69.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In: *Frontiers in econometrics* (ed. Zarembka, P.). Academic Press, New York, NY.

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

Meyer, D. & Thomson, G. (2001). How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics*, 1996, 1–26.

Mockenhaupt, F.P., Cramer, J.P., Hamann, L., Stegemann, M.S., Eckert, J. & Oh, N.R. *et al.* (2006). Toll-like receptor (TLR) polymorphisms in African children: common TLR-4 variants predispose to severe malaria. *Proceedings of the National Academy of Sciences*, 103, 177–182.

Moormann, A.M., Embury, P.E., Opondo, J., Sumba, O.P., Ouma, J.H. & Kazura, J.W. *et al.* (2003). Frequencies of sickle cell trait and glucose-6-phosphate dehydrogenase deficiency differ in highland and nearby lowland malaria-endemic areas of Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 97, 513–514.

Moro, D., Lawson, M.A., Hobbs, R.P. & Thompson, R.C.A. (2003). Pathogens of house mice on arid Boullanger Island and subantarctic Macquarie Island, Australia. *Journal of Wildlife Diseases*, 39, 762–771.

Mughini-Gras, L., Mulatti, P., Severini, F., Boccolini, D., Romi, R. & Bongiorno, G. *et al.* (2014). Ecological niche modelling of potential West Nile virus vector mosquito species and their geographical association with equine epizootics in Italy. *EcoHealth*, 11, 120–132.

Netea, M.G., Wijmenga, C. & O'Neill, L.A. (2012). Genetic variation in Toll-like receptors and disease susceptibility. *Nature Immunology*, 13, 535–542.

Niebuhr, C.N., Poulin, R. & Tompkins, D.M. (2016). Is avian malaria playing a role in native bird declines in New Zealand? Testing hypotheses along an elevational gradient. *PLoS ONE*, 11, e0165918.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P. & McGlinn, D. *et al.* (2018). vegan: Community Ecology Package. R package version 2.5-2. https://CRAN.R-project.org/package=vegan.

Olsson-Pons, S., Clark, N.J., Ishtiaq, F. & Clegg, S.M. (2015). Differences in host species relationships and biogeographic influences produce contrasting patterns of prevalence, community composition and genetic structure in two genera of avian malaria parasites in southern Melanesia. *Journal of Animal Ecology*, 84, 985–998.

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

Ortego, J., Cordero, P.J., Aparicio, J.M. & Calabuig, G. (2008). Consequences of chronic infections with three different avian malaria lineages on reproductive performance of Lesser Kestrels (*Falco naumanni*). *Journal of Ornithology*, 149, 337–343.

Otero, L., Schall, J.J., Cruz, V., Aaltonen, K. & Acevedo, M.A. (2018). The drivers and consequences of unstable *Plasmodium* dynamics: A long-term study of three malaria species infecting a tropical lizard. *bioRxiv*, doi: https://doi.org/10.1101/189696.

Paaijmans, K.P., Read, A.F. & Thomas, M.B. (2009). Understanding the link between malaria risk and climate. *Proceedings of the National Academy of Sciences*, 106, 13844–13849.

Padilla, D.P., Illera, J.C., González-Quevedo, C., Villalba, M. & Richardson, D.S. (2017). Factors affecting the distribution of haemosporidian parasites within an oceanic island. *International Journal for Parasitology*, 47, 225–235.

Pérez-Rodríguez, A., Ramírez, Á., Richardson, D.S. & Pérez-Tris, J. (2013). Evolution of parasite island syndromes without long-term host population isolation: parasite dynamics in Macaronesian blackcaps *Sylvia atricapilla*. *Global Ecology and Biogeography*, 22, 1272–1281.

Piel, F.B., Patil, A.P., Howes, R.E., Nyangiri, O.A., Gething, P.W. & Williams, T.N. *et al.* (2010). Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nature Communications*, 1, 104.

Piertney, S.B. & Oliver, M.K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96, 7–21.

QGIS Development Team. (2017). QGIS Geographic Information System.

Raghwani, J., Rambaut, A., Holmes, E.C., Hang, V.T., Hien, T.T. & Farrar, J. *et al.* (2011). Endemic dengue associated with the co-circulation of multiple viral lineages and localized density-dependent transmission. *PLoS Pathogens*, 7, e1002064.

Ravenhall, M., Campino, S., Sepúlveda, N., Manjurano, A., Nadjm, B. & Mtove, G. *et al.* (2018). Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. *PLoS Genetics*, 14, e1007172.

Reinoso-Pérez, M.T., Canales-Delgadillo, J.C., Chapa-Vargas, L. & Riego-Ruiz, L. (2016). Haemosporidian parasite prevalence, parasitemia, and diversity in three resident bird

species at a shrubland dominated landscape of the Mexican highland plateau. *Parasites and Vectors*, 9, 307.

Richardson, D.S., Jury, F.L., 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, J.L., Urban, M.C., Bolnick, D.I. & Skelly, D.K. (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology and Evolution*, 29, 165–176.

Roiz, D., Ruiz, S., Soriguer, R. & Figuerola, J. (2015). Landscape effects on the presence, abundance and diversity of mosquitoes in mediterranean wetlands. *PLoS ONE*, 10, e0128112.

Rubio-Palis, Y. & Zimmerman, R.H. (1997). Ecoregional classification of malaria vectors in the neotropics. *Journal of Medical Entomology*, 34, 499–510.

Schmidt, P.S., Bertness, M.D. & Rand, D.M. (2000). Environmental heterogeneity and balancing selection in the acorn barnacle *Semibalanus balanoides*. *Proceedings of the Royal Society B: Biological Sciences*, 267, 379–384.

Schröder, N.W. & Schumann, R.R. (2005). Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *The Lancet Infectious Diseases*, 5, 156–164.

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

Signorell, A. (2018). DescTools: Tools for descriptive statistics. R package version 0.99.24. https://cran.r-project.org/package=DescTools.

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

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

Spurgin, L.G. & Richardson, D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277, 979–988.

Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A. & Richardson, D.S. (2014). Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Molecular Ecology*, 23, 1028–1039.

Spurgin, L.G., Illera, J.C., Padilla, D.P. & Richardson, D.S. (2012). Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia*, 168, 691–701.

Steffan-Dewenter, I., Münzenberg, U., Bürger, C., Thies, C. & Tscharntke, T. (2002). Scale-dependant effects of landscape context on three pollinator guilds. *Ecology*, 83, 1421–1432.

Tack, A.J.M., Horns, F. & Laine, A.L. (2014). The impact of spatial scale and habitat configuration on patterns of trait variation and local adaptation in a wild plant parasite. *Evolution*, 68, 176–189.

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

Thye, T., Vannberg, F.O., Wong, S.H., Owusu-Dabo, E., Osei, I. & Gyapong, J. *et al.* (2010). Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nature Genetics*, 42, 739–741.

Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G. & Destro-Bisol, G. *et al.* (2001). Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science*, 293, 455–462.

Tschirren, B., Andersson, M., Scherman, K., Westerdahl, H. & Råberg, L. (2011). 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., Andersson, M., Scherman, K., Westerdahl, H., Mittl, P.R.E. & Råberg, L. (2013). Polymorphisms at the innate immune receptor TLR2 are associated with *Borrelia* infection in a wild rodent population. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20130364.

Turcotte, A., Bélisle, M., Pelletier, F. & Garant, D. (2018). Environmental determinants of haemosporidian parasite prevalence in a declining population of Tree swallows. *Parasitology*, 145, 961–970.

Van Riper, C., Van Riper, S.G., Goff, L.M. & Laird, M. (1986). The epizootiology and ecological significance of malaria in Hawaiian land birds. *Ecological Monographs*, 56, 327–344.

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

Waldenström, J., Bensch, S., Hasselquist, D. & Östman, Ö. (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.

Wang, S., Liu, C., Wilson, A.B., Zhao, N., Li, X. & Zhu, W. *et al.* (2017). Pathogen richness and abundance predict patterns of adaptive major histocompatibility complex variation in insular amphibians. *Molecular Ecology*, 26, 4671–4685.

Warner, R.E. (1968). The role of introduced diseases in the extinction of the endemic Hawaiian avifauna. *The Condor*, 70, 101–120.

Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W. & Künstner, A. *et al.* (2010). The genome of a songbird. *Nature*, 464, 757–762.

Wenzel, M.A., Douglas, A., James, M.C., Redpath, S.M. & Piertney, S.B. (2016). The role of parasite-driven selection in shaping landscape genomic structure in red grouse (*Lagopus lagopus scotica*). *Molecular Ecology*, 25, 324–341.

Werling, D., Jann, O.C., Offord, V., Glass, E.J. & Coffey, T.J. (2009). Variation matters: TLR structure and species-specific pathogen recognition. *Trends in Immunology*, 30, 124–30.

Westerdahl, H., Stjernman, M., Råberg, L., Lannefors, M. & Nilsson, J.Å. (2013). MHC-I affects infection intensity but not infection status with a frequent avian malaria parasite in blue tits. *PLoS ONE*, 8, e72647.

Westerdahl, H., Waldenström, J., Hansson, B., Hasselquist, D., Schantz, T. von & Bensch, S. (2005). Associations between malaria and MHC genes in a migratory songbird. *Proceedings of the Royal Society B: Biological Sciences*, 272, 1511–1518.

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

Wong, S.H., Gochhait, S., Malhotra, D., Pettersson, F.H., Teo, Y.Y. & Khor, C.C. *et al.* (2010). Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathogens*, 6, e1000979.

World Health Organization. (2017). *World Malaria Report 2017*. WHO, Geneva.

Yanoviak, S.P., Paredes, J.E.R., Lounibos, L.P. & Weaver, S.C. (2006). Deforestation alters phytotelm habitat availability and mosquito production in the Peruvian Amazon. *Ecological Applications*, 16, 1854–1864.

Zuur, A.F., Ieno, E.N. & Elphick, C.S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1, 3–14.

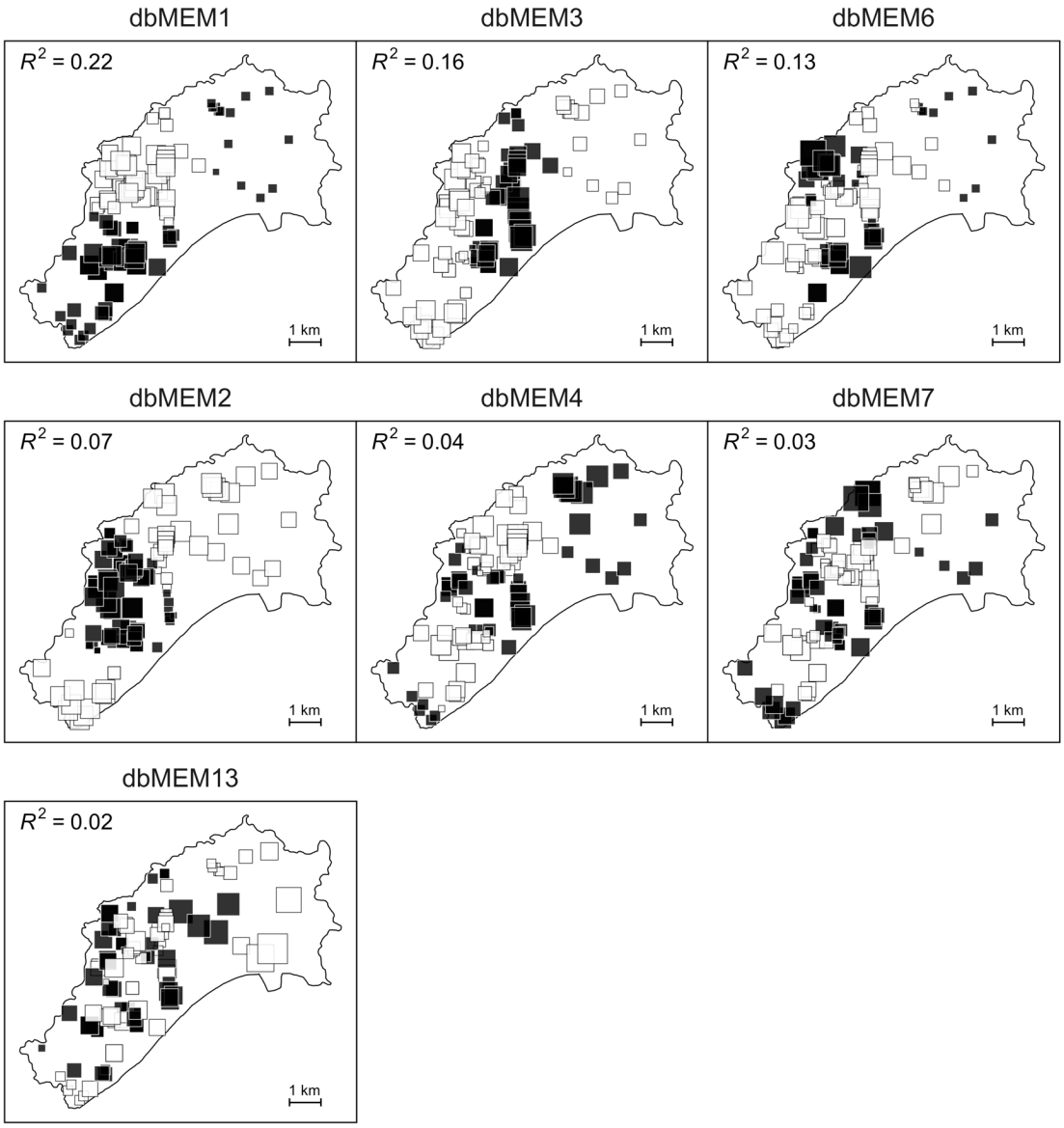
4.7. Supplementary Materials

Supplementary Table S4.1. Summary of general linear models of the association between genetic variants and malaria infection status in Berthelot's pipits on Porto Santo (PS) and Tenerife (TF). Parameter estimates (with standard error in brackets) for each genetic variant were taken from multipredictor models with genetic variants coded as SNP heterozygosity, SNP genotype, or TLR4 protein haplotype presence/absence. Models were performed in each island across all sampling years ('All years estimate'), or in the main sampling year with largest sample size (Porto Santo = 2016; Tenerife = 2011; 'Main year estimate'). Asterisks next to parameter estimates denote significance of the predictor (* $P < 0.05$).

Island	Variant type	Variant	All years estimate	Main year estimate
PS	SNP genotype	5239s1	0.75 (0.34)*	0.60 (0.37)
		TLR4_2	0.90 (0.51)	1.10 (0.58)
		TLR4_4	0.15 (0.40)	-0.16 (0.45)
	SNP heterozygosity	5239s1	0.53 (0.45)	0.61 (0.52)
		TLR4_2	0.94 (0.49)	0.95 (0.55)
		TLR4_4	-0.76 (0.47)	-0.98 (0.53)
	TLR4 protein haplotype	TLR4_P1	-1.62 (0.85)	-1.57 (0.94)
		TLR4_P2	-0.53 (0.52)	-0.82 (0.60)
		TLR4_P3	0.40 (0.51)	0.24 (0.59)
TF	SNP genotype	5239s1	0.08 (0.17)	0.18 (0.18)
		7259s1	-0.02 (0.14)	-0.03 (0.15)
		TLR4_3	-0.10 (0.22)	0.09 (0.26)
		TLR4_4	0.31 (0.25)	0.19 (0.29)
	SNP heterozygosity	5239s1	0.09 (0.21)	0.25 (0.24)
		7259s1	-0.29 (0.20)	-0.30 (0.22)
		TLR4_3	0.20 (0.25)	0.00 (0.29)
		TLR4_4	-0.41 (0.26)	-0.40 (0.29)
	TLR4 protein haplotype	TLR4_P1	-0.34 (0.20)	-0.46 (0.23)*
		TLR4_P2	-0.07 (0.38)	-0.20 (0.43)

Supplementary Table S4.2. Summary of linear models of the association between genetic variants and malaria risk in Berthelot's pipits on Porto Santo (PS) and Tenerife (TF). Parameter estimates (with standard error in brackets) for each genetic variant were taken from multipredictor models with genetic variants coded as SNP heterozygosity, SNP genotype, or TLR4 protein haplotype presence/absence. On Porto Santo, models were performed with just the genetic variants, or with the inclusion of dbMEMs to control for autocorrelation in model residuals ('dbMEM estimate'). dbMEMs were unable to account for autocorrelation in Tenerife models. Asterisks next to parameter estimates denote significance of the predictor (* $P < 0.05$; ** $P < 0.01$).

Island	Variant type	Variant	Estimate	dbMEM estimate
PS	SNP genotype	5239s1	0.69 (0.27)*	0.38 (0.17)*
		TLR4_2	1.03 (0.42)*	0.22 (0.27)
		TLR4_4	0.21 (0.36)	0.09 (0.23)
	SNP heterozygosity	5239s1	0.79 (0.41)	0.74 (0.25)**
		TLR4_2	0.77 (0.42)	0.39 (0.26)
		TLR4_4	-0.59 (0.42)	-0.36 (0.26)
	TLR4 protein haplotype	TLR4_P1	-1.32 (0.62)*	-0.41 (0.41)
		TLR4_P2	-0.36 (0.51)	0.02 (0.33)
		TLR4_P3	0.60 (0.47)	0.14 (0.30)
TF	SNP genotype	5239s1	0.26 (0.73)	
		7259s1	-0.23 (0.60)	
		TLR4_3	-0.03 (1.04)	
		TLR4_4	0.40 (1.13)	
	SNP heterozygosity	5239s1	0.80 (0.96)	
		7259s1	-1.08 (0.86)	
		TLR4_3	0.48 (1.13)	
		TLR4_4	-0.59 (1.14)	
	TLR4 protein haplotype	TLR4_P1	-0.45 (0.89)	
		TLR4_P2	0.22 (1.68)	



Supplementary Figure S4.1. Distance-based Moran's eigenvector maps (dbMEMs) showing strongest associations with *Plasmodium* strain LK6 infection risk in adult Berthelot's pipits on Porto Santo. R^2 values for each dbMEM were calculated from single-predictor LMs of each dbMEM against logit-transformed LK6 risk. Positive eigenvector scores are indicated by black squares, and negative scores are white. The size of the square indicates the magnitude of the score.

Chapter 5. General discussion

Berthelot's pipit. Photo by Philip Lamb

5.1. General discussion

Pathogens are a major selective agent, with the potential to reduce individual survival and fitness, and exert demographic change on populations (Daszak 2000; Morens *et al.* 2004). Although pathogen-mediated selection is predicted to promote high levels of diversity at immune genes (Bernatchez & Landry 2003; Piertney & Oliver 2006; Spurgin & Richardson 2010), this variation is at risk of erosion through genetic drift, especially in bottlenecked populations (Acevedo-Whitehouse & Cunningham 2006). In my thesis I have investigated patterns of immunogenetic diversity at the level of individuals, populations and archipelagos in Berthelot's pipit *Anthus berthelotii*, an island bird which has experienced reduced population sizes and loss of genetic diversity associated with colonisation events, to determine the extent to which pathogen-mediated selection is able to maintain adaptive variation in the face of genetic drift.

Establishing a baseline level of neutral variation was an important first step for determining the evolutionary dynamics of immune genes. In Chapter 2, I used restriction-site associated DNA sequencing (RAD-seq) to assess genome-wide levels of genetic diversity in Berthelot's pipit archipelagos, and in the mainland sister species, the tawny pipit *A. campestris*. I found sharp declines in genetic diversity in Berthelot's pipits in the Canary Islands relative to tawny pipits, followed by further reductions in Madeira and Selvagens, in line with past inferences of the colonisation history of Berthelot's pipits (Arctanderi *et al.* 1991; Illera *et al.* 2007; Spurgin *et al.* 2014). Admixture analysis revealed previously undetected population structuring within archipelagos, with an east–west gradient in the Canary Islands, and divergence within the Madeiran archipelago.

One approach to investigating the role of pathogens in shaping genetic diversity is to study candidate genes, which have known or predicted functions of relevance to the immune response against pathogens, to determine the evolutionary forces driving adaptive variation. In Chapter 3, I focused on avian β-defensins (AvBDs) as immune gene candidates. AvBDs are a key component of the innate immune system, providing broad-scale defence against microbial pathogens (Zasloff 2002). The AvBD mature peptide region destroys pathogens through membrane damage, with polymorphisms in this region having a direct effect on host antimicrobial defence (Hellgren *et al.* 2010; Cadwell *et al.* 2017; Schmitt *et al.* 2017). Preliminary work on AvBDs in Berthelot's pipits had revealed polymorphisms at five AvBD loci (González-Quevedo, unpublished data). I designed primers to cover the entire mature peptide region of these AvBDs, and sequenced individuals across Berthelot's pipit archipelagos, as well as in tawny pipits. For the most part, the highest variation was found in tawny pipits, followed by the Canary Islands, with the lowest in Madeira and Selvagens, as expected given the demographic history inferred previously (Chapter 2). Using the RAD data generated in Chapter 2, I compared

nucleotide diversity at AvBDs and genome-wide RAD loci. This revealed that, despite losses of AvBD variation associated with colonisation history, AvBDs often ranked highly in terms of nucleotide diversity relative to RAD loci. However, AvBD2 was the only locus to maintain nonsynonymous variation in Madeira and Selvagens, suggesting that these populations may have a reduced breadth of protection against microbial pathogens relative to the Canary Islands. I did not detect signatures of balancing selection at AvBD2 in Berthelot's pipits, although this may be under balancing selection in tawny pipits. Nucleotide diversity at AvBD2 was particularly high relative to RAD loci in tawny pipits and in the Canary Islands; it is possible that this locus has maintained nonsynonymous variation in Madeira and Selvagens by chance, owing to initial high diversity prior to colonisation, akin to the 'ghost of selection past' observed in toll-like receptors (TLRs) in the bottlenecked Seychelles warbler *Acrocephalus sechellensis* (Gilroy *et al.* 2017).

In contrast to utilising existing candidate genes as in Chapter 3, in the second part of Chapter 2 I instead used genome-wide polymorphisms to identify potential novel candidates associated with pathogen infection. Only focusing on a small number of loci, such as the MHC or TLR loci, ignores most of the huge range of genes that play a role within the immune system, and may miss important determinants of resistance and susceptibility. Using a genome-wide association study, I tested for genetic variants that showed strong statistical associations with malaria infection in Berthelot's pipits across the Canary Islands. This revealed five outlier SNPs, with the top two situated within or near potential candidate genes. The strongest association was found at SNP 5239s1, situated ca. 2000 bp away from interleukin-16 (IL-16), a cytokine which is expressed as part of the immune response. Serum levels of proinflammatory cytokines such as IL-1β, IL-6 and tumour necrosis factor alpha (TNF-α) have been implicated in malaria severity, due to their modulation of the inflammatory response (Kern *et al.* 1989; Day *et al.* 1999; Lyke *et al.* 2004). The presence of IL-16 triggers the expression of these cytokines in a dose-dependent manner (Mathy *et al.* 2000); it is therefore plausible that polymorphisms that modify the expression rate of IL-16 could be influencing the inflammatory response to malaria infection in Berthelot's pipits. The second strongest association was found at SNP 7259s1, situated within RIMS4, which was shown to be overexpressed in breast cancer cells (Abba *et al.* 2005). As both genes have involvements in the immune system, these may be useful novel candidates for the study of malaria resistance.

Identifying divergent selection between populations at immune genes could indicate the presence of pathogen-mediated balancing selection, due to local adaptation for different pathogen communities (Hill 1991), or rare-allele advantage, with different alleles providing a selective advantage in each population (Slade & McCallum 1992). I performed a genome scan to

test for divergent selection between Berthelot's pipit archipelagos using RAD SNPs (Chapter 2), though this did not reveal signatures of divergent selection at the malaria-associated SNPs identified in the GWAS. Allele frequencies at SNP 7259s1 appeared to follow a pattern of drift (Fig. 2.3), with low minor allele frequency in Madeira and monomorphic in Selvagens. Conversely, SNP 5239s1 had a higher minor allele frequency in Madeira than Canary Islands, against neutral expectations. This could still be due to drift, as the random change of allele frequencies could act to increase the frequency of an allele against the general pattern caused by demographic effects, but could also potentially indicate weak divergent selection. Including the AvBD SNPs into the RAD dataset for a second genome scan (Chapter 3) also found no evidence of divergent selection at AvBD loci. I had already found AvBD variation to be limited in Madeira and Selvagens, suggesting that evolution at these loci is largely shaped by founder effects and genetic drift.

To further investigate the evolutionary dynamics of the malaria-associated SNPs, I used a landscape genetics approach to explore the genetic associations with malaria risk using Porto Santo (Madeira) and Tenerife (Canary Islands) as replicate populations (Chapter 4). Malaria distributions are strongly related to the environmental suitability for their invertebrate vectors (Craig *et al.* 1999; Ferraguti *et al.* 2016; Illera *et al.* 2017). By identifying the environmental risk factors that predict malaria infection, it was possible to calculate a measure of malaria risk at each sampling location, and determine whether genetic variants were associated with areas of high or low risk. A previous study found that malaria risk in Berthelot's pipits on Tenerife was mainly shaped by temperature, distance to artificial water sources, and distance to poultry farms (González-Quevedo *et al.* 2014). Genetic variation at the MHC was also linked to malaria risk and distance to poultry farms, suggesting the presence of fine-scale adaptation at the landscape scale (González-Quevedo *et al.* 2016). For Chapter 4, Porto Santo was chosen as a replicate population as this island had the highest rates of malaria found across Berthelot's pipit populations (Spurgin *et al.* 2012). In addition, with a lack of gene flow between the Canary Islands and Madeira (Chapter 2), it allowed us to explore the repeatability of genetic adaptations to disease in isolated populations.

I tested for associations with malaria infection and risk across Tenerife and Porto Santo in the GWAS malaria SNPs identified in Chapter 2, in addition to four SNPs in TLR4. Codons at this locus were previously found to be under positive selection, with unique haplotypes found in Madeira (González-Quevedo *et al.* 2015). In sequencing > 300 additional samples from Tenerife, I did not find any instances of the Madeiran TLR4 haplotypes (Chapter 4). This suggests that these alleles may indeed be missing from the Canary Islands, although it is possible that they are present in islands outside of Tenerife. I had previously found a large degree of population admixture among

the majority of populations within the Canary Islands (Chapter 2); along with the large size and central location of Tenerife, sharing of TLR4 haplotypes from neighbouring islands may be expected. Two scenarios are therefore plausible: either the additional haplotypes were initially present in the Canary Islands, and have since been lost through drift or purifying selection, or they originated in Madeira following colonisation. In contrast to variation at TLR4, I found no evidence of unique haplotypes at AvBD loci in the Madeiran archipelago (Chapter 3).

The different patterns observed for TLR4 and AvBDs across Berthelot's pipit populations may be indicative of fundamental differences in function between these genes. TLR proteins are pathogen recognition receptors, detecting molecular structures associated with pathogens and triggering an immune response (Janeway & Medzhitov 2002). Polymorphisms in TLRs are associated with susceptibility and resistance to pathogens (Schröder & Schumann 2005; Misch & Hawn 2008; Basu *et al.* 2010), suggesting that having a diverse set of TLR genes is likely to increase the breadth of pathogens an organism is able to recognise and respond to. Biogeographical processes can result in divergent pathogen communities between islands (Fallon *et al.* 2005; Olsson-Pons *et al.* 2015; Wang *et al.* 2017). During the colonisation of the Madeiran archipelago, Berthelot's pipits may have faced new pathogen assemblages. This could have selected for the maintenance of TLR4 haplotypes that have since been lost in the Canary Islands. Alternatively, if existing haplotypes that were adapted to pathogen regimes in the Canary Islands were less suited to Madeiran pathogens, new mutations arising after colonisation could rise in frequency through positive selection. By contrast, the antimicrobial activity of AvBDs is reliant on interruption of the phospholipid membrane, a highly conserved feature of microbial pathogens (Yeaman & Yount 2003). Selection to retain AvBD haplotypes could therefore have been weaker than on TLR loci, as there might potentially be greater overlap in pathogen defence between different AvBD haplotypes than provided by TLR haplotypes.

On Porto Santo, I found some similarities in the environmental predictors of malaria risk with those identified in Tenerife (Chapter 4; González-Quevedo *et al.* 2014). Altitude (which was strongly correlated with temperature) and distance to water sources were both strong predictors of malaria, mirroring the findings in Tenerife. In addition, pipit density, and its interaction with distance to water, played a role, as did the vegetation type (Fig. 4.2). Together, these predictors were used to calculate malaria risk at each sampling location (Fig. 4.3). I did not find associations between malaria infection or risk on Tenerife for either of the malaria SNPs, whereas on Porto Santo, the opposite association was found at 5239s1 compared with in the Canary Islands GWAS (Chapter 2). Where the AA genotype was associated with the highest rates of malaria infection across the Canary Islands, on Porto Santo this was linked to reduced malaria infection and risk. Rare-allele advantage, local adaptation, or a combination of the two, could

potentially be driving divergent associations with malaria between the Canary Islands and Porto Santo.

The findings in Chapter 4 emphasise the importance of considering environmental factors when investigating the genetic adaptations to infectious disease. In contrast to the high rates of malaria found on Porto Santo, no infected Berthelot's pipits were found on the neighbouring islands of Madeira and Deserta Grande (Spurgin *et al.* 2012). Without knowledge of the environmental drivers of malaria risk, it might be assumed that individuals on Porto Santo are highly susceptible to malaria, whereas the other populations have evolved resistance. However, knowing that altitude and artificial water sources are key predictors of infection in Porto Santo, it is likely that the other islands are not conducive to supporting malaria populations. Deserta Grande is uninhabited, lacking in artificial water sources, while most of Madeira is at or above altitudes that would be associated with very low risk on Porto Santo. No evidence of malaria parasites was found on Madeira in a study of spectacled warblers *Sylvia conspicillata* (Illera *et al.* 2015), also suggesting that a lack of malaria in Madeiran Berthelot's pipits is not due to genetic differences between Madeiran and Porto Santo pipits.

Environmental differences between islands may have obscured genetic associations with malaria in the GWAS analysis (Chapter 2). An ideal GWAS study would use individuals from the same population, to reduce spurious effects of population structure (Marchini *et al.* 2004). In the GWAS of malaria infection, to gain a suitably large sample size, I used individuals from all populations across the Canary Islands. F_{ST} analysis demonstrated that population structuring was generally low across this archipelago (Chapter 2), though the use of a relatedness matrix in the GWAS analysis was used to control for structuring. However, this approach did not account for environmental differences between sampling locations, both within and between islands. This could potentially obscure associations: individuals might harbour highly susceptible alleles, but be situated in an area with little actual risk of encountering malaria parasites. On islands such as La Palma and El Hierro, where malaria was found at very low frequency in Berthelot's pipits (Spurgin *et al.* 2012), and absent in spectacled warblers (Illera *et al.* 2015), with such low selection pressure to evolve resistance against malaria, alleles that would otherwise be selected against could rise in frequency through genetic drift, or through selection against a different pathogen. Likewise, chance fluctuations in allele frequencies could inflate an association with malaria, causing spurious results.

The findings of this thesis have identified a number of avenues for future research. Having detected novel SNPs associated with malaria infection, determining the causal factor is an important next step. Though I have identified the genes IL-16 and RIMS4 as being in closest proximity to the SNPs, this does not guarantee that those genes are driving the association. In a

GWAS analysis performed on whole-genome sequencing data, a peak of association would be expected over the causal gene, with associations dropping away to either side as linkage disequilibrium decays. With the low marker density present in the RAD dataset, it is possible that the true, undetected peak would be some distance from the outlier SNPs detected in Chapter 2, situated over different genes. Utilising whole-genome sequencing methods would allow for a more in-depth view of the genetic basis of adaptive traits, and highlight genomic regions to sequence throughout Berthelot's pipit populations. Likewise, the genome scans detected signatures of divergent selection in Berthelot's pipits; having a more complete dataset of SNPs would help to identify possible candidates that are undergoing adaptive divergence.

At present, malaria strains in Berthelot's pipits have been characterised according to mitochondrial DNA (mtDNA; Chapter 2 & 4). However, this disregards adaptive evolution within the nuclear genome. Sequencing malaria strains by functional loci which may affect infection ability would provide more complete information on the diversity of pathogens faced by Berthelot's pipits. By grouping pathogens by mtDNA only, patterns of local adaptation, resistance and susceptibility may be obscured, if divergence within a mtDNA strain is also driving divergent selection on pipit immune loci. A nested PCR approach akin to mtDNA sequencing of malaria strains was used to sequence merozoite surface protein 1, a gene related to invasion biology (Hellgren *et al.* 2015). This could be easily implemented into future studies of malaria in Berthelot's pipits, and may aid in detecting the genes under selection for divergent malaria strains.

In Chapter 3, I found potential indications that AvBD2 might be experiencing positive selection, or that the variation found at this locus was due to past balancing selection, preventing the diversity at this locus from being fully eroded by bottlenecks and drift. Without information of the microbial pathogens faced by Berthelot's pipits, we are currently unaware of the selection pressures upon AvBDs: are pathogen regimes consistent between islands, or do we see reduced variation in smaller, more isolated islands? DNA barcoding can be utilised to detect gut microbiota (David *et al.* 2014), which may be a useful way to characterise microbial infections and determine whether different AvBD haplotypes are associated with the presence or absence of specific gut pathogens.

Berthelot's pipit represents an interesting system for investigating the role of pathogen-mediated selection for maintaining genetic variation through successive colonisation events. This thesis has identified novel candidates for disease resistance and divergent selection, characterised genetic variation and the role of selection and drift in AvBDs, and assessed fine-scale patterns of genetic adaptation to malaria risk. These findings emphasise the key roles

demography and environment play in modulating the evolutionary response to pathogen infection.

5.2. References

Abba, M.C., Hu, Y., Sun, H., Drake, J.A., Gaddis, S. & Baggerly, K. *et al.* (2005). Gene expression signature of estrogen receptor alpha status in breast cancer. *BMC Genomics*, 6, 37.

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

Arctanderi, P., Folmerl, O. & Fjeldså, J. (1991). 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.

Basu, M., Maji, A.K., Chakraborty, A., Banerjee, R., Mullick, S. & Saha, P. *et al.* (2010). Genetic association of Toll-like-receptor 4 and tumor necrosis factor-α polymorphisms with *Plasmodium falciparum* blood infection levels. *Infection, Genetics and Evolution*, 10, 686–696.

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.

Cadwell, K., Niranji, S.S., Armstrong, V.L., Mowbray, C.A., Bailey, R. & Watson, K.A. *et al.* (2017). AvBD1 nucleotide polymorphisms, peptide antimicrobial activities and microbial colonisation of the broiler chicken gut. *BMC Genomics*, 18, 637.

Craig, M.H., Le Sueur, D. & Snow, R.W. (1999). A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today*, 15, 105–111.

Daszak, P. (2000). Emerging infectious diseases of wildlife – threats to biodiversity and human health. *Science*, 287, 443–449.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E. & Wolfe, B.E. *et al.* (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505, 559–563.

Day, N.P.J., Hien, T.T., Schollaardt, T., Loc, P.P., Chuong, L.V. & Chau, T.T.H. *et al.* (1999). The prognostic and pathophysiologic role of pro- and antiinflammatory cytokines in severe malaria. *The Journal of Infectious Diseases*, 180, 1288–1297.

Fallon, S.M., Bermingham, E. & Ricklefs, R.E. (2005). Host specialization and geographic localization of avian malaria parasites: a regional analysis in the Lesser Antilles. *The American Naturalist*, 165, 466–480.

Ferraguti, M., Martínez-De La Puente, J., Roiz, D., Ruiz, S., Soriguer, R. & Figuerola, J. (2016). Effects of landscape anthropization on mosquito community composition and abundance. *Scientific Reports*, 6, 29002.

Gilroy, D.L., Phillips, K.P., Richardson, D.S. & Oosterhout, C. van. (2017). Toll-like receptor variation in the bottlenecked population of the Seychelles warbler: computer simulations see the 'ghost of selection past' and quantify the 'drift debt'. *Journal of Evolutionary Biology*, 30, 1276–1287.

González-Quevedo, C., Davies, R.G. & Richardson, D.S. (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.

González-Quevedo, C., Davies, R.G., Phillips, K.P., Spurgin, L.G. & Richardson, D.S. (2016). Landscape-scale variation in an anthropogenic factor shapes immune gene variation within a wild population. *Molecular Ecology*, 25, 4234–4246.

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

Hellgren, O., Atkinson, C.T., Bensch, S., Albayrak, T., Dimitrov, D. & Ewen, J.G. *et al.* (2015). Global phylogeography of the avian malaria pathogen *Plasmodium relictum* based on MSP1 allelic diversity. *Ecography*, 38, 842–850.

Hellgren, O., Sheldon, B.C. & Buckling, A. (2010). *In vitro* tests of natural allelic variation of innate immune genes (avian β-defensins) reveal functional differences in microbial inhibition. *Journal of Evolutionary Biology*, 23, 2726–2730.

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

Illera, J.C., Emerson, B.C. & Richardson, D.S. (2007). Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology*, 16, 4599–4612.

Illera, J.C., Fernández-Álvarez, Á., Hernández-Flores, C.N. & Foronda, P. (2015). Unforeseen biogeographical patterns in a multiple parasite system in Macaronesia. *Journal of Biogeography*, 42, 1858–1870.

Illera, J.C., López, G., García-Padilla, L. & Moreno, Á. (2017). Factors governing the prevalence and richness of avian haemosporidian communities within and between temperate mountains. *PLoS ONE*, 12, e0184587.

Janeway, C.A. & Medzhitov, R. (2002). Innate immune recognition. *Annual Review of Immunology*, 20, 197–216.

Kern, P., Hemmer, C.J., Damme, J.V., Gruss, H.-J. & Dietrich, M. (1989). Elevated tumor necrosis factor alpha and interleukin-6 serum levels as markers for complicated *Plasmodium falciparum* malaria. *The American Journal of Medicine*, 87, 139–143.

Lyke, K.E., Burges, R., Cissoko, Y., Sangare, L., Dao, M. & Diarra, I. *et al.* (2004). Serum levels of the proinflammatory cytokines Interleukin-1 Beta (IL-1β), IL-6, IL-8, IL-10, Tumor Necrosis Factor Alpha, and IL-12 (p70) in Malian children with severe *Plasmodium falciparum* malaria and matched uncomplicated malaria or healthy cont. *Infection and Immunity*, 72, 5630–5637.

Marchini, J., Cardon, L.R., Phillips, M.S. & Donnelly, P. (2004). The effects of human population structure on large genetic association studies. *Nature Genetics*, 36, 512–517.

Mathy, N.L., Scheuer, W., Lanzendörfer, M., Honold, K., Ambrosius, D. & Norley, S. *et al.* (2000). Interleukin-16 stimulates the expression and production of pro-inflammatory cytokines by human monocytes. *Immunology*, 100, 63–69.

Misch, E.A. & Hawn, T.R. (2008). Toll-like receptor polymorphisms and susceptibility to human disease. *Clinical Science*, 114, 347–360.

Morens, D.M., Folkers, G.K. & Fauci, A.S. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature*, 430, 242–249.

Olsson-Pons, S., Clark, N.J., Ishtiaq, F. & Clegg, S.M. (2015). Differences in host species relationships and biogeographic influences produce contrasting patterns of prevalence, community composition and genetic structure in two genera of avian malaria parasites in southern Melanesia. *Journal of Animal Ecology*, 84, 985–998.

Piertney, S.B. & Oliver, M.K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96, 7–21.

Schmitt, C., Garant, D., Bélisle, M. & Pelletier, F. (2017). Linking innate immunogenetic variation with phenotypic traits in a wild population of tree swallows, *Tachycineta bicolor*. *Biological Journal of the Linnean Society*, 121, 685–697.

Schröder, N.W. & Schumann, R.R. (2005). Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *The Lancet Infectious Diseases*, 5, 156–164.

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

Spurgin, L.G. & Richardson, D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277, 979–988.

Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A. & Richardson, D.S. (2014). Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Molecular Ecology*, 23, 1028–1039.

Spurgin, L.G., Illera, J.C., Padilla, D.P. & Richardson, D.S. (2012). Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia*, 168, 691–701.

Wang, S., Liu, C., Wilson, A.B., Zhao, N., Li, X. & Zhu, W. *et al.* (2017). Pathogen richness and abundance predict patterns of adaptive major histocompatibility complex variation in insular amphibians. *Molecular Ecology*, 26, 4671–4685.

Yeaman, M.R. & Yount, N.Y. (2003). Mechanisms of antimicrobial peptide action and resistance. *Pharmacological Reviews*, 55, 27–55.

Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, 415, 389–395.